Miguel Hernandez University |

Universidad Miguel Hernández

Cyber Places, Crime Patterns, and Cybercrime Prevention: An Environmental Criminology and Crime Analysis approach through Data Science

A doctoral thesis submitted to the Miguel Hernandez University in partial fulfilment of the requirements for the degree of International Doctor of Philosophy in Criminology in the Faculty of Social and Legal Sciences

Asier Moneva

Director: Fernando Miró-Llinares

Doctoral Program in Criminology |

Programa de Doctorado en Criminología

Elche, June 2020

—Blank page—

This doctoral thesis consists of a compendium of the following previously published articles:

Moneva, A., & Caneppele, S. (2019). 100% sure bets? Exploring the precipitation-control strategies of fixed-match informing websites and the environmental features of their networks. *Crime, Law and Social Change*. https://doi.org/10.1007/s10611-019-09871-4

Moneva, A., Miró-Llinares, F., & Hart, T. C. (2020). Hunter or Prey? Exploring the Situational Profiles that Define Repeated Online Harassment Victims and Offenders. *Deviant Behavior*. https://doi.org/10.1080/01639625.2020.1746135

Miró-Llinares, F., Moneva, A., & Esteve, M. (2018). Hate is in the air! But where? Introducing an algorithm to detect hate speech in digital microenvironments. *Crime Science*, *7*(15), 1-12. https://doi.org/10.1186/s40163-018-0089-1

In addition, the following article submitted for publication is included:

Moneva, A., Leukfeldt, E. R., Van de Weijer, S. G. A., & Miró-Llinares, F. (submitted). Repeat victimization by website defacement: A test of Environmental Criminology premises for cybercrime. *Computers in Human Behavior*.

Other publications related to the doctoral thesis are listed below:

Miró-Llinares, F., & Moneva, A. (2019a) Environmental Criminology and Cybercrime: Shifting Focus from the Wine to the Bottles. In T. J. Holt & A. Bossler (Eds.), *The Palgrave Handbook of International Cybercrime and Cyberdeviance* (pp. 1–22). Springer International Publishing. https://doi.org/10.1007/978-3-319-90307-1_30-1

Miró-Llinares, F., & Moneva, A. (2019b). What about cyberspace (and cybercrime alongside it)? A reply to Farrell and Birks "Did cybercrime cause the crime drop?". *Crime Science, 8*(12), 1-5. https://doi.org/10.1186/s40163-019-0107-y

UNIVERSITAS
Miguel Hernández

D. Fernando Miró Llinares, en calidad de director de la presente tesis doctoral,

INFORMO

de que doy mi conformidad para la lectura y defensa de la tesis doctoral *Cyber Places, Crime Patterns, and Cybercrime Prevention: An Environmental Criminology and Crime Analysis approach through Data Science,* presentada por D. Asier Moneva en la modalidad por compendio de publicaciones, regulada en el artículo 18 de la Normativa de Estudios de Doctorado de la Universidad Miguel Hernández, y dirigida por el Dr. Fernando Miró Llinares, y la considero conforme en cuanto a forma y contenido para que sea presentada para su correspondiente exposición pública con el fin de optar al grado de Doctor.

Y para que conste a los efectos oportunos, firmo el presente informe a 11 de junio de 2020.

FERNANDO|MIRO|LLINARES

Firmado digitalmente por FERNANDO|MIRO|LLINARES
Fecha: 2020.06.14 11:31:50 +02'00'

Fdo.: Fernando Miró Llinares

—Blank page—

UNIVERSITAS
Miguel Hernández

D. Fernando Miró Llinares, en calidad de Coordinador del Programa de Doctorado en Criminología de la Universidad Miguel Hernández de Elche,

INFORMO

de que doy mi conformidad para la lectura y defensa de la tesis doctoral *Cyber Places, Crime Patterns, and Cybercrime Prevention: An Environmental Criminology and Crime Analysis approach through Data Science*, presentada por D. Asier Moneva en la modalidad por compendio de publicaciones, regulada en el artículo 18 de la Normativa de Estudios de Doctorado de la Universidad Miguel Hernández, y dirigida por el Dr. Fernando Miró Llinares, y la considero conforme en cuanto a forma y contenido para que sea presentada para su correspondiente exposición pública con el fin de optar al grado de Doctor.

Y para que conste a los efectos oportunos, firmo el presente informe a 11 de junio de 2020.

FERNANDO|MIRO|
LLINARES

Firmado digitalmente por FERNANDO|MIRO|LLINARES
Fecha: 2020.06.14
11:23:09 +02'00'

Fdo.: Fernando Miró Llinares

—Blank page—

FUNDING INFORMATION

This thesis was funded by the Spanish Ministry of Education, Culture and Sports, and the Spanish Ministry of Science, Innovation and Universities under Grant FPU16/01671 for the training of university staff; and under Grant EST18/00043 for brief research stays.

The research in CHAPTER VI was supported by the Spanish Ministry of Science, Innovation and Universities under Grant FPU16/01671; and under Grant EST18/00043.

The research in CHAPTER VII was supported by the Spanish Ministry of Science, Innovation and Universities under Grant FPU16/01671.

The research in CHAPTER VIII was supported by the Spanish Ministry of Science, Innovation and Universities under Grant FPU16/01671; and by the Governing Council of Castile-Leon under Grant A2017/009386.

The research in CHAPTER IX was supported by the Spanish Ministry of Education, Culture and Sports under Grant FPU16/01671; and by European Union's Horizon 2020 research and innovation programme under Grant Agreement No. 740773.

—Blank page—

DEDICATORIA | DEDICATION

A mis padres, Begoña y Juanjo, por confiar siempre en mis capacidades más que yo mismo. Y a mi familia, por ser el sólido pilar sobre el que se asientan mis valores. Sin vosotros nada de esto hubiera sido posible.

También a ti, Cristina, por representar la cualidad de lo incondicional en los momentos más difíciles.

—Blank page—

AGRADECIMIENTOS | ACKNOWLEDGEMENTS

A Fernando Miró-Llinares, por darme mil y una oportunidades para crecer. Con ninguna otra persona podría haber aprendido lo que aprendí contigo.

A José E. Medina-Sarmiento, por ser mi primer mentor. Siempre tendrás mi más profunda admiración.

A todos mis compañeros de Crímina, por estar al pie del cañón cada día. Sois un ejemplo de tesón. De entre ellos, F. Javier Castro-Toledo, Ana B. Gómez- Bellvís, y David Buil-Gil merecen una mención especial por haber sido capaces de soportarme todo este tiempo. También Zoraida Esteve-Bañón, una de las pioneras, por enseñarme el arte de la docencia universitaria, y M. Mar Ruiz-Oliver, por hacerme la vida más fácil. Nunca podré agradeceros lo suficiente vuestra empatía, ni vuestra amistad.

To all my co-authors, from whom I have learned so much. Thanks to E. Rutger Leukfeldt, Steve van de Weijer, Fernando Miró-Llinares, Stefano Caneppele, Miriam Esteve, and Steven Kemp. My special thanks to Timothy C. Hart, with whom I published my first scientific paper. From you I learned the true meaning of co-authorship. You all were a fundamental piece in this puzzle. I honestly hope to continue working with you in the future.

To Marcelo F. Aebi and the School of Criminal Sciences at University of Lausanne, and to E. Rutger Leukfeldt and the Netherlands Institute for the Study of Crime and Law Enforcement for their hospitality during my research stays. It was a pleasure working with you.

—Blank page—

INDEX

—Blank page—

LIST OF ABBREVIATIONS

Acronym	Meaning
5Is	Intelligence, Intervention, Implementation, Involvement, Impact
API	Application Programming Interface
CACC	Conjunctive Analysis of Case Configurations
CPTED	Crime Prevention Through Environmental Design
CRAN	Comprehensive R Archive Network
CRAVED	Concealable, Removable, Available, Valuable, Enjoyable, Disposable
CSEW	Crime Survey for England and Wales
ECCA	Environmental Criminology and Crime Analysis
FMIWs	Fixed-Match Informing Websites
GH	General Hypothesis
H	Hypothesis
HNA	Hyperlink Network Analysis
HTML	Hypertext Markup Language
IP	Internet Protocol
IT	Information Technology
IVI	Introduction, Value, Interaction
JSON	JavaScript Object Notation
LISS	Longitudinal Internet Studies for the Social sciences
OS	Operating System
PDF	Portable Document Format
RQ	Research Question
SARA	Scanning, Analysis, Response, Assessment
SCAREM	Stealth, Challenge, Anonymity, Reconnaissance, Escape, Multiplicity
SCP	Situational Crime Prevention
SNA	Social Network Analysis
SVM	Support Vector Machines
TOR	The Onion Router
URL	Uniform Resource Locator
VIVA	Value, Inertia, Visibility, Accessibility
VOLTAGE	Victims, Offenders, Locations, Times, Attractors, Groups, Enhancers
VPN	Virtual Private Network

—Blank page—

ABSTRACT

For years, academics have examined the potential usefulness of traditional criminological theories to explain and prevent cybercrime. Some analytical frameworks from Environmental Criminology and Crime Analysis (ECCA), such as the Routine Activities Approach and Situational Crime Prevention, are frequently used in theoretical and empirical research for this purpose. These efforts have led to a better understanding of how crime opportunities are generated in cyberspace, thus contributing to advancing the discipline. However, with a few exceptions, other ECCA analytical frameworks —especially those based on the idea of geographical place— have been largely ignored. The limited attention devoted to ECCA from a global perspective means its true potential to prevent cybercrime has remained unknown to date. In this thesis we aim to overcome this geographical gap in order to show the potential of some of the essential concepts that underpin the ECCA approach, such as places and crime patterns, to analyse and prevent four crimes committed in cyberspace. To this end, this dissertation is structured in two phases: firstly, a proposal for the transposition of ECCA's fundamental propositions to cyberspace; and secondly, deriving from this approach some hypotheses are contrasted in four empirical studies through Data Science. The first study contrasts a number of premises of repeat victimization in a sample of more than nine million self-reported website defacements. The second examines the precipitators of crime at cyber places where allegedly fixed match results are advertised and the hyperlinked network they form. The third explores the situational contexts where repeated online harassment occurs among a sample of non-university students. And the fourth builds two metadata-driven machine learning models to detect online hate speech in a sample of Twitter messages collected after a terrorist attack. General results show (1) that cybercrimes are not randomly distributed in space, time, or among people; and (2) that the environmental features of the cyber places where they occur determine the emergence of crime opportunities. Overall, we conclude that the ECCA approach and, in particular, its place-based analytical frameworks can also be valid for analysing and preventing crime in cyberspace. We anticipate that this work can guide future research in this area including: the design of secure online environments, the allocation of preventive resources to high-risk cyber places, and the implementation of new evidence-based situational prevention measures.

KEYWORDS

crime analysis, crime patterns, crime prevention, cyber place, cybercrime, Data Science, Environmental Criminology

RESUMEN

Durante años, los académicos han examinado la potencial utilidad de las teorías criminológicas tradicionales para tratar de explicar y prevenir el cibercrimen. Algunos marcos analíticos de la Criminología Ambiental y el Análisis Delictivo (ECCA), como el Enfoque de las Actividades Cotidianas y las Prevención Situacional del Crimen, se han utilizado frecuentemente en investigaciones teóricas y empíricas con este fin. Estos trabajos han permitido mejorar nuestra comprensión sobre cómo se generan las oportunidades delictivas en el ciberespacio, contribuyendo así al avance de la disciplina. Sin embargo, salvo contadas excepciones, el resto de los marcos analíticos de ECCA —especialmente aquellos basados en la idea de lugar geográfico— han sido ampliamente ignorados. La escasa atención prestada al enfoque desde una perspectiva global ha causado que todavía hoy se desconozca su verdadero potencial para prevenir el cibercrimen. En esta tesis tratamos de superar esta barrera geográfica para mostrar el potencial de algunos de los conceptos esenciales que vertebran el enfoque de ECCA, como los lugares y los patrones delictivos, para analizar y prevenir cuatro crímenes que se cometen en el ciberespacio. Para ello, esta disertación se estructura en dos fases: una primera, en la que se propone la transposición de las proposiciones fundamentales de ECCA al ciberespacio; y una segunda, en la que se derivan algunas hipótesis de este enfoque y se contrastan mediante la realización de cuatro estudios empíricos a través de la Ciencia de Datos. En el primer estudio se analizan una serie de premisas sobre victimización repetida en una muestra de más de nueve millones de desfiguraciones web auto reveladas. En el segundo, se examinan los precipitadores del crimen en los ciber lugares donde se ofertan resultados de partidos supuestamente amañados y la red de hipervínculos que conforman. En el tercero, se exploran los contextos situacionales donde ocurre el acoso en línea repetido en una muestra de estudiantes de enseñanzas no universitarias. Y, en el cuarto, se construyen dos modelos de aprendizaje automático basados en metadatos para detectar discurso de odio en línea en una muestra de mensajes de Twitter recogida tras un atentado terrorista. Los resultados generales muestran (1) que los cibercrímenes no se distribuyen aleatoriamente en el espacio, en el tiempo, ni entre las personas; y (2) que los elementos ambientales de los ciber lugares donde acontecen determinan la aparición de oportunidades delictivas. En conjunto, concluimos que el enfoque de ECCA y, en particular, sus marcos analíticos basados en lugares también pueden ser válidos para analizar y prevenir el crimen en el ciberespacio. Anticipamos que este trabajo puede guiar investigaciones futuras en este ámbito como: el diseño de entornos seguros en línea, la concentración de recursos preventivos en ciber lugares de alto riesgo, o la implementación de nuevas medidas de prevención situacional basadas en la evidencia aportada.

PALABRAS CLAVE

análisis delictivo, ciber lugar, cibercrimen, Ciencia de Datos, Criminología Ambiental, patrón delictivo, prevención del crimen

CHAPTER I

GENERAL INTRODUCTION

This doctoral thesis by compendium of publications has three fundamental defining characteristics that frame both the relevance of its contribution and the innovation of its approach. First, it constitutes a nexus between the old and the new. The four articles presented here seek to bridge the gap between relatively old theoretical bodies (i.e. Environmental Criminology theories) and empirical research applied to a relatively new object of study (i.e. cybercrime). In this statement, what appears as relative actually hides an absolute. One could argue whether Environmental Criminology theories are really old, or whether cybercrime as a phenomenon is really new. But what is out of the question is that there is a chasm between the theoretical construction of Environmental Criminology and its application to crime committed in cyberspace. The main problem is that these theories were conceived in the 70s and the 80s, when cyberspace did not even exist, and the impact of cybercrime could only have been anticipated by a visionary. It was not until two decades later that some scholars showed interest in the adaptability of criminological theory to tackle the new criminal opportunities offered by cyberspace. And despite the noblest attempts to do so since then, there is still much to be done today. This thesis advances one step in that direction.

Second, this thesis is also a repertoire of methodologies for cybercrime analysis. In the current era of datafication "[w]e have increasingly detailed data, very large information files and advanced software that combines each other to analyse detailed information about the offenders, the victims and the places" (Felson, 2015). Throughout the manuscript, a number of Data Science techniques are described and applied to analyse cybercrime data by adopting an eminently quantitative approach. Occasionally though, this approach is complemented and enriched by the application of few qualitative analysis techniques. Both share a crime and place focus. It is important to note that the data collected for hypothesis testing in each of the four articles come from a variety of sources that require, in turn, the use of a wide range of techniques for their adequate contrast. It is widely known that the complexity of the phenomenon under study makes it difficult to collect quality data, such as properly maintained official statistics or carefully designed and systematically administered victimization surveys. Instead, we were forced to rely on databases maintained by third parties, to scrape data from the Internet, to generate our own with surveys, and to obtain permission to access social media content. Quite an odyssey.

Third, this thesis has a strong vocation for applied crime prevention. This is possibly its most promising value, since beyond what it currently adds to praxis, it has great potential to contribute in the future. Besides the theoretical approach that guides it, whose development seeks to advance the discipline, and the different methodologies that accompany each empirical study presented here, whose application intends to illustrate the multiple possibilities offered by the combination of a situational approach and Data Science for cybercrime analysis, each article contains important insights to inform cybercrime prevention. Here lies the importance of the Environmental Criminology and Crime Analysis (ECCA) approach adopted. ECCA does not attempt to understand the root causes of crime, but to solve crime problems. The situational patterns of cybercrime revealed in this doctoral thesis have the potential to be cornerstones in the design of preventive strategies. One must only know where to look.

Here we show time patterns of repeat victimization and offending concentration, hyperlink networks that connect illicit websites with distinct environmental features, online situational contexts that define the risk of victimization and offending, and the microenvironments' characteristics where cybercrime occurs. Every crime pattern has its own form of prevention and, in this thesis, we make some suggestions.

And last, this thesis is the spearhead of an important line of research. Given its intended scope, this is such an ambitious project that a doctoral thesis appears to be insufficient to complete it. To the mammoth task of adapting a criminological approach with so many nuances to an object of study with so many edges, it would be necessary to dedicate not a doctoral thesis, but probably a lifetime devoted to research. And yet, here are the first steps that open the way for future work. In an attempt to satisfy at least the main objectives of this doctoral thesis, a framework is proposed to adapt the approach of ECCA to cyberspace in order to make a preventive and empirical approach to different cybercrimes through the application of the concept of cyber place. Although the empirical task is pioneered here, credit must be given to the previous theoretical work of Miró-Llinares and Johnson (2018), *Cybercrime and Place: Applying Environmental Criminology to Crimes in Cyberspace*, which is the seed of this thesis. All articles presented here share its same philosophy, the ECCA philosophy, which through the application of the cyber place concept to concrete problems —ranging from repeat victimization problems, through environments that precipitate deviant behaviour, or others that constitute contexts of risk for victimization and offending, to the dissemination of harmful content through social media— proposes concrete solutions. Who can say whether further research will put the icing on the cake?

In the following sections of this general introduction, the object of study and the approach adopted are developed in depth. After outlining the objectives of the line of research on which the thesis articles are based, CHAPTER II covers the theoretical development of the thesis, in which a model for adapting ECCA to cybercrime and cyberspace is presented. This whole reflective process serves to raise the general research questions and hypotheses in CHAPTER III. Afterwards, the materials and methods used in the thesis are summarized according to a Data Science framework in CHAPTER IV. As an interlude, CHAPTER V then provides a brief overview of the four articles that comprise the thesis. Each of the following sections from CHAPTER VI to CHAPTER IX corresponds to one article. CHAPTER X discusses the general results obtained in the four articles and their implications in relation to the third characteristic. This allows the research question of the thesis to be addressed. Finally, CHAPTER XI —duplicated in Spanish— assembles a series of conclusions consistent with the objectives initially posed.

1.1 The object of study: From the myth of cybercrime to the reality of many cybercrimes

To properly introduce the reader to the object of study of this thesis, an initial clarification effort must be made. Note that this thesis distinguishes between crime, traditional crime or crime committed in physical space, and cybercrime or crime committed in cyberspace. This strategy will prove useful later, when confronting Environmental Criminology theories applied to traditional crimes, with their application to cybercrime. Such instrumental contrast is based on illustrative rather than ontological criteria, because what differentiates traditional crime from cybercrime is not its sophistication but the environment in which it occurs (Miró-Llinares & Johnson, 2018), which in turn determines it. And this is key to the matter. Rather than focusing on the nature of the phenomenon, emphasis is placed on how the environment where the event occurs influences its manifestation. Despite the profound academic discussion [1], as argued below, cybercrime is ultimately crime committed in cyberspace; as simple as that, or as complicated as that —depending on one's point of view—. If this axiom is understood, the rest of the technicalities that characterize cybercrime take a back seat.

1.1.1 Cybercrime fallacies

It is important to demystify cybercrime in order to approach its study. In this sense, the crime fallacies served to cast doubt on what we thought we knew about crime, but in fact did not (Felson & Eckert, 2019). As with traditional crime, there are a number of misconceptions about cybercrime that prevent this phenomenon from being adequately dimensioned. Such misconceptions are composed of those aspects that are taken for granted in relation to cybercrime and others that are ignored. Only through a correct

[1] Although the concept seems to be well established, actually there is still a debate about the definition of cybercrime (Payne, 2019). In fact, the concept has evolved as technology has. As Payne notes, the first use of the term "computer crime" can be traced back to the book *Crime by Computer* (Parker, 1976). For this author, computer crime is, in short, any crime that is related to a computer in one way or another; either because it is the object of a crime, the environment where a crime occurs, the instrument for committing a crime, or the symbol of a crime (Parker, 1976; Payne, 2019). With a few exceptions regarding its criminalisation (e.g. Hollinger & Lanza-Kaduce, 1988), in the following years the term received limited attention until the advent of the new millennium and, with it, the Budapest Convention on Cybercrime in 2001 and other academic publications that reflect the growing interest in the phenomenon (e.g. Grabosky, 2001). Interestingly, the Budapest Convention does not provide a definition of the general concept but articulates guidelines for a homogeneous response to the phenomenon from a legal perspective. And for Grabosky (2001), cybercrime is nothing more than *Old Wine in New Bottles*; in other words, the biggest change compared to traditional crime is the means of commission, but not the nature of the phenomenon. Since then, at least seven terms have been used to refer to what is now known as cybercrime: computer crime, digital crime, electronic crime, Internet crime, network crime, technocrime, and virtual crime (for a review, see Payne, 2019). In spite of the diverse terminology —and each one with its nuances— few managed to capture the true scope of the phenomenon as Parker did. The main problem with many definitions is that they expect the object of the crime to be a computer or assume that the offender must have computer skills to commit the crime (Miró-Llinares, 2012). This would exclude all those cybercrimes of a social dimension that target people. So, as cybercrime adapted to technological developments, these definitional elements were insufficient to cover the true scope of the phenomenon (Choi et al., 2019). To tackle this obstacle, other authors propose a broader interpretation of cybercrime. For example, Wall points out that "cybercrimes are criminal or harmful activities that are informational, global and networked and are to be distinguished from crimes that simply use computers" (2007, p. 4). Although the debate is not yet over, we believe that this author offers a more comprehensive view of the concept throughout his book by referring to the transformation of traditional criminal activity into a global arena, cyberspace, which provides new criminal opportunities and gives rise to cybercrime, a phenomenon whose control and prevention is more complex. In any case, the debate continues, since —for some— "whilst we might think we know what cybercrime is, we remain far from really understanding it" (M. R. McGuire, 2020, p. 25).

dimensioning of the phenomenon will it be possible to employ adequate mechanisms for its prevention. As an introductory section we believe that there is no better way than to summarize the fallacies that were attributed to crime and, subsequently, adapted to cybercrime (Miró-Llinares, 2015a). Just as there are eight crime fallacies (Felson & Eckert, 2019) [2], there is also a reflection on the cybercrime fallacies (Miró-Llinares, 2015a). Some of them refer to the name of the phenomenon itself, to its spatiotemporal distribution, and to its technification.

The first of which is the Name Fallacy. Miró-Llinares (2015a) points out that when the term cybercrime is mentioned, it is immediately associated with a type of crime that is highly sophisticated and technical, but that this is not true. The author argues that we relate cybercrime with computers, with IT, but that technological advances have improved the accessibility of users to the utilities they offer. This would facilitate the commission of cybercrimes through, for example, applications installed on the mobile phone, while democratizing criminal opportunities (Cullen & Kulig, 2018). In fact, the most prevalent cybercrimes such as the many forms of fraud require a relatively low level of skills and IT knowledge (Button & Cross, 2017), and can be executed from any device that is connected to the Internet (e.g., romance fraud, Nigerian letters, advance fee fraud, lottery scam). Furthermore, speaking of cybercrime gives the impression that it is one thing, a whole. But the reality is that cybercrime can take many forms, just like crime does (Miró-Llinares, 2015a).

[2] Eight are the fallacies that Felson and Eckert (2019) identify in the sixth edition, although this number has varied from previous editions of *Crime and Everyday Life*. For example, in the fourth edition of the book, the authors describe nine fallacies (Felson & Santos, 2010). It may seem that the authors have simply removed one of the nine, but that is not the case. Interestingly, they exclude two of the nine (i.e., the Vague-Boundary Fallacy and the Random Crime Fallacy) and include a new one (i.e., the Big Gang Fallacy) to add up the eight totals (see Miró-Llinares, 2015a). How curious. But as exciting as this whole topic is, we must focus here on what really concerns us, the fallacies of cybercrime. So, we refer the reader to each of the six editions of this magnificent book to learn more about Felson's work and the crime fallacies.

The second fallacy relevant to this thesis is the Random Cybercrime Fallacy (Miró-Llinares, 2015a). The original Random Crime Fallacy is built on the popular belief that crime can occur at any time and place, and affect anyone (Felson & Santos, 2010). So, its cyber counterpart suggests exactly the same thing. However, evidence appears to point in the opposite direction: it appears that crime describes identifiable patterns that cause its concentration in space and time. In fact, crime concentration is considered by many a scientific law (Weisburd, 2015). Although it is unclear whether the same law is observed for cybercrime, what is known to date about cybercrime events is that, like the traditional crime events, they occur more frequently in some places than others and at particular times (Miró-Llinares & Johnson, 2018; Miró-Llinares & Moneva, 2019a). And if cybercrime is not randomly distributed, people and things are not victimized at random either. Both the routine activities people undertake and the environments they transit determine their risk of online victimization. Similarly, offenders' targets possess certain characteristics that define their suitability. According to Clarke (1999), when a target is CRAVED (i.e. concealable, removable, available, valuable, enjoyable, and disposable) it is highly vulnerable and thus becomes a hot product, also in cyberspace (G. R. Newman & Clarke, 2003) [3]. In short, far from occurring randomly, there are many things that condition crime. One of which is the human factor.

The human factor in cybercrime is the central theme of the Cybersecurity Fallacy. Presented as a digression in the book chapter authored by Miró-Llinares

[3] CRAVED is a revamped version of the VIVA acronym originally proposed by Felson (Cohen & Felson, 1979) for better application to targets of crime. Felson originally created VIVA (i.e. value, inertia, visibility, and accessibility) to encompass the characteristics that make a target suitable in a broad sense, which includes both people and objects; whereas CRAVED was designed to be used primarily for objects. CRAVED continues to be used today to analyse the suitability of a wide range of targets for many different crimes. Miró-Llinares (2012) also attempted to adapt VIVA to targets in cyberspace by using the acronym IVI (i.e. introduction, value, and interaction); unfortunately, scientific literature written in Spanish has little outreach.

(2015a), this —third— cybercrime fallacy has become increasingly relevant until today, when the human factor in cybercrime and cybersecurity has become one of the leading perspectives in cybercrime research (Leukfeldt, 2017; Leukfeldt & Holt, 2020). This fallacy refers to the reductionism of the concept of cybersecurity towards the purely technical that distorts its own nature. Some perceive more technical cybercrimes as more dangerous. Maybe that is true, maybe not. Or perhaps it is mere ignorance that attributes the property to the substance. The truth is that the few figures available on the impact of cybercrime show that it is the less technical forms that cause more victims and also more economic losses (Internet Crime Complaint Center, 2018). People seem to forget that cybersecurity is only one side of the coin on whose back the concept of cybercrime is carved; and in cybersecurity —just like in cybercrime— the human component is huge. We are not yet in the age where robots enjoy such autonomy. In the end, whoever is responsible for a cybercrime is human; whoever suffers the consequences of a cybercrime, whether directed at an object or a person, is human; and whoever creates the strategies to control cybercrime is also human. Just as emphasizing the human factor is central to understanding traditional crime, it is just as important regarding the various forms of cybercrime.

1.1.2 Classifying cybercrimes

Cybercrime is diverse as it encompasses many forms of crime (Miró-Llinares, 2012). For example, there is cyber-trespass, cyber-deception, cyber-obscenity, and cyber-violence (Wall, 2001). Each of these categories encapsulates many specific forms of crime that complete the phenomenological puzzle. Profound knowledge of each form of crime requires specific examination because a detailed analysis of the phenomenon as a whole entity is not feasible (G. R. Newman & Clarke, 2003). Note that this is not a fallacy, but a reality. To address this, some attempts have been made to provide a conceptual framework that allows cybercrime to be properly defined and delimited. Although many have tried [4], perhaps the most important classification of cybercrimes —because of its impact on criminological research— was the one prepared by McGuire and Dowling (2013a, 2013b, 2013c) for the Home Office; a straightforward macro categorization that has been widely used by researchers around the world since then. For these authors, there are two types of cybercrime: cyber-dependent crimes, and cyber-enabled crimes. Cyber-dependent crimes "are offences that can only be committed using a computer, computer networks or other form of IT" (McGuire & Dowling, 2013a, p. 4). Cyber-enabled crimes "are traditional crimes, which can be increased in their scale or reach by use of computers, computer networks or other forms of IT" (McGuire & Dowling, 2013b, p. 4). Although the definitions make it clear enough, the main difference is that while cyber-dependent crimes can only be committed online (for a review, see Maimon & Louderback, 2019), cyber-enabled crimes can be committed in both online and offline environments.

Another classification —possibly the most influential written in Spanish— is elaborated by Miró-Llinares (2012). In fact, this author proposes not one, but two overlapping classifications based on two criteria: the incidence of IT on criminal behaviour, and the motive and criminological context. Regarding the former, Miró-Llinares (2012) distinguishes between pure attacks, replica attacks, and content attacks; with respect to the latter, a distinction is made between economic cybercrimes, social cybercrimes, and political cybercrimes. While the first adds small nuances to previous

[4] One of the first and most recognized taxonomies is that produced by Wall (2001b), which has already been referred to in this paragraph. This framework has also been used recently by Holt and Bossler (2016) to structure the phenomenology of cybercrime in their award-winning book *Cybercrime in Progress*. A few years later, Wall (2005) elaborated a new categorization to distinguish between computer integrity crimes, computer related crimes, and computer content crimes. Later on, the US Department of Justice, cited by Clough (2010) in *Principles of Cybercrime*, developed another taxonomy based on three categories. According to this classification, there are computer crimes, computer-facilitated crimes and computer-supported crimes. And so on, multiple classifications have come to light. These are just a few of the many examples found in the literature.

classifications based on a similar criterion, the second brings a new dimension to the phenomenon. Let us focus on the second one. For Miró-Llinares (2012), this is a strong classification that allows to distinguish cybercrimes with diverse criminological features (e.g. just as the nature of a theft is very different from that of an assault, cyberfraud is quite different from online harassment). In addition, each category corresponds to one of the three functional areas of Internet use: the development of economic relations, personal development, and the development of institutional and supranational relations (Miró-Llinares, 2012). Interestingly, each of these contexts leads to different routine activities, which shape the convergence between people, and people and objects. And varying forms of convergence enable distinct crime opportunities. Such distinction is important for crime prevention, as it is likely that cybercrimes in each of these categories will require different strategies [5].

In this doctoral thesis, four cybercrimes are analysed to inform their prevention: website defacement, match-fixing, online harassment, and online hate speech. Why these cybercrimes and not others? As the reader may have noticed, there seems to be an inconsistency in the selection of cybercrimes analysed in this thesis. There is an explanation for this. The short answer is that the four selected cybercrimes have such different characteristics that they provide an ideal scenario on which to put criminological theories to test (i.e. Environmental Criminology). But let us elaborate. This is where the cybercrime fallacies come back into play. Firstly, in order to test the validity of an analytical approach for cybercrime prevention, its application to multiple phenomena rather than a single crime is mandatory (the Name Fallacy). Cybercrime

[5] In his book on successful case studies, Clarke stresses that Situational Crime Prevention (SCP) measures must be "specific in nature, and cater precisely to addressing particular types of crime" (Clarke, 1997, pp. 4–5). Clarke's work has proven to be the embodiment of "preaching by example" both concerning the application of SCP in physical space (Clarke, 1997) and in cyberspace (G. R. Newman & Clarke, 2003). The results of his research speak for themselves.

cannot be prevented by using a general framework no matter how comprehensive or integrated it may seem; only by implementing concrete evidence-based strategies that specific forms of cybercrime can be reduced. Secondly, if we wish to generalise about the nature of a phenomenon (i.e. cybercrime) it is necessary to observe how it behaves in different contexts. Should the concentration property also be attributed to cybercrime in the future, the spatiotemporal distribution in all its forms must be examined (the Random Cybercrime Fallacy). Thirdly and finally, cybercrime has a substantial technical component, but it also has an essential human factor (the Cybersecurity Fallacy). In this thesis we wanted both elements to be represented, for the results of the analysis would be flawed without considering either of them.

So, what should be known about each one? To avoid repetition, in the following lines we present each cybercrime briefly, as they will be examined in depth in their respective article.

- Website defacement is a form of hacking that involves accessing a web server to modify the content displayed on a web page. Some of the most used hacking techniques for defacing are file inclusion, SQL injections, or the exploit of known server vulnerabilities (Romagna & Van den Hout, 2017). Because this cybercrime does not require in-depth technical knowledge it is usually carried out by novice hackers or script-kiddies to gain status among the hacker community (Holt, 2011). Defacements are also part of the repertoire of hacktivists, as many people can be reached relatively easily with an ideological message (Romagna, 2019).
- Match-fixing refers to —in the context of this thesis— the advertisement of results of allegedly manipulated sport events on websites and their subsequent sale. These websites include both the price of the matches and the procedure to

obtain the information. As of 2018, the king of sports in match-fixing is tennis, although a number of incidents have also been reported concerning football and other sports (ESSA, 2018). Interestingly, most of the fixed matches advertised on websites are football games. This cybercrime is just the tip of the iceberg of a criminal network that moves millions in profits (Haberfeld & Sheehan, 2013).

- Online harassment can be defined as the repeated and unwanted contact experienced through IT by an individual. There seems to be consensus on these two basic defining elements, although some nuances may affect this concept (Wolak et al., 2007). For example, some argue that online harassment should not involve an emotion of fear (Baum et al., 2009), or that it should be a repeated, but not continuous, act (Miró-Llinares, 2012). In any case, despite the difficulty in defining the term, research on this topic is already extensive and has been related to many contexts (e.g. professional, educational, sexual).
- Online hate speech is the expression of hatred towards certain groups on discriminatory grounds via the Internet. And such grounds are perfect for radicalization to sprout. Discrimination may refer to: “race, ethnicity, gender, gender identity, sexual orientation, national origin, religion, or other group characteristic” (Costello & Hawdon, 2019, p. 1). Hence, note that there is not one, but many forms of online hate speech (Miró-Llinares, 2016). Both the ease of dissemination of this type of content through social media (e.g. Twitter, Facebook) and the concern about its control by service providers has placed online hate speech at a privileged place in the research agenda in recent years (Miró Llinares, 2017).

The correct categorization of any cybercrime is a critical step for its accurate understanding. Below, Table 1 draws on the classifications of McGuire and Dowling

(2013a, 2013b, 2013c), and Miro-Llinares (2012) to categorize each cybercrime. For clarity, a single category has been assigned to each cybercrime. However, website defacement would admit a multiple classification according to Miró-Llinares' (2012) taxonomy depending on the context. For example, some defacements are executed for political purposes when part of hacktivist activities [6], but they can also be used for extortion, or as a challenge to gain status among peers —or just for fun— (e.g. Holt, Leukfeldt, et al., 2020). We have chosen the economic category for defacements because in CHAPTER VI we approach this form of hacking by comparing it to traditional property crimes such as burglary. In this case, we sacrifice exhaustiveness for conciseness.

Table 1.
Dual classification of the four cybercrimes studied

	Classification according to	
Cybercrime	McGuire and Dowling (2013a, 2013b, 2013c)	Miro-Llinares (2012)
Website defacement	cyber-dependent	economic
Match-fixing	cyber-enabled	economic
Online harassment	cyber-enabled	social
Online hate speech	cyber-enabled	political

1.2 The approach: Environmental Criminology and Crime Analysis

It is common for doctoral theses to address a specific problem, a specific object of study. However, this one does not examine a specific type of crime (the term cybercrime can be misleading, see the Name Fallacy). In fact, it is possible that the least relevant part of this piece of research is its object of study. This thesis examines cybercrime as an event. Here cybercrime itself and not its multiple and different

[6] A recent example is the website defacement sustained by the U.S. Federal Depository Library Program website. Which was attributed to the Iran Cyber Security Group "HackerS" on the occasion of the killing of Iranian commander Qasem Soleimani (Chiu, 2020). The defacement consists of the display of an image of Donald Trump's face, bloodied from being punched by a member of the Iranian army. The image is accompanied by other features such as missiles, a map of Iran, and a short threatening text.

manifestations is the object of interest. Addressing crime as an event requires a particular approach, one that focuses on the environment where crime opportunities emerge (Felson & Clarke, 1998). So certainly, this is a thesis about cybercrime, but also about something else. Here we propose the individualised application of an analytical approach to tackle various cybercrime problems. Rather than a mere exercise in theoretical development, this thesis adopts an approach that has been characterized by delivering practical solutions for crime prevention. Such an approach consists of two elements: a theoretical framework and a tool for its application. The framework is Environmental Criminology theories and the tool is crime analysis; jointly, ECCA. ECCA is the guide proposed by Environmental Criminology theories that is channelled through crime analysis tools in order to solve crime problems (Wortley & Townsley, 2017b) [7].

Environmental Criminology (P. J. Brantingham & Brantingham, 1981) was conceived to provide solutions to crime problems that had not been solved through the application of conventional criminology frameworks oriented towards the study of the individual offender (Jeffery, 1971). Its founding principle was clear: to shift the focus from the individual involved in a crime to the environment where crime occurs; or in other words, change the focus from criminality to crime. A call for change that was announced almost half a century ago by Jeffery (1971) in response to the widespread belief that "Nothing Works" in the criminal justice system to rehabilitate offenders (Martinson, 1974):

> "A new school of environmental criminology must emerge, based on scientific procedures, behaviourism, and environmentalism. The basic

[7] ECCA also gives its name to the annual symposium where the leading proponents of this intellectual movement meet. With almost three decades of tradition, ECCA was first organized in 1992 and continues to be held today (the 2020 edition will take place in Leeds). For an overview of the origin, objectives, and development of the ECCA symposiums, refer to the work of Bichler and Malm (2008).

> principles of the classical school (i.e. prevention of crime before it occurs and certainty of consequences for behaviour) would be retained, but the emphasis would shift from punishment to reinforcement and from the individual offender to the environment. The major form of control would be reinforced of lawful behaviour and the removal of reinforcement for illegal behaviour. The focus would be the environment in which crimes are committed, not the individual offender" (Jeffery, 1971, p. 279).

But conceptual frameworks need tools that enable their implementation. Otherwise how do they prove their usefulness? The instrumental component that enables the materialisation of such ideas is crime analysis. Crime analysis can be defined as "the set of systematic, analytical processes that provide timely, pertinent information about crime patterns and crime-trend correlations" (Emig & Heck, 1980; cited in Wortley & Townsley, 2017b, p. 1). As can be inferred from this definition, there is no closed list of crime analysis techniques, but their nature, quantity, and variety will be determined by the need for their use. Crime analysis techniques must be capable of producing clear and concise results to match the synthetic work of the crime analyst applying them. Its ultimate goal is to transform complex problems into simple solutions, easily understandable by the people responsible for their implementation. Indeed, it seems an easier task than it is.

Environmental Criminology theories and crime analysis tools must mutually feed into each other to better fulfil their function. The synergies between the two have allowed a series of common cornerstones to be identified. Specifically, three are the propositions on which ECCA relies:

> "Criminal behaviour is significantly influenced by the nature of the immediate environment in which it occurs. […]
>
> The distribution of crime in time and space is non-random. […]

> Understanding the role of criminogenic environments and being aware of the way that crime is patterned are powerful weapons in the investigation, control and prevention of crime" (Wortley & Townsley, 2017b, p. 2).

What now seem obvious statements were once revolutionary claims. Such was the turmoil that environmental criminologists caused that their banishment from criminology was considered, labelling them as the proponents of a new discipline unconcerned with the root causes of crime. And there was some truth in that consideration. Environmental criminologists were unconcerned with the root causes of crime because they simply could not do anything to reverse them. Instead, environmental criminologists were concerned with performing small but meaningful manipulations of environments to reduce crime opportunities. Over time, analytical theoretical approaches, robust research designs, and rigorous methodological executions proved that a different kind of criminology was possible. Most notably, the ECCA approach earned special acceptance by crime control practitioners because of its close connection to the reality of police praxis and a profound understanding of law enforcement agencies. In this way, the end users of the approach became its greatest advocates; an achievement that many scientific disciplines cannot boast of. Fortunately, the change of perspective sought by Jeffery has long since taken place and research from ECCA has shown its usefulness in crime prevention (Clarke, 1997). Now, the problem with cybercriminals is not that nothing works, but that we are uncertain about what works. As with traditional crime, we believe that Environmental Criminology has a great potential to prevent cybercrime that remains undiscovered.

In the following sections, the ECCA approach is expanded in two separate parts: first, explaining the main environmental theoretical frameworks that comprise it and, second, illustrating its application through Criminology of Place as an approach to crime analysis. Two additional sections follow showing how this approach has been applied to the object of study of the thesis (i.e. cybercrime), considering the particular characteristics of the new environment in which it manifests (i.e. cyberspace). Finally, the potential of the concept of place for cybercrime prevention is discussed.

1.2.1 Environmental Criminology theories

Environmental Criminology theories in the broad sense have received many names. Some of the most commonly used are: Environmental Criminology (Brantingham & Brantingham, 1981), Crime Science (Clarke, 2010), socio-spatial criminology (Bottoms, 2012), and situational opportunity theories (Wilcox & Cullen, 2018). The reason for such variety is that some of these denominations compile different theoretical bodies, which means that Environmental Criminology is not the same for everyone. For Bottoms (2012), Environmental Criminology is only one of the three schools of thought in socio-spatial criminology [8], although he argues that the term has been misused to encompass all three. For Wilcox and Cullen (2018), situational opportunity theories would have a broader scope than Environmental Criminology, as they would encompass other theoretical bodies that contribute to understanding individual victimisation and offending. For Clarke (2010), Crime Science shares the fundamental premises of Environmental Criminology [9], but possesses certain particularities: aims primarily at reducing crime rather than reducing crime opportunities, supports the incapacitation of

[8] Bottoms (2012) differentiates three schools of thought within socio-spatial criminology: The Neo-Chicagoans, intellectual descendants of the Chicago School of Sociology who are interested in the social organization of neighbourhoods; the Environmental Criminology and Crime Analysis group, closely connected to practitioners by their practical interest in crime events for crime prevention; and a third school interested in culture and semiotics from a signal crimes perspective. For this author, only the second represents the canons of environmental criminology.

[9] Drawing on previous work by Wortley and Mazerolle (2008), Clarke lists five premises of environmental criminology: "[1] Crime is the outcome of the interaction between dispositions and situations, […] [2] Crime is always the product of choice, […] [3] A crime-specific focus is fundamental to understanding the role of situational factors in crime, […] [4] Crime is heavily concentrated, […] [5] Crime can be reduced (often immediately and dramatically) by environmental changes that reduce opportunities and modify precipitators" (Clarke, 2010, pp. 273–275).

prolific offenders, and embraces multi-disciplinarity beyond social sciences. Following the leading scholars of the discipline (Andresen et al., 2010; Bruinsma & Johnson, 2018; Wortley & Townsley, 2017b), the term Environmental Criminology will be used in this thesis to refer to the "family of theories that share a common interest in criminal events and the immediate circumstances in which they occur" (Wortley & Townsley, 2017b, p. 1).

Environmental criminologists identify three fundamental mid-range theoretical bodies for understanding crime events: The Routine Activities Approach (Cohen & Felson, 1979), the Geometry of Crime (Brantingham & Brantingham, 1981), and the Rational Choice Perspective (Clarke & Cornish, 1985). The mid-range label reflects the modesty of the approach. Environmental Criminology does not seek to understand the root causes of crime, but to understand why certain crimes occur in specific contexts in order to prevent them. In this sense, there are two practical frameworks that further deepen the applicability of the Rational Choice Perspective for crime prevention: the SCP measures (Clarke, 1980), and the Situational Precipitators of Crime controlling techniques (Wortley, 2001). The overlap of these theoretical bodies called for an integrative effort, more ambitious in terms of explanatory scope. It is to respond to this challenge that the Crime Pattern Theory (P. L. Brantingham & Brantingham, 1993a) was conceived. Figure 1 aims to illustrate the synergies and dependencies between all these theoretical bodies. Each of them is outlined below.

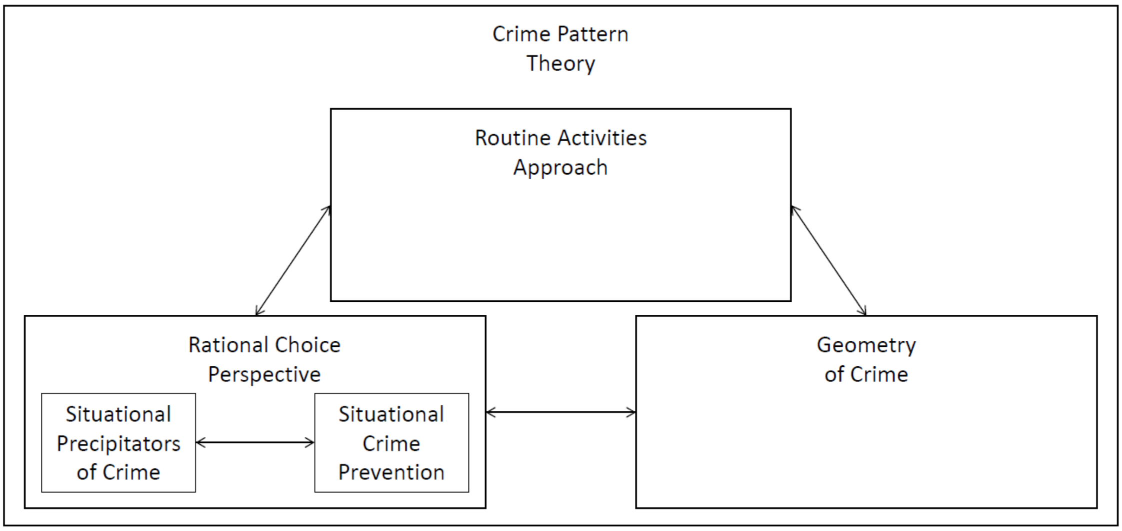
Crime Pattern Theory
Routine Activities Approach
Rational Choice Perspective
Situational Precipitators of Crime
Situational Crime Prevention
Geometry of Crime

Figure 1. Conceptual map of Environmental Criminology theoretical bodies

1.2.1.1 *The Routine Activities Approach*

The Routine Activities Approach provides a macro explanation for the variation in crime rates as a function of technological and social change (Cohen & Felson, 1979). In their seminal work, the authors note that after the Second World War, some crime rates continued to rise despite improved socio-economic conditions. They hypothesized that technological advances, such as the car, and social advances, such as the incorporation of women into the labour market, encouraged people to spend more time on the streets. As a result, contact between strangers would be facilitated and households would be empty for longer. In turn, this would favour specific crime opportunities. To test their hypothesis, they examined different types of robberies, burglaries, larcenies, and murders, and found that both personal crimes perpetrated by strangers and burglaries committed during the day had increased. Such findings were relevant not because they supported their hypothesis, but because they contradicted the theories postulated to date. Of course, Cohen and Felson's paper shook the foundations of mainstream criminology, so initially it caused significant discomfort. Today no one disputes the tremendous impact that the routine activities approach had on criminology, but it was not due to its macro explanation for crime.

The whole macro analysis in this research was later eclipsed by what would become one of the most popular propositions in Environmental Criminology, also known as the chemistry for crime (Felson & Clarke, 1998). A micro causal mechanism that underlies the macro explanation: “The probability that a [crime] will occur at any specific time and place might be taken as a function of the convergence of likely offenders and suitable targets in the absence of capable guardians” (Cohen & Felson, 1979, p. 590). In this equation, a likely offender would be an individual “with both criminal inclinations and the ability to carry out those inclinations” (Cohen & Felson, 1979, p. 590), a suitable target would be defined by the VIVA acronym, and a capable guardian would be a person or an object that can prevent the crime from occurring. Paradoxically, it was this framework for the micro analysis of crime which would propel the approach to notoriety. Cohen and Felson's research would then become the seed from which Environmental Criminology sprang.

1.2.1.2 *The Geometry of Crime*

Chronologically, the second theoretical development in Environmental Criminology is the Geometry of Crime. The Geometry of Crime provides a meso explanation of the geographical distribution of crime opportunities in the urban environment (Brantingham & Brantingham, 1981). Building on previous research in the field of geography, Brantingham and Brantingham established the fundamental premise that crime depicts patterns; that is, crime is not randomly distributed in space. Crime distribution would then be a direct consequence of the emergence of crime opportunities in the environmental backcloth that surrounds an individual (P. L. Brantingham & Brantingham, 1993b). By their mere presence, individuals alter the environmental backcloth and are influenced by it. Note that this backcloth is not only comprised of static elements such as the urban fabric, but also of dynamic ones such as socio-legal norms. Hence, crime opportunities would be determined by virtually immutable elements (e.g., the road network), and by others under constant change (e.g., the time of day) (Andresen, 2010). This results in crime opportunities that vary from time to time and from place to place.

People carry out their routine activities while they travel through this environmental backcloth, visiting some areas more often than others depending on how central they are in their everyday life. The areas that people visit more often constitute their activity nodes (P. L. Brantingham & Brantingham, 1993b). Different types of nodes can be identified according to the type of activity that people carry out in them (e.g. domestic, work, leisure). When people move from one node to another, they usually follow a path in a recurrent way, with little and occasional variations (P. L. Brantingham & Brantingham, 1993b). Paths connect activity nodes and thus configure people’s activity space. As people move through their activity space, they eventually develop a mental map of their environment called awareness space, becoming more comfortable within this space and more uncomfortable outside it (Andresen, 2010). Offenders, like everyone else, have their own activity and awareness space and often commit crimes in these areas (P. L. Brantingham & Brantingham, 1981). In addition, the urban design often presents strong contrasts between neighbouring areas, which can be architectural, functional, or socio-cultural. In these perceptual edges people with different background converge, sometimes leading to conflict and crime (P. L. Brantingham & Brantingham, 1993b). Together, the interaction of people with nodes, paths, and edges constitutes the Geometry of Crime (P. L. Brantingham & Brantingham, 1981). By identifying these settings that constitute the backbone of our daily activity (i.e. places where people spend most of their time), Brantingham and Brantingham are able to outline a model of crime risk. There is where the value of this framework of analysis lies, in its potential to anticipate the geographical distribution of crime opportunities to prevent crime events (Andresen, 2010). The Geometry of Crime will prove to be key in subsequent practical developments that integrate additional theoretical frameworks for crime prediction.

1.2.1.3 *The Rational Choice Perspective*

One of such frameworks is the Rational Choice Perspective. The Rational Choice Perspective provides a micro explanation for the decision-making process of offenders as they interact with their immediate environment (Clarke & Cornish, 1985). To develop their approach, Clarke and Cornish (1985) examine the advances in research from various disciplines (i.e. sociology of deviance, criminology, economics, and cognitive psychology) on rational decision-making to conceive of crime as the outcome of such process; an integrative effort that aims to provide a unified framework for numerous scattered findings. The Rational Choice Perspective states that decisions made by offenders regarding crime perpetration are the result of a cost-benefit calculation. Should the benefit be greater than the cost, then the offender is more likely to commit a crime. To placate criticism, it should be noted that the authors do not assume a complete rationality of offenders, but a limited rationality that takes into account the biases and heuristics inherent in the human mind. Obviously, this perspective implies a non-deterministic approach to crime involvement.

Crime involvement can be defined as a four-stage process including initial involvement, the crime event itself, continuance, and desistance; and at each stage, different decisions are made (Clarke & Cornish, 1985). As these stages may vary greatly from one crime to another, the authors emphasize the need for their model to be crime specific (Clarke & Cornish, 1985). During crime involvement, the offender first determines to commit the crime and then executes it because of a precipitator —usually a chance event—. Then, the crime event occurs upon offender's target selection, which contrary to popular belief, usually corresponds to a clumsy and improvised process. Next, continuance is determined by a sequence of reinforcements that the offender receives while committing the crime. If the offender positively evaluates the situation, the process of crime commission continues. Finally, desistance may occur if the offender perceives an adverse circumstance in the course of the criminal action [10]. As a result, the offender can simply cease the action or move on to a new —and more suitable— target producing crime displacement.

There is crime displacement when offenders alter their criminal activity upon encountering obstacles that are difficult to overcome (Clarke, 1980). Environmental criminologists have identified six types of crime displacement: temporal, spatial, tactical, target, functional, and perpetrator (Barr & Pease, 1990) [11]. Now, crime can be displaced in all these ways, but can such displacement be induced? More importantly, can crime be displaced so effectively that it is reduced or completely suppressed? Apparently, this is what Clarke thought when he began to craft the principles of the

[10] The systematic study of this rational and multi-stage decision-making process is carried out through crime scripts. “The script is generally viewed as being a special type of schema, known as an ‘event’ schema, since it organizes our knowledge about how to understand and enact commonplace behavioural processes or routines” (Cornish, 1994, p. 32). Depending on the scope of the analysis, there are different types of crime scripts. From the most general to the most specific, these are: universal script, metascript, protoscript, script, and track. For Cornish (1994), the most useful for examining criminal involvement processes is the track due to its high degree of detail. However, the universal script is often used for standardisation. This type of script is structured around a sequence of phases with little variability that facilitates comparative analysis (i.e. preparation, entry, precondition, instrumental preconditions, instrumental initiation, instrumental actualization, doing, postconditions, and exit). The excellent paper cited can be consulted for more information on any of these aspects.

[11] There have not always been six types of displacement. Originally, Reppetto (1976) identified just five: Temporal, when a crime is committed at a different time; tactical, when offenders are forced to change their modus operandi; target, when the original is inaccessible; territorial —which would correspond to the spatial—, when the offenders move to another geographical location to commit the crime; and functional, when the offenders decide to commit a different type of crime. To these five types of crime displacement, Barr and Pease (1990) add a sixth: perpetrator, when it is an offender other than the original who ends up committing the crime.

SCP. For Clarke (1980), SCP should rest on two fundamental pillars: reducing the physical opportunities for crime, and increasing the risks of being caught. Subsequently, these pillars became three —each divided into four segments— whose multiplication resulted in twelve SCP measures (Clarke, 1992), and which then became 16 (Clarke, 1997). Following the discussion with Wortley about the role of situational precipitators of crime in the SCP, a final number of five pillars were defined with five sections each, resulting in the 25 SCP measures that are applied today (Cornish & Clarke, 2003) [12]. What caused Clarke's original scheme to be altered? For Wortley, there is a prior step to considering crime opportunities that precipitates criminal behaviour (Wortley, 1998). It is in this initial stage that some factors that induce criminal behaviour come into play, beyond the perceived costs and benefits of criminal acts, and which Clarke had not taken into consideration. These are defined as Situational Precipitators of Crime. Wortley's reasoning not only helped to complement the SCP measures, but introduced a novel scheme of precipitation control strategies (Wortley, 2001) [13]. Together, SCP measures and the situational precipitators of crime controlling strategies provide a cost-

[12] From the first pillar, dedicated to increasing the effort, the following measures were proposed: Target harden, control access to facilities, screen exits, deflect offenders, and control tools/weapons. The second pillar, built to increase the risk, contained the following measures: Extend guardianship, assist natural surveillance, reduce anonymity, utilize place managers, and strengthen forma surveillance. The third pillar was designed to reducing rewards through the following measures: Conceal targets, remove targets, identify property, disrupt markets, and deny benefits. To reduce provocations, the fourth pillar lists the following measures: Reduce frustrations and stress, avoid disputes, reduce emotional arousal, neutralize peer pressure, and discourage imitation. The fifth and final pillar advocates removing excuses through the following: Set rules, post instructions, alert conscience, assist compliance, and control drugs and alcohol. For a list of examples for each measure, see Cornish and Clarke (2003). Alternatively, visit: https://popcenter.asu.edu/sites/default/files/library/25%20techniques%20grid.pdf

[13] The Situational Precipitators of Crime controlling techniques (Wortley, 2001) resemble the 16 SCP measures established by Clarke in the second edition of his book *Situational Crime Prevention: Successful Case Studies*. Like the SCP measures at the time (Clarke, 1997), these techniques are organized in a four by four matrix. First, controlling prompts includes: Controlling triggers, providing reminders, reducing inappropriate imitation, and setting positive expectations. Second, controlling pressures includes: Reducing inappropriate conformity, reducing inappropriate obedience, encouraging compliance, and reducing anonymity. Third, reducing permissibility includes: Rule setting, clarifying responsibility, clarifying consequences, and personalising victims. Lastly, reducing provocations includes: Reducing frustration, reducing crowding, respecting territory, and controlling environmental irritants. Admittedly, some of these categories overlap with the SCP measures —something that Wortley (2001) himself acknowledges— but the fact is that they represented a major improvement in the evolution of the preventive framework.

effective toolkit for crime prevention that is frequently applied today by policy makers and practitioners.

1.2.1.4 *The Crime Pattern Theory*

The Crime Pattern Theory —or Pattern Theory of Crime— provides a multilevel explanation for the occurrence of crime events by integrating the concepts of the three main theoretical bodies of Environmental Criminology explained above (i.e. the Routine Activities Approach, the Geometry of Crime, and the Rational Choice Perspective) (P. L. Brantingham & Brantingham, 1993a). Brantingham and Brantingham observe there are conceptual synergies between these three theoretical bodies that enable the development of their meta-theory. Among them, they consider rationality to be the most important because it constantly shapes our routines (Andresen, 2010). For example, there is rationality in our routine activities when we decide what time we leave home or what means of transport we use; we choose to spend more time in some places than others because they better meet our needs; and we choose the routes that we travel because they are the ones that require less time or because they are more pleasant to transit. After a considerable time performing similar routines, we automate the decision-making processes to release cognitive load by creating templates.

Templates are “generally formed by developing an array of cues, cue sequences, and cue clusters that identify what should be considered a ‘good’ target in specific sites and situations” (P. L. Brantingham & Brantingham, 1993a, p. 370). Similarly, when offenders become accustomed to making decisions conducive to committing crimes, they develop crime templates (P. L. Brantingham & Brantingham, 1993a). Crime templates are not easy to deconstruct, and they vary according to each crime and context. As a result, an offender may not act the same way when committing a crime if the environmental backcloth is different but is likely to do so if the context does not change. According to Andresen (2010), two are the added benefits that Crime Pattern Theory brings to Environmental Criminology: The first and most obvious is that it brings together the previous developments of the discipline in a common framework, and the second is that it provides a framework of analysis that allows the real complexity and dynamism of crime events to be addressed. By understanding the rational processes involved in people's routine activities, it is possible to understand why crime patterns in certain places.

1.2.2 Place-based frameworks for crime analysis

It was stated earlier that, according to the Routine Activities Approach, there are three minimum elements for a crime event to occur: the presence of a likely offender, the presence of a suitable target, and the absence of a capable guardian (L. E. Cohen & Felson, 1979). To produce the indicated result, it is also indispensable that these three elements converge in time and place. By slightly shifting the focus of attention, this simple premise can be reformulated as follows: "[A] crime is highly likely when an offender and a target come together at the same place at the same time, and there is no one nearby to control the offender, protect the target, or regulate conduct at the place (Eck, 2003, p. 88)". If this reformulation of the chemistry of crime were to be illustrated, it would show what has become known as the crime triangle [14]. Through this change of scheme, Eck (2003) attributes more importance to the role that the place plays in the production of crime events. Now, the place replaces the capable guardian as a minimum element of crime, and pushes it to a second level of analysis, where it is converted into three types of guardian —one for each minimum element—: the classic guardian to control the target, the handler to control the offender, and the manager to

[14] Also known as the Problem Analysis Triangle. For more information, visit: https://popcenter.asu.edu/content/problem-analysis-triangle-0

control the place (Eck, 2003). Without adequate guardianship over any of these elements, crime is more likely to occur; thus, without proper management, places may become prone to crime.

1.2.2.1 *Crime Prevention Through Environmental Design*

That certain places are especially vulnerable to crime is something Newman (1972) already advocated in his book *Defensible Space: Crime Prevention through Urban Design*. This author argues that place owners can play a fundamental role in crime prevention and that success in performing this function is related to how the immediate environment is designed. Hence, there would be space designs that would facilitate the place management efforts of their owners, while others would hinder them. Newman's work is considered an application of the previous theories of Jacobs (1961; see Clarke, 2010), who argued that the surveillance of certain places had much to do with how they were designed. In *The Death and Life of Great American Cities*, Jacobs (1961) impetuously criticizes the urban renewal policies of the 1950s due to their inability to create spaces that promote public life. This model was a failure for the involvement of citizens in community life, since it instrumentalises the urban fabric for productive purposes, but forgets its essential function of strengthening social bonds and the feeling of belonging to the territory. According to Newman (1972) and Jacobs (Jacobs, 1961), the surveillance of certain urban environments would be affected by the type of road that connects them, the layout of the public furniture, and the design of the buildings, among other elements. In short, proper urban design could be a great ally to crime prevention in cities, while negligent design could be its worst enemy.

The concept of Crime Prevention Through Environmental Design (CPTED) was coined by Jeffery (1971), who took it from the discipline of Urban Studies and applied it to Criminology and the Criminal Justice system —although he did so in a broad sense—. Importantly, when Jeffery talks about CPTED, he is not just referring to urban design, but actually promoting a paradigm shift in crime prevention: changing the focus from the individual offender involved in crime to the environment where crime occurs (Jeffery, 1971). If it is not possible to undertake effective interventions with offenders so that they commit fewer crimes, then why not manipulate the environment so that they have fewer opportunities to commit them? This insight created a whole new school of thought that attracted many enthusiasts. The first CPTED researchers —1st generation CPTED— adhered to this idea and identified six areas of potential impact in this regard: access control, activity programme support, image/maintenance, target hardening, territoriality, and surveillance (Moffat, 1983). At that time, it was already a tradition that environmental approaches were criticized for their apparent simplicity, so a 2nd generation CPTED responded by including additional social dimensions related to: risk assessment, socio-economic and demographic profiling, and active community participation (for a review, see Cozens et al., 2005). If theory proved effective, then manipulating these elements of the environment would reduce crime opportunities and therefore also crime in places.

1.2.2.2 *The Criminology of Place*

According to Brantingham and Brantingham (1995), there are four types of places in the urban environment that are relevant to the geography of crime: crime generators, crime attractors, crime-neutral sites, and fear generators. Crime generators are places that concentrate many people —or people and objects— in specific time periods, thus generating crime opportunities (P. L. Brantingham & Brantingham, 1995). A fair or a concert would be crime generators. Crime attractors are places known to likely offenders for harbouring specific crime opportunities (P. L. Brantingham & Brantingham, 1995). Examples of crime attractors would be an unattended parking lot or a jewellery store. Crime-neutral sites are places where not many suitable targets converge nor are there particularly attractive crime opportunities (P. L. Brantingham & Brantingham, 1995). Such places usually dominate most of the urban environment. Fear generators are places in which people perceive fear of crime, regardless of whether they objectively harbour crime (P. L. Brantingham & Brantingham, 1995). There are several factors that cause the environment to be perceived as scary, such as darkness, unfamiliarity, and loneliness, among many others. Note that most urban areas do not constitute pure but mixed types of places; this is "they may be crime attractors for some types of crime, crime generators for other types of crime, and neutral with respect to still other types of crime" (P. L. Brantingham & Brantingham, 1995, p. 9). This exercise of conceptualisation of crime places led to the use of micro-geographical units for crime analysis which served, in turn, to develop the Criminology of Place.

The Criminology of Place is concerned with the study of crime events in micro places that constitute the nexus between physical and social environments (Sherman et al., 1989). The reason being that understanding where and when particular crimes occur is critical to their control. But what is the relationship between crime and place? And why is it important to study crime events at the micro level? Like Sherman and colleagues point out, "[t]here is little point in examining variation in crime by place, of course, if such variation is merely random" (Sherman et al., 1989, p. 33). However, if opportunities to commit crimes are not randomly distributed in places, as Environmental Criminology theories suggest, then it is important to know how they are distributed —and why— to inform crime prevention. Furthermore, studying crime events in micro geographical units can help to establish causal relationships between the characteristics of specific places and their likelihood to host crime. This was the research line initiated by Sherman and collaborators (1989). To determine whether crime is concentrated by chance, the authors used police call data as a proxy measure for analysing the distribution of crime on micro places in Minneapolis. Their research revealed that 50% of the calls to the police came from solely 3% of the micro places in the city, although such concentration varied by crime type. Since then, the Criminology of Place has demonstrated its strength by showing similar results over the years despite of being applied to different contexts and using different micro units of analysis (e.g. building blocks, street segments, postal addresses, grids) (Weisburd et al., 2016). Research on crime concentration in micro places has even led to the premise being stated as a scientific law (Weisburd, 2015) —if such a thing can exist in a social science like Criminology—. These micro places where most crimes concentrate are commonly known as hot spots of crime (Sherman et al., 1989).

A crime hot spot is formed because crime occurs repeatedly at a particular micro place. Alternatively, there are also temporal hot spots when crime is concentrated in narrow time frames. However, it is usual to study both dimensions together to determine the presence of spatiotemporal crime patterns that define hot spots. In the end, crime concentration in hot spots is due to a disproportion between the causes that produce it and the results of its occurrence. This axiom involving anomalous distribution is known as the Pareto Principle, sometimes referred to as the 80/20 rule. Although crime data do not always faithfully reflect this 80/20 rule, they do come very close. In Eck's words: "A few targets, places, or offenders are involved in a large proportion of the problem events, and all problems involve repeat offending, repeat victimization, repeat places, or some mixture of these repeats" (2003, p. 88). The phenomenon of repeat victimization that prompts this mathematical disproportion has been observed in both places and people and for both violent and property crimes (Farrell & Pease, 1993). While this is a problem for those affected, it also represents an opportunity for crime prevention. If we are able to identify the targets that suffer the most crime, as well as the offenders that produce the most crime; if we are also able to protect the former and incapacitate the latter, then we can achieve a significant reduction in crime and its impact.

Research on traditional crime has already proven the usefulness of combining Environmental Criminology theories and their practical application through the postulates of Criminology of Place for crime prevention (Bruinsma & Johnson, 2018; Weisburd et al., 2016; Wortley & Townsley, 2017a). But crime evolves and takes advantage of new opportunities such as those offered by technological advances. It offers no respite from preventive measures. Cybercrime challenges academics and practitioners who pursue crime prevention because not only does it possess distinctive characteristics from traditional crime, but also because the environment in which it occurs is different. It appears that a fundamental question must be asked: Does ECCA have a place in preventing crime committed in cyberspace?

1.3 Objectives of the thesis

This doctoral thesis pursues the application of the ECCA approach in general, and the concept of cyber place in particular, to study and prevent different crime events that occur in cyberspace; namely, website defacement, match-fixing, online harassment, and online hate speech. Achieving this main objective requires designing a two-stage research process, the first theoretical and the second empirical. The theoretical stage aims to develop the ECCA approach through two secondary objectives: the review of the adaptation of Environmental Criminology theories to cybercrime, and the transposition of ECCA's propositions into cyberspace. This will be achieved through a comprehensive exercise of literature review and theoretical reflection. In turn, the empirical phase seeks to test the application of various hypotheses derived from the developed ECCA approach through four studies. The most adequate crime analysis techniques are then implemented through Data Science to study each crime event and propose measures for their prevention. A schematic representation of the objectives of the thesis and its structure is illustrated in Figure 2.

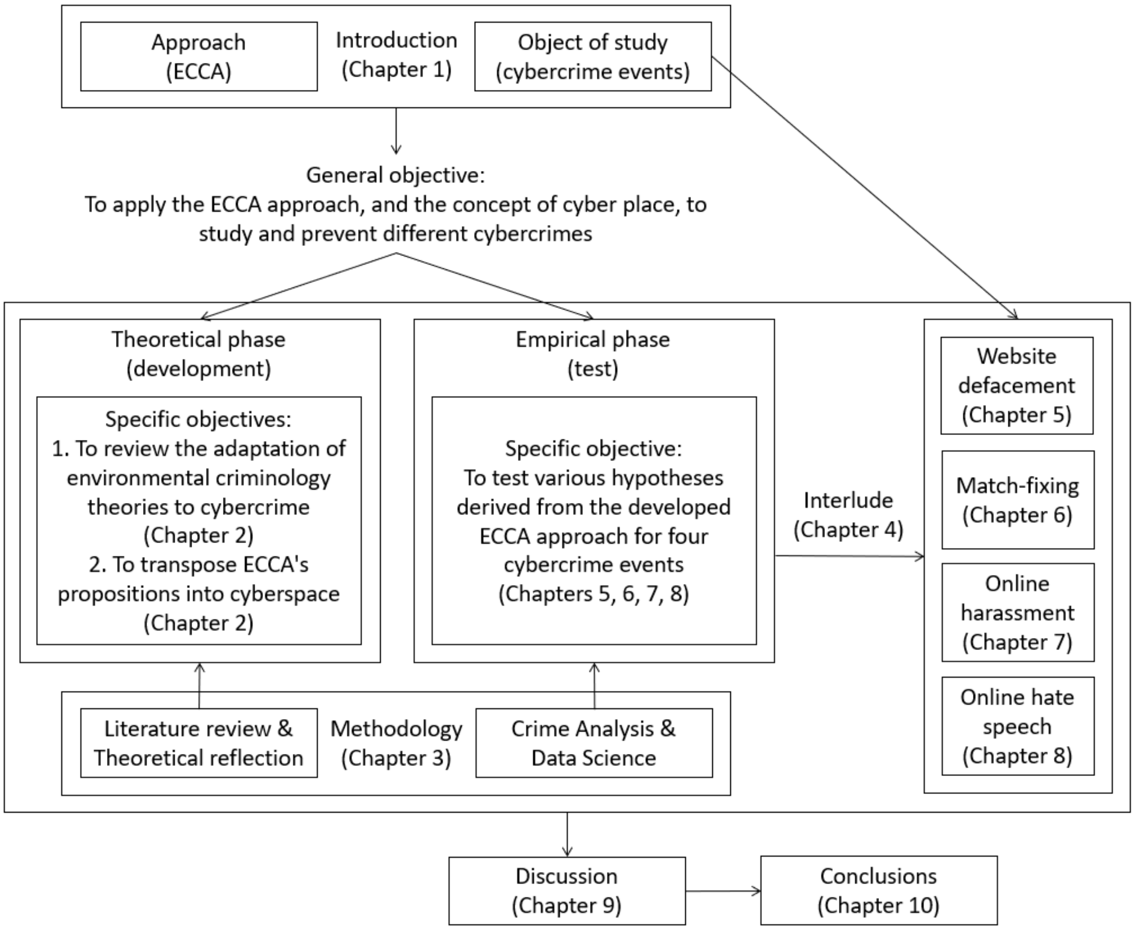
Approach (ECCA)
Introduction (Chapter 1)
Object of study (cybercrime events)
General objective: To apply the ECCA approach, and the concept of cyber place, to study and prevent different cybercrimes
Theoretical phase (development)
Specific objectives: 1. To review the adaptation of environmental criminology theories to cybercrime (Chapter 2) 2. To transpose ECCA's propositions into cyberspace (Chapter 2)
Empirical phase (test)
Specific objective: To test various hypotheses derived from the developed ECCA approach for four cybercrime events (Chapters 5, 6, 7, 8)
Interlude (Chapter 4)
Website defacement (Chapter 5)
Match-fixing (Chapter 6)
Online harassment (Chapter 7)
Online hate speech (Chapter 8)
Literature review & Theoretical reflection
Methodology (Chapter 3)
Crime Analysis & Data Science
Discussion (Chapter 9)
Conclusions (Chapter 10)

Figure 2. Objectives and logical structure of the doctoral thesis

Below, CHAPTER II develops the theoretical phase of the thesis, where the potential applicability of the ECCA approach to the study and prevention of crime in cyberspace is assessed. This step allows the identification of a catalogue of the specific ECCA propositions in CHAPTER III that will be addressed in each of the four articles presented in the thesis. Of course, each article develops —and then tests— its specific theoretical framework derived from the general one to further contextualize the research. After presenting the general methodological approach that will be followed to accomplish this in CHAPTER IV, CHAPTER V presents an outline of the four articles.

—Blank page—

CHAPTER II

GENERAL THEORETICAL FRAMEWORK: APPLYING ECCA TO CRIME COMMITTED IN CYBERSPACE

ECCA has proven to be versatile in addressing different crimes. It is true that this approach has been especially effective in reducing property crime, but it has also been used to counter some forms of violent crime (Welsh & Taheri, 2018). And while it is true that cybercrime can challenge the very limits of the approach, the cybercrime fallacies reveal that, despite its peculiarities, cybercrime is just crime after all (Miró-Llinares, 2015a). So, no matter whether it is economic cybercrime, social cybercrime, or political cybercrime: the object of interest is still crime. The problem that cybercrime poses to ECCA is not the phenomenon itself, but the environment where it occurs. ECCA was conceived to study traditional crime in geographic environments, not in digital environments (Miró-Llinares & Moneva, 2019a). The premises on which Environmental Criminology theories rest have a strong dependence on the geography of crime: the spatiotemporal convergence mechanism behind the chemistry of crime was devised for geographical settings (L. E. Cohen & Felson, 1979), people's activity spaces are delimited by an urban environment (P. L. Brantingham & Brantingham, 1981), and the SCP measures are designed to reduce criminal opportunities in physical spaces (Clarke, 1980). The Criminology of the Place focuses on the analysis of crime hotspots in micro geographical places (Weisburd et al., 2016), and the Law of Crime Concentration is formulated on the basis of the results of a “cross-city comparison of crime concentration using a common geographic unit" (Weisburd, 2015, p. 1). The reason is simple, ECCA is about solving crime problems, and the problems that existed in the 80s were traditional crime problems in geographic settings. Crime problems that occurred on the street were solved by intervening on the street; if the problem is not in that environment, neither are the solutions. Because cybercrime was not a problem then, there was no need to consider cyberspace.

Unfortunately, though, cybercrime is a problem today and the environment where it occurs is cyberspace. A problem for which ECCA was not prepared, at least in theory, but to which it had to adapt. As new crime problems emerged in cyberspace, environmental criminologists were requested to solve them. For example, Newman and Clarke's (2003) research on e-commerce crime analysis and situational prevention stems from an initiative by the British Department of Trade and Industry to bring together a panel of experts to discuss the emerging threats posed by 21st century crime. To face this new challenge, the authors develop the framework of the SCP against a crime that occurs in a completely new environment, while ignoring its preventive effectiveness given the lack of precedents. In their comprehensive crime analysis exercise, the authors had to examine the dynamics of e-commerce in order to understand the new crime opportunities it generates, as well as adjust classic Environmental Criminology concepts such as the acronym CRAVED to take into account the defining characteristics of cyberspace. As a result of their work, a new proposal for SCP measures applied to e-commerce crime was drafted, along with an outline of their implementation involving law enforcement agencies that operate in cyberspace (G. R. Newman & Clarke, 2003). Judging from the outcome, it appears that the authors made an important contribution in applying the ECCA framework to crime committed in cyberspace. What were the milestones of their work?

There are two actions that must be undertaken to successfully apply ECCA to crime committed in cyberspace: revise the theories and transpose their propositions. Regarding the former, previous application of Environmental Criminology theories should be revised to assess their ability to explain cybercrime events and, if necessary, identify their key concepts in order to adequately address the new object of study from a situational angle. Concerning the latter, it should be examined whether the ECCA propositions are successfully transposed to crime in cyberspace. Both actions are further expanded below.

2.1 Revising the application of Environmental Criminology theories to cybercrime

Work on applying Environmental Criminology theories to the understanding of cybercrime began two decades ago. Initial theoretical discussions (e.g. Grabosky, 2001; Yar, 2005) were followed by the first attempts at empirical operationalization of their underlying concepts (e.g. Holt & Bossler, 2008; G. R. Newman & Clarke, 2003). Step by step, the growing interest in the applicability of Environmental Criminology theories to explain and prevent cybercrime was reflected in a growing volume of studies that steadily pushed the discipline forward (Bossler, 2020; Holt & Bossler, 2016). As the theoretical discussion deepened (Miró-Llinares, 2011), the empirical analyses became more sophisticated. Yet, there is still a lot of work to be done. There are overlooked concepts whose usefulness remains to be studied and whose applicability must be measured from an empirical angle. From the theoretical to the empirical, the following sections provide an overview of how Environmental Criminology theories have been used to explain cybercrime, accompanied by some remarks on their possible future application.

2.1.1 The Routine Activities Approach: Just a cybervictimization theory?

Perhaps because the Routine Activities Approach is the backbone of Environmental Criminology theories —for without it there is nothing else— the first debates on the applicability of the situational approach to cyberspace focused on this framework. In his seminal article *Old Wine in New Bottles*, Grabosky (2001) cautioned against the magnification of cybercrime as a completely new phenomenon by arguing that, even though the environment had changed, human nature had not. Thus, he questioned whether it was really necessary to revise the theoretical assumptions about crime as the fundamental premise of convergence and guardianship also applied to cybercrime. He did, however, highlight important changes in crime opportunities in cyberspace that could challenge their control.

The reality of cyberspace, its space-time continuum, is distinct and affects the appearance of crime opportunities in a singular way. One of the first criminologists to notice this was Yar (2005). Drawing on previous research, he argues that space in cyberspace is non-existent and therefore ecological approaches that explain crime opportunities in physical space are of little use in digital environments. The logic being that in the absence of physical convergence, the chemistry of crime stops reacting. In addition, Yar (2005) points out that places in cyberspace are volatile, both in terms of lifespan and ambient population, making it difficult to apply a Routine Activities Approach. For Yar, the time dimension of cyberspace is also a drawback in applying an ecological approach to understanding criminal events. Since cyberspace is a global environment that can be accessed by people from all over the world, it is not possible to identify clear activity patterns or, consequently, to anticipate greater volumes of convergence (Yar, 2005). This argument clashes with the work of other cyberspace theorists such as Grabosky (2001; Grabosky & Smith, 2001) and Miró-Llinares (2011), who argue that, despite its unique characteristics, cyberspace is just another environment where an ecological approach to crime analysis and prevention can be applied.

Miró-Llinares (2011) agrees with Yar that distances in cyberspace shrink, but argues that this hardly invalidates convergence but rather enhances it. Since there are no distance restrictions in cyberspace, personal intercommunication —a form of convergence— becomes easier. Moreover, for this author, the reduced importance of space adds relevance to time. Time that also contracts, or at least is perceived as such. The immediacy of convergence results in a larger number of interactions in cyberspace, many of which occur simultaneously. Because it takes little time to implement actions that previously required greater effort due to geographical distances, convergence is facilitated again. In short, according to Miró-Llinares (2011), the point is not that the unique convergence experienced in cyberspace invalidates the application of an ecological framework to explain crime, but that convergence is different within this environment and must be carefully considered. Grabosky (2001) had already stated that there is no problem in using the routine activities approach to cybercrime, stressing that cyber offenders are still people, that the capacity of guardians must evolve to keep up with the new challenges posed by this environment —as they have always done— and that the first line of defence for suitable targets lies in self-protection mechanisms.

Much like the application of the Routine Activities Approach to explain traditional crime, its macro and micro premises have also been used to explain cybercrime (Bossler, 2020; Holt & Bossler, 2016). Unfortunately, the same flaws have been observed when this approach has been applied to cybercrime as when it has been applied to traditional crime: it has been widely used as a victimization theory while ignoring its other central dimensions such as the role of guardianship and, especially, the routine activities of the offenders (Miró-Llinares & Moneva, 2019a). For this reason, despite the many studies that have employed this approach for understanding various cybervictimization processes, its full potential remains unexplored (Holt & Bossler, 2016). And while it is true that the micro premise of the approach has received considerably more attention in the literature than the macro premise, both deserve to be considered. Interestingly —and contrary to its original formulation to explain traditional crime rates— it was its micro premise that was first studied empirically to explain cybercrime, so following a chronological order we shall start from there.

It has already been mentioned that the Routine Activities Approach is very convenient for researching crime at the micro level because of its simple formulation. All it takes is operationalizing the minimum elements of crime and measuring their convergence. But things may not be so easy considering that almost ten years elapsed since the first debates on its potential applicability to cybercrime and its very first actual application. It was Holt and Bossler (2008; Bossler & Holt, 2009) who managed to take this important first step and many others followed them by replicating their research design. Through a survey design, Holt and Bossler (2008) operationalised the Chemistry of Crime in the following way: target suitability was measured through time spent online by users performing various routine activities (e.g. online shopping, playing video games, using email); capable guardianship was measured in two different ways, social guardianship (i.e. deviant online behaviour of peers) and physical guardianship (i.e. updated security software or hardware) [15]; and likely offenders were measured through a scale comprised of a series of deviant online behaviours (e.g. pirating

[15] Eventually, a triple categorization of capable guardians was established in cyberspace according to their source of implantation: physical guardians, which refer to the security software or hardware installed in a device; social guardians, represented by those who exercise informal social control such as peers and relatives; and personal guardians, based on individual skills in the use of technology (Holt & Bossler, 2014, 2016).

software, pornography consumption, password guessing). In addition, some sociodemographic variables were included in the model. However, the main problem with this model is that the authors did not measure the same behaviour (i.e. online harassment) for both offending and victimization. So, despite a good exercise in operationalising the minimum elements for cybercrime, their convergence could hardly be measured (Reyns et al., 2011). Since then, many studies on online routine activities have followed the trail of these authors with varying degrees of success, amending some aspects and neglecting others (for a review, see Leukfeldt & Yar, 2016) [16].

To a lesser extent, existing scholarship has also examined the applicability of the Routine Activities Approach at the macro level. Studies that have applied this approach can be divided into three groups: those that have analysed users' routines in relation to the volume of attacks experienced in large networks (e.g. Maimon et al., 2013), those that have examined the relationship between socio-economic factors at the country level and their associated probability of experiencing cybervictimization (e.g. Kigerl, 2012), and those who have explored shifting crime trends and the adoption of technologies by people (e.g. Farrell et al., 2011). The former two groups tend to focus their analysis on where large-scale cyber-dependent crimes originate and where they are targeted to reveal patterns of user activity that allow comparisons between regions. The latter often use official data sources collected annually to explore potential correlations between changes in crime rates and the adoption of technologies over time. While possibly the most faithful to Cohen and Felson's (1979) original work, these are the less abundant.

As anticipated, this brief overview shows that the micro and macro application of the Routine Activities Approach to cybercrime is eminently focused on victimization,

[16] For an international comparison study on target suitability among Spanish and Australian Internet users, see also the work of Miró-Llinares, Drew and Townsley (2020). According to this study, although the prevalence of cybervictimization is similar in both samples, the behavioural correlates are not.

be it at the individual or regional level. In this sense, “some scholars may feel that researchers have exhausted this issue with respect to cybercrime victimization” (Holt & Bossler, 2016, p. 73). However, there are still a number of issues —many of them theoretical— that could receive more attention. From the micro paradigm, the role of the different types of guardians in cyberspace still needs to be unravelled as there remain important unanswered questions (Reynald et al., 2018; Vakhitova et al., 2016), such as who are the handlers and the place managers, or what kind of guardianship is most effective in preventing what kind of crime (Vakhitova et al., 2019) [17]. There is also a strong need to consistently measure the offending-victimisation dynamics in specific settings and for specific behaviours to ensure a proper analysis of convergence. In the end, convergence is the cornerstone of this approach. The main difference in terms of prevention is that, despite the offender's ability to converge with targets in different ways, there is a great capacity in the latter to mitigate the impact of cybercrime. Research from the macro paradigm would benefit from revisiting the concepts of human ecology of rhythm, tempo, and timing (Hawley, 1950) on which Cohen and Felson (1979) built their approach. Few researchers reflect on the ecology of crime and security despite its central role in Environmental Criminology theories (Vozmediano & San Juan, 2010). Not to be exhaustive, these are just a few points that, despite the perceived stagnation, would serve to refresh the approach and give renewed impetus to cybercrime research.

[17] A different matter is what makes guardians exert guardianship. In a study interviewing residential guardians, Reynald (2010) synthesizes the information gathered into interesting findings. First, this author finds that the available guardians monitor the environment depending on their sense of responsibility and how they perceive security in it. Second, whether they detect suspicious activity depends on whether they received training and how familiar they are with the environment. Third, guardians intervene or not ─either directly or indirectly─ depending on their sense of responsibility, their physical competence, the availability of tools for their protection, and the severity of the incident to be disrupted. This is important work in terms of crime prevention, as it identifies the key factors for enabling guardianship in particular contexts. A similar exercise would need to be carried out on specific places to better understand cyber guardianship.

2.1.2 The Rational Choice Perspective through SCP

If the Rational Choice Perspective explains people's decision-making processes related to criminal involvement, it should be appropriate for both traditional crime and cybercrime. But of course, such assumption must be tested. Perhaps this is what Higgins (2007) thought when designing his study on the influence of rational choice on digital piracy behaviour. Employing a survey design, the author sought to test whether the rational choice of college students mediated the link between low self-control and digital piracy. The results of the factor analyses showed that the situational factors measured as a subjective measure of the utility of digital piracy (i.e. value) have both a direct and indirect effect on the actual behaviour. Based on previous research, Higgins argues that, despite the few hundred participants in the study and the demographic range chosen, the findings are valuable because of the propensity of youth to engage in digital piracy. This early study was important because it not only showed "that low self-control and rational choice theory maybe compatible theories that can explain digital piracy" (Higgins, 2007, p. 48), but also highlighted the influence of situational factors in crime opportunities. As a result, this study paved the way for exploring the usefulness of SCP for cybercrime.

There are two approaches to exploring the adaptation of SCP measures to cybercrime: the theoretical and the empirical (Holt & Bossler, 2016). Theoretical research on SCP is grounded in reflective processes that address the potential usefulness of such measures to prevent a specific cybercrime whose nature is already well-known. The understanding of the phenomenon is often supported by a review of the literature. For example, Reyns (2010, p. 107) proposes a series of SCP measures that address the two main issues he considers key "in avoiding cyberstalking: exposure and communication". To control online exposure, Reyns (2010) suggests various measures to increase the degree of anonymity of users by limiting the personal information they post (e.g. not making their email address available to unknown parties); and to reduce harmful communications, another set of measures aimed at filtering information received by users online is presented (e.g. not accepting messages from unknown parties). For the sake of usability, SCP measures are then classified among those that can be used by the victims themselves, and those that can be implemented by place managers, depending on the type of cyberstalking to be prevented. Note that no prior analysis of the effectiveness of the measures has been conducted, but rather the proposal is based on the researcher's extensive experience in the field and a short review of studies on cyberstalking [18].

Afterwards, Miró-Llinares (2012) advocates a deeper adaptation of the SCP measures to cybercrime, which included structural changes in the original matrix. At one level, this author proposes suppressing the “reduce provocations” category. The rationale being that the measures it includes focus on emotional aspects —mainly related to the offender— but that cyberspace favours the depersonalisation of the victim, where emotions have little effect. And that although the author acknowledges that such measures can be useful in preventing certain cybercrimes (e.g. online harassment, sexting), they do not represent an added value compared to others. In addition, Miró-Llinares (2012) suggests introducing a new category that could be termed “reduce the influence”. In line with the narrative of the author throughout his book, this category would encompass a set of measures to be implemented by the victim in order to reduce

[18] Other studies in this category have studied the applicability of SCP measures to information security problems such as general vulnerabilities (Hinduja & Kooi, 2013), phishing, auction fraud (Hartel et al., 2011), or insider fraud (Willison, 2000); financial cybercrime (Leukfeldt & Jansen, 2020); and online child sexual abuse (Krone et al., 2020; Wortley, 2012; Wortley & Smallbone, 2012), among other crimes. Although these works have in common the adoption of the same preventive framework, an interesting fact is that they originate from different disciplinary backgrounds (i.e. criminology, criminal justice, computer science, economics, sociology, psychology), which demonstrates its wide acceptance in academia. Yet another proof of the transdisciplinary nature of Crime Science (Wortley et al., 2018).

his exposure and mitigate any potential harm. More specifically, this set of measures would be aimed at impairing the offender's target selection process by reducing the available targets. The SCP model proposed by Miró-Llinares (2012) for cybercrime therefore represents an important shift in the traditional view of implementing the measures, as it places strong emphasis on the self-protection of the victim.

Another way to approach research on SCP for cybercrime is empirical, either through qualitative or quantitative methodologies. Qualitative methodologies are usually based on a previous study of specific crime scripts. This allows for the identification of the offenders' decision-making processes that are most susceptible to intervention (Holt & Bossler, 2016). This is precisely what Hutchings and Holt (2015, 2017) accomplished in their research on online stolen data markets. By using crime scripts on the content of 13 stolen data forums, the authors are able to understand the interaction dynamics of the actors involved in the marketplace, and identify the specific actions they perform (Hutchings & Holt, 2015). It is this initial effort that lays the groundwork for their subsequent research on SCP. In a second paper, Hutchings and Holt (2017) propose a series of disruption initiatives and intervention approaches aimed at both the event itself (e.g. authentication systems), the actors involved (e.g. generate distrust), and the marketplace (e.g. domain deregistration). Neither have these studies, like the previous ones, evaluated the effectiveness of the proposed SCP measures. In fact, there are few quantitative research designs that evaluate SCP measures and they do not always use this framework explicitly. Still, research on the effectiveness of antivirus products and warning banners can be included in this category (Brewer et al., 2020).

The use of antivirus software can be considered as an SCP measure aimed at increasing the effort of offenders when attacking a computer system by target hardening. But evaluating its effectiveness can be tricky. On the one hand, if detection tools are used, a sample selection bias can be incurred because some participants are only equipped with scanning tools; on the other hand, self-reporting measures can be used, but they may lead to inaccuracies (Brewer et al., 2020). A good alternative is natural experiments. In a conference paper, Lévesque and collaborators (2011) presented the results of an experimental design used to measure the effectiveness of the default antivirus products installed on Microsoft machines (i.e. Microsoft Windows Malicious Software Removal Tool and Microsoft Windows Defender) —the controls— versus third party antivirus —the trials—. After monitoring approximately 27 million computers for 4 months, the authors found that despite the good preventive performance of Microsoft's products, third party antiviruses were more effective in preventing malware infections [19].

Warning banners are a deterrent mechanism that can be considered as an SCP measure to alert conscience of offenders under the category of remove excuses [20]. A major advantage of researching their deterrent effectiveness is that they are relatively easy to design and implement. That said, their design must be flawless in order to simulate a genuine stimulus for users. In a double experimental design, Maimon and collaborators (2014) tested the effect of a warning banner on unauthorized access to computer systems through the use of honeypots. Honeypots are computers deliberately programmed with certain vulnerabilities and prepared to collect information from trespassers that exploit them. By randomly assigning trespassers to the experimental warning banner stimulus [21], the authors found that its deterrent effect was insufficient to

[19] It should be noted that one of the authors of the study was working for Microsoft at the time of publication and another was a former employee of the company.

[20] According to its design and the text it displays, a warning banner could also be considered a set rules or post instructions SCP measure.

[21] The warning banner displayed the following text: “The actual or attempted unauthorized access, use, or modification of this system is strictly prohibited. Unauthorized users are subject to institutional disciplinary proceedings and/or criminal and civil penalties under state, federal, or other applicable domestic and foreign laws. The use of this system is monitored and recorded for administrative and security reasons. Anyone accessing this system expressly consents to such monitoring and is advised

reduce the number of events, but enough to reduce their duration. Similar results were also reported in a replication study (Stockman et al., 2015). Further research using available data from the first experiments revealed that the deterrent effects of the warning banner were limited by the type of trespasser involved (Testa et al., 2017). While trespassers capable of hacking a network with administrator privileges did not reduce the amount of harmful commands used against the system, those without administrator privileges did. These results show that warning banners can be effective in reducing the duration of trespassing incidents produced by less skilled hackers, but that they may be ineffective against more proficient offenders.

Overall, even though few studies have yet been conducted on the effectiveness of SCP measures applied to cybercrime, these show promising results. On a positive note, recent studies using quantitative methodologies are using experimental designs, which allows the effect of the SCP measures to be rigorously evaluated. As an aspect to be improved, the amount of existing studies is still reduced, and the variety of measures evaluated is rather small. The main problem that hinders progress in this field is the complexity of working in controlled digital environments, which require sophisticated research designs, a large amount of resources, collaboration with private third parties that have control over them, and often also interdisciplinary collaboration with computer science experts. Despite the above, current research is moving in the right direction and increasing attention is being paid to the usefulness of SCP in controlling cybercrime. And this is no small thing to say, since this privilege is not something that all Environmental Criminology theoretical bodies enjoy.

that if monitoring reveals possible evidence of criminal activity, the Institution may provide the evidence of such activity to law enforcement officials" (Maimon et al., 2014, p. 41).

2.1.3 The forgotten ones: The Geometry of Crime and the Crime Pattern Theory

The most neglected theoretical framework of Environmental Criminology is the Geometry of Crime. If both the Routine Activities Approach and the Rational Choice Perspective —mainly through SCP— are so popular among cybercrime researchers, why is not the Geometry of Crime? Possibly, the main reason is that, when reflecting on the fundamental concepts underpinning the Geometry of Crime, they are often unconsciously associated with a geographical environment. Not by chance, one of the main tasks with which Environmental Criminology has been entrusted is the study of the geography of crime (P. J. Brantingham & Brantingham, 1981). However, when such concepts are studied in depth, one realises that while indeed some are apparently geographical, most are purely spatial (Miró-Llinares & Moneva, 2019a).

For example, paths in the context of the Geometry of Crime refer to a route traced in the urban environment, a network of street segments that connects two geographical points; the distance decay principle states that the more distance an offender travels from an anchor point, the lower the probability of committing a crime; hot spots are used to identify concentrated crime events in micro-geographical locations at specific times, and so on (P. L. Brantingham & Brantingham, 1981). However, with a slight change of perspective it could be argued that these are apparently geographical concepts, but in fact they are spatial. In cyberspace, two activity nodes may also be linked by a path, but it will not be geographic. If a user frequently visits a certain social media network and then accesses a digital newspaper to read the latest headlines, then the path may be the sequence of clicks he had to perform in order to travel from one activity node to another. An offender may not fatigue when travelling distances in cyberspace, but it is possible that such effort must be measured in terms of time resulting in a similar decay function. And hotspots in cyberspace may not form in micro-geographical locations, but they may form in digital microenvironments (Miró-Llinares et al., 2018). Slight modifications serve to adapt these concepts to explain cybercrime events.

In addition, there are other concepts within the framework of the Geometry of Crime that need almost no adaptation because, by definition, they are purely spatial. For example, the environmental backcloth in cyberspace is also defined by the static and dynamic elements of a particular digital environment; as in geographic space, activity spaces in cyberspace can be shaped by the cyber places that people routinely visit and the hyperlink paths that connect them; and, over time, users will become familiar with these environments which will then become their awareness cyberspace; and when they leave their comfort zone and visit underground websites whose users have a different socio-cultural background, they can be involved in conflicts on such digital edges. It would appear there is no concordance between the non-use of this framework and the adaptability of its concepts to the study of cybercrime, and yet this remains the case. We call this perceived barrier the geographical gap (Miró-Llinares & Moneva, 2019a).

The first step in adapting the Geometry of Crime to study cybercrime involves overcoming the geographical gap. Only by overcoming this barrier will there be more research based on the Geometry of Crime, which will in turn lead to an increased use of Crime Pattern Theory to understand cybercrime patterns and will ultimately impact on the advancement of the discipline. However, until more attention is paid to the Geometry of Crime, the Crime Pattern Theory cannot develop its full potential in cyberspace. Just as the Chemistry of Crime needs its three minimum elements to function, the Crime Pattern Theory —as an integrative theory—needs to be nourished by its three essential frameworks to be complete (i.e. the Routine Activities Approach, the Geometry of Crime, and the Rational Choice Perspective). If the premises of each of these frameworks remain consistent when they are used to explain cybercrime, then they can be considered valid. Such premises are reflected in ECCA's propositions and their validity would constitute the first milestone in applying ECCA to explain cybercrime events. To assess the extent to which the ECCA's basic propositions hold true for crime committed in cyberspace, it is necessary to revisit their origins and review each of them.

2.2 Transposing the ECCA propositions into cyberspace

The particularities of the new object of study and those of the environment in which it manifests should not be a problem; if the foundations of the approach are sound, the answer should be yes: ECCA ought to have applicability also to crimes committed in cyberspace. But first, of course, the distinctive aspects of both elements (i.e. cybercrime and cyberspace) should be taken into consideration. And if, despite of that, the three propositions of ECCA remain in place, it can be assumed that the approach continues to be valid. Following this logic, the question of whether criminal behaviour is still substantially influenced by the environment in which it occurs —in this case a digital setting— must be asked first. Answering this question requires to understand the nature of cyberspace, its structure, its space-time continuum. Then, an analysis of how these characteristics can influence human and, therefore, criminal behaviour is needed. Second, it is necessary to determine whether the spatiotemporal patterns described by cybercrime are random or rather subject to crime opportunities emerging from the dynamic environment that concentrates them. A circumstance that will be strongly determined by the type and nature of the everyday activities that people carry out in the places of cyberspace. Third and last, it is imperative to understand the practical relevance of all of the above. If the interaction of people and objects with digital environments produces crime opportunities, then it is possible to control such opportunities by manipulating the environment. In this sense, retrieving the terminology employed in Environmental Criminology literature, we refer to the concept of place — or cyber place in this case— (Table 2). Cyber places can be defined as “discrete nodes or areas of activity on the Internet where one is not physically located but can nevertheless act” (Miró-Llinares & Johnson, 2018, p. 893).

Table 2.
Key issues in transposing the ECCA propositions to crime committed in cyberspace

ECCA propositions (Wortley & Townsley, 2017b, p. 1)	Key issues for application to cybercrime
“Criminal behaviour is significantly influenced by the nature of the immediate environment in which it occurs.”	What is the nature of cyberspace? How does the structural nature of cyberspace affect criminal behaviour?
“The distribution of crime in time and space is non-random.”	Is cybercrime randomly distributed in space and time?
“Understanding the role of criminogenic environments and being aware of the way that crime is patterned are powerful weapons in the investigation, control and prevention of crime.”	Why do cybercrime patterns appear? What role do digital environments play in cybercrime causation? What is the function of place in cybercrime prevention?

2.2.1 The (non-)physical nature of cyberspace

People live their lives in different realities: in their imagination, on the streets, or in cyberspace. There are those who prefer to dream of a better job, those who prefer to go out and look for it, and those who would rather do it through cyberspace. The human perception of a moment in reality is defined by a time and a space. There is no time without space and no space without time. Whatever the reality of everyone, people interact with their environment differently in specific moments. While navigating the physical reality of the street, a person interacts in one way with his car in the garage in the morning to go to work and does it differently in the office with his computer in the afternoon to get the job done. It is true that a person can do two things simultaneously, but then both actions will share a single moment. Due to the dynamism of reality, different moments imply different environments which, in turn, condition the appearance of distinct crime opportunities. This was the underlying message left by Brantingham and Brantingham (1993b) when they coined the concept of environmental backcloth. Yet the environmental backcloth operates in a unique way in cyberspace.

So, what are the implications of a different space-time, a different environmental backcloth, for crime perpetration? The first important aspect is that an offender can be located in as many places as there are possible convergences with suitable targets (Miró-Llinares, 2011). This way, a single offender can be ubiquitous, converging with multiple targets and generating in turn many crime opportunities. Similarly, an offender can take advantage of this circumstance to engage the same target from different places, increasing the likelihood of success. For example, this happens when multiple cyber-attacks are executed from a botnet that connects different computers. Note that cyberspace allows such convergence to occur at the same moment or at different times. Another relevant issue is that the effects of actions in cyberspace can be permanent as opposed to expiring actions in physical space (Miró-Llinares, 2011). This would enhance the harmful effect of certain actions, as they could be reactivated indefinitely. Because it is the target itself who determines its degree of exposure in cyberspace, this is where the role of the victim becomes central. Imagine a download button hosted on a website that contains latent malicious software and is clicked over and over again by unwary individuals: A single criminal action that would be causing multiple victimizations at different moments. Finally, it is worth mentioning the role of the target in spreading the harmful effects of certain cybercrimes (Miró-Llinares, 2011). Like the spread of a contagious disease, targets may unconsciously serve as multipliers of cybercrime. It is common to spread a hate speech message on social media with the intention of making it visible and publicly report it, but such dissemination will generate new victimizations if it is received by a vulnerable person. Only the capable guardians of cyberspace can prevent crime in such cases.

Undoubtedly, these intrinsic features of cyberspace pose a challenge to capable guardianship (Grabosky et al., 2001; Yar, 2005) [22]. Beyond the legal conflicts that arise in determining the jurisdiction responsible for pursuing a cybercrime due to the transnationality of cyberspace (Grabosky, 2001), capable guardians face new challenges. Even more so, in a virtual setting where crime seems more unpredictable and where offenders are harder to control, a question arises as to whether there really are capable guardians in cyberspace (Yar, 2005). There are three ways of exercising guardianship in cyberspace: through formal control, social control, or self-protection (Bossler & Holt, 2009). For Yar (2005, p. 423), maintaining formal control in cyberspace "is well nigh impossible, given the ease of offender mobility and the temporal irregularity of cyber-spatial activities". Even though the issue of the irregularity of routine activities cannot be corroborated yet —and may never be—, the former part of such statement seems accurate. How can this problem of capacity be solved? One possible course of action is to rely on third parties to assist law enforcement agencies in their efforts (Grabosky et al., 2001). Just as corporations are involved in preventing traditional crime, they can also be committed to preventing cybercrime. In fact, service providers are sometimes solely responsible for the content that millions of users generate in their private environment when interacting. Such a volume of interactions generates countless crime opportunities that cannot be controlled with human force. To deal with such a massive threat, third parties must employ automatic detection tools for harmful content capable of filtering out as much noise as

[22] So do the extrinsic features of cyberspace, such as offshoring, transnationality, neutrality, decentralisation, universality, popularisation, anonymity, openness, and ever-changing (Miró-Llinares, 2011). However, we believe that the extrinsic features are partly a consequence of the intrinsic ones and therefore will not be addressed here in detail.

to identify the rotten apple and remove it. Predictive algorithms have proven to be quite effective (e.g. Burnap & Williams, 2015), although they pose ethical-legal dilemmas that are currently under discussion (Mittelstadt et al., 2016). Alternatively, other forms of crime control can be sought.

The most obvious, given its relevance to Environmental Criminology theories, would be social control. In cyberspace, social control is exercised by the users themselves, who witness the online activity of their peers and can reinforce or recriminate it with their reactions. This is the role of handlers in the crime triangle (Eck, 2003). In terms of scope, the effect of peers can reach every corner of cyberspace, as a single user can monitor the activity of thousands. However, the effectiveness of the social norm is undermined by the diffusion of responsibility on the one hand and the anonymity provided by cyberspace on the other (Wortley, 2001). This puts a lot of weight on self-protection systems, the last barrier of guardianship (Grabosky, 2001). Regardless of space and time in cyberspace, of whether many convergences create many crime opportunities, or of whether offenders can launch massive automated attacks, one can always resort to self-protection. Self-protection in cyberspace includes adopting certain safe behaviours, such as blocking users or content, refraining from visiting certain websites, not downloading certain content, or simply keeping security software such as anti-virus or firewalls up to date. Their effectiveness is a different matter.

2.2.2 Cybercrime patterns

According to Environmental Criminology theories, crime is a possible result of the convergence between people and things as they engage in daily activities. This everyday routine is what determines the emergence of crime patterns (P. J. Brantingham & Brantingham, 1981, 1984). We also converge with people and things as we perform our increasingly extensive online routine and, as a result, cybercrime may occur (Miró-Llinares & Moneva, 2019a). Conditioned by their routine, people spend more time in some digital environments than others and do so at specific times. Circadian rhythms determined by time zones mean that, generally, when people go online in Europe, people in Oceania are offline. Even people who are awake at the same time may visit different digital environments because —just like visiting physical environments— it satisfies a range of needs. No wonder people spend more time on their email account during their work hours and more time on social media during breaks (Li et al., 2013). This not only creates different geographic patterns of online activity, but also digital crime patterns.

However, the unique intrinsic characteristics of cyberspace cause online convergence to differ from that in physical space, and this —in turn— influences cybercrime concentration. Because offenders no longer need to travel geographic distances to commit crimes, cybercrimes will not necessarily be concentrated near their anchor points but may be distributed across a wide range of digital environments. Similarly, because it takes less time to move between cyber places, more volatile temporary activity spikes can be formed. This produces cybercrime patterns that are different from those observed for traditional crime, but patterns nonetheless. It may seem that the absence of physical restrictions makes activity in cyberspace less predictable. But are online routines really less predictable? As long as human behaviour remains subject to routines, also online, crime will describe patterns and be predictable. Identifying such routines through the analysis of human behaviour in cyberspace is therefore the key to the issue. It is possible to establish two broad categories of studies on cybercrime patterns: studies on online routine activity patterns, and studies on repeated victimization patterns.

2.2.2.1 *Online routine activity patterns*

Online routines can be reflected in human activity at any aggregation level (i.e. macro, meso, and micro). At the macro level, everyday life can influence the times and places from which cybercrime is most often executed and received, a scenario particularly studied for cyber-dependent crimes such as hacking (Holt, Leukfeldt, et al., 2020; Maimon et al., 2013, 2015). To examine such spatiotemporal patterns on trespassing events and their relationship to the users' routine activities, Maimon and collaborators (2015) collected data from public IP addresses of Chinese and Israeli universities and deployed honeypots in their computer network to register cybercrime events. Two main findings of the study are highlighted. First, based on the assumption that users from nearby geographic regions would share online routines and behaviours to a greater extent than those from distant regions, the authors found that while Chinese IPs were commonly attacked from the same region, Israeli IPs were not so much (Maimon et al., 2015). Similar results were found by Holt and collaborators (2020) in their study on the motivation of hackers to deface Dutch websites. These authors found that politically motivated defacements and those executed as a personal challenge were more likely to target Dutch IP addresses, suggesting that this trend “may be a function of perceptual differences in the nature of these targets” (Holt, Leukfeldt, et al., 2020, p. 15). Second, it should be noted that —surprisingly— Maimon and collaborators (2015) do not expect cybercrimes to describe any daily time pattern [23]. And, as a matter of fact, the authors do not find any daily time pattern between the first trespassing events in the system and the routine activities of the network users. They attribute this to the purely volitional criteria of the offenders as it “seems to depend solely on trespassers’ decisions of when

[23] The reason we say this is surprising is because environmental criminology research has consistently found time patterns of crime at the micro level (e.g. Lersch & Hart, 2015). Instead, it would have been logical to pose as an alternative hypothesis the emergence of daily time patterns of trespassing.

to initiate a trespassing incident" (Maimon et al., 2015, p. 630). This means that although there may be similarities in the online routines of users from nearby geographic regions, there are insufficient grounds to explain the formation of cybercrime patterns.

At the meso level, patterns of routine activities can be found in organizations and in the collective use of certain online services. At the organizational level, Maimon and colleagues (2015) assume that universities will suffer more attacks during office hours (i.e. when there are more employees working and therefore more activity on the network). In line with their previous work (Maimon et al., 2013), the authors observe that cybercrimes are concentrated during office hours (Maimon et al., 2015). Regarding the use of online applications and services, another cross-national study involving four countries analysed several online routine activity models that failed to reveal significant common predictors of cybervictimization except for the use of social media, which is associated with a higher exposure (Näsi et al., 2017). Interestingly, informal social control, as measured by the number of Facebook friends, seemed to have no effect on cybervictimization. Deepening the relationship between the use of social media and the increased likelihood of victimization, Choi and Lee (2017) reported that such risk was associated with engaging in specific behaviours such as publishing habits, opinions, and personal information. Other studies in the same line show that carrying out certain online routine activities is related to specific forms of cybercrime: online banking, online shopping, messaging, and downloading all appear to be associated with an increased likelihood of identity theft (Reyns, 2013); downloading, and using dating sites seems to favour victimization by malware infection (Holt, van Wilsem, et al., 2020); and online shopping may be associated with experiencing consumer fraud (Pratt et al., 2010; van Wilsem, 2013a). Despite the different routines analysed and the multiplicity of digital environments examined, from the results it seems indeed that the more convergence, the more cybercrime. However, looking collectively at cybercrime events in relation to users' everyday activities at meso level, there seem to be few clear crime patterns and many mixed results (for a review, see Leukfeldt & Yar, 2016).

At the micro level, individual users' routine activities are reflected in their communication patterns through, for example, social media. It is the conjunction of such communicative interaction in specific places and moments which allows the observation of activity patterns that, in turn, reveal cybercrime patterns. Due to its —generally— open data policy, the social media Twitter has supplied data to several studies that examine deviant behaviour related to online communication such as hate speech, and fake news (e.g. Grinberg et al., 2019; Williams & Burnap, 2016). For example, Williams and Burnap (2016) used the Twitter streaming Application Programming Interface (API) to gather information on user reaction after the Woolwich terrorist attack in 2013. One of the main aspects on which the authors focus their analysis is on the propagation and survival of online hate speech over time. Using statistical models, they were able to "determine the escalation, duration, diffusion and de-escalation of cyberhate and non-cyberhate information flows" (Williams & Burnap, 2016, p. 232), thus revealing cybercrime time patterns. The authors found that the spread of online hate speech peaked shortly after the event and that it persisted for a short time, which is consistent with a massive reaction from users on social networks immediately after the event occurred in physical space. Grinberg and collaborators (2019) also found strong time patterns in their study on fake news dissemination during the 2016 United States presidential election. Their research shows that the prevalence of fake news was steady during the months leading up to the election, increasing slightly during the two weeks preceding and dropping heavily immediately after the election date. In both cases, cybercrime patterns are related to the occurrence of a relevant event in physical space, which underlines the importance of paying attention to the interconnection between both worlds.

2.2.2.2 *Repeat victimization patterns*

Besides online routine activities, one of the most obvious forms of cybercrime patterns is that produced by repeat victimization. Whether at the macro, meso or micro level, repeat victimization —by definition— generates crime patterns. The mechanism is the same as in traditional crime: either due to particularly prolific offenders or exceptionally criminogenic places, certain targets concentrate an unusual volume of victimizations. These are known as the "boost" and "flag" explanations (Johnson, 2008a; Pease, 1998). Note that both explanations may overlap (Farrell, 2015). The boost explanation accounts for offenders who have successfully committed a crime and mark their target as suitable for future attempts. Applied, for example, to online hate speech, it would serve to explain why a Twitter user repeatedly publishes discriminatory comments against a vulnerable group in the absence of a sanction from the platform's moderators. Suppose that this user offends the victim with a comment and even receives some credit from another user in the form of a like or a retweet: the crime has been successfully committed. As a result, the offender perceives an opportunity to continue targeting that user with more messages in the future. Instead of focusing on the offender, the flag explanation refers to the environment where the crime occurs. According to this explanation, certain environments possess static characteristics that constantly label them as vulnerable to crime. Here is a cybercrime example to illustrate this. Imagine that a vulnerability has been detected in the WordPress system for content management and made public. A defacer familiar with this vulnerability will be in a privileged position to successfully target any website employing this system and will easily commit repeated attacks. Whatever the explanation, the result is an anomalous concentration of crime on specific targets. Analysing such repeat victimization patterns has served to allocate preventive resources efficiently to reduce traditional crime (e.g. Braga, Turchan, et al., 2019) and it is only reasonable to expect the same for cybercrime.

While systematic and comprehensive research on repeat victimization and traditional crime served to identify a number of valuable premises to inform crime prevention (e.g. Farrell & Pease, 1993, 2018), there is little such work on cybercrime [24]. Although not based on the ECCA approach, one of such works is that of Moitra and Konda (2004). In their work, the authors analyse a collection of different cybercrime events (e.g. root break-in, log in attempt, account break-in, password file, password cracking, and many others) that impact a network between 1988 and 1995 to identify time patterns and help improve its security. One of the main findings is precisely that there is a peak of events occurring with little or no time interval between them (i.e. repeat victimization). In a more detailed year-by-year analysis —and excluding 1988 due to a very small sample of recorded events— results show that the average interval between a victimization and its repetition is between 73 and 100 days. Despite their simplicity, repeat victimization patterns revealed by such analyses are essential to understanding the dynamics of cybercrime and addressing its prevention. That said, the main problem with this research is the aggregation level of the analysis, which makes it impossible to distinguish by type of event and, consequently, to propose preventive measures that are effective. Other important limitations of this research are the outdated

[24] Some authors have claimed to conduct literature reviews on repeated victimization and cybercrime but have failed to reference their claims. In such papers, the studies generally cited bear little relation to repeat victimization or its study from an ECCA approach and tend to mention the phenomenon —if at all— only anecdotally or tangentially.

nature of the data, or the inappropriately large time frame established for some of the analyses.

In another exploratory study, and although they do not conduct a repeat victimization analysis per se, Sidebottom and Tilley (2017) analyse romance fraud patterns on dating websites. Using data collected in 2013 and 2014 by the National Fraud Intelligence Bureau, the authors analyse the distribution of romance fraud events reported to the police in England and Wales. The more than 6000 recorded events allow crime patterns to be observed. Their findings show that 14% of the websites examined account for 78% of all reported incidents, proof of a strong cybercrime concentration (Sidebottom & Tilley, 2017). One might think that the higher volume of victimization is simply due to more user traffic on these websites, but the authors show that this is not the case. There are other works on cybercrime that address repeat victimization issues, but not from an analytical time pattern perspective. In fact, most do so by addressing events that, by definition, are repeated (i.e. online harassment, cyberbullying, cyberstalking). Such studies were not included here, as they do not represent the preventive essence of crime analysis (for a review, see Reyns & Fissel, 2019).

Despite their scarcity, it appears that research on online routine activities at the micro-level and repeat victimization studies, possibly assisted by the large volumes of data they handle, do reflect clearer cybercrime patterns compared to online routine activity studies at the macro and meso levels [25]. Beyond the purely anecdotal, this trend underscores the importance of analysing crime events at the micro level, as has been suggested by the Criminology of Place for years (Weisburd, 2015; Weisburd et al.,

[25] However, this is not always an easy task. Distinguishing genuine human activity patterns from synthetic patterns generated by bots is becoming increasingly complex due to the growing sophistication of the latter (Ferrara et al., 2016). Artificial intelligence techniques are contributing to the creation of systems capable of replicating human activity with remarkable precision. Fortunately, their sophistication is not complete, and it is still possible to detect their incidence.

2016). If traditional crime research focusing on the micro level has served to examine crime patterns for prevention, it is only logical to think that cybercrime research is likely to do the same. And if future research is consistent with these findings, a major milestone for cybercrime prevention would have been reached. But for now, empirical research on the Criminology of Place in cyberspace is still in its infancy, although it has already begun to develop theoretically elsewhere (e.g. Miró-Llinares & Johnson, 2018; Reyns, 2010). The work of Miró-Llinares and Johnson (2018) is fundamental in this sense, since it lays the theoretical foundations for the application of the Criminology of Place to crime committed in cyberspace, developing the concept of cyber place while respecting its theoretical precedents. The following section analyses this work in depth.

2.2.3 Places for cybercrime prevention

Crime is concentrated in a few geographical places (Weisburd, 2015) and it also appears to be concentrated in a few cyber places (Sidebottom & Tilley, 2017). Understanding crime concentration in places is important for distributing scarce preventive resources efficiently and reducing crime. But what causes crime to be concentrated in cyber places? In the previous section we showed that both the online routine activities carried out by users and the phenomenon of repeat victimization contribute to the formation of cybercrime patterns. In both cases, cyber places play a fundamental role in explaining crime causation. In the first case, the relevance of cyber places is determined by their functionality. How places are used fulfils specific needs of daily life and human activity is therefore constrained to specific moments of the day. In the second case, places are more or less vulnerable to repeat crime depending on their structural characteristics. This is because the design of digital environments can favour or restrict the emergence of crime opportunities.

According to Miró-Llinares and Johnson (2018), cyber places constitute different activity spaces just as geographic places do. For example, one can find cyber places for leisure, such as online games; consumer places, such as online shopping platforms; or work places, such as institutional email networks. Various forms of theft and fraud are among the most prevalent crimes in online games because the economic system that fuels such cyber places is often built on digital currencies (Chen et al., 2005). Auction fraud occurs at online shopping places because the purchase and sale system is set up asynchronously (G. R. Newman & Clarke, 2003). And ransomware may be notified through the email platforms of important organisations due to the elevated value of the data they handle (e Silva, 2018). What do all these explanations have in common? The answer is simple: crime occurs where opportunity exists. The use that people make of certain cyber places determines the type of crime opportunities that emerge there and, therefore, the type of cybercrimes that are perpetrated. In turn, the use that people make of these spaces is determined by the rhythms that "influence the mix and volume of users at particular cyber places at particular times and hence the opportunities for offending" (Miró-Llinares & Johnson, 2018, p. 896). Resembling the work of Brantingham and Brantingham (1981), cyber places that function as online activity spaces would also constitute a part of the Geometry of Cybercrime to understand the distribution of crime opportunities in cyberspace.

In addition to their purpose, the convergence that cyber places enable —which is embedded in their design— is also key to the type of crime opportunities they harbour. With a few exceptions, most cyber places allow some form of communication between their users, or between their users and their administrators. In other cyber places where users converge with servers, communication is not interpersonal, but computer based. As a result of this convergence, cybercrime is likely to happen. Convergence between offenders and targets at cyber places can occur either asynchronously or in real time (Miró-Llinares & Johnson, 2018). The most common form of convergence is the asynchronous one, also known as store and forward, which involves the delayed transmission of information: a user performs an action that is stored in the cyber place and another user converges with it later on. This is the most common channel for committing crimes in cyberspace, whether through a phishing attempt via email, an auction fraud through an online shopping platform, or spreading hate speech through social media. Some other cyber places incorporate technologies that enable real-time convergence between their users, such as video calling systems. This form of convergence generates opportunities to perpetrate certain cybercrimes such as sexting or online harassment, since the exchange of audio-visual material is immediate. However, these cybercrimes are not exclusive to real-time convergence, as they can also be perpetrated asynchronously. Rather than new forms of cybercrime, the various modes of convergence generate new forms of crime perpetration.

So far, it appears that the type of online activity space determines which cybercrime opportunities emerge in it, and that the type of contact they enable determines the type of crime perpetration (Miró-Llinares & Johnson, 2018). In addition, there are other elements in the design of digital environments that flag them as vulnerable thus facilitating repeat crime. For Sidebottom and Tilley (2017), there are online systems such as social networks or online auctions that unintentionally —due to their design— create crime [26]. Specifically, there are nine ways in which these systems create crime (Sidebottom & Tilley, 2017): (1) systems can furnish rewards for crime by providing incentives for criminal behaviour; (2) systems can make crime easy if they

[26] These authors use the term system to refer “to any set of organised or consciously developed habitual human behaviours” (2017, p. 254). Note, therefore, that they do not refer in their work only to online systems, but to all kinds of systems (e.g. banking systems, navigational systems, health care systems, public transport systems).

are unguarded; (3) systems can make crime less risky by not monitoring user activity frequently; (4) systems can facilitate crime planning if they are predictable; (5) systems can disinhibit and provoke crime if they fail to control triggers; (6) systems can generate needs in their users that push them to commit crimes; (7) systems can create crime networks by bringing together likely offenders; (8) systems can teach crime if they are used to share knowledge about crime commission; and (9) systems can legitimatise crime if certain misconduct is tolerated. Let us consider online betting systems, for example. The design of these cyber places can unintentionally facilitate crime in different ways according to the previous examples. If user activity is not monitored, it is possible that multiple accounts are registered from the same IP address to exploit new user deals. If recurring advertisements are shown about the ease of winning huge amounts of money with easy bets, the need to bet impulsively is being generated. And if a strict registration system is not established, the system may legitimatise minors to engage in illegal gambling.

Newman and Clarke (2003) were also thinking in terms of systems rather than cyber places when they developed their piece on criminogenic digital environments for e-commerce. But this is largely a purely nominal issue. For these authors, some digital environments provide “situations that are imbued with attributes that make certain crimes more possible” (G. R. Newman & Clarke, 2003, p. 61). Such attributes are included in the acronym SCAREM, which stands for: Stealth, Challenge, Anonymity, Reconnaissance, Escape and Multiplicity (G. R. Newman & Clarke, 2003). Each of these attributes connects offenders to their immediate environment to describe emerging crime opportunities in cyber places. For example, stealthy situations can make offenders go unnoticed; challenging situations can motivate hackers to commit crimes; anonymous situations can cause certain people to behave irresponsibly because they cannot be identified (Wortley, 1997); recognizable situations allows the detection of vulnerabilities that can be exploited to commit crime; escaping situations make it easier for offenders to elude responsibility and harder for law enforcement agencies to prosecute them; and multiplying situations present additional opportunities to commit more crimes, and not necessarily of the same nature (G. R. Newman & Clarke, 2003). To summarize, there are online systems that create crime unintentionally by design and situations that facilitate crime in cyber places. Knowing what causes crime in cyber places, how can it be prevented?

How cybercrime can be controlled, depends on the configuration of each cyber place. According to Miró-Llinares and Johnson (2018), there are three features of cyber places that shape the guardianship exerted over them [27]: access restrictions, traffic volume, and their underlying configuration. First, access restrictions refer to whether cyber places are in the public domain or, conversely, private spaces (Miró-Llinares & Johnson, 2018). For example, there are social media that can be accessed after a simple login and videoconference meetings that require an invitation. While guardianship may prove effective in preventing certain crimes in the first case, in the second case it is unlikely to be so. Second, the traffic volume depends on the data flow at a given time (Miró-Llinares & Johnson, 2018). An example is cyberplaces crowded with people when streaming content that are emptied at the end of the broadcast. In this case, the greater the traffic, the greater the guardianship. Third and last, the underlying configuration refers to whether the cyber place is hosted on the clear web or the deep web (Miró-Llinares & Johnson, 2018). For instance, there are commercial websites that

[27] Here guardianship must be understood as "the presence of a human element which acts — whether intentionally or not— to deter the would-be offender from committing a crime against an available target" (Hollis et al., 2013, p. 76). Guardianship is not, therefore, self-protection, as one does not exercise it oneself. Nor is it social control, since the guardians must be present physically and not just symbolically. Moreover, guardianship can be effective without intent, whereas social control is exercised with the will to prevent crime (see Hollis et al., 2013).

can be accessed from any search engine and dark markets that can only be reached through dedicated links. Guardianship is likely to be enhanced in the former, as external observation is easier. It becomes clear therefore that each cyber place requires a crime control strategy that fits its nature.

Following the crime triangle model, different forms of guardianship can be exerted over each of its minimum elements to prevent crime events: offenders can be controlled by handlers, targets by guardians, and places by managers (Eck, 2003) [28]. With cybercrime it is no different. Reyns (2010), one of the cybercrime scholars who has worked most on the concept of cyber place, emphasises the role of place managers in the SCP of cyberstalking. Reyns argues that if digital environments such as websites are cyber places, "then website administrators, webmasters and designers are their place managers" (Reyns, 2010, p. 104). For this author, place managers have significant control over everything that happens in digital environments, as opposed to self-protection mechanisms such as anti-virus or firewalls (i.e. physical guardians) . Cyber place managers "can manipulate the online environment at will, limit access, and set rules for participation in the site" (Reyns, 2010, p. 104). Such is his confidence in the capacity of cyber place managers, that half of the SCP measures proposed in his work are designed to be implemented by them [29]. Miró-Llinares and Johnson (2018) add that the competencies of cyber place managers also extend to advertising and the use of cybersecurity systems available elsewhere (e.g. in browsers). Hartel and collaborators (2011) use intrusion detection systems —a software— as an example of place

[28] In practice, the role of those who discourage crime is not always so well defined, so they sometimes blend into hybrid categories (Felson, 1995). But suppose, for the sake of clarity, that places are controlled by place managers.

[29] Some examples of those measures include embedding personal identifiers into every sent email, monitoring public blogs for misuse, providing a clear code of conduct and reminders for users, and enhancing surveillance by providing more ways to report abuse. For a complete list of SCP measures for cyber place managers in relation to each modality of cyberstalking, see Reyns (2010).

managers, which would be in line with previous work not limiting the guardianship tasks to people (e.g. CCTV). And although not expressly as place managers, Newman and Clarke (2003) point out that internet service providers play a fundamental role in controlling e-commerce crime. It seems that even if there is no consensus when it comes to defining what a place manager is or what is their specific role, many authors have seen in them an important figure for preventing cybercrime.

Literature shows that many factors intervene in the causation of crimes in cyber places (Figure 3). First, we visit specific online activity spaces based on our routine activities. The configuration of these spaces determines the social control present. Then, the environmental design of these activity spaces shapes both the nature of the convergence that can happen in them and their vulnerability to crime. As a result, on the one hand, there would be criminogenic cyber places where crime opportunities proliferate, and on the other hand, safe cyber places. While in the former, crime opportunities would determine the type of crime occurring and the modality of contact the method of crime perpetration, in the latter, crime opportunities would be controlled by cyber place managers. This model reveals three ways of preventing crime in cyber places: through social control, through environmental design, and through place management. Therefore, despite the fact that many factors cause crime at cyber places —and there are many more if we look beyond the purely situational— the distribution of resources to develop preventive strategies based on cyber places would benefit from a thorough study of these three elements for crime control.

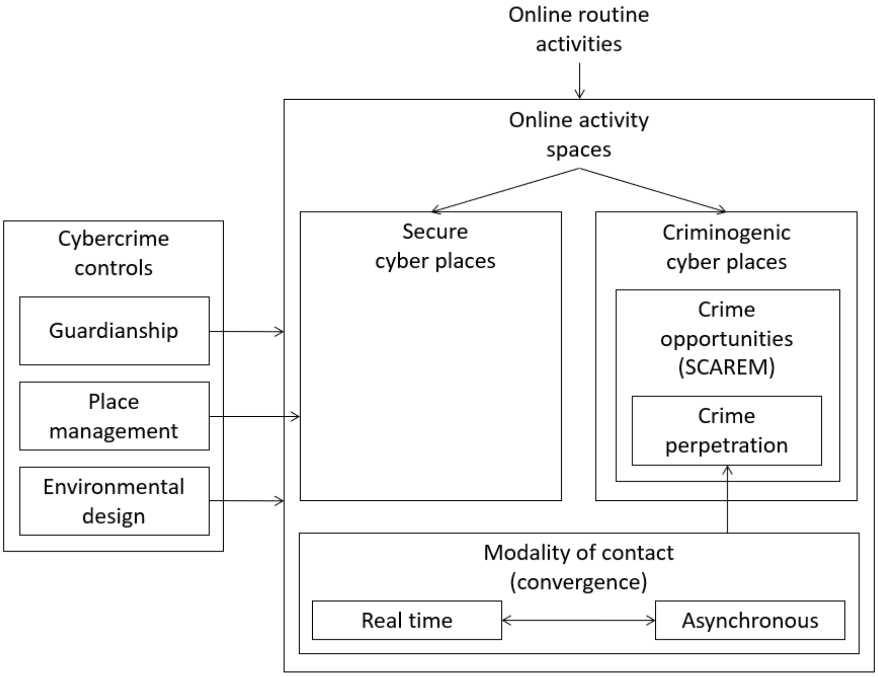

Figure 3. The role of cyber places in cybercrime causation

2.3 Overall assessment of the applicability of the ECCA approach to cybercrime

This chapter examined the applicability of the ECCA approach to understanding crime events in cyberspace. To this end, two sequential research steps were carried out. In a first step, a literature review has been conducted on how Environmental Criminology theories have been adapted to study cybercrime with the aim of detecting both strengths and current gaps. This synthesis work showed that the Routine Activities Approach was adapted more systematically and extensively at the micro level than at the macro level, but that it was generally used as an explanatory approach to cybervictimization while neglecting some of its other essential elements such as the likely offender or the cyber place where the crime occurs. Regarding the Rational Choice Perspective, most cybercrime research adapted it through SCP. In this sense, although there are many studies on theoretical proposals for concrete measures, few have evaluated them. Those that have, tend to use robust research designs and show promising results. Finally, the adaptation of the Geometry of Crime to the study of cybercrime, and consequently the Crime Pattern Theory, is virtually non-existent. The fact that most of the theoretical concepts behind these approaches are geographical has limited the creativity of cybercrime researchers. Future research should address several points: first, analyses from the micro paradigm of the Routine Activities Approach should be extended to the other minimal elements of crime, and be used more frequently from the macro paradigm; second, the SCP measures that have already been proposed should be evaluated —preferably using experimental research designs— before any new ones are suggested; and third, attention should be paid to the forgotten theoretical bodies of Environmental Criminology to extract their full potential for cybercrime analysis and prevention.

In a second step, the ECCA propositions —as articulated by Wortley and Townsley (2017b)— were examined to determine whether their adaptation to cyberspace is plausible. Should this be the case, ECCA can be considered a useful approach for analysing and preventing cybercrime. First, the spatiotemporal nature of cyberspace has been analysed to understand its impact on human behaviour online. In this regard, it should be noted that the convergence between people and people and objects in cyberspace is different from that in physical space, impacting crime opportunities. More specifically, it seems that carrying out actions in cyberspace demands less effort (i.e. they require less time and no distance to be travelled), so crime opportunities may proliferate. Second, ECCA assumes that the distribution of crime is not random, so cybercrime must also describe patterns if the approach is to be applied to cyberspace. After reviewing the empirical research, online routine activity patterns and repeated victimization patterns were found, suggesting that cybercrime distribution is not random and that the ECCA approach would therefore remain valid for cyberspace. Third and last, building on previous work, a theoretical model has been developed to understand the role of cyber place in causing crime. Furthermore, this model aims to identify the key elements for controlling crime events: environmental design and place management. These preventive mechanisms will be at their most effective when they are deployed according to the incidence of crime patterns and always taking into account the type of convergence that exists in each cyber place.

Overall, the adaptation of Environmental Criminology theories and the transposition of the ECCA propositions indicate that this is a valid approach to analyse and prevent cybercrime. The next chapter (CHAPTER III) presents the general research questions and hypotheses derived from transposing the ECCA propositions into cyberspace.

—Blank page—

CHAPTER III

GENERAL RESEARCH QUESTIONS AND HYPOTHESES

The previous chapter identified the key issues that should be addressed to determine whether the ECCA approach can actually be applied to cyberspace. In this chapter, we address the questions of why the four studies presented in this thesis have been carried out —and not others—, and how they relate to the theoretical framework developed. By doing so, we intend to contextualize the articles in a broader framework than their own. Fully addressing all the transposed ECCA propositions here is, however, an unrealistic task that may require not a thesis, but a lifetime. For this reason, each of the studies was designed to answer small but important questions that, even if they are unable to address all the major issues, would allow to lay the foundations of a more ambitious project in the future. In this sense, each article tests a set of hypotheses that have been derived from the transposed ECCA propositions. Such scheme was designed to give broad coverage to the transposition of the ECCA approach into cyberspace while providing practical solutions to crime problems.

In the original paper by Wortley and Townsley (2017b), we identified six fundamental propositions, which were then transposed into cyberspace. These issues are the research questions (RQ) that guide the empirical phase of this thesis, in which we seek to ascertain their empirical observation (Table 3). While some RQ are addressed in a recurrent manner since they are inherent to the ECCA approach, others are addressed specifically. And, particularly, one of them is addressed transversely. In one way or another, all studies are concerned with understanding the role of criminogenic environments in the causation and prevention of crime through the concept of cyber place. To inform crime prevention, we draw on Environmental Criminology theories to understand crime events and identify the situational factors that can be manipulated to reduce crime opportunities. Another recurring RQ that was tested is the non-random distribution of crime events. Drawing on different crime analysis techniques, three articles focus on the identification of crime patterns in cyber places, whether they are temporal, contextual, or configurational. By identifying crime patterns, we suggest how to better allocate preventive resources. There are also two RQ that are addressed twice. In two articles, we gain insight into why crime patterns are formed by analysing repeated offending and victimization. This permits to identify which cyber places and which people are most at risk of engaging in two forms of cybercrime. Finally, we examine what characteristics of online environments are associated with cybercrime in the other two studies. We argue that by manipulating those environmental features —metadata and precipitation-control strategies— it is possible to design more secure cyber places.

Table 3.
Transposed ECCA propositions that are empirically addressed in each article

	Transposed ECCA propositions			
Chapters that empirically address the ECCA propositions	RQ_1 Is cybercrime randomly distributed in space and time?	RQ_2 Why do cybercrime patterns appear?	RQ_3 What role do digital environments play in cybercrime causation?	RQ_4 What is the function of place in cybercrime prevention?
CHAPTER VI	X	X		X
CHAPTER VII			X	X
CHAPTER VIII	X	X		X
CHAPTER IX	X		X	X

Note: The RQs “What is the nature of cyberspace?” and “How does the structural nature of cyberspace affect criminal behaviour?” were not included in the table because they were not empirically addressed.

From these general research questions, the following four general hypotheses (GH) are derived. Regarding RQ_1, we hypothesize that:

GH_1 Cybercrime, like traditional crime, describes identifiable patterns.

To test this hypothesis, we analyse how three different cybercrimes are distributed in space, across time, or among people. Specifically, we analyse the time patterns described by website defacements, the contexts in which online harassment occurs, and the cyber micro places where online hate speech is spread.

Regarding RQ_2, we hypothesize that:

GH_2 There are certain environments that are especially vulnerable to crime.

To test this hypothesis, we explore the environments where two forms of repeat cybercrime occur. To this end, we explore to what extent website defacements are concentrated on certain websites and which are the situational contexts most likely to harbour repeat online harassment.

Regarding RQ_3 we hypothesize that:

GH_3 The configuration of cyber places determines crime emergence in them.

To test this hypothesis, we examine the characteristics of online environments that facilitate cybercrime. In particular, we examine the features of fixed match informing websites (FMIWs) that act as situational precipitators of crime and the configuration of causal recipes that produce online harassment.

Regarding RQ_4 we hypothesize that:

GH_4 There are certain characteristics of the environment whose manipulation could reduce crime.

To test this hypothesis, we identify situational factors potentially associated with the four forms of cybercrime studied (i.e. website defacements, online harassment, match-fixing, and online hate speech). These four general hypotheses in turn lead to a number of specific assumptions in each article.

In the following, CHAPTER IV describes the materials and outlines the methodology employed in this thesis to effectively implement the ECCA approach in the four studies presented from CHAPTER VI to CHAPTER IX.

CHAPTER IV

GENERAL SUMMARY OF MATERIALS AND METHODS USED

If empirical research is a fundamental component of Environmental Criminology, it is even more so for ECCA's practical problem-solving approach for reducing crime (see, originally, Goldstein, 1979). A cornerstone of this approach is that crime solutions must be supported by empirical evidence. Empiricism is based on the observation of reality. Applied to a criminological context, the observation of a crime problem allows its measurement, which in turn enables its analysis. Based on the results of the analysis, a response is articulated to solve the crime problem. A simple process for the implementation of measures based on scientific evidence. However, if any of the preceding steps are performed inadequately, the subsequent steps will be irremediably carried out in an incorrect manner. Hence, the process of empirical research must be systematic.

Aware of this, environmental criminologists and crime scientists have developed several frameworks for evidence-based problem solving such as SARA (i.e. Scanning, Analysis, Response, Assessment) (Eck & Spelman, 1987), 5Is (i.e. Intelligence, Intervention, Implementation, Involvement, Impact) (Ekblom, 2011), and VOLTAGE (i.e. Victims, Offenders, Locations, Times, Attractors, Groups, Enhancers) (Ratcliffe, 2016). These models facilitate the systematic acquisition of knowledge and generate an evidence base for solving crime problems which provide critical assistance to law enforcement. Because their effectiveness in reducing crime and disorder has been widely documented (Weisburd et al., 2010), problem-solving frameworks have been adopted by many law enforcement agencies around the world and are applied daily in many police departments (Tilley & Laycock, 2002). What is the basis for their success? In addition to sharing the systematization of knowledge acquisition, these models all attach the utmost importance to crime analysis: the tool used by crime scientists to understand the reality of crime events.

For decades, crime analysis has served to synthesize complex problems and provide practical, actionable solutions. In *Crime Analysis for Problem Solvers in 60 Small Steps* —one of the most popular manuals for crime analysts published to date—two of the most notable contemporary crime scientists provide a set of guidelines for solving crime problems in a systematic way (Clarke & Eck, 2005). With this manual, and guided by the SARA model, analysts engage in a formative process where theory leads to practice, and crime analysis techniques serve to identify crime solutions. In addition to becoming familiar with the problem-solving approach and Environmental Criminology theories, crime analysts must deploy a whole suite of technical expertise to understand crime data. Among other skills, analysts must learn to use software such as statistical packages and geographic information systems, as well as communication techniques such as rhetoric and data visualization. Only the combination of theoretical and practical knowledge provides the necessary skills to become a proficient crime analyst.

This thesis shares Clarke and Eck's (2005) thinking. For these authors, crime analysts must achieve eight goals to solve crime problems: (1) prepare yourself; (2) learn about problem-oriented policing; (3) study Environmental Criminology; (4) scan for crime problems; (5) analyse in depth; (6) find a practical response; (7) assess the impact; and (8) communicate effectively. These eight specific goals match three more general notions that have been addressed in previous chapters: understanding Environmental Criminology theories, applying crime analysis techniques to crime data, and translating the knowledge acquired into solutions. Accordingly, CHAPTER I and CHAPTER II explained, developed and adapted Environmental Criminology theories to the context of cybercrime represented by the first notion. This chapter covers the second notion by describing materials used [30] and the crime analysis techniques conducted to analyse cybercrime problems. The latter notion is addressed in the following chapters.

4.1 Crime analysis through Data Science

As technologies evolve and large volumes of information are generated, new forms of data open the door to new angles from which to approach traditional crime problems. For example, relevant data on people's routine activities are generated through smartphones, social media, and other mobile applications. When such data is openly available, it “can be used to learn about people’s behaviour and make inferences about exposure to risk by finding different patterns on people’s daily routine activities”, thus advancing criminological understanding (Solymosi & Bowers, 2018, p. 213). There are many ways to access this data, such as crowdsourcing, participatory mapping, volunteered geographic information, and trough APIs, among others (for a review, see Solymosi & Bowers, 2018; see also Solymosi et al., forthcoming). However, these require specialised knowledge. To keep up with the circumstances, crime analysts require dedicated software that is capable of handling large volumes of data to harness its full predictive potential (Williams et al., 2016), while being aware of its limitations

[30] Note that, besides the statistical techniques employed, the empirical studies presented in this doctoral thesis required a series of research supporting materials for their completion. Some of them have been used transversely in all articles while others have served particular objectives of an article.

regarding quality and accessibility (Lynch, 2018). Beyond the software, a specific analytical framework is required to handle these data: The Data Science process.

The crime data used in each of the studies presented here certainly came from a range of sources that required the use of specific materials and crime analysis techniques. Yet, all these fall within a common analytical framework of Data Science. The Data Science concept on which this thesis builds is that described by Grolemund and Wickham (2016) in their seminal work *R for Data Science* [31], in which they define it as a process that allows analysts "to turn raw data into understanding, insight, and knowledge". Figure 4 shows the steps involved in this process. First and foremost, the data must be designed and collected. Then it is necessary to import the data taking into account their structure to be able to work with them. This is when data wrangling begins. Since even structured data needs to be arranged for analysis, the next step involves converting them into tidy data. Once the data are tidied, they are suitable to be understood; yet this requires further wrangling by transforming them. Data transformation is also the first stage in the data understanding cycle which involves data visualization and modelling too. While the first stage of this cycle is aimed at obtaining information from the data, the other two permit the extraction of knowledge from them. After performing as many iterations as deemed appropriate, the analyst will be prepared to understand the data and be ready to communicate it. By communicating the data, the Data Science process comes to an end. Note that most of the process is oriented to hypothesis generation, while only one of its steps —modelling data— is oriented to hypothesis confirmation (Grolemund & Wickham, 2016). Influenced by such bias, this thesis gives great importance to exploratory research, as can be recognized in

[31] This resource is available in open access via the following link: https://r4ds.had.co.nz/.

CHAPTER VII and CHAPTER VIII. In contrast, CHAPTER VI and CHAPTER IX are focused on testing premises and building predictive models respectively.

Since the purpose of this chapter is to provide an overview of the materials and methods used in the thesis, not all the procedures followed and the techniques employed are detailed here —that is the purpose of the methodology sections of each article—. Instead this chapter serves to define, describe, summarise, and organize the materials and methods used in the context of the Data Science process. For the sake of completeness, note that the following sections are full of footnotes containing the URLs that allow for a rapid extension of the information from reliable sources about the software used.

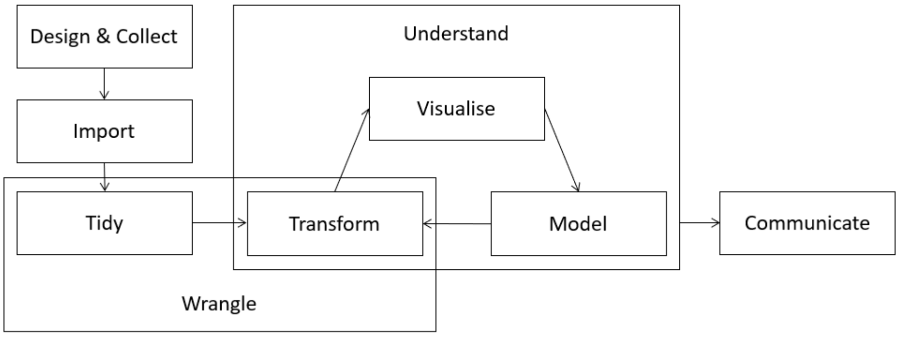

Figure 4. The Data Science process. Adapted from Grolemund and Wickham (2016), and Leek [32].

4.1.1 Design and collect data

An initial step in any Data Science project is the design and collection of data. Before starting with the fancy procedures, one must determine what type of data are required for measuring the reality to be observed. Depending on the demand, a unique instrument is used. Different tools are used to collect self-reported measures (e.g. questionnaires) than to collect objective measures (e.g. official statistics). And choosing the wrong tool

[32] See https://twitter.com/jtleek/status/963064077051408384?s=20.

can condemn the rest of the research to failure. In addition, many forms of data collection require specific materials —from surveys to APIs to web crawlers— which must be properly designed. Generally, there is no strict rule about which specific tools or materials should be used for this task, leaving it to the good judgment of the researcher. Yet the general rule is that the tools chosen must be adequate to fulfil the established need. The materials and methods selected to carry out the specific design and data collection tasks on each of the articles are outlined below.

To test whether the premises of Environmental Criminology theories apply to cybercrime in CHAPTER VI, we relied on Zone-H data on website defacements [33]. The Zone-H archive contains information about website defacements that are either self-registered by the offenders themselves or obtained from public sources of information. This makes Zone-H a unique data source. Their files contain millions of records that collect information about the time at which the cybercrime event is logged, the nick of the notifier, the type of defacement executed, the hacking method employed, the domain that has sustained the attack, its operating system, and the offender's motivation for committing the crime (e.g. Romagna & Van den Hout, 2017). Additionally, a mirror copy of the defaced website is stored for qualitative analysis. The Zone-H team is responsible for reviewing and validating each of the notifications they receive, as well as maintaining the defacements database. There is a section in their website that can be accessed to contact its administrator, its dedicated database maintainer, or to resolve general inquiries. After filling an online form, the person responsible contacts the applicant via email. Through this channel, it is possible to make a formal data request. Once the range of dates and variables to be acquired are specified —and an economic

[33] For more information, visit http://www.zone-h.org/.

agreement is reached— the database maintainer sends the file via email. For the details of the Zone-H dataset used in this thesis, see the section "Data" in CHAPTER VI.

The research on FMIWs in CHAPTER VII also required the use of additional specific materials. Since the detection of FMIWs required visiting risky online environments, certain self-protection measures were taken to preserve the security of both the researchers and their affiliated institutions. The security measures were oriented to preserving the anonymity of the computers used in the research. For this purpose, three tools were used to mask their IP addresses: a virtual private network (VPN), The Onion Router (TOR) browser, and the DuckDuckGo search engine. Regarding the first tool, NordVPN was chosen to use their application for the Windows operating system (OS). This service provider allows connecting through more than 5,500 servers in 59 countries using a sophisticated encryption system that ensures the connection is secured [34]. With regard to the second tool, TOR enhances the anonymity of its users by not revealing the domain they visit, by standardizing their digital fingerprint (i.e. the information related to their internet connected device and the browser they use), by deleting all cookies and search history, and by implementing its own multi-layered encryption system that is maintained by volunteers [35]. The third tool, DuckDuckGo, provides an additional boost to users' privacy, as it does not collect or disclose any personal information, track their browsing, or record their search history [36]. Together, all three tools constitute a thorough protective strategy.

Once the FMIWs were identified, additional methods and resources were deployed to collect and design data. Data collection required, in turn, two separate strategies. First, a systematic observation process was carried out in the FMIWs.

[34] For more information, visit https://nordvpn.com/.
[35] For more information, visit https://www.torproject.org/.
[36] For more information, visit https://duckduckgo.com/.

Systematic observation makes it possible to objectively detect the presence or absence of a number of elements in a given context (Reiss, 1971). In the case of FMIWs, such elements consisted in features integrated in their design that were classified as precipitation-control strategies (Wortley, 2001). But it was not feasible to collect all the data manually, so a second data collection strategy was needed. Retrieving the URLs contained in the FMIWs required the implementation of more sophisticated techniques such as web crawlers: bots capable of systematically inspecting websites for specific elements. To perform this task, the RCrawler R package version 0.1.9 was used, which allows to collect the URLs contained in a website and store them in a structured manner (Khalil & Fakir, 2017). As these URLs formed a network, its composition required another R package specialized in such task. We therefore relied on igraph version 1.2.4, a R package that "contains routines for creating, manipulating and visualizing networks, calculating various structural properties, importing from and exporting to various file formats and many more" (Csárdi & Nepusz, 2006, p. 1). In this way, social science data collection methods are brought together with computer science tools for a comprehensive Data Science approach.

0 presents a survey research design involving the participation of minors, so it was necessary to obtain the informed consent of their parents or tutors (Appendix G). Once prepared, informed consents were sent to the Governing Council of Castile-Leon, the autonomous government of the Spanish region where the survey was administered. From there, upon approval, they were sent to the Provincial Councils of each of the nine provinces of the region (i.e. Ávila, Burgos, León, Palencia, Salamanca, Segovia, Soria, Valladolid, and Zamora) so that these, in turn, could formally send them to the educational centres sampled. Accordingly, only those under-age students who provided a signed informed consent participated in the study. Following a procedure supervised

by the teachers, a questionnaire was administered to this group through an online platform. Although the original questionnaire was more extensive, only a few questions were selected for analysis in our research. These include measurements of the sociodemographic characteristics of the participants, the daily activities they carry out online, as well as their offending and victimisation experiences (Appendix H).

Finally, it was necessary to create a Twitter developer account [37] to access the Twitter social media data used in CHAPTER IX. Twitter is one of the few social media companies that has —more or less— consistently maintained an open data policy, albeit interrupted at certain times. This means that Twitter provides access to data generated by its users for certain cases and provided that certain ethical-legal conditions are met. One of such cases is the analysis of Twitter data for academic research [38]. But first, it is necessary to submit a formal application. To do this, an online form must be filled out detailing the type of data requested by the researcher and the purpose for which it would be used. The Twitter team then evaluates the application and decides whether to grant developer permissions to the applicant. If the evaluation is positive and Twitter grants developer permissions, the researcher can then proceed to create an authorized app, set up a dev environment to connect to Twitter's servers through their APIs, and manage the access level granted. In our research, we used the API streaming to obtain a dataset of tweets in real time. We chose this API among all others because it allowed us to capture the reaction and degree of interaction of Twitter users to certain events, and to monitor their evolution over time. But first, —to start collecting tweets— a set of instructions must be provided to the Twitter server so that the content can be filtered out. For example, it is possible to filter the data by using hashtags or keywords, limiting the

[37] For more information, visit https://developer.twitter.com/.

[38] In their own words: “Twitter believes in the value of an open exchange of information. This is why we are committed to providing academic researchers unparalleled access to our public conversational data”. See https://developer.twitter.com/en/use-cases/academic-researchers.

collection time, or setting a maximum number of tweets to be collected. For details of the Twitter dataset used in this thesis, as well as the filters used for the collection, see the section “Sample and procedure” in CHAPTER IX.

4.1.2 Import data

The second step in the Data Science process involves importing the data. If the data is not imported, it cannot be analysed, so this step is as obvious as it is necessary. Importing data into R, implies reading different file formats (e.g. .csv, .xlsx, .sav, .json), databases, or APIs, to load them as a data frame (Grolemund & Wickham, 2016). But not all data is structured, instead there is semi-structured data such as .json files or even unstructured data such as .pdf files. In the latter two cases, it is necessary to carry out a prior analytical process called parsing that allows for the restructuring of the data. For example, while the Zone-H data in CHAPTER VI and the data collected through the questionnaire in CHAPTER VIII were both directly imported from a .sav and a .xlsx file respectively, the Twitter data in CHAPTER IX was parsed from a .json file —quite complex to read— to a .csv file —far simpler— before being imported to R for its processing. When reading this data into R, it takes the structured shape of data frames. Only after the data is structured it is suitable to be analysed.

Base R provides its own functions for importing data such as .csv, but the Tidyverse offers an even better option: the readr R package. “The goal of readr is to provide a fast and friendly way to read rectangular data” [39]. According to Grolemund and Wickham (2016), there are at least three reasons for using readr functions instead of their baser R counterparts: (1) the data parsing speed is about 10 times faster; (2) the data is structured in an orderly manner, which facilitates subsequent processing; (3) the

[39] See https://cran.r-project.org/web/packages/readr/index.html.

code is more reproducible, since readr does not inherit the behaviour of the OS as base R does. The Tidyverse also offers alternatives for other forms of data. The haven R package reads .sav files, and the readxl R package reads both .xls and .xlsx files. To import data in this thesis, we used haven version ≥ 1.1.0 (CHAPTER VI) (Wickham & Miller, 2019), readxl version ≥ 1.0.0 (CHAPTER VII and CHAPTER VIII) (Wickham & Bryan, 2019), and readr version ≥ 1.1.1 (CHAPTER IX) (Wickham et al., 2018).

4.1.3 Tidy data

After importing the data, the next step is to tidy it up. Note that in the context of Data Science, the concept of tidy data has a special connotation, which goes beyond merely organizing information. According to Wickham (2014, p. 4), "[i]n tidy data: (1) [e]ach variable forms a column; (2) [e]ach observation forms a row; and (3) [e]ach type of observational unit forms a table". Therefore, "[a] dataset is messy or tidy depending on how rows, columns and tables are matched up with observations, variables and types" (Wickham, 2014, p. 4). Because they contain clearly structured data, tidy datasets make the job of analysts easier (Grolemund & Wickham, 2016). Although recording tidy data should be mandatory, the reality is that crime data — much like other types of data— is often messy. This happens in several ways: when the column headers are omitted; when multiple variables are recorded in the same column or in rows and columns indistinctly; and when the observational units get disarranged, either because several are mixed in the same table or because one is divided into several tables (Wickham, 2014). Tidying up the data constitutes the first phase of data wrangling, an arduous process that usually consumes about 80% of the analyst's time (Dasu & Johnson, 2003). Fortunately, there are tools that simplify this task.

The Tidyverse is designed to function optimally with tidy data, but if there are messy data to deal with, it counts on dedicated packages to tidy them up (Grolemund &

Wickham, 2016; Wickham et al., 2019). One of these R packages is tidyr. tidyr provides a set of functions to restructure messy data into tidy (Grolemund & Wickham, 2016). These functions are used to perform five types of tasks: (1) pivoting data, to reshape the data by redistributing the values in rows and columns; (2) rectangling data, to convert semi-structured data into structured; (3) nesting data, to ungroup data by assigning them to their corresponding row; (4) splitting and combining data, to operate with character string data —such as plain text—; and (5) handling missing data [40]. In addition, the Tidyverse includes another package that participates in the task of tidying up the data: tibble. This package allows the creation of objects called tibbles, a type of data frames that possess synergies with many Tidyverse functions (Grolemund & Wickham, 2016). Although the contribution of tibble seems insignificant, besides being a visual improvement of the data frame, it is useful for the early detection of problems in the data, since it does not change the names of the variables nor does it perform partial matching (i.e. when not all the characters of a variable are coincident) [41]. To tidy data in this thesis, we used tidyr version ≥ 0.7.2 (Wickham & Henry, 2020) and tibble version ≥ 1.3.4 (Müller & Wickham, 2019).

4.1.4 Understand data: The transform-visualize-model cycle

The previous three steps of the Data Science process, which require solid technical skills, serve to prepare the data for further analysis. But in order to understand the data, the next step requires complementing such skills with specialized knowledge in whatever field of study the research belongs to. This is arguably the most complex and important step. In turn, understanding data involves a three stages cycle: transforming data, visualizing data, and modelling data (Grolemund & Wickham, 2016). While

[40] See https://tidyr.tidyverse.org/.

[41] See https://tibble.tidyverse.org/.

transformation allows to obtain information from the data, visualization and modelling permit to go further and extract knowledge from them. If the knowledge extracted is incomplete or insufficient, the data can always be transformed again to address the research questions from a different angle. The virtues of this cycle are discussed below, breaking them down and specifying the materials used in each stage.

4.1.4.1 *Transform data*

Data transformation represents the second phase of the costly data wrangling process. Transforming data includes filtering observations, creating new variables from existing ones, and calculating descriptive statistics (Grolemund & Wickham, 2016). In this sense, data can be filtered by establishing comparative criteria (e.g. if a value is greater, equal or less than another), through logical operations (e.g. to select the values that meet one condition and/or another), and by handling missing data (e.g. eliminating observations that contain not assigned values) (Grolemund & Wickham, 2016). Regarding the second aspect, sometimes data contains the information needed but does not show the information desired. For example, some people may be interested in knowing more about crime rates, but the available data only contains crime counts and population figures. In that case, a new variable must be created from the previous to reflect the required information. Finally, no matter how much data is available, it is often necessary to synthesize it in order to obtain useful information. Once again, the Tidyverse is equipped with the right tool to perform all these tasks. Building on its data manipulation grammar, the dplyr R package provides a set of functions to "solve the most common data manipulation challenges" [42]. We relied on dplyr version ≥ 0.7.4 to manipulate data in CHAPTER VI, CHAPTER VII, and CHAPTER VIII (Wickham et al., 2020).

[42] See https://dplyr.tidyverse.org/.

Since a particular type of data —networked data— was manipulated in CHAPTER VII, we resorted to the igraph R package as a specialized solution. The networks generated by igraph constitute a special type of object that cannot be manipulated with other tools such as dplyr. For this reason, in addition to designing networked data as previously stated, igraph also incorporates utilities for transforming and analysing such data. Networks formed by such data are the object of study of the academic interdisciplinary discipline known as Network Science (Barabási & Pósfai, 2016). Criminologists usually draw on Network Science when examining social groups (Bichler, 2019), but there are other types of associations, such as hyperlinked networks, that may also be of interest to the field. In hyperlinked networks, nodes are websites and edges represent the hyperlinks that connect them (H. W. Park, 2003). Although the methodological foundations of Hyperlink Network Analysis (HNA) are the same as Social Network Analysis (SNA) (Wasserman & Faust, 1994), the interpretation of their results is subject to different considerations, as there is a substantial disparity in the nature of the data. It is obvious that a network of hyperlinks is not comparable to a network of people, since both types of nodes do not interact in the same way. Therefore, although the analytical techniques applied may be similar, the implications of the results are quite different. To analyse the FMIWs hyperlinked network, we used igraph version 1.2.4 (Csárdi & Nepusz, 2006).

In CHAPTER VIII a particular process of data transformation known as Conjunctive Analysis of Case Configurations (CACC) (Miethe et al., 2008) has been applied to the questionnaire data collected. It is particular because, actually, the CACC can be considered as both a data transformation process and an analytical technique, whereas the analysis is the outcome of a relatively complex data transformation process. How can that be? While CACC can be defined as a multivariate technique for exploratory data analysis (Miethe et al., 2008), its execution follows a three-step data transformation process: (1) creating a data matrix, a synonym for the tidy data to be analysed; (2) populating the data matrix, by sorting the data and counting how many identical observations exist along with their probability of occurrence; and (3) preparing the observations for further analysis, by establishing a threshold that defines which cases are dominant (Hart, 2014). The output is a table that reveals patterns in the tidy data (Appendixes I and J). A CACC was then applied to the data collected through the questionnaire using the CACC R package version 1.0.0 (Miriam Esteve et al., 2019). This R package contains a set of functions to conduct a CACC and other related analyses to identify situational clustering (Hart, 2019) and the main effect of specific variables (Hart & Moneva, 2018). The functions of the CACC R package are in turn nourished by the Tidyverse, since it relies on some of the functions of dplyr and ggplot for data transformation and visualization, respectively. Although the syntax to conduct a CACC was already available elsewhere for other software (i.e. SPSS, STATA, SAS) (see Miethe et al., 2008), this R package constitutes the first initiative to carry it out with free software.

On an exceptional basis, to carry out data transformation for the research presented in CHAPTER IX, a software additional to R —Python— was used (see Appendixes K and L). Like R, Python is both a programming language and software; unlike R, Python is a general-purpose programming language (i.e. it is designed to handle all kinds of tasks, not just statistics) [43]. We opted to use this tool for a purely operational criterion (i.e. multidisciplinary collaboration), as the same tasks that were performed with Python could have been performed with R. Python also features an integrated development environment called Spyder that "offers a unique combination of

[43] For more information, visit https://www.python.org/.

the advanced editing, analysis, debugging, and profiling functionality of a comprehensive development tool with the data exploration, interactive execution, deep inspection, and beautiful visualization capabilities of a scientific package" [44]. This environment facilitates the programming tasks carried out by the analysts. Python's functionalities can also be extended by drawing on its system of packages, known as libraries. In our research, we used Python version 3.7, and Spyder version 3.3.0. Once data has been transformed, there are two ways to extract knowledge from it: through visualization and by modelling.

4.1.4.2 *Visualize data*

After transforming the data to obtain information, it is necessary to take further steps to extract knowledge from them. One of these steps is taken through data visualization. Visualization is important because it may reveal patterns in the data that are normally hidden (Healy, 2018). This is especially true for Big Data because the bigger the data, the noisier it is. Graphics can help to visually synthesize the information contained in the data by eliminating the noise thus generating knowledge. There are many types of graphics and many more every day [45]. Despite the wide variety, it is important to understand that not any graphic is adequate to represent an idea. There is a logic behind data visualization. Aside from the researcher's creativity and good taste —which of course play an important role in designing a good visualization—, depending on the

[44] For more information, visit https://www.spyder-ide.org/.

[45] Among others: area graph, bar chart, box and whisker plot, bubble chart, bullet graph, candlestick chart, density plot, error bars, histogram, Kagi chart, line graph, Marimekko chart, multi-set bar chart, open-high-low-open chart, parallel coordinates plot, point and figure chart, population pyramid, radar chart, radial bar chart, radial column chart, scatterplot, span chart, spiral plot, stacked area graph, stacked bar graph, stream graph, violin plot, arc diagram, brainstorm, chord diagram, flow chart, illustration diagram, network diagram, non-ribbon chord diagram, Sankey diagram, timeline, tree diagram, Venn diagram, calendar, Gantt chart, heatmap, stem and leaf plot, tally chart, time table, circle packing, donut chart, dot matrix chart, nightingale rose chart, parallel sets, pictogram chart, pie chart, proportional area chart, sunburst diagram, treemap, word cloud, bubble map, choropleth map, connection map, dot map, flow map. See https://datavizcatalogue.com/.

number of variables and the type of data to be represented, a certain type of graphic should be used (Healy, 2018). For example, to visualize the distribution of a single numerical variable it is more appropriate to use a histogram than a bar graph, because the x-axis is continuous in the first case and better reflects the nature of the data. In addition to following some formal rules, an essential element in data visualization is clarity. For this reason, all unnecessary elements in a graphic should be removed. To superfluous graphic elements, Tufte refers by the term "chartjunk" (Tufte, 1999). One of the most typical examples of chartjunk is the use of 3D geometries, although there are many others such as the improper use of gradients, colours, or grids. Fortunately, all these issues in data visualization can be handled by understanding the grammar of graphics.

The grammar of graphics is a system for identifying the components shared by any graphic (Wilkinson, 2005). According to Wickham (2010), these components are: data, aesthetics, geometries, scales, facets, statistics, and coordinates — to which themes should be added—. By manipulating these elements, it is possible to design any graphic. Much to the delight of the analyst, the Tidyverse toolkit features a specialized R package for data visualization that has been designed based on the grammar of graphics called ggplot2 (Wickham, 2016). Basically, to operate with this package one has to "provide the data, tell ggplot2 how to map variables to aesthetics, what graphical primitives to use, and it takes care of the details" [46]. We used ggplot2 version ≥ 2.2.1 for designing the figures displayed in CHAPTER VI, CHAPTER VII, and CHAPTER VIII. Furthermore, we have used other R packages for data visualization that share the philosophy of ggplot2 to function.

[46] See https://ggplot2.tidyverse.org/.

One of these is the ggraph R package (Pedersen, 2020). Pedersen —its developer— noted that ggplot2's Grammar of Graphics does not adjust to the requirements of network visualization due to the different structure of their input data. Therefore, he decided to design an R package that would overcome this obstacle. Thus, ggraph was created: "an extension of the ggplot2 API tailored to graph visualizations and provides the same flexible approach to building up plots layer by layer" [47]. In CHAPTER VII, we used ggraph version 1.0.0 to plot the FMIWs network. Another example is the GGally R package (Schloerke et al., 2020). GGally extends ggplot2's Grammar of Graphics to other designs that, due to their more complex nature, are not supported by the latter package. For that purpose, GGally adds "several functions to reduce the complexity of combining geometric objects with transformed data" [48]. This allows the display of graphics such as pairwise plot matrixes, parallel coordinates plots, and survival plots, among others. In CHAPTER VIII, we used GGally version 1.4.0 to create the parallel coordinates plot. A third package that, unlike the previous, does not extend the functions of ggplot2 was employed. The DiagrammeR R package (Iannone, 2020) provides a set of functions to create graph network structures for further analysis and visualisation. One such structures is the decision tree, which can be converted into a flowchart. In CHAPTER IX, we used DiagrammeR version 1.0.0 to display an example of a decision tree by means of a flowchart. All three packages described above, provide additional flexibility to ggplot2 to, collectively, create the appropriate data visualizations. To complement the extraction of knowledge resulting from data visualization, data modelling can be used.

[47] See https://cran.r-project.org/web/packages/ggraph/index.html.
[48] See https://cran.r-project.org/web/packages/GGally/index.html.

4.1.4.3 *Model data*

In this thesis, more emphasis is placed on description than on prediction. For this reason, readers will note that data modelling does not become too prominent until 0. This has nothing to do with the usefulness of data modelling, but with the fact that models were simply not the best methodologies to answer most of the research questions we posed. In fact, two of the studies presented here are essentially exploratory (CHAPTER VII and CHAPTER VIII). Fortunately, Data Science confers great relevance to data processing and exploratory analysis; there is only one phase in the cycle of data understanding dedicated to modelling (Grolemund & Wickham, 2016).

Models are mathematical tools that can be used for hypotheses testing. By fitting data to functions, models reveal hidden patterns in the data in a way that complements data visualization. There are many different types of models: linear, generalised linear, generalised additive, penalised linear, robust linear, and trees (Grolemund & Wickham, 2016). Unlike the others, trees are models that constantly fit the data while sequentially dividing them into smaller subsets. In this way, trees make “decisions” on the basis of a set of inputs to classify an output (Quinlan, 1986). One of such models are Random Forests (Breiman, 2001). As their name suggests, Random Forests are forests because they generate a multitude of trees, and they are random because they select variable inputs at random whenever they split the data. When Random Forests are implemented from a machine learning approach, they are first fed into a training dataset to learn how to classify, and then their performance is evaluated in a test dataset. In this sense, training data enable the identification of patterns while test data serve to evaluate whether such patterns are robust. Taking advantage of it, we used Random Forests to fit Twitter data and classify online hate speech in CHAPTER IX. To do this, we used the scikit-learn Python library version 0.19.2 which integrates a set of machine learning algorithms (Pedregosa et al., 2011) [49]. Models are helpful in understanding data, but then the findings must be communicated effectively.

4.1.5 Communicate data

The final step of the Data Science process is data communication (Grolemund & Wickham, 2016). This is a crucial step, because unless the knowledge extracted in this process is properly communicated, all the previous steps are in vain. Yet, before communication can begin, it is essential to be aware to whom it is addressed. Note that there are different actors who may be interested in acquiring the knowledge generated: from stakeholders to researchers to the general public. And depending on who the recipient is, the communicative strategy should be different. For example, stakeholders may be interested in receiving a briefing containing the main findings, researchers may be interested in reading a detailed report on the research methodology, and the general public may be interested in obtaining clear information through simple infographics. With R Markdown it is possible to address all three scenarios.

R Markdown provides a working environment in RStudio that represents an evolution of traditional scripts, since it allows not only to save and reproduce code, but also to generate high quality reproducible reports (Xie et al., 2018) [50]. The main benefit of R Markdown is that provides the possibility to generate reports on the go, while performing data wrangling or visualization. With this tool, it is not only possible to annotate the code, but also to integrate it into a comprehensive working document containing text, code, tables and figures. Once finished, it can be exported in several formats (e.g. HTML, PDF, Word), which greatly facilitates its dissemination. Following the previous example, it is possible to produce an executive report for debriefing

[49] For more information, visit https://scikit-learn.org/stable/.

[50] For more information, visit https://rmarkdown.rstudio.com/.

stakeholders in Word format, to share code with other researchers interested in replicating the research in HTML format, or to create infographics in PDF format to share publicly [51]. The rmarkdown R package version ≥ 1.17 (Allaire et al., 2020) was used to carry out and annotate the analyses, as well as display the graphics, contained in CHAPTER VI and CHAPTER VIII.

4.2 A transversal tool: The R free software

There are many software products that can handle Big (and new forms of) Data, but few have as many advantages as R. Besides being a programming language optimized for statistical analysis, R is "a free software environment for statistical computing and graphics" (R Core Team, 2019) [52]. At least, five important reasons support this choice: (1) its accessibility, because it is free; (2) its reliability, as its open source code is constantly reviewed and updated by one of the most active user communities; (3) its performance, as it allows for the convenient handling of large volumes of data; (4) its versatility, as it supports a wide range of functions through its system of packages; and (5) its transparency, since the code of the analyses are captured in scripts that can be replicated by anyone (e.g. journal reviewers, other researchers). The latter represents a particularly important feature due to the growing concern for replicability in Social Sciences and, more specifically, in Criminology (Pridemore et al., 2018). Therefore, we relied on R to implement all the crime analysis techniques detailed in the four empirical studies presented here. By using R, we wanted to take a further step to respond to the so-called Replication Crisis (Baker, 2016). R was further enhanced with RStudio: an integrated development environment that "includes a console, syntax-highlighting editor

[51] It should be noted that while producing rudimentary versions of these documents is relatively simple, the art of communicating science effectively is quite complex. Only practice makes perfect.

[52] See https://www.r-project.org/.

that supports direct code execution, as well as tools for plotting, history, debugging and workspace management" [53]. The purpose was to improve the usability of R by means of the features that RStudio incorporates (e.g. automatic indentation, function definitions, shortcuts). In short, RStudio makes the analyst's job easier.

The functionalities offered by the basic R programming language —known as base R— can be extended by installing "fundamental units of reproducible R code" (Wickham, 2015) called packages. R packages "include reusable R functions, the documentation that describes how to use them, and sample data" (Wickham, 2015). As of today, over 15,000 R packages are available on the Comprehensive R Archive Network (CRAN) [54]. One of these packages is the Tidyverse: "a language for solving data science challenges with R code" (Wickham et al., 2019, p. 1). Actually, the Tidyverse is not a regular package, but a meta-package that supports "a collection of R packages that share a high-level design philosophy and low-level grammar and data structures" (Wickham et al., 2019, p. 1). Tidyverse consists of a series of frequently used core packages for generic tasks, and a set of non-core packages for specialized tasks (Wickham et al., 2019). Generic tasks include import data, tidy data, understand data, and communicate data, while specific tasks include reading particular data files, handling certain types of data, and incorporating additional tools to facilitate programming (Grolemund & Wickham, 2016; Wickham et al., 2019). Table 4 shows the software versions that were used for processing and analysing data in each article [55].

[53] See https://rstudio.com/products/rstudio/#rstudio-desktop.

[54] For more information, visit https://cran.r-project.org/.

[55] Although R is a valuable resource, it is not the most suitable software for all tasks, nor do its packages contain all the data one may need, nor —of course— does it respond to other needs beyond data processing and analysis. For this reason, the transversal task of processing and analysing data with R occasionally needed to be complemented with additional resources in order to complete specific research actions.

Table 4.
Versions of the software used in all the articles of the thesis

	Software version		
Article	R	RStudio	Tidyverse R package
1 (CHAPTER VI)	3.6.1	1.2.5001	1.2.1
2 (CHAPTER VII)	3.6.0	1.2.1335	1.2.1
3 (CHAPTER VIII)	3.6.1	1.2.5019	1.3.0
4 (CHAPTER IX)	3.5.1	1.1.4630	1.2.1

Following a brief introduction, the next chapter presents an outline of the four empirical articles that constitute the backbone of the thesis.

—Blank page—

CHAPTER V

INTERLUDE: OUTLINE OF THE ARTICLES

In music, interludes are introductory pieces that are inserted between plays or acts as an impasse. They are often useful for balancing abrupt leaps between acts or introducing complex plays. In theses by compendium of articles, a sudden leap usually occurs just before presenting the papers. After a general introduction of the thesis in CHAPTER I, a development of the theoretical framework in CHAPTER II, an overview of the research question and the hypotheses posed in CHAPTER III, and a summary of the materials and methods used in CHAPTER IV, directly presenting the first empirical article felt like an abrupt leap. To avoid that, this chapter is intended to be an interlude to introduce the papers. In addition, this chapter serves to address a number of mandatory formal issues. For this reason, we present, along with the abstracts of each article, their reference, as well as the contribution of each co-author to the research. The full content of these is then presented in the next four chapters.

5.1 Article 1 (see CHAPTER VI)

This article has been submitted as: Moneva, A., Leukfeldt, E. R., Van de Weijer, S. G. A., & Miró-Llinares, F. (submitted). Repeat victimization by website defacement: A test of Environmental Criminology premises for cybercrime. *Computers in Human Behavior*.

Abstract: Repeat victimization has been widely studied from the perspective of Environmental Criminology for several decades. During this period, criminologists have identified a set of repeat victimization premises that are observed for many property crimes; however, it is unknown whether these premises are also valid for cybercrime. In this study we employ more than 9 million Zone-H data records from 2010 to 2017 to test whether these premises apply for the cybercrime of website defacement. We show that the phenomenon of repeat victimization is also observed in cyber places where this type of cybercrime occurs. In particular, we found that repeats contributed little to crime rates, that repeats occurred even several years after the original incident, that they were committed disproportionately by prolific offenders, and that few offenders returned to victimize previous targets. The results suggest that traditional premises of repeat victimization may also be valid for understanding cybercrime events such as website defacement, implying that Environmental Criminology theories also constitute a useful framework for cybercrime analysis. The implications of these results in terms of criminological theory, cybercrime prevention and the limitations derived from the use of Zone-H data are discussed.

Keywords: cyber place, cybercrime, Environmental Criminology, hacking, repeat victimization, website defacement, Zone-H

5.2 Article 2 (see CHAPTER VII)

This article has been published as: Moneva, A., & Caneppele, S. (2019). 100% sure bets? Exploring the precipitation-control strategies of fixed-match informing websites and the environmental features of their networks. *Crime, Law and Social Change*. https://doi.org/10.1007/s10611-019-09871-4

Abstract: In recent years, many human activities have made cyberspace their preferred environment. This study focuses on the betting environment, specifically on

fixed-match informing websites (FMIWs). These sites claim to be capable of selling tips about fixed sports events. They essentially act as vendors of confidential sources, allowing punters to place 100% sure bets. We hypothesize that cyber places for match-fixing tips facilitate deviant behaviours. Through systematic observation, we describe and quantify a set of 15 environmental features they share, which do not always belong to regulated online betting platforms. Findings from 78 FMIWs corroborate our hypothesis, as they support the relevance of Environmental Criminology theories applied to cybercrime. Additional exploration through hyperlink network analysis shows that FMIWs are highly homogeneous and have similar characteristics to the Tor network but differ from other illicit online environments such as sexual child exploitation networks or white supremacist communities. The characteristics of the network suggest that the business is more similar to a fraud scheme than an illicit market. Finally, the practical implications of the results for crime prevention and the directions for future research are outlined.

Keywords: fixed-match informing websites, sport betting, cyber place, situational precipitators of crime, hyperlink network analysis

5.3 Article 3 (see CHAPTER VIII)

This article has been published as: Moneva, A., Miró-Llinares, F., & Hart, T. C. (2020). Hunter or Prey? Exploring the Situational Profiles that Define Repeated Online Harassment Victims and Offenders. *Deviant Behavior*. https://doi.org/10.1080/01639625.2020.1746135 [56]

[56] The annotated R code written for the analyses is freely available in a GitHub repository via the following link: https://github.com/amoneva/code_papers/blob/master/articles/moneva_etal_2020_github.Rmd

Abstract: Data collected from a sample of Spanish non-university students (N = 4174) were used to identify unique situational profiles of self-identified repeated online harassment victims and offenders, through a Conjunctive Analysis of Case Configurations (CACC). Repeat victim and offender profiles were constructed using individual-level factors and variables related to the cyber "places" where students go online and their personal information they share while there. Clustering analysis demonstrates that students spent their time online in few situational contexts where online harassment occurs. Dominant situational profiles of students are then provided, along with their associated probabilities for experiencing repeat victimization or committing repeat offending, identifying those at relatively higher and lower risk. Results show that composite profiles associated with victims of repeated online harassment are dissimilar to those associated with offenders of repeated online harassment, suggesting that each form of online harassment occurs in different situational contexts and therefore requires different preventative measures. Our findings are discussed in terms of criminological theory, future online harassment research, cybercrime prevention, and policy implications.

Keywords: online harassment; cyber place, CACC, conjunctive analysis, situational profile

5.4 Article 4 (see CHAPTER IX)

This article has been published as: Miró-Llinares, F., Moneva, A., & Esteve, M. (2018). Hate is in the air! But where? Introducing an algorithm to detect hate speech in digital microenvironments. *Crime Science, 7*(15), 1-12. https://doi.org/10.1186/s40163-018-0089-1

Abstract: With the objective of facilitating and reducing analysis tasks undergone by law enforcement agencies and service providers, and using a sample of digital messages (i.e., tweets) sent via Twitter following the June 2017 London Bridge terror attack ($N = 200,880$), the present study introduces a new algorithm designed to detect hate speech messages in cyberspace. Unlike traditional designs based on semantic and syntactic approaches, the algorithm hereby implemented feeds solely on metadata, achieving high level of precision. Through the application of the machine learning classification technique Random Forests, our analysis indicates that metadata associated with the interaction and structure of tweets are especially relevant to identify the content they contain. However, metadata of Twitter accounts are less useful in the classification process. Collectively, findings from the current study allow us to demonstrate how digital microenvironment patterns defined by metadata can be used to create a computer algorithm capable of detecting online hate speech. The application of the algorithm and the direction of future research in this area are discussed.

Keywords: hate speech, Twitter, cyber place, metadata, random forests

Author’s contributions: The theoretical framework and research question were initially proposed by Fernando Miró Llinares, while Asier Moneva further developed this background. Then, Miriam Esteve obtained and preprocessed the sample required for the analysis. Variables were selected according to Miró-Llinares and Moneva’s approach. Machine Learning techniques were conducted by Esteve and interpreted by Moneva. Finally, Miró-Llinares and Moneva elaborated the discussion section and conclusions. All authors read and approved the final manuscript.

—Blank page—

CHAPTER VI

TESTING REPEAT VICTIMISATION PREMISES TO UNDERSTAND WEBSITE DEFACEMENTS

This chapter has been submitted as: Moneva, A., Leukfeldt, E. R., Van de Weijer, S. G. A., & Miró-Llinares, F. (submitted). Repeat victimization by website defacement: A test of Environmental Criminology premises for cybercrime. *Computers in Human Behavior*.

6.1 Introduction

Our society is gradually becoming increasingly digitized and so is crime. For the past three decades, technological breakthroughs have created new opportunities to commit crimes in digital environments such as the Internet. Sometimes these crimes resemble traditional crimes (i.e. cyber-enabled crimes), but on other occasions they appear as completely new criminal phenomena unparalleled in physical space (i.e. cyber-dependent crimes) (e.g. M. McGuire & Dowling, 2013a). Although cybercrimes have become a regular occurrence, we still know relatively little about them: Why do they occur? How can they be prevented or mitigated? To address these questions, a growing body of research on the human factor of cybercrime has contributed to expanding our knowledge about victims, offenders, policing strategies required for cybercrime control, as well as the role of criminological theory in these three areas (Holt & Bossler, 2014; Leukfeldt, 2017; Leukfeldt & Holt, 2020; Maimon & Louderback, 2019). With regard to criminological theories, it is particularly important to examine whether traditional theories remain useful in explaining cybercrime (Bossler, 2020; Holt & Bossler, 2017). In this sense, this article contributes to the existing literature by exploring the potential applicability of the Environmental Criminology theories to better understand cybercrime as an event.

Crime events have a certain baseline risk of occurring, but research has shown that for some property crimes such as burglary, vandalism, and graffiti, this risk increases after the initial occurrence (Farrell, 2005). Sometimes this increase in risk manifests when a specific crime impacts a target more than once, meaning the target suffers repeat victimization. Established research suggests that repeat victimization typically occurs within a short interval of time after the first victimization (Bowers & Johnson, 2005; Farrell, 2005; Farrell & Pease, 1993; Johnson et al., 1997; Johnson & Bowers, 2004; Pease, 1998), that it has a large impact on crime rates (Farrell & Pease, 2017, 2018; Pease, 1998), and that it is committed by a few prolific offenders (Bernasco, 2008; Farrell, 2005; Farrell & Pease, 1993, 2017; Lammers et al., 2015; Pease, 1998). Repeat victimization has mainly been studied regarding property crimes such as residential burglaries (e.g. Bowers & Johnson, 2005; Johnson, 2008b) or commercial burglaries (e.g. Bowers, 2001), theft from motor vehicles (Johnson et al., 2009), robberies, and shoplifting (Farrell, 2005) [57]. The consistency of the findings on repeat victimization for different types of crime over more than two decades allows their transformation into verifiable premises that can be tested for other crimes. The present study explores whether the traditional premises on repeat victimization also apply to a specific type of cybercrime: website defacements.

[57] In addition to research that has focused on property crime, the phenomenon of repeat victimization has also been observed in interpersonal crimes such as rape, sexual assault, or violent assault (e.g. Nazaretian & Merolla, 2013; Planty & Strom, 2007; Turanovic et al., 2018).

Website defacement is a cyber-dependent crime that involves trespassing on a website to alter its contents (see Maimon & Louderback, 2019 for a review of the current state of research on cyber-dependent crime). “Defacements enable hackers to post messages and images that indicate their perspectives and beliefs, as well as gain status by listing their name and group affiliation” (Holt, 2011, p. 171). When this crime is committed with political motives, it is encompassed within the phenomenon of hacktivism (Romagna, 2019), but there is a wide variety of motives and modus operandi behind defacements, which means it acquires a phenomenological dimension of its own (Madarie, 2017; Romagna & Van den Hout, 2017). For example, some hackers seek recognition after successfully trespassing web servers; the more domains they attack and the greater the difficulty, the more they can flaunt their skills. Recognition is a cornerstone for gaining status in the hacker community (Holt, 2019) that can lead to certain offenders being especially prolific or certain domains being disproportionately victimized.

But how can defacements be studied from the quantitative perspective required by repeat victimization studies when there are no official sources of data nor longitudinal panel studies on this type of crime? One of the few alternatives is to rely on secondary data such as Zone H, a database containing millions of self-reported defacement cases. This data has been used for researching defacements in the past and continues to be used with this aim today (Davanzo et al., 2011; Howell et al., 2019; Maimon et al., 2017; Romagna & Van den Hout, 2017; Woo et al., 2004). Previous quantitative studies on defacements can be divided into two categories: those that rely on the human factor perspective to understand the phenomenon, and those that apply a computational perspective for its prevention and mitigation. The former category of studies, which is scarcer than the latter, tend to approach the issue from a descriptive perspective —with the exception of some recent studies using more advanced methodologies— and from a certain theoretical foundation (Holt, Leukfeldt, et al., 2020; Howell et al., 2019; Romagna & Van den Hout, 2017; van de Weijer et al., submitted). The latter are usually brief or preliminary works with an eminently technical component (e.g. Davanzo et al., 2011; Maimon et al., 2017).

This paper aims to contribute to criminological literature by bridging the gap between the two groups as it introduces a hitherto unexplored theoretical framework for defacements with a distinctly preventive purpose. Based on the idea that traditional criminological theories may be useful in cyberspace (Holt & Bossler, 2017), and particularly Environmental Criminology theories (Miró-Llinares & Johnson, 2018; Miró-Llinares & Moneva, 2019a), this paper explores the applicability of the premises of repeat victimization observed for physical crime to defacements in cyberspace.

The following section presents the theoretical framework for this study. The proposed approach is founded on the applicability of Environmental Criminology theories, and particularly the concept of place, to crimes committed in cyberspace. Next, the objectives of the study are presented together with the traditional repeat victimization premises and their reformulation to be specifically explored for defacements. In the methods section we explain our data source, as well as the measures used in the analysis. The results, which are organized by premises and accompanied by tables and figures for clearer interpretation, are then discussed within the framework of previous research and Environmental Criminology theories. The potential implications of the work in relation to other studies that have used the same data source and its potential to find patterns of repeat victimization for cybercrime prevention are also discussed. The paper briefly concludes with the key insights obtained from the study.

6.2 Environmental Criminology as a theoretical framework for cybercrime

For decades, the Environmental Criminology theories have served to understand the situational aspects of crime events and propose strategies for their prevention (Bruinsma & Johnson, 2018; Wortley & Townsley, 2017). There are three main Environmental Criminology theories: The routine activity approach, whose most popular premise is that crime occurs at the micro level in the absence of capable guardians when a motivated offender and a suitable target converge in space and time (L. E. Cohen & Felson, 1979); the geometry of crime, which postulates that the distribution of crime events is not random, but occurs in places where the activity spaces of offenders and targets intersect (Brantingham & Brantingham, 1981); and the rational choice perspective, which states that the offenders' decision to commit a crime reflects a weighting of costs and benefits (Cornish & Clarke, 1986). The latter has been complemented by situational precipitators of crime (Wortley, 2001) and applied in practice through situational crime prevention strategies (Clarke, 1997). An important advantage of these theoretical bodies is their simple formulation, which allows for synergies and whose interpretation has resulted in analytical frameworks which contribute to a better understanding of crime, such as the crime triangle (Eck, 1994) or the repeat victimization premises (e.g. Farrell & Pease, 2018; Pease, 1998). The application of these theoretical frameworks has always been heavily influenced by the geography of crime, but their potential scope has yet to be discovered for crimes committed in cyberspace.

The pre-digital context in which the Environmental Criminology theories were conceived meant their development was essentially geographical, as little was known about cybercrime at that time. The increase of cybercrime as a problem has caused some scholars who previously focused on geographic crime to pay more attention to crime in cyberspace. This shift in focus has served to theoretically develop the frameworks of Environmental Criminology theories into cybercrime (Miró-Llinares & Johnson, 2018; Miró-Llinares & Moneva, 2019a). First Grabosky (2001; Grabosky & Smith, 2001) and then Yar (2005) developed a theoretical application of the routine activities approach to cyberspace, which Miró-Llinares (2011) and Reyns and colleagues (2011) later discussed, and which Holt and Bossler (2008) pioneered into empirical practice. In this context, whereas some consider that the structural characteristics of cyberspace —the contraction of time and space— complicate the application of environmental theories (e.g. Yar, 2005), others consider that they simply need to be adapted to the particularities of the environment (e.g. Miró-Llinares & Moneva, 2019a).

Nevertheless, since then, dozens of empirical studies have been conducted on the application of this approach to understand the dynamics of different forms of cybercrime (for a review, see Leukfeldt & Yar, 2016; see also Bossler, 2020). The rational choice perspective was also applied to cybercrime when Newman and Clarke (2003) turned the focus of their analysis to e-commerce crimes. Subsequently, situational crime prevention strategies have been applied to different contexts such as those defined by stolen data markets (Hutchings & Holt, 2017), or financial cybercrimes (Leukfeldt & Jansen, 2020), among many others (e.g. Hinduja & Kooi, 2013; Reyns, 2010). Overall, both the routine activity approach and the rational choice perspective have received attention from academics in the last decade and have consequently evolved and contributed to the development of the discipline.

But when it comes to the geometry of crime, there are few studies that apply this theory to cyberspace, except from recent attempts to extend the theory (Miró-Llinares & Johnson, 2018; Miró-Llinares & Moneva, 2019a) and apply it to the prevention of cybercrime (Miró-Llinares et al., 2018). This is probably because this theory depends to a great extent on the concept of place, which is usually associated with a physical space. However, it has been argued that cyber places can be understood as digital spaces of convergence where offenders also interact with the environment that defines crime opportunities (Leukfeldt, Kleemans, & Stol, 2017, 2017, 2017; Miró-Llinares & Johnson, 2018). This reasoning shows that not all concepts within Environmental Criminology are geographical, as some are merely spatial (Miró-Llinares & Moneva, 2019a). An example of purely geographical concepts, at least in their current formulation, are paths and edges. Paths are the routes that connect people's activity nodes, while edges are the boundaries of neighborhoods with distinct socio-cultural characteristics (P. L. Brantingham & Brantingham, 1995). However, there are also spatial concepts such as crime generators —places where many people congregate— and crime attractors —places that create criminal opportunities— (P. L. Brantingham & Brantingham, 1995). Similarly, crime hot spots, which are the result of the repeated occurrence of crime events in a given place and over a certain period of time, have traditionally been measured in physical space, but such concentrations can also be observed in crimes occurring in cyberspace. For example, there may be certain web domains that are more prone to victimization by defacement than others and there may be certain time frames in which the activity of defacers is more intense. In this case, spatiotemporal hot spots of cybercrime will be formed in those cyber places or domains that are repeatedly defaced. What is unknown to date is whether the theory behind repeat victimization is also applicable in cyberspace.

6.3 The present study

In this paper, we aim to test whether the fundamental premises of repeat victimization that apply to some crimes committed in physical space (e.g. burglary) are also observed for defacements in cyberspace. For this purpose, the main premises of some of the most relevant work on repeat victimization have been selected and reformulated as hypothesis for the cybercrime of website defacement.

The first premise states that “high crime rates and hot spots are as they are substantially because of rates of repeat victimization” (Pease, 1998, p. v; see also Farrell & Pease, 2017, 2018). Thus, we derived the following hypothesis for defacements:

H_1 *A substantial share of all defacements and variation in defacements is due to repeat victimization.*

In his original work, Pease (1998) uses the word “substantial” to refer to the fact that repeat victimization accounts for 68% of the total incidents on which the property crime rate is calculated. In a review of 2007 and 2014 studies, Farrell and Pease (2017) find a similar proportion of repeats for personal larceny (58.3%) and robbery (63.9%), but the authors indicate that the real figures are even bigger because they use survey data and surveys under-estimate repeats. By “variation” these authors refer to the year-on-year change in crime rates (Farrell & Pease, 2017). Thus, by adopting a very conservative definition, in order to test this hypothesis, we can define “substantial part of all defacements” as 50%, and to analyze the variation in crime we can examine their distribution over time.

The second premise states that “when victimization recurs it tends to do so quickly” (Pease, 1998; see also Bowers & Johnson, 2005; Farrell, 2005; Farrell & Pease, 1993; Johnson et al., 1997; Johnson & Bowers, 2004). And the hypothesis derived from this premise is the following:

H_2 *After a first defacement event, a repeat incident will occur shortly thereafter.*

Normally, an interval of one year is used to assess repeat victimization (e.g. Chainey, 2012; Farrell & Pease, 1993, 2017). Thus, to determine whether repeat victimization occurs shortly after an initial event, it is necessary to calculate how many domains were defaced more than once within a one-year period.

The third premise states that “repeated crimes are disproportionately the work of prolific offenders” (Pease, 1998, p. vi; see also Farrell & Pease, 2017). Therefore, we can derive the following hypothesis for defacements:

H_3 *Repeat defacements are disproportionately the work of prolific defacers.*

In criminology, this type of Pareto Principle has been studied for both offending and victimization through the analysis of repeat events (Fagan & Mazerolle, 2011; Farrell & Pease, 2017; Pease, 1998), showing that a few victims suffer most crimes, and that a few offenders commit most criminals acts.[58] By testing this hypothesis, we expect to find similar results for website defacements. However, since it can be argued that the type of repeat defacement (i.e. mass or single) can influence the relationship between the number of offenders and the percentage of defacements for which they are responsible, such a distinction should be examined. The reason would be that a single offender could direct mass defacements to many domains, a considerable difference with respect to single defacements —independent events that could only be directed against one domain at a time—. Additionally, total repeat victimization figures could be biased by mass attacks directed to different extensions of the same domain, which would be registered as repeats according to our methodology.

The fourth premise states that “a major reason for repetition is that offenders take later advantage of opportunities which the first offence throws up” (Pease, 1998, p. v; see also Bernasco, 2008; Farrell, 2005; Farrell & Pease, 1993; Lammers et al., 2015). So, the following hypothesis is derived:

[58] Originally, the Pareto Principle —also known as the 80/20 rule— served to establish that about 80% of the results were due to about 20% of the causes.

H_4 *A major reason for repeats is that offenders repeatedly target domains they have defaced previously.*

This premise requires examining how often the same domains are victimized by the same defacers. In addition to this analysis, a distinction made according to the motivation of the offenders seems appropriate, since it may have an impact on their criminal behavior. For example, it would seem logical that those offenders who have no apparent motive for defacing a specific website are not obsessed with targeting the same website again. But in the same way, it could be argued that those with a political motivation or, especially, those who execute their attack for revenge should have an interest in repeatedly directing their attack towards specific targets.

6.4 Materials and methods

6.4.1 Data

We use data from the Zone-H Defacement Archive (http://www.zone-h.org/), a self-reported data source that the defacers themselves supply with their activity. The Zone-H team collects, validates, stores and maintains information about defacement incidents committed by individuals or groups who record their own defacements under a nickname (for an overview of the database, see Romagna & Van den Hout, 2017). Among other variables, this dataset contains information about the date on which defacers submit a request to register an attack, their nickname, their motivation, the type of attack used for the defacement, the URL of the defaced website, and whether the attack is a redefacement of a previously registered domain. In our dataset, the time period in which the defacement incidents are recorded extends from 1 January 2010 to 4 April 2017. After removing 85 records that had incorrectly registered the URL of the defaced website or the type of attack recorded, the dataset contains 9,117,268 registries representing unique defacements to 8,603,658 domains.

6.4.2 Measures

6.4.2.1 *Repeat victimization: Repeat defacements*

To measure repeat victimization, instead of relying on the redefacement variable in the archive,[59] we used the full URLs of the defaced domains; that is, the protocol, the web domain, the path or extension, and additional parameters. We trimmed the URL strings of defaced websites by using the following regular expression:[60]

http://|http://www\\.|https://|https://www\\.|/[:graph:] ∗

This removed all characters except the website domain and we subsequently identified, aggregated, and stored unique domains in a new variable. Thus, repeat victimized domains can be defined as those that appear more than once in the data. By our own calculations we found that repeat defacements represented 5.6% of all attacks, ranging from 1 to 7 repeats.

It is important to note that the Zone-H administrators have established a one-year restriction on the registration of incidents in order to prevent domains from being massively revictimized because their vulnerability is publicly displayed on Zone-H's platform (Zone-H, personal communication, November 21, 2019). So, if a defacer wants to register an attack on a revictimized domain, it is not possible until this period

[59] According to the data, 10.1% of the records are redefacements. However, while inspecting the distribution of the variables that comprise the dataset, we observed an inconsistency in the values of the redefacement variable. We found that 3,301 website domains that appeared more than once in the data (i.e. repeats) were not labelled as redefacements. In addition, we also found 409,183 domains that appeared just once in the data but were labelled as redefacements. This may be due to these domains appearing in previous records that are not part of our dataset.

[60] Regular expressions are sequences of characters that create search patterns in a given field, URLs in our case.

has elapsed, which creates a one-year gap between potential repeats. However, it seems that this restriction does not always work, as some isolated incidents have been recorded within this interval.

The authors are aware that both these circumstances have obvious implications for the phenomenon of repeat victimization explored in this paper. However, to the best of the authors' knowledge, Zone-H remains the best public source of data for studying website defacements and it continues to be valuable to explore patterns of repeat victimization.

6.4.2.2 *Defacers' motivation*

When recording an attack, defacers must fill out a short form that includes a drop-down list of possible reasons that motivated the defacement. Defacers can choose one of the following six categories: "Heh… just for fun!", "as a challenge", "I just want to be the best defacer", "political reasons", "patriotism", and "revenge against that website". Since some of these categories seem not exclusive and may overlap, we have proceeded to regroup them into four categories: "Fun" includes the first category, "challenge" includes the next two; "politics" includes the fourth and fifth; and "revenge" remains alone. Thus, defacements performed for fun represent 54.8% of the records, those executed as a challenge account for 23.4% of the records, while those perpetrated for political reasons account for 9.4%, and those seeking revenge for 4.1% of the total (Holt, Leukfeldt, et al., 2020). The motivation behind the remaining defacements is unknown. Although data aggregation causes some loss of information, we believe that the new categories are better delimited and facilitate the interpretation of the results.

6.4.2.3 *Type of attack: Single and mass defacements*

Another variable that describes the nature of defacements is the type of attack involved, which can be "single" or "mass". As opposed to single attacks, mass defacements represent attacks that target several websites in a short interval of time. Single attacks account for 23.6% of defacements, compared to 76.4% for mass attacks.

6.4.3 Analytic strategy

Repeat victimization has been analyzed in the same consistent manner over the past few decades (Farrell & Pease, 1993, 2017). "The preferred way of analyzing repeat victimization is to establish a set assessment period (usually twelve months), then identify initial victimization of each unique target and determine whether the target was re-victimized in the assessment period following that initial victimization" (Chainey, 2012, p. 1). This strategy, known as the rolling period methodology, is also followed in the present paper with a slight modification. Since Zone-H restricts registrations of defacements of the same website within a one-year period, we were not able to maintain a one-year period to assess whether repeat victimization exists. Thus, the analyses were carried out considering an indefinite time series in order to observe whether there is repeat victimization regardless of the time gap, and to understand its complete scope.

In addition, our third hypothesis requires analyzing the extent to which repeat defacements are concentrated among the defacers in our sample. To that end, we used Fox and Tracy's (1988) proposed coefficient to measure skewness in offense distributions.[61] This measure facilitates comparison of the results with those obtained from other studies.

Data transformation, string manipulation, and data visualization were executed using the tidyverse R package version 1.2.1 (Wickham, 2017) in RStudio version 1.2.5001 for the R free software version 3.6.1. Data transformation involved: Reshaping

[61] By formula, $\alpha = 2\sum P\,n_k\left(\frac{Po_k}{2} + cPo_{k+1}\right) - 1$, where Pn_k is the proportionate size of our sample of defacers with exactly k offenses; Po_k is the proportion of defacements executed by defacers with exactly k defacements, and cPo_{k+1} is the proportion of the defacements executed by defacers with at least $k + 1$ offenses.

data to change its layout; summarizing, grouping, and manipulating cases to return new values; manipulating variables by extracting them or making new ones; and combining data tables. String manipulation was essential in our analyses as it allowed us to define, by means of regular expressions, a new unit of analysis for repeat victimization: web domains. Regarding data visualization we used a staircase or step chart to visualize the results for the first premise, bar charts to compare the results obtained to explore the second premise, and histograms to show the distribution of repeat victimization for premises three and four. Due to the extremely skewed distribution of the data, for some figures we used a transformed y-axis by means of a $\log_{10}(x)$ to facilitate their visualization. Some charts include annotations.

6.5 Findings

This first hypothesis requires calculating which share of total recorded defacements corresponds to repeats, as shown in Figure 5. Because 2010 is the initial year —and there is a one-year gap in repeat victimization—, and 2017 only contains data for the first four months, we omitted these two years from the data and found that repeats per year only represented 7.1% of total defacements (*SD* = 3.3) with a minimum of 2.3% in 2011 and a maximum of 11.1% in 2016.[62] Next, a Pearson correlation test was conducted to assess the relationship between total and repeat counts of defacements using aggregate figures per month. We found a very weak non-significant relationship between the two figures (*r*(70) = 0.072, p = 0.545) showing that the contribution of repeat defacements to the annual variation in the crime rate was minimal.

[62] We repeated the analysis including all data points and obtained a slightly lower proportion of repeats of 6.3% (*SD* = 3.7), with a minimum of 0.2% in 2010, and a maximum of 11.1% in 2016.

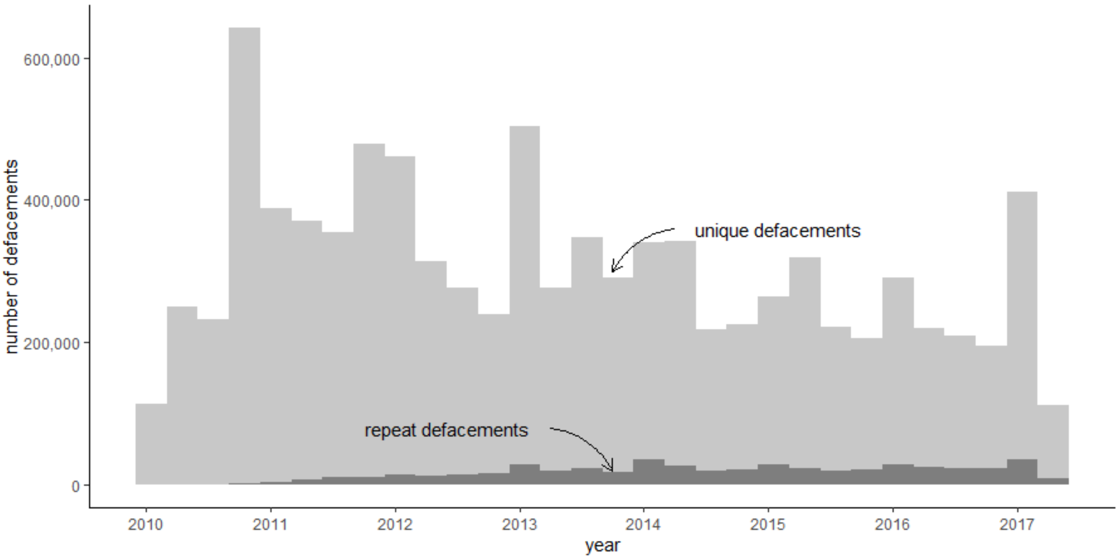

Figure 5. Distribution of repeat defacements to total defacements. Histogram bins = 30

Following the rolling period methodology, repeat victimization sequences were identified and a number was assigned based on their order (i.e. 1st victimization, 2nd victimization, etc.) to test the second hypothesis. Next, the distribution of repeats was analyzed to calculate the amount of time between the intervals (Table 5). This revealed that the average time interval between repeat victimizations was 440.3 days (*SD* = 158.7). Although the average duration between the first defacement and the first repeat victimization was almost 690 days, this figure decreased after each repeat. This seems to be influenced by those defacements that were recorded on the same day as the original victimization causing a reduction in the mean value. Considering the skewed distribution of the repeats, it is worthwhile to highlight the figures corresponding to the first quartile, which are consistently around a year in every interval after the first repeat victimization.

Figure 6 serves to illustrate that these patterns were still visible even after several years, although with each victimization that occurred the number of repeats was lower. Note that Figure 6 displays a modified y-axis to visualize this otherwise unnoticeable pattern. For the same Figure with an unmodified y-axis see Appendix A.

Table 5.
Time lapse between repeat victimization intervals

Repeat victimization interval	Repeats per interval		Time in days between repeats						
	n	%	Min	1Q	Mdn	3Q	Max	M	SD
First	450,278	4.9	0.0	402.8	527.2	832.3	2638.4	689.6	408.0
Second	52,336	0.6	0.0	373.3	426.6	617.1	2347.2	548.9	275.1
Third	9,054	0.1	0.0	369.0	390.2	493.2	1737.9	472.2	185.9
Fourth	1,696	0.0	0.0	366.7	371.8	413.1	2283.7	421.5	127.8
Fifth	218	0.0	0.0	366.4	372.4	398.7	914.7	410.4	102.7
Sixth	26	0.0	0.0	366.2	366.2	378.5	459.3	364.7	78.2
Seventh	2	0.0	0.0	-	-	-	366.2	183.1	258.9
Total mean								440.3	158.7

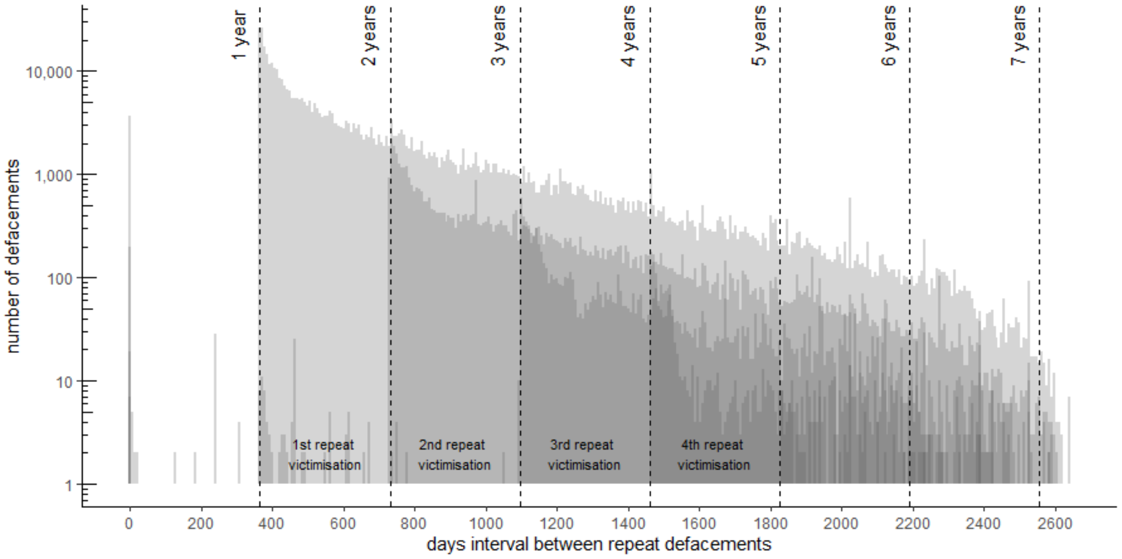

Figure 6. Repeat victimization time pattern for website defacements. The Figure displays a transformed y-axis by means of $log_{10}(x)$. Histogram binwidth = 7

There were 66,648 defacers responsible for 9,117,268 defacements. Of these, 30,935 (46.4%) defacers only executed one attack, suggesting that clustering exists. However, while most defacers performed few attacks, others launched many (*Min* = 1; *Q1* = 1; *Mdn* = 2; *3Q* = 9; *Max* = 303,442; *M* = 136.8; *SD* = 2764.7). And there were others who concentrated their attacks on the same website domains; specifically, 17,026 defacers did so, committing 513,610 repeats. To test our third hypothesis, we analyzed what percentage of these repeats were carried out by a particular percentage of offenders.

The results in Figure 7 show that 1% of redefacers committed 57.8% of repeat defacements, and that 50% of redefacers committed 98.2% of repeat defacements. Fox and Tracy's (1988) measure for skewness shows a very high concentration of repeat defacements among defacers (α = 0.906). The same distribution was also examined according to the type of attack, whether single or mass. As illustrated in Figure 7, this distinction shows that single attacks (α = 0.881) were slightly more concentrated per offender than mass attacks (α = 0.877). Regardless of the type of attack, 1% of redefacers were responsible for more than 46% of repeat defacements, and 50% of redefacers for more than 96% of repeat defacements. Detailed data tables can be found in Appendix B.

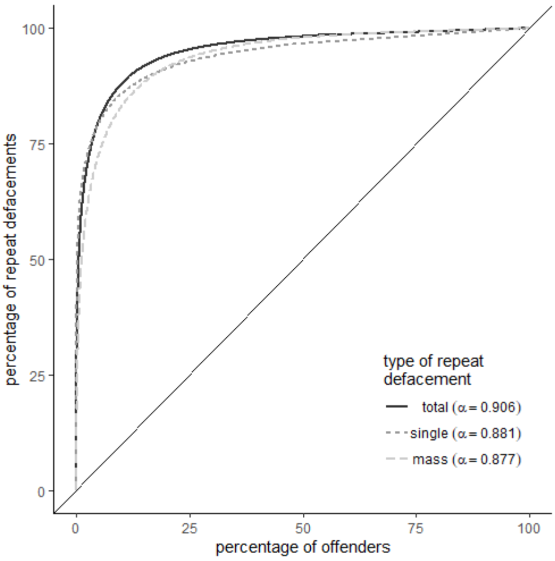

Figure 7. Percentage of offenders responsible for a percentage of defacements

After grouping all defacements by offender, the analysis shows that offenders rarely defaced the same domains that they had previously defaced; in fact, this only occurred 0.3% of the time (Table 6). When repeats are distinguished according to the motivation of the offenders, the results show that most of the defacers who did it did so for fun. Interestingly, revenge-driven defacers committed the least repeat attacks against the same website.

To test the fourth hypothesis, the next step was to analyze whether the number of times the same offenders attacked the same domains was a major reason for repeat victimization. This can be calculated as follows:

$$\frac{\text{n repeats by the same offender to the same domain}}{\text{n total repeats}} = \frac{31{,}841}{513{,}610}$$

The results show that 6.2% of repeat victimization was due to the same offenders defacing the same domains repeatedly.

Table 6.
Frequency with which a domain has been victimized by the same offender and its motivation

Number of times victimized by the same offender	Any motivation		For fun		As a challenge		Political reasons		For revenge	
	n	%	n	%	n	%	n	%	n	%
1	9,052,741	99.7	4,958,735	99.6	2,124,096	99.8	851,171	99.8	367,475	99.9
2	31,036	0.3	17,786	0.4	3,537	0.2	1,477	0.2	369	0.1
3	775	0	341	0.0	58	0.0	23	0.0	4	0.0
4	23	0	13	0.0	4	0.0	2	0.0	0	0.0
5 +	7	0	3	0.0	4	0.0	0	0.0	0	0.0

Note: Total defacements by any motivation = $\sum_{i=m}^{n} i * a$ = 9,117,268; where i = number of times victimized, and a = frequency of victimization. Total defacements in the dataset is greater than total motivations due to a small number of defacements being of unknown motivation.

6.6 Discussion

Drawing on a unique database containing millions of self-reported cases, this paper addresses the important question of whether traditional criminological theories developed in the pre-digital era can still be used to explain cybercrimes. In the same way that criminological research has explored the utility of traditional criminological theories to understand cybercrime (Bossler, 2020; Holt & Bossler, 2017), in this paper we explored whether some of the main premises of Environmental Criminology related to repeat victimization of traditional crimes also apply to the cybercrime of website defacements. After noting that the phenomenon of repeat victimization was also observed for this particular cybercrime, we examined: Whether it constituted a substantial fraction of crime rates and their variation over time, whether it occurred shortly after the first incident, whether a few defacers were responsible for most repeats, and whether this was largely due to the same offenders defacing the same domain over again. The results suggest that some of these premises of traditional repeat victimization could also be valid in the case of website defacements.

Firstly, we observed that, despite the fact that Zone-H does not register defacements on the same domain within one-year after the first defacement, the contribution of repeat events to the total website defacement rate was still relevant. However, the volume it represented is minimal compared to that observed for traditional property crimes. While repeats represented 63.9% of robberies and 58.3% of personal larceny (Farrell & Pease, 2017), repeat defacements represented an average of 7.1% between 2011 and 2016. The most likely explanation for this large discrepancy is that repeat victimization patterns observed in Zone-H are limited by the one-year time interval after the original incident for a phenomenon that is essentially characterized by being temporarily concentrated shortly after the first event (Bowers & Johnson, 2005; Farrell, 2005; Farrell & Pease, 1993; Johnson et al., 1997; Johnson & Bowers, 2004; Pease, 1998). Thus, it is likely that the results are highly underestimating the share of repeats.

This would imply that the formation of spatiotemporal hot spots of defacements in cyberspace could still be caused by repeat victimization. On the rationale that the best predictor of future behavior is the past, hot spot analyses have traditionally been used to predict future crime events for prevention. Since these techniques rely heavily on repeat events, it could be argued that such algorithms would still be effective in the case of website defacements.

Secondly, we observed that some website domains registered in Zone-H suffered between 1 and 7 repeats after the initial defacement, although the prevalence decreased exponentially after each repetition. In addition, our results indicate that some repeat events were still recorded within the one-year restricted registration period, suggesting that this measure established by Zone-H has some flaws. Because of the large data set used in this study, it was possible to detect patterns of repeat victimization that might have gone unnoticed with less data. In this sense, even though we cannot determine whether repeat victimization occurs shortly after the first incident due to the one-year period established as a restriction to record repeat attacks, the sharp distribution of the data, with a number of redefacements shortly after the end of the one-year restriction, suggests that a large volume of defacements would be observed in that initial period if there were no such restriction. This claim is reinforced by the results of a study on network attacks on computer systems, in which researchers found that repeat victimization was most likely to occur within the first week after a previous attack (Moitra & Konda, 2004).

Hence, crime prevention measures such as cyber-attack detection systems should be specifically intensified immediately after the first victimization so that they can have an effect on the peak hours, when most events occur. Prevention efforts could also benefit from enforcing guardianship by incorporating place managers such as SSL security certificates and ensuring they do not expire to prevent man-in-the-middle attacks.

Thirdly, while research on traditional crime shows that most offenses are committed by few offenders, repeat cyber offenders seem to be more prolific. This phenomenon was observed when exploring the third premise and represents an exacerbation of the Pareto Principle identified in previous criminological studies. For example, a cross-national comparative study in London and Stockholm showed that about half of the offenses were committed by 2% of the offenders (Farrington & Wikstrom, 1994), and using data from the Philadelphia birth cohort, researchers found that 6% of young males in the sample accounted for 52% of arrests (Fox & Tracy, 1988). In this particular study, Fox and Tracy show that the concentration of offenses was considerably high in the cohorts of 1945 ($\alpha = 0.816$) and 1958 ($\alpha = 0.838$). Compared to the alpha coefficients described by Fox and Tracy (1988), the concentration of repeat offenses among defacers was even higher, both in absolute terms ($\alpha = 0.906$) and for each type of defacement (single, $\alpha = 0.881$; mass, $\alpha = 0.877$).[63] Our results show that 1% of offenders accounted for over 57% of the repeat offenses. Moreover, when the repeat event was a single attack, these figures were further accentuated, as 1% of defacers were responsible for 64% of repeats.

Note that instead of analyzing which percentage of offenders commits which percentage of crimes, in our study we examined how repeated attacks were concentrated as a function of the percentage of defacers. Looking at these figures, it is likely that defacements will be even more concentrated if we consider all victimizations rather than just repeats. It should also be noted that defacers registered in Zone-H may not exclusively be individual offenders, but groups of offenders jointly registering their attacks. Conversely, hackers may change their identity by registering a new attack using an alternative nickname. In any case, the concentration figures would probably vary.

[63] An important difference between this study and those of Fox and Tracy (1988), and Farrington and Wikstrom (1994), is that their samples also include non-offenders. So, if 50% of their sample does not commit a crime, then 50% of the offenders would be responsible for 100% of the crime. Since in our study we calculated the concentration in a sample of offenders only, it is possible that our figures are even underestimated in comparison.

Considering all possible scenarios, it is safe to claim that the concentration of crime perpetration among a few prolific offenders is also observed for website defacements Therefore, it is possible that focused deterrence strategies that have served to reduce violent crime in physical space (Braga, Zimmerman, et al., 2019; Kennedy, 2012) can be adapted to the particularities of defacers to be effective in reducing the impact of repeats in this type of cybercrime.

Lastly, our analysis shows that a few offenders returned to deface the same domains even one year after their initial attack, regardless of their motivation. It seems that the benefits obtained by these offenders from the first attack were sufficient to again exploit the opportunities that allowed the previous defacement. This suggests that the theoretical rationale for repeat victimization based on the "boost" could still be valid for website defacements. However, we also found that repeat defacements from the same offenders on the same domains contributed little to the total ratio of repeats (6.2%) compared to burglaries (see Bernasco, 2008; Lammers et al., 2015). So, although a few defacers were responsible for a large part of repeat victimizations, these were not concentrated within the same domains. Instead, because defacements occurred on many different websites, it could be argued that their vulnerabilities are constant and can be exploited by any defacer. In fact, hacking through known vulnerabilities is one of the most prevalent hack modes used to deface websites (Holt, Leukfeldt, et al., 2020; Romagna & Van den Hout, 2017). It would seem, therefore, that the "flag" explanation could explain repeat defacements too. Nevertheless, the one-year gap in the data might be a reason for the low number of observed repeats that were also executed by the same offender. After a year defacer's motivations may change: the political agenda may be different, feelings of revenge may ease, and new challenges and sources of fun other than website defacement can be found. In such cases, our findings would be under-representing the phenomenon of repeated victimization.

The adoption of situational crime prevention measures could be a valid option for preventing defacements that has already been explored for other cybercrimes (Hutchings & Holt, 2017; Leukfeldt & Jansen, 2020; Reyns, 2010). These measures would include target hardening techniques such as patches for known vulnerabilities and exploits that would help to prevent SQL injections. Such measures could both discourage the boosted offender and reverse the flagged vulnerability of website domains.

6.7 Conclusion

In this paper we explored four premises of repeat victimization on website defacements from the perspective of Environmental Criminology. Based on the concept of cyber place, we presented an analysis that pivots on the essential premises of repeat victimization. In particular, we found that repeat victimization may contribute to high crime rates of defacement; that it occurred even several years after the initial attack; that most repeat defacements were also committed by only a few offenders; and that in only a few cases offenders repeatedly targeted those domains that they had successfully defaced in the past. These results suggest that some of the traditional premises of repeat victimization may also apply to this type of cybercrime, thus advancing the discipline in the field of criminological theory. This work also contributes to crime prevention by uncovering distinct spatiotemporal patterns of crime that can be tackled with appropriate resources and strategies.

However, this work also has limitations. Although we used the richest existing data source to study website defacements, Zone-H's one-year data recording restriction policy undermines understanding the full extent of repeated victimization. Yet, the more than 9 million website defacements analyzed reveal previously unstudied victimization patterns which are useful to generate both basic knowledge about the phenomenon and applied knowledge for prevention, even when random repeats were not examined (S. min Park & Eck, 2013).

In order to examine the application of criminological theories to cybercrime, more research is needed that focuses on well-defined premises applied to specific cybercrimes. Since this paper presents an initial assessment of repeat victimization, a possible course of action would be an in-depth examination of the explanations regarding the boost. This would contribute to a better understanding of the characteristics of repeatedly targeted cyber places. Until we understand how causal mechanisms work on a small scale, we will be unable to fully grasp the bigger picture of the most complex theories. Future research should also focus on the applicability of the premises identified in this work by contrasting them with other types of cybercrime and better data.

CHAPTER VII

EXAMINING SPORT BETTING CYBER PLACES TO DISRUPT CRIMINAL NETWORKS

This chapter has been published as: Moneva, A., & Caneppele, S. (2019). 100% sure bets? Exploring the precipitation-control strategies of fixed-match informing websites and the environmental features of their networks. *Crime, Law and Social Change*. https://doi.org/10.1007/s10611-019-09871-4

7.1 Introduction

In recent years, many illicit activities (e.g., fraud, child pornography, harassment) have made cyberspace their preferred environment (Holt & Bossler, 2014). Solo offenders, criminal networks, and other groups move to online environments because of new criminal opportunities (Morselli & Décary-Hétu, 2013). These actors use cyber environments to commit crimes, organize their illicit activities, establish new links with other networks, and recruit new members, thereby facilitating the diversification of their criminal activities from, for example, traditional fraud to phishing (Leukfeldt et al., 2017a). For criminals, online fraud is among the most prevalent and beneficial types of cybercrimes and does not require sophisticated skills beyond the motivation for financial gain (Cross & Blackshaw, 2015). Indeed, financial gain is one of the reasons why criminal networks still incorporate low-tech all-round to high-tech specialists (Leukfeldt et al., 2017b). The few requirements for committing crimes in cyberspace in terms of both skill and resources, together with the new criminal opportunities generated by the accessibility to this space, favour the emergence of new forms of online frauds together with online fraud markets. By fraud markets, we refer to markets through which people sell counterfeit goods (e.g., fake passports), stolen data (e.g., carding) or services (e.g., tutorials, botnets, confidential information) to facilitate further fraud. Fixed-match informing websites (FMIWs) belong to this category of markets.

In this study, FMIWs are cyber places where users buy and sell information on alleged fixed sports results. Potential users include sport-betting punters who want to place their money on fixed matches for the highest return on their investment—minimizing the risk—. An alleged market of fixed results, if false, can turn users into defrauded victims and, if true, can fuel corruption in sports. This phenomenon may be enrolled under the larger issue related to match-fixing and sports betting. Match-fixing affairs are not new in sports (Huggins, 2018), but their relevance has grown since the 2000s. Online betting boosted opportunities to place sports bets from around the world on many types of disciplines and competitions. As the size and complexity of the betting market increase, so do the size of opportunities to make money illegally (Forrest, 2012). One fraudulent way deals with adjusting sports results and earning money on sure bets. Although match-fixing is not always related to betting, the betting-related dimension of match manipulation is a crucial concern among sports federations (Moriconi & Almeida, 2019). There are two main reasons for this concern. First, sports betting is now an essential source of revenue for many disciplines, and match-fixing scandals may hamper it (Tak, 2018). Second, as suggested by the Interpol Match-Fixing Task Force, the prospect of big profits with minimal cost—in terms of risk—has led criminals to seek profit opportunities in this area [64] with negative consequences for the sport movement.

By recording recent trends in this matter, the Sports Betting Integrity (ESSA) entity reported growing numbers of match-fixing incidents to competent authorities since 2015 [65], mainly related to tennis and football (ESSA, 2015, 2016, 2017, 2018). FMIWs claim to possess insider information about the outcome of a fixed match that is then fraudulently sold online. This service can be sold via the darknet and the clear web. Some recent indications point to the existence of illegal online betting platforms and darknet forums, where the results of fixed matches are marketed (CK Consulting & Stichting VU-VUmc, 2017). Additionally, taking advantage of the easy dissemination of content in the clear web, and claiming to have privileged information about these fixed matches, some websites also advertise the sale of related information on results, hoping to seduce potential buyers. Although observing the websites promoting such activities cannot confirm whether they actually possess the information they claim, what is quite evident is that, in one way or another, they are promoting fraudulent illegal activities.

Assuming postulates of Environmental Criminology theories are also valid in cyberspace, in this study, we aim to understand which elements of the FMIWs favour the onset of specific criminal opportunities. In particular, we use situational precipitators of crime to examine the extent to which these websites encourage users who visit them to engage in deviant behaviour. This study contributes to the literature by applying an analytical framework to the problem of FMIWs to identify their unique environmental features that facilitate their detection. We also explore the URLs contained in these websites to explore their network of connections, thereby facilitating

[64] https://www.interpol.int/Crimes/Corruption/Corruption-in-sport
[65] ESSA reported 100 incidents in 2015, 130 in 2016, 266 in 2017, and 267 in 2018.

a better understanding of their organization. We employ a network analysis technique that allows us to reveal other cyber places that comprise this network while identifying the primary nodes in this structure. We show a method for disrupting such networks that can be employed by law enforcement agencies.

7.2 Places in cyberspace: An opportunity-precipitation framework

According to environmental criminologists, the place where a crime occurs is the key organizing feature for crime analysis (Weisburd et al., 2016). However, in a review of the book, Place Matters: Criminology for the Twenty-First Century (Weisburd et al., 2016), Clarke (2018) argues that, for crimes committed in cyberspace, as well as some types of fraud, the role of geographic location is hardly relevant. Instead, the important issue is the convergence of an offender with an environment of opportunity, which does not necessarily have to be geographical. When a motivated offender takes advantage of such opportunities in cyberspace to commit crimes, we refer to those digital convergence settings as cyber places (Miró-Llinares & Johnson, 2018).

Similar to physical places, there are different types of cyber places whose characteristics favour or hinder the concentration of specific crimes within them. While in cyber places, such as social media—where personal interaction is more frequent—one can expect a substantial incidence of social cybercrimes such as harassment, sexting, or hate speech. Cyber places devoted to consumer activities, such as shopping or banking, will host a different criminal phenomenology that is more financial. For example, some research reports that older students who spend more time in chatrooms, and younger adults who frequently use social media, are more likely to experience online harassment victimization (Marcum, Higgins, et al., 2010; Näsi et al., 2017). Additionally, individuals who perform online activities like banking or shopping are more likely to experience identity theft (Reyns, 2013) or be defrauded (van Wilsem, 2013a). There are two main interconnected reasons for these findings. First, the configuration of cyber places shapes the range of actions available to their users. Second, the type of activity carried out by users in an online environment affects the criminal opportunities that proliferate there. As for consumption platforms dedicated to selling products or offering services (e.g., eBay, Amazon), their configuration permits certain actions for e-commerce, and the opportunities derived from such activities make them particularly attractive for committing financially motivated cybercrimes.

Cyber places, such as websites that claim to sell results of fixed matches, can be particularly attractive for potential buyers in terms of costs versus benefits. How administrators of these websites advertise the feasibility of profiting from such activity can lead users to buy their services. Without any proper crime control websites, offering fixed matches can quickly become crime attractors. Such places are appealing to offenders because they offer particularly attractive criminal opportunities in terms of cost-effectiveness (Brantingham & Brantingham, 1995). Additionally, FMIWs' configurations are such that the mere act of visiting them constitutes a tempting situation to buy the products they offer.

According to Wortley (Wortley, 1997), there are certain situations that prompt or provoke individuals to engage in criminal behaviour. Some of the features on these websites are situational precipitators of crime and, therefore, the Precipitation-Control Strategies established by Wortley (2001) (See Table 7), can be used classify specific strategies for avoiding them. Following the opportunity-precipitation model (Wortley, 2001), crime may be preventable by (a) avoiding precipitating criminal behaviour initially and (b) reducing opportunities to commit the crime in a subsequent stage. This model operates within Situational Crime Prevention (SCP) strategies—a set of practical measures proven highly effective in reducing crime in particular contexts (Clarke, 1997). Beyond the SCP measures that have been implemented in physical spaces, the foundations on which strategies are built have proven sufficiently robust to develop applications for online environments (G. R. Newman & Clarke, 2003). SCP models have been used to approach problems in online stolen data markets (Hutchings & Holt, 2017), develop preventive strategies for e-commerce crime (G. R. Newman & Clarke, 2003), reduce information security vulnerabilities (Hinduja & Kooi, 2013), and examine DDoS operators (Hutchings & Clayton, 2016). Overall, the literature shows that the adaptability of such strategies to new phenomena is as great as researchers' can imagine, although there is little evidence of the results of their application to crimes committed in cyberspace.

Table 7.
Classification of precipitation-control strategies

Controlling Prompts	Controlling Pressures	Reducing Permissibility	Reducing Provocations
Controlling triggers	Reducing inappropriate conformity	Rule setting	Reducing frustration
Providing reminders	Reducing inappropriate obedience	Clarifying responsibility	Reducing crowding
Reducing inappropriate imitation	Encouraging compliance	Clarifying consequences	Respecting territory
Setting positive expectations	Reducing Anonymity	Personalizing victims	Controlling environmental irritants

Source: Adapted from Wortley (2001)

7.3 Aims of the study

The analysis of FMIWs has received little attention in academia. Most of the research addresses the broader topic of match-fixing in international sports (Haberfeld & Sheehan, 2013). Aiming to fill this gap in the literature, this paper focuses on (alleged) FMIWs and their networks. Adopting Wortley's (2001) Precipitation-Control Strategies framework, we hypothesize the following:

H1. FMIWs offer specific crime opportunities because they incorporate distinctive environmental features that incentivize deviant behaviours (i.e. buying fixed matches results) when compared to regulated sport-betting websites.

H2. Due to the peculiarity of this cyber environment, vending places for fixed matches have a specific network compared to a random network distribution.

7.4 Method

7.4.1 Sampling: Detection and selection of the websites

This study follows a methodology similar to that proposed by Pineau et al. (2016) to obtain a sample of websites from the clear web related to fixed matches. After defining a list of keywords [66], they were entered into the TOR browser using the DuckDuckGo search engine, a strategy followed to improve anonymity. Then, the first 50 results for each keyword were manually checked (i.e. 200 URLs visited) to determine whether these websites offer information in exchange for money about supposedly fixed matches. Through this process, 78 websites that met the inclusion requirements were identified as an FMIW (Appendix D Table 23). To determine the extent to which the characteristics that define FMIWs as crime attractors differ from other cyber places, a second set of websites was selected for comparison purposes. The authors considered a list of 28 regulated sport-betting sites (Appendix E Table 24). To ensure that this second group of websites had a legitimate origin, we referred to the list of members belonging to two official international entities that promote integrity in betting: The World Lottery Association (WLA), and ESSA.

[66] (1) match-fixing, (2) fixed betting tips, (3) fixed matches, (4) fixed-odd sports.

7.4.2 Analytic strategy

Two analysis techniques were used to achieve the established objectives set, including (1) systematic observation, to detect the situational features of the websites and (2) network analysis, to describe the structure of fixed matches vending cyber-places.

7.4.2.1 *Systematic observation*

Systematic observation in the social sciences is based on the identification of a series of items in a specific context whose presence or absence can be objectively determined (Mastrofski et al., 2010; Reiss, 1971). For example, this methodology has been used to quantify the social and physical properties of neighbourhoods such as urban disorder (Raudenbush & Sampson, 1999; Sampson & Raudenbush, 1999), or to study police work in public settings (Mastrofski et al., 1998).

This study proposes a modality of systematic observation to compare cyber places that allows for quantifying the situational features that configure them. Through an observational process, we first identified 15 items that usually define the environmental design of sport-betting websites. Next, we adapted and classified each item as a technique under precipitation-control strategies (Wortley, 2001). The systematic observation was conducted on two subsets: (1) FMIWs and (2) regulated sport-betting websites. After recording the elements observed on both illicit and regulated web pages, we compared the results to determine which of these cyber places incorporate more techniques that regulate behaviours of the users who visit them. In theory, the subset of websites that incorporates fewer of these features in its design will have less control over the behaviour of its users, a circumstance that may turn them into crime attractors. On the contrary, a greater presence of features appears on regulated websites. Table 8 shows a description of the items that were observed and subsequently checked for each of the sampled websites.

Table 8.
Situational precipitators, and specific observed items on sport-betting websites with a description

Situational precipitator typologies by item	Description
Controlling prompts	
Controlling triggers	
Advertisements of other betting sites	The website does not incorporate a banner linked to an external betting site.
Providing reminders	
Self-restriction measures	The website facilitates tools or utilities for users to limit their betting.
Advice on abusive gaming	The website provides tips for detecting signs of or resources for mitigating abusive gambling.
Setting positive expectations	
Operator and contact information	The website exhibits legal information of the site operator as well as visible contact channels.
License number/model	The website displays a license model or number authorizing the activity.
Privacy and cookies policy	The website has a privacy policy that includes a cookie policy.
Controlling pressures	
Reducing anonymity	
Registration/login system	The website integrates a user login system for accessing its services.
Reducing permissibility	
Rule setting	
Required payment methods	The website specifies which payment systems are allowed.
Terms and conditions of use	The website has a guideline of terms and conditions of use of its services.
Protection of minors	The website has a policy of restricting access to minors.
Clarifying consequences	
Copyright information	The website shows the copyright information of its domain.
Reducing provocations	
Reducing frustration	
Help/FAQ section	The website contains a user help section or frequently asked questions.
Site language options	The website allows the user to change the language in which its contents are communicated.
Controlling environmental irritants	
Menu	The website has a menu that facilitates navigation.
Smooth, responsive interface	The website has a pleasant and functional interface that makes navigation enjoyable.

7.4.2.2 *Hyperlink network analysis*

The second objective of the research was to survey the network structure among the sampled FMIWs. We used hyperlink network analysis (HNA) to review the linked websites (H. W. Park, 2003; H. W. Park & Thelwall, 2006; Thelwall, 2004). HNA focuses on relationships among websites and recalls the same techniques and metrics used by social network analysis, which focuses on social relationships. For example, researchers have used HNA to explore the structure of online child sexual exploitation networks in a criminological context (Westlake & Bouchard, 2016), as well as to examine the structure of white supremacist online communities in a sociological one (Burris et al., 2000).

To collect data on websites' relationships, we implemented the use of a web crawler that allows data scraping with the R software using the RCrawler package (Khalil & Fakir, 2017). This package offers a function that facilitates the retrieval of external links from a given website and their storage in a data frame with an appropriate structure (i.e., which websites the links come from and to where they are directed) to apply network analysis (Wasserman & Faust, 1994). We then pre-processed the stored URLs to identify unique domains. This process enabled the creation of a targeted network wherein the nodes are websites, and the edges are their connections, represented by a linking URL. In all, 923 unique cyber places were identified within the network with 2306 links between them. We then calculated several standard network metrics, including density, reciprocity, diameter, and mean distance. We compared the obtained results with those of 1000 simulated networks that share the same characteristics as the observed one (i.e., direction, density, number of nodes, and number of edges). All network analyses were performed with the igraph package in R (Csárdi & Nepusz, 2006).

7.4.3 Ethical issues

Results appear in aggregate format and omit any information that could lead to the individualization of users. However, the researchers cannot assume responsibility if this type of information is publicly available by website administrators on some of the websites that appear listed in Appendix D Table 23.

7.5 Results

7.5.1 Comparison between regulated sport-betting websites and illicit FMIWs

Table 9 shows the different characteristics observed between legitimate cyber places and those supposedly selling fixed-match results. The results indicate that all precipitation-control techniques manifest themselves more often on regulated websites than they do on FMIWs ($\chi^2(14, N = 15) = 400.54, p < .001$). Further, 9 of 15 techniques appear on all regulated websites, and the other six appear in more than 50% of cases. Three techniques never appear on FMIWs, and seven additional techniques are present less than 10% of the time.

Table 9.
Differential presence of precipitation-control strategies by type of cyber place

Precipitation-control strategies and techniques	Regulated websites (n = 28)		FMIWs (n = 76)	
	n	%	*n*	%
Controlling prompts				
Controlling triggers				
Avoid other betting sites advertisements	28	100.0	4	5.3
Providing reminders				
Facilitate self-restriction measures	23	82.1	0	0.0
Advise on abusive gaming	28	100.0	2	2.7
Setting positive expectations				
Exhibit operator and contact information	28	100.0	1	1.3
Display a license number/model	19	67.9	0	0.0
Have a privacy and cookies policy	28	100.0	5	6.7
Controlling pressures				
Reducing anonymity				
Enable registration/login	28	100.0	0	0.0
Reducing permissibility				
Rule setting				
Set payment methods	17	60.7	37	49.3
Establish terms and conditions of use	27	96.4	8	10.7
Discourage the participation of minors	28	100.0	5	6.7
Clarifying consequences				
Show copyright information	19	67.9	44	58.7
Reducing provocations				
Reducing frustration				
Provide help/FAQ	28	100.0	9	12.0
Enable site language options	15	53.6	1	1.3
Controlling environmental irritants				
Embed a menu	28	100.0	57	76.0
Design a smooth responsive interface	28	100.0	2	2.7

There are vast differences in almost all the techniques described. A paradigmatic example is the technique aimed at controlling triggers (e.g., avoid other betting sites' advertisements), which always appear on regulated websites, but only in 5.3% of fixed matches ones. The remaining illicit websites embed advertising banners that redirect users to other fixed matches domains, thereby forming a network of websites.

7.5.2 Analysis of the FMIW network

After visiting each of the FMIWs that compose the nodes of the network, we assigned them an additional attribute that indicates the type of cyber place they are. These assigned attributes indicate whether each site is one of the following: (1) sites that trade fixed match results; (2) regulated sport-betting sites; (3) social media sites; (4) platforms that offer web services or utilities; (5) online payment systems; and (6) other cyber places. The last category includes websites that did not belong to any of the previous categories, as well as links that were outdated, broken, expired, or redirected to different websites. The distribution of nodes, according to the type of cyber place they are, appears in Table 10. When examining the nodes of the network, besides those categorized as fixed match sites, some websites and web applications commonly accessed by Internet users were found. For example, the crawler captured regulated betting sites such as William Hill, Bet365, and 188bet (Appendix F Table 25); web services, including WordPress, SurveyMonkey, and Imgur; social media sites like WhatsApp, Instagram, Facebook, Twitter, and YouTube; payment systems like Western Union, PayPal, Bitcoin, MoneyGram, Skrill and Neteller; and other websites, such as the Gmail email system, the top Spanish football competition LaLiga, the European law enforcement agency Europol, Wikipedia, the iTunes platform, and the Daily Mail newspaper. Although it has undoubtedly been detected that these nodes belong to the observed network, their inclusion does not imply that they do so willingly; instead, they were likely hyperlinked without their consent to some of the FMIWs.

Table 10.
Network composition by type of node

	Composition (*n* = 923)	
Type of node	*n*	%
FMIW	715	77.5
Regulated betting site	26	2.8
Web service	14	1.5
Social media	7	0.8
Payment system	7	0.8
Other	154	16.7

An initial scan of FMIWs shows that they tend to include advertisements of other similar sites, suggesting that they may be connected. Of the 78 websites initially sampled, all are interconnected except two, causing the resulting network to consist of a large graph made up of 866 nodes, a small graph comprising 55 nodes, and a micrograph of 2 (Figure 8). To examine the entire network's cohesiveness, we calculated its density, which measures the ratio of observed edges to the number of possible edges. Our network has a density of 0.003 (0.3%), indicating that its nodes are poorly connected.

We then calculated three additional metrics that help to describe the network further. These metrics included (1) reciprocity, which accounts for the proportion of bidirectional links between nodes; (2) diameter, which measures the size of the network by calculating the length of the longest observed geodesic distance; and (3) mean distance, which represents the mean length of all the shortest paths leading to or coming from each vertex. We compared the obtained results with those of 1000 simulated networks that share the same characteristics as the observed one (i.e., direction, density, number of nodes, and number of edges).

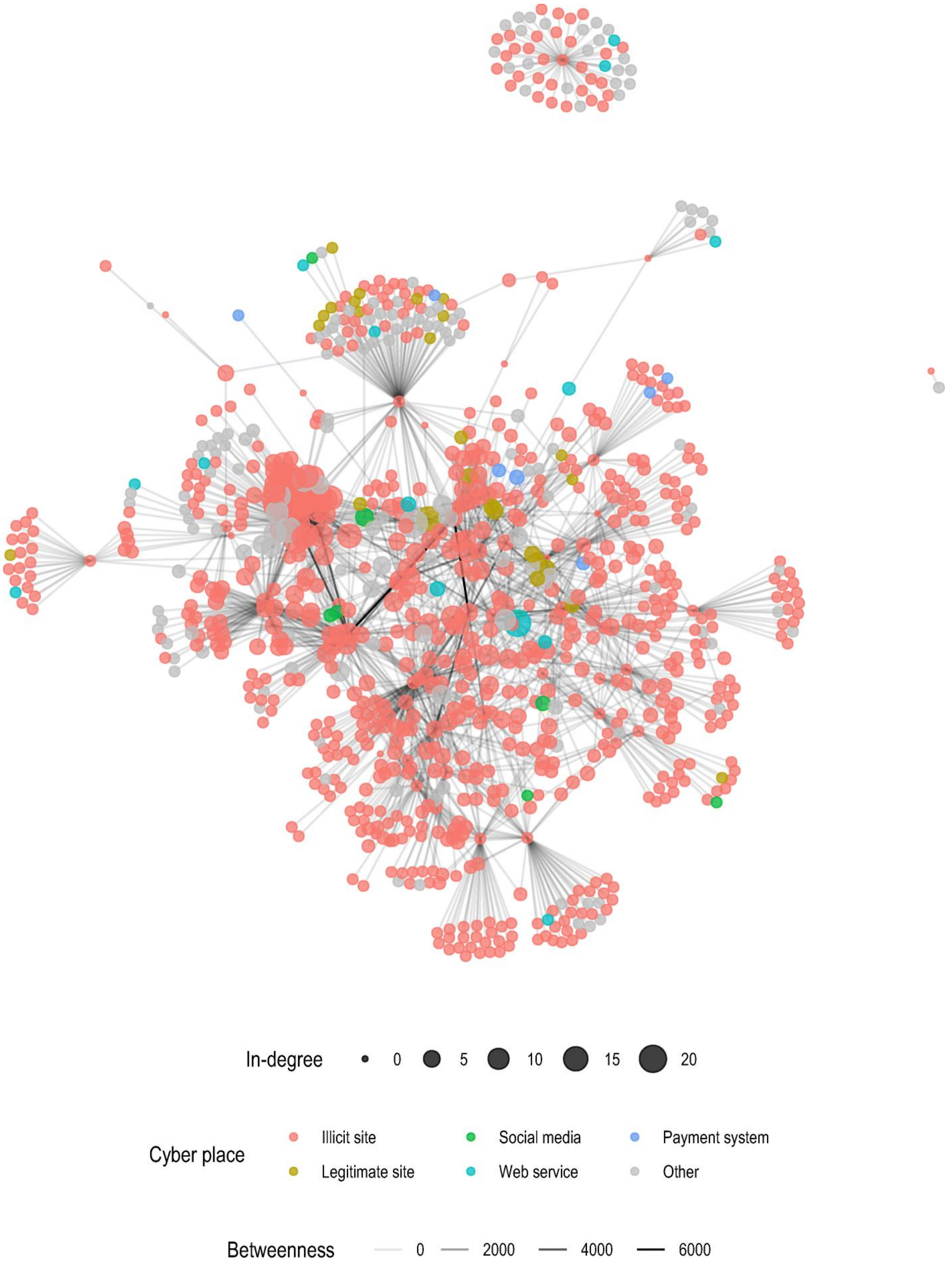

Figure 8. Network of FMIWs. All figures illustrating this manuscript have been created using the ggplot2 R package (Wickham, 2016), and the ggraph R package (Pedersen, 2020).

The observed network presents a reciprocity of 0.09 (9.1%), a diameter of 11, and a mean distance of 4. Compared to simulated networks, these results show that the observed reciprocity is notably larger than expected, whereas the diameter and mean distance are about half of the expected values (Figure 9).

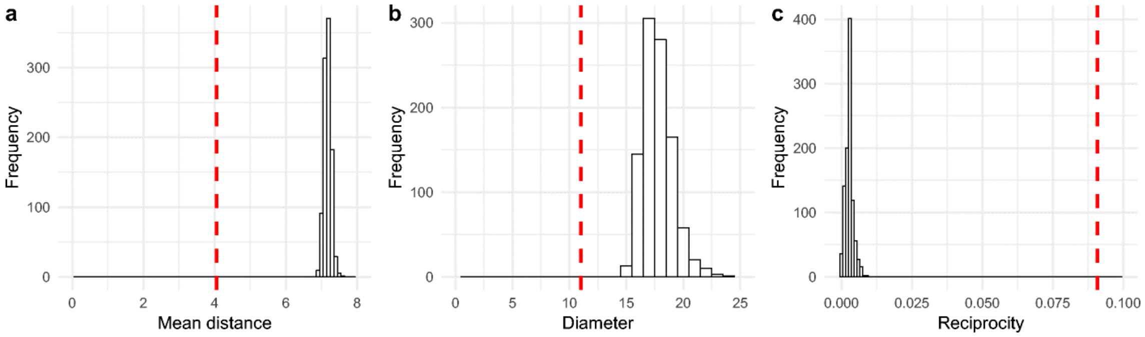

Figure 9. Comparison between metrics of the network observed and 1000 simulated. The dashed red line indicates the values obtained for the FMIW network

At the vector level, we calculated two centrality measures to identify the most salient nodes in terms of accessibility within the network, including in-degree and edge betweenness. In-degree measures the number of adjacent nodes terminating at them, an indicator of the ease with which a given website can be accessed from another. The distribution of the in-degree score by network nodes appears in Figure 10. The average in-degree score of the observed network is 2.49, indicating that most nodes receive few hyperlinks. When a node has a high in-degree score, it is referred to as a receiver node (Wasserman & Faust, 1994). Such nodes are represented with a larger size in the network, as depicted in Figure 8. Edge betweenness measures the number of shorter paths that pass through an edge connecting the key nodes or bridges that are critical for the connectivity of a network (Wasserman & Faust, 1994). In Figure 8, bridges are represented by more opaque lines connecting their nodes; only a few nodes are connected by edges with high betweenness.

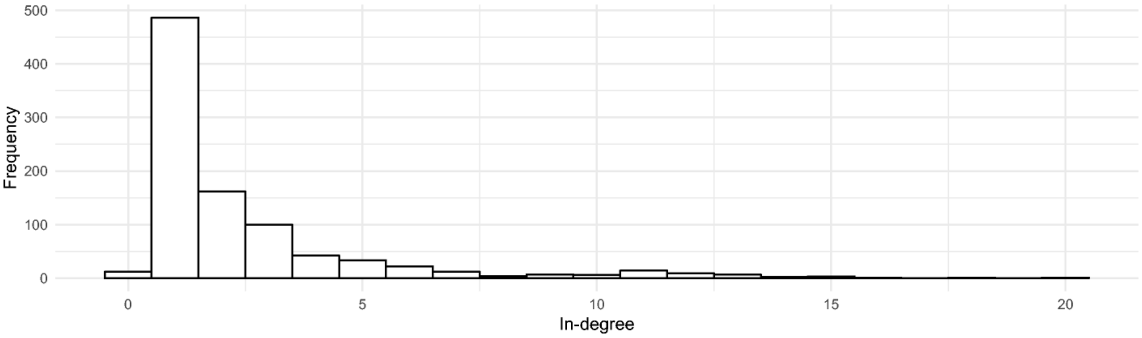

Figure 10. Network in-degree score distribution (min = 0; Mdn = 1; M = 2.49; SD = 2.73; max = 20)

7.6 Discussion

This study applied the concepts of Environmental Criminology to cyber places and, in particular, to FMIWs. These websites, which claim to sell tips on fixed sporting events, were mostly available on the clear web and may be crime attractors because they offer particularly attractive criminal opportunities. We hypothesized that these websites offered distinctive environmental features to incentivize deviant behaviours (buying fixed matches results) compared to regulated sport-betting websites. The results appear to corroborate our hypothesis. In general, FMIWs abound of situational precipitators. They promote triggers by posting advertisements on other betting/fixed match sites; however, they do not provide reminders to discourage pathological gambling, nor they do control prompts displaying contact information or license number (even fakes ones). Additionally, they encourage anonymity without enabling registration and login.

The contrasts are stark compared to regulated sport-betting operators' websites, which usually must comply with established guidelines and regulations. Compliance with regulatory standards facilitates a certain homogeneity in terms of reducing situational precipitators. For example, all regulated sport-betting operators discourage the participation of minors, give advice on abusive gaming, and require registration/login to play. Still, including a banner with an external link to another website has a clear, intentional nature and is not a common practice on regulated betting websites. This component of purposiveness has been evidenced by the existing literature on hyperlinked websites (H. W. Park & Thelwall, 2006). Eventually, FMIWs present distinctive layout signs, which facilitate the precipitation to deviant conducts (in our case to buy illegal tips). Our study does not discuss how much this approach is successful in terms of business, but argues that all FMIWs share a pattern with similar environmental features that characterize them as cyber places which are “located” in the sport-betting cyber environment (Miró-Llinares & Johnson, 2018).

In the present study, we also hypothesized that vending places for fixed matches have a specific network compared to a random network distribution. The results of the study corroborate this second hypothesis as well. The network structure of FMIWs reveals more about the nature of these cyber places. First, the network is highly homogeneous. The majority of nodes (77.4%) are fixed-match sites. Previous studies on hyperlinked networks in political contexts show that they generally form homogeneous communities (Ackland & Shorish, 2009; Burris et al., 2000). The same trend appears in the match-fixing network, which is comprised of 77.4% illicit betting cyber places. However, this trend has not been observed in online child sexual exploitation communities (Westlake & Bouchard, 2016). Compared to traditional criminal networks, studied by Malm and colleagues (Malm et al., 2010), the density of the observed network —0.003— resembles that of a kinship or formal organization networks —0.004 for both— rather than co-offending or legitimate associates. The former networks are characterized as being more cohesive and, thus, not easily disrupted. However, the figures for average and maximum in-degree centrality in the observed network —2.49 and 20, respectively— appear most similar to those of co-offending networks described by Malm et al. (2010). Therefore, it appears that fixed matches website network cannot be included in any of the four categories established by Malm et al., which makes it more reasonable to compare their characteristics with hyperlinked networks instead of social networks.

Hyperlinked networks, such as Tor, show values similar to the observed network—0.002 (Monk et al., 2018). Conversely, other hyperlinked networks, such as child sexual exploitation websites, show a much higher density— an average of 0.45 for sites and 0.34 for blogs (Westlake & Bouchard, 2016); white supremacist communities have a density of 0.11 (Burris et al., 2000). Despite showing a high level of reciprocity with regard to simulated networks, as well as the Tor network (4.9%) (Monk et al., 2018), the reciprocity of the match-fixing network is small (9.1%) when compared to the online child sexual exploitation websites network (23%) (Westlake & Bouchard, 2016). Regarding network connectivity, the average distance of the observed network, 4, is also lower when compared to Tor, which is 4.95 (Monk et al., 2018), meaning that it takes about one less connection on average to move from one node to another. Compared to child sexual exploitation, the analysed network of FMIWs has different characteristics. Specifically, the distance between its nodes is shorter, its connectivity is lower, it lacks communitarian places like forums, and it sells allegedly fixed-match results. Therefore, the network structure should be more similar to network marketplaces.

The characteristics of the network show similarities to those of the Tor network, a network that hosts these marketplaces and favours the anonymity of its users. Still, it is debatable whether FMIWs are an actual illicit market instead of a scam business model. Indeed, there are several points that support the scam hypothesis. Structurally, the alleged fixed matches market does not provide any warranty (such as escrow schemes) to protect buyers from frauds. An escrow system would reinforce trust toward vendors, and it could expand the market. Economically, the business model of selling tips on fixed matches looks weak. Once the information on fixed matches is sold, and many punters bet on the fixed match, the odds will be lowered by betting algorithms. Additionally, it is questionable why a group with insider information would sell tips on fixed matches instead of only using this information internally. The internal use would minimize the risk that fixed matches would be highlighted as suspicious, which may trigger investigations from sports federations or law enforcement. Finally, using the information to place bets on fixed matches, either directly or through a group, may generate significant rewards that the dissemination of confidential information would hamper.

7.7 Conclusion

This study focused on FMIWs and their networks. Through the concepts of cyber places and Wortley's situational precipitators framework, we corroborated the hypothesis that online match-fixing services share a typical pattern in layout design, and that they form a specific cyberenvironment: a niche market where users trade fixed-match information. Our descriptive analysis showed that FMIWs starkly differ from other regulated sport-betting websites and that they are conceived to limit environmental inhibitors and to facilitate deviant behaviours, pushing potential punters to buy fixed-match tips. The HNA provided further insight into this structure. FMIWs form a quite homogeneous environment, they have a lower density, and higher reciprocity compared to similar random networks, but lower compared to other online illicit communities (e.g., child pornography, white supremacist). From a practical point of view, in terms of prevention, it would be interesting to apply the concept of 'secured by design' to cyber places adopting some situational crime prevention measures to avoid crime victimization (Davey et al., 2017). Further, in terms of investigation and repression, law enforcement could use the HNA to highlights those websites that have a higher betweenness centrality. Targeting bridges should be particularly useful in reducing the robustness of the network (Malm et al., 2010; Malm & Bichler, 2011). Targeting nodes with bridging ties could facilitate policing efforts to disrupt networks (McGloin, 2005).

Nevertheless, our research has limitations. We conducted the initial sampling by using keywords, so a new search that includes additional or different words may reveal new fixed-match networks not analysed in this study. Since the search used language with Western letters, our results do not automatically extend to FMIWs in other languages that use different typing characters, whether they exist (e.g., the most spoken languages in the Asian market where sports betting is very important). In analysing the network, it was sometimes difficult to classify web pages within the proposed categories of cyber places.

Further research on different stages may also be useful. Such research should explore the applicability of secured by design principles to cyber places and, through HNA, corroborate whether and why cybercrime places have similar or different network structures, as well as explore the network survivability. Regarding FMIWs, further contributions could compare the business model used by different fixed-match vendors (e.g., prices, warranty, payment methods, types of bets sold), their prediction accuracy and the types of sport matches allegedly fixed and, eventually, establishing contact with vendors to understand how they manage the relationship with customers and the extent to which such vendors are fraudulent. Research could also explore FMIW administrator motivations to determine whether they seek personal profit as a way of promoting illegal betting. Finally, it would be interesting to understand how the punters perceive these sites and how they use them.

CHAPTER VIII

EXPLORING SITUATIONAL CONTEXTS IN SOCIAL MEDIA TO PREVENT ONLINE HARASSMENT

This chapter has been published as: Moneva, A., Miró-Llinares, F., & Hart, T. C. (2020). Hunter or Prey? Exploring the Situational Profiles that Define Repeated Online Harassment Victims and Offenders. *Deviant Behavior*. https://doi.org/10.1080/01639625.2020.1746135

8.1 Introduction

Online harassment among young people is often described differently based on the origin, frequency, and nature of the behaviour. In general, cyberstalking is understood to be a form of continuous online harassment, but may be characterized as cyberbullying when the aggressor is known to the victim (e.g. a classmate) (Miró-Llinares, 2012). Studying these behaviours can be challenging because myriad definitions for similar behaviours have been established within the empirical literature (Wolak et al., 2007). This lack of consensus in defining online harassment can also make measuring the phenomenon a tricky endeavour (Patchin & Hinduja, 2015). For these reasons, it is not surprising that a recent systematic review of online harassment studies found that prevalence rates varied considerably, between 1% and 41% for perpetration and between 3% and 72% for victimization (Selkie et al., 2016).

Researchers have also investigated similarities and differences between traditional or offline harassment and similar behaviours that occurs online (Beran & Li, 2008). These studies often hypothesize that a substantial proportion of online harassment behaviours originate from a previous interpersonal relationship. Researchers also acknowledge that although offline and online behaviours may be related, they also have unique defining characteristics that distinguish them from one another. For example, Henson (2010) describes three main differences between online and offline harassment: (1) the physical proximity between offender and victim (i.e. place); (2) the time of commission of the offence, and (3) the effective prevention measures for each modality. In terms of place, while offline harassment may occur at the workplace or on the street, online harassment occurs in cyber places, including in chat rooms and on social media (Ybarra & Mitchell, 2008). With respect to time, offline harassment requires direct convergence between offenders and victims, but online settings allow communication to be streamed or asynchronous. Additionally, a number of successful strategies aimed at preventing offline harassment (see, for example, Ttofi & Farrington, 2011) may incorporate new measures (e.g. parental monitoring) that can also be effective against online harassment (Khurana et al., 2015). Therefore, to be effective online, preventive measures must be implemented according to the convergent environments defined by the factors described above.

Drawing on the original Routine Activities Approach (L. E. Cohen & Felson, 1979) and inspired by its adaptation to cyberspace (Holt & Bossler, 2008), we demonstrate an alternative method to analysing the place and time dimensions of online harassment among young people. Our aim is to identify the situational patterns in offending and victimization that can inform the creation and implementation of crime prevention measures at the micro level. To accomplish this goal, several online convergence settings (i.e. social media) in which young people spend their time and interact with each other are examined. As a result, the paper makes an innovative contribution to the existing literature in two meaningful ways: first, it contributes to criminological theory by incorporating the concept of cyber place (Miró-Llinares & Johnson, 2018) for the development of studies on routine activities and cybercrime; and second, it adds to applied crime prevention research by exploring the relationship between crime and place using configural thinking and conjunctive data analysis techniques (Miethe et al., 2008).

The next section presents the theoretical framework used in the current study, which aims to help explain the relationship between the cyber places where online harassment manifests and the routine activities that users undertake within them. The theoretical framework serves to contextualize three research questions. Then the methodology used in the present study, the measures used, and the analytical strategy based on the Conjunctive Analysis of Case Configurations (CACC) (Miethe et al., 2008) to answer our questions are presented. Results are structured and presented sequentially, according to the current research questions. Finally, a discussion of the results in relation to criminological theory and the prevention of cybercrime, as well as the implications for policy making, is presented. This section is followed by some concluding comments.

8.2 Routine activities and victimization in cyber places

The Routine Activities Approach (L. E. Cohen & Felson, 1979) is a theoretical framework used in the analysis of contextual opportunities that produce crime events; it has been one of the most frequently empirically tested theories for various forms of cybervictimization (Holt & Bossler, 2016). To help explain victimization processes further, criminologists have also relied on Lifestyle Theory, a theory of criminality that explains the propensity of certain individuals to become victims according to their lifestyle (Hindelang et al., 1978). Some scholars suggest that both theoretical frameworks possess important synergies; and as a result, offer a third integrating construct of both: The Lifestyle-Routine Activities Theory (Holt & Bossler, 2008; Reyns et al., 2011). However, merging these two theories can be confusing because the Lifestyle Theory is a theory of criminality that focuses on individuals, while the Routine Activities Approach focuses on events (Hirschi & Gottfredson, 1986). To address specific crime problems in cyberspace, the Routine Activities Approach has generally been applied to explain the spatiotemporal convergence of motivated offenders and suitable victims when a capable guardian is absent (L. E. Cohen & Felson, 1979).

Debate over the applicability of Routine Activities Approach in cyberspace research has both supporters (e.g. Grabosky, 2001; Pease, 2003) and detractors (e.g. Yar, 2005). This debate was purely theoretical until scholars put the Routine Activities Approach model into practice by operationalizing its essential elements in cyberspace (e.g. Choi, 2008; Holt & Bossler, 2008; Hutchings & Hayes, 2008). Usually, victims were measured with self-reported victimization and their suitability with online exposure measures. Guardians and their absence were measured through personal guardianship (e.g. parent monitoring) and technical guardianship (e.g. antivirus software) variables. However, as with more traditional routine activity studies, the motivated offender has been largely ignored and rarely measured with self-reported offending. Since the Routine Activities Approach was first measured for cybercrime analysis, a growing body of empirical evidence consistently indicates that the approach has contributed to a better understanding of the dynamics of different forms of cybercrime (for a review, see Leukfeldt & Yar, 2016).

While some have not found complete support for the application of the Routine Activities Approach to cyberspace, as it relates various forms of economic cybercrime (Leukfeldt, 2014), others have obtained promising results (Bossler & Holt, 2009; Petrescu et al., 2018). Furthermore, in his study on identity theft, Reyns (2013) found that this framework had explanatory potential beyond the criminality that required physical convergence. These contradictory results could be explained by the fact that there is not a standardized model for applying the Routine Activities Approach to cyberspace, since neither the models used in most studies are not similar, nor are the ways in which the variables included in them are measured. Regarding the various forms of social cybercrime, existing scholarship shows greater consistency between studies using the Routine Activities Approach as an explanatory framework (Marcum, Ricketts, et al., 2010; Reyns et al., 2011; Wolfe et al., 2016). Collectively, these studies show how the application of the Routine Activities Approach to cyberspace has been more successful in explaining cyber-enabled crimes in which the convergence between people in digital spaces is evident and strongly conditioned by everyday offline activities.

In addition to risk factors related to the everyday activities undertaken by victims, findings from other studies suggest that the Routine Activities Approach is an appropriate framework for studying cybervictimization. For example, studies show that people who have admitted to committing a cyber offence, or who have associated with peers who have done so, are more likely to experience a subsequent cybervictimization (e.g. Holt & Bossler, 2008; Ngo & Paternoster, 2011; Reyns et al., 2011). As with certain criminal dynamics in physical space, these findings suggest that some cybercrimes are also likely to generate homogeneous pools of offenders and victims.

Thus, there appears to be elements other than those related to the suitability of potential victims that also affect the likelihood of participating in a cybercriminal dynamic.

Existing scholarship also suggests factors that influence the likelihood of cybervictimization are related to individual and environmental characteristics that define digital spaces where people converge and interact (Leukfeldt & Yar, 2016; Miró-Llinares, 2015b; Miró-Llinares et al., 2018). As in the physical space, these digital places or cyber places have certain characteristics that (1) affect the way people contact each other, (2) define the forms of surveillance and their scope, and (3) condition the different activities carried out in them (Miró-Llinares & Johnson, 2018). Because online harassment requires a specific form of convergence to occur, these elements may configure cyber places in such a way that victimization and offending is more/less likely to occur. For example, prolonged use of chat rooms by teenagers increases their chances of becoming victims of online harassment (Marcum, Higgins, et al., 2010; Ybarra & Mitchell, 2008). Similarly, users who have many social media accounts and add strangers as friends are more likely to be harassed (Henson et al., 2011).

Social media are cyber places mostly transited by teenagers and young adults. When social media users interact, there is an exchange of information that can include both live streaming, and store-and-forward interactions – when information is stored but sent/received later (Miró-Llinares & Johnson, 2018). In addition, social media contain digital microenvironments where natural surveillance and surveillance capacity can vary across platforms as the timelines where users publish their posts are usually public environments, while the spaces for personal messaging are usually private (Miró-Llinares et al., 2018). And while some social media allow thousands of users to interact at the same time, others limit their capacity to a few hundred. The use of social media (e.g. leisure, work), defines the type of activities that users perform in them and, consequently, shapes crime opportunities. Thus, certain activities, such as the publication of opinions, habits of daily life, or personal information, also appear to be related to an increased risk of victimization (Choi & Lee, 2017). Similarly, excessive use of the social media Facebook increases the likelihood of online harassment (Näsi et al., 2017). On the contrary, these same authors found that receiving greater social control, defined by the number of friends in each account, does not have a protective effect against online harassment.

In summary, existing research shows that the application of the Routine Activities Approach as an explanatory framework for studying cybercrime has produced a large and growing body of empirical knowledge, with three key aspects emerging. First, despite highlighting the value of convergence between offenders and targets, this theoretical framework has been applied mainly from a victimological perspective, focusing on variables that constitute both risk and protective factors that influence cybervictimization dynamics. This necessitates more cybercrime research that focuses on offenders (Bottoms, 2012; Cullen & Kulig, 2018; Miró-Llinares & Moneva, 2019a). Secondly, and in line with Vakhitova, Reynald, and Townsley's (2016) interpretation of the studies on cyber abuse and routine activities, these risk factors have been more or less correctly related to one of the three minimum elements for the occurrence of the crime, a combination known as the Chemistry of Crime (Felson & Eckert, 2019): motivated offender, suitable target, and the absence of a capable guardian, but who have separated themselves from the other essential elements to avoid the occurrence of the event that gathers the triangle of the crime (Cullen et al., 2002): the place, the manager, and the handler. In this sense, some have discussed the use of place-based approaches and have contributed to developing a theoretical environmental framework for analysing crime events in cyber places (Miró-Llinares & Johnson, 2018; see also Reyns, 2010).

Thirdly, previous studies show that researchers consider a wide range of digital environments relevant for the study of the criminal opportunity outside the cybercrime object of study, but that their analysis has not been carried out from the prism of the event, emphasizing the context in which cybercrime occurs, but in the individual actors who participate in it (Miró-Llinares & Moneva, 2019a).

8.3 The Present Study

By analysing the influence of the cyber place where online harassment may occur, the present study pursues three objectives: (1) to determine whether online harassment repeat victimization and offending among students is context-dependent, using conjunctive analysis of case configurations; (2) to determine which dominant situational contexts define self-reported online harassment repeat victimization and offending among students; and (3) to determine whether repeat online harassment is defined by a homogeneous pool of victims and offenders, by testing whether distributions of dominant case configurations associated with each group are statistically similar.

8.3.1 Sample

A probabilistic sampling method stratifying by sex, age and area of residence (i.e., rural or urban) in Castile-Leon (Spain) was carried out to select the respondents for this study. Castile-Leon is an Autonomous Community consisting of nine provinces, most of them low density populated. Once the number of participants was calculated for each stratum, the classrooms containing the right number of students were accordingly selected for the survey to be administered. Our sample of Spanish non-university education students (*N* = 4174) was comprised of 1999 males (47.9%) and 2175 females (52.1%), ranging from 12 years to 21 years of age (*M* = 15.44; *SD* = 1.87). All subjects included in the sample use at least one social media on a daily basis and spend at least 1 hour online every day. Relative to the non-university educated population in Spain, our sample was very similar in terms of sex and age according to National Institute of Statistics (INE) official figures (INE, 2018).

8.3.2 Instrument

To collect our sample, an *ad hoc* online survey was administered in local schools, supervised at the time by classroom teachers, which helped ensure students understood survey questions and assist students with questions about the survey when they arose. Given the sensitive content of the survey, its design was elaborated in a joint effort of methodologists, criminologists, and jurists, and then adapted to a language that could be understood by school-aged children. The instrument was comprised of four groups of questions: (1) sociodemographic questions that queried students about their sex and age, (2) questions related to students "routine activities" in cyberspace, which were designed to measure social media use and school-children's habits, (3) questions designed to measure self-reported online harassment victimization, (4) and questions designed to measure self-reported online harassment offending behaviours.

8.3.3 Dependent variable: Online harassment

Existing empirical scholarship fails to provide a consensus definition for online harassment (i.e. cyberharassment). Instead, there is considerable debate on the use and operationalization of this behaviour, with some suggesting it is synonymous to cyberbullying and cyberstalking, which has led to confusion among researchers (Patchin & Hinduja, 2015; Wolak et al., 2007). For the current study, we use a behaviourally-defined definition of online harassment: experiencing repeated, unwanted, harassing behaviour that would likely cause a reasonable person to become fearful or worried (Finn, 2004; Wall, 2001a).

To define online harassment, we refer to five self-reported behaviours related to repeated, unwanted, harassing online contact: (1) insulting and humiliating, (2) spreading rumours, (3) marginalizing, (4) threatening, and (5) pretending to be someone else. Each of these measures is dichotomous. Participants who claimed to commit or suffer at least one of these repeat behaviours were labelled as online harassment "repeat victims" and "repeat offenders". Elements of intent and harm were integrated in the design of each question to identify online harassment offenders. In these questions, we measured intent by asking students whether their online behaviour was "intended" to "cause harm". Following Wolak and colleagues (2007), the questions referred to incidents occurring during the last year. The questions were formulated as follows: "In the last year, have you repeatedly [self-reported behaviour] someone online?" —for measuring repeat offending—; and "In the last year, has anyone repeatedly [self-reported behaviour] you online?" —for measuring repeat victimization—.

8.3.4 Independent variables

A total of 10 predictors of online harassment victimization and offending were used in the analysis that follows. Three of the 10 correspond to individual-level characteristics, whereas seven are related to cyber places where adolescents spend their time online.

8.3.4.1 *Individual factors*

Developmental and life-course criminology literature has found a relationship between sex and specific age intervals and criminal propensity for offending and victimization (Farrington et al., 1990; Moffitt et al., 2001). In addition, it has been found that young adults are those who are most likely to spend most of their time online (Hargittai & Hinnant, 2008) and are also among the age group most likely to be victimized or offend (Cops & Pleysier, 2014). To examine this relationship, three age intervals have been defined as (1) 12 – 14 years, (2) 15 – 17 years, and (3) 18 – 21 years. Note that the legal age of majority in Spain is 18 years old, so these age intervals were set on the recommendation of the Department of Education of the Governing Council of Castile-Leon, accounting for the possible policy-making implications of the findings. The age intervals of the underage participants were further divided into two groups based on a similar recommendation, given their different degree of maturity [67]. Although lower secondary schooling is often completed by the age of 16 in Spain, some of the participants were either repeating grades or studying professional training courses in the same school. Students' sex was also recorded and coded "0" for females and "1" for males.

Previous research also suggests that spending more time online increases the likelihood of exposure to deviant behaviours (Bossler & Holt, 2009; Hinduja & Patchin, 2008). For this reason, and under the category of routine activities, a measure designed to gauge the amount of time students reportedly spent online each day was included in the analysis through the following question "How many hours a day do you spend surfing the Internet?" and possible answers "Less than 1 hour", "From 1 to 3 hours", "From 4 to 7 hours", "From 8 to 15 hours", and "More than 15 hours". For participants it may be difficult to determine exactly how much time they spend on the Internet and, in addition, only 0.8% of participants reported spending less than 1 hour per day on the Internet and none more than 15 hours, so responses were recoded into three categories: (1) less than 4 hours, (2) 4 – 7 hours, and (3) more than 7 hours.

8.3.4.2 *Cyber place-related factors*

Victims play an important role when it comes to determining their own online harassment victimization risk by incorporating certain assets to digital spaces (Miró-

[67] Personal communication.

Llinares, 2015b). Considering adolescents spend much of their time interacting with each other and building online relationships using social media (Subrahmanyam et al., 2008), another set of factors were included in the analysis that follows to help understand the role that these cyber places play in online harassment behaviours. These variables measure (1) whether students used various social media every day through the following question “Which of the following social media do you use daily? (You can choose more than one option)” and possible answers “I do not use social media”, “Snapchat”, “Instagram”, “Facebook”, “Twitter”, and “Another, which one?”; (2) whether students uploaded their name and photos to their social-network profiles through the following question “What kind of personal data do you publish in social media? (You can choose more than one option)” and possible answers “I do not publish any personal data”, “First name and/or surname”, “Personal photos”, and “Another, which one?”; and (3) whether they restrict other users’ access to them through the following question “Do you restrict access to your social media (only your contacts can see your information)?” and possible answers “Yes”, and “No”. Including each of the multiple response options, these variables were coded as dichotomized, where 0 indicates “No” and 1 indicates “Yes”.

The national studies conducted by van Wilsem (2011, 2013b) revealed that online harassment victimization was related to interacting through social media. A matrix question regarding which social media were used daily included a list of seven possible answers: Instagram, Twitter, Snapchat, Facebook, Periscope, Ask.FM, and the option Other as an open answer. According to their popularity among students, the top four social-network sites, in terms of their usage, were then selected and included in the dataset. The others were not included in our analysis.

Table 11 contains all measures used in the analysis that follows, presented by self-reported online harassment victim/offender status.

Table 11.
Descriptive statistics for self-reported online harassment repeat victims and offenders

			Online harassment status			
	Total (*N* = 4174)		Repeat victim (*N* = 1401)		Repeat offender (*N* = 514)	
Variable	*n*	%	*n*	%	*n*	%
Individual factors						
Age						
12-14	1561	37.4	447	31.9	152	29.6
15-17	2148	51.5	753	53.7	304	59.1
18-21	465	11.1	201	14.3	58	11.3
Sex						
Female	2175	52.1	795	56.7	201	39.1
Male	1999	47.9	606	43.3	313	60.9
Time online						
< 4 hours	2105	50.4	585	41.8	213	41.4
4 - 7 hours	1924	46.1	736	52.5	263	51.2
> 7 hours	145	3.5	80	5.7	38	7.4
Cyber place factors						
Reportedly uses						
Snapchat	839	20.1	335	23.9	130	25.3
Instagram	3635	87.1	1287	91.9	472	91.8
Facebook	805	19.3	317	22.6	109	21.2
Twitter	1108	26.5	427	30.5	160	31.1
Profiles contain						
Name	1305	31.3	607	43.3	246	47.9
Photo	595	14.3	319	22.8	133	25.9
Profile access						
Restricted	3348	80.2	1098	78.4	370	72

8.3.5 Analytical strategy: Conjunctive Analysis of Case Configurations

To analyse the situational profiles of online harassment among both offenders and victims, we used Miethe and colleagues' (2008) CACC approach. CACC is a case-oriented analysis technique that can be applied to categorical data. As an alternative to traditional, variable-oriented approaches to data analysis, CACC enables researchers to identify the complex causal recipes of variable attributes that give rise to a particular outcome (i.e. the dependent variable).

Specific details for conducting CACC are available in the extant literature (Hart, 2014; Hart et al., 2017; Hart & Moneva, 2018; Miethe et al., 2008), but can be summarized with a few basic steps. First, a "truth table" is constructed from variables contained in an existing data file. The table's columns reflect each predictor variable included in the analysis, the outcome variable, a column associated with the number of times a case configuration is observed in the existing data file, and one that represents the probability a configuration results in the outcome of interest. Each row in the truth table reflects a unique combination of predictor variable attributes that could be observed in the existing data file (i.e. case configurations). Once the truth table is constructed, all the data from the existing file are aggregated to each case configuration and are prepared for data analysis by applying decision rules for defining dominant case configurations [68]. For the current study, dominant case configurations are defined as 10 or more observed configurations. Finally, analysis of a CACC truth table involves identifying and quantifying patterns of situational clustering (Hart, 2019) and describing patterns of contextual variability [69]. This approach can uncover patterns in one's data that main-effect models commonly used in traditional analysis (e.g. logistic regression) may not be capable of identifying (Hart, 2014; Hart et al., 2017; Miethe et al., 2008).

For the current study, we created two CACC truth tables (i.e. one for victimization and one for offending), following the steps described previously. In doing so, will were able to link the specific situational profiles of online harassment victims with identical profiles of online harassment offenders. As described previously, 10 predictor variables were analysed in the current investigation. The "age" and "time

[68] See Miethe et al. (2008), Hart (2014), and Hart, Miethe, and Rennison (2017) for a discussion on the decision rules for defining dominant profiles.

[69] A chi-square goodness-of-fit test is used to determine whether data from an existing data file cluster among dominant case configurations more than expected and Hart's (2019) Situational Clustering Index (SCI) is used to measure the magnitude of clustering if it is detected. The SCI is a standardized metric, similar to the Gini coefficient.

spent online” measures each were defined by three categories, whereas the other eight measures were dichotomized. This enabled us to compare and contrast the attributes that define the victim and offender group of students simultaneously, in ways that existing empirical scholarship has yet to do.

The next section presents results of our analysis of these variables using the CACC methodology, which answers our three research questions. CACC has been conducted with the CACC R package version 1.0.0 (Miriam Esteve et al., 2019) that incorporates tidyverse (Wickham et al., 2019) data transformation functions. Data visualization uses GGally R package version 1.4.0 (Schloerke et al., 2020). All code was written in R version 3.6.1 (R Core Team, 2019) using RStudio version 1.2.5019.

8.4 Findings

Our first research question is whether repeated online harassment victimization and offending among students is context dependent. The structure of our CACC matrixes could have produced over 2,300 case configurations (i.e. two variables with three attributes and eight dichotomous variables or 3^2 x 2^8 = 2,304). However, when aggregated to our truth tables, our survey data were defined by far fewer situational profiles. Specifically, our entire survey data were defined by a total of 643 repeat online harassment profiles or 27.9% of all observable profiles. This is despite the fact that our sample was large enough that nearly two students could have been associated with each of the theoretically observable configurations. These findings suggest that participants do not visit cyber places randomly. Instead, their behaviour —both as victims and offenders of repeated online harassment— cluster within specific situational contexts defined by the unique combinations of variable attributes examined in the current study.

In addition to our data clustering within a relatively small subset of theoretically observable profiles, our survey data clustered significantly among 94 dominant case configurations (X^2(93, N = 2817) = 3,378.22, p < .001), which were defined by 10 or more observations. Furthermore, based on the Situational Clustering Index (Hart, 2019), the magnitude of clustering among dominant profiles was moderate (*SCI* = 0.451). These findings provide strong evidence that our online harassment survey data is very context dependent.

Our second research question asks, “Which dominant situational contexts define self-reported online harassment victimization and offending among students?” Findings from our CACC indicate that the likelihood of online harassment repeat victimization varies considerably among dominant situational profiles. For example, 82% of female students, age 15 – 17 years, who spend between 4 – 7 hours per day online, who reportedly use Snapchat and Instagram, and share both their names and photos on these social media platforms, but who do not restrict other users’ access to their profiles reported experiencing repeat online harassment. In contrast, none of the male students, age 12 – 14 years, who spend 4 – 7 hours online each day, using Instagram, Facebook, and Twitter, but who do not share their names or photos on social media and who do not allow other users to access their social media profiles reported similar repeat victimization experiences. This 82 percentage-point difference in victimization risk illustrates the extreme contextual variability in online harassment repeat victimization, which is not easily identified using traditional, variable-orientated approaches to data analysis (i.e. HLM, OLS, etc.) because these analytic methods focus on identifying “main effects”, while holding covariates “constant” (Weisburd & Britt, 2014).

Table 12 shows the composite profiles associated with the five dominant case configurations *most* and *least* likely to be associated with repeat online harassment. These profiles illustrate the complex causal recipes that lead/do not lead to online repeat harassment, as many of the predictor variable attributes are associated with profiles found in both groups. For example, all the students who reportedly restrict access to their social media profiles to other users (i.e. Privacy = Yes) are among the *least* likely to report being repeatedly victimized. However, three of the five dominant profiles *most* likely associated with online harassment are also defined by students who allow other users to access their profiles. It is the application of the CACC method that enables us to disentangle the complex causal recipes that give rise to online harassment repeat victimization.

Table 12.
The five dominant case configurations most and least likely to result in online harassment repeat victimization, the probability of being victimized, and the number of students associated with each profile

Sex	Age	Hours	Snapchat	Instagram	Facebook	Twitter	Name	Photos	Privacy	P(V)	*N*
Dominant profiles *most* likely to result in online harassment repeat victimization											
Female	15 - 17	4 - 7	Yes	Yes	No	No	Yes	Yes	No	0.82	11
Female	12 - 14	< 4	No	Yes	No	No	Yes	Yes	Yes	0.70	10
Female	15 - 17	4 - 7	No	Yes	No	No	Yes	Yes	No	0.63	16
Female	15 - 17	4 - 7	No	Yes	No	Yes	Yes	No	No	0.60	10
Female	18 - 20	4 - 7	No	Yes	No	No	No	No	Yes	0.60	10
Dominant profiles *least* likely to result in online harassment repeat victimization											
Female	18 - 20	< 4	No	Yes	Yes	No	No	No	Yes	0.10	10
Male	12 - 14	< 4	No	No	No	No	No	No	Yes	0.09	89
Male	12 - 14	< 4	No	Yes	Yes	No	No	No	Yes	0.08	13
Male	15 - 17	< 4	No	No	No	No	No	No	Yes	0.07	42
Male	12 - 14	4 - 7	No	Yes	Yes	Yes	No	No	Yes	0.00	10
Mean =										0.33	30
SD =										0.15	33

Table 13 shows the composite profiles similar to those in Table 12. In Table 13, however, profiles are associated with the five dominant case configurations most and least likely to be associated with self-reported online harassment repeat offending. Results from a CACC presented in Table 13 show that 44% of females, age 12 – 14 years, spending 4 – 7 hours online each day, and who reportedly use Instagram, and who share their names and photos on social media, but who do not restrict access to their social media profiles are the most likely to report having engaged in online harassment behaviours. In contrast, several different combinations of variable attributes define students who never report harassing others online (i.e. P(O) = 0.00). As with the dominant profiles of repeat victimization, case configurations associated with online harassment repeat offending behaviour are characterized by variable attributes that fail to demonstrate linear main-effects on offending that are assumed by popular traditional analytic approaches.

Table 13.
The five dominant case configurations most and least likely to result in online harassment repeat offending, the probability of offending, and the number of students associated with each profile

Sex	Age	Hours	Snapchat	Instagram	Facebook	Twitter	Name	Photos	Privacy	P(O)	N
Dominant profiles *most* likely to result in online harassment repeat offending											
Female	15 - 17	4 - 7	No	Yes	No	No	Yes	Yes	No	0.44	16
Male	15 - 17	4 - 7	No	Yes	No	No	Yes	Yes	Yes	0.40	20
Male	15 - 17	< 4	No	Yes	No	Yes	Yes	No	Yes	0.39	18
Female	15 - 17	4 - 7	Yes	Yes	No	Yes	Yes	Yes	Yes	0.36	11
Male	15 - 17	4 - 7	No	Yes	No	No	Yes	No	Yes	0.32	31
Dominant profiles *least* likely to result in online harassment repeat offending											
Male	15 - 17	4 - 7	No	Yes	No	Yes	Yes	No	No	0.00	11
Female	12 - 14	4 - 7	Yes	Yes	No	Yes	No	No	Yes	0.00	10
Female	18 - 20	< 4	No	Yes	Yes	No	No	No	Yes	0.00	10
Female	18 - 20	4 - 7	No	Yes	No	No	No	No	Yes	0.00	10
Male	12 - 14	4 - 7	No	Yes	Yes	Yes	No	No	Yes	0.00	10
Mean =										0.12	30
SD =										0.10	33

Finally, our third research question investigates whether the pool of online harassment repeat victims and offenders are homogeneous. To answer this question, we compared the 94 dominant profiles that defined online harassment repeat victims to the 94 profiles that defined repeat offenders, based on the rank-orders of the likelihoods of being a victim/offender. Results of a Wilcoxon’s signed-ranks test revealed that the distributions of matched profiles were significantly different from each another ($W+$ = 22.00, z = 7.91, $p \leq .001$). In other words, offending probabilities are not proportional to victimization probabilities, suggesting that the situational contexts of those who repeatedly engage in online harassment are dissimilar to those who repeatedly experience online harassment.

These findings are illustrated in Figure 11 using parallel coordinates plot, where dominant profiles are presented in descending order along the y-axis according to their offending probabilities and a line drawn from each ordered position to the position along the opposite y-axis that corresponds to the same dominant victimization profile.

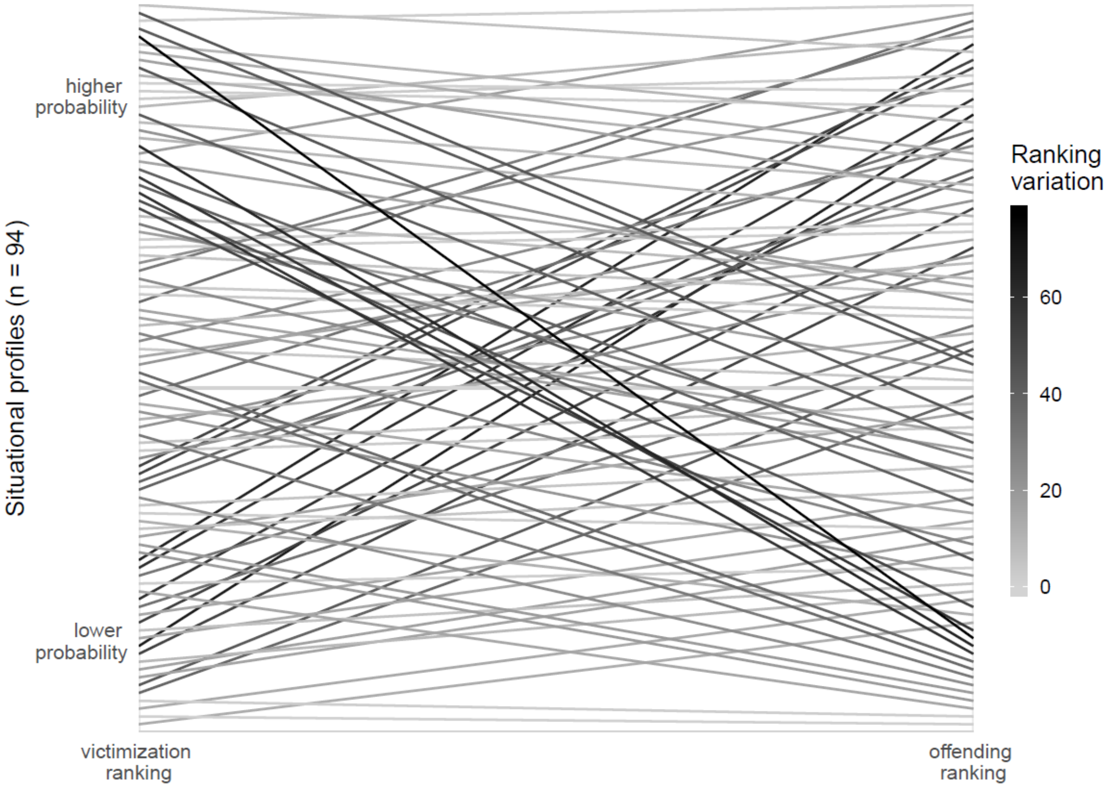

Figure 11. Linkages between dominant situational profile probabilities for repeat victimization and offending. Each line represents matched case configurations across both groups

8.5 Discussion

Although most cybervictimization studies show the explanatory potential of the Routine Activities Approach regarding different cybercrimes (Holt & Bossler, 2016; Leukfeldt & Yar, 2016), to date, they all typically use a variable-oriented approach (e.g. logistic regression) to generate new empirical knowledge. An alternative analytic strategy used in the current study (i.e. CACC) allowed us to explore the routine activities of social media users in relation to repeat online harassment from a new perspective. This new perspective informed us about online harassment repeat victims and offenders by examining the situational profiles or the unique causal recipes defined by all observed variable attributes in combination with one another simultaneously.

With regards to cybervictimization profiles, several points require further discussion. First, the situational profiles of users associated with a *lower* likelihood of victimization spend *less* time navigating through cyberspace daily. This conclusion is consistent with the framework of opportunities offered by the Routine Activities Approach background, since the less time spent online, the fewer opportunities there are for them to become objectives for harassers. Existing literature provides a consensus on this aspect (e.g. Bossler et al., 2012; Hinduja & Patchin, 2015; Reyns et al., 2011). Results show that the visibility of users is also related to victimization likelihood. In line with Reyns and his colleagues (2011), those that do not publicly share personal information, such as their real name or pictures, have lower risk of being repeatedly victimized within the context described. It should also be noted that the top five case configurations observed in data used for the current study were defined by profiles of female students, showing another pattern identified in previous studies (Marcum, Higgins, et al., 2010; Navarro & Jasinski, 2013). Specifically, current findings suggest that sex is a determining factor in online harassment outcomes, since other profiles that were similar —expect where the students were male— had a substantially lower probability of being victimized.

In addition to corroborating findings obtained by much of the existing research into online harassment victimization and offending, our study also produced new insights that are unique. For example, based on our configural analysis, the composition of the top profile associated with online harassment repeat victims, reflects certain types of “context-specific interaction effect” (Miethe et al., 2008, p. 235) because the probabilities of victimization vary greatly when compared to other top profiles (i.e. the outcome varies by 19% between the first and third profile). This could mean that interacting in more digital environments within that context significantly increase the probabilities of suffering online harassment repeatedly. It can also be observed that the two case configurations in which none of the social media measured is used daily by students are among the three situational profiles least likely to produce online harassment repeat victimization (0.10 and 0.07 respectively). The fact that these profiles still have a small probability of victimization associated with their configuration means that this behaviour occurred in different cyber places from others in the CACC matrix (e.g. in Flickr or Ask.fm —see “Independent Variables” section—). That the chances of being victimized are so low when none of the social media examined are present in the CACC matrix is convincing evidence that the selection of the social media included in our analysis is adequate.

Results from the current study also produced findings contrary to what can be found in the existing literature. For example, our CACC analysis shows that the probabilities of repeat offending are lower (*M*(O) = 0.12 versus *M*(V) = 0.33) and more homogeneous than those of repeat victimization (i.e. they vary less; *SD*(O) = 0.10 versus *SD*(V) = 0.15). The former indicates that criminal behaviour is infrequent and concentrated in fewer users, while the latter suggests this is an obsessive and therefore more stable behaviour (Pittaro, 2007). In fact, although not shown in the tables, one of the most representative case configurations comprises 4.3% of the total sample (*n* = 181), with a very low probability of repeated offending (P(O) = 0.03). Whereas traditional research on deviant behaviour among youth populations suggests that males engage in the majority of offending behaviour (Moffitt et al., 2001), our results show a mixed distribution in line with Novo and colleagues (2014). Nonetheless, the age interval for high risk repeat offenders' situational profiles is the same as their analogous, which seems logical considering that many of these criminogenic dynamics happen between peers within the context of conflicts generated at school (Beran & Li, 2008; Hinduja & Patchin, 2008).

Our analysis also show that the top five repeat offender profiles use their real name on the social media that they use frequently and three of them also upload their personal photos. However, from a rational choice perspective, offenders should be expected to describe higher levels of anonymity to reduce their risk of being identified. Similarly, one might assume that some users diversify their offensive opportunities among several social media accounts, but when examining their situational profiles this is not evident. Configurations with almost zero probabilities associated with offending are associated with students that spend less time online daily and who tend not to provide personal information. This could also indicate that users who make up such profiles are less familiar with the use of social media or have restricted access to them.

Like Holt and Bossler (2016) noted, most of the previous research on online harassment victimization has focused on victims, leaving aside both their relationship with offenders and the context in which this dynamic occurs. Some environments where online harassment occurs, such as social media, produce a two-way interaction that increases the opportunities of getting involved into personal conflict with other users, resulting in an offender-victim continuum. Our results show that each situational profile associated to repeat offenders matches a repeat victim profile, meaning that any context that determines an online harassing behaviour also meets the requirements to lead to a cybervictimization. In contrast, 13 of the 94 profiles resulted in victimization only (13.8%). These results underscore the importance of accounting for more situational elements than the victim, since most profiles show that there are not purely victimizing or purely offending environments, but rather mixed contexts that can lead to both situations. However, it should be noted that the probabilities of repeat victimization are higher than the probabilities of repeat offending.

In their literature review on routine activities, Holt and Bossler (2016, p. 70) state that "scholars have consistently found that committing cybercrime or cyber-deviance is one of the strongest risk factors for being harassed or stalked in the virtual world". While previous research has focused on the dynamics of cybervictimization from a broader perspective (Leukfeldt & Yar, 2016), CACC allows us to analyse this link at the profile level, showing that actors involved in offending do not necessarily share the same situational context as those who suffer cybervictimization. In Figure 11, greater differences in the range of links between columns would indicate fewer specific contexts between repeat offenders and victims, while less variance would suggest that there is a more homogeneous dynamic. This means that some of the case configurations analysed in this study are key to defining whether a social media user is more likely to offend repeatedly or become a repeat victim in cyberspace.

Findings from the current study also provide guidance for future research in the area of online harassment. Specifically, scholars undertaking research in the future should go beyond the traditional variable-oriented analysis based on the elements that constitute the Chemistry of Crime. As an alternative, we propose the use of conjunctive analysis techniques, as they allow to generate knowledge in terms of configuration (i.e. unique combinations of multiple variable attributes) (Hart, 2014; Hart et al., 2017; Hart & Moneva, 2018; Miethe et al., 2008). Since an essential component of this type of cyber-enabled crime is the previous relationships between offenders and victims (Beran & Li, 2008; Hinduja & Patchin, 2008), future research on online harassment should also address the connections between the occurrence of these dynamics in cyberspace and physical space. Furthermore, it would be interesting to transfer the study of the homogeneous populations of offenders and victims to a micro-level analysis that would enable us to determine the characteristics that relate both conditions.

In terms of policy implications, our results show which student situational profiles are most likely to repeatedly commit online harassment or suffer a repeat victimization. This information can be used by service providers, teachers, parents, and students themselves to raise awareness about propensity and vulnerability. However, it is important to note that our results showed different situational contexts of risk for repeat offenders and victims, therefore responses to this problem may have to be adapted differently for each of them. These findings stress the importance of responses be "situationally" dependent (i.e. different situations or contexts require different prevention strategies). In this sense, Situational Crime Prevention (SCP) measures are known for their versatility, simplicity and effectiveness, making them an adequate complement to the safety of young students. Based on SCP measures that have been specifically adapted for a similar behaviour (i.e. cyberstalking) such as those proposed by Reyns (2010), those profiles that have obtained a high associated probability of cybervictimization should receive training on self-protection measures while repeat offenders should be controlled by social media service providers (i.e. cyber place managers). These types of measures are also often quite efficient, so they can be implemented even when resources for prevention are scarce.

8.6 Conclusion

In this paper we presented a study on repeat online harassment from a novel situational approach that uses a conjunctive analysis technique (i.e. CACC) to explore the situational contexts where this dynamic occurs. Our work contributes to existing scholarship in two ways: (1) based on the Routine Activities Approach, we introduced the notion of cyber place as an essential element to analyse the convergence of offenders and victims in digital environments where online harassment is known to be found; and (2) we moved beyond victimization to explore through conjunctive analysis techniques the situational profiles of repeat offenders and their possible overlap with those of repeat victims.

In accordance with the specific objectives initially proposed in this paper, several conclusions can be drawn. First, concentration analyses show that the dynamics of repeat online harassment manifest themselves in very specific situational contexts, defined both by the routine activities undertaken by the participants and by the configuration of the cyber places they visit. Secondly, the CACC has allowed us to identify the composition of every situational profile defined by the participants. With this information it is possible to know which exact combination of factors influences a greater probability of being involved in an online harassment dynamic. Finally, this study reveals that the contexts in which a specific user is most likely to suffer repeat victimization are different from those in which another is more likely to offend repeatedly, which suggests that prevention and control strategies to tackle this problem require the adoption of different measures for each form of participation.

However, this research also has limitations. Although the CACC certainly allows patterns to be discovered in the data that other methods cannot, the inclusion of many variables in the matrix increases the variability of the number of the resulting profiles. This makes the interpretation of the results too complicated. For this reason, we excluded from the analysis any factors unrelated to cyberplaces, but equally important for understanding the dynamics of online harassment (e.g. self-control).

Therefore, future research should explore other factors identified in the literature as relevant to the study of online harassment. In addition, the “repeat” offending and victimization dimension should be further investigated to reduce the incidence of this phenomenon. It is also necessary to do more research on the implementation of specific preventive measures for online harassment such as SCP and to evaluate their effectiveness. We also encourage further approaching this problem by adopting the notion of cyber place and using conjunctive analyses.

CHAPTER IX

MODELLING CYBER MICROPLACES' METADATA TO DETECT ONLINE HATE SPEECH

This chapter has been published as: Miró-Llinares, F., Moneva, A., & Esteve, M. (2018). Hate is in the air! But where? Introducing an algorithm to detect hate speech in digital microenvironments. *Crime Science*, *7*(15), 1-12. https://doi.org/10.1186/s40163-018-0089-1

9.1 Introduction

Moments after Khuram Shazad Butt used a van to run down pedestrians along the London Bridge, Twitter was boiling. At 22:01 [70], before the first call for help was received, the hashtag #PrayForLondon was trending [71] on a global level; 2 min later, the first message including the hashtag #StopIslam was posted; and an hour later, 18 million tweets with the hashtag #LondonBridge had been published. In all of these digital messages, users expressed solidarity and indignation over the attack. Unfortunately, some digital content also contained messages of happiness, hatred towards certain groups, and the glorification of violence.

[70] Time in London.

[71] A topic is considered trending in Twitter when it is popular in a specific location at a given moment.

Academic interest inherent in the impact of hate speech on the Internet is not new (Tsesis, 2001). The possibilities of cyberspace to unify users and tear down some of the spatiotemporal barriers that limit the transmission of knowledge in physical space have augured an exponential increase both in the number of potential diffusers of such types of content and its receivers (Levin, 2002). Such quantitative growth, however, has taken place simultaneously with an even more relevant qualitative change. The democratisation of electronic communications and technologies (Brenner, 2017) and, in particular, the emergence of social networks as a brand-new social interrelation environment that has normalised communications through instant messaging systems has created a window of opportunity in which the expression of violent messages is no longer hidden or considered uncharacteristic of an ideological or political discussion.

We reconceptualize the role social networks play in the production of criminal events (e.g. hate speech) based on an adaptation of the principles of Criminology of Place to cyberspace (Miró-Llinares & Johnson, 2018). The present paper addresses the potentially massive dissemination of radicalized content via Twitter through the introduction of an algorithm for the automatic detection of contents that contribute to mitigate their impact. This research demonstrates how patterns of hate speech can be detected in metadata [72], basing the analysis on the relation between crime and place (Eck & Weisburd, 1995; Sherman et al., 1989). Cyberspace, however, is not contained in a single "place" with homogeneous characteristics, but events occur in different cyber places inside of it and at different times (Miró-Llinares & Johnson, 2018). The identification of these spatiotemporal patterns may help us to improve the algorithms based solely on content analysis. This method adds to quantitative efficiency by

[72] The information that defines single data items (e.g., the number of times a tweet has been retweeted, or the number of followers an account has).

automatizing part of the analytic process and thereby reducing the complexity of content analysis needed to identify messages of hate speech. Furthermore, it adds to qualitative efficiency by increasing the ability to limit the attention on content by private entities or public authorities to content that is actually related to high-risk activities, that is the dissemination of hatred or radical content in cyberspace.

In the following section, a review of recent literature is conducted to summarise the existing approaches to hate speech detection in cyberspace. Then, a comprehensive explanation of the concept of "cyber place" based on the idea of convergence is provided to present the theoretical framework in which the algorithm is built on. Afterwards, an empirical study is reported on to show the performance of the system proposed with a sample of tweets. The results are then interpreted and discussed in terms of efficiency and innovation to conclude with a summary of the relevant contributions and developments this work provides.

9.2 Related work

There has been a normalisation of extreme situations in an environment visited daily by millions of users to obtain the latest news and to socialise that is also used for propaganda purposes and the recruitment of radicalised subjects (Berger & Morgan, 2015). This situation has led European authorities who were already focused on social control (M. R. McGuire, 2017) to increase social media surveillance and specially to create and use digital tools that employ complex algorithms to detect propaganda and extremist and hate speech content (Awan & Blakemore, 2012) as well as to identify individuals in the process of radicalising (Edwards, 2017).

Such tools for the early detection of radical content are based on the identification of patterns, but in order to achieve this aim, they utilise a variety of techniques of content analysis, including the following: (1) manual collection

(Gerstenfeld et al., 2003), and sampling methods and crowdsourcing (Chatzakou et al., 2017; Magdy et al., 2015); (2) systematic keyword searches (Décary-Hétu & Morselli, 2011); (3) data mining for sentiment analysis (Cheong & Lee, 2011); (4) natural language processing (Nobata et al., 2016); and (5) different machine learning procedures (Ashcroft et al., 2015; Burnap & Williams, 2015; Malmasi & Zampieri, 2017; Sharma et al., 2018), including logistic regression models (Davidson et al., 2017), and neural networks (Djuric et al., 2015; dos Santos & Gatti, 2014). Although some of these tools employ metadata analysis in combination with semantic or syntactic methods (Schmidt & Wiegand, 2017; Waseem & Hovy, 2016), all of them focus their attention at the core of the analysis on the content of the message, meaning the words themselves or the relations among them, which implies a major drawback when analysing communicative environments as dynamic as social networks (Serrà et al., 2017). To overcome these difficulties when analysing online hate speech, in this paper we focus instead on analysing the metadata features extracted from Twitter digital microenvironments that are relevant for hate speech dissemination.

9.3 Traditional microenvironments, digital microenvironments, and hate speech

Twitter, like other social networks, is not a concrete physical location but can be accessed from many places, and criminal microenvironments are usually thought of as the locations, places, or spaces where crimes occur. Traditionally, the analysis of these micro places has served the purpose to understand how convergence allowed for a criminal event to take place. Social networks are not places in the traditional geographic sense, but they are places in a relational sense, since they are environments “that are visited” in which people converge with other people and with content in different ways, depending on the characteristics of the particular digital environment or network. The combination of the people (i.e., accounts), who say things (i.e., tweets) to other people (i.e., other accounts), define unique digital microenvironments in cyberspace. Indeed, it is in this sense of “place” where some cybercrimes occur in certain digital places more often than in others (Miró-Llinares & Johnson, 2018), which implies that the basic premises of Environmental Criminology in general, and crime patterns in particular, may be true for certain cybercrimes.

In particular, this approach refers to the idea that crime distribution is not random but is based on patterns determined by the different environmental elements of the places where victims and offenders converge and by the relevance of such places to the routine activities developed in the activity spaces (P. L. Brantingham & Brantingham, 1981). This is similarly valid for hate speech and for similar behaviours such as the dissemination of terrorist propaganda and radicalisation messages. It is true that in these types of crimes, the relevant convergence is not occurring between offender and victim but between the sender and receiver of the message. However, the convergence remains necessary: it needs a place where the hate message is reflected, and where another (or others, as the quantity of receivers is irrelevant) perceives it, such that hate speech or radicalisation on the internet will occur in some places more frequently than in others at both macro and micro levels, given certain environmental parameters.

From a macro perspective, that is, in comparison with other “places” or social networks, Twitter is an environment of massive, interactive and immediate communication of content. Although it allows streaming communication (through Periscope) and direct messages to concrete users out of sight of the rest of network, Twitter works essentially as a public square in which stored and forward communication is used to express content that can be observed and shared by a large number of people (Marwick & Boyd, 2011). If we add that political or ideological communication has become increasingly frequent on Twitter (Bode & Dalrymple, 2016), it seems understandable that this social network is commonly used to disseminate hate speech (Schmidt & Wiegand, 2017) and that it has become perhaps the favourite social network of extremist and terrorist groups for propaganda and the promotion of radicalisation to a wider audience (Berger & Morgan, 2015; Villeux-Lepage, 2014; Weimann, 2014).

In addition, Twitter's structural configuration, in particular the restriction on the length of messages (first 140 characters, now 280), limits the possibilities for interaction among users and makes both hate speech, which will not be the same as the content expressed in a different forum or on Facebook (Awan, 2016), and the activities of radicals and terrorists based on such speech less focused on recruitment and more aimed at normalising and magnifying terrorist activity for soft sympathisers (Villeux-Lepage, 2014) as well as disseminating propaganda by redirecting users to other places in cyberspace (Weimann, 2014). Furthermore, Twitter allows anonymity, although it is not the most common way of interacting (see Peddinti et al., 2014). Finally, despite its constant technical modifications, Twitter has not shown much efficiency with regard to withdrawing offensive, hate-related or radical content (Weimann, 2014), either because of the technical ease involved in creating accounts and the immediate publication of tweets or because of its rather vague free speech policy, which makes requests for removal different in each country (Hsia, 2017).

However, Twitter is not a homogeneous place where everything occurs in the same way everywhere inside it. It is well known, for example, that the temporal distribution of messages does not occur randomly (Miró-Llinares & Rodriguez-Sala, 2016); that there are some profiles with more followers than others and that not all of them publish the same number of tweets (Lara-Cabrera et al., 2017); and that there are very different degrees of identity expression on this social network (Peddinti et al., 2014). This indicates that a microanalysis of the configural elements of digital microplaces may be helpful to detect the environmental patterns that determine the occurrence of an event. In addition, it seems similarly obvious that the micro units that are essential for such an analysis are accounts and tweets.

A tweet is the essential microplace because it is where a message is expressed and shown and is where other users can interact with it, while an account is the microplace from which the publication or the viewing of such messages is made available. Like every microplace, a Twitter account has certain characteristics that differentiate it from the rest. For instance, if an account's registration information coincides with the identity of a public personality, Twitter will verify the user account with a blue badge. At the same time, a user can include a brief personal biography in one's profile and even activate an option to geolocate tweets in such a way that when publishing a message, the geographic location of where the tweet was written can be attached. Furthermore, users can include other accounts in thematic groups called "lists", which are useful for seeing only those messages published by selected accounts in chronological order. The number of lists in which an account is included is reflected in its profile together with other parameters such as the number of tweets published, the number of tweets liked, and the number of followers as well as the number of users that the account follows.

Similarly, a variety of elements configure and define a message transmitted by tweet. Tweets have a structural limitation in relation to the extension of their content that permits only a maximum number of characters, whether alphanumeric or in the shape of small icons, known as emojis. The combination of these characters with a variety of other elements will define the content of the microplace and its scope. Such elements include mentions, which act as specific personal notification when they include the @ symbol before the name of the user; Uniform Resource Locators (URL), which allow the inclusion of a hyperlink to additional content, whether an image, a video, a GIF or a link to an external site; or hashtags, which are situational elements that serve to thematically tag the content of a tweet to connect messages and create communicative trends. Indeed, the result of combining all these elements conditions the ways and the frequency with which people interact with a tweet just by seeing it or by interacting with the message and promoting its dissemination through a retweet, which is a feature that allows the dissemination of messages to the followers of an account.

In any case, the relevance of the microplaces where more or less hatred can be found lies in the premise that motivates the present work: that hate speech, similar to other crimes in physical spaces and in cyberspace (Miró-Llinares & Johnson, 2018), will also be distributed in certain patterns conditioned by the characteristics of the digital microenvironments where they occur. Thus, with regard to the special nature of hate speech in the sense of its dissemination via Twitter and taking into consideration the different structural characteristics of the microplaces that integrate it, there exists an opportunity to detect environmental patterns related to hate speech that could help to detect its early appearance in order to prevent, control or mitigate its impact.

9.4 The present study

The present study introduces and evaluates a new algorithm, designed to detect hate speech, through the identification of patterns found in the situational metadata of digital messages. Existing research has discovered various types of patterns on Twitter: linguistic and temporal (Williams & Burnap, 2016), sociodemographic and temporal (Marcum et al., 2011), spatiotemporal and socioeconomic (Li et al., 2013) and sociodemographic (Sloan et al., 2015), among others. In addition, patterns have been found related to the metadata on other social networks: for example, those linked to certain content for the detection of cyberbullying on Instagram (Hosseinmardi et al., 2015), or the tagging of YouTube videos to identify deviant content (Agarwal et al., 2017). What has not yet been analysed, however, is whether such patterns are related to the environmental characteristics of the social media accounts and digital messages in relation to their configuration as microplaces.

To achieve the study's aim, we required a large sample of digital messages from Twitter, upon which data mining techniques could be applied. This would enable us to determine whether characteristics of this social network's microplaces are decisive with regard to determining the types of messages that will be published from or inside them. With the aim of finding a more efficient tweet classification criterion, two classification trees were implemented: one with account metadata as inputs and another with the tweet microplace's metadata. A detailed description of the sampling strategy, variables analysed, and analytic technique follows.

9.4.1 Sample and procedure

The data collection was performed through the Application Programming Interface (API) of Twitter, which allows users with developer permissions access to data for reading, writing or monitoring in real-time. Researchers that work with data from Twitter are already familiar with the constant changes experienced by their API, which may compromise the process of data gathering. To address this problem and to overcome the possible changes caused by the application, an algorithm for data gathering was developed (see Appendix L) that is equipped with sufficient rigidity due to an exception management system: programming techniques that enable researchers to control the appearance of anomalies during the execution of a script. Additionally, a system was implemented that provides immediate alerts if the server experiences any problems, the connection is interrupted, or the API loses or receives new permissions. Through this system, it is possible to quickly resolve any adjustment problems regarding the requests sent to the server via the code and the responses from the API when new updates modifying the composition of the dataset occur.

Once the API access is obtained and after establishing convenient authentication parameters, information about a concrete event can be collected for subsequent analysis by using certain keywords or hashtags as search criteria. In this case, the terrorist attack perpetrated on London Bridge on 3 June 2017 has been selected. Once the data collection process has begun, the API can store up to 1% of the tweets published on Twitter based on pre-set search criteria. Thus, three filtering hashtags were selected to provide balanced sampling (see Miró-Llinares, 2016): #LondonBridge, which refers neutrally to the event; #PrayForLondon, for solidarity content; and #StopIslam, which is a representative hashtag for radical expressions, Islamophobia in this case. The first two hashtags were trending topics at some point during the event, while the last one was also a trending topic during previous attacks, allowing us to make comparisons with other samples collected earlier. Through this procedure, over 3 days, a sample of more than 200,000 tweets was obtained ($N = 200{,}880$) that refer directly or indirectly to the selected event.

9.4.2 Independent variables: microplace characteristics

In addition to the content of the tweets, the semi-structured dataset [in JavaScript Object Notation (JSON) format] contains numerous fields that provide information on different elements of Twitter, including the microplaces of accounts and tweets. Once the dataset was pre-processed and high-value dispersion variables were eliminated together with record identifiers as well as those variables with a percentage of nulls higher than 25–

30% (Hernández et al., 2004), the dataset was built. To build the dataset on which the classification tree was applied, there has been selected, on one hand, those variables that are related to the anonymity and the visibility of accounts and, on the other hand, to the structure and interaction of the tweets. These variables and others that were created from the aforementioned, together with each observation (i.e. tweet), comprise the dataset analysed in the present study.

The users' account has been identified as a microplace intimately related to their anonymity and the visibility of their actions, hence relevant for hate speech dissemination. Table 14 provides a detailed description of the variables related to the anonymity and visibility of the accounts that were used in the present study. Those variables that provide information about the person behind the profile, such as their name, interests, or area of residence were included within the anonymity category. A second set of variables measuring the visibility of the users' activity in Twitter such as message posting, the user's active period on the social network, and different forms of interaction with other users were included within the visibility category. Regarding the characteristics of an account, the variable "description" has been modified because the API returned the entire text field of users' biographies, and since the analysis of its content would have implied a subjective interpretation, a dichotomisation was applied (1, the user has a biography; 0, the user does not have a biography) to enable the classification tree to operate with these data.

Table 14.
Account variables related to users' anonymity and visibility

Variable	Type	Description
Anonymity		
Verified	Boolean	When true, indicates that the user has a verified account
Description [a]	Boolean	When true, indicates that the user has included a biography in his or her account profile
Geoenabled	Boolean	When true, indicates that the user has enabled the possibility of geotagging their tweets
Visibility		
Day_count	Numeric	The number of days since the user account was created
Listed_count	Numeric	The number of public lists in which this user is a member
Statuses_count	Numeric	The number of Tweets (including retweets) issued by the user
Followers_count	Numeric	The number of followers the account currently has
Friends_count	Numeric	The number of users the account is following. Also known as followings
Favourites_count	Numeric	The number of tweets the user has liked in the account's lifetime

[a] New variable

Tweets themselves and their associated metadata have also been identified as potential predictors of hate speech dissemination. Some of these elements are related to the interaction a tweet generates, while others determine its structure. Within the interaction category, some interactive elements that favour the users' engagement in dissemination activities were included together with the timing of the tweet publication. The structure category comprises two variables that constrain the length of the text and consequently the content of the message. The group of variables from the microplace of a tweet is shown in **Error! Not a valid bookmark self-reference.**. Regarding these elements, a few modifications have been made (see Appendix K). Because the restriction on the number of characters when publishing a tweet is one of the most distinctive characteristics of Twitter that has an obvious communicative impact, we measured the length of the text in the messages in the sample. To this effect, short scripts were elaborated to identify both the codification of the emojis on Twitter and the character chains composing URL to subsequently extract them from the body of a message. Thus, it is possible to carry out a character count to determine the actual length of a message, and two new variables are used to measure the presence of emojis and URL. With a similar method, we were able to determine the number of mentions and hashtags in each message, and we codified the results using two more numerical variables.

Table 15.
Tweet variables related to the interaction and the structure of messages

Variable	Type	Description
Interaction		
Mention_count [a]	Numeric	Number of mentions included in the text of the tweet
Hashtag_count [a]	Numeric	Number of hashtags included in the text of the tweet
Url [a]	Boolean	When true, indicates that the tweet includes a URL
Retweet_count	Numeric	Number of times this tweet has been retweeted
Minute_count	Numeric	Number of minutes since the event happened and the tweet was issued
Structure		
Text_count [a]	Numeric	Number of characters in the message, excluding URL, emoji, and retweet structure characters (i.e., 'RT @username')
Emoji [a]	Boolean	Indicates whether the text of the tweet includes an emoji

[a] New variable

9.4.3 Dependent variable: hate speech

With regard to the dependent variable, a tailored reading and the subsequent dichotomisation were carried out to determine whether the content of each tweet was neutral or hate speech. This method was chosen over semantic or syntactic approaches (e.g., Bag of Words) because these have shown weaknesses when dealing with specific messages such as humour or irony (Farías et al., 2016; Reyes et al., 2013). Plenty of investigations have addressed the problem of hate speech detection in social networks with such methodologies (e.g., Burnap & Williams, 2015, in Twitter; Mariconti et al., 2018, in YouTube). Although there exists a profound dogmatic discussion in that regard, in the present study, a broad concept of hate speech was used to classify such messages that comprises all the expressions considered violent or hateful communication in the taxonomy elaborated by Miró-Llinares (2016). According to this classification, for a tweet to be considered hate speech, its content must include the following categories: (1) direct incitement/threat of violence, (2) glorification of physical violence, (3) an attack on honour and human dignity, (4) incitement to discrimination/hate and (5) an offense to the collective sensitivity. This classification task was therefore based on the subjective interpretation of a text, with the limitations derived from this method. To alleviate the effect of judges' subjective analysis of the messages ($n = 100$), the Kappa coefficient (J. Cohen, 1960), which measures the degree of agreement, was applied to ensure accordance in the assessments and thus the reliability of the classification of the tweets. As can be observed in Table 16, and according to the criteria established by Landis and Koch (1977, p. 165), "almost perfect" agreement was obtained among the three pairs of judges (0.81–0.89).

Table 16.
Results of the applications of the Kappa coefficient to the three pairs of judges

Group	Value of κ
Judges A and B	0.81
Judges A and C	0.89
Judges B and C	0.88

Although previous studies that used the same classification methodology removed all retweets from the sample to filter original messages from their redundant replicas (Esteve, Miró-Llinares, & Rabasa, 2018; Miró-Llinares, 2016; Miró-Llinares & Rodriguez-Sala, 2016), this procedure was not adequate in this study because the data collection method through the API did not guarantee that all retweets fit the original tweets that bounced back. Thus, only duplicated tweets were removed, which left 35,433 remaining unique cases to be classified. After the judges classified these messages, duplicates were folded back into the dataset to calculate the hate speech prevalence in our sample: a total of 9488 (4.7%) out of 200,880 tweets.

9.4.4 Analytical strategy

Regarding the characteristics of the sample, to confirm the relevance of places in cyberspace, it is necessary to apply data mining techniques. Therefore, by making use of the Random Forests classifier technique (Breiman, 2001), an algorithm was implemented to create a number of classifiers for tweets that divide the sample based on the filters generated by each of the variables included in the model (i.e., nodes). These classifiers grow from a randomized data set extracted from the main sample to train the model and fit its parameters. 70% of the sample comprises the training set and the remaining 30% constitutes the test set. This division was repeated 10 times to promote randomization. The training set was then balanced favouring the minority class (i.e., hate speech tweets), while the remaining data were included within the unbalanced test set (Table 17).

Table 17.
Training set and test set composition

Class	Training set	Test set
Neutral	6638	184,754
Hate speech	6638	2850
Total	13,276	187,604

This training and testing process allow to control for anomalous or less consistent nodes and, hence, growing a non-overfitted, pruned tree. To define the most appropriate parameters for our algorithm, a series of computational experiments were carried out. These parameters were adjusted to reduce the forest’s sensitivity to their value (Tufféry, 2011).

When going through each node, the model asks each classifier whether the sample fulfils the condition established on it, thereby filtering the main sample and creating two subsamples: one that fulfils the condition and one that does not. The model then selects the best filtering among all trees and averages their individual estimations to produce the final output. By creating several decision trees that learn from a predetermined training set, the Random Forest produces robust predictions. When the condition that defines a node reaches maximum classifying efficiency, it means that the model has reached a leaf node, and it classifies the corresponding subsample to the same class: hate speech or neutral content. This technique intends to demonstrate that the cyber place variables selected can be used to properly classify a part of the sample, thereby contributing to the automation of the process. Additionally, to avoid results to be positively or negatively influenced by the training set composition, we used κ-fold cross validation defining κ = 5 subsamples (Kuhn and Johnson 2013).

An overview of the methodology employed in the present paper can be found in the figure below (Figure 12).

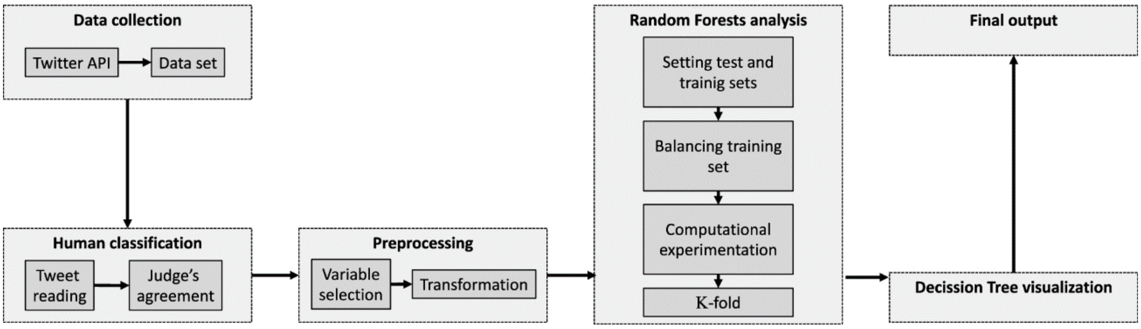

Figure 12. Overview of the methodology employed

9.5 Results

As can be observed in Table 18, two classification models were implemented and then validated for each set of cyber place variables to classify our sample: one used account variables as predictors while the other used tweet variables. Since the vast majority of accounts issued a single message (*Min* = 1.0; *Q1* = 1.0; *Mdn* = 1.0; *M* = 1.3; *Q3* = 1.0; *Max* = 126), their associated metadata can be treated differently and therefore the performance of the algorithm between the two models can be compared. Whereas account variables related to visibility and anonymity of users produce a rather poor model performance, the variables related to interaction and the structure of the tweets produce very promising results. Overall, the ability to avoid false positives (i.e., Precision) is consistently higher when including tweet variables in the algorithm. Regarding the accuracy of the model, results also support the use of tweet metadata over account metadata when it comes to the correct classification of positive cases (i.e., Recall). Mean scores resulting from fivefold validation are also included.

Table 18.
Algorithm maximum precision and validation scores according to account and tweet models

Model	Precision	Recall	F1-score	Fivefold
Account				
Neutral	0.99	0.65	0.79	
Hate speech	0.03	0.62	0.05	
Average/total	0.98	0.65	0.78	0.63
Tweet				
Neutral	1.00	0.87	0.93	
Hate speech	0.09	0.86	0.17	
Average/total	0.98	0.87	0.92	0.86

Parameters: *number of estimators* = 1000; *maximum depth* = 10

More detailed information about the number of correctly and incorrectly classified messages for both models can be found in the resulting confusion matrix (Table 19). Attending to the final purpose of the algorithm, effort was put into reducing the incorrect classification of hate speech messages (i.e., false negatives).

Table 19.
Confusion matrixes according to account and tweet models

Model	Real	Prediction	
		Neutral	Hate speech
Account	Neutral	120,511	64,243
	Hate speech	1078	1772
Tweet	Neutral	160,676	24,078
	Hate speech	397	2453

Regarding the cyber place related variables used to classify the messages, Table 20 shows their specific relevance within the models. The importance score reflects the proportion of nodes that include a condition imposed by each of the variables listed. In the case of account metadata, results show that visibility related variables are more important for the output decision, while anonymity has a negligible impact. On the other hand, two tweet variables influence the decision process over the rest: the number of retweets under the interaction category ($importance = 0.41$), and the length of the text associated to the structure of the message ($importance = 0.34$).

Table 20.
Importance of the variables included in both models

Variable	Importance
Account	
Anonymity	
Verified	0.00
Description	0.02
Geoenabled	0.05
Visibility	
Day_count	0.16
Listed_count	0.12
Statuses_count	0.17
Followers_count	0.14
Friends_count	0.16
Favourites_count	0.17
Tweet	
Interaction	
Mention_count	0.02
Hashtag_count	0.08
Url	0.05
Retweet_count	0.41
Minute_count	0.08
Structure	
Text_count	0.34
Emoji	0.02

To further understand which specific conditions a message must meet to be classified as neutral or hate speech by the algorithm, one of the decision trees produced with the Random Forests has been randomly selected and transformed into a flow chart (Figure 13). As can be observed, the metadata patterns described by hate speech messages are different from those depicted by neutral communication. This flowchart shows some contents that describe clear patterns and can be classified using only one to three variables: retweet count, text count, and minute count. Even if temporal stamps appear to have low influence in the decision process (Table 20), they are crucial to define the content of the messages.

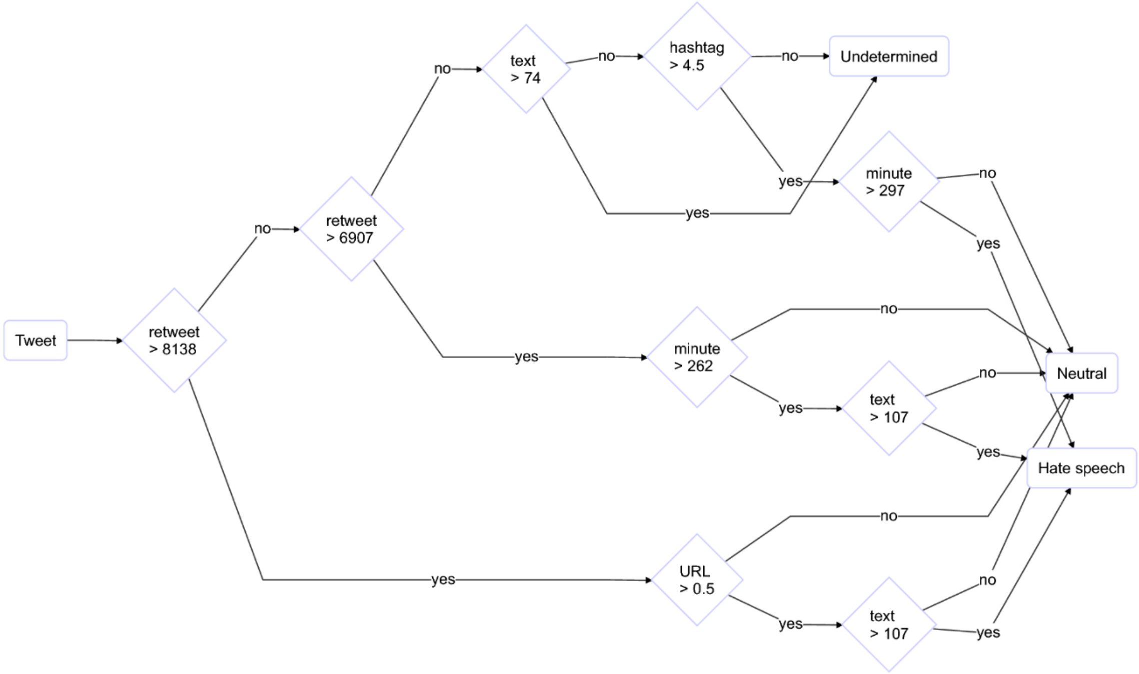

Figure 13. Flowchart for a Random Forest classification tree according to the variables of the tweet (depth = 5)

In summary, and as shown in the previous graph for the analysed sample, it is possible to define the environmental conditions that Twitter microplaces should have in order to differentiate the type of event occurring in them with certainty. These figures allow us to interpret the environmental patterns that arise from the sequential combination of account and tweet metadata associated to concrete messages. For example, if a message in our sample received between 6907 and 8138 retweets, was published 262 min after the attack, and had a text length of more than 107 characters (140 characters was the maximum allowed at the time of sampling), it was classified as a hate speech message; otherwise, it was classified as neutral (see Figure 13).

9.6 Discussion

Based on the results of the present study, we can deduce that (1) digital microenvironment metadata can be used to detect hate speech patterns in cyberspace similar to the way spatiotemporal crime patterns in the physical environment can be found, and that (2) hate speech messages on Twitter describe environmental patterns that are different from neutral messages. This result is derived from the fact that hate speech messages are communicated via tweets, or through accounts, with specific environmental characteristics reflected in concrete metadata associated with the message. In other words, tweets and accounts containing hate speech have different characteristics from tweets and accounts containing neutral messages, which is a logical consequence of the different ways of communication currently available and messages that are expressed differently by taking advantage of the different possibilities of the digital environment.

The performance of the models reported on in this paper demonstrate that not all account variables related to the anonymity and visibility of users are relevant criteria to distinguish whether or not the content of a tweet is hate speech. This is perhaps due to the ease in proving them fake as an identifier element, and therefore, they are not relevant for differentiating between messages. More specifically, anonymity related variables have proven to be almost irrelevant for classification purposes, probably conditioned by their dichotomous categorization as the information gain is biased towards variables with large number of values (Quinlan, 1986). Additionally, it does not seem entirely correct to make use of variables that describe a place where a crime will not occur just to determine the optimal environmental characteristics. As a matter of fact, the account is the microplace from which hate speech is published, but it is not where it manifests. In other words, in the present analysis, we are using the characteristics of houses to define the context of a crime that occurs on that street. For this reason, we argue that the results are far from expected. We also believe that account metadata are not useful for classifying tweets because such data are associated with a dichotomised result of a particular tweet, and in this way, we might be incorrectly attributing radical characteristics to a not-so-radical place, such as an account that might have published just one hateful message. It seems reasonable to conclude that the intention of a user who posts a single hate speech message cannot be considered the same as a radical user who systematically disseminates hatred.

Conversely, in line with the work of Ferrara et al. (2016), the most important element for classifying the contents of a tweet are the retweets it receives, as they are closely related to the interaction generated and the visibility of a message. According to theory, hate speech users seek a greater dissemination of their ideas and might therefore include certain elements such as URL and hashtags that have been found to make messages more appealing to retweeting (Suh et al., 2010). On the other hand, and in the same way that the architectural design of a physical space can condition the occurrence of criminal events in certain places [for a review of Crime Prevention Through Environmental Design (CPTED), see Cozens et al. (2005)], the present study shows that the architecture of a tweet, especially the length of its text, is an essential element to determine the nature of the message. In line with previous research, tweet time stamps have shown that hate speech messages also cluster in time (Miró-Llinares & Rodriguez-Sala, 2016), suggesting that certain cues activate radical responses on individuals more than others do. However, this analytical approach seems insufficient to explain why this is the case. In addition, the results confirm that tweet metadata have proved especially relevant to automatically identifying the specific microplaces where a criminal event will not occur (i.e., neutral tweets). There is no doubt these results are consistent in environmental terms, and we suggest that future investigations examine, for example, the role played by the anonymity variables of accounts in more detail, or the structural elements of a tweet regarding the dissemination of content.

Although the present study represents an initial stage of the investigation, it demonstrates the unquestionable capacity of the social sciences to provide important contributions to the fight against cyberterrorism (Maimon & Testa, 2017), and, since the main goal is to automate the process of classifying messages regardless of platform, it offers relevant information in terms of ways to potentially improve the search algorithms for different content, as it demonstrates that to detect this type of communication, we must focus not only on the content of a message but also on the environment in which it is expressed. In this sense, recent studies applying different lexical approaches for classifying tweets such as Support Vector Machines (SVM), Logistic Regression, or Random Forests, have obtained similar or inferior performances than the algorithm presented in this study, solely fed with metadata. Thus, while our Random Forest tweet model hits a F1-score of 0.92 [73], these previous attempts obtained F-measures of 0.77 (Burnap & Williams, 2015), 0.90 (Davidson et al., 2017), and 0.76 (Sharma et al., 2018) respectively.

We further argue that the use of metadata to classify messages can help to overcome limitations that arise from the application of approaches such as Bag of Words to samples comprising texts in different languages. In this sense, we believe that a combination of lexical and metadata approaches would enhance the ability of state-of-the art approaches to detect radical communication in social networks. From a methodological point of view, it can also be argued that metadata yield benefit both in

[73] Similar F1-scores were obtained in different samples that were not included in this paper but used the same methodology.

the extraction of variables, since they can be obtained through the API, and their simpler computation process compared to text-based variables.

It should be noted that the contribution of the present work is cross-cutting, as it goes beyond the frontiers of Twitter because all social networks host information of major importance in the metadata of their microplaces. However, this raises interesting questions regarding who has access to such metadata and whether the metadata should be made available to any user through open access systems or its access should be somehow limited. In any case, it seems that the current trend for many social networks is restrictive. Indeed, this has been the case for Facebook and Instagram, from which the extraction of information is becoming increasingly difficult. Until now, Twitter has continued to function with an open philosophy that allows researchers to collect a wide range of data.

9.7 Conclusion

Showing that Environmental Criminology can also be applied to cyberspace settings, this paper has introduced a brand-new theoretical framework to underpin online hate speech detection algorithms. Crime Pattern Theory principles and cyber place conceptualizations based on digital spaces of convergence (Miró-Llinares & Johnson, 2018) have been adapted to identify the most relevant characteristics associated to hate speech dissemination in Twitter. This important contribution provides an analytical background that opens the way to study different forms of cybercrime relying on cyber place metadata.

Two relevant cyber places for hate speech dissemination have been identified in Twitter: accounts and tweets. Drawing on the Random Forests technique, tweet metadata proved to be more efficient in the classification of hate speech content than account metadata. This suggests that not all variables should be taken into account when building predictive models, restricting models to those variables which are supported by valid theoretical schemes for solving particular problems. In this case, and given the nature of hate speech, it is crucial to consider the essential variables for content propagation in social networks for predictive modelling. And even if this is not a methodology comparison paper, the precision scores obtained show that this approach is, at least, on par with other methods based on semantic approaches.

Although studying the entire population of digital messages on any platform is an unrealistic task, a sample of over 200,000 tweets gives us the ability to answer our research question, despite our inability to generalise the current findings to all Twitter events. This further leads to the fundamental question of whether hate speech has been properly measured, that is, whether hate speech content has been properly distinguished from what is not. Regardless of the appropriateness of the taxonomy used to identify hate speech or whether the judges properly classified the sample, it is certain that the chosen method differentiates between events, which has been shown in the aforementioned studies.

As an axiological analysis, the sample may not accurately reflect the prevalence of hate speech on Twitter, but it is true that any pragmatic analysis will never lead two researchers to draw identical conclusions given the nature of language and the circumstances of communication. In this sense, this study aimed to achieve the greatest possible accuracy between judges to enable the analysis to interpret each criterion based on an acceptable level of agreement. Further research should be conducted to be able to escalate the application of the idea behind the methodology proposed in the present study.

Finally, despite demonstrating the utility of metadata in terms of precision for classification purposes, future research should aim to (1) compare computational times when using metadata versus text variables to determine which technique is more efficient, (2) test the ability of metadata models to overcome language limitations by comparing their performance in samples of different languages, and (3) merge the application of metadata and lexico-syntactical approaches to reduce the number of false negatives and positives, and to subsequently obtain even higher precisions with hate speech detection algorithms in cyberspace.

—Blank page—

CHAPTER X

GENERAL RESULTS AND DISCUSSION

This thesis sought to determine whether it was possible to apply the ECCA approach and the concept of cyber place to the analysis and prevention of four types of crime committed in cyber space (i.e. website defacement, match-fixing, online harassment, and online hate speech) (see CHAPTER I). In particular, special attention was devoted to the propositions of the ECCA approach related to the Crime Pattern Theory (P. L. Brantingham & Brantingham, 1993a) for identifying crime patterns in cyberspace, and to the developments of other place-based analytical frameworks to examine the role of cyber places in crime prevention (Miró-Llinares & Johnson, 2018). After developing the ECCA approach for its adaptation to cyberspace (see CHAPTER II), and establishing a replicable methodology based on crime analysis through Data Science (see CHAPTER IV), the four articles presented in this thesis applied this framework to four cybercrimes in order to better understand both the potential of cybercrime patterns and the role of cyber places in their prevention (see CHAPTER VI, CHAPTER VII, CHAPTER VIII, and CHAPTER IX). In this chapter, we present a general discussion of the collective findings.

To lay a strong theoretical foundation, we relied on the Environmental Criminology theories that underpin the ECCA approach (Bruinsma & Johnson, 2018; Wortley & Townsley, 2017a), namely: the Routine Activities Approach (L. E. Cohen & Felson, 1979), the Geometry of Crime (P. L. Brantingham & Brantingham, 1981), the

Rational Choice Perspective (Clarke & Cornish, 1985) along with the Situational Precipitators of Crime (Wortley, 2001), and the Crime Pattern Theory (P. L. Brantingham & Brantingham, 1993a). For their theoretical development, we drew on the work of many cybercrime scholars who have paved the ground ahead of us. Some of the key papers that inspired this work are: the discussions on the spatiotemporal configuration of cyberspace in relation to the application of the Routine Activities Approach (Grabosky, 2001; Miró-Llinares, 2011; Yar, 2005), the first operationalisations of this approach for its empirical testing (Bossler & Holt, 2009; Holt & Bossler, 2008), the crime analysis research on the first forms of financial cybercrime (G. R. Newman & Clarke, 2003), the initial discussions on cyber places and SCP (Reyns, 2010), their subsequent theoretical developments (Miró-Llinares & Johnson, 2018; Miró-Llinares & Moneva, 2019a), and —of course— later synthesis and compilation pieces (Bossler, 2020; Brewer et al., 2020; Holt & Bossler, 2016; Leukfeldt & Yar, 2016). All these works took fundamental steps to advance cybercrime scholarship regarding crime prevention. In fact, we learned a lot from crime theory, but we also looked at crime practice.

Since this thesis puts more weight on the empirical than on the theoretical —at least quantitatively—, we also needed referents in this field. So, we relied on the Criminology of Place. From the original studies on crime hot spots (Sherman et al., 1989), to the application of crime analysis to geographic micro units (Eck & Weisburd, 1995; Weisburd et al., 2016), to the consolidation of crime concentration as a criminological law (Weisburd, 2015), many contributions to the framework of the Criminology of Place have served as inspiration to guide the analytical strategy of this thesis. In line with this applied approach, previous work on repeat victimization also helped us to better understand how places and targets that are particularly vulnerable to crime are important in reducing crime (Farrell & Pease, 1993, 2017; Johnson, 2008a; Pease, 1998). Fortunately, both Environmental Criminology theories and the Criminology of Place share a strong applied philosophy, so finding synergies between the two was not a difficult task. By combining both situational frameworks, the four empirical studies presented here contribute to the discipline by following Miró-Llinares and Johnson (2018) in applying the concept of "place" to analyse and prevent four crime problems that occur in cyberspace. And unlike previous research, we do this empirically using a set of crime analysis techniques.

Thanks to crime analysis we were able to apply the aforementioned situational approach to extract knowledge from crime data and thus understand the nature of the four cybercrime problems in order to come up with practical solutions. Yet our crime analysis framework is also novel in the sense that it does not follow a classic statistical procedure typical of Criminology, but rather builds on Data Science. The main advantage of applying crime analysis through Data Science is that we could easily handle both large amounts and different forms of data to answer our research questions. Through a Data Science process (Grolemund & Wickham, 2016), firstly we used existing crime databases, deployed web crawlers —adopting the necessary cyber security measures—, conducted systematic online observation, administered online surveys, and used the Twitter API to collect data. Secondly, we used various parsing tools to import data. Thirdly, we processed the raw data following the philosophy of tidy data. Fourthly, we created several datasets that included networked data and a CACC matrix by transforming data. Fifthly, we used multiple graphs to visualise data. Sixthly, we implemented the random forests machine learning technique to model data. And lastly, we ensured that all this information was properly recorded to reach the right audience when it comes to communicate data. All the previous steps were carried out using free software consistent with our commitment to open science and replicability (Pridemore et al., 2018). In this sense, the material provided by the R community was essential (Grolemund & Wickham, 2016; R Core Team, 2019; Wickham et al., 2019).

Using the ECCA approach through the Data Science process, the general objective initially established was met as shown by several empirical findings that contribute to the advancement of the discipline. Overall, we found that it is indeed possible to apply the ECCA approach to cybercrime, although it needs to be adapted to the unique structural characteristics of cyberspace (i.e. the contraction of space and time). Our empirical studies show that the ECCA approach is useful for analysing different forms of cybercrime from a situational perspective, and that therefore it has an enormous preventive potential that remains unexplored. We merely showed the tip of the iceberg. However, it should be noted that, although there are many similarities, not all findings go in the same direction as other studies have shown for traditional crime. In this sense, we believe that further —but most importantly more rigorous— research is needed to conduct robust ECCA tests that demonstrate its validity. Only a selection of hypotheses derived from the ECCA approach were tested in this thesis and it is possible that the results were influenced by flawed data or insufficiently robust research designs. In particular, to address the specific objective of the thesis, the hypotheses test certain premises of crime concentration and repeated victimization —both in cyber places and among people—, and assumptions about safe and criminogenic online environments. Table 21 shows the results of the tests of the selected hypotheses.

Beyond the dichotomy between the acceptance or rejection of the hypotheses, the results of the four empirical studies provided many insights into the application of the ECCA approach to cyberspace. In the following sections, the implications of such findings for both criminological theory and preventive practice are discussed in the context of the transposed ECCA propositions. The discussion is structured around three important Environmental Criminology paradigms that are framed in a slightly different way for the occasion: Cybercrime Prevention Through Environmental Design, Patterns in Cybercrime, and the Criminology of Cyber Place.

Table 21.
Compilation of the hypotheses tested in the thesis

	Empirical support		
Selected hypothesis derived from the ECCA approach	Yes	Mixed	No
CHAPTER VI			
A substantial share of all defacements and variation in defacements is due to repeat victimization		X	
After a first defacement event, a repeat incident will occur shortly thereafter			X
Repeat defacements are disproportionately the work of prolific defacers	X		
A major reason for repeats is that offenders repeatedly target domains they have defaced previously			X
CHAPTER VII			
FMIWs offer specific crime opportunities because they incorporate distinctive environmental features that incentivize deviant behaviours when compared to regulated sport-betting websites	X		
Due to the peculiarity of this cyber environment, vending places for fixed matches have a specific network compared to a random network distribution	X		
CHAPTER VIII [a]			
Online harassment repeat victimization and offending among students is context-dependent	X		
Repeat online harassment is defined by a homogeneous pool of victims and offenders			X
CHAPTER IX [a]			
Hate speech patterns are related to the environmental characteristics reflected in the metadata of Twitter accounts and tweets	X		

[a] The hypotheses tested on these chapters have been reworded because they were originally raised as research questions.

10.1 Cybercrime Prevention Through Environmental Design

One of the transposed ECCA propositions calls for an understanding of what is the role of cyber places in cybercrime causation. If the immediate physical environment is known to affect criminal decisions (Clarke, 1980, 1992; Cornish & Clarke, 2003; Wortley, 2001), it is possible that a similar effect occurs in cyberspace. In both CHAPTER VII and CHAPTER IX we addressed this question and gained some insights. Initially we assumed that, just as in physical space there are places that are more criminogenic than others due to their environmental characteristics (P. L.

Brantingham & Brantingham, 1995), there would be cyber places that are more criminogenic than others for the same reason. We argue that, if the characteristics of these online environments were to be manipulated, it would be possible to reduce crime opportunities.

Crime opportunities can be measured by the convergence of people at specific times and by the presence of particularly attractive targets in terms of cost benefit for offenders. In addition, there are environmental elements that can influence crime opportunities such as surveillance systems (Welsh & Farrington, 2009)[74], street lightning (Painter & Tilley, 1999), or urban design (Cozens et al., 2005). In cyber places there are no physical elements like these, but there are digital structural elements that configure them. For example, many cyber places have sign in systems, display a license number to operate, and have menus that make navigation easier. In CHAPTER VII we argued that some of these elements may act as situational precipitators of crime in the absence of adequate control strategies. When examining the configurational elements in two comparable types of cyber places —FMIWs and regulated sport-betting sites—, we found that the former were configured to precipitate criminal behaviour, while the latter featured a number of strategies to control it. At the micro level, the configuration of cyberplaces can be reflected in their metadata. For example, the metadata of Twitter microenvironments (i.e. tweets) provide information about the time they were posted, the number of characters they contain, and how many times they have been retweeted by other users. In CHAPTER IX, we found that hate speech propagated in tweets with specific metadata, whereas neutral communication was contained in tweets with a different configuration. We argue that such elements could be manipulated to reduce the incidence of hate (e.g. limiting the number of mentions to other users or restricting the

[74] For the Spanish case, see also Cerezo-Domínguez and Díez Ripollés (2010).

use of certain hashtags). Both scenarios show the potential usefulness of implementing CPTED strategies to design secure cyber places.

The environmental design of cyber places is also related to the ease of access. If it is easier for people to travel between nearby geographical locations, it is also easier for users to navigate through linked websites. Saying that two websites are linked is the same as claiming that two places are connected by a street; going down that street is not mandatory, but it is a simple alternative. In our research we showed that FMIWs were linked to other FMIWs to a greater extent than to any other type of cyber place, forming some sort of illegal sports betting neighbourhood. As these cyber places did not integrate control strategies, crime opportunities were abundant. In addition, our analysis showed that the network structure of these FMIWs neighbourhoods was dissimilar to other online communities, such as those for political extremism or online child sexual exploitation (Ackland & Shorish, 2009; Burris et al., 2000; Westlake & Bouchard, 2016). This suggests that each type of cyber place may have a specific connectivity structure that defines their ease of access. In terms of crime prevention, analysing the structure of these criminogenic neighbourhoods is important, as a potential disruption strategy will differ according to their density (i.e. more or fewer nodes will need to be removed to meaningfully reduce the transitivity of the network) (Malm et al., 2010; Malm & Bichler, 2011). Applying CPTED to cyber places confers a great responsibility upon cyber place managers (e.g. website administrators, forum moderators) (Miró-Llinares & Johnson, 2018; Reyns, 2010). On many occasions, the managers themselves are responsible for the design of their website and, therefore, for the implementation of precipitation-control strategies as well as for the linkage with other cyber places. Proper management can make the difference between criminogenic and secure cyber places.

10.2 Patterns in Cybercrime

One of the main interests of the ECCA approach has always been crime patterns. Environmental criminologists and crime analysts have rejected the random distribution of crime in a myriad of studies, and in many others they have tried to determine what causes crime to form such patterns (P. J. Brantingham & Brantingham, 1984; P. L. Brantingham & Brantingham, 1995; Eck & Weisburd, 1995; Weisburd, 2015). Assuming that traditional crime concentrates in space and time —and, by the way, also on people— in CHAPTER VI, CHAPTER VIII, and CHAPTER IX we sought to determine whether and why different cybercrimes also described identifiable patterns. In our research we used different concentration analysis techniques for three cybercrimes (i.e. website defacement, online harassment, and online hate speech) and identified two main types of cybercrime patterns: those related to crime events and those related to the individuals who commit them. The former, in turn, can be spatial or temporal patterns, while the latter refer to how crimes concentrate per offender.

10.2.1 Temporal patterns

ECCA has often neglected time in favour of space, even though many research questions posed within this approach are about changes inherent in the passage of time. In this thesis, time patterns were examined both in dynamics of repeated victimization in a large sample of website defacements (CHAPTER VI), and in contexts of repeat online harassment offending and victimization in a sample of non-university Spanish students (CHAPTER VIII). In the first case, because research on repeat victimization in traditional property crimes indicates that it often occurs shortly after the original event (Pease, 1998; see also Bowers & Johnson, 2005; Farrell, 2005; Farrell & Pease, 1993; Johnson et al., 1997; Johnson & Bowers, 2004), we expected to find a similar time pattern for website defacements. In fact, temporal patterns were found in other cybercrimes of a similar nature, such as repeated network attacks on computer systems (Moitra & Konda, 2004) or packet transmission in DDoS attacks (Thapngam et al., 2011). However, the limitations of Zone-H data prevented us from testing this aspect conclusively, as the site administrators do not allow a repeated attack to be recorded until one year has elapsed since the initial event [75]. Yet we found distinct time patterns. Interestingly, just after the one-year restriction period, we observed a very skewed distribution of attacks indicating that some domains were experiencing new victimizations. This suggests that these cyber places were victimized extensively even though there were no records of such attacks in Zone-H. In such a case, the premise that applies to traditional patrimonial crime would remain valid for this type of cybercrime.

In the second case, we found rather weak time patterns of online harassment related to users' routine activities. According to the Routine Activities Approach, spending more time online would increase the likelihood of victimization, as the chances of converging with a likely offender in the absence of a capable guardian are greater (e.g. Bossler, 2020; Holt & Bossler, 2016). In fact, empirical research shows that this is the case (Bossler & Holt, 2009; Hinduja & Patchin, 2008). Aware of previous findings, we analysed the situational contexts of higher and lower risk of repeated victimization by online harassment for the students and found that, in general, the dominant situational profiles of students who spent less time online were less likely to suffer repeated victimization; however, there appeared to be no time differences in committing repeat online harassment. These results support those described in previous research, but call for caution, as the analyses were not specifically designed to identify accurate time patterns.

[75] This is a measure to prevent certain domains from being recurrently attacked as their vulnerability is publicly exposed. This information was obtained through personal communication.

10.2.2 Spatial patterns

Obviously, the spatial patterns we observed cannot be expressed in geographic terms as is customary, but they also reveal concentrations in “discrete nodes or areas of activity on the Internet where one is not physically located but can nevertheless act” (Miró-Llinares & Johnson, 2018, p. 893). To understand the spatial distribution of crime in cyberspace, we analysed three scenarios: In CHAPTER VI we examined the extent to which repeat victimization by website defacement were due to offenders who had defaced the same cyber places previously; in CHAPTER VIII we explored the situational contexts in which repeat online harassment occurred; and in CHAPTER IX we modelled the metadata of the micro places (i.e. tweets) where online hate speech spreads.

Regarding website defacements —and in comparison to burglary (Bernasco, 2008)—, we found that few offenders hacked the same cyber places more than once. However, it should be noted that for this circumstance to be reflected in the Zone-H data, the offenders had to target the same cyber place even a year after the first attack (i.e. the findings are most likely underrepresented). The two explanations offered by Environmental Criminology literature for such perseverance may apply (Chainey, 2012; Johnson, 2008a). The “boost” explanation would allude to an initial successful experience, which would make the hacker to persevere in the future. In this case, it is possible that an initial disfigurement helped the hacker gain status in the community, so by repeating the strategy, more credit can be earned (Holt, 2019). According to the “flag” explanation for repeat victimization, vulnerable cyber places may be repeatedly targeted by offenders because committing the crime requires less effort. In this sense, some websites may possess certain characteristics inherent to the place that make them particularly vulnerable to website defacements. For example, Zone-H data shows that known website vulnerabilities were one of the main entry points for defacers, together with file inclusion and SQL injection (Holt, Leukfeldt, et al., 2020) (Appendix A). If, for either or both reasons, offenders repeatedly deface the same websites, cybercrime would be concentrated in specific cyber places, thus creating a virtual hot spot. Although the Zone-H data are not optimal for this purpose, the evidence we found is an important indicator of spatial patterns of repeated victimization.

Repeat online harassment dynamics also describe spatial patterns. Using the CACC (Miethe et al., 2008) and other complementary situational clustering techniques (Hart, 2019), we empirically observed that this cybercrime occurs in a few situational contexts. These contexts were defined both by individual factors and by cyber-place related factors (i.e. daily use of several social media and content management by users) in a framework of routine activities. We also noted that there were some situational contexts that were specific to victimization, but not offending. These findings are important for cybercrime prevention, as they can guide an efficient distribution of resources only in those contexts where crime is most likely (Brewer et al., 2020). In addition, they emphasize specific interventions, as the resources used to prevent repeat online harassment offending (e.g. handlers) are not necessarily the same as those used to prevent repeat victimization (e.g. self-protection). Furthermore, by defining the situational contexts at the level of individuals' routine activity profiles, the findings also orientate the design of SCP measures that can be useful in both cases (Reyns, 2010). Such approaches would also benefit users who have little or no risk of engaging in these dynamics, as they would not be impacted by unnecessary intervention.

Similarly, we found that online hate speech spread unevenly across all cyber places. Previous work revealed that this cybercrime displayed spatial patterns (Li et al., 2013; Williams & Burnap, 2016), but they related to the position that users occupied in geographic space. In contrast, our Random Forest model based on the metadata of tweets shows spatial concentrations of hate speech in cyberspace by identifying an overwhelming majority of secure micro places. This is an unusual but valid way to find cybercrime concentrations by discarding the cyber places where there is no crime. Considering the large volumes of data published daily on social media, this can prove very helpful to law enforcement agencies (Williams et al., 2013). With our algorithm, those in charge of monitoring social media for signs of radicalization would save a lot of time that can be spent on important research and surveillance tasks (Awan & Blakemore, 2012). In this respect, our algorithm also avoids many ethical conflicts. Since the data requirements of our model do not include any reference to the characteristics of the users who post the tweets, it avoids problems related to their identification or abuse of their privacy (see Mittelstadt et al., 2016; Williams et al., 2017). In this way, we provide a useful tool for cybercrime prevention that respects ethical standards.

10.2.3 Perpetrator patterns

Besides the spatiotemporal patterns, we also identified another variety that we called perpetrator patterns. These crime patterns refer to how cybercrime is concentrated among offenders. In this sense, we examined how website defacements were distributed per offender in CHAPTER VI, while in CHAPTER VIII we examined whether the dynamics of repeated online harassment were comprised of a homogeneous population of victims and offenders. By focusing on offenders, we wanted to address one of ECCA's alleged weaknesses (Cullen & Kulig, 2018; Miró-Llinares & Moneva, 2019a).

Regarding website defacements, we found that a few offenders were responsible for a large proportion of crimes. Concentration analyses (Fox & Tracy, 1988) showed that the Pareto principle that has been observed in other criminological studies in relation to traditional offending (Farrington & Wikstrom, 1994) also applied to this form of cybercrime. In fact, the results show that in the case of website defacements the concentration of crime per offender is even greater. We already discussed that there may be several factors related to the Zone-H self-report method that may have affected the results (e.g. groups of hackers may report attacks together or the same hacker may use more than one nickname to report attacks). In any case, the results are overwhelming. Therefore, we consider that the perpetrator patterns observed here are clear and able to guide cybercrime prevention policies. If just a few defacers are so prolific, a well implemented focused deterrence strategy that knocks them out of the game could reduce the cybercrime figures dramatically. However, we do not know to date that any such preventive strategies have been successfully implemented. Although some advances have been made for similar cyber-dependent crimes from the framework of deterrence by employing warning banners (Maimon et al., 2014; Testa et al., 2017), the results of those experiments show that the approach still has ample room for improvement in deterring offenders.

We also explored the role of repeat offenders in the context of online harassment. As there is often consensus that committing this cybercrime increases the likelihood of experiencing it in the future (e.g. Holt & Bossler, 2008; Ngo & Paternoster, 2011), we explored the shared situational contexts of repeat online harassment offending and victimization in search of more evidence. Unlike other studies whose analyses were variable-oriented, we used the CACC method (Miethe et al., 2008). This not only enabled us to observe how sets of variables interacted (i.e. case configurations), but also allowed us to observe the differences between contexts at the subject profile level. When we ranked the profiles of repeat offenders and victims according to the probability of committing or suffering online harassment and compared them, we observed statistically significant differences (i.e. the contexts in which repeat offending was more likely were not the same as those in which repeat victimization was more likely). This indicates that the situational profiles of the offenders were dissimilar to those of the victims. Contrary to previous research, our findings suggest that both actors were not part of a homogeneous population. These results underscore the importance of using different analytical approaches to the same research question, as they may reveal unique insights.

10.3 The Criminology of Cyber Place

Understanding the role of places in causing and preventing crime is at the heart of the ECCA approach. Since criminologists identified that crime is concentrated in small geographical units (Sherman et al., 1989), proponents of the Criminology of Place have advocated for targeted interventions to reduce crime in those micro places (Braga, Turchan, et al., 2019; Weisburd et al., 2016). This topic, however, has been largely overlooked with regard to cybercrime. As an exception, some scholars have recently drawn attention to this aspect (Miró-Llinares & Moneva, 2019a), but empirical research from this framework is non-existent. Acknowledging the difficulty in defining and delimiting places and —even more— micro places in cyberspace, in this thesis we sought to break through this unexplored scenario. The role of place in cybercrime prevention is something that we addressed across all chapters, but how would interventions in micro places be carried out in cyberspace?

The repeat victimization patterns of website defacement identified in CHAPTER VI are a valuable indicator for targeting preventive resources to specific cyber places. In that study, we defined each website that was defaced as a cyber place. And on some of them, we observed that cybercrime was disproportionate. By intervening in such places, website defacements could be substantially reduced. For prevention efforts to be effective, however, it would be essential to distinguish what kind of defacement is being addressed in each case. To do so, it would be necessary to examine the hack mode used. Note that there are many hack modes for performing website defacements, some of which are quite different from each other (Romagna & Van den Hout, 2017) (see also Appendix C). For example, password sniffing requires the use of a specific tool that monitors network traffic and extracts sensitive information such as usernames or passwords. Many of these attacks can be prevented by installing software aimed at implementing secure communication channels or security protocols. But password sniffing is considerably different to exploiting a web application bug. Bugs are unintended errors in the code that can generate vulnerabilities. Programs called debuggers or update patches can be used to fix bugs and thus eliminate vulnerabilities. However, a security protocol would not fix a bug and a debugger probably would not stop a hacker from sniffing a password. So, such preventive interventions, when misplaced, would be useless. Just as reducing theft or robbery would require a different strategy —although both are traditional property crimes that can affect individuals— each form of hacking requires a specific intervention.

We used the same concept of cyber place in CHAPTER VII. However, in this case, cyber places were not defaced websites but FMIWs that formed a network. With the data available, we would not know if there was more match-fixing in one cyber place than in another, so no cybercrime hot spots were identified. Which criteria should guide a preventive intervention then? We suggested to use the centrality of the nodes within the network to prioritize interventions in places (McGloin, 2005). If controllers' efforts were able to take down bridges, the cohesion of the network would be drastically reduced. This, in turn, would diminish the connectivity of the cyber places as much as the transit of users through them, hence reducing the number of available targets. As a result, while cybercrime would not be reduced directly, it would possibly be reduced indirectly. Considering that after each intervention the structure of the network would be different, the centrality of each cyber place in the network would be likely to change. It would therefore be necessary to recalculate the centrality measures in order to reprioritize the cyber places with greater connectivity. As with traditional crime, it is important to constantly monitor interventions in cyber places in order to be cost-effective.

Instead of analysing cyber place units in CHAPTER VIII, we examined situational contexts. Each situational context was comprised of four cyber places where users spent their time and their routine activities in micro places within them. Using self-reported data, we then calculated the risk of suffering or committing repeat online harassment in each context (Appendixes I and J). We argued that implementing SCP measures in the highest risk contexts would help reduce cybercrime —especially considering this involved repeat cybercrime—. However, we found that contexts with a higher risk of victimization were dissimilar from contexts with a higher risk of perpetration. This implies that some contexts could be manipulated to reduce offending opportunities, while others could reinforce self-protection mechanisms. Therefore, because SCP measures must be situation-specific to be effective (Clarke, 1997), they must address each context separately. Previous work has addressed the use of SCP to prevent similar cybercrime problems from a theoretical point of view (Reyns, 2010), but rigorous evaluations of their effectiveness have not yet been conducted. To avoid a waste of resources, it is necessary to implement SCP measures that have been previously evaluated. Otherwise, carefully designed interventions on paper may have little value in practice.

Following the trail set by the Criminology of Place, in CHAPTER IX we moved from analysing cyber places to analysing online microenvironments or cyber micro places. The reasoning was the same, if traditional crime is concentrated in geographical micro places, is it possible that cybercrime is concentrated in cyber micro places? We hypothesised that, if this was the case —and online hate speech spread to specific cyber micro places—, the metadata of the tweets would reflect the characteristics of the online microenvironments that are most prone to crime. In this way, it would be possible to distinguish cyber micro places containing online hate speech from those that do not. To test the hypothesis, we designed a novel machine learning model that used the metadata of the tweets and accounts as input to classify the messages as online hate speech or neutral communication. This is an innovative approach compared to traditional algorithms that classify the messages based on their content (e.g. Burnap & Williams, 2015). The results showed that the metadata of the tweets serve to rule out an overwhelming majority of secure cyber micro places where online hate speech is not spread. This is an important finding for law enforcement agencies and service providers, since they could save a lot of time by filtering irrelevant content with such models (Miró-Llinares, 2018). Thus, they could focus on monitoring a few cyber micro places (i.e. accounts and tweets) at risk of hosting cybercrime.

In short, the findings from the four empirical studies indicate that cybercrime concentrates on few cyber places and that interventions aimed at reducing cybercrime should be: (1) crime-specific, (2) constantly monitored, (3) properly evaluated, and (4) adapted to controllers' needs.

—Blank page—

CHAPTER XI

GENERAL CONCLUSIONS

This doctoral thesis sought to apply the ECCA approach to the analysis and prevention of various cybercrimes (i.e. website defacement, match-fixing, online harassment, and online hate speech). To achieve this, the investigation was developed in two stages: an initial one in which the approach is theoretically developed and a subsequent one in which it is then applied empirically. The former draws on the foundations of ECCA to examine its core propositions and transpose them into cyberspace with the aim of establishing the basis for their implementation. Following the assumptions of place-based approaches for crime analysis, the latter situated the cyber place as the central element of convergence to observe crime patterns in cyberspace in four empirical studies. In each of them, specific premises of the approach were tested on different objects of study to provide external validity to the general findings. The following is what we concluded [76].

First, after weighing arguments for and against, we believe that there is no substantive obstacle to applying the ECCA approach to crime committed in cyberspace (see CHAPTER II). Although their empirical demonstration is still tenuous, the ECCA propositions were indeed transposed to the context of cyberspace while retaining their essence. To address the key issues they reveal, existing research has already paved the

[76] The four articles that comprise this thesis already attest to the conclusions, limitations, and future research directions of each piece in their respective "Conclusions" sections. So, please, refer to each chapter for details (i.e. CHAPTER VI, CHAPTER VII, CHAPTER VIII, CHAPTER IX). In an effort to generalize and —at the same time— synthesize, this chapter presents the main ones.

way for some of the main theoretical frameworks (i.e., the Routine Activities Approach, and the Rational Choice Perspective together with its preventive corollaries —the SCP and the Situational Precipitators of Crime—), but those approaches with a strong geographical component have yet to be adapted to cyberspace (i.e. the Geometry of Crime, the Crime Pattern Theory). A more solid theoretical grounding in this respect will allow the discipline to advance steadily.

Second, methodologies used in traditional non-geographical crime analysis are equally useful for analysing different cybercrimes (see CHAPTER IV). We successfully employed some techniques for crime analysis (i.e. the rolling period methodology, HNA —a version of the SNA—, CACC, and Random Forests) in our empirical studies demonstrating that the essence of their application is not the geographical but the situational. In addition, the Data Science process proved to be crucial for properly handling cybercrime data in its various formats. The application of new methods with different forms of data will help to identify the most effective tools for cybercrime analysis.

Third, it appears that cybercrime is not randomly distributed in space and time, or among people either. Our empirical studies revealed temporal patterns of repeat victimization by website defacement in cyber places, which is committed by few defacers (see CHAPTER VI); situational contexts that concentrate most of the repeated victimization and offending online harassment events, which are specific to each case (see CHAPTER VIII); and concentrations of online hate speech in Twitter micro environments with specific characteristics, which differ from those containing neutral communication (see CHAPTER IX). The identification of such patterns in cyber places and among people is fundamental for cybercrime prevention, as it allows resources to be allocated efficiently.

Fourth, we observed that some environmental characteristics of cyber places favour the emergence of crime opportunities. In this sense, we found that —unlike regulated sport betting websites— the structural design of the FMIWs was almost completely deprived of precipitation-control strategies, and that these cyber places were linked to similar websites forming digital match-fixing neighbourhoods (see CHAPTER VII). We also identified certain metadata in Twitter messages (i.e. tweets) and accounts that were related to the dissemination of online hate speech after the occurrence of specific events in the physical world, and that such features could be used for its automated detection (see CHAPTER IX). It is likely that, by manipulating the characteristics of such online environments, crime opportunities in cyber places can be reduced.

In conclusion, we have shown the usefulness of crime analysis in identifying crime patterns in cyberspace, as well as the value of Environmental Criminology in understanding the role of cyber places in cybercrime prevention. Criminological research stands before an approach with enormous potential for the reduction of cybercrime. ECCA has demonstrated this in the past and, with proper understanding, can do so again in the future. Two options lie ahead: seize the opportunity or miss it. Let us not await a new “Nothing Works” in (cyber) Criminology; let us anticipate it by applying the approach that reverted the paradigm. Research focused on the individual must continue but, in Jeffery’s (1971) words, if we fail to change the individual, then we must try to change the environment in which he acts. Even if that environment is cyberspace.

11.1 Future research directions

The novel approach of this thesis provides a gateway for new research questions focusing on the following areas: developing and testing Environmental Criminology theories for cybercrime, methodologies for crime analysis in cyber places, and situational approaches for cybercrime prevention.

11.1.1 Developing and testing Environmental Criminology theories for cybercrime

Theoretical research on criminological theory applied to cybercrime is extensive (e.g. Bossler, 2020; Holt & Bossler, 2016). However, some theories have received much more attention than others (Holt & Bossler, 2014). For example, the Routine Activities Approach is one of the most examined frameworks —both theoretically and empirically— (Leukfeldt & Yar, 2016). In contrast, the closely related Geometry of Crime has hardly been considered by cybercrime scholars. Yet both are Environmental Criminology theories. Both focus on the crime event and not on the individual offender. Why is then the Routine Activities Approach so popular among cybercrime scholars and the Crime Pattern Theory is not? There are at least two possible explanations for this.

One explanation concerns the difficulty of adapting to cyberspace the strong geographical component of frameworks such as the Geometry of Crime or the Crime Pattern Theory. As discussed elsewhere (Miró-Llinares & Moneva, 2019a), this would be caused by the geographical gap: a conceptual barrier that prevents extracting only the spatial component of geographical assumptions. Thus, the theoretical reflection that requires applying geographical theories to cyberspace is perceived as overly complex, discouraging cybercrime scholars from using geographical frameworks in favour of situational ones.

Future research should concentrate on adapting the spatial elements of the Geometry of Crime and the Crime Pattern Theory to encourage their use. Even these frameworks that lay emphasis on the geography of crime incorporate key concepts that are spatial in nature and therefore also measurable in cyberspace. Discovering, for example, how target and offender activity spaces overlap in cyberspace can yield important insights into why crime opportunities emerge in certain cyber places. Similarly, understanding what types of cyber places can be considered crime attractors or crime generators would help explain why cybercrime is concentrated at certain places and times. Or perhaps a new type of cyber place that is unparalleled in physical space needs to be defined. Traditional explanations for the formation of hotspots may not be applicable to cyberspace, but there are others that may. In either case, there is only one way to shed light on these issues. Additional theoretical research is needed to adapt geographic frameworks to the structure of cyberspace, as well as empirical research to determine whether their premises and concepts are still valid for understanding why cybercrime events occur.

A second explanation is related to the apparent ease of measuring the fundamental premises of some frameworks. Certainly, measuring the convergence of the minimal elements of crime at a given time and place in cyberspace seems much simpler than measuring the activity space of Internet users. In fact, while there are dozens of studies that test the premise of routine activities online, those that test the premise of the Geometry of Crime in the same context are virtually non-existent (Holt & Bossler, 2017). There is a positive and a negative side to this. On the positive side, empirical evidence is piling up and synthesis research is providing increasingly strong evidence. On the negative side, many studies perpetuate existing research designs, so —if there were any— they would be replicating their flaws as well. In fact, this has produced certain stagnation in research on routine activities and cybercrime that has already been highlighted by some scholars (Holt & Bossler, 2016).

To overcome this obstacle, future research should breakdown each theoretical framework into its simplest premises and reformulate them in the form of hypotheses to test them. It is therefore recommended that the original sources of the theories are consulted and not later reinterpretations that may have distorted the initial meaning of the premises. This would allow to lay the foundations of solid basic knowledge on their feasibility, to —later on— conduct strong tests of the theories. It does not seem reasonable to attempt comprehensive tests if the validity of the basic assumptions is unknown. At this point, it is important that future research collects data that can properly test the hypothesis initially proposed, avoiding the malpractice of adjusting the hypothesis to the available data (i.e. hypothesis *ad hoc*). Otherwise, there is a risk of derailing the research from its main purpose. Furthermore, in an ideal scenario, these hypotheses should be tested with data from different cybercrimes, as their rejection may be conditioned by the nature of a specific crime event. Note that an assumption that is true for one cyber-dependent crime may not be so for another; and that it may even be true for one form of hacking but not for another.

11.1.2 Methodologies for crime analysis in cyber places

Unlike the previous directions for future research, those proposed under this heading are virtually uncharted territory. It is true that one of the main defining features of this doctoral thesis is the proposal a series of techniques to analyse crime in cyber places through Data Science (see CHAPTER IV). And that, to this end, we reviewed previous research that addressed similar objects of study. However, none of them worked specifically on the concept of place, nor analysed crime patterns, from the perspective of Environmental Criminology. Thus, we were confronted with a scenario in which there was nothing to compare; this is, we did not know if the methodologies we employed were the most appropriate to answer the research questions we posed.

For this reason, future research should explore the potential usefulness of other methodologies for analysing crime patterns at cyber places. Hence, the application of such methodologies would provide a basis for comparison to identify the best alternative. Special attention should be devoted to methodologies traditionally used in crime and place research that are not dependent on a geographical component for their implementation. Some examples are the rolling period methodology for analysing patterns of repeated victimization (Chainey, 2012), SNA —as well as HLN, its hyperlinked variant— for analysing relationships between networked entities (H. W. Park, 2003; Wasserman & Faust, 1994), the CACC for analysing situational contexts related to crime (Miethe et al., 2008), and the Random Forests for predicting crime at micro places (Wheeler & Steenbeek, 2020). There are also other methods yet to be developed for crime analysis in cyber places that have not been employed here. Namely, crime scripts (Cornish, 1994) to analyse the decision-making process of cyber offenders in specific microenvironments, aoristic signatures (Ratcliffe, 2002) to examine probabilistic spatiotemporal patterns of cybercrime, or agent-based simulation models (Weisburd et al., 2017) on the interaction of the minimum elements of crime and their controllers in cyberspace. It is important to underscore that methodologies designed to answer questions related to how cybercrime varies over time may be especially relevant to understanding a phenomenon that occurs in non-geographical environments.

11.1.3 Situational approaches for cybercrime prevention

Since the advent of the application of Environmental Criminology theories to cyberspace (Grabosky, 2001), there have been many notable attempts to develop situational approaches to cybercrime prevention. Among these efforts, SCP measures stand out as a widespread approach among cybercrime scholars (Brewer et al., 2020). After Newman and Clarke's (2003) work on e-commerce crime —which added a new dimension to crime analysis— other works such as Reyns' (2010) on SCP measures for cyberstalking, or Sidebottom and Tilley's (2017) on leaky systems, adopted a situational approach. Some proposals for preventive measures from this approach are exercises of reflection (e.g. Hinduja & Kooi, 2013), while others support their proposals on empirical data (Hutchings & Holt, 2015, 2017). However, with a few exceptions (Maimon et al., 2014; Testa et al., 2017), most crime prevention research from an situational perspective shares the common feature of being theoretical exercises. Such exercises are necessary, especially in the initial stages of the development of analytical frameworks, but over time efforts should be directed at grounding them in empirical evidence.

For this reason, future research on situational preventive measures for cybercrime should move towards practical implementation in concrete cyber places. In this regard, there are two approaches that could be developed in synergy with each other: SCP and Cybercrime Prevention Through Environmental Design. The former, much more developed than the latter, should be supported by review and synthesis research on existing publications that serve to compile the state of the art (e.g. Hartel et al., 2011) and expedite its implementation. And although the latter approach would admit of an initial theoretical effort, this should always be oriented to a subsequent measurement of the effects produced by the design of secure places in cyberspace. Often, this delicate task will require interdisciplinary collaboration between criminologists and other social scientists responsible for proposing preventive mechanisms, and computer scientists capable of implementing them. Moreover, enlisting private entities for the implementation of such measures and designs in controlled environments would represent an important asset for testing them in a real scenario. Note that, in both cases, the implementation of the designs or measures is as important as their subsequent evaluation. Otherwise, the final purpose of the research would be undermined. Such research projects would be resource-consuming (e.g. design of measures and environments, implementation, form of participation, sample collection), so again collaboration with the third sector would be crucial to secure sufficient funding.

11.2 Limitations

This thesis provides no explanation for the root causes of cybercrime. The truth is it never intended to do so. Rather, this thesis aims to test a number of hypotheses regarding the immediate causes of cybercrime. In particular, such hypotheses relate to the crime opportunities that emerge in a variety of cyber places and produce cybercrime events. This should not come as a surprise, since the ECCA approach has never been interested in the root causes of crime, nor characterised by a large explanatory scope (Bottoms, 2012; Cullen & Kulig, 2018). In contrast, ECCA is a medium-range approach that seeks to understand why crime opportunities occur and how they can be manipulated in very specific contexts (Bruinsma & Johnson, 2018; Wortley & Townsley, 2017a). While reviewing the application of its theories to cybercrime and transposing its propositions to cyberspace here (see CHAPTER II), we tried to maintain the essence of the approach.

Perhaps because of this philosophy of “less is more”, this thesis did not address all the questions it raised either. Of the six questions that were identified as key to ascertaining that the ECCA approach can be applied to cyberspace, only four were addressed empirically (see CHAPTER V), and even these four were not fully answered. As stated before, answering all these questions rigorously is a task that overwhelms this doctoral thesis. Furthermore, there are obviously many other factors beyond the will of the researcher that affect the ability to answer research questions adequately, such as time or available resources. An essential resource when attempting to answer such questions using quantitative methodologies is data. For this reason, this thesis approaches crime analysis from Data Science (see CHAPTER IV), a necessary approach that allowed us to extract the maximum potential from the available data to test our hypotheses.

In general, the available cybercrime data do not yet allow for robust testing of theoretically derived hypotheses from the ECCA framework, nor for unravelling the plausible causal processes underlying crime events in cyber places. In this sense, it should be noted that the true value of the place for traditional crime prevention could not be disentangled until crime data was available at the micro level (e.g. building blocks, street segments, postal addresses, grids). Sherman and collaborators' (1989) first study on crime and place had to rely on a proxy measure (i.e. emergency calls) to estimate the concentration of crime in specific locations. Today, criminologists of place use highly accurate geopositioned data to explain the law of crime concentration (Weisburd, 2015; Weisburd et al., 2016) —even though the data are not yet perfect—. The current state of cybercrime data is still far from this situation, which prevents the full potential of ECCA from being unleashed in cyberspace. This thesis is a good example. In the absence of official cybercrime data sources, we resorted to third party databases, implemented crawlers to scrap data from FMIWs, administered a questionnaire to collect self-reported data, and used the Twitter API to obtain a sample of social media messages (see CHAPTER IV). This is far from an ideal scenario where law enforcement agencies and other security authorities provide quality, open and anonymized data for the public to analyse. Although there is progress in this area [e.g. the Longitudinal Internet Studies for the Social sciences (LISS) panel data in The Netherlands [77], the Crime Survey for England and Wales (CSEW) in the United

[77] For more information, visit https://www.lissdata.nl/.

Kingdom [78], or the public archives on Twitter information operations [79]], limitations in most cybercrime data sources are still significant.

[78] Since 2016, the CSEW has been incorporating a set of questions related to victimization by different types of cybercrime that generate comparable data. For more information, visit https://www.ons.gov.uk/peoplepopulationandcommunity/crimeandjustice/datasets/crimeinenglandandwalesquarterlydatatables.

[79] For more information, visit https://transparency.twitter.com/en/information-operations.html.

—Blank page—

CHAPTER XI

CONCLUSIONES GENERALES

La presente tesis doctoral ha pretendido aplicar el enfoque de ECCA al análisis y prevención de distintos cibercrímenes (i.e. desfiguraciones web, amaño de partidos, acoso online, y discurso de odio online). Para lograrlo, la investigación se ha desarrollado en dos etapas: una inicial que ha servido para desarrollar teóricamente el enfoque y una sucesiva en la que se ha aplicado empíricamente. La primera se inspira en los cimientos de ECCA para examinar sus proposiciones centrales y trasladarlas al ciberespacio con el objetivo de sentar las bases para su implementación. Atendiendo a los presupuestos de los enfoques basados en lugares para el análisis delictivo, la segunda sitúa el ciber lugar como el elemento central de convergencia para observar patrones delictivos en el ciberespacio a través de cuatro estudios empíricos. En cada uno de ellos, se han contrastado premisas específicas del enfoque sobre distintos objetos de estudio para dotar de validez externa a los resultados generales. A continuación, se presentan nuestras conclusiones [80].

En primer lugar, tras sopesar los argumentos a favor y en contra, creemos que no existe ningún obstáculo sustantivo para aplicar el enfoque de ECCA al crimen cometido en el ciberespacio (véase CHAPTER II). A pesar de que su demostración empírica es

[80] Los cuatro artículos que componen esta tesis ya atestiguan cuáles son las conclusiones, limitaciones, y líneas de investigación futura de cada obra en sus respectivas secciones de “Conclusiones”. Por tanto, rogamos acudan a cada capítulo para consultar los detalles (i.e. CHAPTER VI, CHAPTER VII, CHAPTER VIII, CHAPTER IX). En un esfuerzo tanto de generalizar como —al mismo tiempo—de sintetizar, en este capítulo se presentan las principales.

todavía tenue, las proposiciones de ECCA han sido efectivamente trasladadas al contexto del ciberespacio logrando mantener su esencia. A la hora de abordar las cuestiones clave que revelan, la investigación existente ya ha allanado el camino que conduce a algunos de los marcos teóricos principales (i.e. el Enfoque de las Actividades Cotidianas, y la Perspectiva de la Elección Racional junto con sus corolarios preventivos —la SCP y los Precipitadores Situacionales del Crimen—), pero todavía quedan por adaptar al ciberespacio aquellos enfoques que tienen un marcado componente geográfico (i.e. la Geometría del Crimen, la Teoría del Patrón Delictivo). Una fundamentación teórica más sólida a este respecto permitiría a la disciplina avanzar con firmeza.

En segundo lugar, las metodologías empleadas para el análisis delictivo tradicional no geográfico han demostrado ser igualmente útiles para analizar distintos cibercrímenes (véase CHAPTER IV). Hemos utilizado satisfactoriamente algunas técnicas de análisis delictivo (i.e. la metodología de periodos en movimiento, el HNA —una versión del SNA—, el CACC, y los Bosques Aleatorios) en nuestros estudios empíricos demostrando que la esencia de su aplicación no recae en lo geográfico sino en lo situacional. Además, el proceso de Ciencia de Datos ha resultado crucial para manejar adecuadamente los datos de cibercrimen en sus distintos formatos. La aplicación de nuevos métodos sobre distintas formas de datos en el futuro permitirá identificar cuáles son las herramientas más efectivas para el análisis del cibercrimen.

En tercer lugar, parece que el cibercrimen no se distribuye aleatoriamente en el espacio, el tiempo, ni tampoco entre las personas. Nuestros estudios empíricos han revelado patrones temporales de victimización repetida por desfiguración web en los ciber lugares, y que estos son cometidos por unos pocos desfiguradores (véase CHAPTER VI); contextos situacionales donde se concentran la mayoría de eventos de agresión y victimización repetida por acoso en línea, y que estos son específicos en cada caso (véase CHAPTER VIII); así como concentraciones de discurso de odio en línea en microentornos de Twitter con características específicas, y que estos son diferentes a aquellos que contienen comunicación neutral (véase CHAPTER IX). La identificación de tales patrones en los ciber lugares y entre las personas es fundamental para prevenir el cibercrimen, ya que permite ubicar los recursos de manera eficiente.

En cuarto lugar, hemos observado que algunas características ambientales de los ciber lugares favorecen la aparición de oportunidades delictivas. En este sentido, hemos hallado que —a diferencia de las páginas web de apuestas deportivas reguladas— el diseño estructural de las FMIW estaba casi por completo privado de estrategias de control de precipitadores, y que esos ciber lugares estaban vinculados a páginas web similares formando barrios digitales de apuestas amañadas (véase CHAPTER VII). También hemos identificado algunos metadatos en los mensajes (i.e. tuits) y las cuentas de Twitter que estaban relacionados con la diseminación de discurso de odio en línea tras la ocurrencia de ciertos eventos en el espacio físico, y que tales elementos podrían servir para su detección automatizada (véase CHAPTER IX). Es probable que, al manipular las características de estos entornos en línea, se puedan reducir las oportunidades delictivas en los ciber lugares.

En conclusión, hemos mostrado la utilidad del análisis delictivo a la hora de identificar patrones delictivos en el ciberespacio, así como el valor de la Criminología Ambiental para comprender el rol que juegan los ciber lugares en la prevención del cibercrimen. La investigación criminológica se encuentra frente a un enfoque con un enorme potencial para reducir el cibercrimen. ECCA ya lo ha demostrado en el pasado y, con el conocimiento adecuado, puede volver a hacerlo en el futuro. Ante nosotros se revelan dos opciones: aprovechar esta oportunidad o perderla. No esperemos a un nuevo

“Nada Funciona” en (ciber) Criminología; anticipémonos al aplicar el enfoque que logró revertir el paradigma. La investigación centrada en el individuo debe continuar, pero, en palabras de Jeffery (1971), si fracasamos al tratar de cambiar al individuo, entonces debemos tratar de cambiar el entorno donde actúa. Incluso si el entorno es el ciberespacio.

10.1 Líneas de investigación futura

El novedoso enfoque de esta tesis abre la puerta a nuevas preguntas de investigación centradas en los siguientes ámbitos: desarrollo y comprobación de las teorías de la Criminología Ambiental para el cibercrimen, metodologías para el análisis delictivo en los ciber lugares, y enfoques situacionales para la prevención del cibercrimen.

10.1.1 Desarrollo y comprobación de las teorías de la Criminología Ambiental para el cibercrimen

La investigación sobre la teórica criminológica aplicada al cibercrimen es extensa (p. ej. Bossler, 2020; Holt & Bossler, 2016). Sin embargo, algunas teorías han recibido muchas más atención que otras (Holt & Bossler, 2014). Por ejemplo, el Enfoque de las Actividades Cotidianas es uno de los marcos más estudiados —tanto teórica como empíricamente— (Leukfeldt & Yar, 2016). Por el contrario, la estrechamente relacionada Geometría del Crimen apenas ha sido considerada por los investigadores. Y eso que ambas pertenecen a las Teorías de la Criminología Ambiental. Ambas ponen el foco sobre el evento delictivo y no en el infractor individual. Entonces ¿por qué el Enfoque de las Actividades Cotidianas es tan popular entre quienes investigan cibercrimen y la Teoría del Patrón delictivo no?

Una posible explicación tiene que ver con la dificultad de adaptar al ciberespacio el marcado componente geográfico que poseen marcos como la Geometría del Crimen o la Teoría del Patrón Delictivo. Tal y como se discute en otro trabajo (Miró-Llinares & Moneva, 2019a), ello podría estar causado por la denominada brecha geográfica: una barrera conceptual que evita que se extraiga únicamente el componente espacial de los presupuestos geográficos. De esta forma, la reflexión teórica que requiere aplicar las teorías geográficas al ciberespacio se percibe como excesivamente compleja, desalentando a los investigadores para utilizar marcos geográficos en favor de los situacionales.

Las investigaciones futuras deberían centrar sus esfuerzos en adaptar los elementos espaciales de la Geometría del Crimen y la Teoría del Patrón Delictivo para fomentar su uso. Incluso estos marcos que ponen el acento en la geografía del crimen incorporan conceptos clave que son de naturaleza espacial y, por tanto, también observables en el ciberespacio. El hecho de descubrir, por ejemplo, cómo los espacios de actividad de objetivos e infractores se superponen en el ciberespacio puede proporcionar ideas importantes sobre por qué las oportunidades delictivas aparecen en ciertos ciber lugares. Asimismo, comprender qué tipos de ciber lugares se pueden considerar atractores del crimen o generadores del crimen podría ayudar a explicar por qué el cibercrimen se concentra en determinados momentos y lugares. O quizá sea necesario definir un nuevo tipo de ciber lugar que no tiene un homólogo en el espacio físico. Las explicaciones tradicionales para la formación de puntos calientes puede que no sean aplicables al ciberespacio, pero hay otras que pueden serlo. En cualquier caso, sólo hay una forma de arrojar algo de luz sobre estas cuestiones. Es necesario realizar más investigaciones teóricas para adaptar los marcos geográficos a la estructura del ciberespacio, así como investigaciones empíricas para determinar si sus premisas y conceptos siguen siendo válidos para comprender por qué ocurren los eventos delictivos en este entorno.

Una segunda explicación se relaciona con la aparente facilidad para medir las premisas básicas de algunos marcos teóricos. Ciertamente, medir la convergencia de los elementos mínimos del crimen en un momento y lugar determinados en el ciberespacio parece mucho más sencillo que medir los espacios de actividad de los usuarios de Internet. De hecho, mientras que existen docenas de estudios que someten a prueba la premisa de las actividades cotidianas en línea, los que contrastan la premisa de la Geometría del Crimen son virtualmente inexistentes (Holt & Bossler, 2017). Esto tiene un lado positivo y otro negativo. El aspecto positivo es que se está amontonando la evidencia empírica y las investigaciones de síntesis están aportando evidencias cada vez más robustas. El aspecto negativo es que muchos estudios perpetúan los diseños de investigación existentes, por lo que también estaría replicando sus defectos —si es que los tuvieran—. De hecho, esto ha producido cierto estancamiento en la investigación sobre las actividades cotidianas y el cibercrimen que algunos académicos ya han puesto de relieve (Holt & Bossler, 2016).

Para superar este obstáculo, las investigaciones futuras deberían descomponer cada marco teórico en sus premisas más sencillas y reformularlas en forma de hipótesis para contrastarlas. Por ello, se recomienda que se consulten las fuentes iniciales de estas teorías y no reinterpretaciones posteriores que puedan haber distorsionado el sentido original de sus presupuestos. Esto permitiría asentar sólidamente los conocimientos básicos sobre su viabilidad, para —posteriormente— llevar a cabo contrastes robustos de las teorías. Y es que no parece razonable intentar realizar contrastes completos cuando la validez de los presupuestos más básicos todavía se desconoce. En este punto, es importante que la investigación futura recoja datos que permitan contrastar adecuadamente las hipótesis inicialmente planteadas, evitando así la mala práctica de ajustar las hipótesis a los datos disponibles. (i.e. hipótesis *ad hoc*). De lo contrario, existe el riesgo de desviar la investigación de su propósito principal. Además, en un escenario ideal, estas hipótesis deberían ser contrastadas con datos de distintos cibercrímenes, ya que su rechazo podría estar condicionado por la naturaleza de un evento delictivo específico. En este sentido, es importante destacar que, aunque un presupuesto que sea cierto para un cibercrimen puro, puede no serlo para otro; y que incluso puede ser cierto para una modalidad de hacking, pero no para otra.

10.1.2 Metodologías para el análisis delictivo en los ciber lugares

A diferencia de las líneas de investigación futura que se han expuesto previamente, las que se proponen bajo este epígrafe constituyen un territorio virtualmente inexplorado. Es cierto que una de las principales características definitorias de esta tesis es la propuesta de una serie de técnicas de análisis delictivo para ciber lugares a través de la Ciencia de Datos (véase CHAPTER IV) y que, para ello, se han revisado investigaciones previas que abordan objetos de estudios similares. Sin embargo, ninguna de ellas trabaja con el concepto de ciber lugar ni analiza patrones delictivos desde la perspectiva de la Criminología Ambiental. Por ello nos enfrentamos a un escenario que no tiene parangón; es decir, no podemos saber si las metodologías que aquí se utilizan son las más apropiadas para responder las preguntas de investigación que planteamos.

Por este motivo, la investigación futura debería explorar la potencial utilidad de otras metodologías para analizar patrones delictivos en ciber lugares. De esta forma, tales metodologías proporcionarían las bases para realizar comparaciones, y así identificar la mejor alternativa. Se debería prestar especial atención a aquellas metodologías que se ha utilizado tradicionalmente en la investigación sobre crimen y lugar, y que no dependen de un componente geográfico para su implementación. Algunos ejemplos son la metodología de periodos en movimiento para analizar patrones de victimización repetida (Chainey, 2012), el SNA —junto con el HLN, su variante para hipervínculos— para analizar las relaciones entre entidades en red (H. W. Park, 2003; Wasserman & Faust, 1994), el CACC para analizar contextos situacionales relacionados con el crimen (Miethe et al., 2008), y los Bosques Aleatorios para predecir el crimen en micro lugares (Wheeler & Steenbeek, 2020). También existen otros métodos todavía por desarrollar para el análisis delictivo en ciber lugares que no se han empleado aquí. A saber, los guiones delictivos (Cornish, 1994) para analizar los procesos de toma de decisiones de los ciber criminales en microentornos específicos, las firmas aorísticas (Ratcliffe, 2002) para examinar los patrones espaciotemporales probabilísticos de cibercrimen, o los modelos de simulación basados en agentes (Weisburd et al., 2017) sobre la interacción de los elementos mínimos del crimen y sus controladores en el ciberespacio. Es importante subrayar que las metodologías diseñadas para responder a la pregunta de cómo varía el cibercrimen a lo largo del tiempo pueden resultar especialmente relevantes para comprender un fenómeno que ocurre en entornos que no son geográficos.

10.1.3 Enfoques situacionales para la prevención del cibercrimen

Desde los orígenes de la aplicación de las teorías de la Criminología Ambiental al ciberespacio (Grabosky, 2001), han existido muchos intentos notables de desarrollar enfoques situacionales para la prevención del cibercrimen. Entre ellos, las medidas de SCP destacan como un enfoque extendido entre los investigadores (Brewer et al., 2020). Desde el trabajo de Newman y Clarke (2003) sobre el comercio electrónico —que dotó de una nueva dimensión al análisis delictivo— otros trabajos adoptaron el enfoque situacional, como el de Reyns (2010) sobre medidas de prevención situacional para el ciberacoso, o el de Sidebottom y Tilley (2017) sobre sistemas con fugas. Algunas de las propuestas de medidas preventivas que se plantean desde este enfoque consisten en ejercicios de reflexión (p. ej. Hinduja & Kooi, 2013), mientras que otras apoyan sus propuestas en datos empíricos (Hutchings & Holt, 2015, 2017). Sin embargo, salvo algunas excepciones (Maimon et al., 2014; Testa et al., 2017), la mayoría de la investigación en materia de prevención del crimen desde una perspectiva situacional tiene en común ser un ejercicio teórico. Estos ejercicios son necesarios, especialmente en las etapas iniciales de desarrollo de los marcos analíticos, pero con el paso del tiempo se debería hacer un esfuerzo para asentarlos sobre evidencias empíricas.

Por esta razón, la investigación futura sobre medidas de prevención situacional para el cibercrimen debería avanzar hacia su implementación práctica en ciber lugares concretos. En este sentido, hay dos enfoques que se podrían desarrollar sinérgicamente: la SCP y la Prevención del Cibercrimen a Través del Diseño Ambiental. La primera, mucho más desarrollada que la segunda, debería ser apoyada por trabajos de revisión y síntesis de las publicaciones existentes, que servirían para compilar el estado del arte de la cuestión (p. ej. Hartel et al., 2011) y acelerar su implementación. Y a pesar de que el segundo enfoque admitiría un ejercicio teórico inicial, este se debería orientar siempre a la posterior medición de los efectos que produce el diseño de lugares seguros en el ciberespacio. A menudo, esta delicada tarea requerirá colaboraciones interdisciplinares entre criminólogos y otros científicos sociales responsables de proponer mecanismos preventivos, e informáticos capaces de implementarlos. Además, la incorporación de las entidades privadas para la implementación de estas medidas y diseños en entornos controlados representaría un activo importante para comprobar su eficacia en escenarios reales. Cabe destacar que, en ambos casos, la implementación de los diseños o medidas es tan importante como su posterior evaluación. Si no es así, el propósito final de la investigación se vería socavado. Este tipo de proyectos de investigación consumiría muchos recursos (p. ej. el diseño de medidas y entornos, su implementación, las formas de participación, la recogida de muestra) así que, nuevamente, la colaboración con el tercer sector sería crucial para asegurar una fuente de financiación suficiente.

10.2 Limitaciones

Esta tesis no proporciona una explicación para las causas estructurales del cibercrimen. La verdad es que nunca lo ha pretendido. Lo que ha pretendido esta tesis es contrastar una serie de hipótesis relacionadas con las causas inmediatas del cibercrimen. En concreto, estas hipótesis abordan la aparición de oportunidades delictivas en distintos ciber lugares donde se producen eventos delictivos. Esto no debería ser una sorpresa, ya que el enfoque de ECCA nunca se ha interesado por las causas estructurales del crimen, ni se ha caracterizado por tener un gran alcance explicativo (Bottoms, 2012; Cullen & Kulig, 2018). Lo que persigue el enfoque de ECCA con su alcance medio es comprender por qué aparecen las oportunidades delictivas y como se pueden manipular en contextos muy específicos (Bruinsma & Johnson, 2018; Wortley & Townsley, 2017a). Al revisar cómo se han aplicado estas teorías al cibercrimen y trasladar sus proposiciones al ciberespacio (véase CHAPTER II), hemos tratado de mantener la esencia de este enfoque.

Quizá debido a esta filosofía del "menos es más", esta tesis tampoco ha pretendido abordar todas las cuestiones que ha planteado. De las seis preguntas que se identificaron como clave para comprobar que el enfoque de ECCA se puede aplicar al ciberespacio, sólo cuatro se atajaron empíricamente (véase CHAPTER V), e incluso esas cuatro no fueron respondidas por completo. Tal y como se menciona anteriormente, responder a todas estas preguntas de manera rigurosa es una tarea que desborda el alcance de esta tesis doctoral. Además, obviamente, existen muchos otros factores más allá de la voluntad del investigador que influyen en la capacidad de responder a las preguntas de investigación adecuadamente, como el tiempo o la disponibilidad de recursos. Un recurso esencial a la hora de responder tales cuestiones a través de metodología cuantitativas son los datos. Por ello, esta tesis se aproxima al análisis delictivo desde la Ciencia de Datos (véase CHAPTER IV), un enfoque necesario que nos ha permitido extraer el máximo potencial de los datos disponibles para contrastar nuestras hipótesis.

En general, los datos disponibles sobre cibercrimen no permiten todavía realizar contrastes robustos de hipótesis derivadas teóricamente del enfoque de ECCA, ni la revelación de los procesos causales plausibles que subyacen a la ocurrencia de eventos delictivos en el ciberespacio. En este sentido, se deber resaltar que el verdadero valor del lugar para la prevención del crimen tradicional no pudo ser desentrañado hasta que se dispuso de datos de crimen a nivel micro (p. ej. manzanas, segmentos de calle, direcciones postales, rejillas). El primer estudio de Sherman y colaboradores (1989) sobre el crimen y el lugar dependió de una medida indirecta (i.e. llamadas de emergencia) para estimar la concentración del crimen en lugares específicos. A día de hoy, los criminólogos del lugar utilizan datos geoposicionados con extremada precisión para explicar la Ley de la Concentración del Crimen (Weisburd, 2015; Weisburd et al., 2016) —y eso que los datos todavía no son perfectos—. El estado actual de los datos sobre cibercrimen todavía está lejos de esta situación, lo que impide desatar todo el potencial de ECCA en el ciberespacio. Esta tesis es buen ejemplo de ello. Ante la ausencia de fuentes de datos oficiales de cibercrimen, recurrimos a bases de datos de terceras partes, implementamos rastreadores web para extraer datos de las FMIW, administramos un cuestionario para recabar datos auto revelados, y usamos la API de Twitter para obtener una muestra de mensajes de redes sociales (véase CHAPTER IV). Esto queda lejos del escenario ideal en el que las fuerzas del orden y otras autoridades de la seguridad proporcionan datos anónimos, abiertos y de calidad para que el público los pueda analizar. A pesar de que se ha avanzado en este ámbito [p. ej. los datos de panel de los Estudios Longitudinales sobre Internet para las Ciencias Sociales (LISS) en Países Bajos [81], la Encuesta sobre Criminalidad de Inglaterra y Gales (CSEW) en Reino Unido [82], o los archivos públicos sobre las actividades de información de Twitter [83]], las limitaciones en la mayoría de los datos de cibercrimen todavía son significativas.

[81] Para más información, visitar https://www.lissdata.nl/.

[82] Desde 2016, la CSEW ha incorporado un conjunto de preguntas relacionadas con la victimización por distintos tipos de cibercrimen que permiten generar datos comparables. Para más información, visitar https://www.ons.gov.uk/peoplepopulationandcommunity/crimeandjustice/datasets/crimeinenglandandwalesquarterlydatatables.

[83] Para más información, visitar https://transparency.twitter.com/en/information-operations.html.

REFERENCES

Ackland, R., & Shorish, J. (2009). Network Formation in the Political Blogosphere: An Application of Agent Based Simulation and e-Research Tools. *Computational Economics*, *34*(4), 383–398. https://doi.org/10.1007/s10614-009-9173-7

Agarwal, N., Gupta, R., Singh, S. K., & Saxena, V. (2017). Metadata based multi-labelling of YouTube videos. *2017 7th International Conference on Cloud Computing, Data Science & Engineering - Confluence*, 586–590. https://doi.org/10.1109/CONFLUENCE.2017.7943219

Allaire, J. J., Xie, Y., McPherson, J., Luraschi, J., Ushey, K., Atkins, A., Wickham, H., Cheng, J., Chang, W., & Iannone, R. (2020). *rmarkdown: Dynamic Documents for R* (Version 2.1) [Computer software]. https://rmarkdown.rstudio.com

Andresen, M. A. (2010). The Place of Environmental Criminology within Criminological Thought. In M. A. Andresen, P. J. Brantingham, & B. J. Kinney (Eds.), *Classics in environmental criminology* (pp. 5–28). Simon Fraser University Publications; CRC Press/Taylor & Francis.

Andresen, M. A., Brantingham, P. J., & Kinney, B. J. (Eds.). (2010). *Classics in environmental criminology*. Simon Fraser University Publications; CRC Press/Taylor & Francis.

Ashcroft, M., Fisher, A., Kaati, L., Omer, E., & Prucha, N. (2015). Detecting Jihadist Messages on Twitter. *2015 European Intelligence and Security Informatics Conference*, 161–164. https://doi.org/10.1109/EISIC.2015.27

Awan, I. (2016). Islamophobia On Social Media: A Qualitative Analysis Of The Facebook’S Walls Of Hate. *International Journal of Cyber Criminology*, *10*(1), 1–20. https://doi.org/10.5281/zenodo.58517

Awan, I., & Blakemore, B. (Eds.). (2012). *Policing Cyber Hate, Cyber Threats and Cyber Terrorism* (1st ed.). Routledge. https://www.taylorfrancis.com/books/9781315601076

Baker, M. (2016). 1,500 scientists lift the lid on reproducibility. *Nature*, *533*(7604), 452–454. https://doi.org/10.1038/533452a

Barabási, A.-L., & Pósfai, M. (2016). *Network Science*. Cambridge University Press.

Barr, R., & Pease, K. (1990). Crime Placement, Displacement, and Deflection. *Crime and Justice: A Review of Research*, *12*, 277–318.

Baum, K., Catalano, S., Rand, M., & Rose, K. (2009). *Stalking Victimization in the United States* (pp. 1–16). US Department of Justice. https://www.justice.gov/sites/default/files/ovw/legacy/2012/08/15/bjs-stalking-rpt.pdf

Beran, T., & Li, Q. (2008). The Relationship between Cyberbullying and School Bullying. *The Journal of Student Wellbeing*, *1*(2), 16. https://doi.org/10.21913/JSW.v1i2.172

Berger, J. M., & Morgan, J. (2015). *The ISIS twitter census: Defining and describing the population of ISIS supporters on Twitter* (No. 20; The Brookings Project on US Relations with the Islamic World, pp. 1–68). Center for Middle East Policy.

https://www.brookings.edu/wp-content/uploads/2016/06/isis_twitter_census_berger_morgan.pdf

Bernasco, W. (2008). Them Again?: Same-Offender Involvement in Repeat and Near Repeat Burglaries. *European Journal of Criminology*, *5*(4), 411–431. https://doi.org/10.1177/1477370808095124

Bichler, G. (2019). *Understanding criminal networks: A research guide*. University of California Press.

Bichler, G., & Malm, A., E. (2008). A social network analysis of the evolution of the Environmental Criminology and Crime Analysis (ECCA) symposiums. *Crime Patterns and Analysis*, *1*, 5–22.

Bode, L., & Dalrymple, K. E. (2016). Politics in 140 Characters or Less: Campaign Communication, Network Interaction, and Political Participation on Twitter. *Journal of Political Marketing*, *15*(4), 311–332. https://doi.org/10.1080/15377857.2014.959686

Bossler, A. M. (2020). Contributions of criminological theory to the understanding of cybercrime offending and victimization. In E. R. Leukfeldt & T. J. Holt (Eds.), *The human factor of cybercrime* (pp. 29–59). Routledge.

Bossler, A. M., & Holt, T. J. (2009). On-line Activities, Guardianship, and Malware Infection: An Examination of Routine Activities Theory. *International Journal of Cyber Criminology*, *3*(1), 400–420.

Bossler, A. M., Holt, T. J., & May, D. C. (2012). Predicting Online Harassment Victimization Among a Juvenile Population. *Youth & Society*, *44*(4), 500–523. https://doi.org/10.1177/0044118X11407525

Bottoms, A. (2012). Developing Socio-Spatial Criminology. In M. Maguire, R. Morgan, & R. Reiner (Eds.), *The Oxford Handbook of Criminology* (pp. 450–489). Oxford University Press. https://doi.org/10.1093/he/9780199590278.003.0016

Bowers, K. J. (2001). Small Business Crime: The Evaluation of a Crime Prevention Initiative. *Crime Prevention and Community Safety*, *3*(1), 23–42. https://doi.org/10.1057/palgrave.cpcs.8140079

Bowers, K. J., & Johnson, S. D. (2005). Domestic Burglary Repeats and Space-Time Clusters: The Dimensions of Risk. *European Journal of Criminology*, *2*(1), 67–92. https://doi.org/10.1177/1477370805048631

Braga, A. A., Turchan, B., Papachristos, A. V., & Hureau, D. M. (2019). Hot spots policing of small geographic areas effects on crime. *Campbell Systematic Reviews*, *15*(3). https://doi.org/10.1002/cl2.1046

Braga, A. A., Zimmerman, G., Barao, L., Farrell, C., Brunson, R. K., & Papachristos, A. V. (2019). Street Gangs, Gun Violence, and Focused Deterrence: Comparing Place-based and Group-based Evaluation Methods to Estimate Direct and Spillover Deterrent Effects. *Journal of Research in Crime and Delinquency*, *56*(4), 524–562. https://doi.org/10.1177/0022427818821716

Brantingham, P. J., & Brantingham, P. L. (Eds.). (1981). *Environmental criminology*. Sage Publications.

Brantingham, P. J., & Brantingham, P. L. (1984). *Patterns in crime*. Macmillan ; Collier Macmillan.

Brantingham, P. L., & Brantingham, P. J. (1981). Notes on the Geometry of Crime. In P. J. Brantingham & P. L. Brantingham (Eds.), *Environmental criminology* (pp. 27–54). Sage Publications.

Brantingham, P. L., & Brantingham, P. J. (1993a). Environment, Routine, and Situation: Toward a Pattern Theory of Crime. In R. V. Clarke & M. Felson (Eds.), *Routine Activity and Rational Choice* (1st ed., pp. 259–294). Routledge. https://doi.org/10.4324/9781315128788-12

Brantingham, P. L., & Brantingham, P. J. (1993b). Nodes, paths and edges: Considerations on the complexity of crime and the physical environment. *Journal of Environmental Psychology*, *13*(1), 3–28. https://doi.org/10.1016/S0272-4944(05)80212-9

Brantingham, P. L., & Brantingham, P. J. (1995). Criminality of place: Crime generators and crime attractors. *European Journal on Criminal Policy and Research*, *3*(3), 5–26. https://doi.org/10.1007/BF02242925

Breiman, L. (2001). Random Forests. *Machine Learning*, *45*, 5–32.

Brenner, S. W. (2017). Nanocrime 2.0. In M. R. McGuire & T. J. Holt (Eds.), *The Routledge handbook of technology, crime and justice* (pp. 611–642). Routledge.

Brewer, R., de Vel-Palumbo, M., Hutchings, A., Holt, T., Goldsmith, A., & Maimon, D. (2020). *Cybercrime Prevention Theory and Applications.* Palgrave Macmillan UK. http://public.eblib.com/choice/PublicFullRecord.aspx?p=5979244

Bruinsma, G. J. N., & Johnson, S. D. (Eds.). (2018). *The Oxford Handbook of Environmental Criminology*. Oxford University Press. https://doi.org/10.1093/oxfordhb/9780190279707.001.0001

Burnap, P., & Williams, M. L. (2015). Cyber Hate Speech on Twitter: An Application of Machine Classification and Statistical Modeling for Policy and Decision Making: Machine Classification of Cyber Hate Speech. *Policy & Internet*, *7*(2), 223–242. https://doi.org/10.1002/poi3.85

Burris, V., Smith, E., & Strahm, A. (2000). White Supremacist Networks on the Internet. *Sociological Focus*, *33*(2), 215–235. https://doi.org/10.1080/00380237.2000.10571166

Button, M., & Cross, C. (2017). *Cyber frauds, scams and their victims*. Routledge, Taylor & Francis Group.

Cerezo-Domínguez, A. I., & Díez-Ripollés, J. L. (2010). La videovigilancia en las zonas públicas: Su eficacia en la reducción de la delincuencia. *Boletín Criminológico, 16*(121), 1–4.

Chainey, S. (2012). *Repeat victimisation* (JDiBrief Series, pp. 1–5). UCL Jill Dando Institute of Security and Crime Science. https://www.ucl.ac.uk/jdibrief/analysis/repeat_victimisation

Chatzakou, D., Kourtellis, N., Blackburn, J., De Cristofaro, E., Stringhini, G., & Vakali, A. (2017). Mean Birds: Detecting Aggression and Bullying on Twitter. *Proceedings of the 2017 ACM on Web Science Conference - WebSci '17*, 13–22. https://doi.org/10.1145/3091478.3091487

Chen, Y., Chen, P. S., Hwang, J., Korba, L., Song, R., & Yee, G. (2005). An analysis of online gaming crime characteristics. *Internet Research*, *15*(3), 246–261. https://doi.org/10.1108/10662240510602672

Cheong, M., & Lee, V. C. S. (2011). A microblogging-based approach to terrorism informatics: Exploration and chronicling civilian sentiment and response to terrorism events via Twitter. *Information Systems Frontiers*, *13*(1), 45–59. https://doi.org/10.1007/s10796-010-9273-x

Chiu, A. (2020, January 6). A government website was 'defaced' with pro-Iran messaging and an image of a bloodied Trump. Hackers claimed responsibility. *The Washington Post*.

https://www.washingtonpost.com/nation/2020/01/06/american-government-website-defaced-iran-hackers-bloodied-trump/

Choi, K.-S. (2008). Computer crime victimization and integrated theory: An empirical assessment. *International Journal of Cyber Criminology, 2*(1), 308–333.

Choi, K.-S., Lee, C. S., & Louderback, E. R. (2019). Historical Evolutions of Cybercrime: From Computer Crime to Cybercrime. In *The Palgrave Handbook of International Cybercrime and Cyberdeviance* (pp. 1–17). Springer International Publishing. https://doi.org/10.1007/978-3-319-90307-1_2-1

Choi, K.-S., & Lee, J. R. (2017). Theoretical analysis of cyber-interpersonal violence victimization and offending using cyber-routine activities theory. *Computers in Human Behavior, 73*, 394–402. https://doi.org/10.1016/j.chb.2017.03.061

CK Consulting, & Stichting VU-VUmc. (2017). *The monitoring systems of sports betting and warning mechanisms between public and private actors* (pp. 1–245). CK Consulting & Stichting VU-VUmc. http://ethisport.com/wp-content/uploads/sites/28/2017/06/Betmonitalert_Design-NB-DEF-2-06-2017.pdf

Clarke, R. V. (1980). "Situational" Crime Prevention: Theory and practice. *The British Journal of Criminology, 20*(2), 136–147. https://doi.org/10.1093/oxfordjournals.bjc.a047153

Clarke, R. V. (Ed.). (1992). *Situational crime prevention: Successful case studies*. Harrow and Heston Publishers.

Clarke, R. V. (Ed.). (1997). *Situational crime prevention: Successful case studies* (2. ed). Harrow and Heston.

Clarke, R. V. (1999). *Hot products understanding, anticipating and reducing demand for stolen goods*. Home Office, Policing and Reducing Crime Unit, Research, Development and Statistics Directorate. http://www7.bibl.ulaval.ca/doelec/lc2/monographies/2018/a2096210.pdf

Clarke, R. V. (2010). Crime Science. In *The SAGE Handbook of Criminological Theory* (pp. 271–283). SAGE Publications. https://doi.org/10.4135/9781446200926.n15

Clarke, R. V. (2018). Book Review. *Journal of Criminal Justice Education*, *29*(1), 157–159. https://doi.org/10.1080/10511253.2016.1258031

Clarke, R. V., & Cornish, D. B. (1985). Modeling Offenders' Decisions: A Framework for Research and Policy. *Crime and Justice: A Review of Research*, *6*, 147–185.

Clarke, R. V., & Eck, J. E. (2005). *Crime Analysis for Problem Solvers in 60 Small Steps* (p. 150). US Department of Justice.

Clough, J. (2010). *Principles of cybercrime*. Cambridge University Press.

Cohen, J. (1960). A Coefficient of Agreement for Nominal Scales. *Educational and Psychological Measurement*, *20*(1), 37–46. https://doi.org/10.1177/001316446002000104

Cohen, L. E., & Felson, M. (1979). Social Change and Crime Rate Trends: A Routine Activity Approach. *American Sociological Review*, *44*(4), 588. https://doi.org/10.2307/2094589

Cops, D., & Pleysier, S. (2014). Usual suspects, ideal victims and vice versa: The relationship between youth offending and victimization and the mediating influence of risky lifestyles. *European Journal of Criminology*, *11*(3), 361–378. https://doi.org/10.1177/1477370813500886

Cornish, D. B. (1994). Crimes as scripts. *Proceedings of the International Seminar on Environmental Criminology and Crime Analysis*, 30–45.

Cornish, D. B., & Clarke, R. V. (Eds.). (1986). *The reasoning criminal: Rational choice perspectives on offending*. Springer-Verlag.

Cornish, D. B., & Clarke, R. V. (2003). Opportunities, precipitators and criminal decisions: A reply to Wortley´s critique of situational crime prevention. In M. J. Smith & D. B. Cornish (Eds.), *Theory for practice in situational crime prevention* (pp. 41–96). Criminal Justice.

Costello, M., & Hawdon, J. (2019). Hate Speech in Online Spaces. In *The Palgrave Handbook of International Cybercrime and Cyberdeviance* (pp. 1–20). Springer International Publishing. https://doi.org/10.1007/978-3-319-90307-1_60-1

Cozens, P. M., Saville, G., & Hillier, D. (2005). Crime prevention through environmental design (CPTED): A review and modern bibliography. *Property Management*, *23*(5), 328–356. https://doi.org/10.1108/02637470510631483

Cross, C., & Blackshaw, D. (2015). Improving the Police Response to Online Fraud. *Policing*, *9*(2), 119–128. https://doi.org/10.1093/police/pau044

Csárdi, G., & Nepusz, T. (2006). The igraph software package for complex network research. *InterJournal, Complex Systems*, *1695*(5), 1–9.

Cullen, F. T., Eck, J. E., & Lowenkamp, C. T. (2002). Environmental corrections. A new paradigm for effective probation and parole supervision. *Federal Probation*, *66*(2), 28–37.

Cullen, F. T., & Kulig, T. C. (2018). Evaluating Theories of Environmental Criminology: Strengths and Weaknesses. In G. J. N. Bruinsma & S. D. Johnson (Eds.), *The Oxford Handbook of Environmental Criminology* (pp. 160–176). Oxford University Press. https://doi.org/10.1093/oxfordhb/9780190279707.013.7

Dasu, T., & Johnson, T. (2003). *Exploratory data mining and data cleaning*. Wiley-Interscience.

Davanzo, G., Medvet, E., & Bartoli, A. (2011). Anomaly detection techniques for a web defacement monitoring service. *Expert Systems with Applications*, *38*(10), 12521–12530. https://doi.org/10.1016/j.eswa.2011.04.038

Davey, C. L., Wootton, A. B., & Wootton, A. B. (2017). *Design Against Crime: A Human-Centred Approach to Designing for Safety and Security*. Routledge. https://doi.org/10.4324/9781315576565

Davidson, T., Warmsley, D., Macy, M., & Weber, I. (2017). Automated Hate Speech Detection and the Problem of Offensive Language. *ArXiv:1703.04009 [Cs]*. http://arxiv.org/abs/1703.04009

Décary-Hétu, D., & Morselli, C. (2011). Gang presence in social network sites. *International Journal of Cyber Criminology*, *5*(2), 876–890.

Djuric, N., Zhou, J., Morris, R., Grbovic, M., Radosavljevic, V., & Bhamidipati, N. (2015). Hate Speech Detection with Comment Embeddings. *Proceedings of the 24th International Conference on World Wide Web - WWW '15 Companion*, 29–30. https://doi.org/10.1145/2740908.2742760

dos Santos, C., & Gatti, M. (2014). Deep Convolutional Neural Networks for Sentiment Analysis of Short Texts. *COLING*, 69–78.

e Silva, K. K. (2018). Vigilantism and cooperative criminal justice: Is there a place for cybersecurity vigilantes in cybercrime fighting? *International Review of Law, Computers & Technology*, *32*(1), 21–36. https://doi.org/10.1080/13600869.2018.1418142

Eck, J. E. (1994). *Drug markets and drug places: A case-control study of the spatial structure of illicit drug dealing*. University of Maryland.

Eck, J. E. (2003). Police problems: The complexity of problem theory, research and evaluation. In J. Knutsson (Ed.), *Problem-Oriented Policing: From Innovation to Mainstream* (Vol. 15, pp. 79–113). Criminal Justice Press.

Eck, J. E., & Spelman, W. (1987). *Problem-solving problem-oriented policing in Newport News*. Police Executive Research Forum.

Eck, J. E., & Weisburd, D. (1995). Crime places in crime theory. In J. E. Eck & D. Weisburd (Eds.), *Crime and place* (pp. 1–33). Criminal Justice Press.

Edwards, A. (2017). Big data, predictive machines and security: The minority report. In M. R. McGuire & T. J. Holt (Eds.), *The Routledge handbook of technology, crime and justice* (pp. 451–461). Routledge.

Ekblom, P. (2011). *Crime prevention, security and community safety using the 5Is framework*. Palgrave Macmillan. http://search.ebscohost.com/login.aspx?direct=true&scope=site&db=nlebk&db=nlabk&AN=356832

Emig, M. N., & Heck, R. O. (1980). *Crime Analysis: A Selected Bibliography* (pp. 1–26). US Department of Justice.

ESSA. (2015). *ESSA Q4 2015 Integrity Report* (pp. 1–6). ESSA. http://www.eu-ssa.org/wp-content/uploads/QR4-BROCHURE-WEB.pdf

ESSA. (2016). *ESSA 2016 Annual Integrity Report* (pp. 1–6). ESSA. http://www.eu-ssa.org/wp-content/uploads/QR1-BROCHURE-2017-SINGLE.pdf

ESSA. (2017). *ESSA 2017 Annual Integrity Report* (pp. 1–6). ESSA. http://www.eu-ssa.org/wp-content/uploads/ESSA-2017-annual-integrity-report.pdf

ESSA. (2018). *ESSA 2018 Annual Integrity Report* (pp. 1–6). ESSA. http://www.eu-ssa.org/wp-content/uploads/ESSA-2018-Annual-Integrity-Report.pdf

Esteve, M., Miró-Llinares, F., & Rabasa, A. (2018). Classification of tweets with a mixed method based on pragmatic content and meta-information. *International Journal of Design & Nature and Ecodynamics*, *13*(1), 60–70. https://doi.org/10.2495/DNE-V13-N1-60-70

Esteve, Miriam, Moneva, A., & Hart, T. C. (2019). *CACC R package: Conjunctive Analysis of Case Configurations* (Version 1.0.3) [Computer software]. https://doi.org/10.5281/zenodo.3472967

Esteve, Z., Moneva, A., & Miró-Llinares, F. (2019). Can metadata be used to measure the anonymity of Twitter users? Results of a Confirmatory Factor Analysis. *International E-Journal of Criminal Science, 13*, 1–16.

Fagan, A. A., & Mazerolle, P. (2011). Repeat Offending and Repeat Victimization: Assessing Similarities and Differences in Psychosocial Risk Factors. *Crime & Delinquency*, *57*(5), 732–755. https://doi.org/10.1177/0011128708321322

Farías, D. I. H., Patti, V., & Rosso, P. (2016). Irony Detection in Twitter: The Role of Affective Content. *ACM Transactions on Internet Technology*, *16*(3), 1–24. https://doi.org/10.1145/2930663

Farrell, G. (2005). Progress and prospects in the prevention of repeat victimization. In T. Nick (Ed.), *Handbook of crime prevention and community safety* (pp. 145–172). Willan.

Farrell, G. (2015). Crime concentration theory. *Crime Prevention and Community Safety*, *17*(4), 233–248. https://doi.org/10.1057/cpcs.2015.17

Farrell, G., & Pease, K. (1993). *Once bitten, twice bitten: Repeat victimisation and its implications for crime prevention* (No. 46; Crime Prevention Unit Series, pp. 1–38). Home Office, Police Research Group.

https://webarchive.nationalarchives.gov.uk/20110218140829/http://rds.homeoffice.gov.uk/rds/prgpdfs/fcpu46.pdf

Farrell, G., & Pease, K. (2017). Preventing repeat and near repeat crime concentrations. In N. Tilley & A. Sidebottom (Eds.), *Handbook of Crime Prevention and Community Safety* (2nd ed., p. 626). Routledge. https://doi.org/10.4324/9781315724393

Farrell, G., & Pease, K. (2018). Repeat Victimization. In G. J. N. Bruinsma & D. Weisburd (Eds.), *Encyclopedia of Criminology and Criminal Justice* (pp. 4371–4381). Springer New York. https://doi.org/10.1007/978-1-4614-5690-2_128

Farrell, G., Tseloni, A., Mailley, J., & Tilley, N. (2011). The Crime Drop and the Security Hypothesis. *Journal of Research in Crime and Delinquency*, *48*(2), 147–175. https://doi.org/10.1177/0022427810391539

Farrington, D. P., Loeber, R., Elliott, D. S., Hawkins, J. D., Kandel, D. B., Klein, M. W., McCord, J., Rowe, D. C., & Tremblay, R. E. (1990). Advancing Knowledge about the Onset of Delinquency and Crime. In B. B. Lahey & A. E. Kazdin (Eds.), *Advances in Clinical Child Psychology* (pp. 283–342). Springer US. https://doi.org/10.1007/978-1-4613-9835-6_8

Farrington, D. P., & Wikstrom, P.-O. H. (1994). Criminal Careers in London and Stockholm: A Cross-National Comparative Study. In E. G. M. Weitekamp & H.-J. Kerner (Eds.), *Cross-National Longitudinal Research on Human Development and Criminal Behavior* (pp. 65–89). Springer Netherlands. https://doi.org/10.1007/978-94-011-0864-5_2

Felson, M. (1995). Those Who Discourage Crime. In J. E. Eck & D. Weisburd (Eds.), *Crime and place* (pp. 53–66). Criminal Justice Press.

Felson, M. (2015). Lectio Doctoralis. El estudio científico del delito. In F. Miró-Llinares, J. R. Agustina-Sanllehí, J. E. Medina-Sarmiento, & L. Summers (Eds.), *Crimen, oportunidad y vida diaria* (Dykinson, pp. 23–28).

Felson, M., & Clarke, R. V. (1998). *Opportunity makes the thief: Practical theory for crime prevention* (No. 98; Police Research Series, pp. 1–44). Home Office. https://popcenter.asu.edu/sites/default/files/opportunity_makes_the_thief.pdf

Felson, M., & Eckert, M. (2019). *Crime and everyday life: A brief introduction* (Sixth Edition). SAGE Publications.

Felson, M., & Santos, R. B. (2010). *Crime and everyday life* (4th ed). SAGE Publications.

Ferrara, E., Wang, W.-Q., Varol, O., Flammini, A., & Galstyan, A. (2016). Predicting Online Extremism, Content Adopters, and Interaction Reciprocity. In E. Spiro & Y.-Y. Ahn (Eds.), *Social Informatics* (Vol. 10047, pp. 22–39). Springer International Publishing. https://doi.org/10.1007/978-3-319-47874-6_3

Finn, J. (2004). A Survey of Online Harassment at a University Campus. *Journal of Interpersonal Violence*, *19*(4), 468–483. https://doi.org/10.1177/0886260503262083

Forrest, D. (2012). The Threat to Football from Betting-Related Corruption. *International Journal of Sport Finance*, *7*(2), 99–116.

Fox, J. A., & Tracy, P. E. (1988). A measure of skewness in offense distributions. *Journal of Quantitative Criminology*, *4*(3), 259–274. https://doi.org/10.1007/BF01072453

Gerstenfeld, P. B., Grant, D. R., & Chiang, C.-P. (2003). Hate Online: A Content Analysis of Extremist Internet Sites. *Analyses of Social Issues and Public Policy*, *3*(1), 29–44. https://doi.org/10.1111/j.1530-2415.2003.00013.x

Goldstein, H. (1979). Improving Policing: A Problem-Oriented Approach. *Crime & Delinquency*, *25*(2), 236–258. https://doi.org/10.1177/001112877902500207

Grabosky, P. N. (2001). Virtual Criminality: Old Wine in New Bottles? *Social & Legal Studies*, *10*(2), 243–249. https://doi.org/10.1177/a017405

Grabosky, P. N., & Smith, R. G. (2001). Telecommunication fraud in the digital age: The convergence of technologies. In D. S. Wall (Ed.), *Crime and the Internet* (pp. 29–43). Routledge.

Grabosky, P. N., Smith, R. G., & Dempsey, G. (2001). *Electronic theft: Unlawful acquisition in cyberspace*. Cambridge University Press.

Grinberg, N., Joseph, K., Friedland, L., Swire-Thompson, B., & Lazer, D. (2019). Fake news on Twitter during the 2016 U.S. presidential election. *Science*, *363*(6425), 374–378. https://doi.org/10.1126/science.aau2706

Grolemund, G., & Wickham, H. (2016). *R for Data Science*. O'Reilly Media, Inc. https://proquest.safaribooksonline.com/9781491910382

Haberfeld, M. R., & Sheehan, D. (Eds.). (2013). *Match-Fixing in International Sports*. Springer International Publishing. https://doi.org/10.1007/978-3-319-02582-7

Hargittai, E., & Hinnant, A. (2008). Digital Inequality: Differences in Young Adults' Use of the Internet. *Communication Research*, *35*(5), 602–621. https://doi.org/10.1177/0093650208321782

Hart, T. C. (2014). Conjunctive Analysis of Case Configurations. *JDiBrief Series*. http://www.ucl.ac.uk/jdibrief/analysis/CACC

Hart, T. C. (2019). Identifying Situational Clustering and Quantifying Its Magnitude in Dominant Case Configurations: New Methods for Conjunctive Analysis. *Crime & Delinquency*, *66*(1), 143–159. https://doi.org/10.1177/0011128719866123

Hart, T. C., & Moneva, A. (2018). Conjunctive Analysis of Case Configurations: An introduction to Configural Thinking. *Revista Española de Investigación Criminológica*, *16*, 1–19.

Hart, T. C., Rennison, C. M., & Miethe, T. D. (2017). Identifying Patterns of Situational Clustering and Contextual Variability in Criminological Data: An Overview of Conjunctive Analysis of Case Configurations. *Journal of Contemporary Criminal Justice*, *33*(2), 112–120. https://doi.org/10.1177/1043986216689746

Hartel, P., Junger, M., & Wieringa, R. (2011). *Cyber-crime Science = Crime Science + Information Security*.

Hawley, A. H. (1950). *Human Ecology: A Theory of Community Structure*. Ronald Press.

Healy, K. (2018). *Data visualization: A practical introduction*. Princeton University Press.

Henson, B. (2010). Cyberstalking. In B. Fisher & S. P. Lab (Eds.), *Encyclopedia of victimology and crime prevention* (pp. 253–256). SAGE Publications.

Henson, B., Reyns, B. W., & Fisher, B. S. (2011). Security in the 21st Century: Examining the Link Between Online Social Network Activity, Privacy, and Interpersonal Victimization. *Criminal Justice Review*, *36*(3), 253–268. https://doi.org/10.1177/0734016811399421

Hernández, O. J., Ferri, R. C., & Ramírez, Q. M. J. (2004). *Introducción a la minería de datos*. Pearson Educación S.A.

Higgins, G. E. (2007). Digital Piracy, Self-Control Theory, and Rational Choice: An Examination of the Role of Value. *International Journal of Cyber Criminology*, *1*(1), 33–55.

Hindelang, M. J., Gottfredson, M. R., & Garofalo, J. (1978). *Victims of personal crime: An empirical foundation for a theory of personal victimization*. Ballinger Pub. Co.

Hinduja, S., & Kooi, B. (2013). Curtailing cyber and information security vulnerabilities through situational crime prevention. *Security Journal*, *26*(4), 383–402. https://doi.org/10.1057/sj.2013.25

Hinduja, S., & Patchin, J. W. (2008). Cyberbullying: An Exploratory Analysis of Factors Related to Offending and Victimization. *Deviant Behavior*, *29*(2), 129–156. https://doi.org/10.1080/01639620701457816

Hinduja, S., & Patchin, J. W. (2015). *Bullying beyond the schoolyard: Preventing and responding to cyberbullying* (Second edition). Corwin.

Hirschi, T., & Gottfredson, M. R. (1986). The distinction between crime and criminality. In T. F. Hartnagel, R. A. Silverman, & G. Nettler (Eds.), *Critique and explanation: Essays in honor of Gwynne Nettler* (pp. 44–69). Transaction Books.

Hollinger, R. C., & Lanza-Kaduce, L. (1988). The Process of Criminalization: The case of Computer Crime laws. *Criminology*, *26*(1), 101–126. https://doi.org/10.1111/j.1745-9125.1988.tb00834.x

Hollis, M. E., Felson, M., & Welsh, B. C. (2013). The capable guardian in routine activities theory: A theoretical and conceptual reappraisal. *Crime Prevention and Community Safety*, *15*(1), 65–79. https://doi.org/10.1057/cpcs.2012.14

Holt, T. J. (2011). The attack dynamics of political and religiously motivated hackers. In T. Saadawi & L. Jordan Jr. (Eds.), *Cyber infrastruture protection* (pp. 159–180). Strategic Studies Institute.

Holt, T. J. (2019). Computer Hacking and the Hacker Subculture. In *The Palgrave Handbook of International Cybercrime and Cyberdeviance* (pp. 1–18). Springer International Publishing. https://doi.org/10.1007/978-3-319-90307-1_31-1

Holt, T. J., & Bossler, A. M. (2008). Examining the Applicability of Lifestyle-Routine Activities Theory for Cybercrime Victimization. *Deviant Behavior*, *30*(1), 1–25. https://doi.org/10.1080/01639620701876577

Holt, T. J., & Bossler, A. M. (2014). An Assessment of the Current State of Cybercrime Scholarship. *Deviant Behavior*, *35*(1), 20–40. https://doi.org/10.1080/01639625.2013.822209

Holt, T. J., & Bossler, A. M. (2016). *Cybercrime in progress: Theory and prevention of technology-enabled offenses* (First Edition). Routledge.

Holt, T. J., & Bossler, A. M. (2017). *Cybercrime in progress: Theory and prevention of technology-enabled offenses*. Routledge, Taylor & Francis Group.

Holt, T. J., Leukfeldt, R., & Van De Weijer, S. (2020). An Examination of Motivation and Routine Activity Theory to Account for Cyberattacks Against Dutch Web Sites. *Criminal Justice and Behavior*, *47*(4), 487–505. https://doi.org/10.1177/0093854819900322

Holt, T. J., van Wilsem, J., van de Weijer, S. G. A., & Leukfeldt, E. R. (2020). Testing an Integrated Self-Control and Routine Activities Framework to Examine Malware Infection Victimization. *Social Science Computer Review*, *38*(2), 187–206. https://doi.org/10.1177/0894439318805067

Hosseinmardi, H., Mattson, S. A., Rafiq, R. I., Han, R., Lv, Q., & Mishra, S. (2015). Detection of Cyberbullying Incidents on the Instagram Social Network. *ArXiv:1503.03909 [Cs]*. http://arxiv.org/abs/1503.03909

Howell, C. J., Burruss, G. W., Maimon, D., & Sahani, S. (2019). Website defacement and routine activities: Considering the importance of hackers' valuations of potential targets. *Journal of Crime and Justice*, 1–15. https://doi.org/10.1080/0735648X.2019.1691859

Hsia, J. (2017). Twitter Trouble: The Communications Decency Act in Action. *Columbia Business Law Review*, *2017*(1), 399–452.

Huggins, M. (2018). Match-Fixing: A Historical Perspective. *The International Journal of the History of Sport*, *35*(2–3), 123–140. https://doi.org/10.1080/09523367.2018.1476341

Hutchings, A., & Clayton, R. (2016). Exploring the Provision of Online Booter Services. *Deviant Behavior*, *37*(10), 1163–1178. https://doi.org/10.1080/01639625.2016.1169829

Hutchings, A., & Hayes, H. (2008). Routine activity theory and phishing victimisation: Who gets caught in the net. *Current Issues in Criminal Justice*, *20*, 432–451.

Hutchings, A., & Holt, T. J. (2015). A Crime Script Analysis of the Online Stolen Data Market. *British Journal of Criminology*, *55*(3), 596–614. https://doi.org/10.1093/bjc/azu106

Hutchings, A., & Holt, T. J. (2017). The online stolen data market: Disruption and intervention approaches. *Global Crime*, *18*(1), 11–30. https://doi.org/10.1080/17440572.2016.1197123

Iannone, R. (2020). *DiagrammeR: Graph/Network Visualization* (Version 1.0.5) [Computer software]. https://CRAN.R-project.org/package=DiagrammeR

INE. (2018). *Cifras de población y censos demográficos*. Instituto Nacional de Estadística.

https://www.ine.es/dyngs/INEbase/es/categoria.htm?c=Estadistica_P&cid=1254735572981

Internet Crime Complaint Center. (2018). *2018 Internet Crime Report* (pp. 1–28). Federal Bureau of Investigation. https://pdf.ic3.gov/2018_IC3Report.pdf

Jacobs, J. (1961). *The death and life of great American cities*. Random House.

Jeffery, C. R. (1971). *Crime Prevention Through Environmental Design*. SAGE Publications.

Johnson, S. D. (2008a). Repeat burglary victimisation: A tale of two theories. *Journal of Experimental Criminology*, *4*(3), 215–240. https://doi.org/10.1007/s11292-008-9055-3

Johnson, S. D. (2008b). Repeat burglary victimisation: A tale of two theories. *Journal of Experimental Criminology*, *4*(3), 215–240. https://doi.org/10.1007/s11292-008-9055-3

Johnson, S. D., & Bowers, K. J. (2004). The Burglary as Clue to the Future: The Beginnings of Prospective Hot-Spotting. *European Journal of Criminology*, *1*(2), 237–255. https://doi.org/10.1177/1477370804041252

Johnson, S. D., Bowers, K. J., & Hirschfield, A. (1997). New insights into the spatial and temporal distribution of repeat victimization. *British Journal of Criminology*, *37*(2), 224–241. https://doi.org/10.1093/oxfordjournals.bjc.a014156

Johnson, S. D., Summers, L., & Pease, K. (2009). Offender as Forager? A Direct Test of the Boost Account of Victimization. *Journal of Quantitative Criminology*, *25*(2), 181–200. https://doi.org/10.1007/s10940-008-9060-8

Kemp, S., Miró-Llinares, F., & Moneva, A. (2020). The Dark Figure and the Cyber Fraud Rise in Europe: Evidence from Spain. *European Journal on Criminal Policy and Research*. https://doi.org/10.1007/s10610-020-09439-2

Kemp, S., & Moneva, A. (2020). Comparing correlates of offline and online fraud victimisation and impacts. *InDret*, *20*(1), 424–444.

Kennedy, D. M. (2012). *Deterrence and Crime Prevention: Reconsidering the Prospect of Sanction* (1st ed.). Routledge. https://doi.org/10.4324/9780203892022

Khalil, S., & Fakir, M. (2017). RCrawler: An R package for parallel web crawling and scraping. *SoftwareX*, *6*, 98–106. https://doi.org/10.1016/j.softx.2017.04.004

Khurana, A., Bleakley, A., Jordan, A. B., & Romer, D. (2015). The Protective Effects of Parental Monitoring and Internet Restriction on Adolescents' Risk of Online Harassment. *Journal of Youth and Adolescence*, *44*(5), 1039–1047. https://doi.org/10.1007/s10964-014-0242-4

Kigerl, A. (2012). Routine Activity Theory and the Determinants of High Cybercrime Countries. *Social Science Computer Review*, *30*(4), 470–486. https://doi.org/10.1177/0894439311422689

Krone, T., Spiranovic, C., Prichard, J., Watters, P., Wortley, R., Gelb, K., & Hunn, C. (2020). Child sexual abuse material in child-centred institutions: Situational crime prevention approaches. *Journal of Sexual Aggression*, 1–20. https://doi.org/10.1080/13552600.2019.1705925

Lammers, M., Menting, B., Ruiter, S., & Bernasco, W. (2015). Biting once, twice: The influence of prior on subsequent crime location choice. *Criminology*, *53*(3), 309–329. https://doi.org/10.1111/1745-9125.12071

Landis, J. R., & Koch, G. G. (1977). The Measurement of Observer Agreement for Categorical Data. *Biometrics*, *33*(1), 159. https://doi.org/10.2307/2529310

Lara-Cabrera, R., Gonzalez-Pardo, A., Barhamgi, M., & Camacho, D. (2017). Extracting Radicalisation Behavioural Patterns from Social Network Data. *2017 28th International Workshop on Database and Expert Systems Applications (DEXA)*, 6–10. https://doi.org/10.1109/DEXA.2017.18

Lersch, K. M., & Hart, T. C. (2015). *Space, time, and crime* (Fourth edition). Carolina Academic Press.

Leukfeldt, E. R. (2014). Phishing for Suitable Targets in The Netherlands: Routine Activity Theory and Phishing Victimization. *Cyberpsychology, Behavior, and Social Networking*, *17*(8), 551–555. https://doi.org/10.1089/cyber.2014.0008

Leukfeldt, E. R. (Ed.). (2017). *The human factor in cybercrime and cybersecurity: Research agenda*. Eleven International Publishing. https://www.elevenpub.com/criminology/catalogus/research-agenda-the-human-factor-in-cybercrime-and-cybersecurity-1

Leukfeldt, E. R., & Holt, T. J. (Eds.). (2020). *The human factor of cybercrime*. Routledge.

Leukfeldt, E. R., & Jansen, J. (2020). Financial cybercrimes and situational crime prevention. In E. R. Leukfeldt & T. J. Holt (Eds.), *The Human Factor of Cybercrime* (pp. 216–239). Rutledge.

Leukfeldt, E. R., Kleemans, E. R., & Stol, W. P. (2017a). Cybercriminal Networks, Social Ties and Online Forums: Social Ties Versus Digital Ties within Phishing and Malware Networks. *British Journal of Criminology*, *57*(3), 704–722. https://doi.org/10.1093/bjc/azw009

Leukfeldt, E. R., Kleemans, E. R., & Stol, W. P. (2017b). A typology of cybercriminal networks: From low-tech all-rounders to high-tech specialists. *Crime, Law and Social Change*, *67*(1), 21–37. https://doi.org/10.1007/s10611-016-9662-2

Leukfeldt, E. R., Kleemans, E. R., & Stol, W. P. (2017c). The Use of Online Crime Markets by Cybercriminal Networks: A View From Within. *American Behavioral Scientist*, *61*(11), 1387–1402. https://doi.org/10.1177/0002764217734267

Leukfeldt, E. R., & Yar, M. (2016). Applying Routine Activity Theory to Cybercrime: A Theoretical and Empirical Analysis. *Deviant Behavior*, *37*(3), 263–280. https://doi.org/10.1080/01639625.2015.1012409

Lévesque, F. L., Fernández, J. M., Young, G., & Batchelder, D. (2011). *Are They Real? Real-Life Comparative Tests of Anti-Virus Products*. 1–11.

Levin, B. (2002). Cyberhate: A Legal and Historical Analysis of Extremists' Use of Computer Networks in America. *American Behavioral Scientist*, *45*(6), 958–988. https://doi.org/10.1177/0002764202045006004

Li, L., Goodchild, M. F., & Xu, B. (2013). Spatial, temporal, and socioeconomic patterns in the use of Twitter and Flickr. *Cartography and Geographic Information Science*, *40*(2), 61–77. https://doi.org/10.1080/15230406.2013.777139

Lynch, J. (2018). Not even our own facts: Criminology in the era of Big Data. *Criminology*, *56*(3), 437–454. https://doi.org/10.1111/1745-9125.12182

Madarie, R. (2017). Hackers' Motivations: Testing Schwartz'S Theory Of Motivational Types Of Values In A Sample Of Hackers. *International Journal of Cyber Criminology*, *11*(1), 1–20. https://doi.org/10.5281/zenodo.495773

Magdy, W., Darwish, K., & Abokhodair, N. (2015). Quantifying Public Response towards Islam on Twitter after Paris Attacks. *ArXiv:1512.04570 [Cs]*. http://arxiv.org/abs/1512.04570

Maimon, D., Alper, M., Sobesto, B., & Cukier, M. (2014). Restrictive deterrent effects of a warning banner in an attacked computer system. *Criminology*, *52*(1), 33–59. https://doi.org/10.1111/1745-9125.12028

Maimon, D., Fukuda, A., Hinton, S., Babko-Malaya, O., & Cathey, R. (2017). On the relevance of social media platforms in predicting the volume and patterns of web defacement attacks. *2017 IEEE International Conference on Big Data (Big Data)*, 4668–4673. https://doi.org/10.1109/BigData.2017.8258513

Maimon, D., Kamerdze, A., Cukier, M., & Sobesto, B. (2013). Daily Trends and Origin of Computer-Focused Crimes Against a Large University Computer Network: An Application of the Routine-Activities and Lifestyle Perspective. *British Journal of Criminology*, *53*(2), 319–343. https://doi.org/10.1093/bjc/azs067

Maimon, D., & Louderback, E. R. (2019). Cyber-Dependent Crimes: An Interdisciplinary Review. *Annual Review of Criminology*, *2*(1), 191–216. https://doi.org/10.1146/annurev-criminol-032317-092057

Maimon, D., & Testa, A. (2017). On the Relevance of Cyber Criminological Research in the Design of Policies and Sophisticated Security Solutions against Cyberterrorism Events. In G. LaFree & J. D. Freilich (Eds.), *The Handbook of the Criminology of Terrorism* (pp. 553–567). John Wiley & Sons, Inc. https://doi.org/10.1002/9781118923986.ch36

Maimon, D., Wilson, T., Ren, W., & Berenblum, T. (2015). On the Relevance of Spatial and Temporal Dimensions in Assessing Computer Susceptibility to System Trespassing Incidents. *British Journal of Criminology, 55*(3), 615–634. https://doi.org/10.1093/bjc/azu104

Malm, A., & Bichler, G. (2011). Networks of Collaborating Criminals: Assessing the Structural Vulnerability of Drug Markets. *Journal of Research in Crime and Delinquency*, *48*(2), 271–297. https://doi.org/10.1177/0022427810391535

Malm, A., Bichler, G., & Van De Walle, S. (2010). Comparing the ties that bind criminal networks: Is blood thicker than water? *Security Journal*, *23*(1), 52–74. https://doi.org/10.1057/sj.2009.18

Malmasi, S., & Zampieri, M. (2017). Detecting Hate Speech in Social Media. *ArXiv:1712.06427 [Cs]*. http://arxiv.org/abs/1712.06427

Marcum, C. D., Higgins, G. E., Freiburger, T. L., & Ricketts, M. L. (2011). Battle of the sexes: An examination of male and female cyber bullying. *International Journal of Cyber Criminology*, *6*(1), 904–911.

Marcum, C. D., Higgins, G. E., & Ricketts, M. L. (2010). Potential Factors of Online Victimization of Youth: An Examination of Adolescent Online Behaviors Utilizing Routine Activity Theory. *Deviant Behavior*, *31*(5), 381–410. https://doi.org/10.1080/01639620903004903

Marcum, C. D., Ricketts, M. L., & Higgins, G. E. (2010). Assessing Sex Experiences of Online Victimization: An Examination of Adolescent Online Behaviors Using Routine Activity Theory. *Criminal Justice Review*, *35*(4), 412–437. https://doi.org/10.1177/0734016809360331

Mariconti, E., Suarez-Tangil, G., Blackburn, J., De Cristofaro, E., Kourtellis, N., Leontiadis, I., Serrano, J. L., & Stringhini, G. (2018). 'You Know What to Do': Proactive Detection of YouTube Videos Targeted by Coordinated Hate Attacks. *ArXiv:1805.08168 [Cs]*. http://arxiv.org/abs/1805.08168

Martinson, R. (1974). What works? Questions and answers about prison reform. *The Public Interest*, *35*, 22–54.

Marwick, A. E., & Boyd, D. (2011). I tweet honestly, I tweet passionately: Twitter users, context collapse, and the imagined audience. *New Media & Society*, *13*(1), 114–133. https://doi.org/10.1177/1461444810365313

Mastrofski, S. D., Parks, R. B., & McCluskey, J. D. (2010). Systematic Social Observation in Criminology. In A. R. Piquero & D. Weisburd (Eds.), *Handbook of Quantitative Criminology* (pp. 225–247). Springer New York. https://doi.org/10.1007/978-0-387-77650-7_12

Mastrofski, S. D., Parks, R. B., Reiss, A. J., Worden, R. E., De Jong, C., Snipes, J. B., & Terrill, W. (1998). *Systematic social observation of public police: Applying field research methods to policy issues.* [Research Report]. National Institute of Justice. https://www.ncjrs.gov/pdffiles/172859.pdf

McGloin, J. M. (2005). Policy and intervention considerations of a network analysis of street gangs. *Criminology & Public Policy*, *4*(3), 607–635. https://doi.org/10.1111/j.1745-9133.2005.00306.x

McGuire, M., & Dowling, S. (2013a). *Cyber-dependent crimes* (No. 75; Cyber Crime: A Review of the Evidence, pp. 1–35). Home Office. https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/246751/horr75-chap1.pdf

McGuire, M., & Dowling, S. (2013b). *Cyber-enabled crimes—Fraud and theft* (No. 75; Cyber Crime: A Review of the Evidence, pp. 1–27). Home Office. https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/248621/horr75-chap2.pdf

McGuire, M., & Dowling, S. (2013c). *Cyber-enabled crimes—Sexual offending against children* (No. 75; Cyber Crime: A Review of the Evidence, pp. 1–26). Home Office.

https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/246754/horr75-chap3.pdf

McGuire, M. R. (2017). Technology crime and technology control: Contexts and history. In M. R. McGuire & T. J. Holt (Eds.), *The Routledge handbook of technology, crime and justice* (pp. 35–60). Routledge.

McGuire, M. R. (2020). It ain’t what it is, it’s the way that they do it? Why we still don’t understand cybercrime. In E. R. Leukfeldt & T. J. Holt (Eds.), *The Human Factor of Cybercrime* (pp. 3–28). Rutledge.

Miethe, T. D., Hart, T. C., & Regoeczi, W. C. (2008). The Conjunctive Analysis of Case Configurations: An Exploratory Method for Discrete Multivariate Analyses of Crime Data. *Journal of Quantitative Criminology*, *24*(2), 227–241. https://doi.org/10.1007/s10940-008-9044-8

Miró Llinares, F. (Ed.). (2017). *Cometer delitos en 140 caracteres: El derecho penal ante el odio y la radicalización en Internet*. Marcial Pons.

Miró-Llinares, F. (2011). La oportunidad criminal en el ciberespacio. Aplicación y desarrollo de la teoría de las actividades cotidianas para la prevención del cibercrimen. *Revista Electrónica de Ciencia Penal y Criminología*, *13*(7), 1–55.

Miró-Llinares, F. (2012). *El cibercrimen. Fenomenología y criminología de la delincuencia en el ciberespacio*. Marcial Pons.

Miró-Llinares, F. (2015a). Cibercrimen y vida diaria en el mundo 2.0. In F. Miró-Llinares, J. R. Agustina-Sanllehí, J. E. Medina-Sarmiento, & L. Summers (Eds.), *Crimen, oportunidad y vida diaria* (Dykinson, pp. 415–456).

Miró-Llinares, F. (2015b). That Cyber Routine, That Cyber Victimization: Profiling Victims of Cybercrime. In R. G. Smith, R. C.-C. Cheung, & L. Y.-C. Lau (Eds.), *Cybercrime Risks and Responses* (pp. 47–63). Palgrave Macmillan UK. https://doi.org/10.1057/9781137474162_4

Miró-Llinares, F. (2016). Taxonomía de la comunicación violenta y el discurso del odio en Internet. *IDP: Revista de Internet, Derecho y Política*, *22*, 82–107.

Miró-Llinares, F. (2018). La detección de discurso radical en Internet. Aproximación, encuadre y propuesta de mejora de los análisis de Big Data desde un enfoque de Smart Data criminológico. In A. Alonso Rimo, M. L. Cuerda Arnau, & A. Fernández Hernández (Eds.), *Terrorismo, sistema penal y derechos fundamentales* (pp. 617–648). Tirant lo Blanch.

Miró-Llinares, F., Drew, J., & Townsley, M. K. (2020). Understanding Target Suitability in Cyberspace: An International Comparison of Cyber Victimization Processes. *International Journal of Cyber Criminology*, *14*(1), 139–155.

Miró-Llinares, F., & Johnson, S. D. (2018). Cybercrime and Place: Applying Environmental Criminology to Crimes in Cyberspace. In G. J. N. Bruinsma & S. D. Johnson (Eds.), *The Oxford Handbook of Environmental Criminology* (pp. 883–906). Oxford University Press. https://doi.org/10.1093/oxfordhb/9780190279707.013.39

Miró-Llinares, F., & Moneva, A. (2019a). Environmental Criminology and Cybercrime: Shifting Focus from the Wine to the Bottles. In T. J. Holt & A. M. Bossler (Eds.), *The Palgrave Handbook of International Cybercrime and Cyberdeviance* (pp. 1–22). Springer International Publishing. https://doi.org/10.1007/978-3-319-90307-1_30-1

Miró-Llinares, F., & Moneva, A. (2019b). What about cyberspace (and cybercrime alongside it)? A reply to Farrell and Birks "Did cybercrime cause the crime drop?" *Crime Science*, *8*(1), 12. https://doi.org/10.1186/s40163-019-0107-y

Miró-Llinares, F., Moneva, A., & Esteve, M. (2018). Hate is in the air! But where? Introducing an algorithm to detect hate speech in digital microenvironments. *Crime Science*, *7*(15), 1–12. https://doi.org/10.1186/s40163-018-0089-1

Miró-Llinares, F., & Rodriguez-Sala, J. J. (2016). Cyber hate speech on twitter: Analyzing disruptive events from social media to build a violent communication and hate speech taxonomy. *International Journal of Design & Nature and Ecodynamics*, *11*(3), 406–415. https://doi.org/10.2495/DNE-V11-N3-406-415

Mittelstadt, B. D., Allo, P., Taddeo, M., Wachter, S., & Floridi, L. (2016). The ethics of algorithms: Mapping the debate. *Big Data & Society*, *3*(2), 205395171667967. https://doi.org/10.1177/2053951716679679

Moffat, R. E. (1983). Crime Prevention Through Environmental Design—A Management Perspective. *Canadian Journal of Criminology*, *25*(4), 19–31.

Moffitt, T. E., Caspi, A., Rutter, M., & Silva, P., A. (2001). *Sex differences in antisocial behaviour: Conduct disorder, delinquency, and violence in the Dunedin longitudinal study*. Cambridge University Press.

Moitra, S. D., & Konda, S. L. (2004). An empirical investigation of network attacks on computer systems. *Computers & Security*, *23*(1), 43–51. https://doi.org/10.1016/S0167-4048(04)00067-7

Moneva, A., & Caneppele, S. (2019). 100% sure bets? Exploring the precipitation-control strategies of fixed-match informing websites and the environmental features of their networks. *Crime, Law and Social Change*, 1–19. https://doi.org/10.1007/s10611-019-09871-4

Moneva, A., Miró-Llinares, F., & Hart, T. C. (2020). Hunter or Prey? Exploring the Situational Profiles that Define Repeated Online Harassment Victims and

Offenders. *Deviant Behavior*, 1–16. https://doi.org/10.1080/01639625.2020.1746135

Monk, B., Mitchell, J., Frank, R., & Davies, G. (2018). Uncovering Tor: An Examination of the Network Structure. *Security and Communication Networks*, *2018*, 1–12. https://doi.org/10.1155/2018/4231326

Moriconi, M., & Almeida, J. P. (2019). Portuguese Fight Against Match-Fixing: Which Policies and What Ethic? *Journal of Global Sport Management*, *4*(1), 79–96. https://doi.org/10.1080/24704067.2018.1493357

Morselli, C., & Décary-Hétu, D. (2013). Crime facilitation purposes of social networking sites: A review and analysis of the 'cyberbanging' phenomenon. *Small Wars & Insurgencies*, *24*(1), 152–170. https://doi.org/10.1080/09592318.2013.740232

Müller, K., & Wickham, H. (2019). *tibble: Simple Data Frames* (Version 2.1.3) [Computer software]. https://CRAN.R-project.org/package=tibble

Näsi, M., Räsänen, P., Kaakinen, M., Keipi, T., & Oksanen, A. (2017). Do routine activities help predict young adults' online harassment: A multi-nation study. *Criminology & Criminal Justice*, *17*(4), 418–432. https://doi.org/10.1177/1748895816679866

Navarro, J. N., & Jasinski, J. L. (2013). Why Girls? Using Routine Activities Theory to Predict Cyberbullying Experiences Between Girls and Boys. *Women & Criminal Justice*, *23*(4), 286–303. https://doi.org/10.1080/08974454.2013.784225

Nazaretian, Z., & Merolla, D. M. (2013). Questioning Canadian Criminal Incidence Rates: A Re-analysis of the 2004 Canadian Victimization Survey. *Canadian Journal of Criminology and Criminal Justice*, *55*(2), 239–261. https://doi.org/10.3138/cjccj.2012.E18

Newman, G. R., & Clarke, R. V. (2003). *Superhighway robbery preventing e-commerce crime*. Routledge Chapman & Hall.

Newman, O. (1972). *Defensible space: Crime prevention through urban design*. Macmillan.

Ngo, F., T., & Paternoster, R. (2011). Cybercrime Victimization: An examination of Individual and Situational level factors. *International Journal of Cyber Criminology*, *5*(1), 773–793.

Nobata, C., Tetreault, J., Thomas, A., Mehdad, Y., & Chang, Y. (2016). Abusive Language Detection in Online User Content. *Proceedings of the 25th International Conference on World Wide Web - WWW '16*, 145–153. https://doi.org/10.1145/2872427.2883062

Novo, F., Pereira, F., & Matos, M. (2014). Cyber-Aggression among Portuguese Adolescents: A Study on Perpetration, Victim Offender Overlap and Parental Supervision. *International Journal of Cyber Criminology, 8*(2), 94–110.

Painter, K., & Tilley, N. (Eds.). (1999). *Surveillance of public space: CCTV, street lighting and crime prevention*. Criminal Justice Press.

Park, H. W. (2003). Hyperlink network analysis: A new method for the study of social structure on the web. *Connections*, *25*(1), 49–61.

Park, H. W., & Thelwall, M. (2006). Hyperlink Analyses of the World Wide Web: A Review. *Journal of Computer-Mediated Communication*, *8*(4). https://doi.org/10.1111/j.1083-6101.2003.tb00223.x

Park, S. min, & Eck, J. E. (2013). Understanding the Random Effect on Victimization Distributions: A Statistical Analysis of Random Repeat Victimizations. *Victims & Offenders*, *8*(4), 399–415. https://doi.org/10.1080/15564886.2013.814612

Parker, D. B. (1976). *Crime by computer*. Scribner.

Patchin, J. W., & Hinduja, S. (2015). Measuring cyberbullying: Implications for research. *Aggression and Violent Behavior, 23,* 69–74. https://doi.org/10.1016/j.avb.2015.05.013

Payne, B. K. (2019). Defining Cybercrime. In *The Palgrave Handbook of International Cybercrime and Cyberdeviance* (pp. 1–24). Springer International Publishing. https://doi.org/10.1007/978-3-319-90307-1_1-1

Pease, K. (1998). *Repeat vicimisation: Taking stock* (No. 90; Crime Detection and Prevention Series, pp. 1–40). Home Office, Police Research Group.

Pease, K. (2003). Crime futures and foresight: Challenging criminal behaviour in the information age. In D. S. Wall (Ed.), *Crime and the Internet* (1st ed., pp. 30–40). Routledge. https://doi.org/10.4324/9780203299180

Peddinti, S. T., Ross, K. W., & Cappos, J. (2014). 'On the internet, nobody knows you're a dog': A twitter case study of anonymity in social networks. *Proceedings of the Second Edition of the ACM Conference on Online Social Networks - COSN '14*, 83–94. https://doi.org/10.1145/2660460.2660467

Pedersen, T. L. (2020). *ggraph: An Implementation of Grammar of Graphics for Graphs and Networks* (Version 2.0.2) [Computer software]. https://CRAN.R-project.org/package=ggraph

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M., & Duchesnay, É. (2011). Scikit-learn: Machine Learning in Python. *Journal of Machine Learning Research, 12,* 2825–2830.

Petrescu, M., Gironda, J. T., & Korgaonkar, P. K. (2018). Online piracy in the context of routine activities and subjective norms. *Journal of Marketing Management*, *34*(3–4), 314–346. https://doi.org/10.1080/0267257X.2018.1452278

Pineau, T., Schopfer, A., Grossrieder, L., Broséus, J., Esseiva, P., & Rossy, Q. (2016). The study of doping market: How to produce intelligence from Internet forums. *Forensic Science International*, *268*, 103–115. https://doi.org/10.1016/j.forsciint.2016.09.017

Pittaro, M., L. (2007). Cyber stalking: An Analysis of Online Harassment and Intimidation. *International Journal of Cyber Criminology*, *1*(2), 180–197.

Planty, M., & Strom, K. J. (2007). Understanding the Role of Repeat Victims in the Production of Annual US Victimization Rates. *Journal of Quantitative Criminology*, *23*(3), 179–200. https://doi.org/10.1007/s10940-007-9026-2

Pratt, T. C., Holtfreter, K., & Reisig, M. D. (2010). Routine Online Activity and Internet Fraud Targeting: Extending the Generality of Routine Activity Theory. *Journal of Research in Crime and Delinquency*, *47*(3), 267–296. https://doi.org/10.1177/0022427810365903

Pridemore, W. A., Makel, M. C., & Plucker, J. A. (2018). Replication in Criminology and the Social Sciences. *Annual Review of Criminology*, *1*(1), 19–38. https://doi.org/10.1146/annurev-criminol-032317-091849

Quinlan, J. R. (1986). Induction of decision trees. *Machine Learning*, *1*(1), 81–106. https://doi.org/10.1007/BF00116251

R Core Team. (2019). *R: A Language and Environment for Statistical Computing* (Version 3.6.1) [Computer software]. https://www.R-project.org/

Ratcliffe, J. H. (2002). Aoristic Signatures and the Spatio-Temporal Analysis of High Volume Crime Patterns. *Journal of Quantitative Criminology*, *18*(1), 23–43. https://doi.org/10.1023/A:1013240828824

Ratcliffe, J. H. (2016). *Intelligence-led policing* (Second Edition). Routledge.

Raudenbush, S. W., & Sampson, R. J. (1999). Ecometrics: Toward a Science of Assessing Ecological Settings, with Application to the Systematic Social Observation of Neighborhoods. *Sociological Methodology*, *29*(1), 1–41. https://doi.org/10.1111/0081-1750.00059

Reiss, A. J. (1971). Systematic Observation of Natural Social Phenomena. *Sociological Methodology*, *3*, 3. https://doi.org/10.2307/270816

Reppetto, T. A. (1976). Crime Prevention and the Displacement Phenomenon. *Crime & Delinquency*, *22*(2), 166–177. https://doi.org/10.1177/001112877602200204

Reyes, A., Rosso, P., & Veale, T. (2013). A multidimensional approach for detecting irony in Twitter. *Language Resources and Evaluation*, *47*(1), 239–268. https://doi.org/10.1007/s10579-012-9196-x

Reynald, D. M. (2010). Guardians on Guardianship: Factors Affecting the Willingness to Supervise, the Ability to Detect Potential Offenders, and the Willingness to Intervene. *Journal of Research in Crime and Delinquency*, *47*(3), 358–390. https://doi.org/10.1177/0022427810365904

Reynald, D. M., Moir, E., Cook, A., & Vakhitova, Z. (2018). Changing perspectives on guardianship against crime: An examination of the importance of micro-level factors. *Crime Prevention and Community Safety*, *20*(4), 268–283. https://doi.org/10.1057/s41300-018-0049-4

Reyns, B. W. (2010). A situational crime prevention approach to cyberstalking victimization: Preventive tactics for Internet users and online place managers.

Crime Prevention and Community Safety, *12*(2), 99–118. https://doi.org/10.1057/cpcs.2009.22

Reyns, B. W. (2013). Online Routines and Identity Theft Victimization: Further Expanding Routine Activity Theory beyond Direct-Contact Offenses. *Journal of Research in Crime and Delinquency*, *50*(2), 216–238. https://doi.org/10.1177/0022427811425539

Reyns, B. W., & Fissel, E. R. (2019). Cyberstalking. In *The Palgrave Handbook of International Cybercrime and Cyberdeviance* (pp. 1–24). Springer International Publishing. https://doi.org/10.1007/978-3-319-90307-1_57-1

Reyns, B. W., Henson, B., & Fisher, B. S. (2011). Being Pursued Online: Applying Cyberlifestyle–Routine Activities Theory to Cyberstalking Victimization. *Criminal Justice and Behavior*, *38*(11), 1149–1169. https://doi.org/10.1177/0093854811421448

Romagna, M. (2019). Hacktivism: Conceptualization, Techniques, and Historical View. In T. J. Holt & A. M. Bossler (Eds.), *The Palgrave Handbook of International Cybercrime and Cyberdeviance* (pp. 1–27). Springer International Publishing. https://doi.org/10.1007/978-3-319-90307-1_34-1

Romagna, M., & Van den Hout, N. J. (2017). *Hacktivism and website defacement: Motivations, capabilities and potential threats*. 1–11.

Sampson, R. J., & Raudenbush, S. W. (1999). Systematic Social Observation of Public Spaces: A New Look at Disorder in Urban Neighborhoods. *American Journal of Sociology*, *105*(3), 603–651. https://doi.org/10.1086/210356

Schloerke, B., Crowley, J., Cook, D., Briatte, F., Marbach, M., Thoen, E., Elberg, A., & Larmarange, J. (2020). *GGally: Extension to 'ggplot2'* (Version 1.5.0) [Computer software]. https://CRAN.R-project.org/package=GGally

Schmidt, A., & Wiegand, M. (2017). A Survey on Hate Speech Detection using Natural Language Processing. *Proceedings of the Fifth International Workshop on Natural Language Processing for Social Media*, 1–10. https://doi.org/10.18653/v1/W17-1101

Selkie, E. M., Fales, J. L., & Moreno, M. A. (2016). Cyberbullying Prevalence Among US Middle and High School–Aged Adolescents: A Systematic Review and Quality Assessment. *Journal of Adolescent Health*, *58*(2), 125–133. https://doi.org/10.1016/j.jadohealth.2015.09.026

Serrà, J., Leontiadis, I., Spathis, D., Stringhini, G., Blackburn, J., & Vakali, A. (2017). Class-based Prediction Errors to Detect Hate Speech with Out-of-vocabulary Words. *Proceedings of the First Workshop on Abusive Language Online*, 36–40. https://doi.org/10.18653/v1/W17-3005

Sharma, S., Agrawal, S., & Shrivastava, M. (2018). Degree based Classification of Harmful Speech using Twitter Data. *ArXiv:1806.04197 [Cs]*. http://arxiv.org/abs/1806.04197

Sherman, L. W., Gartin, P. R., & Buerger, M. E. (1989). Hot spots of predatory crime: Routine activities and the Criminology of Place. *Criminology, 27*(1), 27–56. https://doi.org/10.1111/j.1745-9125.1989.tb00862.x

Sidebottom, A., & Tilley, N. (2017). Designing systems against crime. In N. Tilley & A. Sidebottom (Eds.), *Handbook of Crime Prevention and Community Safety* (Second, pp. 254–273). Routledge. https://doi.org/10.4324/9781315724393

Sloan, L., Morgan, J., Burnap, P., & Williams, M. (2015). Who Tweets? Deriving the Demographic Characteristics of Age, Occupation and Social Class from Twitter User Meta-Data. *PLOS ONE*, *10*(3), e0115545. https://doi.org/10.1371/journal.pone.0115545

Solymosi, R., & Bowers, K. (2018). The role of innovative data collection methods in advancing criminological understanding. In G. J. N. Bruinsma & S. D. Johnson (Eds.), *The Oxford Handbook of Environmental Criminology* (pp. 210–237). Oxford University Press. https://doi.org/10.1093/oxfordhb/9780190279707.013.38

Solymosi, R., Buil-Gil, D., Vozmediano, L., & Sousa-Guedes, I. (forthcoming). Towards a place-based measure of fear of crime: A systematic review of app-based and crowdsourcing approaches. *Environment and Behavior*.

Stockman, M., Heile, R., & Rein, A. (2015). An Open-Source Honeynet System to Study System Banner Message Effects on Hackers. *Proceedings of the 4th Annual ACM Conference on Research in Information Technology*, 19–22. https://doi.org/10.1145/2808062.2808069

Subrahmanyam, K., Reich, S. M., Waechter, N., & Espinoza, G. (2008). Online and offline social networks: Use of social networking sites by emerging adults. *Journal of Applied Developmental Psychology*, *29*(6), 420–433. https://doi.org/10.1016/j.appdev.2008.07.003

Suh, B., Hong, L., Pirolli, P., & Chi, E. H. (2010). Want to be Retweeted? Large Scale Analytics on Factors Impacting Retweet in Twitter Network. *2010 IEEE Second International Conference on Social Computing*, 177–184. https://doi.org/10.1109/SocialCom.2010.33

Tak, M. (2018). Too big to jail: Match-fixing, institutional failure and the shifting of responsibility. *International Review for the Sociology of Sport*, *53*(7), 788–806. https://doi.org/10.1177/1012690216682950

Testa, A., Maimon, D., Sobesto, B., & Cukier, M. (2017). Illegal Roaming and File Manipulation on Target Computers: Assessing the Effect of Sanction Threats on

System Trespassers' Online Behaviors. *Criminology & Public Policy*, *16*(3), 689–726. https://doi.org/10.1111/1745-9133.12312

Thapngam, T., Yu, S., Zhou, W., & Beliakov, G. (2011). Discriminating DDoS attack traffic from flash crowd through packet arrival patterns. *2011 IEEE Conference on Computer Communications Workshops (INFOCOM WKSHPS)*, 952–957. https://doi.org/10.1109/INFCOMW.2011.5928950

Thelwall, M. (2004). *Link analysis: An information science approach*. Elsevier Academic Press.

Tilley, N., & Laycock, G. (2002). *Working Out What to Do: Evidence-Based Crime Reduction* (No. 11; Crime Reduction Research Series, pp. 1–74). US Department of Justice.

Tsesis, A. (2001). Hate in Cyberspace: Regulating Hate Speech on the Internet. *San Diego Law Review*, *38*, 818–873.

Ttofi, M. M., & Farrington, D. P. (2011). Effectiveness of school-based programs to reduce bullying: A systematic and meta-analytic review. *Journal of Experimental Criminology*, *7*(1), 27–56. https://doi.org/10.1007/s11292-010-9109-1

Tufféry, S. (2011). *Data Mining and Statistics for Decision Making*. John Wiley & Sons, Inc.

Tufte, E. R. (1999). *The visual display of quantitative information* (Seventeenth). Graphics Press.

Turanovic, J. J., Pratt, T. C., & Piquero, A. R. (2018). Structural Constraints, Risky Lifestyles, and Repeat Victimization. *Journal of Quantitative Criminology*, *34*(1), 251–274. https://doi.org/10.1007/s10940-016-9334-5

Vakhitova, Z. I., Alston-Knox, C. L., Reynald, D. M., Townsley, M. K., & Webster, J. L. (2019). Lifestyles and routine activities: Do they enable different types of cyber abuse? *Computers in Human Behavior, 101*, 225–237. https://doi.org/10.1016/j.chb.2019.07.012

Vakhitova, Z. I., Reynald, D. M., & Townsley, M. (2016). Toward the Adaptation of Routine Activity and Lifestyle Exposure Theories to Account for Cyber Abuse Victimization. *Journal of Contemporary Criminal Justice*, *32*(2), 169–188. https://doi.org/10.1177/1043986215621379

van de Weijer, S. G. A., Leukfeldt, R., & Holt, T. J. (submitted). Developmental Trajectories of Defacements: A Longitudinal Study among Hackers in the Netherlands. *Tijdschrift Voor Criminologie*.

van Wilsem, J. (2011). Worlds tied together? Online and non-domestic routine activities and their impact on digital and traditional threat victimization. *European Journal of Criminology*, *8*(2), 115–127. https://doi.org/10.1177/1477370810393156

van Wilsem, J. (2013a). 'Bought it, but Never Got it' Assessing Risk Factors for Online Consumer Fraud Victimization. *European Sociological Review*, *29*(2), 168–178. https://doi.org/10.1093/esr/jcr053

van Wilsem, J. (2013b). Hacking and Harassment—Do They Have Something in Common? Comparing Risk Factors for Online Victimization. *Journal of Contemporary Criminal Justice*, *29*(4), 437–453. https://doi.org/10.1177/1043986213507402

Villeux-Lepage, Y. (2014). *Retweeting the Caliphate: The role of soft-sympathizers in the Islamic State's social media strategy*. 1–14.

Vozmediano, L., & San Juan, C. (2010). *Criminología ambiental: Ecología del delito y de la seguridad*. Editorial UOC.

Wall, D. S. (2001a). *Crime and the Internet* (1st ed.). Routledge. https://doi.org/10.4324/9780203299180

Wall, D. S. (2001b). Cybercrimes and the Internet. In D. S. Wall (Ed.), *Crime and the Internet* (pp. 1–17). Routledge.

Wall, D. S. (2005). The Internet as a Conduit for Criminal Activity. In *Information Technology and the Criminal Justice System* (pp. 77–98). SAGE Publications, Inc. https://doi.org/10.4135/9781452225708.n4

Wall, D. S. (2007). *Cybercrime: The transformation of crime in the information age*. Polity Press.

Waseem, Z., & Hovy, D. (2016). Hateful Symbols or Hateful People? Predictive Features for Hate Speech Detection on Twitter. *Proceedings of the NAACL Student Research Workshop*, 88–93. https://doi.org/10.18653/v1/N16-2013

Wasserman, S., & Faust, K. (1994). *Social network analysis: Methods and applications*. Cambridge University Press.

Weimann, G. (2014). *New terrorism and new media* (Research Series, pp. 1–20). Commons Lab of the Woodrow Wilson International Center for Scholars.

Weisburd, D. (2015). The Law of Crime Concentration and the Criminology of Place. *Criminology*, *53*(2), 133–157. https://doi.org/10.1111/1745-9125.12070

Weisburd, D., Braga, A. A., Groff, E. R., & Wooditch, A. (2017). Can Hot Spots Policing reduce crime in urban areas? An agent-based simulation. *Criminology*, *55*(1), 137–173. https://doi.org/10.1111/1745-9125.12131

Weisburd, D., & Britt, C. (2014). *Statistics in criminal justice* (4. rev. ed). Springer.

Weisburd, D., Eck, J. E., Braga, A. A., Telep, C. W., Cave, B., Bowers, K., Bruinsma, G., Gill, C., Groff, E., Hibdon, J., Hinkle, J. C., Johnson, S. D., Lawton, B., Lum, C., Ratcliffe, J., Rengert, G., Taniguchi, T., & Yang, S.-M. (2016). *Place Matters: Criminology for the Twenty-First Century*. Cambridge University Press. https://doi.org/10.1017/CBO9781139342087

Weisburd, D., Telep, C. W., Hinkle, J. C., & Eck, J. E. (2010). Is problem-oriented policing effective in reducing crime and disorder? *Criminology & Public Policy*, *9*(1), 139–172. https://doi.org/10.1111/j.1745-9133.2010.00617.x

Welsh, B. C., & Farrington, D. P. (2009). Public Area CCTV and Crime Prevention: An Updated Systematic Review and Meta-Analysis. *Justice Quarterly*, *26*(4), 716–745. https://doi.org/10.1080/07418820802506206

Welsh, B. C., & Taheri, S. A. (2018). What have we learned from Environmental Criminology for the prevention of crime? In G. J. N. Bruinsma & S. D. Johnson (Eds.), *The Oxford Handbook of Environmental Criminology* (pp. 757–775). Oxford University Press. https://doi.org/10.1093/oxfordhb/9780190279707.013.31

Westlake, B. G., & Bouchard, M. (2016). Liking and hyperlinking: Community detection in online child sexual exploitation networks. *Social Science Research*, *59*, 23–36. https://doi.org/10.1016/j.ssresearch.2016.04.010

Wheeler, A. P., & Steenbeek, W. (2020). Mapping the Risk Terrain for Crime Using Machine Learning. *Journal of Quantitative Criminology*. https://doi.org/10.1007/s10940-020-09457-7

Wickham, H. (2010). A Layered Grammar of Graphics. *Journal of Computational and Graphical Statistics*, *19*(1), 3–28. https://doi.org/10.1198/jcgs.2009.07098

Wickham, H. (2014). Tidy Data. *Journal of Statistical Software*, *59*(10), 1–23.

Wickham, H. (2015). *R packages* (First edition). O’Reilly Media.

Wickham, H. (2016). *ggplot2: Elegant graphics for data analysis* (Second edition). Springer.

Wickham, H. (2017). *Tidyverse: Easily Install and Load the ‘Tidyverse’*. https://CRAN.R-project.org/package=tidyverse

Wickham, H., Averick, M., Bryan, J., Chang, W., McGowan, L., François, R., Grolemund, G., Hayes, A., Henry, L., Hester, J., Kuhn, M., Pedersen, T. L., Miller, E., Bache, S., Müller, K., Ooms, J., Robinson, D., Seidel, D., Spinu, V., … Yutani, H. (2019). Welcome to the Tidyverse. *Journal of Open Source Software*, *4*(43), 1686. https://doi.org/10.21105/joss.01686

Wickham, H., & Bryan, J. (2019). *readxl: Read Excel Files* (Version 1.3.1) [Computer software]. https://CRAN.R-project.org/package=readxl

Wickham, H., François, R., Henry, L., & Müller, K. (2020). *dplyr: A Grammar of Data Manipulation* (Version 0.8.5) [Computer software]. https://CRAN.R-project.org/package=dplyr

Wickham, H., & Henry, L. (2020). *tidyr: Tidy Messy Data* (Version 1.0.2) [Computer software]. https://CRAN.R-project.org/package=tidyr

Wickham, H., Hester, J., & François, R. (2018). *readr: Read Rectangular Text Data* (Version 1.3.1) [Computer software]. https://CRAN.R-project.org/package=readr

Wickham, H., & Miller, E. (2019). *haven: Import and Export ‘SPSS’, ‘Stata’ and ‘SAS’ Files* (Version 2.2.0) [Computer software]. https://CRAN.R-project.org/package=haven

Wilcox, P., & Cullen, F. T. (2018). Situational Opportunity Theories of Crime. *Annual Review of Criminology*, *1*(1), 123–148. https://doi.org/10.1146/annurev-criminol-032317-092421

Wilkinson, L. (2005). *The grammar of graphics* (Second edition). Springer.

Williams, M. L., & Burnap, P. (2016). Cyberhate on Social Media in the aftermath of Woolwich: A Case Study in Computational Criminology and Big Data. *British Journal of Criminology*, *56*(2), 211–238. https://doi.org/10.1093/bjc/azv059

Williams, M. L., Burnap, P., & Sloan, L. (2016). Crime Sensing with Big Data: The Affordances and Limitations of using Open Source Communications to Estimate Crime Patterns. *British Journal of Criminology*, *57*(2), 320–340. https://doi.org/10.1093/bjc/azw031

Williams, M. L., Burnap, P., & Sloan, L. (2017). Towards an Ethical Framework for Publishing Twitter Data in Social Research: Taking into Account Users' Views, Online Context and Algorithmic Estimation. *Sociology*, *51*(6), 1149–1168. https://doi.org/10.1177/0038038517708140

Williams, M. L., Edwards, A., Housley, W., Burnap, P., Rana, O., Avis, N., Morgan, J., & Sloan, L. (2013). Policing cyber-neighbourhoods: Tension monitoring and social media networks. *Policing and Society*, *23*(4), 461–481. https://doi.org/10.1080/10439463.2013.780225

Willison, R. (2000). Understanding and Addressing Criminal Opportunity: The Application of Situational Crime Prevention to IS Security. *Journal of Financial Crime*, *7*(3), 201–210. https://doi.org/10.1108/eb025940

Wolak, J., Mitchell, K. J., & Finkelhor, D. (2007). Does Online Harassment Constitute Bullying? An Exploration of Online Harassment by Known Peers and Online-

Only Contacts. *Journal of Adolescent Health*, *41*(6), S51–S58. https://doi.org/10.1016/j.jadohealth.2007.08.019

Wolfe, S. E., Marcum, C. D., Higgins, G. E., & Ricketts, M. L. (2016). Routine Cell Phone Activity and Exposure to Sext Messages: Extending the Generality of Routine Activity Theory and Exploring the Etiology of a Risky Teenage Behavior. *Crime & Delinquency*, *62*(5), 614–644. https://doi.org/10.1177/0011128714541192

Woo, H., Kim, Y., & Dominick, J. (2004). Hackers: Militants or Merry Pranksters? A Content Analysis of Defaced Web Pages. *Media Psychology*, *6*(1), 63–82. https://doi.org/10.1207/s1532785xmep0601_3

Wortley, R. (1997). Reconsidering the role of opportunity in situational crime prevention. In G. R. Newman, R. V. Clarke, & S. G. Shoham (Eds.), *Rational Choice and Situational Crime Prevention* (pp. 65–81). Ashgate Publishing.

Wortley, R. (1998). A two-stage model of situational crime prevention. *Studies on Crime and Crime Prevention*, *7*, 173–188.

Wortley, R. (2001). A Classification of Techniques for Controlling Situational Precipitators of Crime. *Security Journal*, *14*(4), 63–82. https://doi.org/10.1057/palgrave.sj.8340098

Wortley, R. (2012). Situational prevention of child abuse in the new technologies. In E. Quayle & K. M. Ribisl (Eds.), *Understanding and Preventing Online Sexual Exploitation of Children* (pp. 188–204). Routledge.

Wortley, R., & Mazerolle, L. G. (2008). Environmental Criminology and Crime-Analysis: Situating the theory, analytic approach and application. In R. Wortley & L. G. Mazerolle (Eds.), *Environmental criminology and crime analysis* (pp. 1–15). Willan.

Wortley, R., Sidebottom, A., Tilley, N., & Laycock, G. (Eds.). (2018). *Routledge Handbook of Crime Science*. Routledge, Taylor & Francis Group.

Wortley, R., & Smallbone, S. (2012). *Internet child pornography: Causes, investigation, and prevention*. Praeger.

Wortley, R., & Townsley, M. (Eds.). (2017a). *Environmental Criminology and Crime Analysis* (Second Edition). Routledge, Taylor & Francis Group.

Wortley, R., & Townsley, M. (2017b). Environmental criminology and crime analysis: Situating the theory, analytic approach and application. In R. Wortley & M. Townsley (Eds.), *Environmental Criminology and Crime Analysis* (Second Edition, pp. 1–25). Routledge, Taylor & Francis Group.

Xie, Y., Allaire, J. J., & Grolemund, G. (2018). *R Markdown: The definitive guide*. Taylor & Francis, CRC Press.

Yar, M. (2005). The Novelty of 'Cybercrime': An Assessment in Light of Routine Activity Theory. *European Journal of Criminology*, *2*(4), 407–427. https://doi.org/10.1177/147737080556056

Ybarra, M. L., & Mitchell, K. J. (2008). How Risky Are Social Networking Sites? A Comparison of Places Online Where Youth Sexual Solicitation and Harassment Occurs. *Pediatrics*, *121*(2), 350–357. https://doi.org/10.1542/peds.2007-0693

APPENDICES

A. Additional information for the second hypothesis (CHAPTER VI)

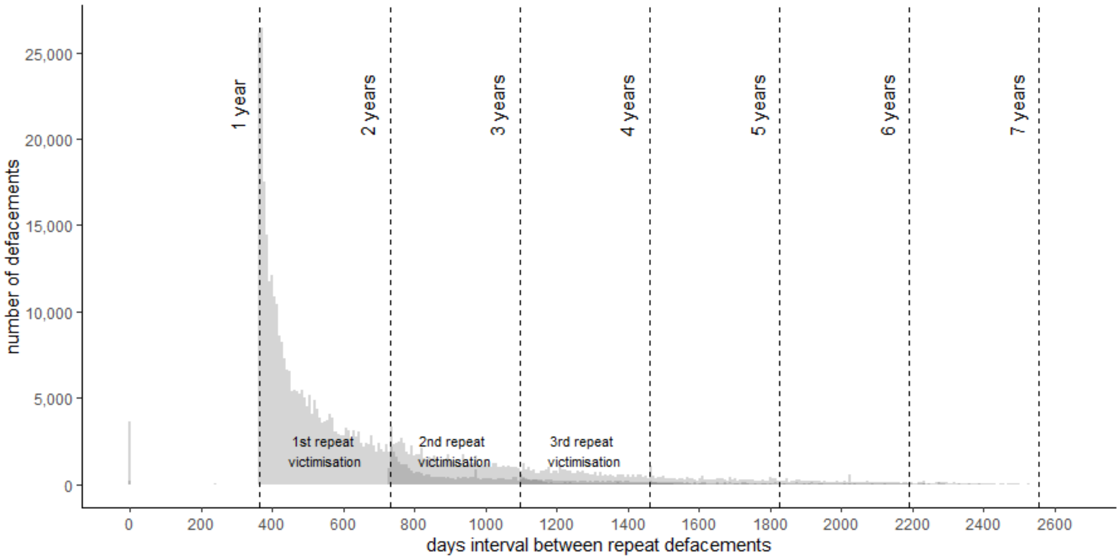

Figure 14. Repeat victimization time pattern for website defacements. Histogram binwidth = 7

B. Additional information for the third hypothesis (CHAPTER VI)

Table 22.

Percentage of offenders responsible for each type of defacement

			Repeat defacements			
	Total		Single		Mass	
Percentage of offenders	*n*	%	*n*	%	*n*	%
1	297,062	57.8	126,920	64.0	145,436	46.1
2	346,187	67.4	139,796	70.5	181,728	57.6
5	410,351	79.9	157,399	79.3	228,863	72.6
10	451,778	88.0	170,115	85.8	261,788	83.0
50	504,461	98.2	191,683	96.6	308,770	97.9
100	513,610	100.00	198,365	100.00	315,245	100.00

C. Hacking modes for website defacement

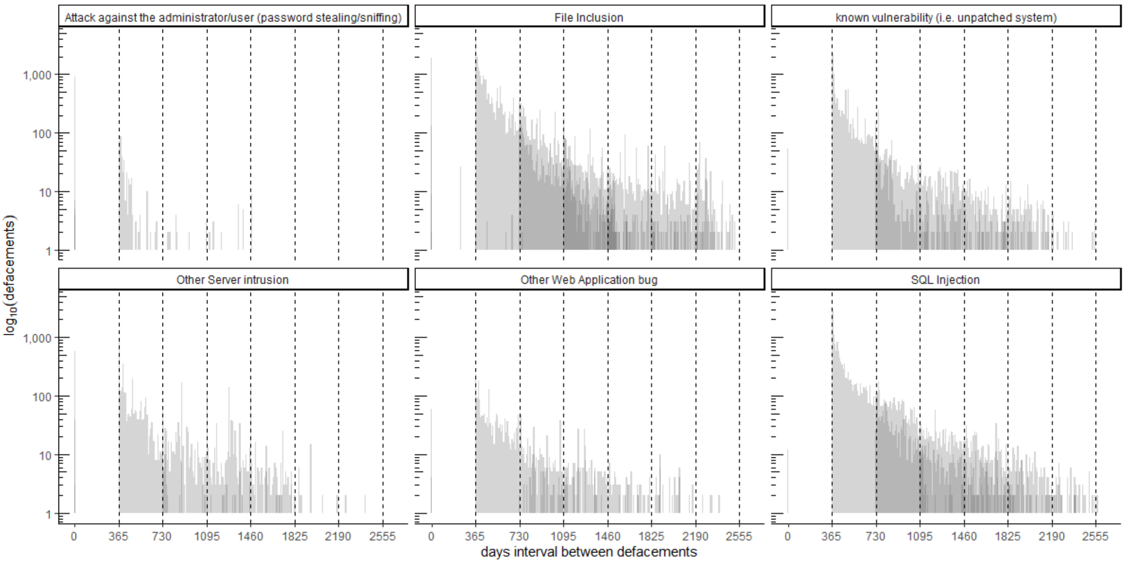

Figure 15. The 6 most common hacking modes to commit website defacement

D. Selected illicit FMIWs by number of external URL

Table 23.

Selected illicit FMIWs by number of external URL

Address	External URL
http://www.fixed-match.us/	163
https://master-fixed.com/	125
https://europol-fixed.com/	111
http://fixed-tips1x2.com/	99
https://1x2bettingtips.com/	85
https://prelazi-dojavi.com/	69
https://9ja-fixed.com/	68
http://juventus-bet.com/	64
http://williamhill1x2.com/	60
http://www.asia-fixedmatches.com/	59
http://fixed-advisor.com/	58
http://soccer-predictions.co.uk/	57
https://fixedmatchtip.com/	55
http://tips-free.com/	55
http://www.verifiedsoccertips.com/	55
https://fixedmatches.tips/	55
https://mata-fixed.tips/	51
https://bettingfixed.com/	50
https://fixedsafematches.com/	49
https://topbet-fixed.com/	48
https://america-fixedmatches.com/	48
https://fcfixedmatches.com/	47
https://fixed-matches.football/	46
https://fixedmatches.website/	46
https://fixed-matches.sportal.tips/	45
https://betting-predictions.football/	43
http://falcao1x2.com/	42
https://match-fixing.sportal.tips/	41
http://www.bestfixedmatches1x2.com/	41
http://matchesfixing.com/	38
http://betting-fixed.com/	37
http://matches-fixed.com/	37
https://manipulated-fixed-matches.sportal.tips/	36
https://www.realfixedmatches.com/	36
http://betyetu-fixed.com/	34
https://truefixedmatches1x2.com/	34
http://viti-bet.com/	33
https://www.freesupertips.co.uk/free-football-betting-tips/	33
http://fixedinsider.com/	32
http://supabet-fixed.com/	31
https://helena1x2.sportal.tips/	30
https://soccer-fixed.tips/	30
http://fixed-odd.com/	30
http://fixedmatches.uk/	28
https://adibet.tips/	27
https://fixed-matches.tips/	27
http://stat-area.com/	26
https://xanthi-fixed.matches.sportal.tips/	25
http://fixed-scores.com/	23
https://hotfixedmatches.com/	21
https://sonkotips.com/free-betting-tips/	20
http://site-fixed-matches.com/	20
https://real-fixedmatches.com/	20

Address	External URL
https://www.qatar-fixed.com/	19
https://basel-fixedmatches.com/	19
https://fixed.tips/	18
https://larsbetting.com/	14
https://fixedmatches.mobi/	13
https://www.paidpicks1x2.com/	13
https://strongfixedmatches.com/	13
https://predictz-tips.com/fixed-match/	10
https://fixedmatches-betting.com/	8
https://fixedmatches-seller.com/	8
https://fixedmatches-betting.com/	8
https://thefixedmatches.com/	7
https://www.fixed-betting-tips.com/	5
http://swiss-fixed.com/	5
https://fixedmatches.football/	5
https://matchfixed.com/	5
https://www.fixdrawsoccer.com/	3
https://fixedmatches.today/	3
https://www.oddstips.co.uk/	3
https://betting1x2.football/	2
http://fixed.matches1x2.com/	2
https://theopicks.com/fixed-matches-single/	2
http://www.fixedbetting.tips/	1
http://www.fixedodd.tips/	1
http://www.bestfootballtips.net/	1

E. Selected regulated sport-betting websites by operator

Table 24.

Selected regulated sport-betting websites by operator

Address	Operator
888sport.com	Cassava Enterprises
bet365.com	Hillside (Sports) ENC
betclic.fr	BetClic Enterprises
betfred.com	Petfre
betsson.com	BML Group
betvictor.com	BetVictor
betway.com	Betway
danskespil.dk/oddset	Danske Spil
e-lotto.be	Loterie Nationale
enligne.parionssport.fdj.fr	Parions Sport En Ligne
e-stave.com	Športna loterija
fonbet.com	Leofon
interwetten.com/en/sportsbook	Interwetten Gaming
jeux.loro.ch/sports	Loterie Romande
lottomatica.it/scommesse/avvenimenti	Lottomatica
pamestoixima.gr	Pamestoixima
sisal.it	Sisal Entertainment
skybet.com	Bonne Terre
spela.svenskaspel.se/europatipset	Svenska Spel Sport & Casino AB
sportingbet.com	ElectraWorks
sportingindex.com	Sporting Index
sports.williamhill.com/bet/en-gb	WHG
swisslos.ch/fr/sporttip/parissportifs/prognostics.html	Swisslos Lotería Intercantonal
tipkurz.etipos.sk	Tipos
tippmixpro.hu	Szerencsejáték Zrt
unibet.com	Trannel International
veikkaus.fi/fi/live-veto	Veikkaus
win2day.at/sportwetten	Österreichische Lotterien GmbH

F. List of regulated betting sites found in the FMIW network

Table 25.

List of regulated betting sites found in the FMIW network

Address	Operator
williamhill.com	WHG
bet365.com	Hillside (Sports) ENC
188bet.com	Cube Limited
betboro.com	Webmedia Development N.V.
betvictor.com	BetVictor
bigbetworld.com	M-Hub Gaming Operations (inactive)
ladbrokes.com.au	GVC Australia
paddypower.com	PPB Counterparty Services
sbobet.com	SBOBET
12bet.com	TGP Europe
betsson.com	BML Group
betway.com	Betway
bwin.com	ElectraWorks
betfred.com	Petfre
mybet.com	Rhinoceros Operations
nordicbet.com	BML Group
odds.betsafe.com	BML Group
skybet.com	Bonne Terre
sportingbet.com	ElectraWorks
stanjames.com	Platinum Gaming
unibet.com	Trannel International
betway.com	Betway
sports.coral.co.uk	Coral Interactive
bet9ja.com	KC Gaming Networks
marathonbet.com	Marathonbet Spain
sports.ladbrokes.com	Ladbrokes Betting & Gaming

G. Informed consent model

Informed Consent [84]

The Education Council of Castile-Leon, on behalf of the Crímina Research Centre at Miguel Hernández University of Elche, requests the participation of your child in the study that is being conducted on the habits and possible risks to which minors are exposed when using new technologies.

In order to collect real and objective data, a completely anonymous questionnaire will be administered in relation to the aforementioned topic. The information collected will be confidential and will only be used to accomplish the objectives of the research. If you have any questions, please contact Crímina Research Centre by phone at 966 658 406 or by e-mail at saf_e@crimina.es.

For your child's participation in this project, it is mandatory that you sign the attached consent form. We ask you to submit the consent form as soon as possible to your educational centre by the same means of delivery, so that it can be received before the survey is administered.

Thank you very much for your attention,

Fernando Miró-Llinares, PhD

Dean of the Faculty of Social and Legal Sciences

Professor of Criminal Law and Criminology

Director of the Crímina Research Centre

Having been informed of these conditions, I voluntarily accept that my child [complete name], student of the course [course] of the centre [educational centre], participates in this project.

[84] Translated from the original document in Spanish.

Complete name of the parent/tutor:

ID card:

Signature:

In [city] on [date].

Miguel Hernandez University

University Avenue

Helike Building

03202 Elche

Tel. 966 658 406

Email: saf_e@crimina.es

H. Selected questionnaire

1. Are you a:

   ☐ Boy

   ☐ Girl

2. What is your age?

   [Open-ended question]

3. How many hours a day do you spend surfing the Internet?

   ☐ Less than 1 hour

   ☐ From 1 to 3 hours

   ☐ From 4 to 7 hours

   ☐ From 8 to 15 hours

   ☐ More than 15 hours

4. Which of the following social media do you use daily? (You can choose more than one option)

   ☐ I do not use social media

   ☐ Snapchat

   ☐ Instagram

   ☐ Facebook

   ☐ Twitter

   ☐ Another, which one?

5. What kind of personal data do you publish in social media? (You can choose more than one option)

   ☐ I do not publish any personal data

☐ First name and/or surname

☐ Personal photos

☐ Another, which one?

6. Do you restrict access to your social media (only your contacts can see your information)?

☐ Yes

☐ No

7. In the last year, has anyone repeatedly insulted or humiliated you online?

☐ Yes

☐ No

8. In the last year, has anyone repeatedly told rumours or lies about you online?

☐ Yes

☐ No

9. In the last year, has anyone repeatedly marginalized you online?

☐ Yes

☐ No

10. In the last year, has anyone repeatedly threatened you online?

☐ Yes

☐ No

11. In the last year, has anyone repeatedly pretended to be you online?

☐ Yes

☐ No

12. In the last year, have you repeatedly insulted or humiliated someone online?

☐ Yes

☐ No

13. In the last year, have you repeatedly told rumours or lies about someone online?

☐ Yes

☐ No

14. In the last year, have you repeatedly marginalized someone online?

☐ Yes

☐ No

15. In the last year, have you repeatedly threatened someone online?

☐ Yes

☐ No

16. In the last year, have you repeatedly pretended to be someone else online?

☐ Yes

☐ No

I. Complete data of repeat victimization dominant profiles

Table 26.

Complete table for dominant case configurations likely to result in online harassment victimization, the probability of being victimized, and the number of students associated with each profile (n = 94)

ID	Sex	Age	Hours	Snapchat	Instagram	Facebook	Twitter	Name	Photos	Privacy	P(V)	N
1	Female	15 - 17	4 - 7	Yes	Yes	No	No	Yes	Yes	No	0.82	11
2	Female	12 - 14	< 4	No	Yes	No	No	Yes	Yes	Yes	0.70	10
3	Female	15 - 17	4 - 7	No	Yes	No	No	Yes	Yes	No	0.63	16
4	Female	15 - 17	4 - 7	No	Yes	No	Yes	Yes	No	No	0.60	10
5	Female	18 - 20	4 - 7	No	Yes	No	No	No	No	Yes	0.60	10
6	Female	18 - 20	4 - 7	No	Yes	Yes	Yes	No	No	Yes	0.60	10
7	Female	15 - 17	4 - 7	No	Yes	No	Yes	Yes	Yes	Yes	0.59	22
8	Female	15 - 17	4 - 7	No	Yes	No	No	Yes	Yes	Yes	0.58	24
9	Female	15 - 17	4 - 7	Yes	Yes	No	Yes	Yes	No	Yes	0.55	11
10	Male	12 - 14	< 4	No	Yes	No	No	Yes	Yes	Yes	0.55	11
11	Male	15 - 17	< 4	No	Yes	No	Yes	Yes	Yes	Yes	0.54	13
12	Female	18 - 20	4 - 7	No	Yes	Yes	No	Yes	No	Yes	0.50	14
13	Male	15 - 17	< 4	No	Yes	No	No	Yes	Yes	Yes	0.48	31
14	Male	15 - 17	4 - 7	No	Yes	No	No	Yes	No	Yes	0.48	31
15	Female	15 - 17	4 - 7	No	Yes	Yes	Yes	No	No	Yes	0.48	23
16	Male	15 - 17	< 4	No	Yes	No	No	Yes	No	Yes	0.47	34
17	Male	18 - 20	< 4	No	Yes	No	No	No	No	Yes	0.47	17
18	Female	15 - 17	4 - 7	Yes	Yes	No	No	Yes	No	Yes	0.46	13
19	Male	15 - 17	4 - 7	No	Yes	No	Yes	Yes	No	No	0.45	11
20	Male	15 - 17	4 - 7	No	Yes	No	No	Yes	Yes	Yes	0.45	20
21	Female	15 - 17	4 - 7	Yes	Yes	No	Yes	No	No	Yes	0.44	16
22	Male	15 - 17	4 - 7	No	Yes	Yes	No	No	No	Yes	0.44	16
23	Female	15 - 17	4 - 7	No	Yes	No	Yes	No	No	Yes	0.43	51
24	Female	15 - 17	4 - 7	No	Yes	No	No	Yes	No	Yes	0.42	33
25	Female	12 - 14	< 4	Yes	Yes	No	No	No	No	No	0.42	12
26	Female	15 - 17	< 4	No	Yes	No	No	Yes	No	Yes	0.40	40
27	Female	15 - 17	< 4	No	Yes	Yes	No	No	No	Yes	0.40	25
28	Female	12 - 14	4 - 7	No	Yes	No	No	No	No	No	0.40	15
29	Male	18 - 20	4 - 7	No	Yes	Yes	No	No	No	Yes	0.40	15
30	Female	15 - 17	< 4	Yes	Yes	No	Yes	Yes	Yes	Yes	0.40	10
31	Female	12 - 14	4 - 7	No	Yes	No	No	Yes	Yes	Yes	0.38	13
32	Male	18 - 20	4 - 7	No	Yes	No	No	No	No	Yes	0.38	13
33	Female	15 - 17	< 4	No	Yes	No	No	No	No	No	0.38	16
34	Female	15 - 17	4 - 7	No	Yes	No	No	Yes	No	No	0.38	16
35	Female	15 - 17	4 - 7	Yes	Yes	No	Yes	Yes	Yes	Yes	0.36	11
36	Female	12 - 14	4 - 7	Yes	Yes	No	No	No	No	Yes	0.35	62
37	Female	15 - 17	< 4	No	Yes	No	No	Yes	Yes	Yes	0.35	17
38	Male	12 - 14	4 - 7	No	No	No	No	No	No	Yes	0.35	17
39	Male	15 - 17	< 4	No	Yes	No	Yes	Yes	No	Yes	0.33	18
40	Female	12 - 14	4 - 7	No	Yes	No	No	No	No	Yes	0.33	89

ID	Sex	Age	Hours	Snapchat	Instagram	Facebook	Twitter	Name	Photos	Privacy	P(V)	N
41	Male	12 - 14	4 - 7	No	Yes	No	Yes	No	No	Yes	0.32	25
42	Male	15 - 17	4 - 7	No	Yes	No	No	No	No	Yes	0.32	85
43	Female	15 - 17	4 - 7	No	Yes	No	Yes	Yes	No	Yes	0.32	19
44	Female	15 - 17	< 4	No	Yes	No	Yes	Yes	No	Yes	0.31	16
45	Female	15 - 17	4 - 7	No	Yes	No	No	No	No	Yes	0.31	100
46	Female	12 - 14	4 - 7	Yes	Yes	No	No	Yes	No	Yes	0.31	13
47	Male	15 - 17	< 4	No	Yes	No	No	No	No	No	0.31	36
48	Female	15 - 17	4 - 7	No	Yes	No	No	No	No	No	0.30	23
49	Female	12 - 14	< 4	No	Yes	No	No	Yes	No	Yes	0.30	20
50	Female	12 - 14	< 4	No	Yes	No	Yes	No	No	Yes	0.30	10
51	Female	15 - 17	4 - 7	Yes	Yes	No	No	Yes	Yes	Yes	0.30	10
52	Female	18 - 20	4 - 7	No	Yes	Yes	No	No	No	Yes	0.29	17
53	Male	12 - 14	4 - 7	No	Yes	No	No	No	No	Yes	0.29	66
54	Male	15 - 17	< 4	No	Yes	No	Yes	No	No	Yes	0.29	35
55	Male	12 - 14	< 4	No	Yes	No	No	Yes	No	Yes	0.29	21
56	Female	15 - 17	< 4	No	Yes	No	Yes	No	No	Yes	0.28	25
57	Female	12 - 14	< 4	Yes	Yes	No	No	No	No	Yes	0.27	62
58	Male	15 - 17	4 - 7	No	Yes	No	Yes	No	No	Yes	0.27	33
59	Male	15 - 17	< 4	No	Yes	Yes	No	No	No	Yes	0.27	22
60	Male	15 - 17	< 4	No	Yes	No	Yes	No	No	No	0.27	11
61	Male	15 - 17	< 4	Yes	Yes	No	No	No	No	Yes	0.27	11
62	Male	15 - 17	4 - 7	Yes	Yes	No	Yes	No	No	Yes	0.27	11
63	Male	15 - 17	4 - 7	Yes	Yes	Yes	Yes	No	No	Yes	0.27	11
64	Female	12 - 14	4 - 7	No	Yes	No	No	Yes	No	Yes	0.27	15
65	Male	15 - 17	4 - 7	No	No	No	No	No	No	Yes	0.27	15
66	Female	15 - 17	4 - 7	Yes	Yes	No	No	No	No	Yes	0.26	35
67	Female	12 - 14	< 4	No	Yes	No	No	No	No	Yes	0.25	181
68	Female	15 - 17	< 4	Yes	Yes	No	No	No	No	Yes	0.25	32
69	Female	15 - 17	4 - 7	No	Yes	Yes	No	No	No	Yes	0.25	32
70	Male	15 - 17	< 4	No	No	No	No	No	No	No	0.25	12
71	Male	15 - 17	< 4	Yes	Yes	No	Yes	No	No	Yes	0.25	12
72	Male	12 - 14	4 - 7	No	Yes	No	No	No	No	No	0.24	17
73	Male	15 - 17	4 - 7	No	Yes	No	Yes	Yes	No	Yes	0.23	22
74	Male	12 - 14	< 4	No	Yes	No	No	No	No	Yes	0.22	174
75	Female	12 - 14	4 - 7	No	No	No	No	No	No	Yes	0.20	20
76	Female	12 - 14	4 - 7	Yes	Yes	No	Yes	No	No	Yes	0.20	10
77	Male	15 - 17	4 - 7	No	Yes	Yes	Yes	No	No	Yes	0.20	10
78	Male	15 - 17	< 4	No	Yes	No	No	No	No	Yes	0.19	150
79	Female	15 - 17	< 4	No	Yes	No	No	No	No	Yes	0.19	114
80	Male	15 - 17	4 - 7	No	Yes	No	No	No	No	No	0.19	43
81	Male	12 - 14	< 4	No	Yes	No	Yes	No	No	Yes	0.18	22
82	Male	12 - 14	< 4	No	Yes	No	No	No	No	No	0.18	34
83	Female	12 - 14	< 4	No	Yes	No	No	No	No	No	0.16	19
84	Male	12 - 14	< 4	No	No	No	No	No	No	No	0.15	13
85	Female	12 - 14	< 4	No	No	No	No	No	No	Yes	0.14	66
86	Male	15 - 17	4 - 7	No	Yes	No	Yes	No	No	No	0.11	18
87	Male	12 - 14	< 4	Yes	Yes	No	No	No	No	Yes	0.11	19

ID	Sex	Age	Hours	Snapchat	Instagram	Facebook	Twitter	Name	Photos	Privacy	P(V)	N
88	Female	15 - 17	< 4	No	No	No	No	No	No	Yes	0.10	29
89	Female	15 - 17	4 - 7	No	No	No	No	No	No	Yes	0.10	10
90	Female	18 - 20	< 4	No	Yes	Yes	No	No	No	Yes	0.10	10
91	Male	12 - 14	< 4	No	No	No	No	No	No	Yes	0.09	89
92	Male	12 - 14	< 4	No	Yes	Yes	No	No	No	Yes	0.08	13
93	Male	15 - 17	< 4	No	No	No	No	No	No	Yes	0.07	42
94	Male	12 - 14	4 - 7	No	Yes	Yes	Yes	No	No	Yes	0.00	10

J. Complete data of repeat offending dominant profiles

Table 27.

Complete table for dominant case configurations likely to result in online harassment offending, the probability of offending, and the number of students associated with each profile (n = 94)

ID	Sex	Age	Hours	Snapchat	Instagram	Facebook	Twitter	Name	Photos	Privacy	P(O)	N
1	Female	15 - 17	4 - 7	No	Yes	No	No	Yes	Yes	No	0.44	16
2	Male	15 - 17	4 - 7	No	Yes	No	No	Yes	Yes	Yes	0.40	20
3	Male	15 - 17	< 4	No	Yes	No	Yes	Yes	No	Yes	0.39	18
4	Female	15 - 17	4 - 7	Yes	Yes	No	Yes	Yes	Yes	Yes	0.36	11
5	Male	15 - 17	4 - 7	No	Yes	No	No	Yes	No	Yes	0.32	31
6	Male	12 - 14	4 - 7	No	Yes	No	No	No	No	No	0.29	17
7	Female	15 - 17	4 - 7	Yes	Yes	No	No	Yes	Yes	No	0.27	11
8	Male	15 - 17	< 4	No	Yes	No	Yes	No	No	No	0.27	11
9	Male	15 - 17	< 4	Yes	Yes	No	No	No	No	Yes	0.27	11
10	Male	15 - 17	< 4	No	Yes	No	No	Yes	Yes	Yes	0.26	31
11	Female	15 - 17	4 - 7	No	Yes	No	No	Yes	No	No	0.25	16
12	Male	15 - 17	< 4	No	Yes	No	Yes	Yes	Yes	Yes	0.23	13
13	Male	15 - 17	4 - 7	No	Yes	No	Yes	Yes	No	Yes	0.23	22
14	Female	18 - 20	4 - 7	No	Yes	Yes	No	Yes	No	Yes	0.21	14
15	Female	12 - 14	< 4	No	Yes	No	No	No	No	No	0.21	19
16	Female	15 - 17	4 - 7	Yes	Yes	No	No	Yes	Yes	Yes	0.20	10
17	Female	18 - 20	4 - 7	No	Yes	Yes	Yes	No	No	Yes	0.20	10
18	Male	15 - 17	4 - 7	No	Yes	Yes	Yes	No	No	Yes	0.20	10
19	Female	15 - 17	< 4	No	Yes	No	Yes	Yes	No	Yes	0.19	16
20	Female	15 - 17	4 - 7	No	Yes	No	Yes	Yes	Yes	Yes	0.18	22
21	Male	12 - 14	< 4	No	Yes	No	No	Yes	Yes	Yes	0.18	11
22	Male	15 - 17	4 - 7	Yes	Yes	No	Yes	No	No	Yes	0.18	11
23	Male	15 - 17	4 - 7	Yes	Yes	Yes	Yes	No	No	Yes	0.18	11
24	Male	15 - 17	< 4	No	Yes	No	No	Yes	No	Yes	0.18	34
25	Male	15 - 17	< 4	No	Yes	No	No	No	No	No	0.17	36
26	Female	15 - 17	4 - 7	No	Yes	No	No	Yes	Yes	Yes	0.17	24
27	Male	15 - 17	4 - 7	No	Yes	No	No	No	No	No	0.16	43
28	Female	12 - 14	4 - 7	No	Yes	No	No	Yes	Yes	Yes	0.15	13
29	Female	12 - 14	4 - 7	Yes	Yes	No	No	Yes	No	Yes	0.15	13
30	Female	15 - 17	4 - 7	Yes	Yes	No	No	Yes	No	Yes	0.15	13
31	Male	12 - 14	< 4	No	No	No	No	No	No	No	0.15	13
32	Male	18 - 20	4 - 7	No	Yes	No	No	No	No	Yes	0.15	13
33	Male	12 - 14	< 4	No	Yes	No	No	Yes	No	Yes	0.14	21
34	Male	15 - 17	4 - 7	No	Yes	No	No	No	No	Yes	0.14	85
35	Male	15 - 17	< 4	No	Yes	Yes	No	No	No	Yes	0.14	22
36	Female	12 - 14	4 - 7	No	Yes	No	No	No	No	No	0.13	15
37	Female	15 - 17	< 4	No	Yes	No	No	No	No	No	0.13	16
38	Female	15 - 17	4 - 7	Yes	Yes	No	Yes	No	No	Yes	0.13	16
39	Female	15 - 17	< 4	No	Yes	No	No	Yes	Yes	Yes	0.12	17
40	Male	12 - 14	4 - 7	No	No	No	No	No	No	Yes	0.12	17

ID	Sex	Age	Hours	Snapchat	Instagram	Facebook	Twitter	Name	Photos	Privacy	P(O)	N
41	Male	18 - 20	< 4	No	Yes	No	No	No	No	Yes	0.12	17
42	Male	12 - 14	< 4	No	Yes	No	No	No	No	Yes	0.11	174
43	Male	15 - 17	< 4	No	Yes	No	Yes	No	No	Yes	0.11	35
44	Male	15 - 17	< 4	No	Yes	No	No	No	No	Yes	0.11	150
45	Female	15 - 17	< 4	No	No	No	No	No	No	Yes	0.10	29
46	Female	15 - 17	4 - 7	No	Yes	No	No	No	No	Yes	0.10	100
47	Female	12 - 14	< 4	No	Yes	No	No	Yes	Yes	Yes	0.10	10
48	Female	12 - 14	< 4	No	Yes	No	Yes	No	No	Yes	0.10	10
49	Female	15 - 17	< 4	Yes	Yes	No	Yes	Yes	Yes	Yes	0.10	10
50	Female	15 - 17	4 - 7	No	No	No	No	No	No	Yes	0.10	10
51	Female	15 - 17	4 - 7	No	Yes	No	Yes	Yes	No	No	0.10	10
52	Female	12 - 14	< 4	Yes	Yes	No	No	No	No	Yes	0.10	62
53	Female	15 - 17	4 - 7	No	Yes	Yes	No	No	No	Yes	0.09	32
54	Male	15 - 17	4 - 7	No	Yes	No	Yes	No	No	Yes	0.09	33
55	Female	15 - 17	4 - 7	Yes	Yes	No	Yes	Yes	No	Yes	0.09	11
56	Female	12 - 14	4 - 7	No	Yes	No	No	No	No	Yes	0.09	89
57	Female	15 - 17	4 - 7	No	Yes	Yes	Yes	No	No	Yes	0.09	23
58	Male	12 - 14	4 - 7	No	Yes	No	Yes	No	No	Yes	0.08	25
59	Male	15 - 17	4 - 7	No	No	No	No	No	No	Yes	0.07	15
60	Male	18 - 20	4 - 7	No	Yes	Yes	No	No	No	Yes	0.07	15
61	Female	12 - 14	4 - 7	Yes	Yes	No	No	No	No	Yes	0.06	62
62	Female	15 - 17	< 4	Yes	Yes	No	No	No	No	Yes	0.06	32
63	Male	15 - 17	4 - 7	No	Yes	Yes	No	No	No	Yes	0.06	16
64	Female	15 - 17	< 4	No	Yes	No	No	No	No	Yes	0.06	114
65	Female	15 - 17	4 - 7	No	Yes	No	No	Yes	No	Yes	0.06	33
66	Male	12 - 14	< 4	No	Yes	No	No	No	No	No	0.06	34
67	Female	18 - 20	4 - 7	No	Yes	Yes	No	No	No	Yes	0.06	17
68	Female	15 - 17	4 - 7	Yes	Yes	No	No	No	No	Yes	0.06	35
69	Male	15 - 17	4 - 7	No	Yes	No	Yes	No	No	No	0.06	18
70	Female	15 - 17	4 - 7	No	Yes	No	Yes	Yes	No	Yes	0.05	19
71	Male	12 - 14	< 4	Yes	Yes	No	No	No	No	Yes	0.05	19
72	Female	15 - 17	< 4	No	Yes	No	No	Yes	No	Yes	0.05	40
73	Female	12 - 14	4 - 7	No	No	No	No	No	No	Yes	0.05	20
74	Female	12 - 14	< 4	No	No	No	No	No	No	Yes	0.05	66
75	Male	12 - 14	4 - 7	No	Yes	No	No	No	No	Yes	0.05	66
76	Male	12 - 14	< 4	No	Yes	No	Yes	No	No	Yes	0.05	22
77	Male	12 - 14	< 4	No	No	No	No	No	No	Yes	0.04	89
78	Female	15 - 17	< 4	No	Yes	Yes	No	No	No	Yes	0.04	25
79	Female	12 - 14	< 4	No	Yes	No	No	No	No	Yes	0.03	181
80	Male	15 - 17	< 4	No	No	No	No	No	No	Yes	0.02	42
81	Female	15 - 17	4 - 7	No	Yes	No	Yes	No	No	Yes	0.02	51
82	Female	15 - 17	< 4	No	Yes	No	Yes	No	No	Yes	0.00	25
83	Female	15 - 17	4 - 7	No	Yes	No	No	No	No	No	0.00	23
84	Female	12 - 14	< 4	No	Yes	No	No	Yes	No	Yes	0.00	20
85	Female	12 - 14	4 - 7	No	Yes	No	No	Yes	No	Yes	0.00	15
86	Male	12 - 14	< 4	No	Yes	Yes	No	No	No	Yes	0.00	13
87	Female	12 - 14	< 4	Yes	Yes	No	No	No	No	No	0.00	12

ID	Sex	Age	Hours	Snapchat	Instagram	Facebook	Twitter	Name	Photos	Privacy	P(O)	N
88	Male	15 - 17	< 4	No	No	No	No	No	No	No	0.00	12
89	Male	15 - 17	< 4	Yes	Yes	No	Yes	No	No	Yes	0.00	12
90	Male	15 - 17	4 - 7	No	Yes	No	Yes	Yes	No	No	0.00	11
91	Female	12 - 14	4 - 7	Yes	Yes	No	Yes	No	No	Yes	0.00	10
92	Female	18 - 20	< 4	No	Yes	Yes	No	No	No	Yes	0.00	10
93	Female	18 - 20	4 - 7	No	Yes	No	No	No	No	Yes	0.00	10
94	Male	12 - 14	4 - 7	No	Yes	Yes	Yes	No	No	Yes	0.00	10

K. Pseudocode for pre-processing the variables

When implementing the script “cont_mention_hashtag_url_emoji”, an algorithm is executed that counts the number of mentions, hashtags, URL, and emojis included by the user in the body of the tweet. Additionally, this script extracts the URL, emojis and retweet structure (i.e., RT @username).

The first loop goes through the tweets in the sample together with a second loop nested within the first one that goes through the text characters of each tweet with the condition that when the nested loop finds a mention (@), a hashtag (#) the count variable increases its value to 1. Subsequently, this count is inserted into “vector_mention_hashtag”, which has the same dimension as the vector including all the texts from the tweet (txt_twt), and it is finally inserted as a new column in the dataframe of our data. In the case of URL and emojis, there is a similar structure, with the exception that the use of a nested loop is not needed: finding the string “http” and “<” in the text of the tweet and inserting it into “stopword” is enough. To extract the structure of a retweet it is only necessary to take the words after the second division of the text and finally insert them in the stopword of the array.

```
    ALGORITHM cont_mentions_hashtag_url_emoji
VAR
    DATAFRAME data;
    STRING txt_twt;
    ARRAY vector_mention_hashtag, vector_url_emoji, stopword;
    INTEGER cont_mention_hashtag_emoji;
    BOOL cont_url_emoji;
    BEGIN
    txt_twt <- data[‘text’];
    for i from 0 to len(txt_twt)
    cont_mention_hashtag <- 0;
        cont_url_emoji <- 0;
    for j from 0 to txt_twt[i]
            if (txt_twt[i].find(mention) or txt_twt[i].find(hashtag))
                cont_mention_hashtag <- cont_mention_hashtag + 1
        vector_mention_hashtag <- cont_mention_hashtag;
        if(txt_twt[i].find(url) or txt_twt[i].find(emoji))
        stopword <- txt_twt[i].split();
        cont_url_emoji;
    vector_url_emoji <- cont_url_emoji;
    if(txt_twt[i][0-4] == retweet)
        txt_twt <- txt_twt[i][0-txt_twt.split(2)]
END
```

The following script was applied to recalculate the length of the body of a message once the URL, emojis, and RT structure were removed. In this sense, when the algorithm detects a URL, an emoji, or an URL structure, according to the specifications of the variable “url_emojis”, and “retweet”, it writes both the string of the URL and the codification of the emoji on a stopword file in such a way that after readjusting the type of variable, all elements stored in the file are deleted from the text of the tweet.

```
ALGORITHM delete_urls_emojis
VAR
DATAFRAME data, df;
STRING txt_twt;
STRING url_emojis, word;
INTEGER stopword[], a[], l[], ;
INTEGER i, detect;
FILE file_stopword;
BEGIN
file_stopword.WRITE(stopword)
df <- pd.DataFrame(stopword)
a <- np.array(df)
l <. a.ravel()
loadStopWordsFile <- lambda f : [re.sub('\r|\n', “”, l) for l in codecs.open(f)]
stop <- np.unique(loadStopWordsFile('stopword.txt'))
    data['tweet_without_stopwords'] <- data[0].apply(lambda x: ' '.join([word
for word in x.split() if word not in (stop)]))
END
```

L. Pseudocode for obtaining the sample

The algorithm capture_tweets was applied within the general-purpose programming language Python using the libraries “sys”, “email”, “smtplib” and “tweepy”. The first, “sys”, is used to store the hashtags introduced as arguments on the Linux dashboard on the strings Q vector; the second and third, “email” and “smtplib”, are used to implement the function called email(), which sends emails to specific addressees with the name of the file in which an error has occurred; and finally, “tweepy” is the library that enables access to the Twitter API. Thus, during the application of the algorithm, the variables access_token, access_token_secret, consumer_key and consumer_secret were declared to store authentication codes provided to the user with developer permissions by Twitter to establish a connection to the Twitter API using the method “tweepy”.

OAUTHHANDLER and STREAM. Furthermore, the listener class STDOUTLISTENER was implemented, using the method StreamListener of tweepy.streaming as a parameter. Thus, when a tweet with the hashtag filter specified on the stream variable from STREAM type is published, the aim is to store it in the outfile file in the JSON format and, if there is an error, to name the function email().

```
ALGORITHM capture_tweets
VAR
INTEGER i <- 0;
STRING Q[];
FILE outfile;
STRING access_token, access_token_secret, consumer_key, consumer_secret;
STDOUTLISTENER l;
TWEEPY.OAUTHHANDLER auth;
TWEEPY.STREAM stream;
CLASS STDOUTLISTENER(StreamListener)
FUNCTION on_data(self, data)
    global i;
    outfile.WRITE(data);
    i <- i + 1
END FUNCTION
```

```
FUNCTION on_error(self, status)
        email();
END FUNCTION
END CLASS
BEGIN
READ(Q);
l <- STDOUTLISTENER();
auth <- OAUTHHANDLER(consumer_key, consumer_secret);
auth.set_access_token(access_token, access_token_secret);
stream <- STREAM(auth, l);
stream.filter(track = Q)
END
```

LIST OF TABLES

LIST OF FIGURES