The Formation, Structure, and Multiplexity of Criminal Networks

George Wood

Nuffield College
University of Oxford

A thesis submitted for the degree of
Doctor of Philosophy

Trinity 2016

Abstract

This thesis presents two empirical papers investigating collaboration in a drug trafficking network, and one empirical paper on the role of social network embeddedness on the incidence of homicide arrests, re-arrests, and co-arrests.

The first paper, *The Structure and Vulnerability of a Drug Trafficking Collaboration Network*, examines the network structure of a heroin trafficking operation. Structural features of the network are characterized and compared to non-criminal collaboration networks, and a simulation model of the vulnerability of the trafficking network to node removal in various degree-mixing states is presented. Paper two reports the results of a *Bayesian Exponential Random Graph Model of Collaborative Tie Formation in a Drug Trafficking Network*, which explores the formation of collaborative ties in the trafficking organization, testing whether collaboration is subject to endogenous shared partner effects and homophily effects by subgroup membership and task specialization. The third paper investigates *The Influence of Network Embeddedness and Multiplexity on Homicide Arrests and Re-Arrests* by analyzing how the probability of arrests for homicide is influenced by prior membership of co-arrest dyads and higher-order structures, and estimating the interdependence of co-arrests for homicide and co-arrests for other crimes. The final chapter concludes by drawing out the main results of the thesis and commenting on their implications. Challenges in the analysis of criminal networks are outlined, such as missing data, non-random sampling, the selection of appropriate models, and the integration of network and contextual data. Strategies and modeling options for dealing with these challenges are discussed.

This thesis uncovers the structure of collaboration in a drug trafficking operation and how this affects the resilience of the trafficking organization to interference, analyzes the formation of collaborative ties in complex criminal operations, and presents the first results on how the social relationships underlying arrests influence the probability of re-arrests and co-arrests.

The Formation, Structure, and Multiplexity of Criminal Networks



George Wood

Nuffield College

University of Oxford

A thesis submitted for the degree of

Doctor of Philosophy

Trinity 2016

Acknowledgements

First and foremost, I am grateful to my supervisor Federico Varese, who kick-started my interest in criminal networks during my M.Sc. and has continued to provide support and guidance throughout my D.Phil. His enthusiasm and commitment to understanding organized crime, and to communicating the results of research to audiences outside of academia, is inspiring. I am thankful to Colin Mills for acting as my second supervisor and, before taking on that role, for being an engaging teacher.

During my D.Phil, I have been fortunate to spend time at the Department of Sociology, Yale University, and at the Yale Institute for Network Science. I thank Andrew Papachristos for hosting me at Yale and greatly appreciate his mentorship. I have also been fortunate to spend time as a visiting scholar at the Department of Sociology, Columbia University. I thank Peter Bearman for facilitating this visit.

At various stages, I have benefited from receiving rigorous feedback on parts of this thesis. I thank Michael Biggs, Paolo Campana, David Kirk, Christiaan Monden, and Felix Reed-Tsochas for their comments and suggestions.

I gratefully acknowledge the Economic and Social Research Council and the Sociology Group at Nuffield College for financial support.

Abstract

This thesis presents two empirical papers investigating collaboration in a drug trafficking network, and one empirical paper on the role of social network embeddedness on the incidence of homicide arrests, re-arrests, and co-arrests.

The first paper, *The Structure and Vulnerability of a Drug Trafficking Collaboration Network*, examines the network structure of a heroin trafficking operation. Structural features of the network are characterized and compared to non-criminal collaboration networks, and a simulation model of the vulnerability of the trafficking network to node removal in various degree-mixing states is presented. Paper two reports the results of a *Bayesian Exponential Random Graph Model of Collaborative Tie Formation in a Drug Trafficking Network*, which explores the formation of collaborative ties in the trafficking organization, testing whether collaboration is subject to endogenous shared partner effects and homophily effects by subgroup membership and task specialization. The third paper investigates *The Influence of Network Embeddedness and Multiplexity on Homicide Arrests and Re-Arrests* by analyzing how the probability of arrests for homicide is influenced by prior membership of co-arrest dyads and higher-order structures, and estimating the interdependence of co-arrests for homicide and co-arrests for other crimes. The final chapter concludes by drawing out the main results of the thesis and commenting on their implications. Challenges in the analysis of criminal networks are outlined, such as missing data, non-random sampling, the selection of appropriate models, and the integration of network and contextual data. Strategies and modeling options for dealing with these challenges are discussed.

This thesis uncovers the structure of collaboration in a drug trafficking operation and how this affects the resilience of the trafficking organization to interference, analyzes the formation of collaborative ties in complex criminal operations, and presents the first results on how the social relationships underlying arrests influence the probability of re-arrests and co-arrests.

Contents

List of Figures **ix**

List of Tables **xi**

1 Introduction **1**

2 The Structure and Vulnerability of a Drug Trafficking Collaboration Network **9**

2.1 Introduction 11
2.2 Previous research 13
2.2.1 Topological features of collaboration networks 13
2.2.2 Topological features of criminal networks 15
2.3 Data 18
2.3.1 Background 19
2.3.2 Document selection 21
2.3.3 Actor and relational coding 22
2.3.4 Limitations and bias 24
2.4 Methods 25
2.4.1 Structural analysis 25
2.4.2 Vulnerability analysis 27
2.5 Results 30
2.5.1 Structural properties 30
2.5.2 Vertex removal 35
2.6 Conclusion 42

3 A Bayesian Exponential Random Graph Model of Collaborative Tie Formation in a Drug Trafficking Network **45**

3.1 Introduction 47
3.2 Data 50
3.2.1 Actor covariates 50
3.3 Methods 52
3.3.1 Dependence in structured data 52
3.3.2 Estimating an ERGM 54

3.3.3 Bayesian ERGMs . . . 57
3.3.4 Model specification . . . 58
3.4 Results . . . 61
3.4.1 Structural parameters . . . 61
3.4.2 Individual-level parameters . . . 63
3.4.3 Homophily . . . 67
3.4.4 Model comparison . . . 69
3.5 Conclusion . . . 71

4 The Influence of Network Embeddedness and Multiplexity on Homicide Arrests and Re-Arrests 75
4.1 Introduction . . . 77
4.2 The social structure of arrests . . . 80
4.2.1 Suspect profiling based on previous arrests . . . 80
4.2.2 Co-arrest network embeddedness . . . 83
4.2.3 Contribution to the literature . . . 86
4.3 Data . . . 88
4.3.1 Limitations . . . 90
4.4 Methods . . . 92
4.4.1 Network projection . . . 92
4.4.2 Temporal network autocorrelation model . . . 93
4.4.3 Multiplex interdependence model . . . 95
4.5 Results . . . 98
4.5.1 Temporal network dependence . . . 98
4.5.2 Multiplex co-arrest interdependency . . . 106
4.6 Discussion . . . 116
4.7 Conclusion . . . 119

5 Conclusion 123
5.1 Summary and implications . . . 123
5.2 Challenges in studying criminal networks . . . 130
5.2.1 Missing data and non-random sampling . . . 130
5.2.2 The selection and application of models to network data . . . 133
5.2.3 Integrating network data, outcome variables, and contextual data . . . 137
5.3 Directions for future research . . . 141

Appendices

A Robustness of structural parameters to changes in τ 149

References 153

List of Figures

2.1 Example of the graph sampling algorithm 29
2.2 Fits to the cumulative distribution function of degree in the trafficking network 31
2.3 Size of the largest component, S, as a function of the fraction of vertices removed, f 36
2.4 The number of components, P, as a proportion of the maximum possible number, in response to the fraction of vertices removed, f . 37
2.5 Change in the number of communities U in response to the fraction of vertices removed, f 38
2.6 The shift in local clustering, c, and assortativity, r, in the graph sampling algorithm used to reshuffle the network into a more dissortative configuration 41

3.1 The Pizza Connection network 53
3.2 Structural statistics, $\boldsymbol{s}_m$, included in the BERGM 56
3.3 Estimated marginal posterior densities for the structural statistics . 62
3.4 Posterior density for 1,000 iterations of the MCMC algorithm and Traceplot of 4 chains at 100-iteration intervals from the MCMC algorithm 64
3.5 Estimated marginal posterior densities for the actor-level statistics . 66
3.6 Estimated posterior densities for the separated homophily statistics in the alternative model 68
3.7 Plots of the goodness-of-fit for the main model and alternative model 70
3.8 Posterior model probabilty of the main and alternative models . . . 72

4.1 Frequency of arrests for homicide in Chicago, 2000–2013 89
4.2 Mean count of individuals per arrest for different types of crime . . 91
4.3 Co-arrest network following endogenous stratified sampling 102
4.4 Estimates for the individual-level parameters in the temporal network autocorrelation models 103
4.5 Estimates for the network terms in the temporal network autocorrelation models 104
4.6 Multiplex edges in the Chicago co-arrest network, 2000–2013 107

4.7 Parameter estimates for the first group of multiplex interdependence models . . . 112
4.8 Parameter estimates for the second group of multiplex interdependence models . . . 114

A.1 Mean posterior estimates for the GWESP and GWDSP statistics in the main model, with varying τ_{gwdsp} . . . 151
A.2 Mean posterior estimates for the GWESP and GWDSP statistics in the main model, with varying τ_{gwesp} . . . 152

List of Tables

2.1 The clustering coefficient c, degree assortativity r, and degree assortativity error σ_r in several collaboration and criminal networks . . . 16
2.2 Parameters for the competing models of the degree distribution . . 32
2.3 Comparison of alternative models of the degree distribution 34

3.1 Membership by subgroup and task specialization in the Pizza Connection Network 51

4.1 Summary of data in the temporal network autocorrelation model . . 100
4.2 Summary of the layer co-arrest networks in the multiplex interdependence models 109

1 Introduction

Research into criminal networks has proliferated in the last fifteen years (Morselli, 2009; Papachristos, 2011a; McGloin and Kirk, 2010). Indeed, much of this research has focussed on collaboration in the form of co-offending ties (Mastrigt and Carrington, 2013; McGloin and Piquero, 2010; Schaefer, 2012) and in organized criminal activity (Natarajan, 2006; Varese, 2013; Campana, 2015; Campana, 2016; Morselli, 2005)

However, despite the "network turn" in criminology, there have been few attempts to integrate research on collaboration in criminal networks with research on collaboration in the broader social networks literature. One of the most robust findings in the networks literature is that collaboration networks exhibit certain shared characteristics, such as local clustering and a heavy-tailed distribution of degree (Barabási, Jeong, et al., 2002; Ramasco, Dorogovtsev, and Pastor-Satorras, 2004). So well supported are these findings that social network theorists have sought to develop universal generative principles, such as preferential attachment, homophily, and transitivity, to explain why social groups so often self-organized with such consistency (Watts and Strogatz, 1998; Barabási and Albert, 1999; Newman, 2003b; Newman, 2005).

Indeed, research on collaboration in criminal networks has largely bypassed these findings, either assuming that they extend to criminal networks or, conversely, that

criminal groups organize in a manner distinct from non-criminal groups (Williams, 2001; Albanese, 2011). Clearly, the idea that criminal networks exhibit exceptional structural features is discordant with the orthodoxy that collaboration networks trend towards common structural features, and the assumption that these features extend to criminal networks requires empirical support.

It is thus important to revisit the structure of criminal networks for at least two reasons. First, if criminal networks indeed exhibit features that are not widely observed in social networks, this would challenge social network theory. Second, by failing to establish the fundamental structural features of criminal collaboration networks empirically, it is difficult to reconcile social network theory, with its focus on the underlying structural and inter-personal tendencies towards certain collaborative structures (Ramasco and Morris, 2006), with criminological explanations of collaboration in criminal settings, that instead focuses on problems of distrust Gambetta, 1988, credible commitment (Campana and Varese, 2013), and secrecy (Erickson, 1981).

To address this discord, the first paper (*The Structure and Vulnerability of a Drug Trafficking Collaboration Network*) in this thesis examines the structure of collaboration in a drug trafficking network, comparing its features to those of several non-criminal collaboration networks that are routinely studied in the social networks literature. In line with several previous studies on criminal networks (Morselli, 2010; Varese, 2013; Campana, 2015), I draw on data from archival court records that document collaboration events over a six-year period. I find evidence that the drug trafficking network exhibits two of the three most common structural characteristics in non-criminal collaboration networks, supporting the assumption that criminal networks are not fundamentally distinct. However, I also find evidence that the trafficking networks diverge from broader collaboration networks in one of the three features: degree assortativity.

Although the paper sets out to revisit structural features of criminal collaboration specifically with reference to non-criminal collaboraton, it also provides evidence relevant to the debate over whether criminal networks are flat in nature. Several

scholars have argued that, in theory, the structure of criminal organizations should be "non-hierarchical" (Shelley, 2006) and decentralized (Williams, 2001). Responding to the "flat structure thesis," Varese (2013) points out that criminal organizations could not maintain such structures while simultaneously maintaining authority structures, which are routinely observed in organized criminal groups, and while carrying out complex tasks. In examining the structural properties of the trafficking network, the first paper presented here finds evidence against the flat thesis.

Aside from the fundamental structural features of criminal networks, the rise of network analysis in studies of organized crime has led to debate over the extent to which criminal networks organize around the principle of maintaining security (Baker and Faulkner, 1993; Erickson, 1981), such that individuals may reduce their risk of exposure to law enforcement agencies, versus maintaining efficiency (Krebs, 2002), so that individuals can successfully execute criminal operations, what (Morselli, Giguére, and Katia, 2007, p. 144) calls the "efficiency/security trade-off." As Morselli argues, the risky conditions under which criminals operate should lead them to prioritize security, but at some stage, the demands of carrying out complex criminal tasks will necessitate some compromise. The first paper analyzes the extent to which the drug trafficking network is vulnerable to intervention, should the security of the network be breached. To do this, the paper uses a graph sampling method to create simulated variants of the trafficking network with different structural features and measure the impact of targeted node removals on these variants, comparing the effects to targeted node removals on the observed trafficking network. I find that the trafficking network is more vulnerable to targeted node removals than one would expect in a non-criminal collaboration network due to its dissortative degree mixing. Thus, the paper also makes a broader contribution on the type of network features that one should expect to be associated with higher and lower levels of vulnerability.

Expanding on the results in the first paper, the second paper (*A Bayesian Exponential Random Graph Model of Collaborative Tie Formation in a Drug Trafficking Network*) in this thesis analyzes the factors that influence the formation of collaborative ties within the trafficking network. In the networks literature, it has

long been held that tie formation in social networks is influenced by homophily effects, which represent the tendency for individuals with similar salient characteristics to preferentially form ties with each other (McPherson, Smith-Lovin, and Cook, 2001), and shared partner effects, which capture the influence of mutual third parties on tie formation, as in the notion that a "friend of a friend is my friend" (Snijders, Pattison, et al., 2006). However, most studies within the criminology literature that aim to explain collaborative patterns between individuals do not account for these effects. This could lead to important partner selection effects being overlooked. Additionally, as Grund and Densley (2015) note, failing to account for well-known network effects could result in correlations between the characteristics of individuals that do and do not share collaborative ties being overstated. For example, if two individuals with different attributes are not engaged in collaboration, it might simply be the case that they do not share mutual co-conspirators or that they occupy different regions of the network, rather than the result of deliberate choice driven by trafficking strategy.

In the second paper, I therefore set out to disentangle the structural effects and trafficking-driven selection effects that influence the formation of collaborative ties among drug traffickers. The process of tie formation is particularly difficult to observe in covert networks. In the absence of longitudinal data, the study evaluates the network data used in the first paper using Bayesian exponential random graph models (BERGMs) to infer structural and selection effects (Caimo and Friel, 2011). Varese (2013) and Campana and Varese (2013) highlight the diversity of tasks performed in organized criminal operations. I find strong evidence in support of this view. Although specialization homophily has no aggregate effect, heroin suppliers and money launderers are more likely to form within-specialty collaborative ties, whereas heroin distributors are more likely to form between-specialty ties. Moreover, there are strong shared partner effects, suggesting that mutual third parties may be a conduit for overcoming the low trust environment engendered by illicit activity. The paper contributes to the literature on the determinants of collaboration in organized criminal groups.

Methodologically, the second paper demonstrates the utility of BERGMs for analyzing criminal networks. Unfortunately, it is often the case that criminologists do not have access to longitudinal data on criminal networks, and thus the process of tie formation cannot be observed directly, placing greater emphasis on the use of models that infer this process from cross-sectional or static data. The most popular models within this class are exponential random graph models. However, criminologists may find that ties are particularly sparse relative to typical social networks due to the nature of the criminal activity under study, or due to the use of unconventional sources of data. This could contribute to model degeneracy, where an ERGM either does not converge or fails to converge on reasonable parameter estimates. Under the BERGM approach, the analyst can specify a prior distribution on the coefficient and variance of each parameter in the model, enabling fine-tuning of the model. The thesis contributes to the applied literature on criminal networks by discussing dependency in network data, outlining issues of degeneracy with reference to criminal networks, and demonstrating a method that may improve the prospects of convergence and therefore prove useful to researchers investigating criminal networks whom are encountering degeneracy.

To expand the contributions of the thesis beyond the domain of drug trafficking, the third paper (*The Influence of Network Embeddedness and Multiplexity in Homicide Arrests and Re-Arrests*) investigates criminal behavior more broadly on a larger, city-wide scale. This paper is distinct from the first and second papers in three important ways. Firstly, rather than exploring the structure and formation of ties, the third paper models an individual-level outcome: the incidence of arrest. Secondly, the third paper draws on thirteen years of arrest data in Chicago to build a longitudinal co-arrest network, enabling the analysis of temporal dynamics in criminal behavior. In this way, the paper examines the extent to which future individual-level outcomes are influenced by the structure of criminal networks, shifting emphasis to the interplay of social relationships and behavior, rather than exploring the dyadic-level outcome captured by the presence of ties. Thirdly, whereas the first and second papers focus on a specific type of criminal

activity—collaboration related to drug trafficking—the third paper explores several criminal behaviors, including violence-involved crime, such as homicide, weapons violations, and lower level drug offenses. In this way, the paper broadens the overall contribution of the thesis by analyzing criminal behaviors that are considerably more prevalent than multinational drug trafficking and by demonstrating that the structure of ties is substantively important for understanding these behaviors.

Ever since Sutherland's (1947) theory of "differential association," in which he argued that the adoption and maintenance of criminal behavior, just as non-criminal behavior, is engendered by social interaction and peer-groups, social structure has been a cornerstone of criminological theory. Despite the emphasis that such foundational theory, as well as more contemporary theory (Venkatesh, 1997; Hagan, 1993; Wright et al., 2001), places on the contribution of social ties to criminal behavior, there is little empirical research on the role that social networks play in persistant contact with the criminal justice system, reoffending, and recidivism—empirical topics in the mainstream of contemporary criminology.

The third paper addresses this gap by investigating the influence of embeddedness in networks of co-arrest ties on arrests and re-arrests for homicide vis-a-vis non-homicide crimes. Co-arrests are defined as events involving two or more individuals arrested on the same charge. Over time, repeated arrests on the same individuals and co-arrests involving groups with varying composition amalgamate into a large, complex network of co-arrest ties. The paper argues that co-arrest ties are not merely artifacts of previous arrests, but rather that these ties represent underlying social relationships. Such relationships are an important factor in differential patterns of arrest and re-arrest for homicide versus non-homicide crimes because they form part of an individual's criminal opportunity structure.

The study provides an empirical strategy for evaluating the influence of social ties on the probability of sustaining re-arrests in a network setting. The strategy involves the use of temporal network autocorrelation models to estimate the extent to which the incidence of homicide arrest in a one-year period is influenced by embeddedness in the co-arrest network in the previous year, and the use of multiplex

interdependence models to capture the interdependence of co-arrests for homicide and non-homicide crimes.

The third study thus contributes to the literature by providing empirical support to long held theories that social ties influence patterns of criminal behavior. I find that the probability of sustaining a homicide arrest is smaller for individuals that are members of co-arrest dyads and transitive co-arrest triads. However, the results are not unambiguous, at least over short time intervals. For instance, membership in a co-arrest tie with an individual that sustained a homicide arrest had little influence on the odds of sustaining a homicide arrest in the following year.

Finally, the fourth chapter concludes by summarizing the main results of the thesis and commenting on their implications. I discuss several methodological challenges in the study of criminal networks, including those encountered in this thesis, such as the problem of missing data, non-random sampling, boundary specification, model selection, and integrating network and contextual data. This part of the thesis highlights several strategies and modeling options for dealing with these challenges, including emerging methods that are presently under development. I close the thesis with a discussion of directions for future research.

2

The Structure and Vulnerability of a Drug Trafficking Collaboration Network

Contents

2.1 Introduction 11
2.2 Previous research 13
2.2.1 Topological features of collaboration networks 13
2.2.2 Topological features of criminal networks 15
2.3 Data 18
2.3.1 Background 19
2.3.2 Document selection 21
2.3.3 Actor and relational coding 22
2.3.4 Limitations and bias 24
2.4 Methods 25
2.4.1 Structural analysis 25
2.4.2 Vulnerability analysis 27
2.5 Results 30
2.5.1 Structural properties 30
2.5.2 Vertex removal 35
2.6 Conclusion 42

Abstract[1]

Prior research has established that collaboration networks exhibit heavy-tailed degree distributions, assortative degree mixing, and large clustering coefficients. Using court record data, I assess these properties in a collaboration network among heroin traffickers. Consistent with previous results, I find an exponential degree distribution and strong local clustering. However, the traffickers mix dissortatively by degree rather than assortatively. Using a graph sampling method, I show that a consequence of dissortative mixing is that targeted vertex removals have a greater impact on the connectivity and cohesion of the trafficking network. I also note the importance of degree mixing for characterizing and identifying topological weaknesses.

[1]A version of this paper will appear in: Wood, G. (2017) "The Structure and Vulnerability of a Drug Trafficking Collaboration Network," *Social Networks*, 48: 1–9.

2.1 Introduction

Many criminal endeavors are complex and demand a large degree of collaboration and coordination. Criminals rely on each other for tasks that cannot be performed alone, such as the trafficking of drugs and humans, money laundering, fixing sports events, and rigging markets. In this regard, however, criminals are not unique. After all, many non-criminal endeavors are just as complex. What makes criminal collaboration extraordinary is that, in most cases, criminals operate under the constant threat of detection and the punitive consequences that this entails.

Much research has focused on how these extraordinary conditions shape the structure of collaboration in criminal networks. For instance, it has long been proposed that, in order to remain covert, criminal networks adopt particular topological features, such as greater sparsity and decentralization (Baker and Faulkner, 1993; Krebs, 2002; Kenney, 2007; Williams, 2001). Further work has addressed the trade-off between efficiency and security in criminal networks (McCormick and Owen, 2000; Morselli and Petit, 2007; Lindelauf, Borm, and Hamers, 2009). This work has, in part, motivated research into the vulnerability and disruption of criminal networks (Ayling, 2009; Duijn, Kashirin, and Sloot, 2014; Morselli, 2010; Malm and Bichler, 2011).

Meanwhile, there has been a great deal of interest in the structure of collaboration networks in non-criminal domains, such as academia, music, and business. It is recognized that such collaboration networks typically share three properties in common (Ramasco, Dorogovtsev, and Pastor-Satorras, 2004). First, the vertex degree distribution is heavy-tailed (Newman, 2001b; Barabási, Jeong, et al., 2002; Guimerá et al., 2005) such that a minority of actors have a disproportionately large number of collaborators. Second, the degree of connected vertices is positively correlated (Newman, 2005; Chang et al., 2007). Well-connected actors tend to work with other well-connected actors, while the poorly-connected work with the poorly-connected. Third, collaboration networks exhibit strong local clustering. Actors who share at least one mutual partner are more likely to become partners themselves,

whether as a matter of choice or chance, than would be the case if vertices were randomly connected (Watts and Strogatz, 1998; Newman, 2001a; Moody, 2004).

However, although these properties are routinely examined in collaboration networks outside of the criminal networks literature, whether collaboration among criminals shares the same three properties has not been explicitly addressed. Although studies on criminal networks often consider degree and clustering coefficients, they seldom characterize the degree distribution formally, and degree mixing is typically overlooked entirely.

Yet, there are good reasons to be interested in whether collaboration between criminals exhibits the same topological properties as collaboration elsewhere. First, such comparison provides a systematic means of identifying irregularities in the structure of criminal networks. Research on clandestine networks has drawn criticism for lacking empirical benchmarks against which to assess whether the observed properties are, in fact, exceptional (Varese, 2013; Crossley, Edwards, and Stevenson, 2012). By drawing on empirical results in the wider literature on collaboration networks, such benchmarks are made explicit.

Second, the combination of the topological properties, and in particular degree mixing, are interesting in their own right. It has long been known that assortatively mixed networks are more robust to the removal of high degree vertices, whereas this strategy is considerably more effective in dissortative networks (Newman, Forrest, and Balthrop, 2002). In light of growing research interest in disrupting criminal networks, degree mixing is therefore a property of particular relevance.

In this paper, I begin to address this gap by examining the degree distribution, degree mixing, and clustering of a heroin trafficking network. I remain agnostic about how these properties should manifest in the network, and instead simply evaluate each property in turn. I find that the trafficking network has a heavy-tailed degree distribution, which may be classified as either exponential or log-normal. Additionally, the network exhibits strong local clustering. However, unlike other collaboration networks, the traffickers mix dissortatively by degree. Traffickers

with a large number of collaborators tend to be connected to traffickers with relatively few, and vice versa.

The dissortative mixing has substantial consequences for the vulnerability of the network to targeted vertex removals. I demonstrate this by analyzing the impact of vertex removals on the heroin trafficking network under different mixing configurations using a graph sampling method. In variants of the trafficking network which are more assortative, but where the degree distribution is preserved, the impact of targeted vertex removals on the cohesion and connectivity of the network diminishes. Conversely, targeted vertex removals have a greater impact in more dissortative variants of the network. Thus, the property on which the trafficking network deviates from other types of collaboration network appears to be critical for its robustness to attacks.

The remainder of this chapter is structured as follows. In section **??**, I briefly review previous research on the structure of collaboration networks and criminal networks. In section 2.3, I describe the data. In section 2.4, I outline the methods for evaluating each property in the network, along with the graph sampling method for generating networks with varying degree mixing configurations, and the measurements used to assess the impact of vertex removal. In section 2.5, I present the results, before concluding in section 2.6.

2.2 Previous research

2.2.1 Topological features of collaboration networks

Studies on the structure of collaboration networks abound. Here, I focus on three topological features that feature regularly in these studies: clustering, a heavy-tailed degree distribution, and assortative degree mixing. In a well-known study, Watts and Strogatz (1998) used data on the co-appearence of actors in movies and found that when an actor i has appeared in a movie with another actor j, and i has also appeared with a third-actor k, it is far more likely than one would expect by chance that j and k will have also appeared in a movie together. This clustering

phenomenon is now a routine finding in social networks, and is a feature of co-authorship among scientists (Newman 2001 Newman 2001) and social scientists (Moody, 2004), in collaboration among jazz musicians (Gleiser and Danon, 2003), and in the Broadway musical industry (Guimerá et al., 2005).

Second, collaboration networks often exhibit a heavy-tailed distribution of degree. Newman (2001a) demonstrated a heavy-tailed distribution in co-authorship networks among physicists, biomedical researchers, and computer scientists. Across disciplines, a minority of authors have a disproportionately large number of co-authors, while the majority have few. The heavy-tail in degree persists outside of scientific co-authorship. For example, collaboration among jazz musicians and in the Broadway musical industry also exhibit a heavy-tailed distribution of vertex degree. Moreover, this property holds for collaboration networks among organizations. Butts, Acton, and Marcum (2012) show that a network of inter-organizational collaboration, developed in response to Hurricane Katrina, has a characteristic right-skewed distribution of degree, while collaboration among firms in the biotechnology industry is characterized by a small number of high-degree firms and much larger group of poorly-connected firms (Gay and Dousset, 2005).

Barabási and Albert (1999) propose that preferential attachment is part of the explanation for this pattern of inequality. Preferential attachment is often associated with a power law in the tail of the degree distribution (Barabási and Albert, 1999; Albert, Jeong, and Barabási, 1999), such that the probability that an actor has precisely x collaborators abides by $P(x) \sim x^{-\alpha}$, where the α scaling parameter relates to the extent of inequality in the number of collaborators per vertex. This work, along with Huang et al. (2008) and Moody's (2004) research on academic collaboration, point to a power law in the tail of degree. Moreover, using dynamic data, Newman (2005) shows that the growth of a co-authorship network is consistent with preferential attachment.

The third property typical of collaboration networks is the positive correlation in the degree of connected actors. Newman (2003a) demonstrates that the partners of academics, movie actors, and businesspeople who have large vertex degree are likely

to have many partners themselves, calling this property "degree assortativity." Chang et al. (2007) and Ramasco and Morris (2006) show the degree-degree correlation holds in the movie actor network for both weighted and unweighted degree, while Ramasco, Dorogovtsev, and Pastor-Satorras (2004) finds assortative degree mixing in company co-directorships. I reanalyze some of the aforementioned networks and report the local clustering coefficient c, assortativity r, and standard error for the assortativity coefficient σ_r in Table 2.1.[2]

All of these networks are, of course, qualitatively different in kind. The scientific collaboration networks are made up of thousands of actors partitioned into smaller co-authoring teams, where the high-performing teams cluster into a large component (Newman, 2001a; Guimerá et al., 2005). In contrast, those in the Broadway musical industry are divided into somewhat larger teams on a show-by-show basis, where actors cross-over or belong to several teams (Guimerá et al., 2005). Movie actors and jazz musicians assemble around records and movies, respectively. However, in spite of these differences, all of the networks exhibit the clustering, heavy-tailed degree, and assortativity characteristics. Whatever the specific purpose of collaboration, whether producing jazz records, co-authoring papers, or operating a business, similar patterns emerge. A minority of actors represent a large portion of the collaborative activity, well-connected actors tend to work with other well-connected actors, and cooperation is more likely between actors who share a mutual partner.

2.2.2 Topological features of criminal networks

There has been a concerted effort during the last two decades to understand the structure of criminal organizations (Sparrow, 1991; Coles, 2001; Lampe, 2009). However, the role of networks in organized crime has been conceptualized in two contrasting ways. Campana (2016) refers to these conceptualizations as the "substantive" and "instrumental" perspectives. The substantive perspective, Campana (2016) argues, views networks as a fundamental organizing principle. For example, it is proposed that criminal groups vary from hierarchical structures to

[2]The network data used in Table 2.1 were obtained from the Network Data Repository (Rossi and Ahmed, 2015).

Table 2.1: The clustering coefficient c, degree assortativity r, and degree assortativity error σ_r in several collaboration and criminal networks.

	Network	c	r	σ_r
a	Network Science	0.638	0.462	0.072
b	Condensed Matter Physics	0.636	0.186	0.004
c	High Energy Physics	0.442	0.294	0.019
d	Astrophysics	0.639	0.235	0.005
e	Jazz	0.617	0.020	0.026
f	Students	0.636	0.366	0.113

network forms of organization, where network is conceived as a type of organizational form that is more "fluid" (Shelley, 2006) and flexible (Williams, 2001). The instrumental perspective, on the other hand, conceptualizes networks as a set of nodes and edges. This more conventional understanding translates directly from the social networks literature (Papachristos, 2011b; Papachristos, 2014). Importantly, the instrumental approach does not assume that criminal groups take on any particular organizational structure, but rather it treats the structure as a target of inference to be explained and understood (Campana, 2016; Morselli, 2009). In this paper, I take the instrumental approach to understanding the structure of organized crime groups.

Indeed, there have been several notable studies into the structure of organized criminal networks utilizing the instrumental approach. Previous research demonstrates that skewed degree distributions are a characteristic feature of criminal networks. Varese (2013) finds a heavy-tailed degree distribution in a Russian Mafia group, while Krebs (2002) and Morselli, Giguére, and Katia (2007) show that degree varies considerably in a terrorism network. Campana (2015) finds that degree is heavy-tailed in a human trafficking organization operating in Nigeria and Europe (Campana, 2015; Campana, 2016). In addition, Duijn, Kashirin, and Sloot (2014) shows that degree follows a power law distribution in a large co-offending network of cannabis cultivators, and Qin et al. (2005) finds a truncated power law in a terrorism network. In comparable studies on drug trafficking networks, Natarajan (2006) finds that a small number of disproportionately high degree traffickers among a majority with low degree, and Bright, Greenhill, et al. (2015) finds a skewed distribution of degree.

However, it is common for criminal network research to focus on identifying key players (Ballester, Aremngol-Calvó, and Zenou, 2004; Lindquist and Zenou, 2014), rather than classifying network topology. This focus is perhaps unsurprising given the policy implications and substantive interest in understanding the roles of influential criminals. Nevertheless, since key player analyses are less common elsewhere in the social networks literature, many of the empirical findings on criminal

networks are not directly comparable to their social network counterparts. Thus, although degree is nearly always part of the analysis, it is often tabulated at the actor-level or truncated to include only highly central actors. When the distribution of degree is visualized, it is rarely classified formally. Similarly, higher-order structural statistics, such as clustering, are less frequently considered.

Most notably, degree mixing is typically overlooked as a property of interest. This may owe to the focus on key players, where statistics such as eigenvector centrality are more often used to account for the connectivity of actors' neighbors (Catanese, Ferrara, and Fiumara, 2013). However, there are several reasons to be interested in degree mixing. Firstly, it provides a relatively simple summary of the tendency for actors to collaborate with other members of the network with similar degree, and is therefore indicative of the internal organization of criminal groups. Secondly, it has been proposed that assortative mixing is one of two properties, the other being high levels of clustering, that distinguishes social networks from other kinds of networks (Newman, 2003b). Should there be criminal networks that do not exhibit this property, it may inform our understanding of the underlying mechanics of tie formation in criminal organizations. Thirdly, degree mixing has long been associated with the vulnerability of networks to actor removal (Newman, Forrest, and Balthrop, 2002; Holme and Zhao, 2007), which is particularly relevant to criminal networks. Below, I investigate these properties in the case of a drug trafficking network, which I introduce next.

2.3 Data

A major challenge in studying criminal networks lies in obtaining reliable and informative data. Standard tools like name generators are impractical for criminal networks, while observing clandestine groups is, by definition, not ordinarily an option. However, law enforcement agencies dedicate considerable resources to the surveillance of criminal groups, and parts of the information these agencies collect are often documented in court records. Archival court records have therefore become a substantial source of data on criminal networks (Natarajan, 2006; Morselli

and Petit, 2007; Varese, 2013; Campana and Varese, 2012; Campana and Varese, 2013; Berlusconi, 2013; Campana, 2015). The materials used in this paper were produced during the criminal trial of a group of actors involved in a large-scale heroin trafficking operation.

2.3.1 Background

Before turning to the data and coding strategy, I describe the background to the trial and the context in which collaboration emerged. In the late 1970s, the Bonanno family, one of the five mafia groups operating in New York, sought to expand into the lucrative heroin industry. Charged with leading this expansion was Salvatore Catalano, the ambitious head of a faction of Sicilian-born mafiosi within the Bonanno family. To the wider Bonnano family, the members of Catalano's faction were known as the "zips," a term used by the predominately second-and third-generation members of the family as a derogatory slur against more recent migrants. However, it was Catalano's ties to his native Sicily that proved critical to establishing the heroin importation pipeline, as Catalano and his faction leveraged their connections to a Cosa Nostra group based in Ciaculli, Sicily. Salvatore Greco, the leader of the Ciaculli group, sourced multi-kilogram quantities of morphine base in Turkey, from where it was transported to Sicily and refined into heroin. In coordination with Catalano, Greco arranged for the heroin to be shipped into ports in New York, New Jersey, and Philadelphia.

During the early stages of the trafficking operation, the source of the heroin pipeline confounded federal investigators. However, a major clue emerged in June 1979, when aiport border agents in Sicily intercepted a suitcase containing $497,000 wrapped in pizzeria aprons (Alexander, 1988). Bouyed by the recent expansion of prosecutorial powers under the Racketeer Influenced and Corrupt Organizations Act (RICO), the FBI expanded its investigation and pursued warrants for a large scale surveillance campaign that would ultimately result in the interception of more than 55,000 wiretapped telephone conversations. By mid-1983, investigators had pieced together the major elements of the Turkey-Sicily-East Coast pipeline when,

in August, the Drugs Enforcement Agency sucessfully purchased 2.5 kilograms of heroin from the owner of a pizzeria in Philadelphia. This breakthrough was shortly followed by another: investigators observed a meeting between a known figure within the Bonanno family, whom was under heavy surveillance, and three Midwest-bound passengers at La Guardia Airport, New York. Further investigation revealed the three individuals to be pizzeria owners in Illinois, Wisconsin, and Michigan, each under the command of Gaetano Badalamenti, the former head of the Sicilian Cosa Nostra commission who had become a fugative.

When charges were brought in 1984, the case was dubbed the "Pizza Connection," owing to the use of pizza parlors as fronts for the heroin distribution operation and the role that pizzerias had played in uncovering the trafficking network. The trial lasted from September 1985–May 1987 and featured twenty-two defendants, rendering it among the largest in US criminal history. Prosecutors delivered evidence pertaining to the importation and wholesale distribution of heroin in New York, New Jersey, Philadelphia, and the Midwest, with a valuation of more than 330 million dollars, in addition to a money laundering operation leading to cash deposits in Swiss bank accounts amounting to more than 60 million dollars. The lead prosecutor later described the Pizza Connection as "the FBI's first major, transnational criminal enterprise investigation and prosecution" (Rooney and Schiliro, 2016, p. 21).

Collaboration in the Pizza Connection trafficking operation thus occurred with the very deliberate aim of establishing a long-term infrastructure to import and distribute heroin in huge quantities. To be successful in this endeavour, the traffickers were required to carry out complex tasks including sourcing the raw products required for manufacturing heroin, negotiating prices, transporting narcotics across national borders, distributing heroin to secondary dealers in multiple US cities, and laundering the cash proceeds into legally transferable tender. In short, the collaborative activity required to establish and maintain the trafficking operation is comparable in scale to that of any organization, whether legal or illicit, operating across multiple stages of a sophisticated supply chain. However, unlike organizations in the legal economy, the traffickers faced the additional complications associated

with remaining covert and working in an unregulated and, at times, violent industry (Varese, 2013). It is within this context that the traffickers forged the collaborative relationships that I investigate in this paper.

2.3.2 Document selection

The source document for the network data was selected from a large corpus of documents produced during the Pizza Connection trial. This corpus is publicly available through the United States National Archives and Records Administration. The source document is the Government's Sentencing Memorandum (GSM): a 107-page summary of the prosecution's evidence against the defendants produced towards the conclusion of the trial. The memorandum details evidence on the supply, distribution and laundering activities of the network, in addition to the activities and roles of each individual, with whom they collaborated, and the purpose of that collaboration. To build this body of evidence, the FBI carried out an extensive wiretapping program on the public, business and domestic telephone calls of the network's members, intercepted shipments and samples of heroin, carried out surveillance on meetings, searched the defendants households, businesses and property, and used the testimony of witnesses. Futher information on the activities of the agents in Italy was provided by Italian law enforcement agencies.

In the majority of cases, the memorandum describes the basis of each particular collaboration event. For example, the document describes an event in which Salvatore Salamone used "his own $1,500 cash" to "fraudulently purchase powerful, semi-automatic weapons which he turned over to Ganci, Castronovo, and Greco." In another example, Fillipo Casamento "purchased multikilogram quantities of high-purity heroin from Ganci for sale to his own customers." While these two examples provide contextual detail for the purpose of the collaboration events, such detail is not always presented. Indeed, for a minority of collaborative events, the details may be as vague as "the money laundering evidence proved that Castronovo worked hand-in-hand with Catalano in New York." Nevertheless, the existence

or non-existence of a collaborative relationship between any two given actors is typically unamiguous within the text.

The rationale for selecting one document from the wider corpus is to twofold. First, the corpus is extremely large and almost entirely non-digitized. Coding the entire corpus would therefore involve manually working through hundreds of thousands of pages. Even automating the coding process using, for instance, researcher-generated search terms or natural language processing, would first require digitizing the documents. Selecting a subset of the documents is therefore the only practical way to make the data extraction manageable.

Second, there is no straightforward way to sample the documents that would provide both sufficient coverage and consistency across sources. The indexing of the corpus does not describe the contents of the documents in detail. Many of the documents, particularly in the early stages of the trial, are stenographers minutes that describe various procedural arrangements such as juror selection and contain no information relevant to the network. In order to sample the documents therefore, one would have to draw from the entire corpus either randomly, on the basis of date, or document type. In either case, the documents may have been produced by different parties, such as the defense, prosecution, or the court itself, and would thus be subject to different types of bias or may simply be irrelevant.

The GSM has the advantage of being both concise and consistent, and it is therefore practical to extract a large quantity of relevant data and employ a coding method that is fine-tuned to the document. Finally, any bias in the data has greater tractability insofar as it is reasonable to expect sources of bias to be consistent within-document.

2.3.3 Actor and relational coding

The data are an unweighted, undirected one-mode network of contacts between the individuals involved in the Pizza Connection heroin trafficking network. These contacts, which are the network's edges, cover the full range of activities of the drug trafficking network. More specifically, the contacts cover the supply chain

of heroin from the bulk purchase of base morphine, the refinement and shipment of heroin into the United States, to distribution at the wholesale and secondary wholesale level, in addition to the laundering of cash proceeds.

The coding of the data proceeded in two steps. The first step involved a census of the actors named in the GSM. The GSM refers exclusively to actors involved in trafficking activities and does not refer to witnesses or other individuals related to the case. In total, 82 traffickers were identified in the document. Information on group affiliation is provided in the memorandum and is treated as an agent-level attribute. Although these groups operated primarily in separate locations, they were required to maintain contact between one another in order to orchestrate the trafficking of narcotics and the distribution of proceeds. These edges can be thought of as between-group edges, whereas contacts between members of the same group are within-group edges.

The second step involves extracting the collaborative edges among the actors. Collaboration is defined as an event in which two or more actors interact with each other for the purpose of a drug trafficking task, which broadly fall into three types: sourcing and arranging the supply of heroin, distributing the heroin, and laundering cash proceeds. Time resolution is unavailable for most collaboration events. Although the GSM reports that some actors were involved in several collaboration events together, as discussed below, weighting the edges may induce bias due to the non-random inclusion of collaboration events in the GSM. Thus, the edges are unweighted. In total, there are 432 edges in the network.

Similar coding methods are used in other studies on criminal networks (Morselli, Giguére, and Katia, 2007; Natarajan, 2006; Morselli and Petit, 2007; Varese, 2013; Campana and Varese, 2012; Campana and Varese, 2013; Berlusconi, 2013; Campana, 2015; Papachristos and Smith, 2014) and for other types of social network that are difficult to observe (Butts, Acton, and Marcum, 2012).

2.3.4 Limitations and bias

The data in this chapter derive from a law-enforcement investigation. As such, the coverage of actors and edges depends, in part, on the focus of the investigation and may therefore be affected by sample selection bias. This could manifest in two ways. First, due to the nature of criminal investigations, the sampling of actors may more closely approximate a snowball sample than a random sample. In a criminal investigation, it is typical for a subset of actors to be placed under surveillance whom subsequently lead investigators to other actors as the investigation expands. This may affect the findings in the current study to the extent that the degree of initial targets is inflated relative to later targets of the investigation. Second, the inclusion of collaboration events may be influenced by the extent to which each event supports the prosecution's case against the defendants. It is possible that some collaboration events were omitted, leading to the degree of some actors being underrepresented.

The selected investigation and the coding strategy are designed to mitigate these possible sources of bias. First, a central plank of the investigative strategy was to secure prosecutions through the Racketeer Influenced and Corrupt Organizations Act. Under RICO, it is necessary to demonstrate the existence of a long-term structure (Lynch, 1987, p. 703).[3] Accordingly, the investigation sought to collect evidence of connections between members of the trafficking network so that prosecutors could tie these individuals to the broader criminal enterprise. Hence, while early targets of the investigation may initially be placed under greater investigatory focus, the RICO strategy depended on uncovering the collaborative ties of actors who were identified at later stages of the investigation. This focus on identifying ties throughout the investigation should reduce any inflation in the degree of initial targets vis-a-vis latterly identified targets.

Second, I restrict the coding of edges in the GSM to collaboration events that relate directly to drug trafficking activities and I treat the edges as unweighted. Thus, I am concerned only with events that are eligible for inclusion in the GSM, and an omitted event would not affect the findings unless that event is the only

[3]See Lynch (1987) for a review of the RICO act.

evidence of a tie between two actors. In principle, removing such events would weaken both the evidence that an enterprise exists and the evidence tying these actors to that enterprise, thus jeopardizing RICO prosecution. Moreover, in order for event omission to substantially affect the shape of the degree distribution or assortativity coefficient, it is necessary that events including particular actors are systematically selected against. It is not clear that this would benefit the prosecution. Nevertheless, I cannot rule out the possibility that collaboration events involving actors viewed as a priority for prosecution are selected for, which could affect the shape of the degree distribution. Unfortunately, non-random sampling is a feature data from criminal investigations, which remain one of the few detailed sources of information on criminal networks.

2.4 Methods

The analysis is formed of two parts. First, I examine the degree distribution, clustering, and degree mixing in the network. Second, I analyze the vulnerability of the network to vertex removals (i) in its observed state, and (ii) under several degree mixing configurations.

2.4.1 Structural analysis

Classifying the degree distribution

In order to classify the degree distribution in the trafficking network, I fit power law, log-normal, exponential, and Poisson models to the distribution using maximum-likelihood methods (Clauset, Shalizi, and Newman, 2009). The fits proceed as follows. Each observed vertex degree, k, in the trafficking network represents a candidate threshold value, k_{min}, above which the scaling behavior associated with each model may plausibly begin. Then, for each model, I find the associated parameter values which maximize the fit of that model to the degree distribution above each candidate threshold value. I then select the power law, log-normal, exponential, and Poisson model with the best overall fit, given its corresponding threshold and parameter values.

To test whether each model is plausible given the data, I synthesize 2,500 degree distributions from a "true" version of each model with the threshold and parameter values equal to those estimated for the observed degree distribution. I then refit the models, independently, to each of the synthetic degree distributions. The models are ruled out if 10% (i.e. $p < 0.10$) or fewer of the fits to the synthetic data are poorer than the best fit to the real data.[4] Finally, I compare the log-likelihood ratio of the degree distribution, given each model, in pairwise fashion. I thus arrive at a classification of the degree distribution by considering several competing models.

Clustering

There are several measures of clustering in networks. Here, I use the clustering coefficient, c, defined by Watts and Strogatz (1998). In short, the coefficient is based on the ratio of the edges that exist between a vertex and its neighbors to all of the possible edges that could exist. Averaging across all vertices gives the coefficient, which varies between 0 and 1, and captures the average local connectivity or "cliquishness" (Watts and Strogatz, 1998, p. 440) of the network.

Degree mixing

The final property I examine is the vertex-vertex degree correlation. A network is assortatively mixed when this correlation is positive, i.e. high degree vertices are connected to other high degree vertices, and dissortative when the correlation is negative. Here, I calculate the assortativity coefficient, r, defined in Newman (2003a). The intuition behind this coefficient is that, denoting the degree distribution of the trafficking network p_k, I select an edge and then take the degree of one of the vertices, i, incident to that edge. I then calculate the excess degree of i, which is the degree of i relative to our expectation given p_k. If the excess degree of i is frequently large as I move from one edge to the next, then a relatively large number of edges must end at a high degree vertex. Conversely, if the excess degree

[4]Further details of this procedure are outlined by Clauset, Shalizi, and Newman (2009) and Virkar and Clauset (2014) for continuous and binned data, respectively. For the implementation of this process in R, see Gillespie (2015).

is often negative, then edges frequently terminate at low degree vertices. Such networks are assortative $r \to 1$ and dissortative $r \to -1$, respectively. A network is neither assortative or dissortative when $r = 0$.

2.4.2 Vulnerability analysis

Vertex removal

For the second part of the analysis, I estimate the impact of targeted vertex removals on the cohesion and connectivity of the trafficking network. I implement a simple degree-based removal process, whereby traffickers are removed in descending order of degree (Albert, Albert, and Nakarado, 2004). Three metrics are used to assess the impact of each vertex removal: (1) the number of components, (2) the size of the largest component, and (3) the number of communities. The first two metrics are straightforward, capturing the number of groups that are split-off entirely from the largest connected group of traffickers and the extent to which each split reduces the size of this largest component, respectively.

In order to capture the impact of vertex removals beyond "clean" breaks in the components, the third metric is based on community dissolution. To identify the communities, I use a community detection algorithm based on random walks. Each walk begins at a random vertex and takes a step to a randomly selected adjacent vertex, with the maximum number of steps per walk set to four. After multiple iterations, patterns in the sequences of visited vertices are used to classify each vertex into a community. Put simply, the more often the walks are trapped in certain regions of the network, the more likely those regions are classified as communities (Pons and Latapy, 2006). There may be many candidate classifications, so I select the classification that maximizes a quality function known as modularity (Newman, 2004). Thus, I capture community dissolution in addition to fractures in the network.

Graph sampling

To evaluate how degree mixing influences the impact of targeted vertex removals on the trafficking network, I generate synthetic variants of the network at several

benchmark levels of assortativity, r_b. Following Newman and Park (2003), the synthetic networks are produced by swapping edges while preserving the degree distribution. My algorithm can be described as follows. First, I calculate the assortativity coefficient, r, and clustering coefficient, c, for the trafficking network. Next, two randomly selected edges, (i_1, j_1) and (i_2, j_2), are toggled off. I propose two randomly selected replacement edges, (i_3, j_3) and (i_4, j_4), with the restriction that the degree sequence, or k-sequence, prior to the toggle must be preserved exactly under the proposed edges. Updated assortativity and clustering coefficients, $r\prime$ and $c\prime$, are then calculated under the proposed edges. The proposed edges are accepted if $r\prime \to r_b$ and $c\prime \geq c$. This is designed to preserve the degree distribution and clustering of the observed trafficking network, while reshuffling each synthetic network into a more assortative or dissortative state, as desired. This process is iterated until the desired level of assortativity is reached. I generate 1000 synthetic networks at each r_b.

Of course, the $c\prime \geq c$ condition means that clustering can increase, but not decrease, in each iteration. This is due to limits on the assortativity coefficient, r_{min} and r_{max}, imposed by c.[5] Essentially, the problem is that as $c \to 1$ or $c \to 0$, r_{min} and r_{max} converge. Thus, it is not feasible to reshuffle networks with large clustering coefficients into highly assortative or dissortative states. Since the observed trafficking network has a large clustering coefficient, c tends downwards as I reshuffle towards such states. The $c\prime \geq c$ condition attempts to mitigate this downward trend, unless reducing c is necessary for reaching r_b. More precisely, I accept a proposed reshuffle if $r\prime \to r_b$ but $c\prime < c$ with probability 0.1. This acceptance rate strikes a balance between the computational burden of finding edge reshuffles which satisfy the stringent $c\prime \geq c$, on the one hand, and allowing c to decrease as required in order to reach r_b, on the other. I return to the reshuffling of r and c in section 2.5.2.

[5] See Holme and Zhao (2007) for further details on the relationship between assortativity and clustering.

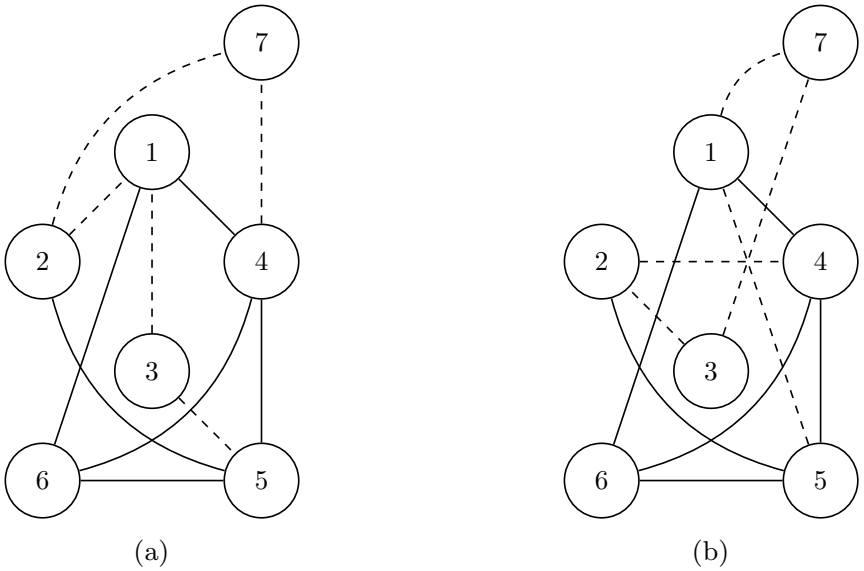


Figure 2.1: Example of graph sampling. (a) Network prior to reshuffling. The dashed edges are ultimately toggled off. (B) The same network after reshuffling. The dashed edges have been toggled on. Note that each vertex has equal degree in both configurations, but the assortativity coefficient has increased.

2.5 Results

2.5.1 Structural properties

I begin by classifying the degree distribution in the trafficking network. Figure 2.2 shows the power law, log-normal, exponential, and Poisson fits to the cumulative distribution function (CDF) of degree, k. As is clear in the CDF, the degree distribution is heavy-tailed, such that most traffickers have fewer than 10 collaborators while the top tenth of the traffickers fall in the range $25 < k < 50$.

Table 2.2 summarizes the parameter estimates, goodness-of-fit, and results of the hypothesis tests for each fit. The p-value exceeds 0.1 in all four models, thus the hypothesis test fails to rule out any of the competing classifications. However, n is small in both the power law and Poisson fits due to the selection of a large k_{min} threshold. The parameter estimates are therefore unreliable, while the hypothesis test has little power to discriminate against either fit.[6] Consequently, while a power law or Poisson model cannot be ruled out, neither represents a parsimonious classification of the degree distribution.

Table 2.3 shows the results of a direct comparison of the log-likelihood of the degree distribution for each model. Where the optimal k_{min} threshold differs between two models, I select the larger of the two, therefore "rewarding" the model which describes a larger section of the distribution. The log-likelihood ratio indicates that the log-normal and exponential models provide considerably stronger classifications of the degree distribution than the power law model. In addition, Figure 2.2 shows that both the log-normal and exponential models are optimized at the lowest possible threshold value, $k_{min} = 1$, thus providing a parsimonious description of the degree distribution in the network as a whole.

Regarding whether the log-normal or exponential model offers the more compelling classification, the direct comparison favors neither. Although the sign of the log-likelihood ratio indicates that the exponential model is a better fit, the magnitude of the ratio does not decisively favor either model. However, one

[6] Clauset, Shalizi, and Newman (2009, p. 669) recommend $n > 50$ for obtaining reliable parameter estimates.

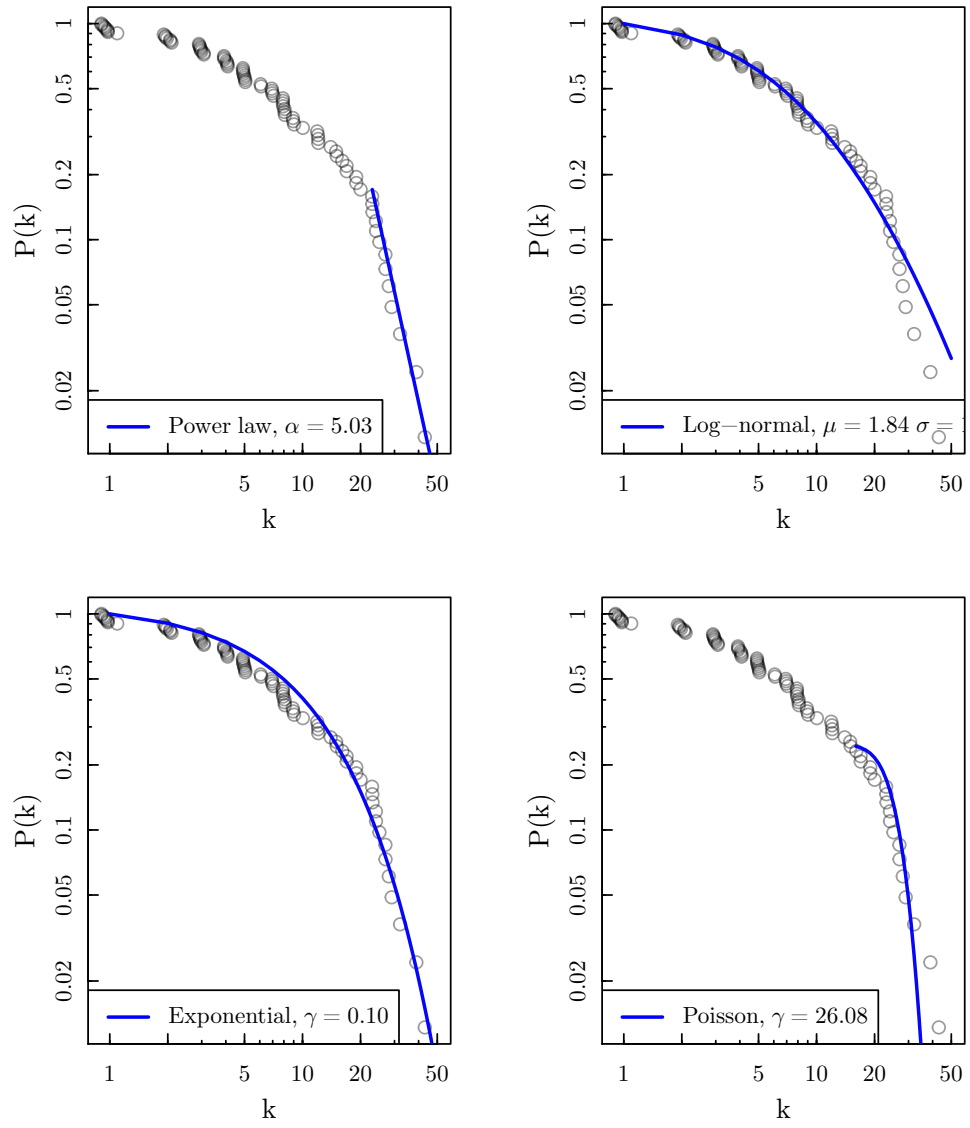


Figure 2.2: Top-left: Power law fit to the cumulative distribution function (CDF) of degree in the trafficking network. Top-right: Log-normal fit. Bottom-left: Exponential fit. Bottom-right: Poission fit.

Table 2.2: Parameters for the competing models of the degree distribution. The k_{min} represents the threshold above which the fit applies to the distribution, where n is the number of traffickers above that threshold. The goodness-of-fit (GOF) is the Kolmogorov-Smirnov statistic. The p-value represents the proportion of fits to synthetic data drawn from a true version of each model that are poorer than the fit to the real data. A fit is ruled out when $p < 0.10$.

Model	n	k_{min}	α	γ	μ	σ	GOF	p
Power law	14	23	5.03				0.099	0.751
Log-Normal	82	1			1.84	1.07	0.053	0.915
Exponential	82	1		0.10			0.517	0.489
Poisson	20	16		26.08			0.177	0.628

small advantage of the exponential model is that it rests on one less parameter. Overall, the exponential decay in degree is consistent with the heavy-tailed degree distribution observed in other collaboration networks.

The exponential classification implies a probability distribution on the formation of collaborative ties in the network. For an actor with degree k, the probability of gaining y collaborators is given by $P(y) = \gamma^{\frac{y}{k}}$. If the network were to continue operating, traffickers with larger degree would therefore be expected to increase their degree at a faster rate than poorly connected traffickers, as in a preferential attachment process.

Turning to clustering, the trafficking network exhibits a large local clustering coefficient $C = 0.801$. This reflects the large proportion of complete triads in the network in comparison to two-paths and triads containing only one edge. The clustering coefficient also implies that around 80% of connected trafficker pairs share at least one mutual collaborator. This finding is consistent with other collaboration networks where it has been observed that actors have a strong tendency to collaborate with the partners of their partners (Newman, 2001a), as noted in section **??**. However, this tendency appears to be particularly strong in the trafficking network, indicating that the network is highly cliquish and suggesting that mutual partners may play an important role in brokering new collaborative relationships.

Despite the characteristic heavy-tailed degree distribution and strong clustering, the mixing pattern in the trafficking network is dissortative rather than assortative. The assortativity coefficient is $r = -0.192$ with a standard error of $\sigma_r = 0.052$, which I estimated via the bootstrapping method detailed in Newman (2002), and therefore the dissortativity is a 3.6σ result and statistically significant. Thus, while ordinarily in collaboration networks large degree actors are connected to other high degree actors and low with low, in the trafficking network those with many partners often collaborate directly with their poorly-connected peers.

Table 2.3: Comparison of alternative models of the degree distribution. A positive log-likelihood ratio indicates that model 1 provides a better fit to the data than model 2. Neither model is preferred when $p > 0.1$.

Model 1	Model 2	k_{min}	LLR	p
Log-Normal	Power Law	1	5.57	< 0.001
Exponential	Power Law	1	5.45	< 0.001
Poisson	Power Law	23	-1.63	0.101
Exponential	Log-Normal	1	0.58	0.560

2.5.2 Vertex removal

Having established the properties of the trafficking network, I now turn to the results of vertex removal on the topology and cohesiveness of the network. Figure 2.3 shows the size, S, of the largest component in the observed trafficking network and in the reshuffled networks as a function of the proportion of traffickers removed, f, where the traffickers are removed in order of their degree. Each band covers the range of values of S given f across 1,000 synthetic networks, which each have an identical degree distribution to the observed network but have been vary in their average levels of degree dissortativity, $r_b = \{0, -0.5, -0.8\}$.

Varying the level of dissortativity changes both the shape and rate at which S falls as a function of f. For example, removing $f = 0.1$ of the highest-degree vertices in the network reduces the size of the largest component to $S = 0.79$ of its original size. This compares to $S = \{0.86, 0.70, 0.46\}$ for $r_k = \{0, -0.5, -0.8\}$, respectively. The results of vertex removal in the observed network overlaps with the $r_k = -0.5$ synthetic networks for small values and large values of f, but S decreases faster in the middle range of f. The reduction in S is slower in the $r_k = 0$ networks across the range of f. Conversely, in the most dissortative case, the size of the largest component steeply declines in response to high-degree removals; removing 13 percent of the traffickers reduces the size of the largest component to approximately half of its original size.

Figure 2.4 presents the number of components, P, as vertices are removed in order or degree. Again, P increases faster in the trafficking network than in the reshuffled networks with benchmark assortativity $r_b = 0$. Indeed, removing the top $f = 0.1$ traffickers from the $r_b = 0$ networks leads to an average split of 2.3 components, while removing the same fraction from the observed network leads to a split into 9 components. Thus, the dissortative mixing in the trafficking network appears to be associated with an increased vulnerability to targeted removals, insofar as the network fractures into a larger number of distinct components across all values of f. The number of components increases faster still as reshuffled to

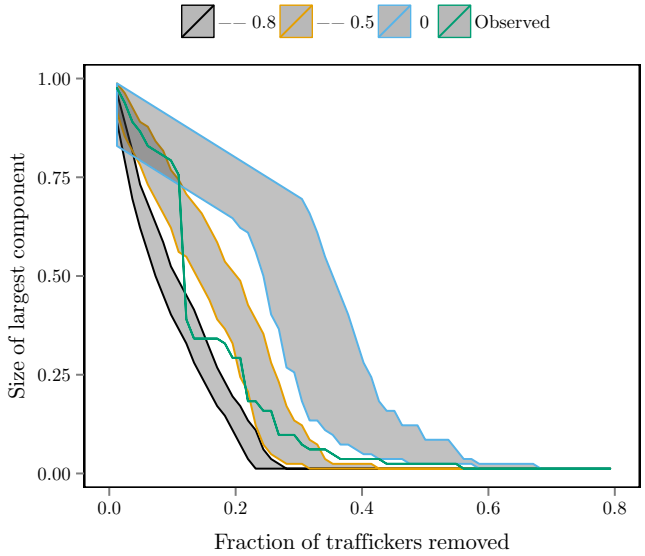


Figure 2.3: Size of the largest component, S, as a function of the fraction of vertices removed, f. Each band represents a reshuffled version of the trafficking network to a specific average level of degree dissortativity, where the degree distribution of the observed network is preserved exactly. The green line represents the observed trafficking network. Since the rewiring process is stochastic, the bands cover the range of values of S given f in 1,000 rewired networks.

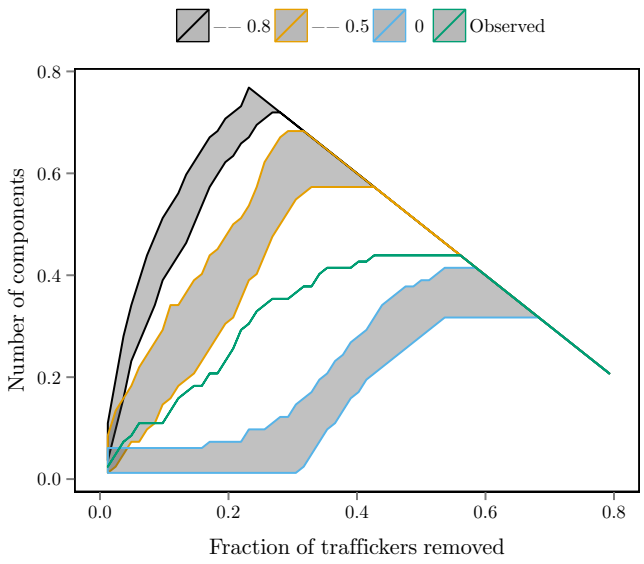


Figure 2.4: The number of components, P, as a proportion of the maximum possible number, in response to the fraction of vertices removed, f. The vertices are removed in order of degree, from highest to lowest.

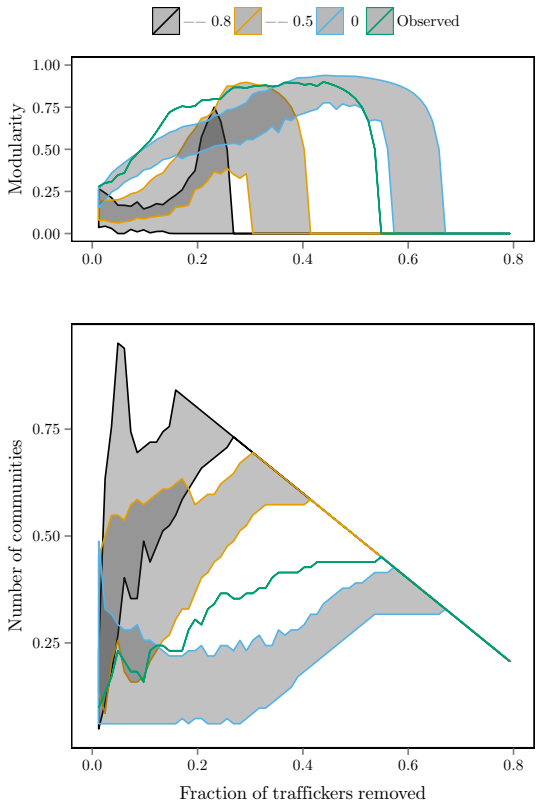


Figure 2.5: Lower: Change in the number of communities, U, as a proportion of the maximum possible number, in response to the fraction of vertices removed, f. Top: The modularity score associated with the classification of communities in the bottom panel. The vertices are removed in order of degree, from highest to lowest.

more dissortative states: removing $f = 0.1$ vertices in the $r_b = 0.8$ networks leads to a split of 37 components, on average.

Finally, Figure 2.5 shows the change in the number of communities, U, as vertices are removed in degree order. As above, varying the dissortativity of the degree mixing has a profound impact on the results of vertex removal. When $r_k = 0$, U remains relatively stable up to $f > 0.3$. However, in the observed network, U begins to increase as $f > 0$, albeit non-monotonically. In the more dissortative networks, the increase in U is faster still, but the modularity scores remain generally lower, indicating that the classification of the communities is not as clear cut. One explanation for this outcome is that, since the cohesiveness of the network is so radically reduced by targeted vertex removals in the most dissortative networks, the number of reasonable classifications of the remaining traffickers into communities remains relatively large, and ultimately approaches the point where all vertices are isolates at considerably smaller values of f. Thus, in addition to the faster reduction in cohesiveness in the more dissortative variants of the observed network, the communities that remain also seem to be less tight-knit.

It is important to note that, while the k-sequence is preserved during reshuffling, assortativity is not the only property that I manipulate. As discussed in the section 2.4.2, the level of clustering in a network constraints the maximum and minimum assortativity that can be obtained through reshuffling. At the extreme, placing an exact constraint on c throughout reshuffling, such that edge proposals are accepted only if they maintain c precisely, places a stringent upper bound on the number of potentially acceptable proposals. Thus, I allowed c to increase or decrease, as required, in order to reach the benchmark assortativity coefficients above.

Figure 2.6 shows the change in c during the graph sampling procedure used to reshuffle the observed network into the most dissortative configuration, $r_b = -0.8$. Note that this represents only one out of 1,000 sampling iterations. To reach $r_b = -0.8$, the clustering is reduced to approximately $c = 0.2$ in this particular network. Although the precise change in c varies across sampling iterations due to the stochastic nature of the edge toggling process, this reduction is fairly typical.

The rate of reduction in c per edge toggle ("iterations" in Figure 2.6) is considerably higher during early iterations of the sampling algorithm. Almost half of the total decrease in r is achieved using only 1000 toggle iterations. However, around 4000 toggles were required for the reduction $r = -0.7$ to $r = -0.8$. This is indicative of the small set of graph configurations that exist for satisfy extreme r values given the constraint on the k-sequence, but it perhaps also suggests that my algorithm could be improved by fine-tuning the acceptance rate to be more stringent for reductions in c when r has not yet reached extreme values.

Due to the decrease in c, the increasing vulnerability of the network to targeted vertex removals in more dissortative states cannot be put down entirely to the degree mixing configuration of the network. Indeed, there is a complex dependence between assortativity, clustering, and vulnerability (Holme and Zhao, 2007), examination of which goes beyond the scope of this paper.

Nevertheless, there are intuitive reasons to suspect that dissortativity is driving much of the vulnerability of the trafficking network. In a dissortative network, there is a greater tendency for high and low degree actors to be connected. Removing a small number of high-degree traffickers can therefore disconnect a large number of low degree traffickers. Indeed, the unusually large clustering coefficient in the trafficking network, which indicates that most traffickers are members of complete triadic structures, may be a critical property for protecting the network from further fragmentation. Moreover, dissortative networks exhibit a different type of vulnerability not examined here. Due to the mixing of high and low degree actors, the detection of any low degree trafficker, of which there are many due to the heavy-tailed degree distribution, may lead investigators directly to highly central, more positionally important traffickers. As such, dissortative mixing places the entire network at greater risk of exposure. Indeed, this may have contributed to the eventual failure of the drug trafficking network.

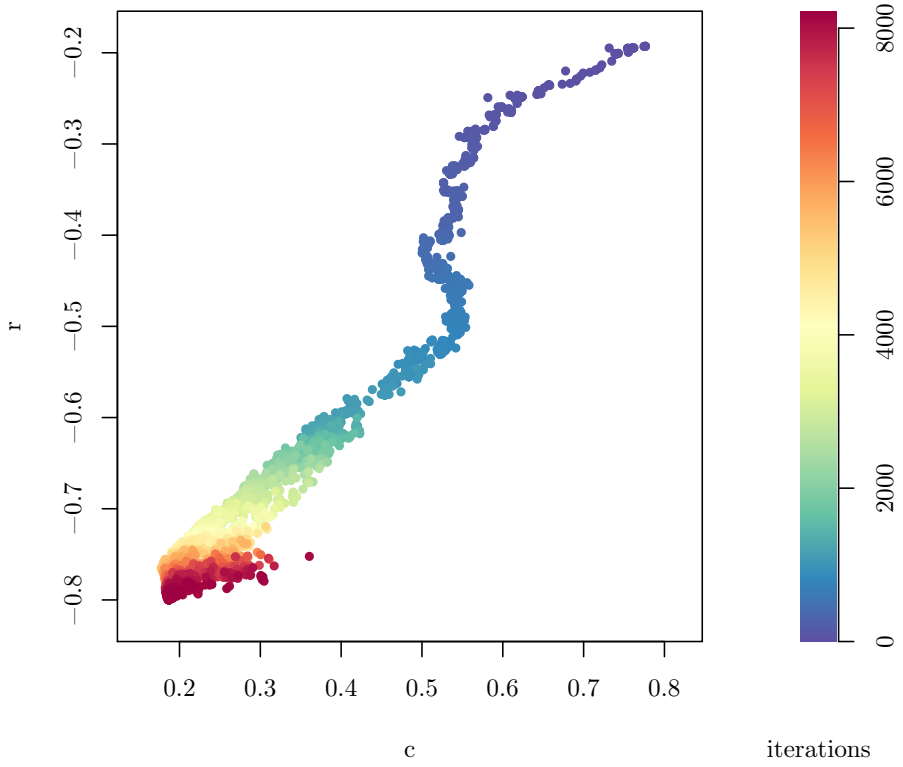


Figure 2.6: The shift in local clustering, c, and assortativity, r, in the graph sampling algorithm used to reshuffle the network into a more dissortative configuration.

2.6 Conclusion

It has long been proposed that criminal networks exhibit abnormal topological properties due to the conditions under which criminals operate. This paper investigates whether a collaboration network among heroin traffickers has a heavy-tailed distribution of degree, strong clustering, and assortative degree mixing – three characteristic properties of collaboration networks outside of the criminal domain. The trafficking network exhibits exponential decay in degree and strong clustering, but dissortative degree mixing, such that high-degree traffickers collaborate preferentially with low-degree traffickers, and vice versa.

The dissortative mixing, in addition to being unusual in a collaboration network and indeed in social networks more generally (Newman and Park, 2003), seems to increase the vulnerability of the network to degree-targeted vertex removals. Indeed, reshuffling the network into more dissortative states increased the impact of such removals, although part of this effect is likely explained by the decrease in clustering that accompanies highly dissortative states. However, degree mixing has received relatively little attention in the empirical literature on criminal networks, despite the fact it can be calculated, in most cases, without additional data. Since degree mixing has demonstrable consequences for the vulnerability of the network to targeted attacks, a growing area of research interest (Morselli, 2010; Malm and Bichler, 2011; Duijn, Kashirin, and Sloot, 2014), it is hoped this study encourages researchers to include degree mixing in future analyses of criminal networks.

To the extent that dissortativity is observed in other criminal networks, the findings raise several questions that were not explored here, particularly regarding the tie formation processes that underpin dissortative mixing, but also regarding the possible upsides of dissortativity for the coordination of criminal activity. One plausible upside lies in the capacity that dissortative networks confer for well-connected criminals to monitor and coordinate peripheral actors directly, for example.

The study is not without limitations. First, although the data source is internally consistent in its reporting of actors and collaboration events, the possibility remains that the data capture an unobserved nonrandom selection process. If the proclivity

to include edges in the source document is influenced by which actors are incident to those edges, the assortativity coefficient could be affected. It would be overestimated if, for example, prosecutors wished to link peripheral members of the network with highly active members. On the other hand, it would be underestimated if peripheral members truly have a tendency to collaborate with high-degree actors, but are also less likely to be included in the sample. Second, while I preserve the degree distribution, rewiring the network to different levels of dissortativity effects other properties in the network such as clustering. Hence the results of vertex removal on the rewired networks do not derive entirely from degree mixing.

However, the results are consistent with previous research that demonstrates that dissortative networks are more vulnerable to targeted attacks than assortative networks. Further research is required to evaluate whether dissortativity is a common property of collaboration among criminals.

3

A Bayesian Exponential Random Graph Model of Collaborative Tie Formation in a Drug Trafficking Network

Contents

3.1 Introduction 47
3.2 Data 50
3.2.1 Actor covariates 50
3.3 Methods 52
3.3.1 Dependence in structured data 52
3.3.2 Estimating an ERGM 54
3.3.3 Bayesian ERGMs 57
3.3.4 Model specification 58
3.4 Results 61
3.4.1 Structural parameters 61
3.4.2 Individual-level parameters 63
3.4.3 Homophily 67
3.4.4 Model comparison 69
3.5 Conclusion 71

Abstract

Building on the previous chapter, this paper investigates the formation of collaborative ties among the heroin traffickers in the Pizza Connection network. Following

Goodreau, Kitts, and Morris (2009), tie formation is treated as a selection problem, where actors choose who to collaborate with from the set of available traffickers in the network. Using data on a real-world heroin trafficking network, I model this selection problem using Bayesian exponential random graph models. The models include dyadic statistics, which capture homophily effects according to traffickers' specialization and membership in subgroups, in addition to structural statistics that account for higher-order patterns of ties. The results show that traffickers are considerably more likely to collaborate when they share mutual partners, while homophily effects vary markedly by specialization and subgroup membership. I conclude that the presence of shared-partners is particularly important in partner selection, which may partly be a consequence of low trust among criminals. This paper also demonstrates the utility of BERGMs for research on criminal networks.

3.1 Introduction

Collaboration is fundamental to organized crime. As in any organization, the success of a criminal group is linked closely to the success of collaboration among its constituent members. However, as in any organization, criminals do not collaborate with one another by mere happenstance, and co-membership in an organized criminal group does not necessarily imply that two people actually work together on the ground. Moreover, collaboration among criminals is notoriously high stakes: defection can be costly (Campana and Varese, 2013), but the payoffs can be considerable (Reuter, MacCoun, and Murphy, 1990), at least for those in large scale drug trafficking, match fixing, or market rigging, for example.

Yet, how criminals come to form the network of collaborative ties that makes up an organized criminal group remains largely unknown. Empirical work has started to map the topology of such networks (Morselli, 2001; Morselli, 2010; Varese, 2013; Campana, 2015; Papachristos and Smith, 2014; Natarajan, 2006; Bright, Hughes, and Chalmers, 2012; Kenney, 2007; Xu and Chen, 2008),[1] but the process of tie formation that underpins this topology is relatively under studied.

This owes, in no small part, to the complexity of modeling tie formation among criminals. In the first place, longitudinal data on collaboration in organized criminal groups are usually unavailable, since criminal networks often have to be "rebuilt" from court records or other source documents (Campana and Varese, 2012) that rarely supply time resolution. Thus, research is typically restricted to static data, which provides a view only of the accumulation of ties up to a particular point in time.

In light of this constraint, one must draw inference on tie formation without having observed tie formation as it occurred. Put differently, one has to estimate the relative importance of factors that may have contributed to the pattern of ties that we observe, in the absence of information regarding the order or circumstances in which those ties were actually formed. Such inference on unobserved processes using observed outcomes is commonplace in social science. But social scientists

[1] See also section 2.

usually rely on a large number of observations in order to estimate trends in the data, whereas research on criminal networks is usually limited to a single network.

Additionally, estimating the factors that underly tie formation is made more difficult still by the complex dependence structure of network data. The assumption that observations are conditionally independent, which is important in many statistical analyses of social science data, such as in generalized linear modeling, does not hold for network data (Holland and Leinhardt, 1981; Anderson, Wasserman, and Crouch, 1999; Robins, Snijders, et al., 2007; Stewart, 2014; Aronow, Samii, and Assenova, 2015; Chandrasekhar and Jackson, 2014). Rather, in a network, a change in the state of an individual or a tie between two individuals may affect the state of several individuals and ties elsewhere in the network (Lazega and Dujin, 1997). For example, suppose that a criminal is imprisoned: each individual who collaborated with that criminal will simultaneously lose a tie. But each lost tie should clearly not be considered an independent event, rather all of the lost ties were due to a single change in the network.

Exponential Random Graph Models (ERGMs) were developed precisely to deal with the above problems (Robins, Snijders, et al., 2007; Goodreau, Kitts, and Morris, 2009), and are beginning to be applied to criminal networks (Papachristos, Hureau, and Braga, 2013; Schaefer, 2012; Grund and Densley, 2015; Malm, Bichler, and Van De Walle, 2010). However, it is well documented that ERGMs may be problematic to estimate for certain network configurations (Snidjers, 2002; Handcock, 2003).[2] Researchers may encounter problems with degeneracy, where the models either do not converge on reasonable parameter estimates given the observed network, or when the maximum likelihood estimate is difficult to find (Krivitsky, 2012). This may lead researchers to ad hoc modeling decisions, such as dropping particular statistics in order to attain an estimable model. However, dropping statistics that are substantively interesting is clearly undesirable, and one still needs to decide which model is preferable in cases where more than one specification is estimable.

[2]ERGMs may also be problematic for large, sparse networks. The recently developed Statistical Exponential Random Graph Models (SERGMs) and Subgraph Generation Models (SUGMs) provide several solutions for analysing such networks (Chandrasekhar and Jackson, 2014).

Much progress has been made to resolve problems of degeneracy, including the development of new statistics (Snijders, Pattison, et al., 2006; Robins, Snijders, et al., 2007; Goodreau, 2007) that are less prone to issues of degeneracy, along with the curved variant of exponential family models (Hunter, 2007). Most recently, a Bayesian framework for ERGMs has been developed that provides further solutions to the problem of degeneracy in ERGM estimation (Caimo and Friel, 2011; Koskinen, Robins, and Pattison, 2009) and an explicit methodology for model selection (Caimo and Friel, 2013).

In this paper, I present a Bayesian Exponential Random Graph Model (BERGM) of collaborative tie formation in a heroin trafficking network. This is, to my knowledge, the first application of the BERGM framework to a criminal network.

The model captures different rates of tie formation according to traffickers' specialization in three key areas of drug trafficking: supply, distribution, and money laundering. In order to account for the structural dependence between traffickers, I also estimate the overall propensity for tie formation, the tendency for some traffickers to form a larger number of ties than others, and the role of mutual partners in brokering new partnerships. Of particular interest in this regard are shared-partner effects, which captures the tendency for pairs of traffickers to collaborate when they share a common prior collaborator versus when they do not. Additionally, I estimate a second model which captures the tendency for within-specialization versus between-specialization ties, as well as within- versus between-subgroup ties, via differential homophily effects. Thus, the model accounts for partnership selection effects by actor specialization and subgroup membership, while simultaneously estimating the effects of structural factors that define the dependence between traffickers, enabling these effects to be disentangled.

The paper is structured as follows. In section 3.2 I describe the data. In section 3.3 I discuss the motivation behind ERGMs, outline the estimation procedure, and note the problems with degeneracy that motivates the Bayesian approach. I then outline the BERGM framework and detail the model specification for the

drug trafficking network. In section 3.4 I present the results. Finally, I make some concluding remarks in section 3.5.

3.2 Data

This paper studies the Pizza Connection Network (PCN). See section 2.3 for details on the source document, coding strategy, and further discussion of the data. In addition to the actor and edge information documented in section 2.3, the data in this paper include actor covariates on the subgroup membership and task specialization of each heroin trafficker, also extracted from the Government's Sentencing Memorandum source document. I expand on these covariates below.

3.2.1 Actor covariates

The source document reports that 63 traffickers belonged to one of three subgroups, organized around Leonardo Greco, Salvatore Catalano, and Gaetano Badalamenti. Contextual information within the source suggests that each group was operational prior to their involvement in the Pizza Connection heroin trafficking operation. However, the source provides little information on the internal structure of each group, beyond identifying the aforementioned actors as leadership figures. The remaining members of the network are not assigned a subgroup classification. Table 3.1 shows the breakdown of actors into the three subgroups. Figure 3.1 (a) shows the subgroup membership in the PCN.

In addition to the subgroup classification, the reports of collaboration events in the source are often accompanied by a short description of the activities associated with that event. I separate these activities into three categories: supply, distribution, and money laundering. In the vast majority of instances the activity is unambiguous. When two or more activities are associated with a collaboration event, I count each separately. Multiple counting is unproblematic since the activities are not associated with the edge, but rather with the actors incident to the edge. Thus, each collaboration events is counted only once, even if it has two or more associated activities. Each actor is subsequently assigned a task

	Subgroup				Total
	Greco	Catalano	Badalamenti	Unclassified	
Supply	5	3	3	4	15
Distribution	1	17	14	5	37
Laundering	6	12	1	4	23
Unclassified	0	1	0	6	7
Total	12	33	18	19	82

Table 3.1: Membership by subgroup and task specialization in the Pizza Connection Network.

specialization based on the activity in which they most frequently participated. Table 3.1 reports the membership of each specialization. Figure 3.1 (b) shows actor specialization in the PCN.

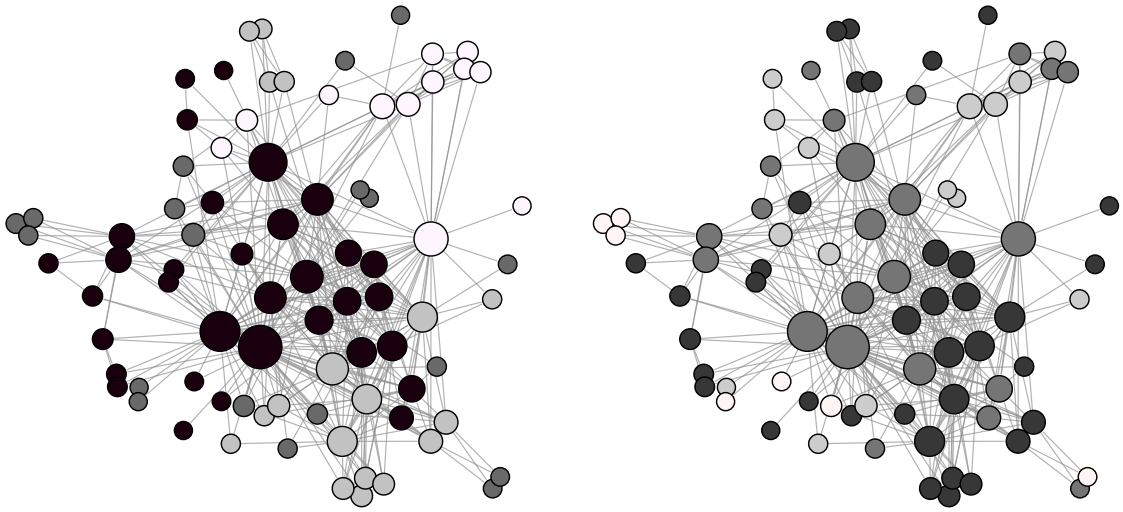
3.3 Methods

The formation of collaborative ties between individuals is a dynamic process and, as such, fully disentangling the predictors of tie formation requires longitudinal data. However, it is often the case that longitudinal data on criminal networks are unavailable, as in this study, and we must therefore infer the structural tendencies underlying tie formation using cross-sectional data. Exponential random graph models (ERGMs) are one means of estimating these tendencies using an observed static network. In this section, I describe the motivation behind ERGMs, give a non-technical outline of the ERGM and Bayesian ERGM frameworks, and specify the precise models used in this paper.

3.3.1 Dependence in structured data

Social scientists work with a variety of types of data. Perhaps the most common data are generated by cross-sectional surveys, where the units of observation are individuals, i, at a fixed point in time, t. Variables of interest are measured on each individual, i.e. X_{1i} and X_{2i}, in the sample S. Taken together, these covariates are vectors $X_1 = (X_{1i}, ..., X_{1S})$ and $X_2 = (X_{2i}, ..., X_{2S})$ of observations for each individual, $i \in S$. When estimating a generalized linear model, for example, the observations are typically assumed to be conditionally independent given the covariates included the model (Stewart, 2014). This assumption underpins many of the standard statistical tests and methods used in social science research (Kenny and Judd, 1986).

However, many types of data have a more complex structure in which the assumption of conditional independence is unreasonable. A network G, where the observations are edges between individuals G_{ij} and therefore dyadic, is one such type. Whereas in cross-sectional data we can often assume that observations



(a) Subgroup. (b) Specialization.

Figure 3.1: The Pizza Connection network. In both graphs, the size of each vertex is proportional to degree. (a) Color indicates subgroup membership: Badalamenti ○, Catalano ●, Greco ◍, unclassified ○. (b) Color indicates task specialization: supply ◍, distribution ●, laundering ◍, unclassified ○.

on each individual i are conditionally independent of observations on j, whether or not i and j share an edge G_{ij} may affect the state of edges elsewhere in the network upon which i or j are incident. Moreover, a change in some covariate of i may simultaneously affect the covariates of several $j \in G$ (Aronow, Samii, and Assenova, 2015). Under the assumption of conditional independence, each of these covariate changes in $j \in G$ could be considered an independent response to the change in i, resulting in model misspecification. However, in the network context, the changes in $j \in G$ may be considered instead as some form of "clustered" response. When modeling a network, it is therefore typical to assume a more complex dependence between observations.

Indeed, for many questions, one may wish to make inferences about the dependence between observations explicitly, as in the present study. Modeling the dependence between observations is the motivation behind the ERGM approach. While ERGMs are not the only approach – modeling dependence between observations has attracted considerable methodological interest and innovation in recent years (King and Roberts, 2014; Aronow, Samii, and Assenova, 2015) – they are perhaps the most widely applied for network data. Under the ERGM framework, we model the state of an observed network as a function of both the covariates of the individuals within the network and the structural features that define their relationships with each other. Crucially, we do so without assuming the independence of observations (Cranmer and Desmarais, 2011), but rather by making precise assumptions regarding the dependence between observations.

3.3.2 Estimating an ERGM

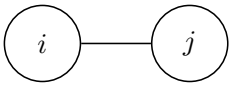
In the ERGM framework, we treat the observed network as one realization, out of many possible realizations, of an unknown stochastic process (Robins, Snijders, et al., 2007), where the set of possible realizations includes only networks that have the exact number of actors as in the observed network. Each one of these realizations, including the observed one, occurs with some probability, and hence there is a probability distribution of realizations. The aim is to recover parameters

of a stochastic process under which the observed network is the most probable realization. These parameters reflect the structural- and individual-level tendencies that underlie tie formation in the network.

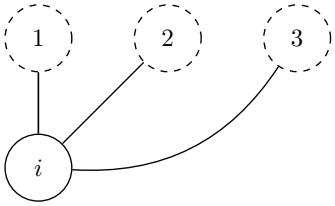
The estimation of an ERGM can be separated into three stages (Robins, Snijders, et al., 2007). First, we specify a model (the ERGM) of the stochastic process by specifying network statistics, y. These statistics are subgraph motifs, or configurations, as shown in Figure 3.2, which occur with a countable frequency in the observed network and which we assume are important factors underlying tie formation. The set of statistics that we select reflects our assumptions about the dependence between observations in the network. Thus, if we think there is a transitivity effect in the network whereby "a friend of a friend is my friend," we may include the geometrically-weighted edgewise-shared partner shown in Figure 3.2.

Second, we estimate parameters θ for these statistics so that, of all the possible realizable networks, the observed network has the greatest probability of occurring. The statistics selected in the first step define a parameter space from which we sample parameter values using Markov chain Monte Carlo (MCMC) simulation methods (Goodreau, Kitts, and Morris, 2009). These parameter values reflect, in part, the frequency with which the statistics in our model occur in the network. We aim to maximize the likelihood of the ERGM, which is defined by the probability distribution of the observed network given the statistics and associated parameters, $f(y|\theta)$. Thus, once we have specified a set of statistics that constitute our ERGM, we then sample parameter values for the set of statistics and select the values under which the likelihood of the observed data, given the model, is maximized.

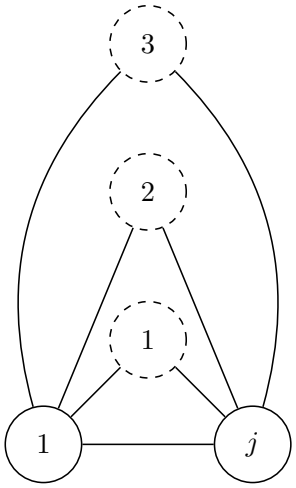
Third, we evaluate the goodness-of-fit of the model by simulating networks based on the ERGM that we estimated in step two, and then comparing the simulated networks to the observed network. The goodness-of-fit improves as more of the simulated networks resemble our observed network. For a more detailed description of the estimation process, see Snidjers (2002), Robins, Snijders, et al. (2007), Snijders, Pattison, et al. (2006), Goodreau (2007), Lubbers and Snidjers (2007), and Robins, Pattison, and Wang (2009).



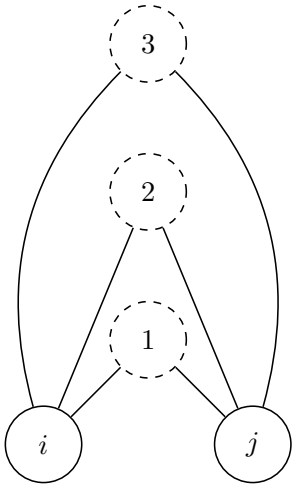
(a) Edges.



(b) Geometrically-weighted degree.



(c) Geometrically-weighted edgewise-shared partner.



(d) Geometrically-weighted dyadwise-shared partner.

Figure 3.2: Structural statistics, $\boldsymbol{s}_m$, included in the BERGM. Note that, for each statistic, we are interested in pairs of actors, i and j, with the exception of (b) where we focus on the degree of a single actor i.

Degeneracy

A substantial issue with ERGM estimation is degeneracy. A model is degenerate when either the probability of the observed network, given the model, is very low, the maximum likelihood estimate cannot be found, or when only a small number of network configurations have a substantial probability of being realized and these configurations have very different structural properties (Snidjers, 2002; Krivitsky, 2012).

Several strategies exist to resolve problems of degeneracy. The most straightforward strategy is to respecify the model ad hoc to include a different group of statistics, for which there may be parameter values where the observed network is realized with greater probability. However, it remains possible that any respecified model will also be degenerate, and this strategy may involve removing statistics of interests from the model. Snijders, Pattison, et al. (2006) and Hunter (2007) proposed variants of commonly used statistics in ERGMs, such as the alternating k-triangle variant of the triangle statistic, as shown in Figure 3.2, which reduce problems with degeneracy. Additionally, Handcock (2003) proposed a Bayesian framework for ERGM estimation in which one can use priors to reduce the assignment of probability mass to certain regions of the parameter space that produce degenerate models. In this paper, I implement the variant statistics within the BERGM framework.

3.3.3 Bayesian ERGMs

Until recently, the Bayesian framework for ERGMs was not practicable due to the "double-intractability" (Caimo and Friel, 2011, pp.9) of sampling parameter values from the posterior distribution and calculating the maximum likelihood within the posterior. However, Koskinen (2004) and later Caimo and Friel (2011) introduced methods for estimating BERGMs that overcome the intractability problem.[3]

The major advantage of the Bayesian framework for ERGMs is that we can specify priors on the parameters in the model. These priors modify the assignment of probability mass to certain regions of the posterior distribution, from which

[3]These methods are now implemented in the `BERGM` package in `R` (Caimo and Friel, 2014).

we sample in order to simulate realizations of networks. The priors reflect our information regarding the values that the parameters can reasonably take on. For instance, suppose we wish to estimate a simple model including only an edge statistic, v. Following Caimo and Friel (2011), the model can be written:

$$\pi(\theta|y) \propto f(y|\theta, v)\pi(\theta|v)\pi(v) \tag{3.1}$$

where π are priors, edge is the only statistic in the model, θ is the parameter estimate associated with the edge statistic, and y is the observed network. If the network is sparse, i.e. there are relatively few edges for the number of actors in the network, then we would expect the edge parameter to be negative. Thus, when specifying the BERGM, we would place a negative prior on the edge parameter, thereby changing the assignment of probability mass in the parameter space so that negative draws from that space are more probable than positive draws. In taking this step, parameter values that would give rise to dense networks, which do not resemble our sparse observed network, are less probable, therefore mitigating the problem of degeneracy.

A further strength of the BERGM framework is that one can select between competing models through direct comparison of the posterior probabilities of each candidate model (Caimo and Friel, 2013). Thus, if there are several combinations of structural- and individual-tendencies that one believes to be theoretically plausible models of tie formation, we can evaluate which combination has the greatest probability of producing the observed network. Direct model comparison may be used to discriminate against, or rule out, certain combinations of network statistics. I take advantage of this capacity in section 3.4.4.[4]

3.3.4 Model specification

Turning to the heroin trafficking network, we are interested in the association between both structural and individual-level statistics and the selection of collaborative partners. Starting with an empty network of heroin traffickers, we assume that

[4]There is also a Bayesian approach for augmenting missing data in ERGMs (Koskinen, Robins, and Pattison, 2009).

each pair of traffickers can, in principle, select to collaborate together. Selection is influenced by four structural parameters: (1) the overall propensity to form collaborative ties in the network, (2) the tendency for some traffickers to belong to more ties than other traffickers, (3) the formation of ties between pairs of traffickers who share a mutual collaborator, and lastly (4) the formation of ties between pairs of traffickers who do not. These structural parameters are represented by an edge, geometrically-weighted degree (GWDegree), geometrically-weighted edgewise-shared parter (GWESP), and geometrically-weighted dyadwise-shared partner statistic (GWDSP), respectively. Figure 3.2 shows each of these statistics.

In addition to the structural parameters, I hypothesize that selection is influenced by three individual-level parameters. Firstly, the tendency for traffickers to collaborate preferentially with members of their own subgroup within the trafficking operation. Secondly, the tendency for traffickers to collaborate preferentially with those who do not have the same specialization as themselves. Put differently, we expect some heterogeneous matching by specialization, reflecting the vertical integration of the trafficking operation. For operational success, it is clear that suppliers of narcotics must collaborate with distributors and the distributors with money launderers, whereas collaboration between distributors is not strictly necessary. Thirdly, I include a group of parameters to allow for differential rates of activity based on subgroup membership and specialization. I use this group of statistics to control for varying levels of involvement according to subgroup membership, and to estimate the demand for collaborators based on trafficker specialization, thus allowing for the possibility that distributors require more partners than suppliers, for example.

The first two individual-level parameters are represented by attribute-homophily statistics, where the attributes, q, are subgroup membership and specialization, respectively. I include the third group of parameters as binary vertex-factors, again reflecting each traffickers subgroup membership and specialization.

In summary, following Caimo and Friel (2011) the model can be formally stated

$$\pi(y|\theta) = \frac{1}{z(\theta)} exp[\boldsymbol{\theta}_m \boldsymbol{s}_m(y) + \boldsymbol{\theta}_n \boldsymbol{s}_n(y)] \tag{3.2}$$

where z is a normalizing constant, $\boldsymbol{\theta}_m$ is a vector of parameters associated with the structural statistics $\boldsymbol{s}_m$, and $\boldsymbol{\theta}_n$ is a vector of parameters associated with the individual-level statistics $\boldsymbol{s}_n$. Using the definitions in Hunter (2007), the structural statistics are given by

$$\text{edge} = \sum_{i \neq j} y_{ij} \tag{3.3}$$

$$\text{gwdegree} = e^{\tau_{m1}} \sum_{i=1}^{n-1} \left\{ 1 - (1 - e^{-\theta_{m1}})^i \right\} D_i(y) \tag{3.4}$$

$$\text{gwesp} = e^{\tau_{m2}} \sum_{i=1}^{n-2} \left\{ 1 - (1 - e^{-\theta_{m2}})^i \right\} EP_i(y) \tag{3.5}$$

$$\text{gwdsp} = e^{\tau_{m3}} \sum_{i=1}^{n-2} \left\{ 1 - (1 - e^{-\theta_{m3}})^i \right\} DP_i(y) \tag{3.6}$$

where D is degree, EP is the number of connected pairs who share exactly k mutual partners, and DP is the number of unconnected pairs who share exactly k mutual partners. Likewise, the individual-level statistics $\boldsymbol{s}_n$ are

$$\text{attribute-homophily} = \sum_{i \neq j} y_{ij}, \mathrm{q}_i = \mathrm{q}_j \tag{3.7}$$

$$\text{vertex-activity} = \sum_{i \neq j} y_{ij} \mathrm{q}_i \tag{3.8}$$

where q represents the subgroup membership or specialty of the trafficker.

A further, important part of the model specification lies in setting the τ parameters in the geometrically-weighted statistics (3.4–3.6). Corresponding to each of these statistics, the τ parameter dictates the decay in influence of each additional unit of D_i, EP_i, or DP_i on the probability of the edge (Goodreau, Kitts, and Morris, 2009). Hence, if we set $\tau = 0$, the probability of an edge that adds one shared partner to a dyad will only be affected if, without that edge, the dyad would have zero shared partners (Hunter, 2007). Conversely, if we set a large τ, the probability of an edge that adds one shared partner to a dyad would be affected even if that dyad already had several. Following Snijders, Pattison, et al. (2006) and Hunter (2007), I set $\tau_{m1} = 0.3$, $\tau_{m2} = log(2)$, and $\tau_{m3} = log(2)$ for the GWDegree, GWESP, and GWDSP statistics, respectively. The results are robust to the specification of τ, as demonstrated in appendix A.

Lastly, I assign normal priors $\pi_{edge} = \mathcal{N}(-1, 5)$ and $\pi_{gwesp} = \mathcal{N}(1, 5)$, reflecting the sparsity and strong local clustering in the heroin trafficking network, as demonstrated in section 3.5.

3.4 Results

3.4.1 Structural parameters

I begin by presenting parameter estimates for the structural statistics in the BERGM. Figure 3.3 shows the estimated marginal posterior densities for the edge, GWDegree, GWESP, and GWDSP statistics. The negative edge parameter indicates the network has low density, whereby the majority of trafficker-dyads do not collaborate. The GWDegree parameter is positive, though the interval straddles zero. Thus, if one were to randomly add an edge to the network, the degree of the actors incident to that edge does not definitively increase or decrease the probability of the observed network.

The structural effects of most interest are the GWESP and GWDSP statistics, which represent third-party motifs, in addition to fourth-party and even higher order motifs, within the network. The posterior mean estimate for the $\theta_{gwesp} = 2.46$ term indicates that, if a collaborating pair of traffickers have a shared mutual partner, they are more likely than would be expected by chance to have at least one additional shared partner. In terms of the magnitude of this increase, suppose we were to count the number of triangles in the network before and after a new collaborative relationship is formed. Given our model of the tie formation process, the GWESP parameter value of 2.46 implies the log-odds that the new edge completes a triangle are 7.38 times higher than the log-odds that it adds no new triangles to the network (Goodreau, Kitts, and Morris, 2009). As per the τ parameter, for each triangle that the new edge would add, the marginal increase in log-odds grows smaller.

In contrast, the GWDSP term is near zero, $\theta_{gwdsp} = -0.03$. Therefore, a new edge is no more or less likely to complete an additional two-path, i.e. a triadic structure with exactly two edges, within the network than would be expected by chance. Taking the GWESP and GWDSP terms together, the model suggests a

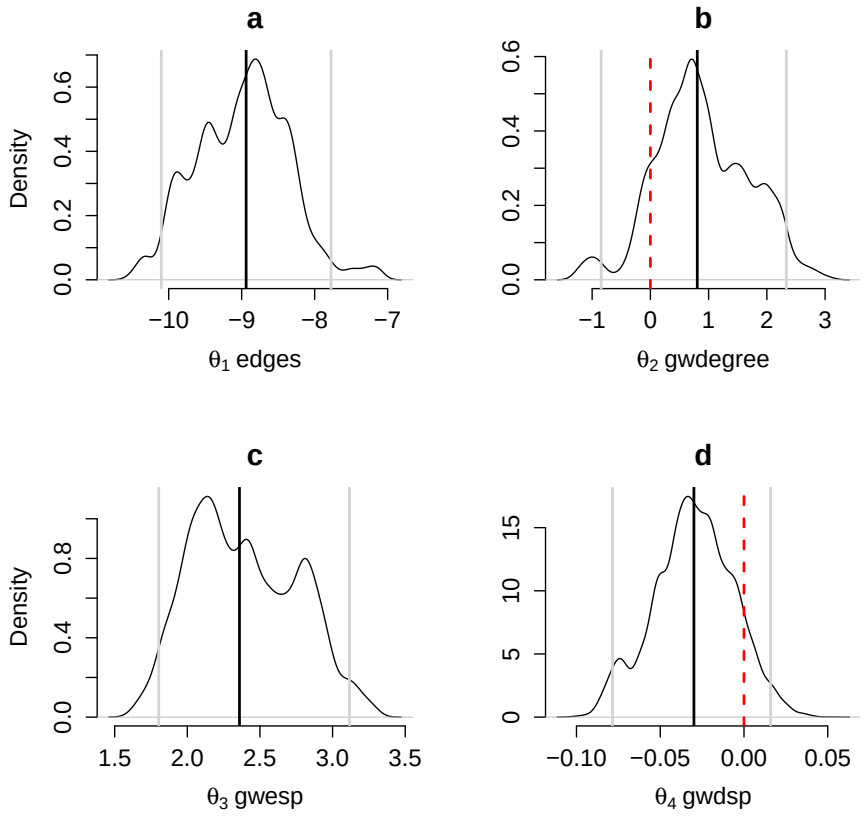


Figure 3.3: Estimated marginal posterior densities for the structural statistics: (a) edge, (b) geometrically-weighted degree, (c) geometrically-weighted edgewise-shared partner, and (d) geometrically-weighted dyadwise-shared partner. The black line represents the posterior mean. The grey lines represent a 95% credible interval. The dotted red line indicates zero.

very strong tendency for collaborative relationships to be formed between pairs of traffickers whom already share a mutual partner. This tendency persists for pairs of traffickers who share multiple partners, but at a diminishing rate.

To guide intuition behind the estimation procedure, Figure 3.4 shows the posterior density of the GWESP and GWDSP statistics from 1,000 iterations of the MCMC algorithm. Figure 3.4 also shows the trace plot for 4 chains of the MCMC sampler at 100-iteration intervals.[5] Each of these traces begins with an initial set of parameter values. For each chain at each iteration, a move to a new set of parameter values is proposed, which is accepted if the new parameter values return a greater probability of the observed network (Caimo and Friel, 2011). One can understand from the traces how uncertainty is estimated in the model. Visits to what ultimately end up as sparse regions of the posterior provided, at that iteration of the model, the greatest probability of the observed network. However, in each chain, later iterations moved back towards more "central" regions, which therefore have greater density. The regularity of visits to different regions of the posterior thus reflects the uncertainty around our parameter estimates. If the traces visit similar regions of the posterior only rarely, and the chains follow divergent paths, then the uncertainty around our parameters will be larger. Conversely, more regular visits to similar regions of the posterior reflects lower uncertainty.

3.4.2 Individual-level parameters

Turning to the individual-level parameters, Figure 3.5 shows the estimated marginal posterior densities for subgroup membership and specialization, alongside homophily effects for pairs of traffickers belonging to the same group and, separately, with the same specialization. The posteriors (e–f) indicate that members of the Catalano (0.83) and Badalamenti subgroups (1.11) tend to form more collaborative ties, compared with their unaffiliated counterparts. The mean of the posterior is positive for traffickers in the Greco subgroup (0.63), however the credible interval includes

[5]It is worth noting that, while only 2 dimensions of the posterior density are shown in Figure 3.4, the posterior density for the entire model has 12 dimensions, one for each statistic. Additionally, 8 chains were used for MCMC sampling. Thus, values for all twelve parameters are proposed simultaneously at each iteration.

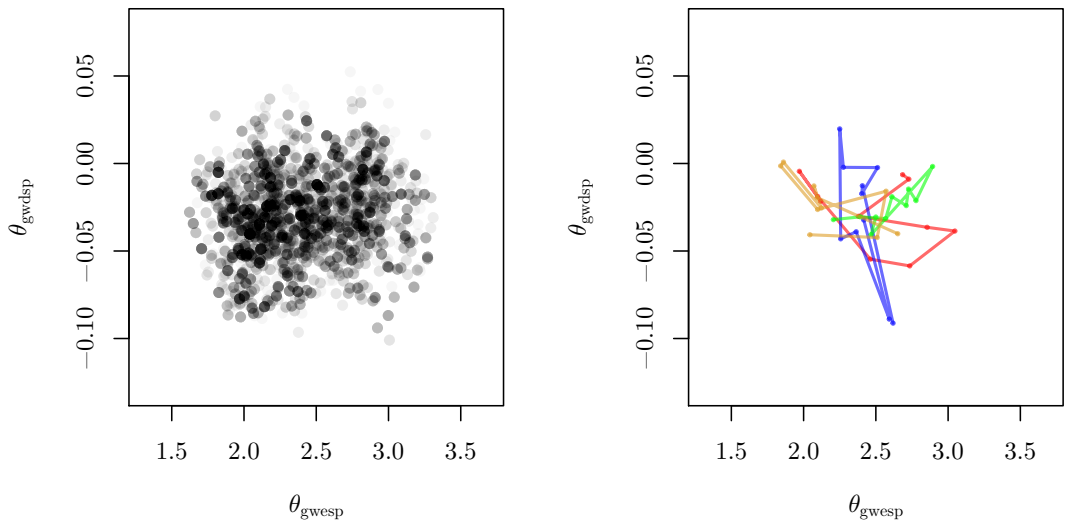


Figure 3.4: Left: Posterior density for 1,000 iterations of the MCMC algorithm. Right: Traceplot of 4 chains at 100-iteration intervals from the MCMC algorithm.

negative parameter values. Thus, the model indicates that Greco members are no more or less likely to form ties than unaffiliated traffickers.

In terms of the size of the subgroup-activity effects, as with the structural parameters, each estimate can be interpreted by supposing that a new collaborative tie is formed in the trafficking network: each subgroup parameter give the log-odds that a trafficker who belongs to that respective subgroup is incident to the new tie (Snijders, Pattison, et al., 2006). The odds of a Catalano member being incident to the tie is therefore 2.29 higher than an unaffiliated trafficker, compared with 3.03 higher for a Badalamenti member. Overall, unaffiliated traffickers seem to be involved in far fewer collaborative relationships. This likely reflects their role in the trafficking operation; one only needs to work with outsiders for specific tasks that cannot be carried out by the personnel within the group. Beyond the collaborative ties formed for these specific tasks, the unaffiliated members seem to be relatively inactive.

Figure 3.5 (h–j) shows the activity parameters by specialization. Traffickers specializing in the supply of narcotics (-0.05) and in money laundering (0.08) have a similar level of activity to traffickers with no specialization, with both posterior densities straddling zero. However, traffickers specializing in distribution have a considerably stronger tendency (1.20) than others to form collaborative ties. The odds of a distributor being incident to a newly formed tie are 3.49, 3.06, and 3.32 higher than a supplier, launderer, and trafficker with no specialization, respectively. These results suggest that specialization is an important factor in collaborative tie formation, in that distributors have a greater demand for partners than other traffickers. It is perhaps to be expected that distributors have a greater demand given that, in principle, they must work with the suppliers to source the narcotics, on the one side, and the money launderers to deal with the proceeds, on the other. Nonetheless, it is important to note that distributors are the largest of the specialization groups, thus distribution-related activities appear to drive a large portion of the connectivity of the network.

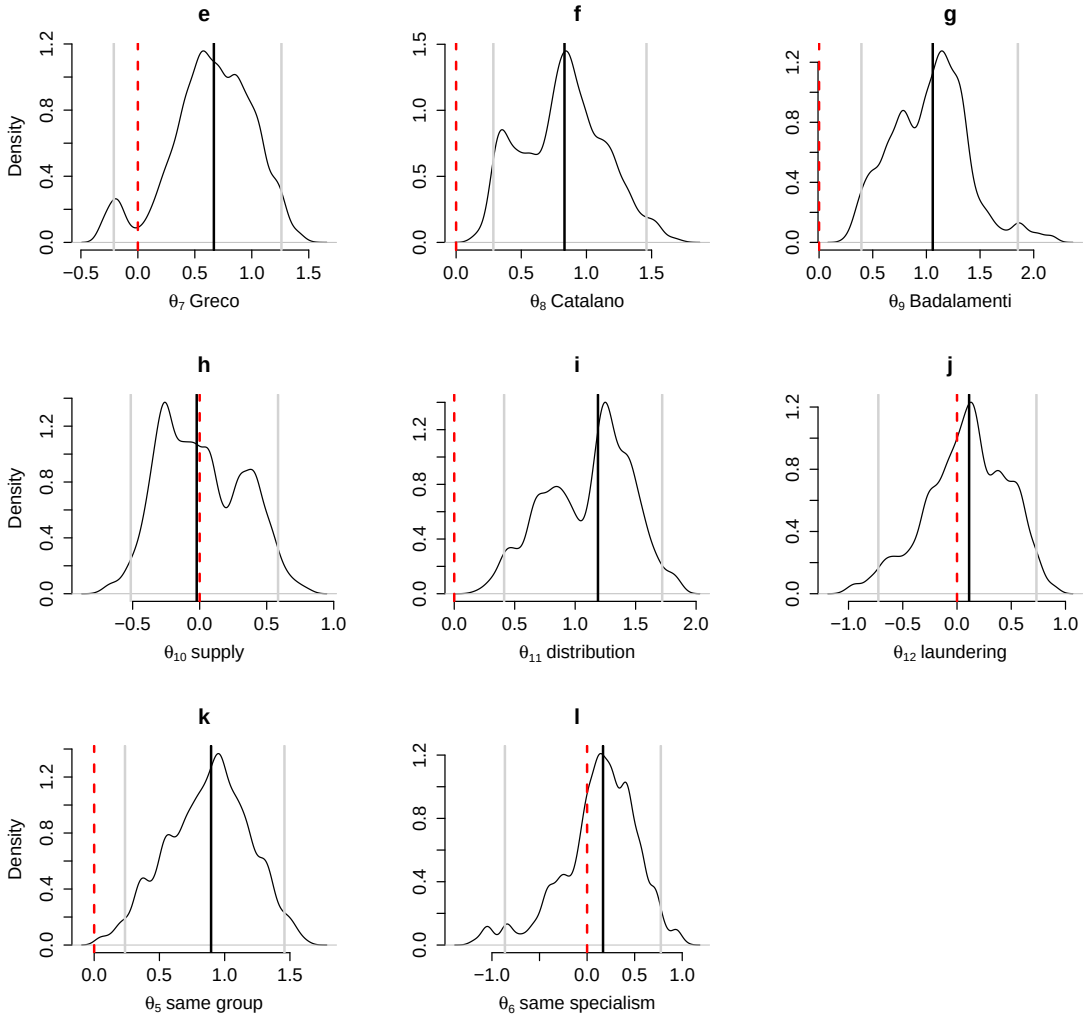


Figure 3.5: Estimated marginal posterior densities for the actor-level statistics: (e, f, g) group activity, (h, i, f) speciality activity, (k) group homophily, and (l) speciality homophily. The black line represents the posterior mean. The grey lines represent a 95% credible interval. The dotted red line indicates zero.

In addition to the overall activity parameters, we are also interested in the tendency for for collaboration in trafficker-dyads that belong to the same subgroup or have the same specialism. Figure 3.5 (k–l) shows posterior densities for the homophily statistics. The first, and perhaps most surprising, finding to note is that traffickers with the same specialism (0.17) are neither more nor less likely to collaborate with each other than traffickers with different specialisms. Thus, supplier-supplier ties are no more likely than supplier-distributor ties, for example. The absence of homophily by specialty suggests the network strikes a strong balance between vertical- and horizontal-integration, in that between-specialism collaboration is as common as within-specialism collaboration. On the other hand, subgroup-homophily (0.90) is relatively strong. The odds of collaboration between traffickers in the same subgroup are 2.46 higher than for traffickers in different subgroups.

3.4.3 Homophily

The main model, reported above, is limited in that it does not estimate separate homophily parameters by subgroup and specialism, and instead estimates a single aggregate parameter for both attributes. In other words, the main model is interested only in the prevalence of ties between members of the same subgroup and specialism, regardless of which subgroup or specialism that may be.

An alternative model includes a set of homophily statistics, one for each distinct subgroup and specialism, in addition to the structural and activity statistics from the main model. Here, I estimate this alternative model using identical priors and τ parameters as in the main model. A consequence of the additional homophily statistics is that the parameter estimates for the structural and individual-level statistics will likely differ from those estimated in the main model. However, the homophily parameters in the alternative model can be compared directly with those in the main model, so long as they are interpreted net of the other statistics.

Figure 3.6 shows the posterior densities for the separated homophily statistics in the alternative model. While in the main model, being a member of the same

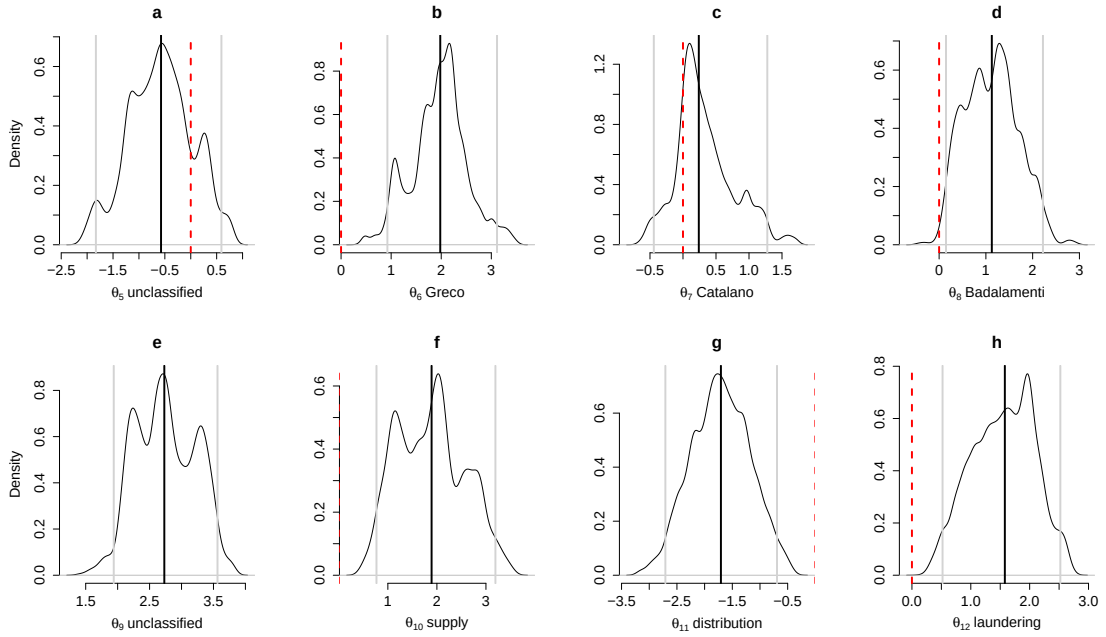


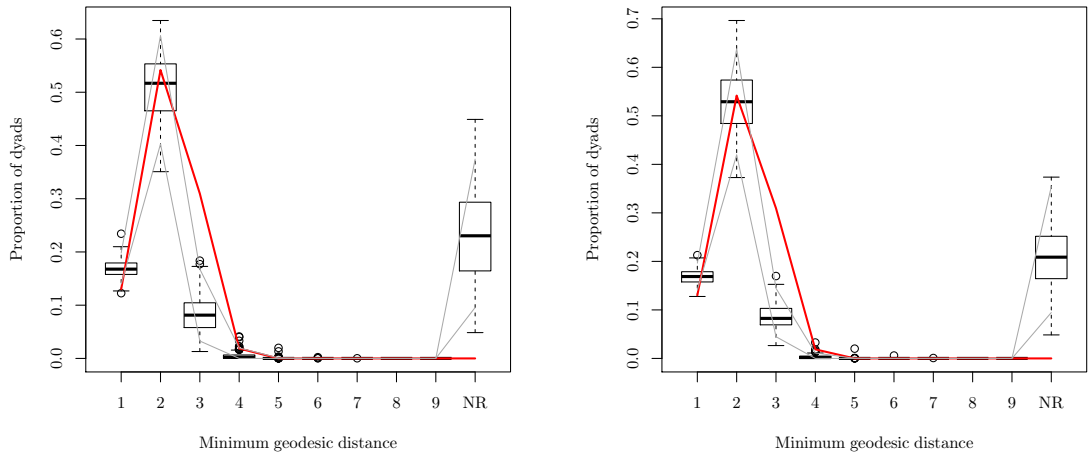
Figure 3.6: Estimated posterior densities for the separated homophily statistics in the alternative model: (a–d) homophily by subgroup membership, (e–h) homophily by specialism. The black line represents the posterior mean. The grey lines represent a 95% credible interval. The dotted red line indicates zero.

subgroup increased the odds of collaboration, the alternative model provides a more nuanced view. Specifically, unclassified traffickers and members of the Catalano subgroup, the largest of three, appear no more or less likely to form within-subgroup than between-subgroup ties. On the other hand, dyads in the Greco (1.95) and Badalamenti (1.12) subgroups are considerably more likely to form within-group collaborative ties, the odds of such ties being 7.03 and 3.06 higher than ties to traffickers outside of each subgroup, respectively. It is worth noting that the Greco and Badalamenti subgroups are much smaller than the Catalano subgroup. It may be the case that influential traffickers within these smaller subgroups were more able to monopolize partnerships with members of the network outside of their own subgroup.

For the separated homophily effects by specialization, Figure 3.6 (e–h) indicate the presence of within-specialization homophily among suppliers, money launderers, and unclassified traffickers. The odds of two suppliers being incident to a newly formed tie are 6.55 higher than a supplier to non-supplier partnership. The corresponding odds for launders and unclassified traffickers are 4.66 and 15.6, respectively. Such homophily effects are striking given the absence of an aggregate specialization-homophily effect in the main model. However, the lack of an aggregate effect is likely explained by the large negative parameter for distributor homophily (-1.7), indicating the odds of two distributors being incident to a newly formed tie are 0.18 that of a distributor to non-distributor tie. Since distributors are the largest specialty group, and also appear to have the greatest demand for partners, these longer odds translate into a larger overall tendency against specialism-homophily within the network. A final point to note is that the tendency for distributors to collaborate with traffickers who have a different specialization to themselves supports our expectation that it is uniquely incumbent upon distributors to work with both sides of the operational chain: the suppliers and launderers.

3.4.4 Model comparison

Turning to the goodness-of-fit of the main and alternative models, Figure 3.7 shows the distribution of geodesic distances, u, between all dyads in the observed trafficking



(a) Main Model (b) Alternative Model

Figure 3.7: Plots of the goodness-of-fit for the main model and alternative model. The red line represents the proportion of dyads in the observed network that are separated by the geodesic distance indicated on the x-axis. The box-plots reflect the equivalent proportions for 100 networks simulated from the respective model.

network, alongside the corresponding distribution for 100 simulated networks drawn under each model. Insofar as the geodesic distance is concerned, the models have a similar goodness-of-fit. Both models provide good estimates for the proportion of dyads at $u < 3$ and $u > 3$. However, both estimate the proportion of dyads at $u = 3$ to be short of 0.1, whereas in the observed network it exceeds 0.3. The models allocate this "missing" proportion instead to dyads that are entirely disconnected. By implication, a portion of the simulated networks are made up of more than one component, which does not reflect the single component in the observed trafficking network. A possibly remedy to this problem would be to assign smaller probability mass to regions of the parameter space that are known to produce realizations with more than one component. Alternatively, it may be desirable to place a fixed constraint on the realizations so that parameter combinations that produce simulated networks with more than one component are simply discarded.

Figure 3.8 shows the posterior model probability of the main and alternative models. The alternative specification, with separated homophily effects, appears to produce network realizations that are similar to the observed network with much greater probability. However, the Bayes factor, which gives the ratio of the posterior odds of the network given the alternative model to the corresponding probability given the main model, $^{P(y|alt)}/_{P(y|main)}$, is just 2.94. A Bayes factors of this magnitude is not typically considered to constitute strong evidence against the main model (Kass and Raftery, 1995; Caimo and Friel, 2013), indicating that the separated homophily effects do not substantially improve the model.

3.5 Conclusion

I presented two Bayesian Exponential Random Graph models of tie formation in a heroin trafficking network. These models specify precise structural- and individual-level statistics, which are treated as the building blocks of the drug trafficking network. The aim of the models was to estimate the relative influence of these statistics in the unobserved tie formation process, based on their prevalence in the observed network.

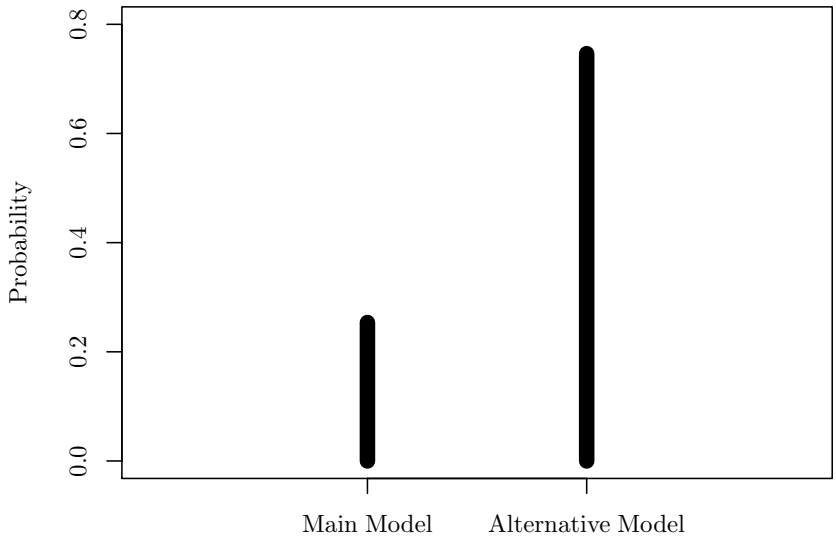


Figure 3.8: Posterior model probabilty of the main and alternative models.

The results of the models highlight several features of tie formation among drug traffickers. From a structural point of view, tie formation is underpinned by strong shared-partner effects. New collaborative partnerships are considerably more likely to develop between pairs of traffickers whom already share a mutual partner. However, tie formation between traffickers at more than two degrees of separation is relatively uncommon. In one sense, this shared-partner effect is to be expected in a social network, where transitivity effects are normal. In this respect, the findings here affirm that criminal networks share an important structural feature with other types of social network. However, the magnitude of the shared-partner effect may point to a particularly important role for third-parties in drug trafficking operations in their capacity for trust-building. Such low trust, high-risk environments may multiply transitivity effects, as forming new ties carries greater risk than in ordinary social situations. Thus, it is plausible that traffickers place a particular emphasis on the information or vouch that only a third-party can provide. The results suggest that, without this information, new partnerships rarely develop.

Another notable tendency is for traffickers to vary in how many ties they form according to their specialization. Distributors have a considerably larger tendency to create new partnerships than suppliers and money launderers. Additionally, distributors form fewer ties with other distributors than with suppliers, launderers, and unclassified traffickers, reflecting the fact they interact with both sides of the trafficking operation. Insofar as distributors are systematically more connected, and connected both up and down the operational chain, they may therefore be the most likely to expose large portions of the network if placed under surveillance. On the other hand, suppliers and money launders are more inclined to form within-specialty ties, thus creating more insular subgraphs.

The BERGM framework can find many applications in research on criminal networks, where tie formation is often of direct substantive interest, but longitudinal data are typically difficult, and sometimes all but impossible, to obtain. Already, ERGMs are being applied in this vein (Papachristos, Hureau, and Braga, 2013; Schaefer, 2012; Grund and Densley, 2015; Malm, Bichler, and Van De Walle,

2010). The BERGM approach represents a promising avenue for mitigating, or even resolving entirely, some of the most common obstacles that researchers encounter when attempting to estimate ERGMs. In particular, the problem of degeneracy may be overcome by assigning priors to the parameter space, thereby shifting probability mass away from degenerate regions. In light of the growing research trend in criminology towards statistical analyses of criminal networks, such priors should be forthcoming. Moreover, the BERGM framework provides a comprehensive solution to the problem of model selection. Many questions of substantive interest to criminologists depend on weighing competing hypotheses against one another, such as which structural- and individual-level statistics play a greater role in patterns of co-offending. Using the Bayesian toolbox, the probabilities that alternatively specified models produce network realizations similar to the observed network under scrutiny can be compared directly (Caimo and Friel, 2013), providing an intuitive means of evaluating model evidence. This paper demonstrated both the prior specification and model selection procedures.

This study is not without limitations. Most important is the non-random sampling of data, as discussed in section 2.3.4. Secondly, the models may be improved by information on the geographical region in which the traffickers were active. Although geographical separation is not a strict constraint on tie formation, since collaboration can and does occur via telephone conversations, and some geographical variation is likely captured by subgroup membership, it would be preferable to include within- and between-region effects in the model. Lastly, dynamic data would, of course, allow for specific tie formation mechanisms to be explored more precisely.

Nonetheless, the models provide several insights into the organization of a drug trafficking operation. In particular, traffickers exhibit different levels of tie formation activity and homophily according to their specialization, while shared-partner effects emerge as an important factor in partnership formation. Whether large shared-partner effects are characteristic of low-trust environments warrants further attention.

4

The Influence of Network Embeddedness and Multiplexity on Homicide Arrests and Re-Arrests

Contents

4.1 Introduction 77
4.2 The social structure of arrests 80
4.2.1 Suspect profiling based on previous arrests 80
4.2.2 Co-arrest network embeddedness 83
4.2.3 Contribution to the literature 86
4.3 Data 88
4.3.1 Limitations 90
4.4 Methods 92
4.4.1 Network projection 92
4.4.2 Temporal network autocorrelation model 93
4.4.3 Multiplex interdependence model 95
4.5 Results 98
4.5.1 Temporal network dependence 98
4.5.2 Multiplex co-arrest interdependency 106
4.6 Discussion 116
4.7 Conclusion 119

Abstract

Research examining the risk factors and predictors of arrest usually examine the

influence of individual-level variables, such as age, ethnicity, and prior arrests. However, the role of social dynamics is largely unaccounted for. Criminologists have long noted that the opportunity structure for criminal behavior is shaped, in part, by social ties. Meanwhile, suspect profiling may lead to repeated co-arrests on the same pairs and groups of individuals. Using data on arrests in Chicago during 2000–2013, I examine whether embeddedness in co-arrest social structures differentially influences the probability of re-arrest for homicide relative to non-homicide violent crimes, drug offenses, and weapons violations. The co-arrest network is comprised of all arrests in Chicago, where individuals are tied if they sustained an arrest together for the same crime. I find that individuals embedded in co-arrest relationships and transitive co-arrest triads were less likely to be arrested for homicide than non-homicide crimes in the following year. I then examine the multiplex interdependence between co-arrests for homicide and co-arrests for non-homicide violent crimes, drug offenses, and weapons violations. Multiplex edges are present when a pair of individuals are co-arrested together for two different crimes. I find that pairs of individuals that sustained a co-arrest for a non-homicide violent crime, drug offenses, or weapons violations, are considerably more likely to be co-arrested for homicide (or to have been co-arrested for homicide) than would be expected by chance. This paper sheds light on the role of social ties in the risk of arrests, re-arrests, and recidivism.

4.1 Introduction

The majority of arrest events involve police contact with one suspect. As such, the risk of arrest and criminal offending is usually analyzed at the individual-level, accounting for traditional risk factors such as age, gender, ethnicity, and prior criminal records (Ousey and Lee, 2008; Steffensmeier et al., 1989; Harer and Steffensmeier, 1992; LaFree and Drass, 1996). However, a substantial fraction of arrest events involve two or more individuals, who are suspected of jointly committing a crime. In some cases, these arrests do not lead to charges or incarceration. Over periods of time, recurrent arrests on the same individuals and groups leaves behind a complex structure of co-arrests.

Despite the theoretical expectation that social ties play an important role in recurrent criminal behavior (Gendreau, Little, and Goggin, 1996), the structure of co-arrests is typically overlooked as a risk factor in empirical analyses of re-arrest and re-offending. However, incidents of co-arrest are fundamentally underpinned by social relationships between individuals, which form part of the opportunity structure for future criminal behavior (Laub and Sampson, 1993; Carrington, 2011). Moreover, prior co-arrests may inform suspect profiling by police (Kohler-Hausmann, 2013; Tyler, Jackson, and Mentovich, 2015), further contributing to the risk of re-arrest. Indeed, previous studies have noted the potential contribution of social ties to the incidence of crime and re-offending (Kirk, 2009; Wright et al., 2001).

This paper examines the influence of the structure of co-arrests on re-arrests for homicide vis-a-vis other crimes. The influence of embeddedness in co-arrest social structures may vary according to the type of crime for which a person is re-arrested. For instance, membership in a co-arrest relationship may have a different impact on the probability of future arrests for homicide relative to other violent crimes due to differences in the nature of these criminal behaviors, or due to differences in the opportunity structure afforded by co-arrest relationships. This paper puts this idea to the test by evaluating the effect of membership in several characteristic network structures, such as dyads and triads, on the odds of re-arrest for homicide relative to other violent crimes such as aggravated assault, as well as non-violent crimes

such as drug offenses. Additionally, the paper analyzes the effect of isolated prior arrests, where an individual is arrested alone, on the odds of re-arrest for these different crimes. Differential effects of co-arrest embeddedness would suggest that co-arrest ties are not only a risk factor in recurrent contact with the police, but also that this risk varies for different types of criminal behavior.

Drawing on the social networks literature, I therefore analyze the dependency of homicide re-arrests relative to non-homicide re-arrests on individuals' membership in co-arrest network structures. Sociologists have long held that network position and network embeddedness are important for understanding social phenomena (Burt, 1982; Granovetter, 1973; Granovetter, 1985). I argue that embeddedness in co-arrest networks is not merely an artifact of prior contact with criminal enforcement agents, but also reflects underlying social relationships that form part of the backdrop for future criminal behavior. Moreover, contemporary policing strategies, such as "maintenance-order policing," which involve intensive repeated contact with individuals deemed likely to be involved in crime, contribute to differential levels of individual embeddedness according to crime types. For example, recurrent arrests are more commonplace for minor crimes, such as vandalism, than for major crimes, such as homicide. This leads, in turn, to complex associations between embeddedness in the co-arrest and the probability of re-arrest for homicide and non-homicide crimes.

In most cases, arrests carry severe negative consequences for individuals, even if the charges associated with the arrest are subsequently dropped. Repeated contact with the criminal justice system negatively impacts employability (Bushway, 1998; Pager, 2003), while arrests contribute to social marginalization where individuals are marked as "deviants" with "all the attendant social meanings, physical discomforts, and civil burdens" (Kohler-Hausmann, 2013, p. 374). Furthermore, repeated police contact can engender widespread legal cynicism, particularly when that contact involves police violence (Desmond, Papachristos, and Kirk, 2016), which adversely effects the legitimacy and effectiveness of policing and the judicial system (Kirk and Matsuda, 2011). At the individual-level, arrests increase the risk of re-arrest

(Sherman et al., 1992), indicating that sustaining arrests can have long-term and increasingly negative consequences. It is therefore important to understand the social risk-factors that underlie recurrent arrests.

My research design uses arrest data in Chicago over a fourteen year period, 2000–2013. The data comprise individuals nested in arrest events, a classification of the type of crime relating to each arrest, and demographic characteristics of arrestees. A substantial fraction of arrest events involve more than one individual and, combined with repeated arrest events on the same individuals, these arrests amalgamate into a large co-arrest network. Using temporal network autocorrelation models, I first analyze the influence of embeddedness in the co-arrest network over one year periods on the incidence of homicide arrests vis-a-vis arrests for other crimes in the following year.

I then examine the interdependency of co-arrests for homicide and co-arrests for other violent crimes, drug offenses, and weapons violations. This "multiplex interdependency" captures the extent to which homicide co-arrests for a particular pair of individuals depends on their isolated arrests and embeddedness in the structure of co-arrests for non-homicide crimes. The results show that homicide arrests are more likely, relative to non-homicide arrests, for individuals who were arrested in the previous year. However, prior embeddedness in co-arrest dyads and in higher-order structures, such as transitive co-arrest triads, reduces the likelihood of homicide arrest vis-a-vis other crimes. I also find multiplex interdependency between homicide co-arrests and co-arrests for other violent crimes, drugs offenses, and weapons violations. Pairs of individuals that were co-arrested for aggravated assault, for example, are substantially more likely to be (or to have been) co-arrested for homicide, while co-arrests with mutual third-parties for non-homicide crimes is associated with increased odds of homicide co-arrest.

The paper proceeds as follows. In Section 4.2 I outline the roles of suspect profiling based on prior arrests and co-arrest network embeddedness in the incidence of re-arrest. I highlight how the study addresses a gap in the literature concerning the role of social ties in re-arrest and re-offending. In Section 4.3 I introduce

the data and note limitations of the study. I explain the methods for projecting the arrest network, before describing the temporal network autocorrelation model and the multiplex interdependence model that make up the empirical analyses in section 4.4. I describe the results in Section 4.5, and discuss the implications of these results in section 4.6. Finally, I conclude in Section 4.7.

4.2 The social structure of arrests

I begin by drawing out mechanisms through which prior isolated arrests and the social structure of arrests may influence re-arrest and co-arrest patterns. I define re-arrest as individuals sustaining two or more arrests and co-arrest as instances where two or more individuals are arrested together. I outline two channels: suspect profiling based on isolated prior arrests, and embeddedness in co-arrest relationships. Although I am not able to analytically separate these channels in the empirical part of this paper, I outline how each can be expected to influence patterns of arrest and re-arrest.[1]

4.2.1 Suspect profiling based on previous arrests

The first channel focuses on how prior isolated arrests—that is, arrests pertaining to an individual rather than a group of individuals—shape police profiling of suspects for a given crime. Police profiling and disproportionate patterns in police arrests have long been the subject of criminology research (Smith, 1986). Broadly, this research has focused on individual-covariates and neighborhood-level factors associated with increased suspicion and police-community contact. For example, a meta-analysis of 131 studies into criminal recidivism (Gendreau, Little, and Goggin, 1996) showed that age, criminal history, gender, and substance abuse are significant predictors of re-offending. Generally, younger people, black males, and

[1]For the purposes here, I define re-arrests as subsequent arrest events for individuals with a previous arrest. Individuals may be re-arrested for the same crime or for a different crime to that of the original arrest. Moreover, individuals who are arrested or re-arrested have not necessarily committed a criminal offense. I refer to recidivism, on the other hand, as criminal re-offending. For a definitional discussion of criminality, recidivism, and re-offending, see Laub and Sampson (2009).

those with multiple prior arrests are more likely to be arrested and re-arrested (Ousey and Lee, 2008; Pettit and Western, 2004; Steffensmeier et al., 1989), and low socioeconomic status is related to the risk of repeated criminal justice contact (Harer and Steffensmeier, 1992; LaFree and Drass, 1996; Messner, Raffalovich, and McMillan, 2001; Western, 2006).[2] Many studies show that black individuals are more likely to be arrested than whites, and blacks are more likely to be arrested for serious crimes (Black, 1971; Blumstein, 1982; Cureton, 2000; Decker and Kohfeld, 1985). Tyler, Jackson, and Mentovich (2015) report that individuals living in disadvantaged neighborhoods, minorities, young people, and poorer people are more likely to report feeling suspected by the police. Moreover, Smith (1986) showed that, after controlling for suspect behavior and type of crime, suspects in disadvantaged neighborhoods were more likely to be arrested.

There is evidence that suspect profiling by police is based, in part, on prior arrests and criminal records. Indeed, a longstanding problem in criminology is that arrests increase the frequency of re-arrest for arrested individuals (Sherman et al., 1992). In the case of minor crimes, arrests frequently do not lead to prosecution or incarceration. Tyler, Jackson, and Mentovich (2015) argue that police frequently use arrests, and in particular minor-crime arrests, to place individuals into a period of "tracking" in order to gain information on whether individuals can adhere to the law, in effect allowing police to build suspect profiles based on uncharged arrests. Charges relating to these arrests may later be dismissed after a period of time involving no or few infractions (Tyler and Wakslak, 2004). Consistent with police using arrests in suspect profiling, individuals with prior criminal records and contact with the judicial system are more likely to report feeling suspected (Tyler, Jackson, and Mentovich, 2015).

Kohler-Hausmann (2013) describes the explicit police strategy of intensive arrest for low-level offenses as an instrumental "crime-control strategy," variously

[2]There are important differences in rates of recidivism and arrest within these covariate groups. For example, the precise influence of age on the risk of arrests varies across crime types (Steffensmeier et al., 1989). However, examining these differences is beyond the scope of this review.

known as "broken windows policing," "intensive community engagement," and "order-maintenance policing." This strategy allows criminal justice actors to "mark and classify recent police contacts" that can be "built upon by successive arrests" (Kohler-Hausmann, 2013, p. 371). In essence, prior arrests can be drawn upon as a resource to inform future suspect profiling, incentivising the use of first-time arrests to build a pool of suspect profiles. The use of prior arrests to identify suspects for reported crimes, and the increased police contact with individuals who have been arrested, may both therefore contribute to high incidence of re-arrest.

In line with the role of arrests as a suspect profiling strategy in policing, re-arrests are somewhat common occurrences for minor crimes and misdemeanors. However, empirical evidence for whether re-arrests for major crimes are based, in part, on a similar type of suspect profiling through prior arrest records is largely missing from the criminology literature. Analytically, re-arrests for major crimes could follow arrests for minor crimes: a form of arrest escalation. Alternatively, re-arrests for major crimes could follow arrests for other major crimes: a form of serial major arrests. Or, major crime re-arrests could be relatively rare events.

In the present study, I analyze re-arrests for major crimes by examining the incidence of arrests for homicide in a particular year, and assessing the influence of arrests and non-arrest in the previous year. The homicide arrests pertain to individuals being arrested as isolates, but prior arrests could be either sustained in isolation or in co-arrest groups. In a network setting, individuals who were previously arrested alone would be represented as isolates, i.e. they would have no adjacent nodes, in a lagged co-arrest network. In the case of previous arrests sustained in groups involving two or more individuals, the arrested individuals would be connected together by edges in the lagged co-arrest network, forming dyadic, triadic, and higher-order groups. From the perspective suspect profiling based on prior arrests, as outlined above, I hypothesize that prior isolated arrests exert a significant influence on the likelihood of short-term homicide re-arrests.

4.2.2 Co-arrest network embeddedness

The second channel addresses the extent to which social embeddedness in co-arrest networks influences re-arrests. Arrests are typically understood as contacts between police and a single individual. However, individuals are often arrested in pairs or in larger groups. When individuals are recurrently arrested, sometimes in the same pair combinations or in groups with new individuals, these co-arrest relationships amalgamate into larger co-arrest network structures. Importantly, the incidence of arrest at any give time is not blind to this co-arrest network; there are two ways in which the network itself may influence future arrests.

First, from the perspective of the police, the co-arrest network is effectively searchable. While it may not be the case that police map the co-arrest network per se, arrest records nevertheless contain information on the people with whom an individual has been arrested, and who those people have been arrested with, and so on. This suggests that police can "leverage" the co-arrest network in deriving suspects for arrest.

Secondly, the network is an artifact of "on-the-ground" social relationships. Arrested individuals are embedded in complex patterns of co-arrest, and the incidence of future arrests is plausibly influenced by endogenous effects stemming from the social relationships underpinning the co-arrest network. Prior to the present study, the dependence of arrests on existing co-arrest networks has not been formally investigated.

To understand how arrests may depend on embeddedness in the co-arrest network, I draw on the sociological literature on tie formation in social networks. In the remainder of this section, I explain how network effects may apply to the co-arrest network, and how I operationalize these effects in the present study. It is important to note from the outset a key difference in co-arrest tie formation and tie formation in the majority of social network studies. Specifically, whereas in social networks emphasis is placed on endogenous social processes that lead to tie formation, in co-arrest networks tie formation is based in large part on exogenous police decisions. While in the case of friendship, for example, individuals are often

assumed to decide, to a certain extent, with whom they will become friends, in the case of arrest, the police decide whether individuals will be co-arrested. Nevertheless, co-arrest ties are based on underlying social relationships; the formation of social ties is a precondition of co-arrest.

The formation of co-arrest ties may be influenced by endogenous structural processes, such as transitivity, and individual-oriented processes such as homophily. Transitivity is the tendency for local clustering in networks and is routinely observed in social networks, including in the previous two chapters in this thesis. In the case of co-arrests, this tendency means that two individuals who have both been co-arrested with the same third-party are more likely to be co-arrested themselves. This could be due to the existence of an unobserved but underlying social relationship between that pair of actors, which is more likely when those actors share a mutual social tie (Watts and Strogatz, 1998). Homophily captures the broad phenomena that individuals are more likely to be socially connected when they share characteristics (Goodreau, Kitts, and Morris, 2009). For instance, co-arrests may be homophilous by race, which would lead to the observed network exhibiting a greater frequency of co-arrests involving individuals of the same race than would be expected by chance, conditional on the racial composition of the arrest network. Previous research finds evidence for intra-gang ethnic homophily (Grund and Densley, 2015). Similarly, co-arrests may be homophilous by age and gender. Specifically interesting in the case of homophily is the role of gangs; one might expect members of the same gang to be substantially more likely to be co-arrested together than non-gang members. Broadly, I hypothesize that transitivity and homophily effects positively influence the future incidence of co-arrests.

Crucially, co-arrest ties between individuals may be present for more than one type of crime. For example, the same pair of individuals may have been co-arrested for homicide and aggravated assault, and therefore share a "multiplex co-arrest." In such a situation, it is possible that the co-arrest for homicide and the separate co-arrest for aggravated assault are interdependent. In other words, one co-arrest made the other more likely. The same multiplex interdependency may apply to

higher order network effects such as transitivity. Moreover, being arrested as an isolate for a particular crime may increase the likelihood of involvement in a co-arrests for a different crime. This could be due to the "criminal careers" of such individuals (Laub and Sampson, 2009), to the increased likelihood of being considered a suspect by the police, or to the presence of unobserved relationships with other individuals who are involved in criminal activity. Furthermore, the extent of multiplex interdependency in co-arrests may vary by the type of crime. Individuals with arrests for drugs violations, for example, may be less likely to be co-arrested with another individual for aggravated assault relative to an individual previously arrested for battery. I hypothesize that multiplex interdependency is a feature of co-arrests. In essence, pairs of individuals who have been co-arrested for one crime should more frequently be co-arrested for other crimes. Similarly, I expect nodes with a mutual co-arrestee for one crime to be more often co-arrested for another crime than would be expected by chance.

I operationalize the influence of network embeddedness on the incidence of co-arrest by examining the co-arrest network for homicide formed by 14 years of homicide arrests. If network embeddedness influences the formation of co-arrest ties, I expect to find significant transitivity and homophily effects for homicide. To account for multiplex co-arrest interdependency, I construct co-arrest networks for six other violent crimes and for drug violations. I model whether homicide co-arrests and arrests, co-arrests, and higher-order arrest structures in these non-homicide co-arrest networks are interdependent. If there is multiplex interdependency between co-arrests for homicide and arrests for different crimes, the likelihood of co-arrests for homicide should be positively associated with arrests for non-homicide crimes, and transitivity in non-homicide arrests should influence the likelihood of co-arrests for homicide.

Finally, it is important to point out that the suspect profiling and the social structure of arrests channel are not entirely distinct processes. Rather, they may jointly influence the likelihood of re-arrests. For instance, individuals who are suspected of committing a crime due to their prior criminal records may not only

themselves be more likely to be re-arrested, but the people with whom they have previously been arrested may also face greater risk of re-arrest. This is analogous to a "social contagion" in social networks (Burt, 1987; Shalizi and Thomas, 2011), where phenomena pertaining to a particular node spreads to surrounding nodes. In the case of re-arrest, such contagion may be likely to occur for crimes with suspected gang involvement, for example; suspicion that a particular gang-affiliated individual may have been involved in perpetrating a crime could lead to re-arrests of that person's associates. In this way, the social structure of previous arrests and suspicion profiling operate jointly.

4.2.3 Contribution to the literature

The richness of the data used in this paper stems from the social ties between individuals underlying co-arrest events, which is typically unaccounted for in cross-sectional studies of crime and re-offending. Indeed, this gap in the literature was identified two decades ago. Gendreau, Little, and Goggin (1996) called for research into "dynamic risk factors" in re-offending, noting the plausible influence of "companions," defined as socialization with other offenders, while it has long been thought that social ties may contribute to criminal behavior (Cohen and Felson, 1979; Venkatesh, 1997; Wright et al., 2001; Hagan, 1993). More recent empirical studies suggest that social risk factors may be pertinent in criminality. Finding that resettlement in new residential locations following prisoner re-entry reduces recidivism rates, Kirk (2009) notes that the "causal mechanism underlying the effect of migration may be related to social ties," which are more likely to be discontinued with greater residential separation, providing "less opportunity and provocation for engaging in crime" (Kirk, 2009, p. 500). Papachristos (2009) demonstrated a clear network structure in the perpetration of, and victimization by, homicide in Chicago, in which the inter-connected structure of gang conflicts was central to understanding the incidence of gang-involved homicide.

The present study addresses this gap by focusing on the social dynamics of homicide arrests and co-arrests vis-a-vis other crimes. It is my contention that arrest

rates are driven largely by individual-level and neighborhood-level factors such as age, gender, ethnicity, and economic and social disadvantage. However, criminal behavior and the policing response to criminality are intrinsically social in nature. Crimes are not always committed alone, and policing is not blind to the social structure among suspected criminals. The purpose of this study is to shed light on the social factors influencing the incidence of arrests, specifically in the context of homicide.

An exogenous influence of the co-arrest structure for non-homicide crimes on homicide arrests and co-arrests would indicate that arrests, or at least major-crime arrests, take place in a network context. This has implications for how we understand the risks and consequences of arrest, particularly in terms of the heightened levels of arrest and police community interaction in disadvantaged neighborhoods. For individuals, the consequences of arrest can negatively affect employment prospects and wellbeing (Pager, 2003). If the risks of arrest are intrinsically linked to embeddedness in the co-arrest network, and the influence of embeddedness varies by the type of crime, then individuals' risk of arrest may not only be influenced by their previous arrest record, but also by arrests within their local co-arrest network.

In addition to the consequences of network embeddedness on the risks of arrest, recent studies indicate that repeated police-community contact and heightened suspicion can undermine trust in police (Tyler, Jackson, and Mentovich, 2015). Moreover, repeated arrests and intensive, but ineffective, police intervention can contribute to legal cynicism, adversely affecting the effectiveness and legitimacy of policing and the law as individuals no longer feel they can rely on agents of law enforcement (Kirk and Papachristos, 2011; Kirk and Matsuda, 2011; Sierra-Arévalo, 2016). Although I do not examine the consequences of network embeddedness on legal cynicism in this study, it is plausible that repeated arrest contact on specific individuals, pairs of individuals, and groups—particularly arrests for which charges not brought—could contribute to increased cynicism towards the role of police. I return to the consequences of network embeddedness in co-arrests networks in Section 4.6.

4.3 Data

To examine the temporal dependency of homicide arrests at time t on the prior co-arrest network at $t-1$, and the multiplex interdependency of homicide and non-homicide co-arrests, I use data on city-wide arrests in Chicago during the period 2000–2013. The data are based on arrest records compiled by the Chicago Police Department (CPD), and comprise 712,050 individuals, nested within 2,255,764 arrest events. Each arrest event is assigned a unique identifier, ensuring that individuals arrested together in groups, containing at least two people, are correctly interpreted as having been co-arrested. Similar data are used in Papachristos, Braga, and Hureau (2012).

The data include information on age, ethnicity, and gender for each individual arrested during the time period. I couple this information with CPD database on gang affiliation. This allows me to identify individuals who have been identified, by the CPD, as gang affiliated. Each arrest event is time-resolved by day, allowing the order of arrests to be established. Importantly, the data also contain a classification of the suspected crime associated with each arrest event. In the present paper, in order to examine the temporal dependence of homicide arrests on the prior co-arrest network, and the multiplex interdependency of homicide co-arrests and co-arrests for non-homicide crimes, I rely on arrest events related to the following categories of crime: simple assault, aggravated assault, simple battery, aggravated battery, miscellaneous drugs offenses, robbery, weapons, and homicide.

To illustrate the data, Figure 4.1 shows the number of arrests per month and the number of individuals per homicide arrest event during the period 2000–2013. During 2000–2004, there were approximately 40 individuals arrested for homicide within 35 arrest events per month. The number of arrests generally trended downwards over the period, with the lowest number of arrests taking place in 2010–2011. Homicide arrests exhibit a seasonal trend; the number of arrests is greater in the summer but decreases in the winter.[3] The majority of homicide arrest events involved just one individual. However, more than 500 arrest events

[3]See Cheatwood (1988) for a review of the evidence on the seasonality of homicide.

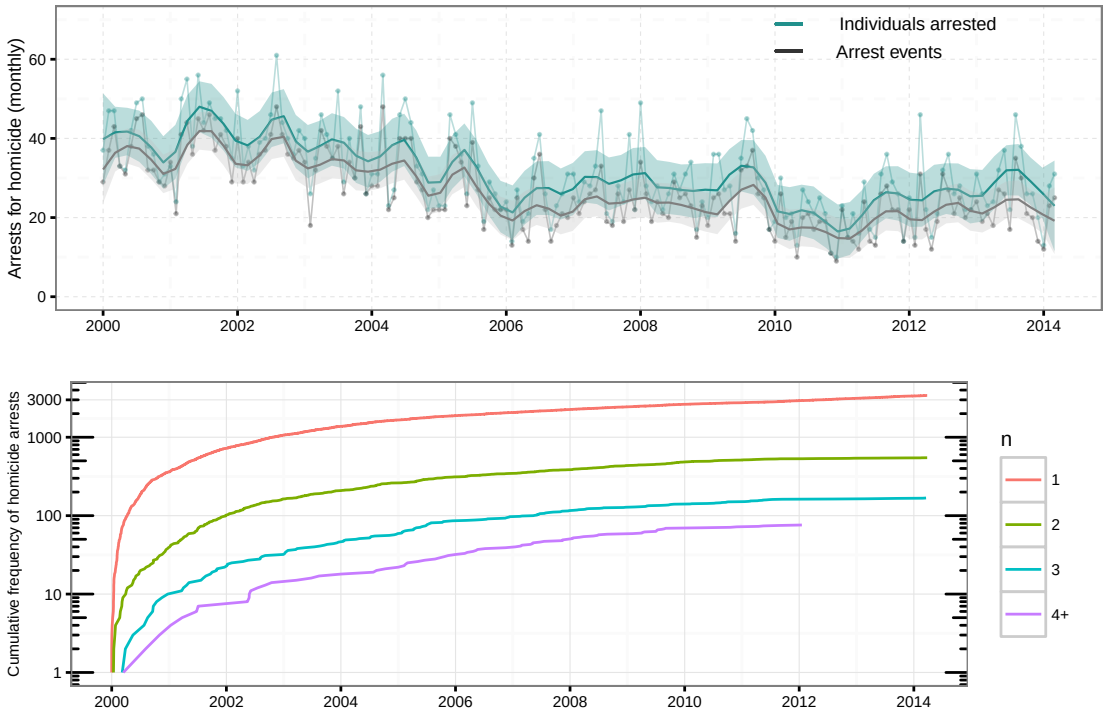


Figure 4.1: Top: Monthly frequency of arrests for homicide in Chicago, 2000–2013. The green line counts the number of individuals arrested for homicide each month, and the grey line counts the number of arrest events per month. The frequencies are smoothed to more clearly demonstrate trends. Bottom: Cumulative frequency of homicide arrest events involving n individuals. Note the log scale for the y-axis.

involved two individuals, 140 events involved three individuals, and around 80 arrest events involved four or more individuals. Thus, while isolated arrests—arrests involving just one individual—constitute the majority of homicide arrest events, co-arrests nevertheless represent a substantial share of all arrest events. To reinforce the prevalence of homicide co-arrests, Figure 4.2 shows the average number of individuals nested within arrest events by crime type. Homicide ranks second in the count of unique individuals per arrest, averaging around 1.20. Accounting for re-arrests, there are approximately 1.27 individuals per homicide arrest event.

4.3.1 Limitations

Although the study makes a unique contribution on the temporal and multiplex interdependency of homicide arrests and other arrests based on these data, it nevertheless has an important limitation. Specifically, I do not have access to incarceration data relating to the arrests. That is, the data do not include an indicator of whether the individuals were incarcerated, nor the period of this incarceration, following arrest. Additionally, I do not have data on attrition from the arrest network due to other causes, such as death or geographic mobility—i.e., leaving Chicago permanently. Consequently, I cannot include an "at-risk" control to indicate whether individuals' are capable of sustaining subsequent arrests. This means that, in principle, individuals could be observed as not sustaining an arrest event due to withdrawal from criminal activity, incarceration, or because the person had left the jurisdiction of the CPD. This is particularly relevant to the present study due to my focus on homicide arrests, which typically carry long periods of incarceration for individuals found guilty. Non-arrest is therefore very likely for a given individual, a priori, in the years following a homicide arrest. That the reason for non-arrest cannot be identified precludes a full examination of the temporal dependence of non-arrests for homicide on prior homicide arrests.

I make two modeling decisions to address this limitation. First, to examine temporal dependency, I include only individuals with an arrest at time t. The upside of this decision is that it ensures all individuals were at-risk of arrest. However, the

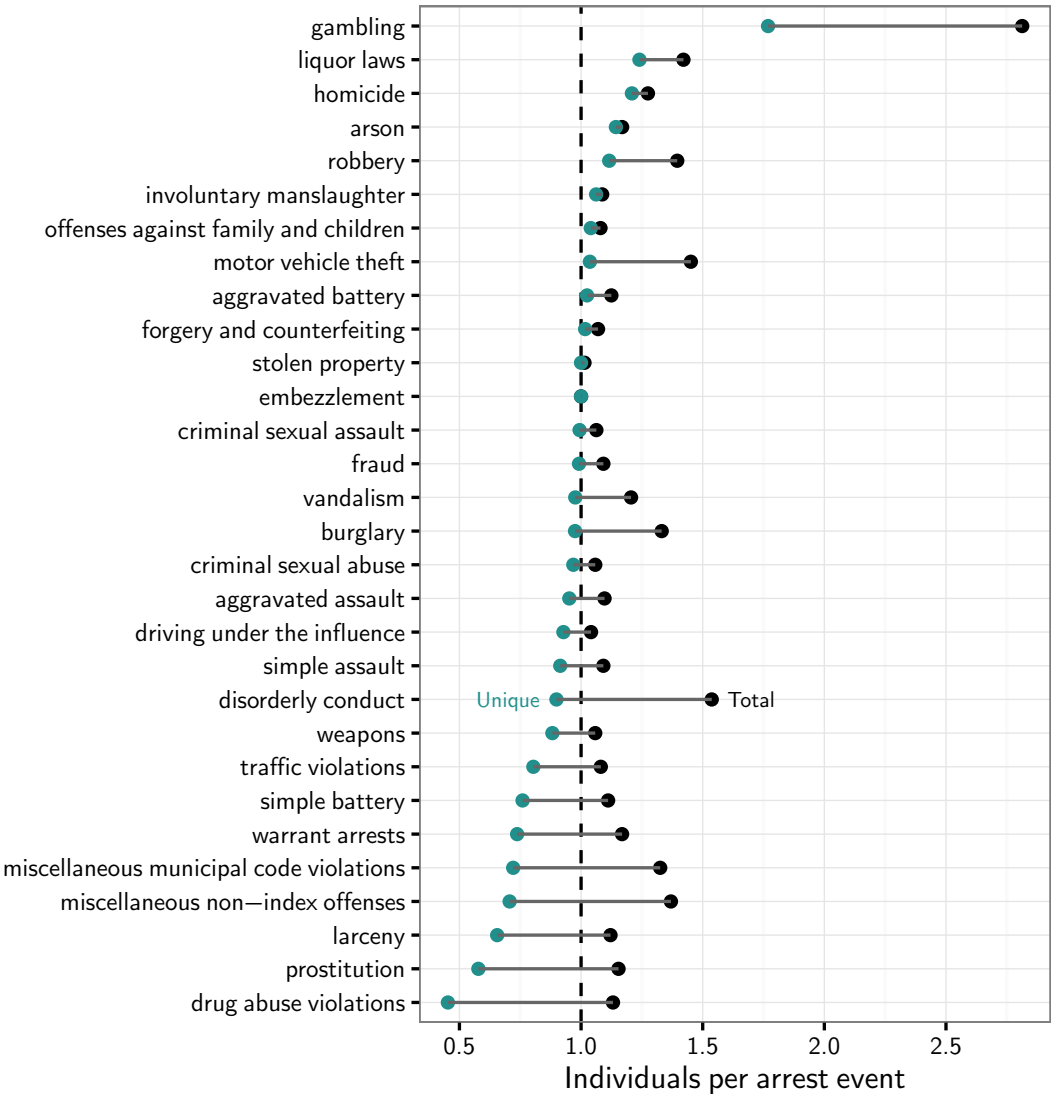


Figure 4.2: Mean count of individuals per arrest for different types of crime. The green points are the mean number of unique individuals per arrest event; smaller values indicate that individuals sustain more recurrent arrests, on average. The black points give the average number of people nested in arrest events for each crime; values closer to one indicate that co-arrests involving two or more individuals are less common.

downside is that the exclusion of non-arrests constitutes selection on the dependent variable. Importantly, a consequence of this decision is that I do not evaluate the likelihood of homicide arrests relative to non-arrests, but instead relative to arrests for non-homicide crimes. Secondly, to examine multiplex interdependency, I pool across all 14 years of data rather than including temporal parameters. Thus, I analyze the interdependency of co-arrests for homicide and for non-homicide crimes, rather than the directional dependency of homicide co-arrests on non-homicide co-arrests. These modeling choices are explained in more detail below.

4.4 Methods

The analysis involves two parts. In the first part, I analyze the temporal dependency of homicide arrests in a given year, t, on the structure of co-arrests for all crimes in the prior year, $t-1$. In the second, I examine the multiplex interdependency between homicide co-arrests and co-arrests for non-homicide crimes during the period 2000–2013. The first part requires the co-arrest network for all crimes, separated into one year intervals, while the second part requires the separation of the co-arrest network for homicide and co-arrest networks for a subset of non-homicide crimes, pooled over 2000–2013.

4.4.1 Network projection

To arrange the co-arrest network into the above specifications, I first construct a bipartite network of arrest events. In the bipartite network, or "two-mode" network, individuals are connected to the arrest event in which they are nested. Although one could analyze the bipartite network itself, for the purposes here, I project the bipartite network into a unipartite network, or "one-mode" network, in which individuals whom have been arrested together are directly connected. This full unipartite network, spanning all arrests throughout 2000–2013, includes 712,050 individuals and 1,004,669 co-arrest connections. The full co-arrest network is reasonably sparse; the mean number of co-arrestees per individual is 1.41. The

network includes arrests for 30 different classes of crime, including arson, aggravated assault, and fraud, for example.

4.4.2 Temporal network autocorrelation model

To examine the dependence of homicide arrests in year t on arrests in the previous year $t-1$, I split the co-arrest network into one-year periods and estimate a series of rare-event temporal network autocorrelation models (TNAM). The boundaries for each time period are January 1–December 31. Although the selection of these boundaries is arbitrary, as any 12-month period could be selected in principle, a split point in the winter ensures that arrests in the summer, which are more common for homicide, as noted above, do not fall into separate time periods. By splitting at a time when the density of homicides arrests is relatively low, we ensure that arrests that occurred in proximity in real-time but which fall into separate time periods due to the boundaries represent as small a fraction as possible of all arrests in each year period.

The year-by-year networks are very large, which presents challenges for the temporal network autocorrelation model. The TNAM is analogous to a generalized linear model (Doreian, Teuter, and Wang, 1984), in which we model a left-hand side dependent variable by estimating a set of coefficients associated with right-hand side parameters, plus an error term. The problem with the generalized linear framework is that, with temporal dyadic data, the errors cannot be assumed to be independently and identically distributed (i.i.d.), but are instead correlated. To account for autocorrelation in the outcome y_t on the lagged outcome y_{t-1}, the TNAM includes an autocorrelation parameter, which is estimated as part of the linear equation:

$$y_t = \rho y_{t-1} W + X\beta + Z\gamma + \epsilon, \quad \epsilon \sim N(0, \sigma^2 I) \tag{4.1}$$

where y is a binary vector that indicates whether an actor has been arrested for homicide, W is a matrix which accounts for differential influence of y_{t-1} on y_t across actors, β is a vector of coefficients associated with the individual-level covariates X, γ is a vector of coefficients associated with the network covariates Z, and ϵ is the

error term (Fujimoto, Chou, and Valente, 2011). The ρ parameter estimates the autocorrelation. The W matrix specification is similar to allowing random effects on the nodes. The formula in 4.4.2 is known as the "network effects" variant of the TNAM (Doreian, Teuter, and Wang, 1984). In short, I estimate the incidence of homicide arrests at t as a function of homicide arrests at $t-1$ and a set of parameters capturing individual-level covariates, such as age and ethnicity, and network-level covariates, such as transitivity, in the co-arrest network at $t-1$. The estimates are expressed as odds, as in a logit model for a binary outcome.

The first challenge emerges due to the network statistics in the $t-1$ co-arrest network that I include in the right-hand side of 4.4.2. Each network statistic must be a $n * n$ matrix, where n is the number of actors in the co-arrest network at $t-1$, in order to account for each dyad. The larger the network at $t-1$, the larger these matrices become, making the likelihood estimation procedure used to fit the model to the data computationally burdensome.[4] For the model to be tractable, the co-arrest data must therefore be sampled.

However, sampling the co-arrest data poses a second challenge: ensuring that the sample contains a sufficient number of homicide arrests. The problem is that homicide arrests are rare events, accounting for less than 0.5 percent of all arrest events. A random sample of arrests will therefore lead to few homicide arrests being modeled on the left-hand side, which in turn means that the parameter estimates will be noisy, and in the worst case scenario could result in artificial model separation due to little or no variation in the outcome variable conditional on the covariate matrices X and Z (Heinze and Schemper, 2002). To ensure that the sample includes a sufficient number of homicide arrests, I therefore use a sampling strategy known as endogenous stratified sampling (ESS) (in econometrics) or case-control design (in epidemiology) (King and Zeng, 2001b).

Endogenous stratified sampling involves selecting on the dependent variable: here, homicide arrest at time t. I sample all of the homicide arrests at t and then randomly sample an equal number of non-homicide arrests at t. I then identify

[4]The likelihood function for the TNAM is detailed in Doreian, Teuter, and Wang (1984).

the individuals nested within these arrest events, taken together, and identify their position in the co-arrest network at $t-1$, provided they were arrested at $t-1$. Indeed, to capture whether nodes were present in the co-arrest network at $t-1$, I include a parameter for lagged arrests in the TNAM. The network terms in the right-hand side are then calculated using the subgraph of the co-arrest network defined by the sampled individuals plus their adjacent neighbors—the nodes with whom the sampled individuals have been co-arrested at $t-1$. The inclusion of adjacent neighbors in the calculation of the transitivity term ensures that transitivity is not understated through the exclusion of third parties due to the sampling framework.

The ESS framework leads to the third and final challenge. Nonrandomly sampled data can bias coefficient estimates. To resolve this problem, I use a statistical correction known as "prior correction." Prior correction has a long history in both the econometrics literature (Manski and Lerman, 1977) and the biostatistics literature (Prentice and Pyke, 1979), and was introduced in an applied social science setting by King and Zeng (2001b). Essentially, the correction requires prior information on the proportion of homicide arrests in the population of all arrests, denoted τ. Although there can be uncertainty regarding τ in many cases, particularly when applying the ESS framework to a sample, in the present study I sample from the population of arrests. Accordingly, I correct the parameter estimates using the exact population τ (Tomz, King, Zeng, et al., 2003). This application of the ESS framework and prior correction to a TNAM (i.e., a rare-event temporal network autocorrelation model) is, to my knowledge, the first of its kind.

4.4.3 Multiplex interdependence model

To analyze the interdependence of homicide arrests and non-homicide arrests, I separate the full co-arrest network into eight network "layers." Each layer represents the co-arrest network for a specific type of crime. Thus, taken together, the networks constitute a multiplex set of co-arrest relationships. Individuals who have been arrested for more than one crime, and therefore appear in at least two of the layer networks, are multiplex nodes. Similarly, co-arrest relationships

that exist in more than one layer are multiplex edges. The eight layer networks are made up of co-arrests in the period 2000–2013 for homicide, simple assault, aggravated assault, simple battery, aggravated battery, robbery, drugs violations, and weapons violations.

I estimate the interdependence of co-arrests in the homicide network and co-arrests in a non-homicide layer network using a set of exponential random graph models (ERGMs) (Robins, Snijders, et al., 2007; Snijders, Pattison, et al., 2006). In the left-hand side of the ERGM I model the homicide co-arrest network. In the right-hand side, I include two groups of parameters. The first group are endogenous parameters capturing the propensity for co-arrests in the homicide network as a function of the homicide network itself. The second group are exogenous multiplex parameters, which capture the influence of arrests and co-arrests for a specified non-homicide crime on co-arrests for homicide. A similar approach based on exogenous multiplex parameters is used in Heaney (2014). I estimate a separate ERGM for each type of multiplexity, i.e. homicide–aggravated assault multiplexity. This modeling approach enables a comparison of homicide interdependence by crime type. The ERGM is estimated as follows:

$$p(x) = \exp\{\theta^{\intercal} z(x) - \psi(\theta)\} \tag{4.2}$$

where $p(x)$ is the probability of the observed homicide co-arrest network, θ is a vector of coefficients associated with the graph statistics $z(x)$, and $\psi(\theta)$ is a normalizing constant. For the multiplex interdependence models, $z(x)$ is comprised of the first group of graph statistics that capture features of the homicide network and the second group of statistics capturing the structure of layer network co-arrests between the nodes in the homicide graph. For example, I include a "homicide edges" term in $z(x)$ to measure the occurrence of homicide co-arrests in addition to an "aggravated assault edges" term to capture the influence of co-arrests in the aggravated assault network (between nodes that are also members of the homicide network) on homicide co-arrests.

An alternative approach to modeling multiplex interdependency is to estimate a bivarate ERGM. Whereas the above approach essentially parameterizes each layer network as an exogenous property of the homicide network, in the bivariate ERGM approach one would model the endogenous features of homicide network and the layer network jointly alongside cross-network terms to account for interdependence (Smith and Papachristos, 2016). The latter approach has several desirable properties. Specifically, it allows for endogenous effects in the right-hand side network to influence the endogenous terms in the left-hand side network through cross-network interdependency, so that endogenous features of co-arrests for aggravated assault could influence the incidence of co-arrests for homicide, and vice versa. The approach here, on the other hand, does not model the endogenous features of the layer co-arrests, but more simply assumes they exert an exogenous influence on homicide co-arrests. The bivariate ERGM approach is infeasible in the present study due to the large number of nodes in the homicide network—the network includes more than 5,000 nodes—which renders estimation computationally burdensome.

In order to estimate the multiplex interdependence models with several endogenous and exogenous terms on this relatively large network, I use maximum psuedo-likelihood estimation (MPLE). The MPLE ignores the dependence between edges that we model, for example, through transitivity or other higher-order terms in the ERGM. Thus, for models in which one assumes that edges are independent, the MPLE and MLE are indentical. In this situation, the MPLE is tantamount to the maximum likelihood estimate (MLE) for a logit model. Because the MPLE does not account for higher-order dependency, it is more straightforward to estimate. The MPLE is therefore advantageous for modeling large networks (Duijn, Gile, and Handcock, 2009).

However, simulation studies (Corander, Dahmstrom, and Dahmstrom, 1998; Robins, Snijders, et al., 2007) and an applied analysis (Lubbers and Snijders, 2007) have shown that MPLE can produce biased estimates and standard errors. Robins, Snijders, et al. (2007) show that the divergence between the MPLE and MLE is larger in networks with greater dependency between edges, such as might be the case in a

network comprised of a complex system of edges involving many nodes. Conversely, the MPLE and MLE are similar in networks with low dependency (Saul and Filkov, 2007). As noted in section 4.3, in the co-arrest data, higher order structures are relatively uncommon because the majority of individuals are arrested as isolates or in pairs. Accordingly, the level of higher-order dependency is relatively low in the co-arrest networks. MPLE is therefore reasonable in the present application and should produce estimates approximately consistent with MLE. It is nevertheless important to note that, despite the low level of higher-order dependency in the co-arrest network, the parameter estimates and errors could be non-trivially biased.

4.5 Results

This section first presents the results of the temporal network autocorrelation model and then presents the multiplex interdependence model of arrests for homicide and arrests for other violent crimes. I begin by reconstructing the co-arrest network in Chicago during the years 2000–2013. In the network autocorrelation models, the co-arrest network is split into one year time periods. For the multiplex interdependence models, the co-arrest network is split by crime type into a set of network layers. Descriptive statistics and model estimates are presented below.

4.5.1 Temporal network dependence

Recall that in the temporal dependence model, I estimate separate network autocorrelation models for each year. I do this for two reasons. Firstly, even under the endogenous stratified sampling (ESS) framework discussed in section 4.4, estimating a pooled model is computationally impractical due to an excessive number of dyads. Secondly, estimating separate models allows the parameters to vary between time steps, capturing the implicit variation in year-to-year autocorrelation, thus also permitting the temporal dependence of homicide arrests at t on arrests at $t-1$ to strengthen or weaken over time. As noted in the methods section, individuals must have been arrested in year t, whether for homicide or a different crime,

in order to be included in the model. This partial selection on the dependent variable is justified in section 4.4.

Table 4.1 summarizes the yearly co-arrest data after ESS. The mean frequency of homicide arrests per year was 362. Overall, homicide arrests trended downwards, with an up-tick in 2013, in line with the frequency of homicide victimizations in Chicago during the same period (Chicago, 2016). The proportion of nodes in the co-arrest network at time t who were also arrested at $t-1$, after sampling, was relatively stable from year to year, averaging 0.71. This partially reflects the sampling schema, which over-samples repeat offenders by selecting on homicide arrests. A small fraction of individuals arrested at t were also arrested for homicide at $t-1$. These individuals were presumably acquitted or charges were dropped. The majority of those arrested at t in the sample had an adjacent co-arrestee at $t-1$, while around 40 percent were arrested alone—i.e, were isolates—at $t-1$. The average fraction of homicides per arrest-year is just $\tau = 0.003$, clearly marking homicide as a rare event. Under the ESS framework we exclude approximately 99% of nodes without a homicide arrest at t, similar to the proportion dropped in King and Zeng (2001a). An average of 2288 individuals remain, including those arrested for homicide, leaving a post-ESS fraction of homicide arrests per arrest-year of $\bar{y} = 0.158$. Note that the year 2000 is excluded due to the year 1999 falling outside of the data range.

To demonstrate the temporal dependence networks more clearly, Figure 4.3 shows the networks at $t = 2012$ and $t = 2013$. Individuals arrested for homicide in year t are highlighted in purple. Referring to Table 4.1, a proportion of 0.79 of the nodes shown in the 2013 network were arrested in 2012, and 0.69 of the nodes shown in the 2012 network were arrested in 2011.[5] The majority of nodes in the temporal networks were isolates—they were arrested alone. However, clear localized structures are present in both networks, ranging from simple connected dyads and triads to larger, more complex structures. In particular, the 2012 network contains a large connected component suggesting, perhaps unsurprisingly, that while arrest events are discrete and separate events, usually involving fewer than four individuals,

[5]For nodes to appear in the 2013 and 2012 networks, they must have been arrested in both years because inclusion in the network is conditional upon arrest at t.

Homicide					All Arrests					
t		$t-1$			$t-1$					
Year	Homicide	Arrest	Homicide	Adjacent	Year	Nodes	Edges	Dyads	ES	Sample
2001	513	0.68	0.003	0.52	2000	2841	4429	8 071 281		0.991
2002	496	0.66	0.005	0.49	2001	2432	8024	5 914 624		0.992
2003	442	0.72	0.006	0.61	2002	3153	15110	9 941 409		0.988
2004	417	0.69	0.008	0.58	2003	2356	2990	5 550 736		0.992
2005	368	0.73	0.004	0.61	2004	2477	12862	6 135 529		0.992
2006	300	0.72	0.005	0.62	2005	2286	2597	5 225 796		0.993
2007	339	0.71	0.008	0.60	2006	2508	2504	6 290 064		0.992
2008	339	0.73	0.008	0.64	2007	2613	8210	6 827 769		0.990
2009	353	0.71	0.003	0.59	2008	2303	2316	5 303 809		0.991
2010	231	0.71	0.013	0.61	2009	1494	1668	2 232 036		0.994
2011	267	0.71	0.002	0.60	2010	1649	4634	2 719 201		0.992
2012	295	0.72	0.005	0.60	2011	1773	11040	3 143 529		0.990
2013	347	0.69	0.018	0.55	2012	1855	3933	3 441 025		0.991
Mean	362	0.71	0.007	0.59		2288	6178	5 445 908		0.992

Table 4.1: Summary of data in the temporal network autocorrelation model. The left side of the table shows the dependent variable: the incidence of homicide arrests in one-year periods (t) and summarizes the proportion of individuals arrested for homicide who were: arrested; arrested for homicide; and co-arrested with at least one other person in the previous year ($t-1$). The right side of the table summarises the right-hand side of the TNAM, including the number of nodes and edges in the stratified sampled co-arrest network at $t-1$ and the proportion of all arrests included in the stratified sample.

as the co-arrest network developed over time it began to exhibit the small-world features of short average path length and local clustering.

Figure 4.4 shows the demographic parameters of the year-by-year temporal autocorrelation models for homicide arrests. In these models, the dependent variable is an arrest for homicide, and the predictors include individual covariates as well as lagged network terms, the latter capturing the network structure among the nodes at $t-1$. For each parameter, we plot the point estimate and 95 percent confidence interval by year. Importantly, the parameter estimates and confidence intervals reflect the odds of an arrest for homicide relative to an arrest for a non-homicide crime.

The odds of arrest for homicide, relative to arrest for a different crime, were approximately two times greater for males than females. These odds were relatively stable over time, with the odds ratio exceeding three in 2006 and 2009. Age was associated with decreasing odds of homicide arrest. Gang membership appeared to have little influence on the odds of arrest for homicide—the parameter estimate is not significant in 10 of the 13 years. As noted in section 4.3, the data on gang membership is somewhat sparse, particularly in the earlier years under study. Caution is therefore advised in interpreting the odds associated with gang membership. Notably, gang membership significantly and substantively increased the odds of homicide arrest in 2012 and 2013, the last two years under study. This may be attributable to increasing accuracy in the identification of gang members, a shift in policing behavior, or a shift in the homicide activity of gangs, such as inter-gang conflicts. Although black individuals had an odds-ratio exceeding 1 in all but one year, 2012, only one odds estimate is significant—black males were approximately 1.5 times more likely than non-blacks to be arrested for homicide as compared to other crimes in 2010. Previous research notes that black males are at particularly high risk of arrest (LaFree and Drass, 1996). The model presented here suggests that, conditional upon being arrested, the difference between black and non-black individuals in the odds of that arrest pertaining to homicide is minimal.

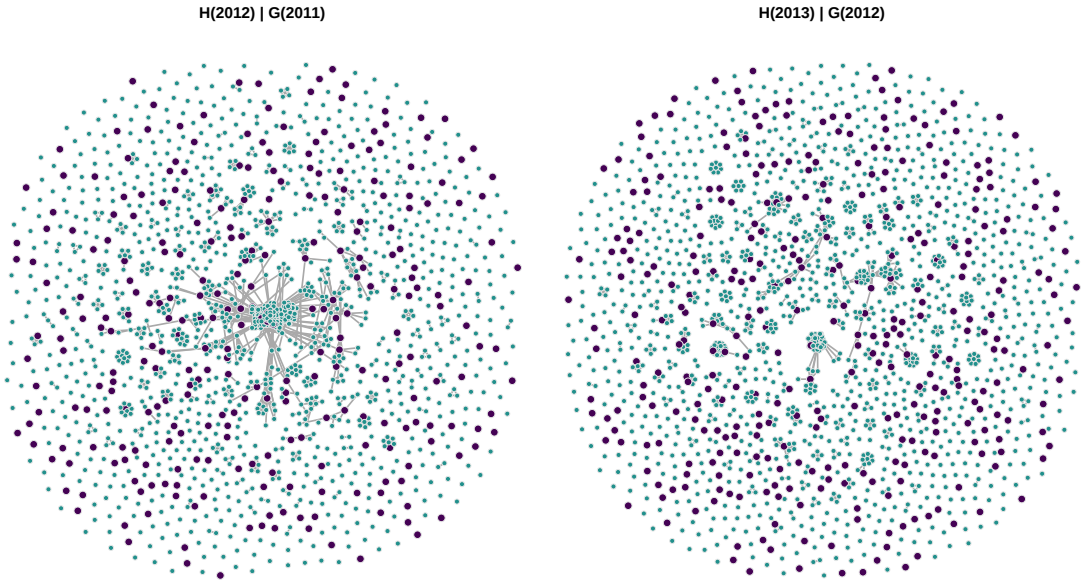


Figure 4.3: Co-arrest network following endogenous stratified sampling in $t = 2011$ (left) and $t = 2012$ (right). Nodes arrested for homicidein at $t + 1$ are highlighted in purple. Edges represent co-arrest relationships.

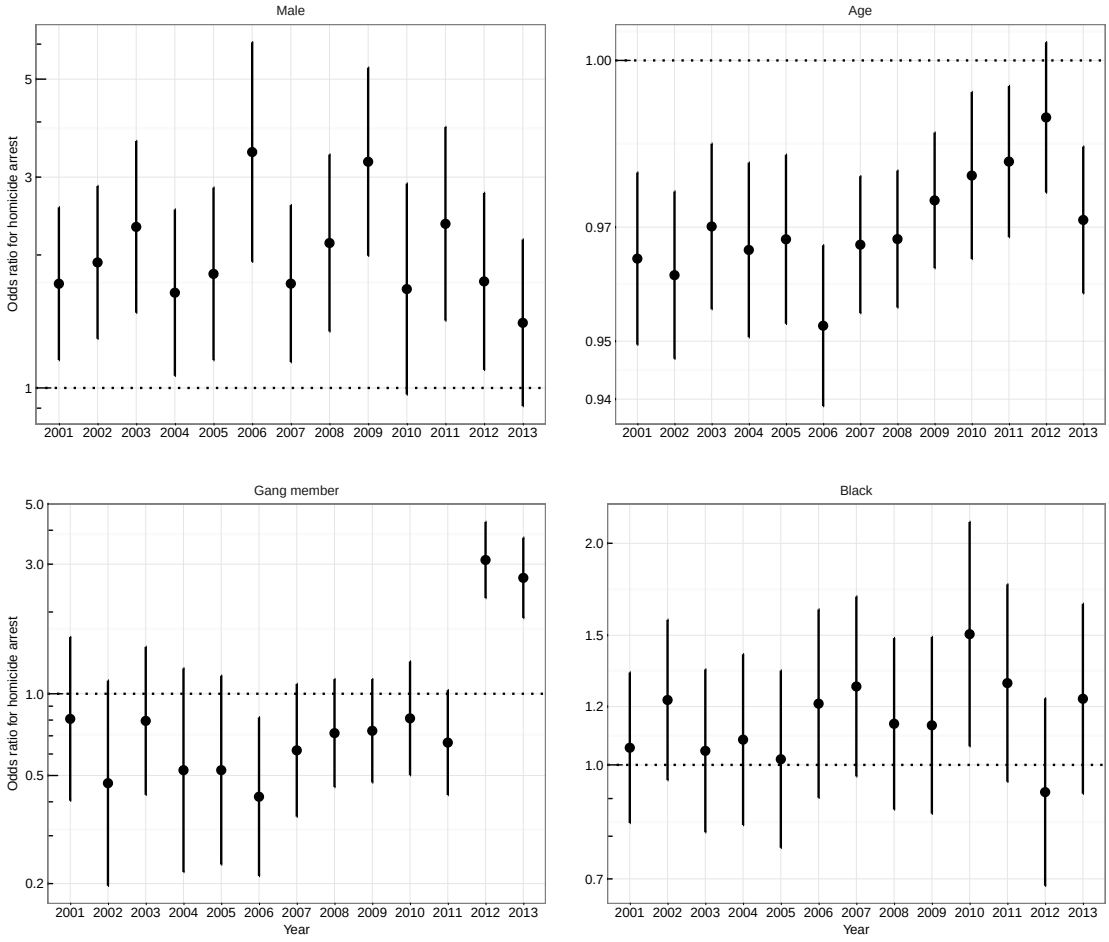


Figure 4.4: Estimates for the individual-level parameters in the temporal network autocorrelation models by year, with 95 percent confidence intervals. The parameters represent the odds-ratio of sustaining a homicide arrest versus a non-homicide arrest at t given the co-arrest network at $t-1$.

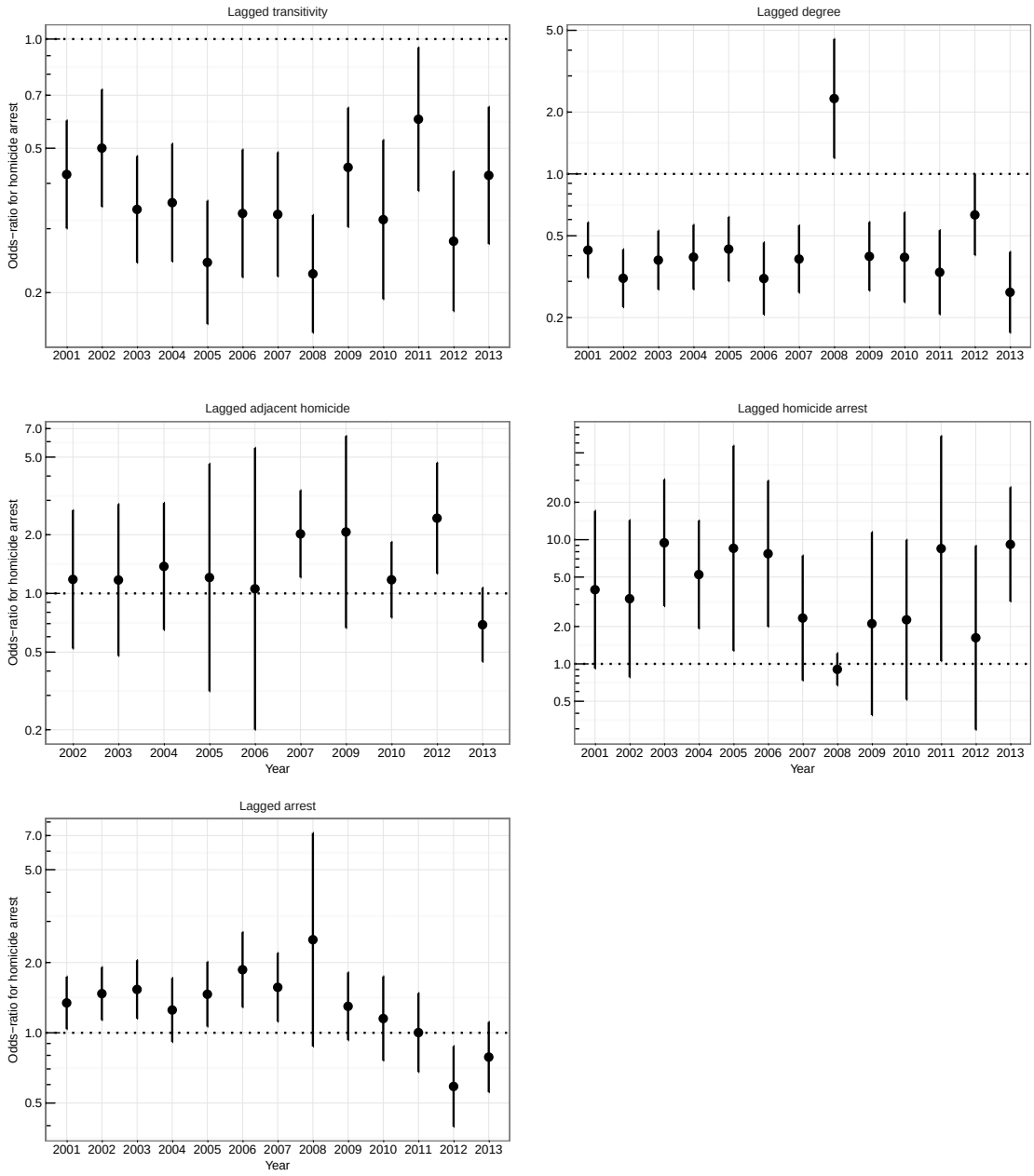


Figure 4.5: Estimates for the network terms in the temporal network autocorrelation models by year, with 95 percent confidence intervals. The parameters are the odds-ratio of homicide arrests versus non-homicide arrests at t. Lagged transitivity relates to individuals embeddedness in the transitive co-arrest triads at $t-1$; lagged degree indicates that the node had at least one co-arrestee at $t-1$; lagged adjacent homicide indicates the node was connected to a person arrested for homicide at $t-1$; lagged homicide arrest indicates the node was arrested for homicide at $t-1$; and lagged arrest accounts for whether the node was arrests for a non-homicide crime at $t-1$.

Figure 4.5 presents the estimated lagged network parameters, again presented by year with 95 percent confidence intervals. These parameters capture the effects of prior network embeddedness on present homicide arrests. A prior isolated arrest for a non-homicide crime at $t-1$ increased the odds of homicide arrest in the following year throughout the period 2001–2007. From 2007 onwards, the effect of a prior isolated non-homicide arrest was not significant, and decreased the odds of homicide arrest in 2012. Transitivity is negatively associated with the odds of homicide arrest. That is, individuals embedded in complete triads at $t-1$ were less likely to be arrested for homicide at t than for other crimes. The odds of homicide arrest were lower for individuals with a co-arrestee at $t-1$ than for individuals arrested alone at $t-1$, with the exception of 2008, where the effect was reversed. Thus, homicide arrests could be differentiated from other types of arrest along two endogenous network dimensions. First, homicide arrests are negatively dependent on lagged co-arrest transitivity and popularity network effects—homicide arrests were less likely among nodes embedded in dyadic and triadic network structures in the previous year. Second, prior non-homicide arrests increased the odds of homicide arrest disproportionately compared to non-homicide crimes throughout 2001–2007. Nodes that were members of the lagged network as isolates, rather than nodes in dyadic or triadic network structures, therefore had the greatest odds of homicide arrest.

The remaining two parameters capture the dependence of a homicide arrest at t on a prior homicide arrest at $t-1$ and on a prior homicide arrest among an adjacent node at $t-1$. The odds of a homicide arrest were approximately three times greater, on average, for nodes with a prior homicide arrest, compared to the odds of a non-homicide arrest. However, nodes connected to an individual arrested for homicide at $t-1$ were not significantly more or less likely to be arrested for homicide at time t with the exception of 2007 and 2012, where an alter homicide arrest increased the odds of homicide arrest twofold. Thus, a connection to a node arrested for homicide, whether through a co-offending tie for homicide or for a different crime, had a spill-over effect on the odds of arrest vis-a-vis other crimes in 2007 and 2012.

4.5.2 Multiplex co-arrest interdependency

The next set of models estimate the interdependence between homicide arrests and six other violence-related arrests: simple assault, aggravated assault, simple battery, aggravated battery, robbery, and weapons violations. Each violent crime is modeled separately, i.e. the first model captures the interdependence between homicide and aggravated assault, for example. Additionally, I model the interdependence between homicide and a non-violent crime, miscellaneous drug violations. As noted in section 4.4, I refer to these non-homicide co-arrest networks as "layer networks." In these models, the dependent variable is the homicide network over the period 2000–2013. The right-hand side parameters include individual-level covariates—capturing the variation in homicide co-arrest attributable to age and gender homophily, for example—and endogenous network terms—capturing the influence of structural features within the homicide network on homicide co-arrest. Most importantly, the right-hand side includes additional parameters measuring the influence of the co-arrest structure in the layer networks on co-arrest for homicide. In essence, this group of parameters measure the extent of co-arrest multiplexity, i.e. the interdependence between co-arrests for homicide, on the one hand, and co-arrests for another crime, on the other.

Recall that the interdependence models are pooled over time and therefore I cannot estimate the causal influence of co-arrests for non-homicide crimes on co-arrest for homicide, as in the above temporal autocorrelation models. Specifically, the direction of dependence is not modeled, which means that homicide co-arrests could influence non-homicide co-arrests or vice versa, or influence could operate in both directions. I therefore caution against interpreting the estimates as causal effects, but instead as the interdependence of homicide and non-homicide arrests.

Figure 4.6 (H) shows the homicide co-arrest network during 2000–2013. Referring to Table 4.2, the homicide network includes 1546 edges among 5084 nodes. Young black males are disproportionately represented in the homicide network relative to city demographics and around 8 percent of the nodes are gang members.

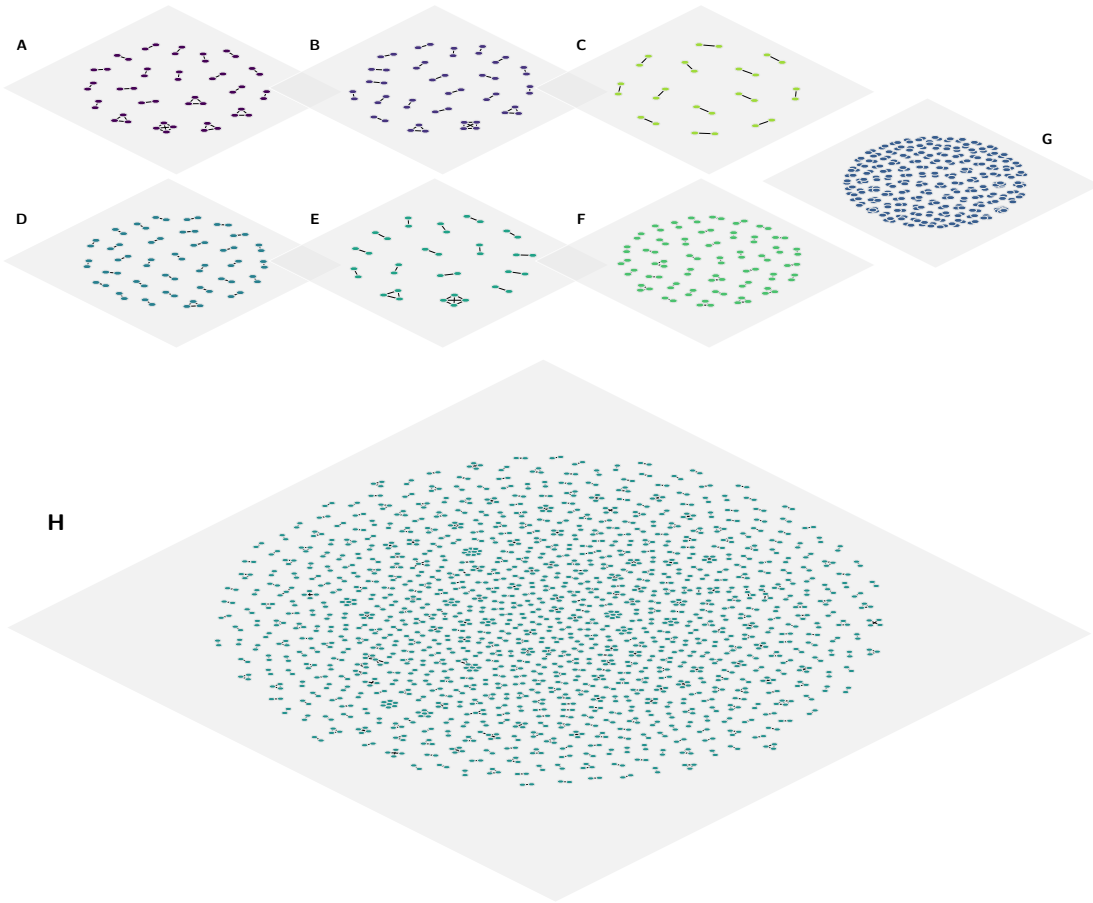


Figure 4.6: Multiplex edges in the Chicago co-arrest network, 2000–2013. (A) Aggravated assault; (B) aggravated battery; (C) weapons violations; (D) robbery; (E) simple assault; (F) simple battery; (G) drugs offenses; and (H) homicide co-arrests. All co-arrest edges in (A–G) are also present in the homicide co-arrest network (H). The plots exclude isolated nodes, but these are included in the multiplex interdependence models.

Figure 4.6 (A–G) show the nodes with co-occurring edges in the homicide network and the given layer network. A co-occurring edge denotes that the same two individuals were co-arrested together for homicide and at least one other crime. As listed in Table 4.2, such edges are relatively uncommon. This is likely influenced by incarceration; in the temporal models, I conditioned on arrests at time t, ensuring that all individuals were at-risk of arrest. The interdependence models do not use this conditioning criteria, thus individuals could have been incarcerated for homicide and are therefore unable to be re-arrested at a later time for a different crime. The remainder of the multiplex nodes, i.e. individuals arrested for homicide and a non-homicide crime, without co-occurring edges are not shown, but represent the majority of nodes in the layer networks. The overall count of multiplex nodes and multiplex edges differs substantially across layers. Multiplex homicide–drugs nodes are commonplace. In fact, approximately 60 percent of individuals arrested for homicide also have an arrest for a drugs violation during 2000–2013. In contrast, the proportion of multiplex nodes is 12 percent for homicide–simple assault, 14 percent for homicide–aggravated assault and homicide–aggravated battery, 15 percent for homicide–robbery, 28 percent for homicide–weapons, and 39 percent for homicide–simple battery.

In terms of covariates, gang members are overrepresented among multiplex nodes, indicating that gang members are particularly likely to be arrested for multiple crimes. This is particularly the case for weapons violations and robbery arrests. Multiplex nodes are younger and more frequently male than individuals with homicide-only arrests. The distribution by race varies by layer. Black individuals are more frequently arrested for homicide and weapons, simple battery, and drugs. However, non-black individuals are more frequently arrested for homicide and aggravated assault and robbery. With the exception of robbery and weapons, these differences are relatively minor at around one or two percentage points in each direction.

Fig	Layer	Nodes	Edges	Age	Black	Gang	Male
	Left-hand side						
H	Homicide	5084	1546	29	0.78	0.08	0.94
	Right-hand side						
	Co-occurring network:						
A	Aggravated assault	724	25	25	0.72	0.17	0.97
B	Aggravated battery	700	28	26	0.78	0.16	0.97
C	Weapons	1404	13	24	0.84	0.18	0.97
D	Robbery	757	28	27	0.70	0.18	0.94
E	Simple assault	618	17	28	0.80	0.12	0.95
F	Simple battery	1963	34	26	0.79	0.16	0.98
G	Drugs	3069	90	28	0.79	0.11	0.96
	Co-occurring plus neighborhood network:						
A	Aggravated assault	1071	704	25	0.65	0.16	0.97
B	Aggravated battery	961	497	26	0.74	0.17	0.95
C	Weapons	1693	419	26	0.78	0.15	0.97
D	Robbery	1431	1104	23	0.82	0.17	0.97
E	Simple assault	919	554	25	0.65	0.20	0.95
F	Simple battery	3134	2399	25	0.79	0.12	0.93
G	Drugs	6280	9881	28	0.80	0.10	0.94

Table 4.2: Summary of the layer co-arrest networks in the multiplex interdependence models. The left-hand side of each model is the homicide co-arrest network. Network terms derived from the layer networks are used as exogenous variables in the right-hand side. The co-occurring plus neighborhood network is used to calculate shared partner statistics, but is not used in the final models. The table reports the mean age, proportion black, proportion gang members, and proportion male in each layer co-arrest network.

Endogenous network effects for homicide co-arrest

In contrast to the temporal dependence networks, the interdependence models are restricted to nodes arrested for homicide who were also arrested for a non-homicide crime. However, as mentioned above, the interdependence models include parameters capturing the network structure of the layer network. Two of these terms, the geometrically weighted edgewise shared partner (GWESP) and geometrically weighted dyadwise shared partner (GWDSP) measure the number of "third-party" nodes with whom a pair of nodes have been co-arrested. These third-party nodes are not necessarily multiplex nodes themselves. That is, they may not have been arrested for homicide and therefore may only appear in a layer network. To account for the possibility that these nodes nevertheless exert influence on the odds of co-arrest for homicide, the GWESP and GWDSP statistics are derived from "co-occurrence plus neighbors" layer networks, which include third-party nodes even if they have not been arrested for homicide. Such nodes are excluded from the other right-hand side parameters.

Figure 4.7 shows the parameter estimates of the multiplex interdependence model for simple assault, aggravated assault, simple battery, and aggravated battery. Figure 4.8 shows the equivalent estimates for drugs, robbery, and weapons arrest multiplexity.[6] The log-odds relate to co-arrests for homicide. Thus, unlike the network autocorrelation model, the interdependence model parameters are dyadic and refer to the odds of co-arrest formation. Terms in the upper part of the plot relate to endogenous statistics within the homicide co-arrest network, while the terms in the lower part of the plot relate to the multiplex interdependence between homicide co-arrests and non-homicide layer arrests.

I first report the results for the endogenous parameters within the homicide co-arrest network. To control for the overall propensity for homicide arrests without co-arrestees and the homogeneous homicide co-arrest probability, isolate and edge terms were included in the model (not shown). Across all seven models, the isolates parameter (log-odds: $[-11.7, -11.1]$) and edge parameter are negative (log-odds:

[6]The models are equivalent and are presented separately to ease interpretation.

$[-1.7, -1.6]$), reflecting the low density of the homicide co-arrest network and the large fraction of isolates.

The homicide network exhibits a strong GWESP effect. That is, there is a clear tendency towards local clustering into triadic structures. One way to interpret this parameter is to hypothesize that, for a pair of individuals co-arrested together for homicide, the chance of that pair also having a homicide co-arrest tie to a third individual—or fourth, or fifth, and so on—is substantially greater than would be the case by chance. Conversely, the log-odds for GWDSP are negative. Thus, individuals that have not been co-arrested together are less likely than chance to share co-arrests with the same individuals. This may be explained by the absence of large connected components in the homicide network, which means that, in the vast majority of cases, unconnected pairs of nodes have not been arrested with the same third-party. The GWESP and GWDSP estimates are consistent across all seven models (including those in Figure 4.8), indicating that the influence of the multiplex parameters on the endogenous structural tendencies within the homicide network is trivial.

Figure 4.7 and Figure 4.8 show that the propensity for individuals to be co-arrested varies across covariates. While males represent the overwhelming majority in the homicide network, as shown in Table 4.2, males had marginally lower odds of being arrested alongside another individual. Thus, conditional upon being arrested for homicide in the first place, women were more likely to be arrested in pairs, although the difference is small in substantive terms. In fact, when the network parameters for aggravated battery, simple assault, and robbery are included in the model, the male term is not significant. Conversely, gang members are more likely to be arrested in pairs or higher-order groups than non-gang members. This is an intuitive result; in homicide cases which are thought to be gang motivated or gang-involved, it may be the case that multiple members of that gang had suspected involvement. Moreover, it is plausible that gang-involved homicides involve more individuals, on average. Overall, black individuals are less likely to be arrested in pairs than non-black individuals, while older people are marginally

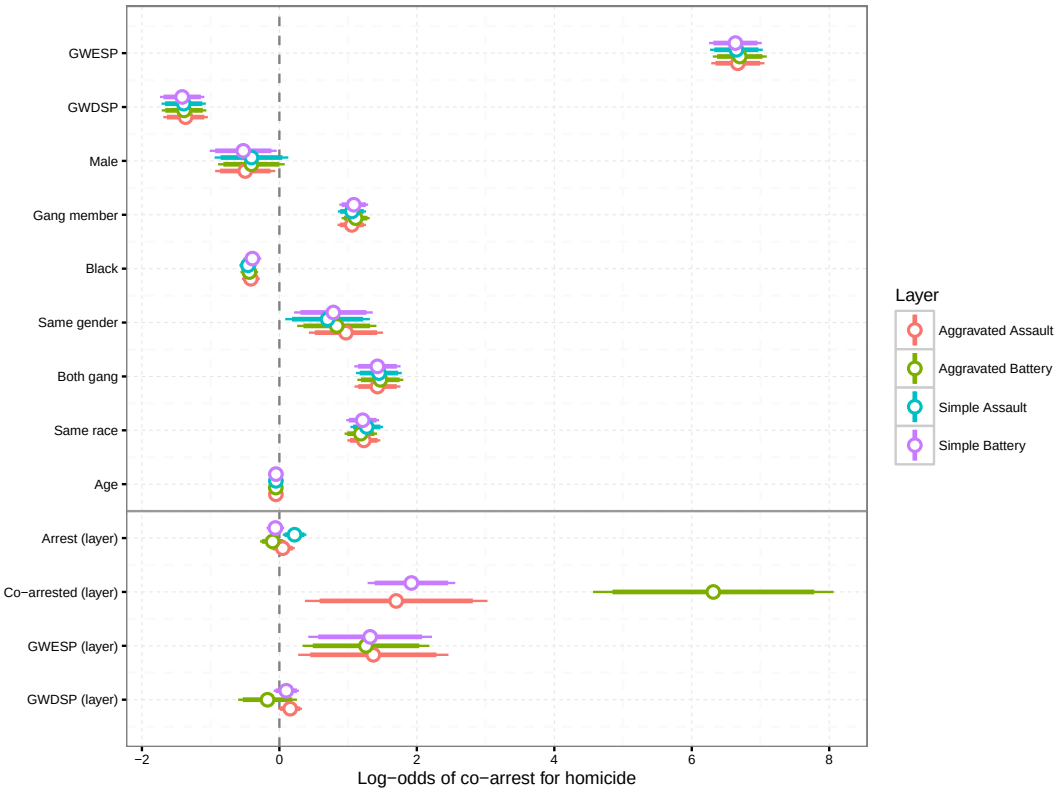


Figure 4.7: Parameter estimates for the multiplex interdependence models with 95 percent confidence intervals. The top part of the plot shows log-odds estimates for the endogenous parameters in the homicide co-arrest network. The lower part shows log-odds estimates for the exogenous parameters, which capture the dependence of homicide co-arrests on network statistics in the right-hand side co-arrest networks designated by the legend.

less likely to be co-arrested together. The gender, gang, race and age estimates are stable across all seven models.

Three homophily parameters are included in the models to assess the tendency for nodes with similar covariates to be co-arrested. There are positive homophily effects for gender, gang membership, and race. Males are more likely to be arrested for homicide with other males, and women with other women, after controlling for the overall propensity of males and females to be arrested for homicide. The race homophily parameter is marginally larger, indicating that ethnic-mixing in homicide co-arrests is substantially less likely than ethnic homogeneity. It is important to note here that race is categorized into just two groups, black and non-black. Thus, the model does not allow me to capture the full influence of different racial group identity on the likelihood of homicide co-arrest. Gang membership is the largest source of co-arrest homophily. The odds of a co-arrest between two gang members, or two non-gang members, is approximately 5.7 times larger than a gang member with non-gang member co-arrest for homicide. Gang members are therefore more likely to be co-arrested, in general, and considerably more likely to be co-arrested with other gang members.

Multiplex interdependence of homicide co-arrest

To estimate the multiplex interdependence of homicide co-arrests and the non-homicide layer co-arrests, I included four terms in the interdependence models. The parameter estimates for these terms are shown in the bottom part of Figure 4.7 and Figure 4.8. The first term (arrest) captures the influence of an isolate arrest in a particular layer network on the odds of co-arrest for homicide. That is, the term measures the interdependence of co-arrests for homicide and arrests—rather than non-arrests—for non-homicide crimes, such as aggravated assault. The arrest term is not significant for aggravated assault, aggravated battery, simple battery, and weapons violations multiplexity, indicating that there is no interdependence between arrests for these crimes and homicide co-arrests. However, the odds of homicide co-arrest are positively associated with arrests for simple assault, drug

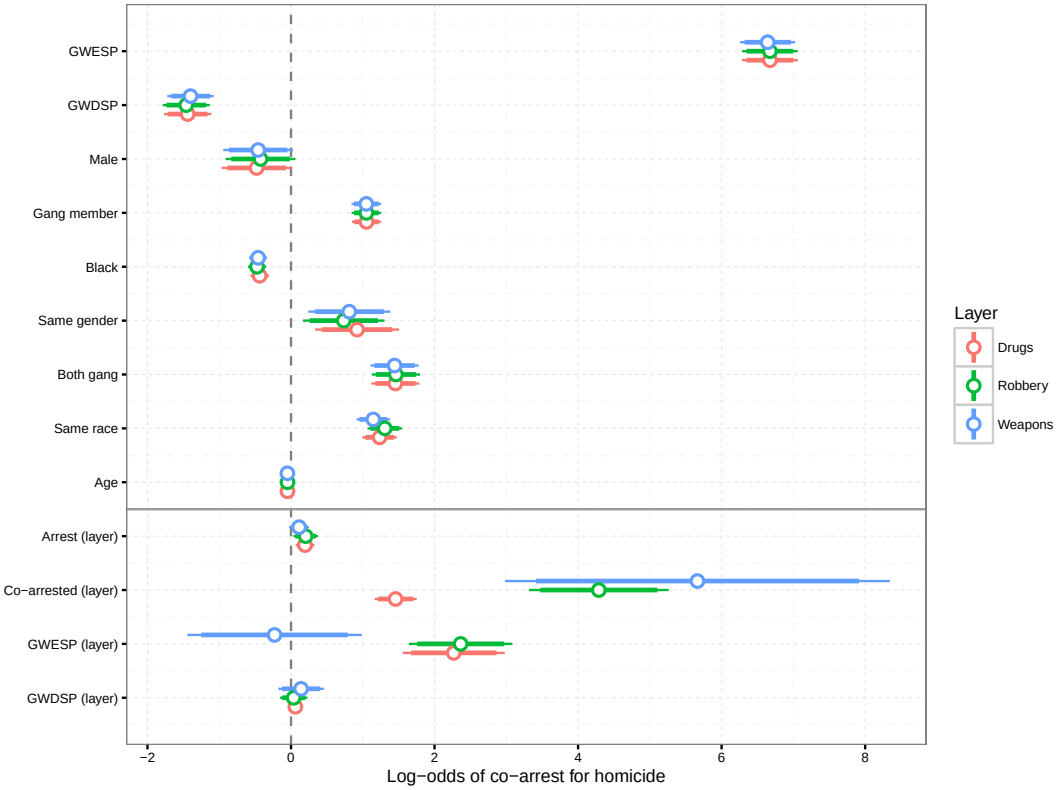


Figure 4.8: Parameter estimates for the multiplex interdependence models with 95 percent confidence intervals. The top part of the plot shows log-odds estimates for the endogenous parameters in the homicide co-arrest network. The lower part shows log-odds estimates for the exogenous statistics in the right-hand side co-arrest networks designated by the legend.

offenses, and robbery. In substantive terms, this association is marginal. Thus, there is little interdependence between homicide co-arrests and isolated arrests in the non-homicide layer networks.

The second term (co-arrested) estimates the dependence between homicide co-arrests and co-arrests in the other layer networks.[7] The co-arrest term produced large positive interdependency estimates. This means that the structure of homicide co-arrests depends on the structure of co-arrests for non-homicide crimes. Thus, in addition to the temporal dependence between co-arrests for homicide in the present on previous homicide co-arrests reported in section 4.5.1, I also find evidence of dependency between homicide co-arrests and co-arrests for aggravated assault, simple battery, aggravated battery, drug offenses, robbery and weapons violations. This dependency is particularly strong in the case of robbery, weapons, and aggravated battery multiplexity.

The third and fourth terms (GWESP and GWDSP) measure higher-order dependency between homicide co-arrests and the presence of shared partners in the layer networks. The significant GWESP parameters reveal interdependency between homicide co-arrests and the frequency of edgewise shared partners in the aggravated assault, aggravated battery, simple battery, drugs and robbery layer networks. Thus, I find that shared mutual co-arrestees in the layer networks—the presence of individuals with whom a pair of nodes have both been co-arrested for a non-homicide crime—is associated with greater odds of that pair sustaining a co-arrest for homicide. That is, the probability that a pair of individuals are co-arrested for homicide is greater if those individual have been arrested with the same third-party for a non-homicide crime. The odds of homicide co-arrest are 3.5, 3.7, and 3.9 times greater for co-arrested individuals with shared co-arrestees in the aggravated battery, simple battery, and aggravated assault networks, respectively. The odds are 9.7 and 10.6 times greater in the case of drugs and robbery arrests, respectively. However, the GWESP term is not significant in the case of weapons multiplexity, indicating that homicide co-arrests are not associated with the presence

[7] In the case of simple assault multiplexity, the co-arrest, GWESP, and GWDSP terms are excluded due to separability.

of shared co-arrestees for weapons violations. With this exception, the positive parameter estimates for the GWESP term, taken together with the positive co-arrest terms, indicate that the structure of arrests for non-homicide crimes plays a significant role in the development of homicide co-arrests, and vice versa.

In contrast to the GWESP term, the GWDSP term is non-significant in all of the layer networks. Pairs of individuals that have not been co-arrested for a non-homicide crime are therefore no more likely to be co-arrested for homicide if they share mutual co-arrestees in a layer network. Thus, local clustering of co-arrests around pairs of individuals whom have not themselves been co-arrested is not associated with the odds of homicide co-arrest.

4.6 Discussion

These results demonstrate how repeated arrests and co-arrests develop into a social structure that is significantly associated with the incidence of future arrests for homicide and non-homicide crimes. In particular, the results show that the social structure of arrests for homicide and the social structure for several non-homicide crimes are interdependent. Analysis based on the individual-level incidence of arrest ignores this structure, and therefore overlooks how patterns of arrest are associated with embeddedness in co-arrest networks. By focusing on dyadic and higher-order arrest structures, this paper quantifies the multiplex interdependence of homicide arrests and provides estimates of the extent to which homicide arrest depends on the structure of previous co-arrests.

The temporal network autocorrelation models reveal that embeddedness in the co-arrest network is negatively associated with the odds of homicide arrest in the following year, relative to arrests for non-homicide crimes. Individuals arrested alone for a non-homicide crime in the previous year, whom are therefore isolates in the co-arrest network, are less likely to be arrested for homicide than for a non-homicide crime. However, individuals arrested as isolates for homicide in the previous year are more likely to be re-arrested for homicide than for a different crime. Those arrested with at least one other person are less likely to sustain a homicide

re-arrest, and individuals embedded in transitive co-arrest triads are more likely to be re-arrested for non-homicide crimes. Furthermore, an individual connected to a person that sustained a homicide arrest is no more or less likely to be arrested for homicide themselves, controlling for prior homicide arrests on the individual.

To what extent is prior co-arrest network embeddedness influencing future arrests? Homicide arrests could be associated with the prior co-arrest network in a number of ways. For example, the fact that homicide arrests are less likely to occur for individuals in prior transitive co-arrest triads, reflected in the negative transitivity estimate, could be due to trends in how many people are typically involved in arrest events for different types of crime. If arrests for non-violent crimes frequently involve two or more individuals, and re-arrests for such crimes are common within a 12-month time frame, one consequence would be that individuals involved in non-violent criminal behavior form into transitive co-arrest triads more often than is the case for violent crimes. The finding that such individuals are less likely to be re-arrested for homicide could therefore owe to the consistency in criminal behavior; these individuals were not suspected of violent crimes in the previous year, and they are not suspected of violent crimes in the present year. Conversely, the fact that individuals embedded in transitive co-arrest triads are more likely to be re-arrested for non-homicide crimes could owe to the "order-maintenance" policing strategy discussed in section 4.2, in which some individuals suspected to be involved in minor crimes face repeated police contact, but are often not charged. The transitivity estimate in the temporal network autocorrelation models may thus reflect a strategy in which targeted individuals are recurrently arrested, sometimes in groups, on suspicion of minor crimes. One could argue that individuals are both more likely to form into transitive triads and more likely to sustain arrests for non-homicide crimes again in the following year simply by virtue of interventionist policing strategy, and that network embeddedness is simply a by-product.

When interpreted in full, however, the results suggest that network embeddedness is important for understanding the likelihood of future homicide arrests. Specifically, the lagged arrest parameter is positively associated with the odds of homicide

arrest vis-a-vis other crimes. Taking previous arrests alone, and ignoring the co-arrest structure, would therefore lead to the conclusion that arrests increase the probability of homicide re-arrest relative to non-homicide re-arrest. However, taking the co-arrest structure into account through the negative lagged transitivity and negative lagged degree parameter reveals a more complex picture in which previous isolated arrests generally increase the probability of homicide re-arrests, but previous co-arrests and membership of triadic co-arrest structures decreases the probability of homicide re-arrest vis-a-vis re-arrests for other crimes. This highlights the importance of analyzing not only the influence of prior arrests and offending on re-arrest and re-offending, but also the influence of prior co-arrest embeddedness.

The multiplex interdependence models indicate that co-arrests for homicide and co-arrests for other violent crimes are interdependent. Moreover, the models also indicate interdependence between co-arrests for homicide and drugs offenses, and homicide and weapons violations. That is, the odds that a pair of individuals are co-arrested for homicide is positively associated with the odds of co-arrest for non-homicide crimes. The odds of homicide co-arrest are also positively associated with the presence of shared partners, with the exception of weapons violations. In contrast, the odds of homicide co-arrest is only marginally associated with the odds of an isolated arrest for simple assault, drugs offense, and robbery, and the association is not statistically significant in the case of aggravated assault, aggravated battery, simple battery, and weapons violations. Thus, embeddedness in co-arrest network structures again emerges as an important aspect of homicide arrests. Whereas an individual sustaining an isolated arrest for aggravated assault, for example, is no more or less likely to be co-arrested with another individual for homicide, a pair whom are co-arrested for aggravated assault are around six times more likely to sustain (or to have sustained) a co-arrest for homicide. Similarly, a co-arrest for homicide is nearly four times more likely for a co-arrested pair whom have both been co-arrested for aggravated assault with the same third-party actor.

The multiplex interdependence of arrests speaks directly to the role of “dynamic risk factors” and the influence of “companions” in criminal behavior raised by

Gendreau, Little, and Goggin (1996) two decades ago. The occurrence of homicide co-arrests does not simply depend on the prior criminal behavior of each individual involved in that co-arrest, but on their co-behavior. From this perspective, a substantial fraction of individuals in the Chicago co-arrest network are embedded in pairs and groups, and these pairs and groups are recurrently arrested. While previous research has justifiably highlighted the role of prior arrests and criminal history on re-offending, recidivism, and re-arrest, the nesting of individuals within pairs and groups that are arrested collectively provides a new angle on the risks of repeated contact with law enforcement agents. Even though isolated arrests occur more commonly than co-arrests, the results here suggest that individuals' embeddedness in the social structure of arrests is an important risk factor in re-arrest.

4.7 Conclusion

The role of social ties in criminal behavior is a longstanding area of criminology research. This paper investigates social ties in the context of co-arrests, in which individuals are nested in arrest events due to being arrested together for the same crime. The findings suggest that the social structure revealed by the amalgamation of co-arrest events over time influences the likelihood of re-arrest for homicide vis-a-vis other crimes, and that the structure itself is a pertinent risk factor in co-arrests for homicide. These ties of arrest are not just artifacts of individuals' contact with police, they also reflect underlying social relationships that may persist following arrest, influencing future criminal behavior. The structure of co-arrest may also inform suspect profiling, contributing to the future risk of arrest by providing a searchable record of prior police-community contact. While the results relate to Chicago and I focus on homicide, I believe the findings are likely to generalize to other cities and types of crime.

Drawing on social network methods, I examined how embeddedness in the previous years' co-arrest network influenced the risk of homicide re-arrest vis-a-vis re-arrest for other crimes. I found that homicide re-arrests are more likely than non-homicide re-arrests for individuals who were arrested in the previous

year, particularly if the prior year arrest pertained to homicide. However, greater embeddedness in the co-arrest network decreased the risk of homicide re-arrest vis-a-vis other crimes. Individuals embedded in co-arrest relationships and transitive co-arrest triads were more likely to be re-arrested for non-homicide crimes. These results indicate that, in addition to prior arrest records and criminal history, the social structure of prior arrests is important for assessing the risk of future arrests for different types of crime. One possible explanation for this phenomenon is that, in capturing underlying, pertinent social relationships, co-arrest ties account for part of the opportunity structure for future criminal behavior.

I then investigated the interdependency between homicide co-arrests and arrests for other crimes by modeling arrest ties at the dyadic-level. I found that pairs of individuals' sustaining homicide co-arrests are considerably more likely to also sustain (or have sustained) a co-arrest for other violent crimes, drug offenses, and weapons violations, than would be expected by chance. There is evidence that higher-order social structures for non-homicide arrests, such as the presence of mutual co-arrestees, are also associated with an increased likelihood of homicide co-arrest. This multiplex interdependency indicates that the dyadic structure for assault, battery, and other violent crimes influences the structure of arrests for homicide, beyond the individual-level effects of prior criminal history. Further investigation is needed to shed light on the influence of arrest multiplexity on the social structure of arrests for minor crimes; whereas long periods of incarceration following homicide prosecutions may limit the incidence of further multiplex co-arrests, multiplexity may be more common in the case of minor crime arrests.

This study is the first investigation of how re-arrests are influenced by the social structure of prior arrests. Although the results relate strictly to arrests, which are distinct in that crimes have not necessarily taken place, it is likely that the findings have considerable overlap with re-offending and recidivism. The results suggest that the social relationships underlying criminal behavior are indeed an important aspect of prolonged and recurrent criminal behavior, as has long been theorized in the criminology literature, and these relationships are in some cases subject

to recurrent police interventions. However, the study is not without limitations. Arrests are only one type of contact with the criminal justice system, and are in many cases a precursor to incarceration. Long periods of incarceration effectively censor individuals from further arrests and therefore, to understand the full picture of how the social structure of arrests influences recurrent police contact, periods of incarceration should be accounted for. In particular, further research is required on the risks of re-arrest following prisoner re-entry, where the social structure of prior arrests may be particularly consequential.

To further investigate the extent to which the probability of recurrent arrest depends upon individuals' embeddedness within co-arrest structures, as well as the characteristics of the individuals with whom a person has been co-arrested, this paper can be extended in two directions. First, the analysis could focus on arrests for minor crimes, such as petty theft, vandalism, or trespassing, that are unlikely to result in incarceration, or which carry only short-term prison sentences. This strategy would substantially mitigate the censoring problem by reducing the possibility that a person is ineligible for re-arrest due to incarceration but the ineligiblity is unobserved. Empirically, the probability of re-arrest for minor crimes, or for all crimes, could subsequentally be evaluated at year intervals, as in the TNAM model. Alternatively, the probability could be evaluated using survival models, where instead of modeling the incidence of arrest within a specified year period, the time-to-arrest is investigated directly. One challenge in this approach would be specifying the network statistics in the survival model framework.

Second, data on incarceration could be incorporated into the analysis. While these data are routinely collected, the difficulty lies in accessing the data and linking individuals' arrest and incarceration records. However, including incarceration data would remove the unobserved censoring problem and enable the direct analysis of prisoner re-entry and recidivism.

Notwithstanding the above limitations, the study provides a new angle on the multiplexity of homicide co-arrest relationships and the network structure of arrests. The results emphasize that many arrests are not isolated contacts between an

individual and police, but are instead part of, and contribute to, a wider structure of arrest ties. The approach used in this paper to uncover the social structure of arrests by projecting bipartite arrest networks provides a means of examining the social dynamics behind repeated contact with agents of law enforcement.

5 Conclusion

Contents

5.1 Summary and implications 123
5.2 Challenges in studying criminal networks 130
5.2.1 Missing data and non-random sampling 130
5.2.2 The selection and application of models to network data 133
5.2.3 Integrating network data, outcome variables, and contextual data 137
5.3 Directions for future research 141

5.1 Summary and implications

For this thesis, I presented three studies on criminal networks using two sets of relational data. In the first study, I collected archival data on collaboration in a heroin trafficking organization operating in Europe and the United States. The study had two aims: first to provide a systematic analysis of the structure of collaboration among the traffickers, and second to investigate the vulnerability of that structure to the removal of nodes. In line with empirical results relating to non-criminal collaboration networks, the trafficking network had a heavy-tailed degree distribution, that I characterized as exponentially distributed after statistical testing, and strong local clustering. While the study did not set out to investigate the

view that organized criminal networks should be decentralized and non-hierarchical, known as the "flat thesis," my findings offer no support to such claims. Instead, centralization emerges as a key organizing principle of the heroin trafficking network.

Despite the resilient finding in the networks literature that collaboration networks exhibit degree assortativity, I found that the trafficking network exhibited dissortative degree mixing, where high degree nodes—whom have more collaborators than average—are more likely to connect to low degree nodes—whom have relatively few collaborators—than would be expected by chance. This contrasts with collaboration between non-criminal actors, where collaboration is routinely found to be assortative.

The dissortativity of collaboration in the trafficking organization had profound consequences for its vulnerability to targeted node removal. I demonstrated this by simulating modified versions of the trafficking network using a graph sampling algorithm. The algorithm toggled some collaborative relationships "off" and switched others "on," thus reconfiguring the network into more assortative and more dissortative states while preserving the degree distribution. I then evaluated the effects of degree-targeted node removal on the number of components, the size of the largest component, and the number of walktrap-communities. The size of the largest connected component reduced sharply and the network separated into a greater number of components and communities when the network was configured into more dissortative states, while the opposite effect occurred in more assortative states. To understand whether these results generalize to organized criminal networks, further studies should evaluate degree assortativity as a matter of course.

The extent to which criminal networks are structured to reduce vulnerability has been debated in the literature on organized crime. Viewed from a perspective of targeted node removal, the drug trafficking network had no specifically resilient properties. This finding is consistent with the view that the structure of organized criminal networks is not driven by concerns over security and secrecy (Varese, 2013; Morselli, Giguére, and Katia, 2007). Instead, the structure appears largely driven by the tasks at hand. For instance, highly central, positionally-important traffickers collaborated directly with individuals on the periphery. While remaining covert

may have been a concern for these traffickers, they nevertheless appeared willing to increase the risk of exposure by reaching out to peripheral actors when it was required to efficiently conduct trafficking tasks.

Viewed from a broader criminological perspective, the heavy-tailed distribution of collaborators per trafficker, and the presence of poorly-connected traffickers whom are not involved in day-to-day trafficking operations, could owe more to well-known patterns in criminal careers. Criminologists have long established that criminal activity follows a curvilinear relationship with age, reaching a peak in late teenage years and then decreasing (Farrington, 1986). Lower levels of collaborative activity on the part of some traffickers could be a result of older individuals detaching from regular criminal behavior, while the most collaborative traffickers could simply be at the peak of their criminal activity. Further research on the age-crime curve in drug trafficking, and in organized criminal groups more generally, can address this hypothesis.

Building on the first paper, the second study investigated the formation of collaborative ties in the drug trafficking network using Bayesian exponential random graph models. Conceptually, collaboration is critical to the operations of a drug trafficking group; narcotics must be sourced and distributed, and the proceeds of this distribution must be laundered and re-allocated. But relatively little is known about how criminal groups self-organize to perform these endeavours. This paper aimed to address this gap by modeling tie formation as a nonrandom process in which collaboration between traffickers is influenced by endogenous structural features of the network such as transitivity, dyadic-level parameters such as homophily, and individual-level parameters such as task specialization and subgroup membership.

Empirically, transitivity emerged as an important factor in the formation of collaborative ties. Reflecting the strong local clustering uncovered in the first paper, traffickers that collaborated with the same third-party were considerably more likely to collaborate themselves. Moreover, intra-group homophily substantially influenced tie formation; traffickers, who were organized into small groups around three individuals; were more likely to form intra-group collaborative relationships

than inter-group ties. Thus, the results support the argument forwarded by Grund and Densley (2015) that network effects such as homophily and transitivity play an important yet overlooked role in criminal networks.

The paper uncovered systematic variation in the overall propensity for traffickers to seek collaborators according to their task specialization. Traffickers specializing in supplying bulk quantities of narcotics and those charged primarily with money laundering had no greater overall demand for collaborators. In contrast, distributors were more likely to form ties. Thus, the trafficking operation appeared to place greater cooperative demands on distributors, whom are charged with sourcing narcotics from the suppliers, on the one hand, and with forwarding the proceeds of narcotics sales to the launderers, on the other. The differential homophily parameters associated with task specialization supports this view—distributors' greater propensity for tie formation is not explained by a surplus of collaborative relationships with other distributors. Rather, distributors were more likely to collaborate with suppliers and launderers, while traffickers specializing in the latter two activities exhibited within-specialization homophily. Varese (2013) and Campana and Varese (2013) develop a versatile schema for classifying the collaborative activities undertaken by organized criminal groups into protection activities, resource acquisition, group management, and economic investments. The results outlined above suggest that resource acquisition and economic investments should be viewed in the context of a supply chain, particularly if the group is engaged in complex activities requiring a clear division of labor.

During the last 15 years, studies of organized criminal networks have increasingly used network data. Indeed, the field of criminal networks is developing in its own right. The second paper contributes to this development by signposting several methodological issues that analysts working in this field are likely to encounter, such as the limitations of linear regression models in the context of high-dependency data and problems of degeneracy when fitting exponential-family models to infer tie formation. Methodologically, the paper demonstrates the Bayesian approach to fitting exponential-random graph models. This framework offers at least two

advantages over the more widely used ERGM. First, the Bayesian approach allows the analyst to specify independent prior distributions for each parameter in the model. Analysts encountering problems with degeneracy when fitting exponential-family models to criminal network data may find that specifying priors mitigates these problems. Second, the framework allows one to fine tune the process of model fitting by, for example, varying the chain length and the number of chains used to converge on parameter estimates, which can reduce the time needed to find suitable models.

Together, the first and second papers provide a view into the inner-workings of an organized criminal group operating in one of the largest illicit markets. When considered together, the trafficking group emerges as profoundly unequal in terms of the distribution of collaborative ties. A small number of traffickers actively collaborated with a large fraction of their peers, and therefore seem to be particularly capable of coordinating trafficking activities. But the most collaborative traffickers tended not to collaborate directly with each other, suggesting that the network contained at least two authority structures that acted concurrently. Further investigation is needed to determine whether these features are unique to the trafficking network studied in this thesis, perhaps owing to its subgroups or operations in more than one geographical location, rather than a general property of drug trafficking organizations.

The third paper in the thesis analyzed the incidence of homicide arrests in Chicago during 2000–2013. For this study, I projected a bipartite network of arrest events, constructed using court records, into a unipartite network of co-arrest ties between individuals. The aim of the study was to examine the social dynamics of homicide arrests and re-arrests. More specifically, I sought to investigate the influence of embeddedness in the co-arrest network on the incidence of homicide re-arrests—where individuals whom have previously been arrested for any crime are subsequently arrested for homicide—and on homicide co-arrests—where two individuals are arrested together. This paper makes a empirical, conceptual, and methodological contribution.

Empirically, embeddedness in the co-arrest network was shown to influence the incidence of homicide arrests vis-a-vis non-homicide arrests over a thirteen year period. Specifically, the results showed that the probability of homicide arrest in a single-year period was negatively associated with embeddedness in co-arrest dyads and transitive co-arrest triads in the previous year. This association persists after accounting for individual-level covariates, such as age, gang membership, and ethnicity, and the independent influence of prior homicide and non-homicide arrests. Moreover, the findings indicate that co-arrests for homicide were ethnically homophilous and that gang members were more likely to be co-arrested with other gang-members.

Importantly, I demonstrated multiplex interdependence between co-arrest dyads in the homicide network and co-arrest dyads in non-homicide arrest networks, such as for aggravated assault, robbery, and weapons violations. Pairs of individuals whom have been co-arrested for homicide are more likely to be (or to have been) co-arrested for non-homicide violent crimes, weapons violations, and drugs offenses. Farrington (1998) found that violent offenders tend to be frequent or persistent offenders and that individual factors, such as socioeconomic circumstances, and situational factors, such as the escalation of apparently trivial altercations, are predictors of repeated involvement in violent crime. The results regarding multiplex interdependence suggest that membership of violence-involved co-arrest ties may also be a relevant factor.

Conceptually, the focus on embeddedness in co-arrest networks shifts our view beyond individual-level determinants of re-arrest, such as age, ethnicity, and prior criminal record, to the relational factors that influence re-arrests. Criminologists have long theorized that social dynamics play an important role in arrests, offending, and recidivism. However, these dynamics have been difficult to study empirically due to the difficulty of obtaining relational data on the social ties between individuals engaged in criminal behavior. The third paper used co-arrest data as a measure of social ties. I aimed to demonstrate that co-arrest networks are not simply artifacts left behind by arrest events, but that they instead represent underlying

social relationships that constitute part of the opportunity structure for criminal behavior. Viewed in this way, co-arrest ties are tantamount to real, on-the-ground social ties. The results in the third paper demonstrated that these ties have a discernible influence on re-arrest and co-arrest patterns for homicide relative to other violent crimes.

Methodologically, in addition to the analytical strategy of modeling arrest outcomes conditional upon the prior co-arrest network, I also provided a modeling strategy for estimating the influence of social ties using temporal network autocorrelation models and multiplex ERGMs. The TNAM models can be used to evaluate the probability of arrest conditional upon status in prior co-arrest networks. I incorporated the rare-event analysis framework into the TNAM models in order to reduce the computational problems associated with estimating models with complex network parameters on large data sets. Smith and Papachristos (2016) study multiplex edges in the context of organized crime using bivariate ERGMs. While this approach allows for the dependency between different two types of edges to be modeled directly, it is limited to pairwise comparisons. The multiplex ERGM model I presented represents an alternative to this approach in which several edge types can be modeled simultaneously. The basis of this method is to use network statistics for one type of edge as exogenous parameters for modeling the existence of a different type of edge between a particular set of nodes.

While the paper focussed on re-arrests, and did not set out to analyze first-time arrests or recidivism, in principle, such data could be coupled with data on convictions and incarceration to analyze the role of ties in re-offending following prisoner re-entry. Moreover, my findings suggest that the structure of co-arrests is likely to influence the incidence of repeated arrests for crimes other than those explored in this thesis. Indeed, as discussed in the third paper, assessing repeated arrests for minor crimes, such as vandalism and petty theft, would reduce problems of non-observed censoring due to incarceration. Crucially, this would permit a direct analysis of arrest versus non-arrest because non-arrests could be reliably modeled without introducing censoring bias.

5.2 Challenges in studying criminal networks

This thesis, and the study of criminal networks more generally, is subject to a range of methodological challenges that, while similar in kind, often exceed in scale those faced in the study of non-criminal social networks. There are three challenges that emerged in this thesis that warrant further discussion: the problem of missing data and non-random sampling; the selection and application of models to network data; and the integration of network data with individual-level outcome variables and contextual data. I outline these challenges below and signpost several recently proposed methodological solutions and modeling options.

5.2.1 Missing data and non-random sampling

The problem of missing (or incomplete) data is not unique to criminal networks. However, the problem is often particularly severe in studies on criminal organizations due, in part, to the clandestine nature of criminal behavior and the associated problem that data derived from survey methods, such as name generators, is usually unattainable. In this thesis, incompleteness emerged in the drug trafficking data. While the problems associated with missing and incomplete are numerous,[1], there are two challenges related to the trafficking data that may recur in future studies on organized crime groups using archival data, and which therefore merit closer attention.

The first problem relates to network boundary specification. Whereas in the third paper, the boundaries of the co-arrest network were clearly defined—arrest events must have occurred within the jurisdiction of the Chicago Police Department, and the study used the entire population of arrests—in the trafficking data, it was less clear where the boundaries should be drawn.[2] The archival data named defendants, but these actors formed only part of the trafficking organization. Moreover, the actors in the wider archival data included only those captured in the net of the investigation. Thus, whether the "true" boundaries of the network, defined as all

[1] See Ibrahim et al. (2005) for a review of methods for dealing with missing data.

[2] See Campana (2016) and Morselli (2009, pp. 44–50) for a more general discussion of the problem of boundary specification in criminal networks.

actors involved in the trafficking operation, were covered by the data is unknown to the researcher. In the case of the trafficking data, the specification problem is "outward-oriented" in nature. That is, at issue was whether some actors who were within the true boundaries of the trafficking network, and the edges associated with those actors, were missing from the data. In contrast, an "inward-oriented" specification problem arises when the researcher has data that may include actors and edges that are not members of the criminal organization under study, but which are nevertheless included in the data.

Inward-oriented boundary specification problems are considerably less likely to be encountered when drawing on archival data based on law enforcement investigations in which, by virtue of being caught in the net of the investigation, the actors included in the data are likely to fall within the boundaries of the criminal network. However, outward-oriented specification problems are very likely to occur, as the investigation and related data are unlikely to capture the entire network. This is especially problematic because one cannot be sure about the extent of missing data, even by closely scrutinizing the available records (Berlusconi, 2013).

This outward-oriented boundary specification problem gives rise to the second problem: the incompleteness of data obtained from archival sources and criminal investigations is non-random. With incompleteness, one can think of these data as a sample of the full data. In typical non-dyadic data analyses, sampling in this way does not constitute a major problem, particularly when the researcher can reasonably assume a distribution from which the sample was drawn (Rubin, 1976). However, when data are dyadic in structure, as in the case of a criminal network, a missing node may contribute missing edges, which affects the observed state of any other nodes whom would have been (i.e. if the researcher had the full data) attendant to those edges.[3] Still, if one can assume the network data are sampled at random, this problem is surmountable using network data imputation methods, such as imputation by reconstruction, imputation using preferential attachment (Huisman, 2009), or Bayesian data augmentation (Koskinen, Robins, and Pattison, 2010).

[3]For a discussion of the effects of missing data in networks, see Kossinets (2006).

However, when network data missingness is non-random, the process of imputation must be more carefully considered. This is particularly the case with data based on criminal investigations, where it is plausible a priori that nodes with low degree centrality are more likely to evade criminal investigation and therefore be missing from the data, leading the researcher to be faced with "missingness that depends on the variable itself" (Gelman and Hill, 2007, p. 530).[4] Conventional measures such as degree, clustering, and assortativity, as well as broader properties such as the core-periphery structure of the network, are therefore prone to bias.

A recent study by Berlusconi et al. (2016) provides a strategy for missing-edge prediction based on node similarity that can mitigate the non-random missing data problem. This strategy relies on coding edges in the network which are "marginal," where a marginal edge is defined as one that appears early in the proceedings of a criminal investigation, but which is discarded at some later stage of the investigation. In other words, marginal edges are relationships that were suspected to exist early on in the investigation, but which were later deemed to either not exist or to be irrelevant to the criminal operations of the network under investigation. The properties of these marginal edges, particularly the similarity of the nodes connected by the marginal edge, as captured by individual-level covariates, can then be used as a basis for predicting unobserved or missing edges. Berlusconi et al. (2016) argue that missing edges are likely to exhibit "opposite characteristics" to marginal edges. Thus, if one finds that marginal edges are often characterized by node dissimilarity, nodes with high similarity are more likely to share a missing edge. While the veracity of this approach requires further testing, initial evidence indicates that the method correctly predicts edges that were initially missing from the data, but which could later be verified to exist from further investigatory data.

In the case of the trafficking network data used in the first two chapters of this thesis, missing-edge prediction would require revisiting the wider investigative data and developing a coding schema for classifying marginal edges. While this would be possible, a clear downside is that such a process would place a large coding

[4]See also Gelman and Hill (2007, pp. 529–543) on the impossibility of proving that data are missing at random.

burden on the researcher. This may be problematic due not only to researchers' time and resource constrains, but also because developing the coding schema introduces greater choice on the part of the researcher, which carries with it the potential for research-induced bias. Nevertheless, missing-edge prediction using node similarity is a promising method for mitigating the problem of non-random missing edges. The imputation of missing nodes for investigatory data requires further methodological innovation.

When data are particularly scarce and incompleteness constitutes a major problem, Campana (2015) and Campana (2016) suggest collecting bipartite network data in which individuals are nested in a set of events—akin to the data in the third paper in this thesis. By centering the data around events, the boundaries are implicitly specified. In essence, a study using this methodology becomes an investigation of the network involved in the events covered by the data, rather than a study pertaining to the unipartite network of an entire criminal organization with the attendant bias introduced by non-random missing data.

5.2.2 The selection and application of models to network data

In analyses of criminal network data, issues often arise relating to model selection, estimation, and convergence. This thesis applied several prominent models, including Bayesian Exponential Random Graph Models, Temporal Network Autocorrelation Models, and Exponential Random Graph Models with exogenous network parameters. Although these choices were appropriate to the data and research aims, they were not the only available options. Due in part to the proliferation of network analysis and related statistical software in the last decade, researchers now have a rich toolkit for modeling criminal network data in ways that are amenable to multivariate statistical inference and hypothesis testing.

One family of models not used in this thesis is Quadratic Assignment Procedure (QAP) multivariate regression (Krackhardt, 1988). In a QAP model, a left-hand side network is regressed on right-hand side matrices of equal dimension. For

instance, the adjacency matrix of collaborative ties in the trafficking network could be regressed on a similarity matrix using a logistic QAP model, where traffickers with the same task specialization are assigned a one in the matrix, and those with different specializations are assigned zero. To account for the dyadic structure of the data which would bias the parameter and standard error estimates in a generalized linear model, as discussed in Section 3.3.1, a large number of permutations (usually 10,000 or a similar order of magnitude) are applied to the left-hand side adjacency matrix such that the order of rows and columns are randomly shuffled (Krackhardt, 1987). Each time the left-hand side is shuffled, the regression model is re-estimated. Thus, in effect, a distribution of correlations between the left-hand side response network and the right-hand side predictor parameters are generated, and these correlations are subsequently built in to the test statistic for whether the response network and predictors are uncorrelated (Hubert and Schultz, 1976; Dekker, Krackhardt, and Snijders, 2007; Krackhardt, 1988). QAP models and well suited to unweighted, undirected network data where interactions between predictor variables are not required. Indeed, QAP regression has recently been applied to criminal network data (Campana, 2015; Campana and Varese, 2013). However, estimation on large networks can be computationally intensive.

In the third paper, I applied a temporal network autocorrelation model to estimate the influence of embeddedness in the co-arrest network on the incidence of homicide arrests in the following year. An alternative modeling strategy would have been to use a Temporal Exponential Random Graph Model (TERGM). This strategy is an extension of the ERGM, where one considers the network at time t to be an instance of a Markov process that underpins the development of the network over discrete time-steps (Hanneke, Fu, and Xing, 2010). Essentially, the model allows for the estimation of the inter-temporal dependence of ERGM parameters, which would permit estimation of the probability of homicide co-arrests, rather than the isolated homicide arrests estimated in the TNAM, given endogenous network statistics at t and lagged network statistics at $t - 1$. More recently, Cranmer, Heinrich, and Desmarais (2014) extended the TERGM to account for temporal variation

in the parameter estimates, which can be used in conjunction with a bootstrap maximum pseudolikelihood estimator for efficient computation (Desmarais and Cranmer, 2012). One downside of the TERGM is that estimation on large networks is computationally burdensome, particularly where nodes are being added and removed from the network between time steps, and where there are several time steps to be modeled. This is due to the fact that each adjacency matrix for the network at time t must be expanded to the dimensions of the combined network over all time steps, ordinarily by adding structural zeros. Alternatively, the matrices can be reduced to equal dimension, but this effectively discards nodes and edges that are not present in each time step. Thus, the TERGM was not appropriate for the co-arrest data in the third paper, but provides an excellent choice for cross-sectional criminal networks observed over time in which the scale of change in the actor composition of the network between time steps is small.

A separate approach to modeling longitudinal network data is to apply stochastic actor-oriented models (SOAMs). This approach conceptualizes changes in the network structure between time-steps as an outcome of the actors' decisions within the network (Snijders, 1996; Snijders, Bunt, and Steglich, 2010). As such, each time step offers an opportunity to actors to change, or not to change, their current set of ties. Similarly to the ERGM, actors tie formation and dissolution choices are the product of endogenous network effects, such as transitivity and related terms used throughout this thesis, exogenous effects, for example individual-level covariates such as gender and ethnicity, and a stochastic component. An advantage of SOAMs is that they can readily be applied to directed networks, as the actor-oriented analytical basis of the model lends to individuals making decisions regarding their ties, rather than two actors jointly deciding whether to form a tie, as would have to be the case in an undirected network. For this reason, SOAMs represent a good choice for modeling reciprocity dynamics in criminal behavior, as in Papachristos (2009).

Finally, researchers applying ERGMs to criminal network data may encounter problems of degeneracy, as was the case in the second paper in this thesis (see Section 3.3.2), while some of the methodological options discussed above are computationally

impractical on very large criminal networks.[5] The BERGM framework used in the second paper may provide relief to issues of degeneracy, particularly for researchers with ample priors regarding the structural tendencies an dyadic parameters, such as the extent of homophily for a given covariate, of the criminal network under investigation. The rare-event TNAM framework used in the third paper provides one avenue for modeling outcomes in large networks, although it has the drawback of relying on maximum pseudo-likelihood estimation, which is likely to bias parameter estimates and standard errors, particularly for criminal networks with greater levels of dependency and structural complexity than the co-arrest network.

There are, moreover, several recent and on-going methodological developments that may provide support for modeling large criminal networks. One such development is latent factor regression (Stewart, 2014). The general form of this approach is to project nodes into a low-dimensional latent space, define a metric that captures the space between the projected nodes, and use this metric as the basis for probabilistic estimates regarding the probability of edges. Nodes that are close together in the latent space are more likely to share an edge, and this extends to higher-order effects such as transitivity in the case of, for example, three nodes in close proximity.[6] The upside of the latent factor approach is that structural terms for network data can be estimated in a fraction of the time required to estimate an ERGM or SOAM.

A second development is the use of cluster-robust variance estimation for dyadic data. The application of cluster-robust errors to account for dependency between observations is now relatively commonplace in regression modeling across the social sciences, such as when units are clustered into geographic regions or units are clustered into unit-time observations. Aronow, Samii, and Assenova (2015) extends cluster robust errors to dyadically structured data.[7] This approach applies an ad hoc adjustment to the estimated variance of the parameters in a generalized linear model (GLM) in order to account for dyadic clustering of observations, such as that

[5] See Lusher, Koskinen, and Robins (2013, pp. 141–166) for an applied discussion of ERGM degeneracy.

[6] See Hoff (2005) for earlier developments of network models based on latent factors.

[7] Cluster-robust standard errors have been applied to large-scale network data in the case of parliamentary networks. This working paper (of which I am co-author) is available on request.

encountered in network data. The advantage of this approach is that GLMs, like ERGM fits using maximum pseudo-likelihood estimation, can be estimated for large network data in a matter of seconds. The plug-in estimator can then be applied to recalibrate the errors in the fitted model. However, a downside to this approach is that, while it accounts dyadic dependence, it does not account for higher-order dependence that may exist in complex criminal networks. Further investigation is required to establish the veracity of the cluster-robust approach compared to MPLE, particularly regarding the extent of bias reduction in the errors and the remaining bias in the parameter estimates due to higher-order dependence.[8]

5.2.3 Integrating network data, outcome variables, and contextual data

The modeling options used in this thesis, and discussed above, provide a set of tools for decomposing the complex structure of criminal networks into a set of parameters of interest to researchers. Whereas in regression analyses in non-network applications of criminal behavior, the researcher is typically interested in examining the association between an outcome variable and a set of covariates, in the analysis of criminal networks the network itself is usually a target of inference. In the third paper, for example, where I examined the incidence of homicide arrests conditional upon embeddedness in co-arrest networks, the dependency in the network was not simply a source of a nuisance dependence that had to be accounted for in order to obtain reliable parameter estimates. Instead, the network was a relevant factor for explaining the occurrence of homicide arrests. Indeed, the strength of network analysis is that it can model the relationship between an outcome of interest and the structure of relationships between individuals for whom that outcome may arise. Using this "instrumental" approach outlined by Campana

[8] Although not discussed here, considerable progress has been made towards models for valued networks. See Krivitsky (2012) for the application of valued ERGMs. Moreover, stochastic actor-oriented models have recently been applied to multilevel networks (An, 2015), to the co-evolution of positive and negative ties (Rambaran et al., 2015; Huitsing et al., 2014), and to the co-evolution of unipartite and bipartite social networks (Stadtfeld et al., 2016). Recent work on dynamic line-graphs presents a promising avenue for analyzing the joint involvement of individuals in criminal events over time (Broccatelli, Everett, and Koskinen, 2016).

(2016), and discussed in Section 2.2.2, much progress has been made towards explaining how the inner-dynamics of criminal networks can shape outcomes such as cooperation (Campana and Varese, 2012), and how the structure of criminal networks is itself shaped by the demands of illicit operations (Morselli, 2009; Crossley, Edwards, and Stevenson, 2012).

An important challenge in the analysis of criminal networks is to integrate contextual data into the relationship between outcome variables and the structure and evolution of criminal networks. For instance, in the case of homicide arrests and the co-arrests network, I did not account for the spatial distribution of the arrests, or neighborhood-level factors that are known to influence arrest rates. As demonstrated by Kirk (2009), the residential location and spatial distance between individuals is likely to be a relevant factor in re-arrests and re-offending. Similarly, in the second paper, the model did not account for the range of geographic locations in which the heroin traffickers operated. Although in both cases this owed to absence of data, in some cases geospatial data and contextual data may be available to the researcher, and considering how to integrate this type of data with network methods therefore bears consideration.

The data that may accompany criminal networks can be differentiated into three types. The first type are nodal attributes, which may influence the outcome of interest as well as the propensity for tie formation in the network. For example, in the analysis of homicide arrests in this thesis, individuals' ethnicity influenced the probability of re-arrest and how likely nodes were to be co-arrested, through conditional disparities in the likelihood of arrest by ethnic group and through ethnic homophily. Typically, node-level data are the most straightforward to integrate into analyses of criminal networks as the modeling options are analogous to regression modeling. Node attributes can be incorporated as propensity, or popularity, effects, capturing how the tendency to form ties varies across attributes. These attributes can also be modeled dyadically, by using homophily parameters for example, to account for influences on tie formation due to similarity and difference in attribute values. Anti-homophily parameters enable the researcher to model the inverse relationship,

i.e. negative relationships between nodes. Of course, the consideration of how to parameterize node-level attributes depends largely on the type of network data. With directional data (which were not used in this thesis) the researcher may wish to account for the association between node-level variables and reciprocity dynamics, in addition to in- verses out-popularity effects. With valued data, a target of inference may be how the strength of a relationship varies according to node attributes.

The second type of accompanying data are edge attributes. While nodal attributes can be viewed as within-individual random variables, edge attributes characterize the ties between individuals. In many cases, whether data are more appropriately treated as node or edge attributes is clear. In the trafficking data, subgroup membership was treated as a nodal attribute, while in the co-arrest data, the type of crime pertaining to each co-arrest was treated as an attribute of those co-arrest edges. However, in some cases, the decision is less clearcut. In the case of the trafficking data, the source document contained information on the activities for which the traffickers were collaborating and characterizations of the role of traffickers as suppliers, distributors, or money launderers. I thus had three options: including the activity as an edge attribute of the collaborative tie, coding the characterizations into a nodal attribute, or both. Ultimately, the activity data were sparse, so I opted to code specialization as a nodal attribute and omit the characterization of the collaborative edges. This restricted my modeling options. Specifically, although the influence of specialization could still be estimated using homophily parameters, I was unable to model the relationship between nodal attributes and the frequency of collaboration for different types of activity. Moreover, I could not investigate the multiplexity of collaborative relationships among the traffickers, which would have added greater context to the patterns of collaboration in the trafficking group.

The above example of a coding decision regarding node and edge attributes calls attention to perhaps the most challenging aspect of studying criminal networks using researcher-coded archival and investigatory data. For studies drawing on this kind of data, such as the first and second papers in this thesis, it is important to consider, prior to coding, the inferential aims of forthcoming analyses in order to

guide data collection and coding decisions. Archival data often lacks comprehensive indexing, particularly in the case non-digitized source documents, which means that, to a certain extent, the researcher cannot know what data are contained in the archival documents without first undertaking a review. This is problematic when the source documents are large in quantity, and is exacerbated further when the informational content of the documents is highly variable. For instance, the archival documents behind the trafficking data included more than ten thousands non-digitized pages. Many were stenographers notes from court proceedings, which were information poor with respect to the traffickers and to collaboration between traffickers. But one of the documents, the Government's Sentencing Memorandum, was information rich and ultimately contained the data for the first two papers.

Clearly, it is desirable to obtain as much data as possible. However, the researcher has to balance this against time investment in searching through the source documents and coding. I offer three general guidelines for the collection of criminal network data from archival and investigatory sources. Firstly, it is important to be flexible with respect to the specific criminal organization or investigation one wishes to study. Rich data that provide comprehensive network coverage may not be accessible for many organizations and clandestine groups. Secondly, if the level of incompleteness of the data is high, particularly for some covariates or types of edges, the researcher should nevertheless extract as much data on these covariates and edges as possible. By virtue of methodological advances in the imputation of missing data, as discussed above, problems of missing covariates values and edges can be mitigated. Coding "marginal edges" may be an effective strategy for imputing edges not covered by the data. Thirdly, consistency in the source documents renders the data more amenable to computer-assisted coding, and more importantly, means that the information regarding social ties is more likely to be consistent. The sentencing memorandum source for the trafficking data used the same identifiers for individuals throughout the document and the textual descriptions of collaboration events were formatted similarly across events. When different types of documents are used in

combination, individual identifiers may vary across the documents, and it is likely to be more difficult to maintain stability in the coding definition for ties.

The third type of data that may accompany criminal network data are contextual (or exogenous) variables. These data may be spatial in nature, or neighborhood-level and aggregate-level variables. For example, Crossley, Edwards, and Stevenson (2012) use aggregate-level arrest data to divide a covert network of suffragette activism into phases. Morselli (2001) draws on data on the frequency and size of narcotics consignments to assess the influence of network brokerage on productivity in drug distribution. The co-arrest data used in this thesis contain spatial identifiers of the region in which each arrest event occurred in Chicago. In principle, this data could be coupled with survey data on social and economic disadvantage in those regions to investigate variation in the likelihood of co-arrests across neighborhoods, for instance. Integrating this type of contextual data is challenging, but adds richness to network studies by providing additional modeling options and adding nuance to research findings.

5.3 Directions for future research

This thesis suggests several lines for future research. The first study found degree dissortativity in a drug trafficking network and demonstrated that dissortativity is positively associated with the vulnerability of the network to node removal. Whether this property generalizes to other criminal networks remains unclear. Incorporating dissortativity measures into future analyses would help to clarify whether it is commonplace for criminal organizations to exhibit dissortative mixing. Conveniently, this measure can be calculated using basic data on the presence of edges, as constituted in an adjacency matrix or edgelist. Thus, the first paper suggests a concrete step that can be readily undertaken without increasing demands on data. To the extent that dissortativity is a general property, which leaves the network weaker with respect to node removal, one has to consider why dissortativity might be an emergent property of criminal networks. Perhaps there are advantages in terms of the efficiently performing complex criminal operations, or benefits to

the individuals whom are able to monopolize connectivity in the network; brokerage roles in networks have long been linked to performance and innovation in the social networks literature (Burt, 2000; Gnyawali and Madhavan, 2001). A further line of inquiry relating to assortativity is the association between degree mixing and the strength of ties. This is consequential to the extent that ties between high degree nodes may be stronger, on average, than ties between high degree and low degree actors, or vice versa. This would imply that, while less frequent, the ties connecting highly central actors are "activated" more often, enabling criminal operations to resolve coordination problems, for example.

In terms of the structural vulnerability of criminal networks to node removal, this thesis used a process of sequential degree-targeted removal to produce the analytical results in the first paper. This is just one of several options for analyzing the resilience of networks. Several alternative methods have been proposed, particularly outside of the criminal networks literature, including targeting by other centrality metrics such as betweenness (Crucitti et al., 2004), random targeting (Callaway et al., 2000; Albert, Jeong, and Barabási, 2000), and edge rather than node targeting (He, Li, and Hongru, 2009). Another option is to investigate cascading removal, where nodes are removed in a sequence determined partly by their connections, in such a way that removals spread out through the network. Using these methods in addition to degree-targeted removal may lead to greater insights on the structural resilience of criminal networks.

However, from an analytical perspective, each of these methods is based on the sequential removal of nodes or edges, which may not be the most suitable strategy for evaluating the effects of real-world law enforcement interventions into criminal networks. The arrest of individual drug traffickers one after another, for instance, is unlikely to occur in a real intervention by law enforcement agents, because it would forewarn other members of the network that they were under investigation. Traffickers' may then destroy evidence, become more covert in their operations, and generally adopt behaviors that would reduce the capacity for legal repercussions. Instead, to more closely approximate the type of interventions that criminal networks

may be adapted to prevent, it would be helpful to develop an algorithm that removes nodes in groups or waves. The composition of these groups should be determined, in part, by the structural properties and position of individuals within the criminal network. It may be reasonable, moreover, for the algorithm to account for the knock-on effects of group arrests. Specifically, law enforcement agents may gain information from the first group of arrests, through plea-bargain agreements and similar strategies, leading to a greater risk of exposure for nodes connected to the individuals in the arrested group. Developing an analytical strategy along these lines, i.e. that more closely simulates real-world interventions, would be useful for investigating the resilience of criminal networks, and therefore examining the extent to which the structures of criminal networks are adapted in response.

The second paper uncovered a pattern of differential collaboration by task specialization among the drug traffickers. Distributors had a greater overall demand for collaborative ties, but were less likely to share ties with other distributors. To better understand the drivers of cooperation in drug trafficking, a next step would be to examine the collaboration events themselves. One strategy would be to code collaboration events using the schema developed in Varese (2011) and Varese (2013) for categorizing the various tasks involved in organized criminal activities. It may be that distributors require greater coordination than suppliers and launderers, and therefore many of the collaboration events involving distributors relate to "group management" (Varese, 2013). Moreover, an investigation of the geographical constraints and determinants of collaboration in the narcotics industry would advance our understanding of how traffickers negotiate the myriad complexities involved in conducting trans-national criminal activities (Campana, 2011). This would require exogenous data on the geographical location of collaboration events, as discussed above, including information on the frequency of communications across territories.

The work in the third paper of thesis could be extended along four lines. Analytically, focusing on re-arrests overlooks the social circumstances surrounding first-time arrest. While individual-level risk factors for arrest and criminal offending

have been investigated elsewhere, the social risk factors are less well understood. However, the findings in the third paper suggest that social factors, specifically embeddedness in social networks that are routinely subject to arrest events, may play an important role in the risk of arrest. From an empirical perspective, this is challenging because co-arrest data cannot capture the social embeddedness of individuals prior to sustaining arrests. Thus, extensions along this line would require data on social ties drawn from a different source.

Secondly, the study of social dynamics and arrests could be extended to minor crimes. The third paper focused specifically on the incidence of homicide arrests as a function of co-arrests for other crimes and embeddedness in the prior co-arrest network. However, homicide arrests are relatively rare events relative to some crimes, particularly drug offenses, theft, and low-level misdemeanors. One might reasonably expect the influence of network embeddedness to be somewhat different in the case of minor crimes for two reasons. Firstly, the co-arrest network for such crimes is substantially larger, and is constituted by more complex structures than for homicide arrests, for which the co-arrest network is largely made up of dyads and triads. Secondly, repeat arrests for minor crimes are much more common, partly due to policing strategy as discussed in Section 4.2.1, but also because incarceration is less likely to follow minor crime arrests, or at least be shorter in duration, than is ordinarily the case for violent crimes such as homicide. Thus, the influence of embeddedness in the co-arrest network on the risk of minor crime re-arrests and co-arrests may be greater than for homicide. Importantly, frequent re-arrests for minor crimes, particularly those not followed up by charges, may negatively impact police-community relations and adversely affect the legitimacy of policing in neighborhoods with high arrest rates.

Thirdly, an important aspect of re-arrests and co-arrests not analyzed empirically in this thesis is the role of policing strategy. One avenue for examining the policing side of arrests would be to analyze data on which police conducted arrests. Using this data, one could investigate how often recurrent within-individual arrests are carried out by the same law enforcement agent, or by law enforcement agents based

in the same subdivision of the city police department. As an analytical strategy, controlling for the beat in which arrests occurred might enable arrests based on explicit suspect profiling to be disentangled from arrests relating to emergency-call responses and more routine beat policing. A qualitative study into contemporary policing may also shed light on the process of suspect profiling. At present, this process is relatively obtuse to outside observers.

A final empirical extension would be to investigate recidivism following prisoner re-entry using co-arrest data. The relationship between prior position in the co-arrest network and subsequent re-arrests demonstrated in this thesis suggests that recidivism and re-offending following re-entry may be similarly associated with prior embeddedness in co-arrest ties. A link between recidivism and social ties has previously been proposed (Kirk, 2009). Although arrest data are unlikely to account for all of the social ties that could provide opportunities for re-offending, they nevertheless provide a reliable means of capturing at least some particularly relevant social ties that could be reactivated following prisoner re-entry. An extension along these lines may ultimately provide avenues for developing policy programs and prisoner re-entry strategies that mitigate recidivism and re-offending.

Appendices

A

Robustness of structural parameters to changes in τ

Recall that the τ parameter for the geometrically-weighted statistics defines the decay in the contribution to the parameter estimate of each unit increase in that statistic. For example, in Figure 3.2 (c) in Section 3.3, τ defines the weight of the contribution of vertices 1, 2, and 3 to the odds of the edge G_{ij}.

Note that, in terms of interpretation, $\tau \to 0$ reduces the contribution of additional shared partners to the parameter estimate for the statistic concerned. Conversely, $\tau \to \infty$ assigns a similar contribution of each additional shared partner to the parameter estimate. Thus, when τ is near zero, the contribution of 2 shared partners is minimal, and 3 is much smaller still. Whereas if τ is very large, the contribution of 2 shared partners is only marginally smaller than the contribution of 1, and the contribution of 3 is only marginally smaller than 2, and so on. Ordinarily, τ is set somewhere between 0.1 and 1 (Snijders, Pattison, et al., 2006), so I explore values within this range only.

First, I show that both the GWESP and GWDSP parameter estimates are robust to different specifications of τ_{gwdsp}. Figure A.1 shows the mean posterior estimates for both statistics in the main model, as detailed in section 3.3.4. I fix τ_{gwesp} at

log(2) and re-estimate the model with $\tau_{gwdsp} = \{0.1, 0.2, ..., 0.9\}$. Clearly, varying τ_{gwdsp} only trivially affects the GWDSP and GWESP mean posterior estimates.

Second, I show that GWDSP is robust to changes in τ_{gwesp}, while GWESP is largely unaffected except at low values of τ_{gwesp}. Figure A.2 shows the mean posterior estimates, but with τ_{gwdsp} fixed at log(2) and $\tau_{gwesp} = \{0.1, 0.2, ..., 0.9\}$. The posterior estimate for GWDSP is only trivially affected by τ_{gwesp}, and the estimate for gwesp is largely unaffected when $\tau_{gwesp} \geq 0.4$. However, at $\tau_{gwesp} = 0.3$, the credible interval widens considerably, encompassing posterior estimates in the range $1.25 < \theta < 3.15$. At $\tau_{gwesp} < 0.3$, the posterior mean estimate falls below 2. Thus, the magnitude of the geometrically-weighted shared partner effect becomes smaller as we impose stricter limits on the contribution of multiple shared partners, suggesting it is common for connected pairs of traffickers to share a large number of mutual collaborators. Nevertheless, the odds of tie formation in trafficker dyads remains substantially higher when they share share mutual partners.

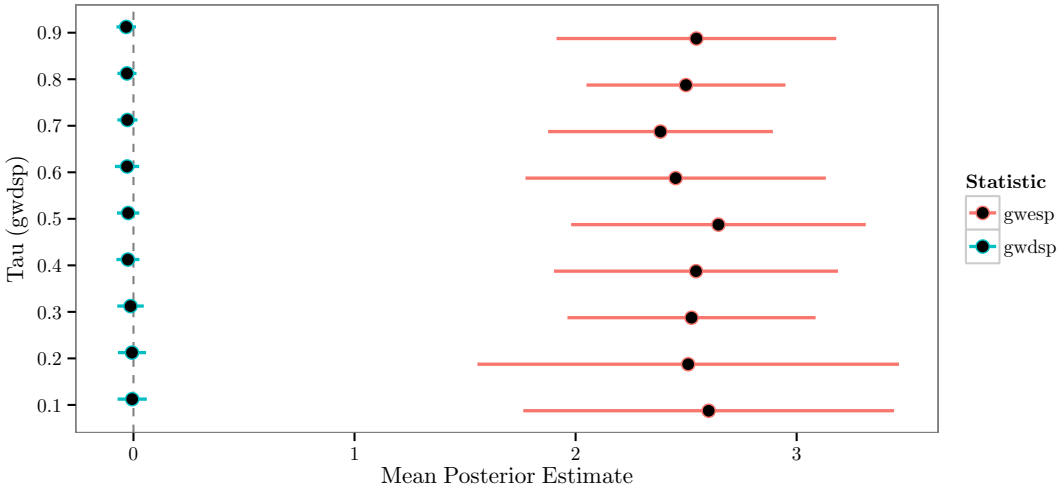


Figure A.1: Mean posterior estimates for the GWESP and GWDSP statistics in the main model. Here, τ_{gwesp} is fixed at log(2) and τ_{gwdsp} is set at values in the range 0.1–0.9 (the reverse of Figure A.1). Each line shows a 95% credible interval around the mean posterior estimate.

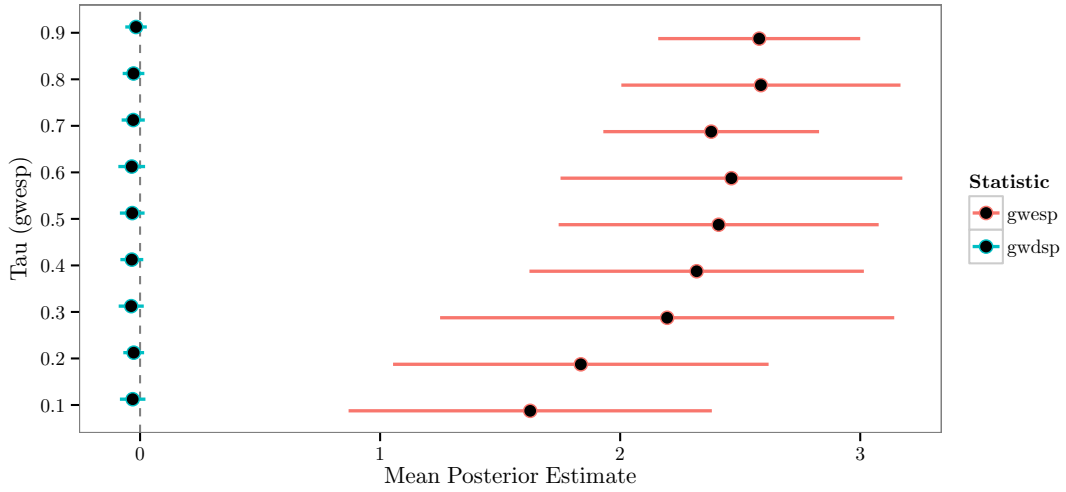


Figure A.2: Mean posterior estimates for the GWESP and GWDSP statistics in the main model. Here, τ_{gwdsp} is fixed at log(2), but τ_{gwesp} is set at values in the range 0.1–0.9. Each line shows a 95% credible interval around the mean posterior estimate.

References

Albanese, Jay S (2011). *Transnational Crime and the 21st Century*. Oxford: Oxford University Press.

Albert, Réka, István Albert, and Gary L Nakarado (2004). "Structural vulnerability of the North American power grid". *Physical Review E* 69(2), p. 025103.

Albert, Réka, Hawoong Jeong, and Albert-László Barabási (1999). "Diameter of the World-Wide Web". *Nature* 401, pp. 130–131.

— (2000). "Error and attack tolerance of complex networks". *Nature* 406, pp. 378–381.

Alexander, Shana (1988). *The Pizza Connection*. New York, NY: Weidenfeld and Nicolson.

An, Weihua (2015). "Multilevel meta network analysis with application to studying network dynamics of network interventions". *Social Networks* 43, pp. 48–56.

Anderson, C J, S Wasserman, and B Crouch (1999). "A p* primer: logit models for social networks". *Social Networks* 21, pp. 37–66.

Aronow, Peter M., Cyrus Samii, and Valentina A. Assenova (2015). "Cluster-Robust Variance Estimation for Dyadic Data". *Political Analysis* 23(4), pp. 564–577.

Ayling, Julie (2009). "Criminal organizations and reslience". *International Journal of Law, Crime and Justice* 37, pp. 182–196.

Baker, Wayne E and Robert R Faulkner (1993). "The Social Organization of Conspiracy: Illegal Networks in the Heavy Electircal Equipment Industry". *American Sociological Review* 58(6), pp. 837–860.

Ballester, Coralio, Antoni Aremngol-Calvó, and Yves Zenou (2004). *Who's who in crime networks: wanted - the key player*. Centre for Economic Policy Research.

Barabási, Albert-László and Réka Albert (1999). "Emergence of Scaling in Random Networks". *Science* 286, pp. 509–512.

Barabási, Albert-László, Hawoong Jeong, et al. (2002). "Evolution of the social network of scientific collaborations". *Physica A* 311, pp. 590–614.

Berlusconi, Giulia (2013). "Do all the pieces metter? Assessing the reliability of law enforcement data sources for the network analysis of wire taps". *Global Crime* 14(1), pp. 61–81.

Berlusconi, Giulia et al. (2016). "Link Prediction in Criminal Networks: A Tool for Criminal Intelligence Analysis". *PLOS ONE* 11(4).

Black, Donald J (1971). "The social organization of arrest". *Stanford Law Review*, pp. 1087–1111.

Blumstein, Alfred (1982). "On the racial disproportionality of United States' prison populations". *J. Crim. l. and Criminology* 73, p. 1259.

Bright, David A, Catherine Greenhill, et al. (2015). "The Use of Actor-Level Attributes to Identify Key Actors: A Case Study of an Australian Drug Trafficking Network". *Journal of Contemporary Criminal Justice* 31(3), pp. 262–278.

Bright, David A, Caitlin E Hughes, and Jenny Chalmers (2012). "Illuminating dark networks: a social network analysis of an Australian drug trafficking syndicate". *Crime, Law and Social Change* 57, pp. 151–176.

Broccatelli, Chiara, M G Everett, and J Koskinen (2016). "Temporal dynamics in covert networks". *Methodological Innovations* 9, pp. 1–14.

Burt, Ronald S (1982). "Toward a structural theory of action: network models of social Structure, Perception, and Action."

— (1987). "Social contagion and innovation: Cohesion versus structural equivalence". *American journal of Sociology*, pp. 1287–1335.

— (2000). "The Network Structure of Social Capital". *Research in Organizational Behavior* 22, pp. 345–423.

Bushway, Shawn D. (1998). "The impact of an arrest on the job stability of young white American men". *Journal of research in Crime and Delinquency* 35(4), pp. 454–479.

Butts, Carter T, Ryan M Acton, and Christopher Steven Marcum (2012). "Interorganizational collaboration in the hurricane Katrina response". *Journal of Social Structure* 13, pp. 1–36.

Caimo, Alberto and Nial Friel (2011). "Bayesian inference for exponential random graph models". *Social Networks* 33(1), pp. 41–55.

— (2013). "Bayesian model selection for exponential random graph models". *Social Networks* 35(1), pp. 11–24.

— (2014). "Bergm: Bayesian Exponential Random Graphs in R". *Journal of Statistical Software* 61(2), pp. 1–25.

Callaway, Duncan S et al. (2000). "Network robustness and fragility: Perlocation on random graphs". *Physical Review Letters* 85(25), p. 5468.

Campana, Paolo (2011). "Eavesdropping on the Mob: the functional diversification of mafia activities across territories". *European Journal of Criminology* 8(3), pp. 213–228.

— (2015). "The structure of human trafficking: Lifting the bonnet on a Nigerian trafficking network". *British Journal of Criminology* 56(1), pp. 68–86.

— (2016). "Explaining criminal networks: Strategies and potential pitfalls". *Methodological Innovations* 9, pp. 1–10.

Campana, Paolo and Federico Varese (2012). "Listening to the wire: criteria and techniques for the quantitative analysis of phone intercepts". *Trends in Organized Crime* 15, pp. 13–30.

— (2013). "Kinship and Violence as Credible Commitments in Criminal Organisations: An Empirical Test". *Rationality and Society* 25(3), pp. 263–289.

Carrington, Peter J. (2011). "The SAGE Handbook of Social Network Analysis". Ed. by J Scott and Peter J. Carrington. Thousand Oaks, CA: SAGE. Chap. Crime and Social Network Analysis, pp. 236–255.

Catanese, Salvatore, Emilio Ferrara, and Giacomo Fiumara (2013). "Forensic analysis of phone call networks". *Social Network Analysis and Mining* 3(1), pp. 15–33.

Chandrasekhar, A and Matthew O Jackson (2014). "Tractable and Consistent Random Graph Models". *arXiv:1210.7375v4*.

Chang, Hui et al. (2007). "Assortativity and act degree distribution of some collaboration networks". *Physica A* 383, pp. 687–702.

Cheatwood, Derral (1988). "Is there a season for homicide?" *Criminology* 26(2), pp. 287–306.

Chicago, City of (2016). *Crimes, City of Chicago Data Portal*. Online. (Visited on 2016).

Clauset, Aaron, Cosma R. Shalizi, and M. E. J. Newman (2009). "Power-law distributions in empirical data". *SIAM Review* 51(4), pp. 661–703.

Cohen, Lawrence E and Marcus Felson (1979). "Social change and crime rate trends: A routine activity approach". *American sociological review*, pp. 588–608.

Coles, Nigel (2001). "It's not what you know – It's who you know that counts: Analysing serious crime groups as social networks". *British Journal of Criminology* 41, pp. 580–594.

Corander, J, K Dahmstrom, and P Dahmstrom (1998). "Maximum likelihood estimation for Markov graphs". *Research report, Department of Statistics, Stockholm University*.

Cranmer, S J and B A Desmarais (2011). "Inferential Network Analysis with Exponential Random Graph Models". *Political Analysis* 19, pp. 66–86.

Cranmer, S J, T Heinrich, and B A Desmarais (2014). "Reciprocity and the structural determinants of the international sanctions network". *Social Networks* 36, pp. 5–22.

Crossley, Nick, Gemma Edwards, and Rachel Stevenson (2012). "Covert social movement networks and the secrecy-efficiency trade off: The case of the UK suffragettes (1906–1914)". *Social Networks* 34, pp. 634–644.

Crucitti, Paolo et al. (2004). "Error and attack tolerance of complex networks". *Physica A: Statistical Mechanics and its Applications* 340(1), pp. 388–394.

Cureton, Steven R (2000). "Justifiable arrests or discretionary justice: Predictors of racial arrest differentials". *Journal of Black Studies* 30(5), pp. 703–719.

Decker, Scott H and Carol W Kohfeld (1985). "Crimes, crime rates, arrests, and arrest ratios: Implications for deterrence theory". *Criminology* 23(3), pp. 437–450.

Dekker, David, David Krackhardt, and T A B Snijders (2007). "Sensitivity of MRQAP Tests to Collinearity and Autocorrelation Conditions". *Psychometrika* 72(4), pp. 563–581.

Desmarais, B A and S J Cranmer (2012). "Statistical mechanics of networks: estimation and uncertainty". *Physica A* 391(4), pp. 1865–1876.

Desmond, Matthew, Andrew V. Papachristos, and David S Kirk (2016). "Police Violence and Citizen Crime Reporting in the Black Community". *American Sociological Review* 81(5), pp. 857–876.

Doreian, Patrick, Klaus Teuter, and Chi-Hsein Wang (1984). "Network autocorrelation models: Some Monte Carlo results". *Sociological Methods and Research* 13(2), pp. 155–200.

Duijn, Marijtje AJ van, Krista J Gile, and Mark S Handcock (2009). "A framework for the comparison of maximum pseudo-likelihood and maximum likelihood estimation of exponential family random graph models". *Social Networks* 31(1), pp. 52–62.

Duijn, Paul AC, Victor Kashirin, and Peter MA Sloot (2014). "The relative ineffectiveness of criminal network disruption". *Scientific Reports* 4.

Erickson, B. H. (1981). "Secret societies and social structure". *Social Forces* 60, pp. 188–210.

Farrington, D (1986). "Age and Crime". *Crime and Justice* 7, pp. 189–250.

— (1998). "Predictors, Causes, and Correlates of Male Youth Violence". *Crime and Justice* 24, pp. 421–475.

Fujimoto, Kayo, Chih-Ping Chou, and Thomas W Valente (2011). "The network autocorrelation model using two-mode data: Affiliation exposure and potential bias in the autocorrelation parameter". *Social networks* 33(3), pp. 231–243.

Gambetta, Diego (1988). *Trust: Making and Breaking Cooperative Relations*. New York: Basil Blackwell.

Gay, Brigitte and Bernard Dousset (2005). "Innovation and network structural dynamics: Study of the alliance network of a major sector of the biotechnology industry". *Research Policy* 34(10), pp. 1457–1475.

Gelman, Andrew and Jennifer Hill (2007). *Data Analysis Using Regression and Multilevel/Hierarchical Models*. Cambridge, UK: Cambridge University Press.

Gendreau, Paul, Tracy Little, and Claire Goggin (1996). "A meta-analysis of the predictors of adult offender recidivism: What works!" *Criminology* 34(4), pp. 575–608.

Gillespie, Colin S. (2015). "Fitting Heavy Tailed Distributions: The poweRlaw Package". *Journal of Statistical Software* 64(2).

Gleiser, Pablo and Leon Danon (2003). "Community structure in jazz". *Advances in Complex Systems* 6(4), pp. 565–573.

Gnyawali, Devi R and Ravindranath Madhavan (2001). "Cooperative networks and competitive dynamics: A structural embeddedness perspective". *Academy of Management Review* 26(3), pp. 431–445.

Goodreau, Steven M, James A Kitts, and Martina Morris (2009). "Birds of a feather, or friend of a friend? Using exponential random graph models to investigate adolescent social networks". *Demography* 46(1), pp. 103–125.

Goodreau, Steven M. (2007). "Advances in exponential random graph (p*) models applied to a large social network". *Social Networks* 29(2), pp. 231–248.

Granovetter, Mark S (1973). "The strength of weak ties". *American journal of sociology*, pp. 1360–1380.

— (1985). "Economic action and social structure: The problem of embeddedness". *American journal of sociology*, pp. 481–510.

Grund, Thomas U. and James A. Densley (2015). "Ethnic Homophily and Triad Closure: Mapping Internal Gang Structure Using Exponential Random Graph Models". *Journal of Contemporary Criminal Justice* 31(3), pp. 354–370.

Guimerá, Roger et al. (2005). "Team Assembly Mechanisms Determine Collaboration Network Structure and Team Perfomance". *Science* 308(5722), pp. 697–702.

Hagan, John (1993). "The Social Embeddedness of Crime and Unemployment". *Criminology* 31(4), pp. 465–491.

Handcock, M S (2003). "Assessing degeneracy in statistical models of social networks". *Center for Statistics and the Social Sciences*, Working paper no 39.

Hanneke, S, W Fu, and E P Xing (2010). "Discrete temporal models of social networks". *The Electronic Journal of Statistics* 4, pp. 585–605.

Harer, Miles D and Darrell Steffensmeier (1992). "The differing effects of economic inequality on black and white rates of violence". *Social Forces* 70(4), pp. 1035–1054.

He, Shan, Sheng Li, and Ma Hongru (2009). "Effect of edge removal on topological and functional robustness of complex networks". *Physica A: Statistical Mechanics and its Applications* 388(11), pp. 2243–2253.

Heaney, Michael T (2014). "Multiplex networks and interest group influence reputation: an exponential random graph model". *Social Networks* 36, pp. 66–81.

Heinze, Georg and Michael Schemper (2002). "A solution to the problem of separation in logistic regression". *Statistics in medicine* 21(16), pp. 2409–2419.

Hoff, Peter D (2005). "Bilinear mixed-effects models for dyadic data". *Journal of the American Statistical Association* 100(469), pp. 286–295.

Holland, P. W. and S. Leinhardt (1981). "An Exponential Family of Probability Distributions for Directed Graphs". *Journal of the American Statistical Association* 76, pp. 33–65.

Holme, Petter and Jing Zhao (2007). "Exploring the assortativity-clustering space of a network's degree sequence". *Physical Review E* 75(046111).

Huang, Jian et al. (2008). "Collaboration over time: Characterizing and modeling network evolution". *International Conference on Web Search and Data Mining (WSDM)*, pp. 107–116.

Hubert, Lawrence and James Schultz (1976). "Quadratic Assignment Procedure as a General Data Analysis Strategy". *British Journal of Mathematical and Statistical Psychology* 29(2), pp. 190–241.

Huisman, M (2009). "Imputation of missing network data: some simple procedures". *Journal of Social Structure* 10(1), pp. 1–29.

Huitsing, G et al. (2014). "Victims, bullies, and their defenders: a longitudinal study of the coevolution of positive and negative networks". *Development and Psychopathology* 26, pp. 645–659.

Hunter, David R. (2007). "Curved exponential family models for social networks". *Social Networks* 29, pp. 216–230.

Ibrahim, J. G. et al. (2005). "Missing-Data Methods for Generalized Linear Models". *Journal of the American Statistical Association* 100(469), pp. 332–346.

Kass, Robert E and Adrian E Raftery (1995). "Bayes Factors". *Journal of the American Statistical Association* 90(430), pp. 773–795.

Kenney, Michael (2007). "The Architecture of Drug Trafficking: Network Forms of Organisation in the Colombian Cocaine Trade". *Global Crime* 8(3), pp. 233–259.

Kenny, David A and Charles M Judd (1986). "Consequences of violating the independence assumption in analysis of variance". *Psychological Bulletin* 99(3), pp. 422–431.

King, Gary and Margaret E. Roberts (2014). "How Robust Standard Errors Expose Methodological Problems They Do Not Fix, and What to Do About It". *Political Analysis* 23, pp. 159–179.

King, Gary and Langche Zeng (2001a). "Explaining rare events in international relations". *International Organization* 55(03), pp. 693–715.

— (2001b). "Logistic regression in rare events data". *Political Analysis* 9(2), pp. 137–163.

Kirk, David S (2009). "A natural experiment on residential change and recidivism: Lessons from Hurricane Katrina". *American Sociological Review* 74(3), pp. 484–505.

Kirk, David S and Mauri Matsuda (2011). "Legal cynicism, collective efficacy, and the ecology of arrest". *Criminology* 49(2), pp. 443–472.

Kirk, David S and Andrew V Papachristos (2011). "Cultural Mechanisms and the Persistence of Neighborhood Violence". *American Journal of Sociology* 116(4), pp. 1190–1233.

Kohler-Hausmann, Issa (2013). "Misdemeanor Justice: Control without Conviction". *American Journal of Sociology* 119(2), pp. 351–393.

Koskinen, J (2004). "Bayesian Analysis of Exponential Random Graphs - Estimation of Parameters and Model Selection". *Research Report: Department of Statistics, Stockholm University* (2004:2).

Koskinen, J, G Robins, and P Pattison (2009). "Analysing exponential random graph (p-star) models with missing data using Bayesian data augmentation". *Statistical Methodology* 7, pp. 366–384.

Koskinen, J, G Robins, and P E Pattison (2010). "Missing data in social networks: Problems and prospects for model-based inference". *MelNet Social Networks Laboratory Technical Report* 09-01.

Kossinets, Gueorgi (2006). "Effects of missing data in social networks". *Social Networks* 28(3), pp. 247–268.

Krackhardt, David (1987). "QAP Partialling as a Test of Spuriousness". *Social Networks* 9, pp. 171–186.

— (1988). "Predicting with Networks: Nonparametric Multiple Regression Analysis of Dyadic Data". *Social Networks* 10, pp. 359–381.

Krebs, Valdis E (2002). "Mapping Networks of Terrorist Cells". *Connections* 24(3), pp. 43–52.

Krivitsky, Pavel N. (2012). "Exponential-family random graph models for valued networks". *Electronic Journal of Statistics* 6(1100).

LaFree, Gary and Kriss A Drass (1996). "The effect of changes in intraracial income inequality and educational attainment on changes in arrest rates for African Americans and whites, 1957 to 1990". *American Sociological Review* 61(4), pp. 614–634.

Lampe, Klaus von (2009). "Organised Crime: Norms, markets, regulation and research". Ed. by Lundgren S. Oslo: Unipub. Chap. The study of organised crime: An assessment of the state of affairs, pp. 165–211.

Laub, John H and Robert J Sampson (2009). *Shared beginnings, divergent lives: Delinquent boys to age 70*. Harvard University Press.

— (1993). "Turning points in the life course: Why change matters to the study of crime". *Criminology* 31(3), pp. 301–325.

Lazega, E and M A J van Dujin (1997). "Position in formal structure, personal characteristics and choices of advisors in a law firm: a logistic regression model for dyadic network data". *Social Networks* 19, pp. 375–397.

Lindelauf, Roy, Peter Borm, and Herbert Hamers (2009). "The influence of secrecy on the communication structure of covert networks". *Social Networks* 31, pp. 126–137.

Lindquist, Matthew J and Yves Zenou (2014). "Key Players in Co-Offending Networks". *CEPR Discussion Paper* (DP9889). URL: `http://ssrn.com/abstract=2444910`.

Lubbers, Miranda J and Tom AB Snijders (2007). "A comparison of various approaches to the exponential random graph model: A reanalysis of 102 student networks in school classes". *Social Networks* 29(4), pp. 489–507.

Lubbers, Miranda J. and T A B Snidjers (2007). "A comparison of various approaches to the exponential random graph model: A reanalysis of 102 student networks in school classes". *Social Networks* 29, pp. 489–507.

Lusher, D, J Koskinen, and G Robins (2013). *Exponential random graph models for social networks: theory, methods, and applications*. Structural Analysis in the Social Sciences. Cambridge, UK: Cambridge University Press.

Lynch, Gerard E (1987). "RICO: The Crime of Being a Criminal, Parts I & II". *Columbia Law Review* 87(4), pp. 661–764.

Malm, Aili and Gisela Bichler (2011). "Networks of Collaborating Criminals: Assessing the Structural Vulnerability of Drug Markets". *Journal of Research in Crime and Delinquency* 48(2), pp. 271–297.

Malm, Aili, Gisela Bichler, and Stephanie Van De Walle (2010). "Comparing the ties that bind criminal networks: Is blood thicker than water?" *Security Journal* 23, pp. 52–74.

Manski, Charles F and Steven R Lerman (1977). "The estimation of choice probabilities from choice based samples". *Econometrica: Journal of the Econometric Society*, pp. 1977–1988.

Mastrigt, S van and Peter J. Carrington (2013). "Crime and networks". Ed. by Carlo Morselli. Abingdon, UK: Routledge. Chap. Sex and Age Homophily in Co-Offending Networks: Opportunity or Preference?

McCormick, G H and G Owen (2000). "Security and coordination in a clandestine organization". *Mathematical and Computer Modelling* 31, pp. 175–192.

McGloin, J and David S Kirk (2010). "An Overview of Social Network Analysis". *Journal of Criminal Justice Education* 21, pp. 169–181.

McGloin, J and A R Piquero (2010). "On the Relationship between Co-Offending Network Redundancy and Offending Versatility". *Journal of Research in Crime and Delinquency* 47(1), pp. 63–90.

McPherson, Miller, Lynn Smith-Lovin, and James M Cook (2001). "Birds of a Feather: Homophily in Social Networks". *Annual Review of Sociology* 27, pp. 415–444.

Messner, Steven F, Lawrence E Raffalovich, and Richard McMillan (2001). "Economic deprivation and changes in homicide arrest rates for white and black youths, 1967–1998: A national time-series analysis". *Criminology* 39(3), pp. 591–614.

Moody, James (2004). "The Structure of a Social Science Collaboration Network: Disciplinary Cohesion from 1963 to 1999". *American Sociological Review* 69(2), pp. 213–238.

Morselli, Carlo (2001). "Structuring Mr. Nice: Entrepreneurial opportunities and brokerage positioning in the cannabis trade". *Crime, Law and Social Change* 35, pp. 203–244.

— (2005). *Contacts, Opportunities and Criminal Enterprise*. Toronto: University of Toronto Press.

— (2009). *Inside Criminal Networks*. New York: Springer.

— (2010). "Assessing Vulnerable and Strategic Positions in a Criminal Network". *Journal of Contemporary Criminal Justice* 26(4), pp. 382–392.

Morselli, Carlo, Cynthia Giguére, and Petit Katia (2007). "The efficiency/security trade-off in criminal networks". *Social Networks* 29, pp. 143–153.

Morselli, Carlo and Katia Petit (2007). "Law-Enforcement Disruption of a Drug Importation Network". *Global Crime* 8(2), pp. 109–130.

Natarajan, Mangai (2006). "Understanding the Structure of a Large Heroin Distribution Network: A Quantitative Analysis of Qualitative Data". *Journal of Quantitative Criminology* 22, pp. 171–192.

Newman, M. E. J. (2001a). "Clustering and preferential attachment in growing networks". *Physical Review E* 64.

— (2001b). "The Structure of Scientific Collaboration Networks". *PNAS* 98(2), pp. 404–409.

— (2002). "Assortative Mixing in Networks". *Phys. Rev. Lett.* 89(20), p. 208701.

— (2003a). "Mixing Patterns in Networks". *Physical Review E* 67(2), p. 026126.

— (2003b). "The structure and function of complex networks". *SIAM Rev.* 45, pp. 167–256.

— (2004). "Fast algorithm for detecting community structure in networks". *Physical Review E* 69(6), p. 066133.

— (2005). "Power laws, Pareto distributions and Zipf's law". *Contemporary Physics* 46(5), pp. 323–351.

Newman, M. E. J., Stephanie Forrest, and Justin Balthrop (2002). "Email networks and the spread of computer viruses". *Physical Review E* 66(035101).

Newman, M. E. J. and Juyong Park (2003). "Why social networks are different from other types of networks". *Physical Review E* 68(3), p. 036122.

Ousey, Graham C and Matthew R Lee (2008). "Racial Disparity in Formal Social Control An Investigation of Alternative Explanations of Arrest Rate Inequality". *Journal of Research in Crime and Delinquency* 45(3), pp. 322–355.

Pager, D. (2003). "The mark of a criminal record". *American Journal of Sociology* 108(5), pp. 937–975.

Papachristos, Andrew V (2009). "Murder by Structure: Dominance Relations and the Social Structure of Gang Homicide". *American Journal of Sociology* 115(1), pp. 74–128.

— (2011a). "The Coming of a Networked Criminology". *Advances in criminological theory* 17. Ed. by John MacDonald, pp. 101–121.

— (2014). "The network structure of crime". *Sociology Compass* 8(4), pp. 347–357.

Papachristos, Andrew V, Anthony A Braga, and David M Hureau (2012). "Social networks and the risk of gunshot injury". *Journal of Urban Health* 89(6), pp. 992–1003.

Papachristos, Andrew V. (2011b). "Measuring Crime and Criminality: Advances in Criminological Theory". Ed. by John MacDonald. New Brunswick, NJ: Transaction Publishers. Chap. The Coming of a Networked Criminology?

Papachristos, Andrew V., David M. Hureau, and Anthony A. Braga (2013). "The Corner and the Crew: The Influence of Geography and Social Networks on Gang Violence". *American Sociological Review* 78(3), pp. 417–447.

Papachristos, Andrew V. and Chris M. Smith (2014). "Crime and Networks". Ed. by Carlo Morselli. New York: Routledge. Chap. The Small World of Al Capone: The Embedded and Multiplex Nature of Organized Crime.

Pettit, Becky and Bruce Western (2004). "Mass imprisonment and the life course: Race and class inequality in US incarceration". *American sociological review* 69(2), pp. 151–169.

Pons, Pascal and Matthieu Latapy (2006). "Computing Communities in Large Networks Using Random Walks". *Journal of Graph Algorithms and Applications* 10(2), pp. 191–218.

Prentice, Ross L and Ronald Pyke (1979). "Logistic disease incidence models and case-control studies". *Biometrika* 66(3), pp. 403–411.

Qin, Jialun et al. (2005). "Intelligence and Security Informatics". Ed. by Paul Kantor et al. Vol. 3495. Lecture Notes in Computer Science. Springer Berlin Heidelberg. Chap. Analyzing Terrorist Networks: A Case Study of the Global Salafi Jihad Network, pp. 287–304.

Ramasco, José J., S. N. Dorogovtsev, and Romualdo Pastor-Satorras (2004). "Self-organization of collaboration networks". *Phys. Rev. E* 70(036106).

Ramasco, José J. and S. A. Morris (2006). "Social inertia in collaboration networks". *Physical Review E* 73, p. 016122.

Rambaran, J A et al. (2015). "The development of adolescents' friendships and antipathies: A longitudinal multivariate network test of balance theory". *Social Networks* 43, pp. 162–176.

Reuter, Peter, Robert MacCoun, and Patrick Murphy (1990). *Money From Crime: A Study of the Economics of Drug Dealing in Washington, D.C.* Santa Monica, CA: RAND.

Robins, G, P Pattison, and P Wang (2009). "Closure, connectivity and degree distributions: Exponential random graph (p*) models for directed social networks". *Social Networks* 31(2), pp. 105–117.

Robins, G, T A B Snijders, et al. (2007). "Recent Developments in Exponential Random Graph (p*) Models for Social Networks". *Social Networks* 29(2), pp. 192–215.

The Pizza Connection (2016). Society of Former Special Agents of the FBI.

Rossi, Ryan A and Nasreen K Ahmed (2015). *The Network Data Repository with Interactive Graph Analytics and Visualization*. URL: `http://networkrepository.com`.

Rubin, Donald B (1976). "Inference and missing data". *Biometrika* 63(3), pp. 581–592.

Saul, Zachary M and Vladimir Filkov (2007). "Exploring biological network structure using exponential random graph models". *Bioinformatics* 23(19), pp. 2604–2611.

Schaefer, David R. (2012). "Youth co-offending networks: An investigation of social and spatial effects". *Social Networks* 34(1), pp. 141–149.

Shalizi, Cosma Rohilla and Andrew C Thomas (2011). "Homophily and contagion are generically confounded in observational social network studies". *Sociological methods and research* 40(2), pp. 211–239.

Shelley, Louise (Feb. 2006). "The Globalization of Crime and Terror". *EJournal USA*, pp. 42–45.

Sherman, Lawrence W et al. (1992). "Crime, punishment, and stake in conformity: Legal and informal control of domestic violence". *American Sociological Review*, pp. 680–690.

Sierra-Arévalo, Michael (2016). "Legal Cynicism and Protective Gun Ownership among Active Offenders in Chicago". *Cogent Social Sciences* 2(1), pp. 1–21.

Smith, Chris M and Andrew V Papachristos (2016). "Trust Thy Crooked Neighbor Multiplexity in Chicago Organized Crime Networks". *American Sociological Review* 81(4), pp. 644–667.

Smith, Douglas A (1986). "The neighborhood context of police behavior". *Crime and Justice* 8, pp. 313–341.

Snidjers, T A B (2002). "Markov chain Monte Carlo estimation of exponential random graph models". *Journal of Social Structure* 3(2).

Snijders, T A B (1996). "Stochastic actor-oriented models for network change". *Journal of Mathematical Sociology* 21, pp. 149–172.

Snijders, T A B, Gerhard G van de Bunt, and Christian E G Steglich (2010). "Introduction to stochastic actor-based models for network dynamics". *Social Networks* 32, pp. 44–60.

Snijders, T A B, Philippa E Pattison, et al. (2006). "New specifications for exponential random graph models". *Sociological Methodology* 36(1), pp. 99–153.

Sparrow, M K (1991). "The application of network analysis to criminal intelligence: an assessment of the prospects". *Social Networks* 13, pp. 251–274.

Stadtfeld, C et al. (2016). "Assimilation and differentiation: A multilevel perspective on organizational and network change". *Social Networks* 44, pp. 363–374.

Steffensmeier, D. et al. (1989). "Age and the distribution of crime". *American Journal of Sociology* 94, pp. 803–831.

Stewart, Brandon (2014). "Latent Factor Regression in the Social Sciences". *Working Paper*.

Sutherland, Edwin H (1947). *Principles of Criminology*. 4th. Philadelphia, PA: Lippincott.

Tomz, Michael, Gary King, Langche Zeng, et al. (2003). "ReLogit: Rare events logistic regression". *Journal of Statistical Software* 8(2).

Tyler, Tom R, Jonathan Jackson, and Avital Mentovich (2015). "The consequences of being an object of suspicion: Potential pitfalls of proactive police contact". *Journal of Empirical Legal Studies* 12(4), pp. 602–636.

Tyler, Tom R and Cheryl J Wakslak (2004). "Profiling and police legitimacy: Procedural justice, attributions of motive, and acceptance of police authority". *Criminology* 42(2), pp. 253–282.

Varese, Federico (2011). *Mafias on the Move: How Organized Crime Conquers New Territories*. Princeton: Princeton University Press.

— (2013). "The Structure and the Content of Criminal Connections: The Russian Mafia in Italy". *European Sociological Review* 29(5), pp. 899–909.

Venkatesh, Sudhir Alladi (1997). "The social organization of street gang activity in an urban ghetto". *American Journal of Sociology* 103(1), pp. 82–111.

Virkar, Yogesh and Aaron Clauset (2014). "Power-law distributions in binned empirical data". *The Annals of Applied Statistics* 8(1), pp. 89–119.

Watts, Duncan J. and Steven H. Strogatz (1998). "Collective Dynamics of 'Small World' Networks". *Nature* 393, pp. 440–442.

Western, Bruce (2006). *Punishment and Inequality in America*. Russell Sage Foundation.

Williams, Phil (2001). "Transnational Criminal Networks". *Networks and Netwars: The Future of Terror, Crime, and Militancy*. Ed. by John Arquilla and David F Ronfeldt. Washington, D.C.: Rand Coorporation, pp. 61–97.

Wood, George (2017). "The Structure and Vulnerability of a Drug Trafficking Collaboration Network". *Social Networks* 48, pp. 1–9.

Wright, B R E et al. (2001). "The effects of social ties on crime vary by criminal propensity: a life-course model of interdependence". *Criminology* 39(2), pp. 321–348.

Xu, Jennifer and Hsinchun Chen (2008). "The Topology of Dark Networks". *Communications of the ACM* 51(10), pp. 58–65.