PAF 593 Final Project Report - Co-Offending Networks in Phoenix (2018-2023)

Jacob T.N. Young

Title: Co-Offending Networks in Phoenix (2018-2023)

Abstract

This project shows how co-offending networks can be created from open-access data. Specifically, using data from the City of Phoenix Open Data Portal (https://www.phoenixopendata.com/dataset/arrests), two-mode networks representing individuals who are arrested together are built and examined. Attention is dedicated to examining reachability and clustering of the networks, as these are important properties identified in the research literature. Examining the reachability and the clustering in the co-offending networks, indicated that there is a low level of reachability and clustering for all years. However, the results did show an increase in clustering from 2018 to 2023. Clustering and trust are important in co-offending networks as these structures may facilitate group solidarity and greater criminal organization. The policy recommendation is that police agencies focus greater attention on the mechanisms that may lead to the observed increase in clustering.

Introduction

By its nature, crime tends to be a group activity (Warr, 2002). Nearly a century of research points to the important role that social relationships play in criminal behavior, indicating that crime is a "network" phenomenon (Papachristos, 2014). While this body of research has covered diverse networks such as gangs (Grund and Densley, 2014), drug distribution (Bright, Koskinen, and Malm, 2018), and mafia syndicates (Andris, et. al., 2021), the bulk of the work focuses on co-offending networks. That is, the relational structure by which individuals participate in crime together. Concurrently, the evolving landscape of policing strategies has witnessed a paradigm shift towards intelligence-driven methodologies. Within this context, the scrutiny of co-offending networks has emerged as an indispensable tool in the arsenal of crime reduction efforts, as expounded by the scholarly contributions of Rostami and Mondani in 2015. This confluence of research and practical application underscores the intrinsic value attributed to understanding the structural dynamics of co-offending networks, given their pivotal role in informing law enforcement strategies aimed at curbing criminal activities.

Research on co-offending has identified two fundamental structural characteristics, reachability and clustering, that bear significant relevance in the formulation of policies geared towards crime reduction. To embark upon a comprehensive exploration of these structural nuances, this undertaking commences with an extensive review of the existing literature pertaining to the intricacies of reachability and clustering. This meticulous review serves as a foundational step in building a nuanced understanding of the theoretical underpinnings and empirical insights associated with these structural properties within the context of co-offending networks.

Subsequently, the project seamlessly transitions into a granular analysis of reachability and clustering within the spectrum of arrests recorded for individuals in the city of Phoenix, spanning the temporal expanse from January 2018 to November 2023. This empirical investigation delves into the intricate patterns and variations exhibited by these structural characteristics, providing a comprehensive snapshot of the evolving dynamics within the co-offending networks prevalent in the specified geographic and temporal domain.

The culmination of this project entails a reflective and forward-looking discussion on the intrinsic value of the “open-science” approach. Emphasis is placed on the collaborative synergy between researchers and practitioners operating in the domains of crime reduction and network analysis. The discourse underscores the enduring potential for fruitful collaborative endeavors in deciphering the complexities of crime reduction strategies, thus affirming the perpetual significance of cohesive activities in this multifaceted arena. In essence, this project not only contributes empirically to our understanding of co-offending networks but also advocates for the sustained collaborative engagement of scholarly and practical stakeholders in the perpetual pursuit of efficacious crime reduction strategies.

Project Goal

This project seeks to develop an open-science approach by examining adult arrest data from the City of Phoenix Open Data Portal (https://www.phoenixopendata.com/dataset/arrests) to construct the co-offending network of individuals arrested in Phoenix between January 2018 and November 2023. This project makes two important contributions. First, the structure of these networks will be analyzed, and policies examined as they pertain to reachability and clustering. Second, it will document how to wrangle and analyze the open-access data.

Literature Review

Studies of co-offending have shown that two structural characteristics are relevant for developing crime reduction policy: *reachability* and *clustering*. These structural characteristics define important properties of the network and help provide insight about the processes that drive co-offending networks. This section reviews these studies.

Reachability and Diffusion

Networks that have high reachability are those that connect many individuals to each other. For example, while Jim and Mike may not be directly linked, Jim may be linked to

Kate, who is linked to Steve, who is linked to Mike. Thus, Jim can "reach" Mike through this chain. Such a structural property facilities the spread of information, diseases, or risk. In contrast, a network with low reachability makes it difficult for content to spread. For example, if Jim cannot reach Mike through some chain of relationships, then Jim cannot indirectly influence Mike. This feature of a network facilitates "diffusion" or the spread of content.

Diffusion views crime as a consequence of an epidemic like process where some causal aspect is transmitted through the network. Networks that are more amenable to diffusion will influence which policies are adopted to prevent crime. For example, Green, Horel, and Papachristos (2017) examined whether individuals are more susceptible to gun violence if they are connected to other individuals who experience gun violence. Using co-arrests networks among individuals in Chicago from 2006 to 2014, they found that the likelihood of being fatally shot or injured by a firearm increased for an individual if someone whom they had been arrested with experienced this event. They argued that this suggests gun violence diffuses through co-offending networks like a virus. Similar processes of diffusion have been found in diverse outcomes from gang violence (Brantingham, Yuan, and Herz, 2020) to police misconduct (Ouellet, Hashimi, & Vega Yon, 2022).

Networks characterized by high reachability, where causal factors can diffuse, require different interventions than those that are characterized by less reachability. Networks in which individuals are reachable may require identifying individuals or sets of individuals who serve as "bridges" between other individuals. These individuals, when targeted for intervention, may disconnect the network reducing the spread of the causal factor in question. In contrast, networks that are not highly reachable will have disconnected groups, requiring a different intervention approach. Therefore, it is important to examine the reachability of a co-offending network.

Clustering and Trust

Another important feature of co-offending networks for policy considerations is the level of clustering in the network. Networks that have high clustering are those in which two individuals share a common third person. For example, Jim and Mike are friends and Mike and Kate are friends. If Jim and Kate are friends this forms a “triadic” structure. The more triadic structures in the network, the more clustering. For a co-offending network, there is clustering if Steve, Trevor, and Ross all engage in offending together. Such a structural property facilities the spread of information and experience with individuals leading to the development of trust relationships. When individuals know about the behavior of others, they are better positioned to make trust judgments about those individuals. In contrast, a network with low clustering makes it difficult for actors to develop trust.

Clustering and trust are important in co-offending networks as these structures may facilitate group solidarity and greater criminal organization. Moreover, networks where there is high clustering tend to create modularity or sets of individuals that are only connected to members of that set. For example, Grund and Densley (2014) examined the level of clustering among gang members in London. They found that racial homogeneity among gangs was largely explained by clustering in co-offending networks. That is, individuals of the same race tended to engage in crime together, leading to groups of racially segregated gangs. Clustering has also been shown to be important for future offending. Charette and Papachristos (2017) found that individuals were more likely to engage in an offense together if they had similar co-offenders in the past. Therefore, it is important to examine the level of clustering in a co-offending network.

An Open-Science Approach

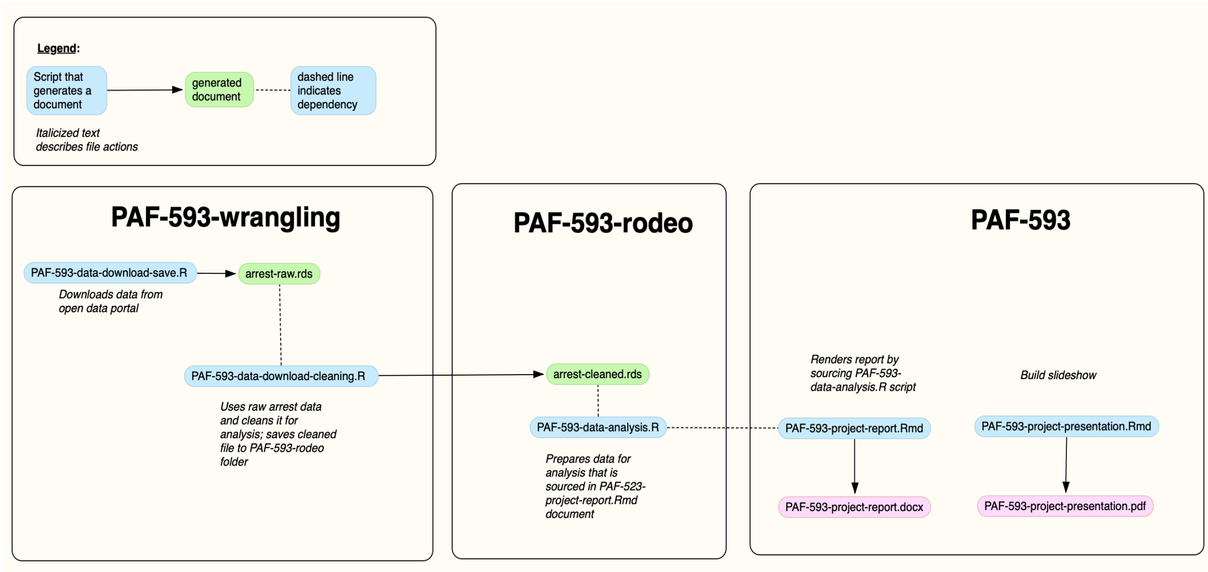
Collaboration among researchers and practitioners in the area of crime reduction and network analysis remains a fruitful area for synergistic activity. However, a key barrier to synergistic activity of this sort is accessing data and sharing information between researchers and departments. As Bright, Brewer, and Morselli (2021: 54) note, negotiating data access is “one of the greatest challenges” for advancing research and policy in this area. Many of the studies on co-offending networks are the consequence of long-term relationship building between researchers and police departments. Such limited access to data on co-offending networks has resulted in researchers using the same data for different projects and multiple studies. The consequence is that the knowledge base for studies examining co-offending networks, and the policies that are gleaned from those studies, rely on a limited or single source of data when making inferences about the processes of interest. Moving toward an approach to research where data are available and accessible would be beneficial to help overcome this barrier between researchers and agencies and generate better policy and better crime reduction efforts.

In the last quarter century, there has been a monumental shift toward making scientific practice more "open". A key component of this so-called open science movement is access to data collected and held by government agencies, often referred to as open government. The open science and open government movements have created a watershed moment for scholars, policymakers, and citizens concerned about public safety. A key feature of the embrace of open government is in how agencies think about data. Specifically, agencies view their data as a strategic asset to achieve their missions. The availability of large amounts of digital data ("big data") amassed in agencies marks a transition toward "governance by data" which is the "gathering, assemblage, formatting, analysis, transmission, and storage of digital data to have governance effects" (Johns, 2021: 54-55). Specific to co-offending networks, an "open science" approach would aid in forming collaborations among researcher and agencies, but would also provide a roadmap for individuals working with other agencies to follow and perhaps even coordinate.

Project Methodology

Data Acquisition and Workflow

Data were downloaded from the City of Phoenix Open Data Portal. A workflow was created whereby the data were downloaded, cleaned, and then analyzed. The workflow and all code used for this project are available at the following GitHub repository: https://github.com/jacobtnyoung/PAF-593-project. As emphasized above, a strength of the open-science approach is that the entire workflow can be examined and reproduced by other researchers or agency analysts. For this project, the workflow is shown in the figure below:



In the upper-left corner, a legend shows the meaning of the line types and colors of the bubbles. Blue bubbles are scripts that execute code to build objects (e.g. a data file, a document). Green bubbles are objects created by the script. This creation process is shown by the connecting solid arrow. Finally, the dashed line indicates a dependency. That is, a blue bubble with a dash line shows that it depends on some other file (e.g. a data set).

In the figure, the workflow for the project is shown. Starting from the left, the PAF-593-wrangling box is a folder of which the files “wrangle” the raw data. As described, the *PAF-593-data-download-save.R* file downloads the data available through the open data portal. This file creates a data file called *arrest-raw.rds*. Next, the *PAF-593-data-download-cleaning.R* file uses the *arrest-raw.rds* file (i.e. a dependency) to clean and prepare a file for analysis. This file, *arrest-cleaned.rds* is stored in the PAF-593-rodeo folder. In that folder, the *PAF-593-data-analysis.R* file uses the *arrest-cleaned.rds* file to build the networks that are analyzed in this paper. This file contains a set of instructions for handling the data and is called from the main folder, PAF-593, with the *PAF-593-project-report.Rmd* file. This file creates the project report document. Finally, the *PAF-593-project-presentation.Rmd* file creates the project presentation document.

As can be seen, the workflow for the project is fully documented and can be reproduced. This is a valuable property for agency analysts who may want to confirm the findings of this study. But, more importantly, for analysts who want to build or modify the approach adopted here. In other words, beyond the analysis presented below, the repository of code and workflow is a valuable asset for an agency.

Wrangling

Data wrangling is the process by which raw data (in this case the data available on the Open Data Portal) are transformed through a series of steps into a format or structure that is ideal for an end user. To build the networks, the raw data required significant "wrangling" in order to get it in a form to use. The major reforming, as processed in the *PAF-593-data-cleaning.R* file, was to restructure the data into an **edgelist**. An edgelist is a data object in which the rows correspond to edges in a network and the columns refer to the nodes. For example, if Jim, Steve, and Louis are involved in a single co-offending event, then the edgelist would have three rows representing, respectively, the tie from Jim to Steve, Jim to Louis, and Steve to Louis.

In order to create this file structure from the raw data, several manipulations were needed. The code to render the edgelist is shown below:

```
dat.edgelist <- dat %>%
  select( YEAR, UNIQUE_NAME_ID, ARST_OFFICER, HUNDREDBLOCKADDR ) %>%
  group_by( HUNDREDBLOCKADDR, ARST_OFFICER ) %>%
  filter( n() > 1 ) %>%
  arrange( ARST_OFFICER ) %>%
  mutate( incident_id = cur_group_id() ) %>%
  ungroup() %>%
  group_by( UNIQUE_NAME_ID ) %>%
  mutate( person_id = cur_group_id() ) %>%
  ungroup()
```

The code above creates an object called `dat.edgelist` that is an edgelist. This object is created by executing the following steps. First, take the object `dat` which represents the raw data pulled from the Open Data Portal. Second, keep the variables YEAR, `UNIQUE_NAME_ID`, `ARST_OFFICER`, and `HUNDREDBLOCKADDR`. These variables are kept because they are used to identify unique co-arrest incidents. Third, the data are "grouped" or sorted according to the `HUNDREDBLOCKADDR` variable and then by the `ARST_OFFICER` variable. Fourth, cases that only have 1 individual are removed. In other words, incidents in which only a single person was arrested are dropped because these are not co-offending incidents. Fifth, a variable that serves as a unique id for the incident is created, `incident_id`, as well as a unique id for the individuals involved in the incident, `person_id`, is created.

This procedure can be seen by reviewing the raw data below:

```
##   YEAR   UNIQUE_NAME_ID  ARST_OFFICER HUNDREDBLOCKADDR
## 1 "2018" "MNI-3643657"   "09865"      "26XX W DUNLAP AVE"
## 2 "2018" "MNI-3441008"   "09352"      "10XX N 48TH ST"
## 3 "2018" "MNI-1668929"   "09685"      "N 27TH AVE & W MCDOWELL RD"
## 4 "2018" "MNI-3295853"   "10101"      "27XX N BLACK CANYON FWY"
## 5 "2018" "MNI-100421207" "09901"      "7XX W INDIAN SCHOOL RD"
## 6 "2018" "MNI-216520"    "10043"      "28XX W MADISON ST"
```

After executing the code discussed above, the wrangled data look like this:

```
##      person_id incident_id
## [1,]      6383        8205
## [2,]     10761        8205
## [3,]      3000        8205
## [4,]      6120       18899
## [5,]      6121       18899
## [6,]     44106       18600
```

As shown, the data have been wrangled into a edgelist where the first column represents the unique id of an individual arrested at that incident and the second column represents a unique id for an incident. For example, the incident “8205” connects individuals “6383”, “10761”, and “3000”. This edgelist can then be used to construct a network object.

Network Construction and Measures

The co-offending data are used to construct two-mode networks. A two-mode network has two sets of nodes: incidents (i.e. arrests) and individuals (i.e. people arrested together) and edges only occur between these partitions (i.e. not within). The definition of a two-mode network is the following: $G = (N, M, L)$ where G is the graph and is defined by two node sets $N = \{n_1, n_2 \dots, n_g\}$ and $M = \{m_1, m_2 \dots, m_g\}$ and an edge set: $L = \{l_1, l_2 \dots, l_L\}$. In this definition, there are N nodes in the first set, M nodes in the second set, and L lines/edges in the network.

In a typical, one-mode network, adjacency is defined between two nodes, n_i and n_j as a link $l_k = (n_i, n_j)$. These data can be represented as an **adjacency matrix**, where each node is listed on the row and the column. The i_{th} row and the j_{th} column of X_{ij} records the value of a tie from *i* to *j*. In two-mode networks, an **adjacency matrix** can also be used to represent the data. The difference is that there are two different sets of nodes N and M. Two nodes from separate node sets, n_i and m_j are adjacent if there is a link $l_k = (n_i, m_j)$. The i_{th} row and the j_{th} column of X_{ij} records the value of a tie from n_i to m_j.

Reachability

Based on the adjacency matrix created above, a measure of reachability can be calculated based on how “reachable” individual **i** is to individual **j** where these nodes are connected through a co-arrest. Reachability is the inverse of the distance between the two nodes where distance is defined as the number of events that connect two individuals. In this analysis, the minimum, mean, and maximum reachabiilty will be examined.

Clustering

In a two-mode network, the extent to which individuals overlap in their co-offenders is capture by the clustering in the network. Specifically, clustering coefficient is:

$$\frac{4 \times C_4}{L_3}$$

where C_4 is defined as a cycle (i.e. individuals 1 and 2 are both connected to events A and B) and L_3 defines a 3-path structure (i.e. individual 1 is connected to events A and B). This ratio is bound between 0 and 1. The ratio represents the proportion of cycles that are closed. A value of 1 indicates that every 3-path is closed (i.e., is embedded in a cycle). Networks with values at or close to 1 will have considerable redundancy in ties. In other words, higher values of the clustering coefficient indicate greater overlap in shared co-offenders.

Analytic Approach

First, descriptive information for each network will be presented. Then, the reachability scores and clustering for the co-offending networks by year will be examined.

Results

Descriptive Information for Co-Offending Networks

Year	Events	Individuals	Edges
2018	7353	13044	15979
2019	7992	12962	16043
2020	6580	10118	12594
2021	5405	8201	10246
2022	4330	6422	7648
2023	3582	6048	7359

The table above shows the descriptive information for the networks. The average number of incidents per year was 5874 and the average number of individuals involved in those events was 9466. The table also shows a general decline in the number of co-offending events from 2018 to 2023.

Description of Reachability for Co-Offending Networks

Year	Mean	Minimum	Maximum
2018	5.941779	0.2542611	16.27271
2019	7.052026	0.2039933	21.41930
2020	6.983182	0.2571481	21.60044
2021	6.778950	0.2085411	21.37547
2022	6.222526	0.1499775	17.32240
2023	5.748357	0.2657048	17.35938

Reachability

Recall that **reachability** in a two-mode network is defined as the number of common nodes that two individuals connect through. In this case, it represents the number of incidents in which two individuals were arrest together. Across all years, the average reachability (i.e. the mean of the mean reachability) is 6.45. This means that the average individual can reach 6.45 other individuals through their co-arrest history. The trend over time has been for reachability to remain roughly similar from 2018-2023.

Description of Clustering for Co-Offending Networks

Year	Clustering
2018	0.0067226
2019	0.0079311
2020	0.0139256
2021	0.0186057
2022	0.0295114
2023	0.0252911

Clustering

The table also shows the scores for the level of clustering in the network. Recall that networks with values at or close to 1 will have considerable redundancy in ties. In other words, higher values of the clustering coefficient indicate greater overlap in shared co-offenders. Overall, the clustering is very low. Across all years, the average clustering is 0.02. This suggest that there is very little overlap in co-offenders. However, analysis of the trend across time shows that there has been an increase in clustering from 2018 (i.e. 0.007) to 2023 (i.e. 0.025). This increase in clustering over time may be indicative of an increase in the tendency for co-offenders to trust other co-offenders.

Discussion and Policy Recommendations

This project embarked upon the development of an open-science methodology with the primary objective of scrutinizing adult arrest data gleaned from the City of Phoenix Open Data Portal. This involved the meticulous construction of a co-offending network, encompassing individuals subjected to arrest within the temporal span from January 2018 to November 2023. Noteworthy in its pursuits, this project stands as a significant contributor on two distinct fronts.

Firstly, an analysis of the *reachability* and *clustering* characteristics within the co-offending networks. These network properties, acknowledged as pivotal in comprehending the intricacies of criminal organizations, revealed a consistently low level of both reachability and clustering across all years under consideration. However, the analysis did bring to light a discernible uptick in clustering from the inception of 2018 to the culmination of 2023. The implications of clustering and trust within co-offending networks are paramount, as these structural elements are perceived to foster group solidarity and contribute to the overarching framework of criminal organization. Furthermore, networks exhibiting heightened clustering tendencies are prone to generating modularity, wherein sets of individuals exclusively interconnect within their designated groups. A salient policy recommendation emanating from these findings advocates for intensified focus by agencies on the underlying mechanisms contributing to the observed surge in clustering.

Secondly, this study demonstrated a systematic approach to navigating and dissecting open-access data. The elucidation of this process underscores the intrinsic value embedded within the City of Phoenix Open Data Portal (accessible at https://www.phoenixopendata.com/dataset/arrests). The portal, by serving as a conduit for external analysis, evaluation, and recommendations, transcends the confines of city agencies and extends its utility to a broader audience. A pivotal revelation emanating from this endeavor is the facilitation of collaborative relationships between external entities and those vested in the pursuit of crime reduction within Phoenix. Consequently, the policy recommendation posits that the continued public accessibility of such data, coupled with an unwavering commitment to maintaining data quality, is imperative for perpetuating the advantages derived from fostering external engagements in the ongoing effort to curb criminal activities.

Conclusion

Criminal activities have long been recognized as inherently social endeavors, as underscored by the scholarly insights of Warr (2002), whose nearly century-spanning research attests to the pivotal role social relationships play in shaping criminal behavior. The elucidation of the networked nature inherent in co-offending, as expounded by

Papachristos (2014) and echoed in the discourse of this report, not only accentuates this sociocentric dimension but also underscores the growing significance of scrutinizing co-offending networks as a cornerstone in endeavors aimed at reducing crime, a sentiment echoed by the scholarly contributions of Rostami and Mondani (2015).

This study, conducted within the framework of an open-science approach, stands as a testament to the capability of leveraging publicly accessible data to delve into the structural characteristics of co-offending networks. The strategic utilization of an open-science paradigm not only showcases the viability of employing such methodologies for scrutinizing network structures relevant to the formulation of crime reduction policies but also positions it as a pioneering avenue for future research endeavors. The intrinsic value of this approach lies in its capacity to unveil and dissect intricate network dynamics that contribute to our understanding of criminal activities.

In summation, the "open-science" approach exemplified in this study emerges as a crucial and forward-looking trajectory for subsequent investigations. It accentuates the need for collaborative endeavors among researchers and practitioners within the realms of crime reduction and network analysis, emphasizing the continued potential for synergistic activities that can yield profound insights into the multifaceted landscape of criminal behavior and inform effective strategies for crime reduction.

References

Andris, C., DellaPosta, D., Freelin, B. N., Zhu, X., Hinger, B., & Chen, H. (2021). To racketeer among neighbors: spatial features of criminal collaboration in the American Mafia. *International Journal of Geographical Information Science*, 35(12), 2463-2488. https://doi.org/10.1080/13658816.2021.1884869.

Brantingham, P. J., Yuan, B., & Herz, D. (2020). Is Gang Violent Crime More Contagious than Non-Gang Violent Crime? *Journal of Quantitative Criminology*, 37(4), 953-977. https://doi.org/10.1007/s10940-020-09479-1.

Bright, D., Koskinen, J., & Malm, A. (2018). Illicit Network Dynamics: The Formation and Evolution of a Drug Trafficking Network. *Journal of Quantitative Criminology*, 35(2), 237-258. https://doi.org/10.1007/s10940-018-9379-8.

Bright, D., Brewer, R., & Morselli, C. (2021). Using social network analysis to study crime: Navigating the challenges of criminal justice records. *Social Networks*, 66, 50-64. https://doi.org/10.1016/j.socnet.2021.01.006.

Charette, Y., & Papachristos, A. V. (2017). The network dynamics of co-offending careers. *Social Networks*, 51, 3-13. https://doi.org/10.1016/j.socnet.2016.12.005.

Green, B., Horel, T., & Papachristos, A. V. (2017). Modeling Contagion Through Social Networks to Explain and Predict Gunshot Violence in Chicago, 2006 to 2014. *JAMA Intern Med*, 177(3), 326-333. https://doi.org/10.1001/jamainternmed.2016.8245.

Grund, T. U., & Densley, J. A. (2014). Ethnic Homophily and Triad Closure. *Journal of Contemporary Criminal Justice*, 31(3), 354-370. https://doi.org/10.1177/1043986214553377.

Johns, F. 2021. Governance by Data. *Annual Review of Law and Social Science*, 17(1), 53-71.

Ouellet, M., Hashimi, S., & Vega Yon, G. G. (2022). Officer Networks and Firearm Behaviors: Assessing the Social Transmission of Weapon-Use. *Journal of Quantitative Criminology*, 39(3), 679-703. https://doi.org/10.1007/s10940-022-09546-9.

Papachristos, A. V. (2014). The Network Structure of Crime. *Sociology Compass*, 8(4), 347-357. https://doi.org/10.1111/soc4.12147.

Rostami, A., & Mondani, H. (2015). The complexity of crime network data: a case study of its consequences for crime control and the study of networks. *PLoS One*, 10(3), e0119309. https://doi.org/10.1371/journal.pone.0119309.

Warr, M. (2002). *Companions in crime: The social aspects of criminal conduct*. New York, NY: Cambridge University Press.