1990s Crimes in the "Gateway to the West"

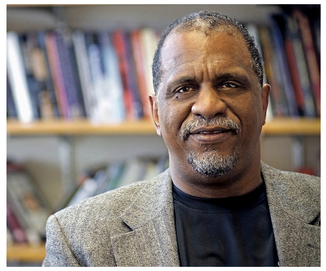
Austin Powell, PhD Candidate | Namrata Kolla, MPA Candidate
University of Washington Industrial & Systems Engineering | Evans School of Public Policy and Administration

RESEARCH QUESTIONS

1. What did this particular crime network look like?
2. Who were the most central individuals and crimes?
3. Can we determine the clustering of crimes [that share at least one individual]?

BACKGROUND

In the 1990s Rick Rosenfeld and Norm White used police records to collect data on crime in St. Louis. They began with 5 homicides and recorded the names of all the individuals who had been involved as victims, suspects or witnesses, and their sex.



Left: Rick Rosenfeld **Above:** Norm White

They explored the records of all other crimes in which those same individuals appeared. This snowball process was continued until they had data on **551 crime events** and **829 unique individuals**. It appears a few more crimes were added outside of the original 5 (and appear as satellite points in our visualizations).

In total there were **1476 involvements**, of which there were:

- 369 victimizations
- 412 suspect accusations
- 189 witnessings
- 40 combined victimizations and suspect accusations

In summary:

- We have a **829 x 551 individual-crime adjacency matrix**
- This is an **undirected bipartite network** where **nodes are crimes or individuals** and **edges are involvements of individuals in crimes**
- The single node attribute is the sex of individuals
- The single edge attribute is the person's involvement as a victim, suspect, witness, or victim-suspect

Course: Statistical Analysis of Social Network Data
Instructor: Dr. Tyler McCormick

June 6, 2019

RQ 1: NETWORK VISUALIZATION

In the circle graph (Figure 1), **nodes for individuals are circles** and **nodes for crimes are squares**. Female individuals are yellow and male individuals are green. A **standard black line for an edge describes a victim**, a **red line describes a suspect**, a **dashed line describes a witness,** and a **bold line describes a dual victim-suspect**.

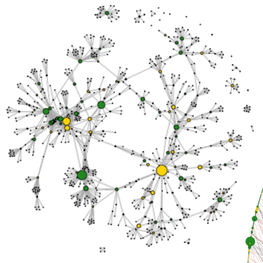


Figure 2. Individual Network, original bipartite network converted to one-mode

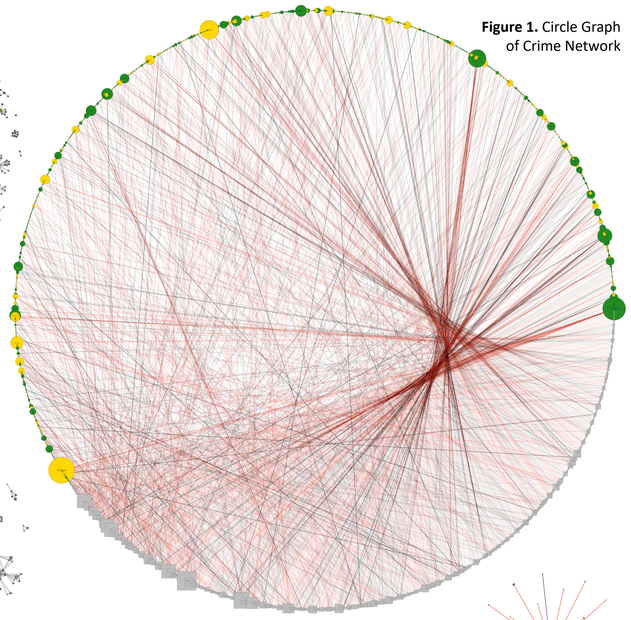


Figure 1. Circle Graph of Crime Network

Figure 3. Crime Network, original bipartite network converted to one-mode

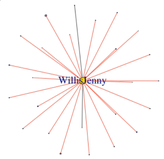


Figure 4. First degree sub-graph of Jenny Willis

Figure 1 suggests that the majority of people are involved in one crime while a few individuals are involved in many crimes, usually as a suspect (i.e. multiple offenders are more likely than multiple witnessings or multiple victimizations). To start looking for potential clustering, we needed to apply methods that are not traditionally used for bipartite networks. Thus, we converted our bipartite graph into two, one-mode graphs by removing intermediate nodes (Figures 2 and 3).

The Individual Network (Figure 2) suggests clustering around a few individuals. These are people who may make good contacts for information on newly committed crimes in St. Louis because they are already connected with many individuals involved in crimes. The Crime Network (Figure 3) suggests that there is clustering in crimes as well. For example, a small group of individuals committing multiple crimes in a similar area would share many suspects and witnesses.

RQ 2: CENTRAL INDIVIDUALS & CRIMES

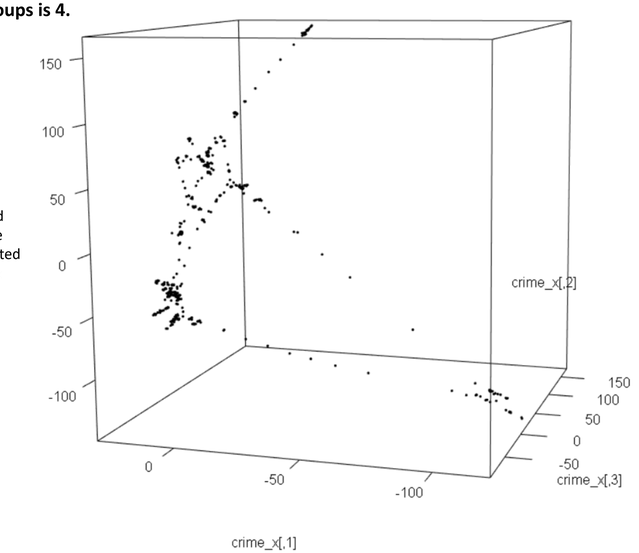
Measure of Centrality	Degree	Closeness	Between-ness	Eigen-centrality
Top 5 Crimes	110, 153, 14, 43, 95	47, 160, 23, 14, 432	110, 23, 160, 47, 46	110, 95, 417, 419, 43
Top 5 Individuals	Willis, Abrams, Katz, Dickson, Bendix	Slattery, Willis Abrams, Steiner, Hemphill	Willis, Slattery, Abrams, Steiner, King	Katz, Steiner, Smith, Johnston, Mitchell

Degree is the most intuitive centrality measure for our problem. The highest degree crimes involved the most people. The highest degree individuals were involved in the most crimes. Often times these high-degree individuals were suspects in multiple crimes (see Figure 4 of Jenny Willis' subgraph).

RQ 3: CLUSTERING CRIMES

While Figure 3 showed that clustering in the one-mode crime network seems to be occurring, we can use statistical methods to confirm that clusters exist and what the optimal number of groups are. **We apply Variational Bayesian Inference on a Latent Position Cluster Model** (VBLPCM) on the largest connected crime involvement network and find that **the optimal number of groups is 4.**

Figure 5. Fitted VBLPCM of the largest connected crime network



In order to choose the optimal number of clusters, we calculate the Bayes information criterion (BIC) for each number of groups and choose the number of groups with the lowest BIC value. We calculate the latent positions of each node in **three dimensions** and use **10 steps in the maximum likelihood estimation**. As shown in the table below, **four groups is the optimal number.** This means that in the crime network, there are four clusters of crimes that share many suspects, witnesses, and victims with each other, but they have relatively few individuals in common with crimes in the other clusters.

Clusters	2	3	4	5
BIC	189,146	31,359	**29,359**	44,169

FUTURE WORK

Including attributes for crimes (e.g. the types of crimes, where they occurred) would take this analysis to the next level by evaluating the driving factors of why crimes cluster together and why certain individuals are involved in clustered crimes. It would also be interesting to use edge attributes to isolate the suspect network and determine whether individuals segregate by sex. Figure 2 suggests that women are most likely to be involved in crimes with other women while men are more likely to be involved with crimes with men and women, but this cannot be proven without further statistical analysis.

Data Source:

"ROSENFELD,WHITE--ST. LOUIS CRIME." University of California Irvine, School of Social Sciences. Accessed 23 May 2019. Link: http://moreno.ss.uci.edu/data.html