February 2017 Chicago Crime Network Analysis

CSC 495 Final Project

Kari Palmier, Michelle Tang

Table of Contents

I. Introduction

This project studies how the different neighborhoods in Chicago are connected to each other through crime. All, violent, and non-violent crime were studied to see what effect crime type had on neighborhodd networks. Bipartite and similarity projections also were conducted, which found what neighborhoods had the the most crimes in common with others (highest number of crimes) and which neighborhoods had the most similar types and numbers of crimes as others.

II. Data Description

The initial data was the 2017 Chicago crime dataset and the 2010 Chicago Census data, which was sourced the City of Chicago website. The crime dataset was used to generate the edge list; connecting neighborhoods to crimes. The node data was created from the census data. The neighborhood attributes used were neighborhood geo IDs, names, total population, average household size, proportion renter, proportion owned, proportion occupied, proportion vacant, proportion African American, proportion Asian, proportion White, proportion Hispanic, and median age. The crime node attributes were description and crime type (violent or non-violent). The violent crimes present were criminal sexual assault, battery, human trafficking, weapons violation, assault, sex offense, homicide, kidnapping, robbery, arson, offense involving children, intimidation, interference with public officer, and stalking. The non-violent crimes present were deceptive practice, theft, motor vehicle theft, public peace violation, liquor law violation, obscenity, criminal damage, narcotics, concealed carry license violation, public indecency, burglary, prostitution, gambling, and criminal trespass. Three different datasets were created: all crime (violent and non-violent), violent crime, and non-violent crime. The month of February was chosen to use for analysis based on the number of crimes present, as well as the types of crimes (it was the month with the lowest crime that still had the majority of crimes occuring). There were 6,356 crimes in the violent dataset and 11,445 in the non-violent dataset (17,801 crimes total).

Bipartite and similarity neighborhood projections were created on all three datasets (all, violent, and non-violent crimes). Weighted degree centrality was used because the analysis focused on what neighborhoods are immediately connected to each other through crime. The concepts of betweenness and closeness were examined but were hard to define (a neighborhood close to others has any type of influence from a crime standpoint was hard to determine, as was a neighborhood being between others).

All bipartite and similarity neighborhood projections for all crime types were very dense (density value of 1 and global transitivity of 1), whichmeans that all of the edges possible were already present in the data. This caused issues with assortativity calculations, as well as graph visualizations. In order to focus on the neighborhoods of interest (higher weighted degree), the bipartite projections were all filtered such that 95% of the edges were removed and similarity projections were filtered such that 90% of the edges were removed. Several different filtering levels were attempted (results for all filtering levels are in the appendix). The final levels were chosen because they generated giant components that captured at least 80% of the nodes, had acceptable transitivity and assortativity values, and were able to be successfully visualized in Gephi. After edge filtering, all singleton nodes were removed. The networks were then decomposed and the giant components were extracted and used for the rest of the analysis. Assortativity was calculated for all node attributes, along with the assortativity CUG and QAP tests. Local and global transitivity were also calculated, along with the CUG tests. The results of local transitivity were not used for analysis due to the presence of NaN values at some nodes. Finally, modularity analysis was performed and modularity clusterings were generated using the edge betweenness, fast greedy, leading eigenvector, Louvain, spinglass, and walktrap with 3 steps, 5 steps, and 7 steps. The modularity clustering algorithms used in visualizations were chosen based on modularity value and how the clusters divided the nodes (good clusterings had nodes of a cluster all close to each other on the visualizations). The descriptive statistics for the violent and non-violent bipartite projections are given in the tables at the end of this section.

Bipartite Analysis

Below are the descriptive statistics for the violent and non-violent bipartite neighborhood projections.

Violent Bipartite Neighborhood Projection Summary												
Percent Edges Filtered Out	Number Initial Nodes	Number Initial Edges	Initial Min Edge Weight	Initial Median Edge Weight	Initial Max Edge Weight	Initial Density	Initial Weighted Degree Min	Initial Weighted Degree Median	Initial Weighted Degree Max	Initial Number of Components		
95	77	2926	6	1070.5	44381	1	9015	116687	940526	1		
Edge Filtering Threshold	Number Final Nodes	Number Final Edges	Final Min Edge Weight	Final Median Edge Weight	Final Max Edge Weight	Final Density	Final Weighted Degree Min	Final Weighted Degree Median	Final Weighted Degree Max	Final Number of Components	Number Final Largest Component Nodes	Number Final Largest Component Edges
8950.25	36	147	8958	12384	44381	0.233333	9535	52938	74870	1	36	147
Global Transitivity												
	CUG Pr >=	CUG Pr <=										
0.556591	0	1										

Table 1: Violent crime bipartite summary

Violent Bipartite Neighborhood Projection Assortativity														
95 Percent Edges Filtered Out														
Prop Occupied					Prop Rented					Prop Vacant				
	CUG Pr >=	CUG Pr <=	QAP p >=	QAP p <=		CUG Pr >=	CUG Pr <=	QAP p >=	QAP p <=		CUG Pr >=	CUG Pr <=	QAP p >=	QAP p <=
-0.0046	0.355	0.645	0.133	0.867	0.0222	0.238	0.762	0.058	0.942	-0.0046	0.366	0.634	0.141	0.859
Prop Owned					Median Age					Total Population				
	CUG Pr >=	CUG Pr <=	QAP p >=	QAP p <=		CUG Pr >=	CUG Pr <=	QAP p >=	QAP p <=		CUG Pr >=	CUG Pr <=	QAP p >=	QAP p <=
0.0222	0.231	0.769	0.056	0.944	-0.1019	0.849	0.151	0.821	0.179	-0.0704	0.693	0.307	0.632	0.368
Avg Household Size					Prop African					Prop White				
	CUG Pr >=	CUG Pr <=	QAP p >=	QAP p <=		CUG Pr >=	CUG Pr <=	QAP p >=	QAP p <=		CUG Pr >=	CUG Pr <=	QAP p >=	QAP p <=
-0.0219	0.427	0.573	0.218	0.782	-0.1651	0.961	0.039	0.960	0.040	-0.0924	0.812	0.188	0.791	0.209
Prop Asian					Prop Hispanic									
	CUG Pr >=	CUG Pr <=	QAP p >=	QAP p <=		CUG Pr >=	CUG Pr <=	QAP p >=	QAP p <=					
-0.1029	0.858	0.142	0.840	0.160	-0.1293	0.926	0.074	0.930	0.070					

Table 2: Violent crime bipartite assortativity

Non-Violent Bipartite Neighborhood Projection Summary												
Percent Edges Filtered Out	Number Initial Nodes	Number Initial Edges	Initial Min Edge Weight	Initial Median Edge Weight	Initial Max Edge Weight	Initial Density	Initial Weighted Degree Min	Initial Weighted Degree Median	Initial Weighted Degree Max	Initial Number of Components		
95	77	2926	41	2272	138304	1	22468	272209	1696784	1		
Edge Filtering Threshold	Number Final Nodes	Number Final Edges	Final Min Edge Weight	Final Median Edge Weight	Final Max Edge Weight	Final Density	Final Weighted Degree Min	Final Weighted Degree Median	Final Weighted Degree Max	Final Number of Components	Number Final Largest Component Nodes	Number Final Largest Component Edges
17100	38	147	17114	24488	138304	0.209104	17114	108029	1319977	1	38	147

Global Transitivity		
	CUG Pr >=	CUG Pr <=
0.389332	0	1

Table 3: Non-Violent crime bipartite summary

Non-Violent Bipartite Neighborhood Projection Assortativity														
95 Percent Edges Filtered Out														
Prop Occupied					Prop Rented					Prop Vacant				
	CUG Pr >=	CUG Pr <=	QAP p >=	QAP p <=		CUG Pr >=	CUG Pr <=	QAP p >=	QAP p <=		CUG Pr >=	CUG Pr <=	QAP p >=	QAP p <=
-0.0265	0.467	0.533	0.294	0.706	-0.0564	0.630	0.370	0.503	0.497	-0.0265	0.484	0.516	0.267	0.733
Prop Owned					Median Age					Total Population				
	CUG Pr >=	CUG Pr <=	QAP p >=	QAP p <=		CUG Pr >=	CUG Pr <=	QAP p >=	QAP p <=		CUG Pr >=	CUG Pr <=	QAP p >=	QAP p <=
-0.0031	0.328	0.672	0.139	0.861	-0.0874	0.785	0.215	0.718	0.282	-0.1151	0.882	0.118	0.835	0.165
Avg Household Size					Prop African					Prop White				
	CUG Pr >=	CUG Pr <=	QAP p >=	QAP p <=		CUG Pr >=	CUG Pr <=	QAP p >=	QAP p <=		CUG Pr >=	CUG Pr <=	QAP p >=	QAP p <=
-0.1974	0.996	0.004	0.965	0.035	-0.0548	0.612	0.388	0.503	0.497	-0.1919	0.990	0.010	0.954	0.046
Prop Asian					Prop Hispanic									
	CUG Pr >=	CUG Pr <=	QAP p >=	QAP p <=		CUG Pr >=	CUG Pr <=	QAP p >=	QAP p <=					
-0.0421	0.580	0.420	0.398	0.602	-0.1294	0.935	0.065	0.847	0.153					

Table 4: Non-Violent crime bipartite assortativity

Below are the initial and filtered edge weights for both the violent and non-violent bipartite networks. The lower value edge weights were filtered out, leaving the higher edge weights in the final networks. Note that the violent data had lower edge weights than the non-violent. This is due to the number of crimes in the non-violent data being nearly twice the number in the violent data.

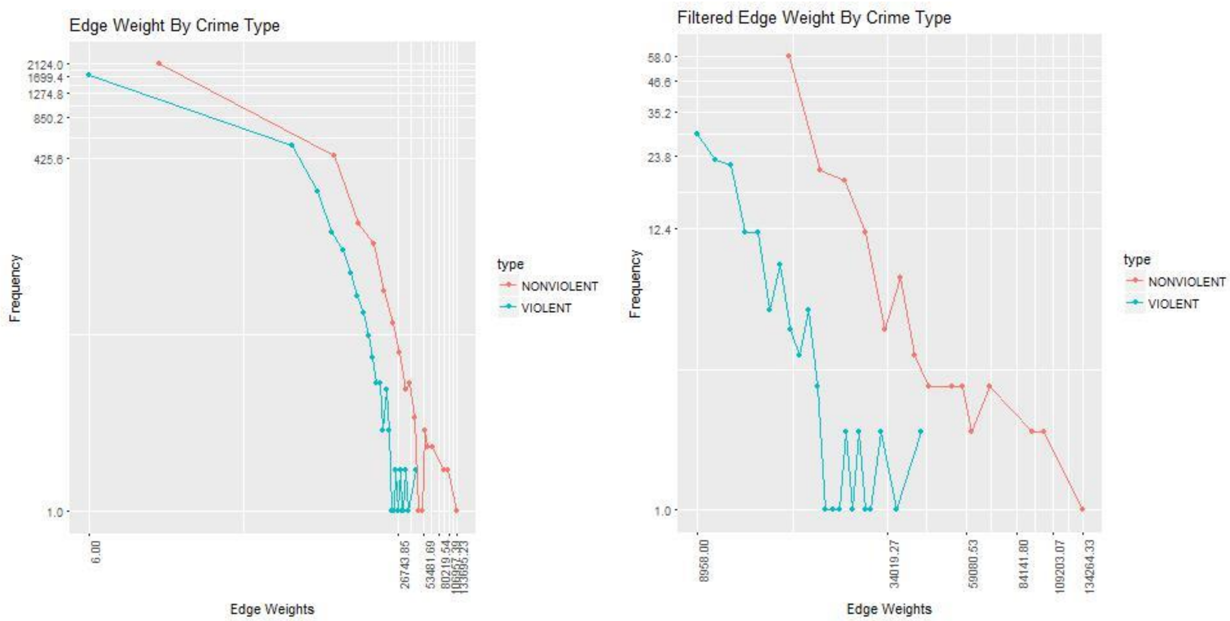

Figure 1: Edge weight filtering by Crime Type

Below are the initial and filtered weighted degrees for both the violent and non-violent bipartite networks.

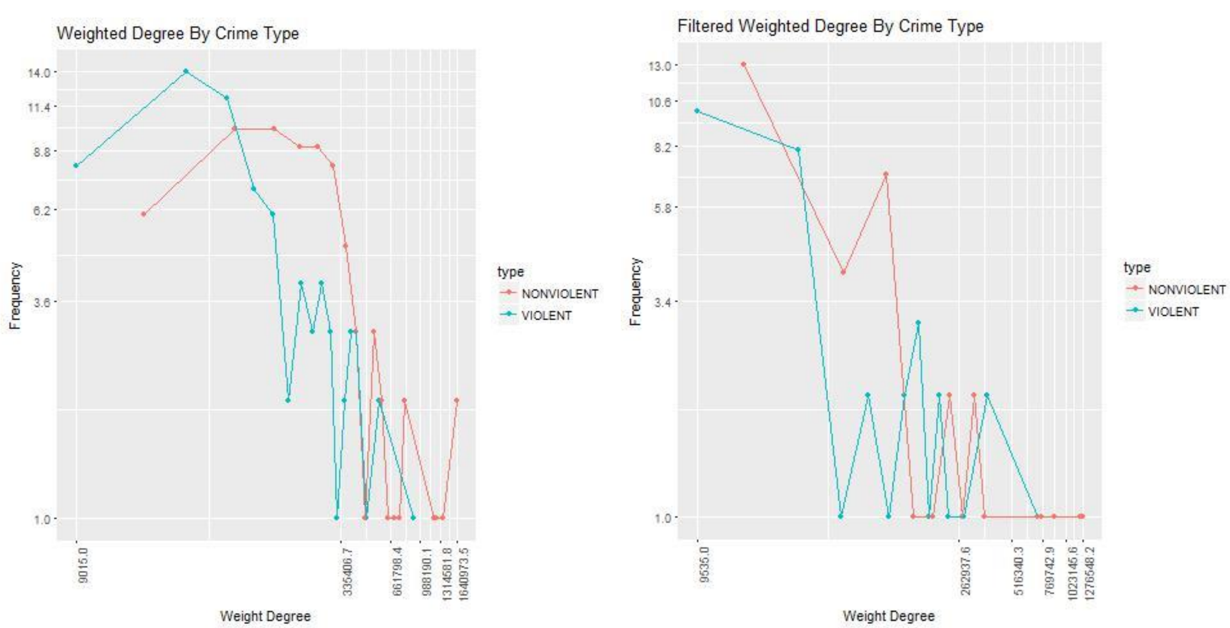

Figure 2: Weighted degree by Crime Type

The modularity results for the violent and non-violent bipartite neighborhood projections are displayed below. The Louvain clustering was chosen for the violent projection and the fast greedy was chosen for the non-violent projection.

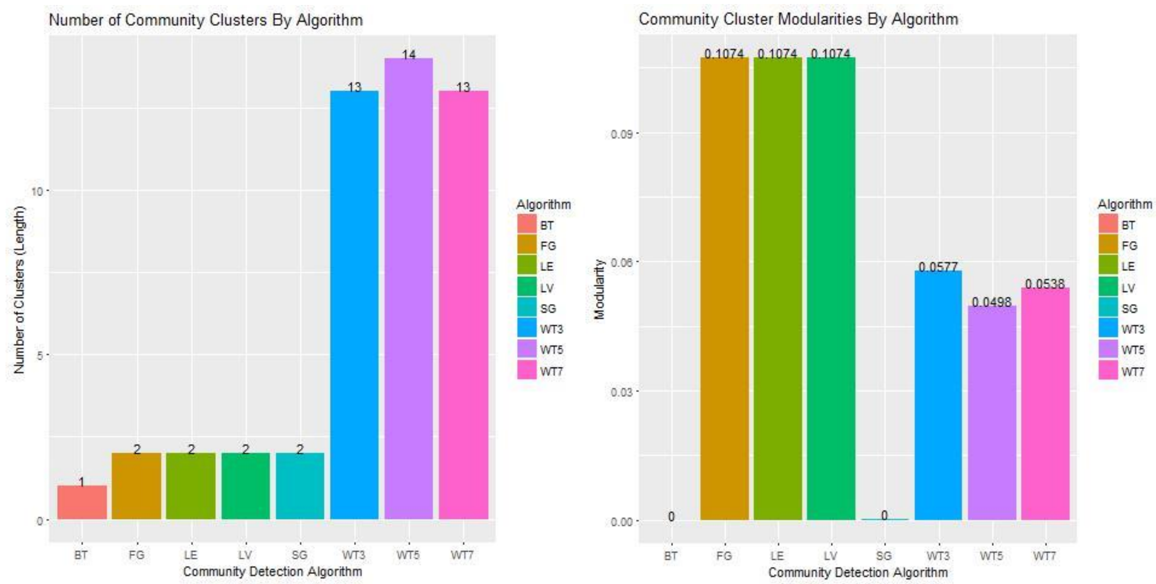

Figure 3: Violent crime modularity

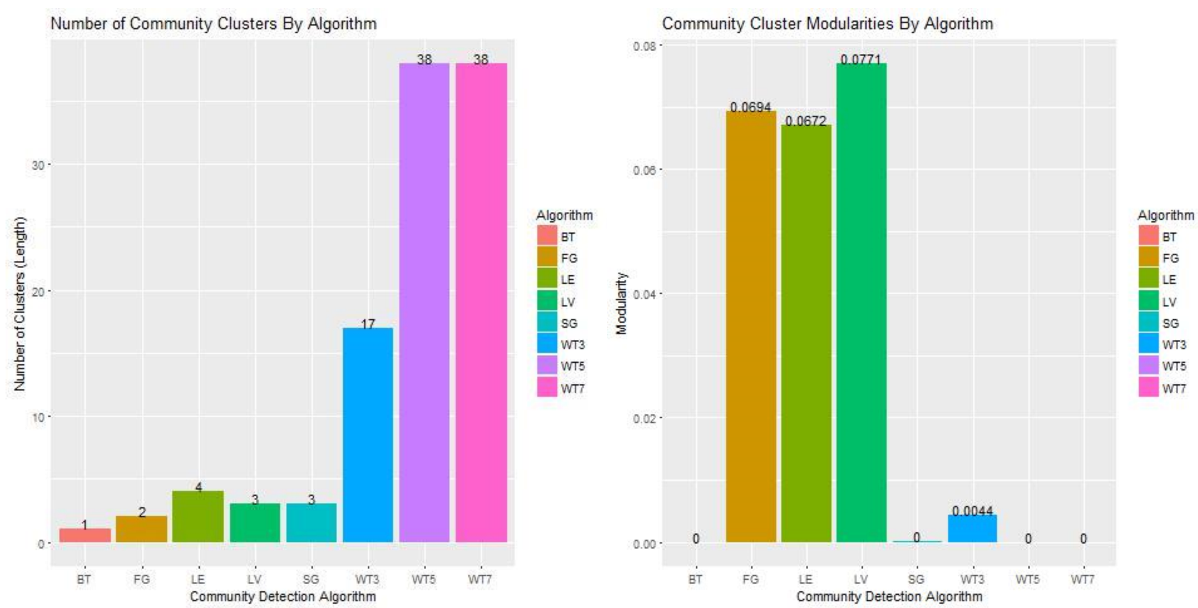

Figure 4: Non-violent crime modularity

Similiarity Analysis:

Below are the descriptive statistics for the violent and non-violent similarity neighborhood projections.

Violent Similarity Neighborhood Projection Summary													
Percent Edges Filtered Out	Number Initial Nodes	Number Initial Edges	Initial Min Edge Weight	Initial Median Edge Weight	Initial Max Edge Weight	Initial Density	Initial Weighted Degree Min	Initial Weighted Degree Median	Initial Weighted Degree Max	Initial Number of Components			
90	77	3003	0.6896	0.9659	1	1.026316	65.75	75	75.86	1			
Edge Filtering Threshold	Number Final Nodes	Number Final Edges	Final Min Edge Weight	Final Median Edge Weight	Final Max Edge Weight	Final Density	Final Weighted Degree Min	Final Weighted Degree Median	Final Weighted Degree Max	Final Number of Components	Number Final Largest Component Nodes	Number Final Largest Component Edges	Percent of Nodes in Largest Component
0.991939	53	224	0.9919	0.9945	0.9995	0.162554	0.9919	5.9677	24.8709	3	49	222	0.9245
Global Transitivity													
	CUG Pr >=	CUG Pr <=											
0.548257	0.000	1.000											

Table 5: Violent crime **s**imilarity analysis summary

Violent Similarity Neighborhood Projection Assortativity														
90 Percent Edges Filtered Out														
Prop Occupied					Prop Rented					Prop Vacant				
	CUG Pr >=	CUG Pr <=	QAP p >=	QAP p <=		CUG Pr >=	CUG Pr <=	QAP p >=	QAP p <=		CUG Pr >=	CUG Pr <=	QAP p >=	QAP p <=
0.0442	0.122	0.878	0.069	0.931	-0.0396	0.598	0.402	0.534	0.466	0.0442	0.121	0.879	0.076	0.924
Prop Owned					Median Age					Total Population				
	CUG Pr >=	CUG Pr <=	QAP p >=	QAP p <=		CUG Pr >=	CUG Pr <=	QAP p >=	QAP p <=		CUG Pr >=	CUG Pr <=	QAP p >=	QAP p <=
-0.0094	0.445	0.555	0.333	0.667	0.0360	0.170	0.830	0.129	0.871	-0.0572	0.734	0.266	0.680	0.320
Avg Household Size					Prop African					Prop White				
	CUG Pr >=	CUG Pr <=	QAP p >=	QAP p <=		CUG Pr >=	CUG Pr <=	QAP p >=	QAP p <=		CUG Pr >=	CUG Pr <=	QAP p >=	QAP p <=
0.2223	0.000	1.000	0.000	1.000	0.1052	0.019	0.981	0.013	0.987	0.2223	0.000	1.000	0.000	1.000
Prop Asian					Prop Hispanic									
	CUG Pr >=	CUG Pr <=	QAP p >=	QAP p <=		CUG Pr >=	CUG Pr <=	QAP p >=	QAP p <=					
0.1260	0.005	0.995	0.008	0.992	0.1045	0.019	0.981	0.013	0.987					

Table 6: Violent crime similarity assortativity summary

Non-Violent Similarity Neighborhood Projection Summary													
Percent Edges Filtered Out	Number Initial Nodes	Number Initial Edges	Initial Min Edge Weight	Initial Median Edge Weight	Initial Max Edge Weight	Initial Density	Initial Weighted Degree Min	Initial Weighted Degree Median	Initial Weighted Degree Max	Initial Number of Components			
90	77	3003	0.6896	0.9283	1	1.02631	45.22	69.98	72.63	1			
Edge Filtering Threshold	Number Final Nodes	Number Final Edges	Final Min Edge Weight	Final Median Edge Weight	Final Max Edge Weight	Final Density	Final Weighted Degree Min	Final Weighted Degree Median	Final Weighted Degree Max	Final Number of Components	Number Final Largest Component Nodes	Number Final Largest Component Edges	Percent of Nodes in Largest Component
0.975257	65	224	0.9753	0.9824	0.9976	0.107692	0.9781	6.8666	18.6849	3	61	222	0.9385

Global Transitivity		
	CUG Pr >=	CUG Pr <=
0.486961	0.000	1.000

Table 7: Non-Violent crime similarity analysis summary

Non-Violent Similarity Neighborhood Projection Assortativity														
90 Percent Edges Filtered Out														
Prop Occupied					Prop Rented					Prop Vacant				
	CUG Pr >=	CUG Pr <=	QAP p >=	QAP p <=		CUG Pr >=	CUG Pr <=	QAP p >=	QAP p <=		CUG Pr >=	CUG Pr <=	QAP p >=	QAP p <=
0.1649	0.004	0.996	0.005	0.995	0.0374	0.185	0.815	0.142	0.858	0.1649	0.003	0.997	0.001	0.999
Prop Owned					Median Age					Total Population				
	CUG Pr >=	CUG Pr <=	QAP p >=	QAP p <=		CUG Pr >=	CUG Pr <=	QAP p >=	QAP p <=		CUG Pr >=	CUG Pr <=	QAP p >=	QAP p <=
0.0353	0.204	0.796	0.170	0.830	-0.0165	0.482	0.518	0.429	0.571	-0.0726	0.809	0.191	0.806	0.194
Avg Household Size					Prop African					Prop White				
	CUG Pr >=	CUG Pr <=	QAP p >=	QAP p <=		CUG Pr >=	CUG Pr <=	QAP p >=	QAP p <=		CUG Pr >=	CUG Pr <=	QAP p >=	QAP p <=
0.5257	0.000	1.000	0.000	1.000	0.1558	0.002	0.998	0.009	0.991	0.4722	0.000	1.000	0.000	1.000
Prop Asian					Prop Hispanic									
	CUG Pr >=	CUG Pr <=	QAP p >=	QAP p <=		CUG Pr >=	CUG Pr <=	QAP p >=	QAP p <=					
0.0342	0.150	0.850	0.130	0.870	0.2597	0.000	1.000	0.000	1.000					

Table 8: Non-Violent crime similarity **a**ssortativity summary

The initial and filtered edge weights for both the violent and non-violent similarity networks are shown below. The lower value edge weights were filtered out, leaving the higher edge weights in the final networks. This indicates more similar neighborhoods are present after filtering since the edge weights are the cosine similarity of the crimes of the connecting neighborhoods. Note the edge weights of the violent data are higher than those of the non-violent data, which implies there are more similar crimes present between neighborhoods in the violent data.

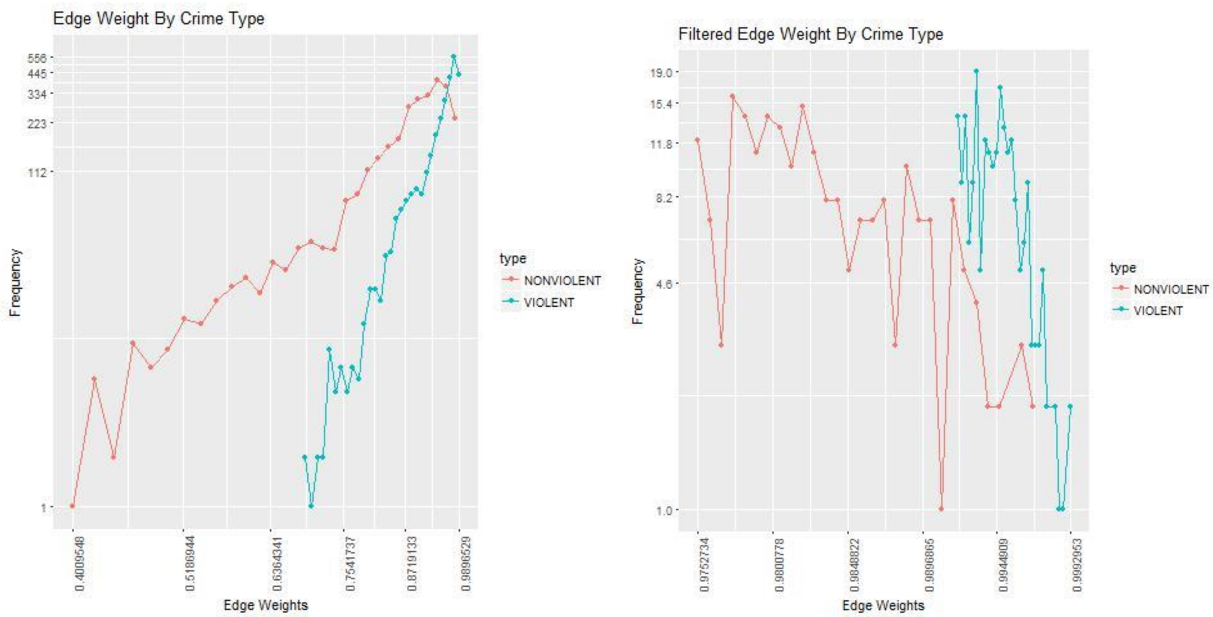

Figure 5: Edge weight filtering by Crime Type

Below are the initial and filtered weighted degrees for both the violent and non-violent similarity networks.

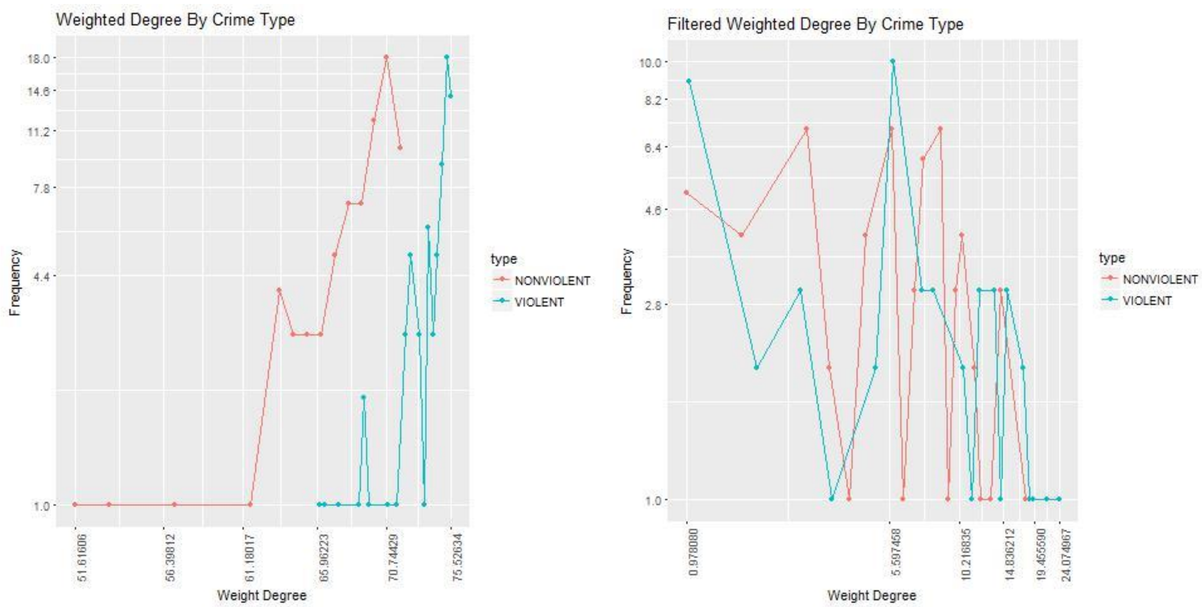

Figure 6: Weighted degree by Crime Type

The modularity results are shown below for the violent and non-violent similarity neighborhood projections. The Louvain clustering was chosen for both the violent and non-violent projections.

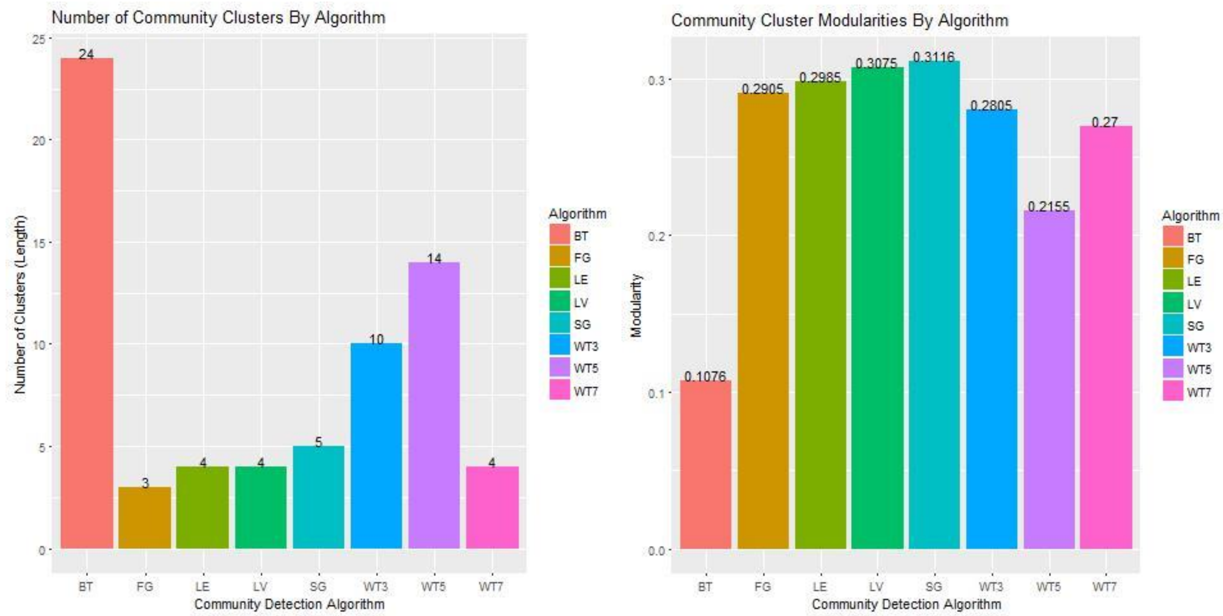

Figure 7: Violent crime modularity

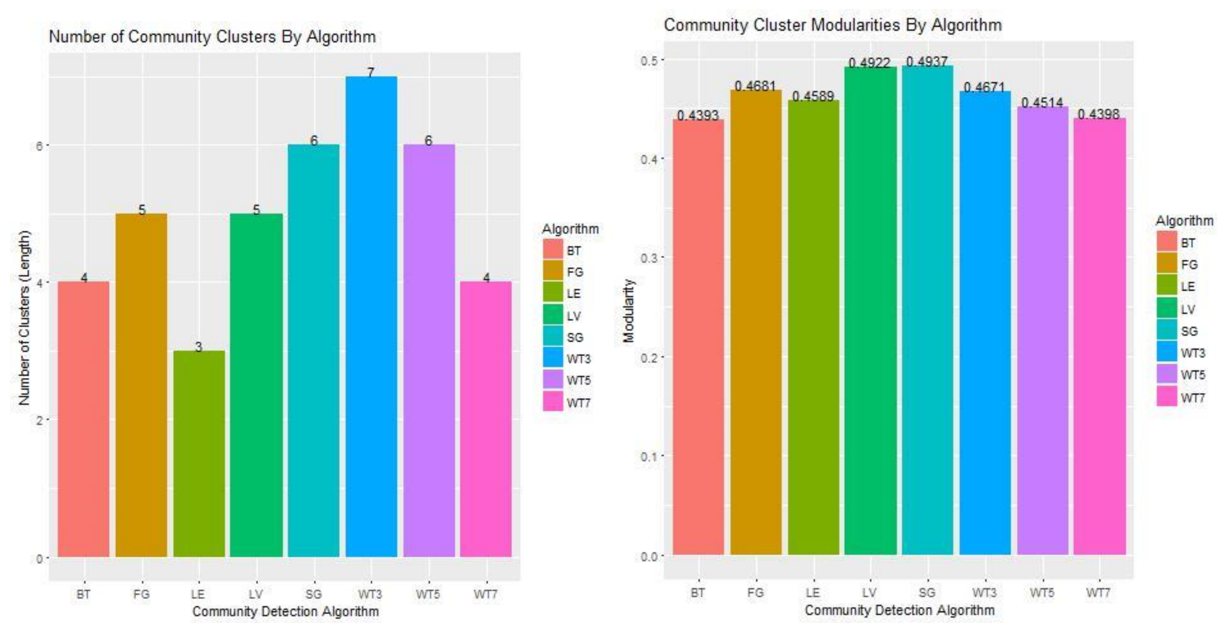

Figure 8: Non-violent crime modularity

III. Analysis

The goal of this analysis is to examine the effects of different types of crime on the neighborhood projections. The neighborhood bipartite projections contain edges that are the number of crimes in common between neighborhoods. Higher weighted degree nodes in the bipartite projections correspond to neighborhoods with higher numbers of crimes (higher weighted degree indicates the node is connected to more neighborhoods through crimes). The neighborhood similarity projections contained edges that connect neighborhoods with similar types and numbers of crimes together. Higher weighted degree nodes in the similarity projections correspond to neighborhoods have the most similar crimes (number and types) as other neighborhoods. Note that the plots on the left (or top) of all visualizations depict weighted degree centrality with both color and node size. The plots on the right (or bottom) depict modularity with color and weighted degree centrality with node size.

The neighborhood bipartite projection of violent crime is depicted in Figure 9. The weighted degree plot on the left shows the Austin, North Lawndale, South Shore, Auburn-Gresham, and Near North Side neighborhoods have the highest number of violent crimes. Higher degree nodes are more interconnected than lower degree ones. The modularity plot on the right shows two clusters: blue/green cluster is made up of higher degree nodes that are all interconnected (dense) and the pink cluster is primarily made up of lower degree nodes that are only connected to the other (Austin on most cases). Figure 10 depicts the non-violent neighborhood bipartite projection. The weighted degree plot on the left shows that Near North Side, the Loop, West Town, Near West Side, and Austin have the highest non-violent crime. Most of these neighborhoods are more affluent (with the exception of Austin). The non-violent modularity clusters in the right plot appear to be divided primarily by geographic locations. The blue/green cluster consists of neighborhoods that are mostly on the west and south sides, where the pink cluster is mostly the north side. Figure 11 describes the bipartite projection for all types of crime. The highest weighted degree nodes and modularity clusterings are very similar to the non-violent projection. This is because the number of non-violent crimes is almost twice the number of violent crimes (non-violent outweighs the violent and takes over the network structure).

The neighborhood similarity projection of violent crime is depicted in Figure 12. The left plot in shows that the highest weighted degree nodes are Chicago Lawn, Austin, Englewood, Grand Boulevard, and South Shore. The modularity plot on the right shows clusters that are separated based on similar crime types and counts. All clusters have high battery and assault crimes. The blue cluster also has high criminal sexual assault, weapons violation, offense involving children, and interference with public officer occurrences. The orange cluster has the same as blue and high sexual offenses. The pink has high criminal sexual assult, weapons violation, and offense involving children. The similarity projection of non-violent crime is depicted in Figure 13. The neighborhoods with highest weighted degree are Chicago Lawn, Near West Side, South Shore, Archer Heights, and Garfield. The modularity plot on the bottom also shows clusters that are separated based on similar crime types and counts. All of the clusters have high theft, criminal damage, and burglary. The blue cluster also has high deceptive practice and criminal trespassing. The teal cluser has the same as blue and high motor vehicle theft and narcotics. The green cluster high theft, high deceptive practice, and motor vehicle theft. Figure 14 depicts the similarity projection for all crime types. The neighorhoods with the highest weighted degree are the same as the non-violent similarity projection. The modularity plot (bottom plot) shows clusters that appear to be divided by economic satus. The blue/green cluster is made up of mostly affluent neighborhoods, where orange is more middle class and purple is lower.

The similarity projections showed that there was assortativity present in the networks. For violent crime, average household size, proportion African American, proportion Asian, proportion White, and proportion Hispanic were all positively assortative (refer to Table 6). For non-violent crime, proportion occupied, proportion vacant, average household size, proportion African American, proportion White, and proportion Hispanic were all positively assortative (refer to Table 8). The average household size and proportion White were the most assortative attributes for violent and non-violent crime. All CUG test results for attributes listed showed assortativity values that were all significantly different than from a random graph. All QAP test results for attributes listed showed assortativity was dependent on specific node placement of the original network.

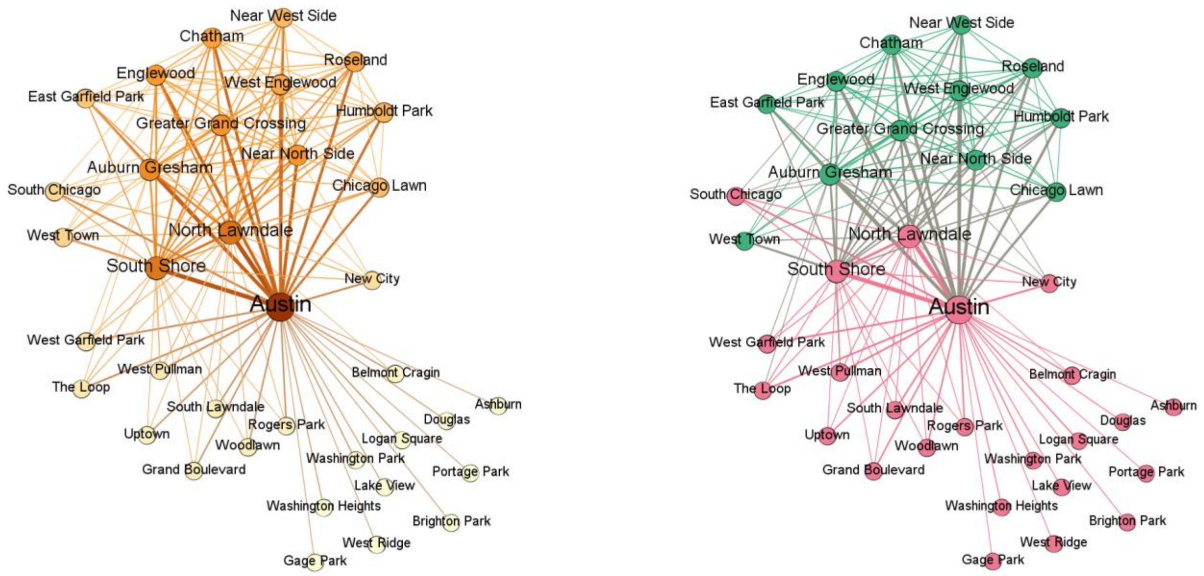

Figure 9: Violent Bipartite Weighted Degree (left) and Louvain Modularity (right)

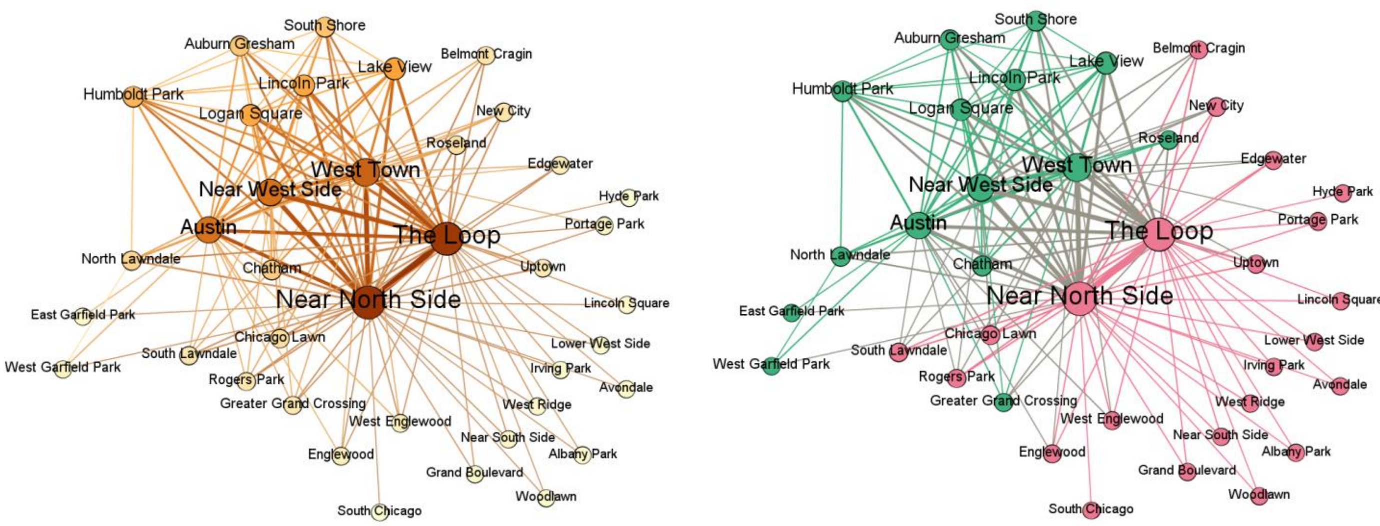

Figure 10: Non-Violent Bipartite Weighted Degree (left) and Fast Greedy Modularity (right)

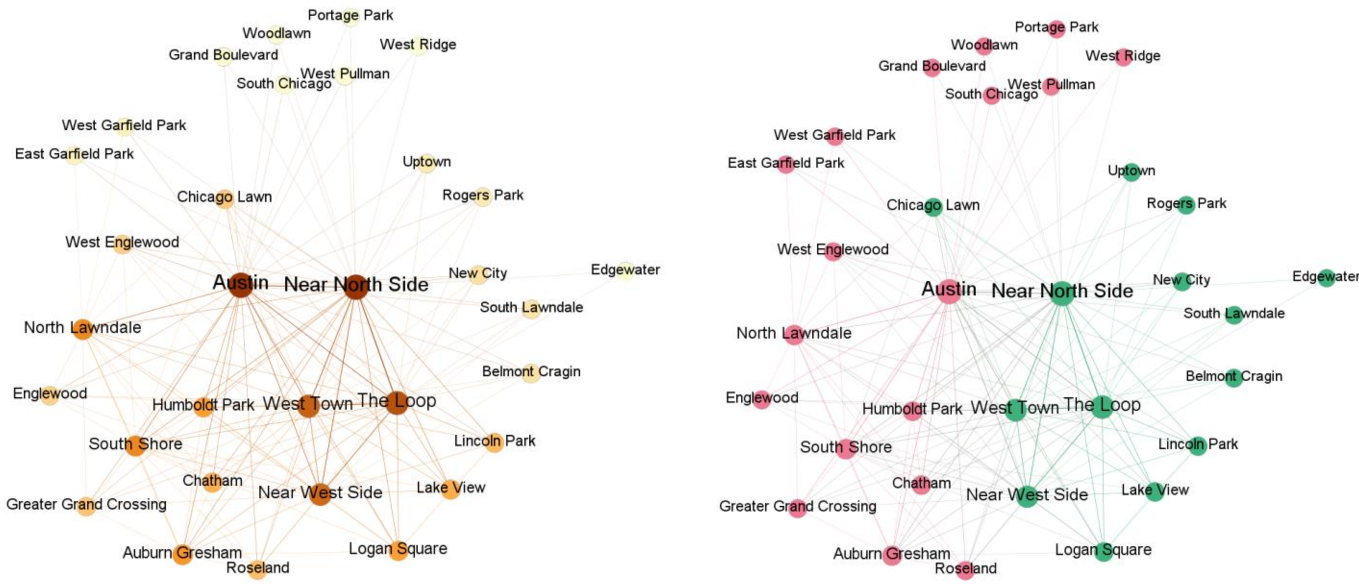

Figure 11: All Crime Bipartite Weighted Degree (left) and Louvain Modularity (right)

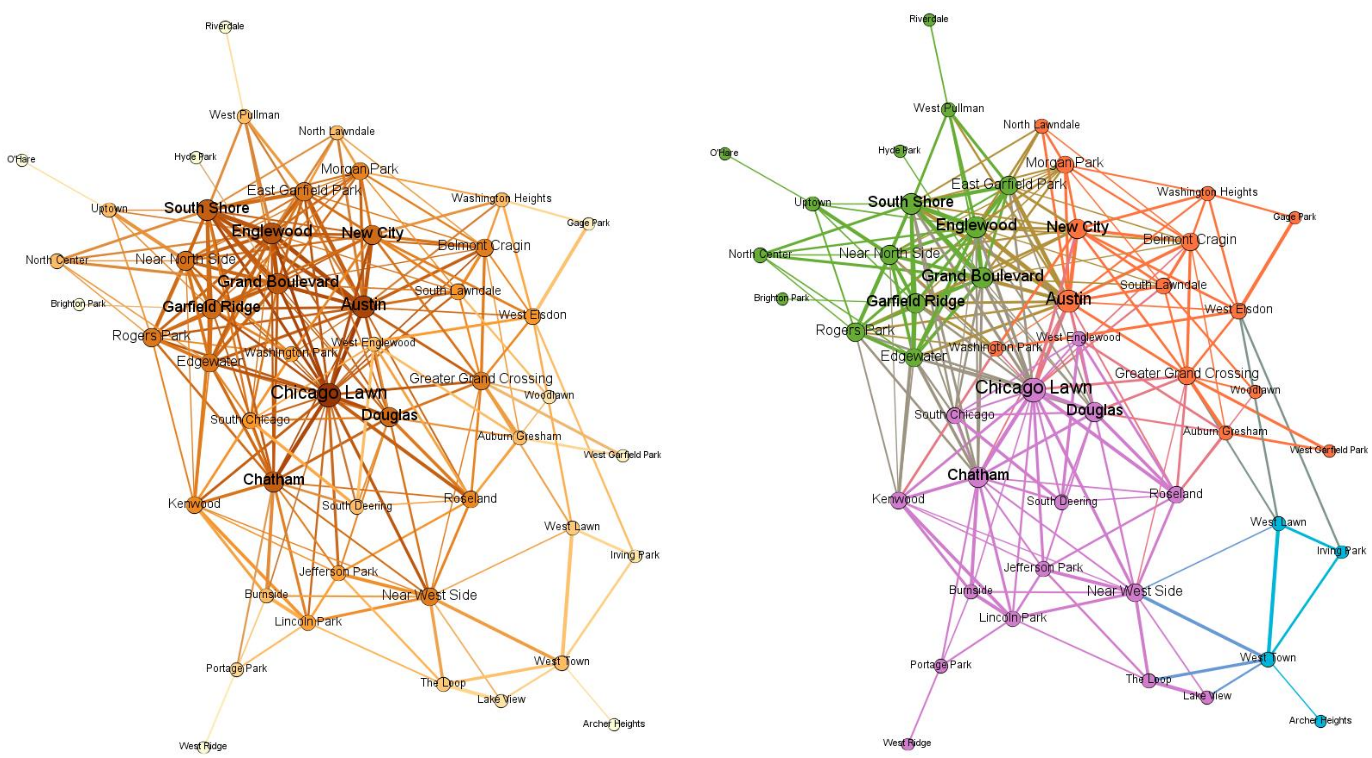

Figure 12: Violent Similarity Weighted Degree (left) and Louvain Modularity (right)

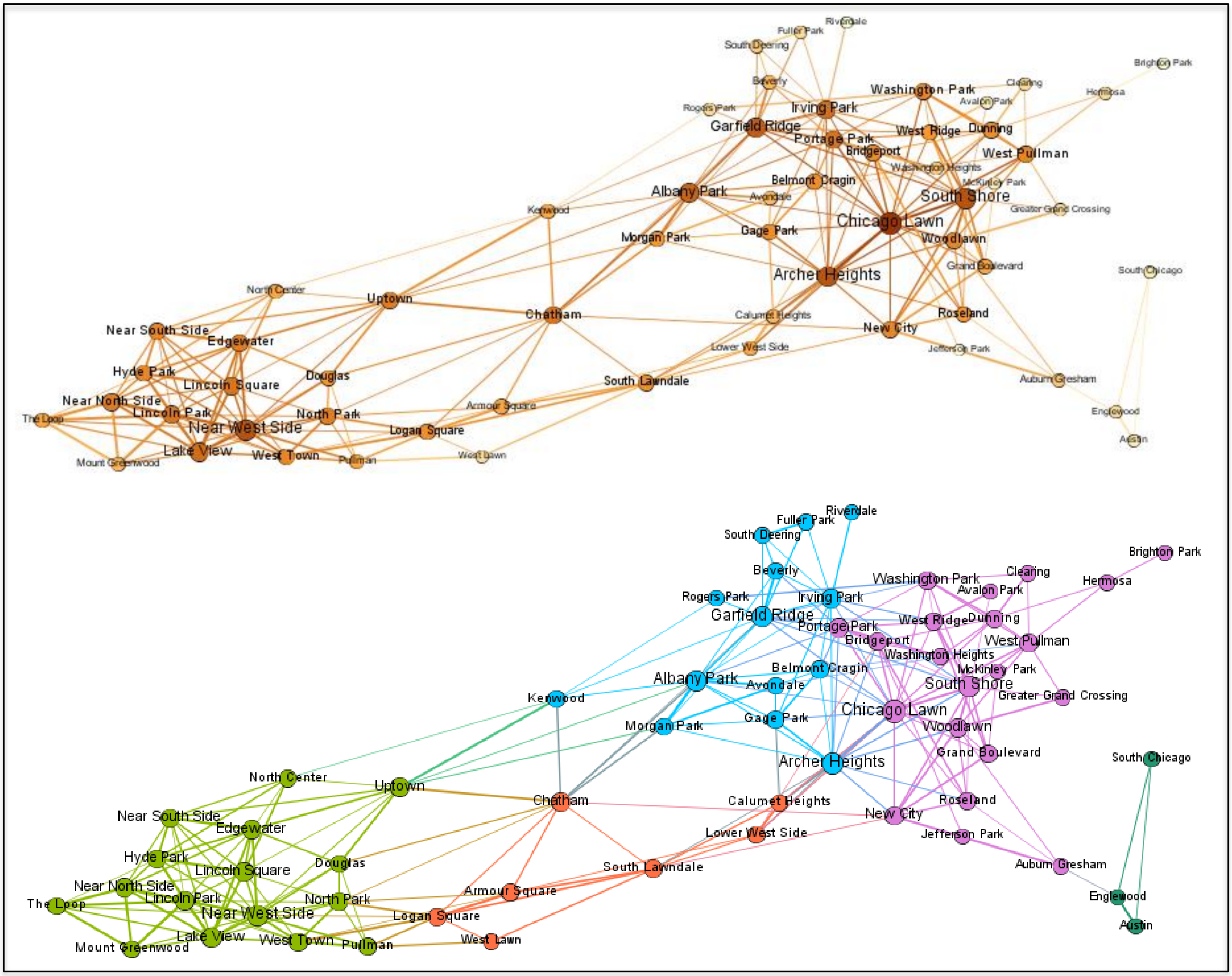

Figure 13: Non-Violent Similarity Weighted Degree (top) and Louvain Modularity (bottom)

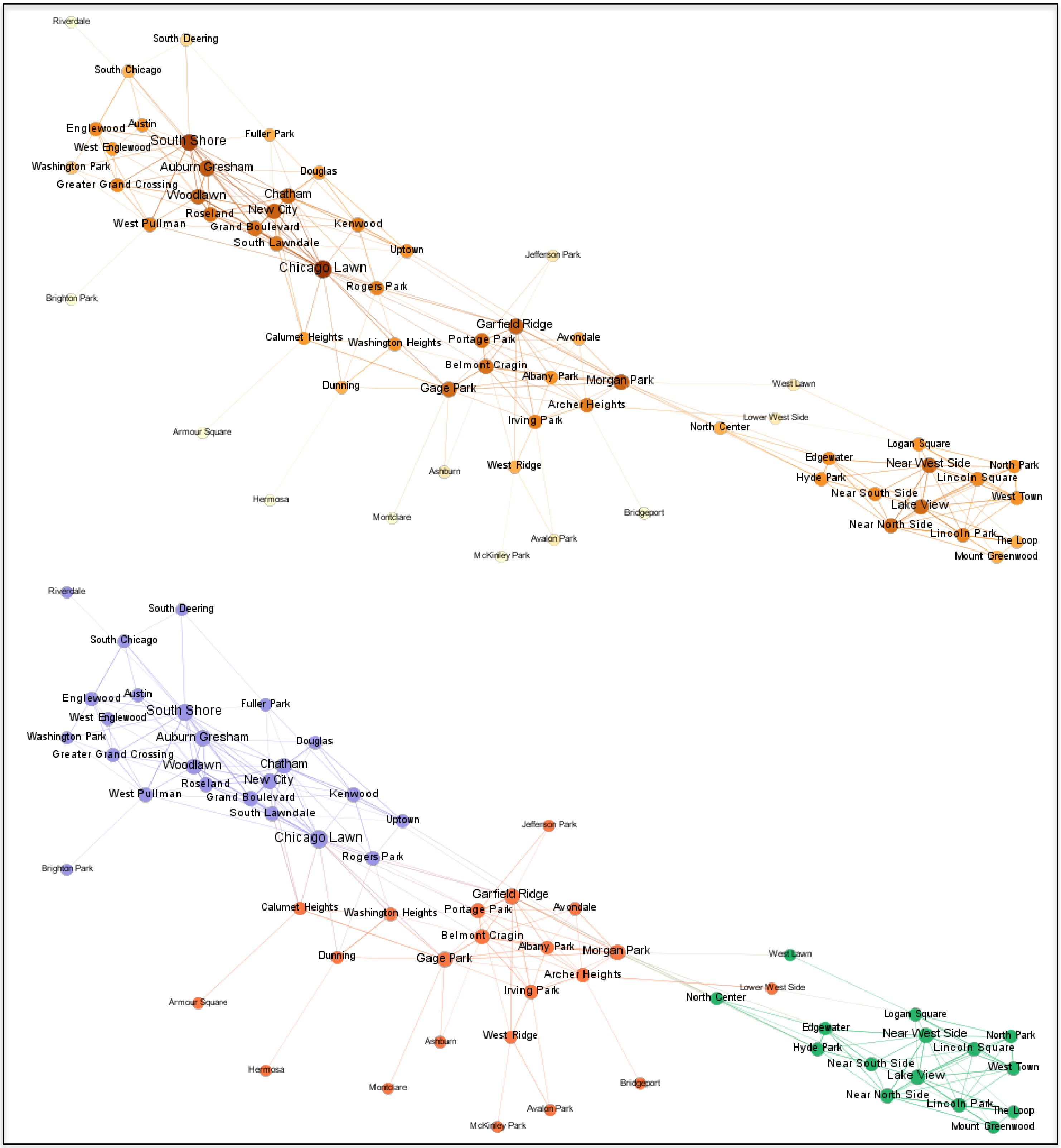

Figure 14: All Simliarity Weighted Degree (top) and Louvain Modularity (bottom)

IV. Conclusion

This analysis showed that the types of crime did have an impact on the structure of the neighborhood networks. The bipartite analysis showed that neighborhoods with higher non-violent crime were more affluent areas, where higher violent crime were on the south and west sides. Theft was the most common non-violent crime, so it makes sense that more expensive areas would have higher non-violent crime. Similarity analysis showed that neighborhoods with similar types and numbers of crime appear to be divided along economic status (upper, middle, and lower classes). This is based on the proportion owned, proportion occupied, and proportion vacant for these areas along with knowledge of the neighborhoods. Neighborhood income data is needed to confirm this hypothesis (is not part of the census data). Similarity projections were shown to be assortative along average household size and proportion of white.

Areas of analysis that would be interesting but were unable to be performed due to time are how the neighborhood networks change throughout the year (summer versus winter, etc) and modeling the similarity projections using ERGM to determine what assortative attributes contribute to determining the network the most.

One limitation of the data is that the node attributes were taken from 2010 census data where the crime data is from 2017. The 2010 census data was the latest that could be located. The assumption made is that neighborhoods have not significantly changed between 2010 and 2017. If there were significan changes, this would affect the assortativity of the attributes (different assortativities may actually be present in 2017 versus 2010).