Empirical Analysis of Inter-district Crime Network

Rais Iqbal
rais.iqbal@gmail.com
Västerås, Sweden

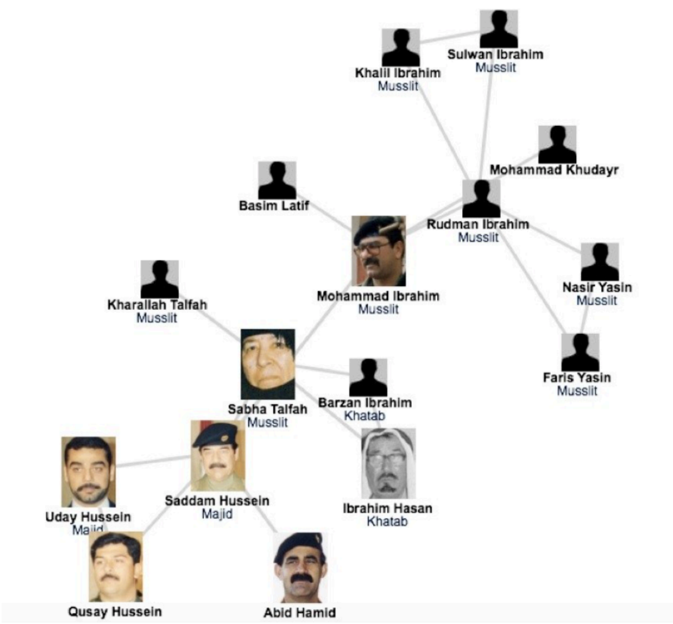
Introduction

The emergence of network science as a separate but inter disciplinary field of study has brought up many new and interesting insights in variety of different domains such as telecommunication, computer science, biological studies, complex systems and social context etc. Moreover, things, information and life in practice have become more and more inter connected. We are generating tons of information and connections among such information without even noticing and thus, ending up building all possible kinds of social networks around us. Companies like Facebook, LinkedIn, Twitter and Google whose business deeply reply on social network have also contributed a lot accelerate the interest in social network analysis (SNA) in particular and network science in general.

Crimes and Networks

In order to make the world a place there is a consensus to eradicate crimes. Every year, each country spend huge amount of the budgets in law enforcement and yet this phenomenon looks unstoppable. In the case mafias, the underlying structure how operates and interact has not changed much but still the results the not satisfactory [2].

Barabási highlighted the story how social network analysis (or network science) helped the US forces to capture Saddam Husain when intelligence failed to do so [1].



Saddam's social network with high-ranking officials. To find Saddam, US was looking for someone close to him who could tell about his whereabouts. His direct connection could not help US to tell about his whereabouts. Network drawn from his family album revealed the second level high-rankings officials know more than his first level high-ranking officials.

Clearly the power of networks is huge. This provides us enough motivation to practice network analysis in those domains in which typical or standards techniques does not produce desired results.

Västerås Inter-District Crime Network

Background

Västerås is a medium size city located in central Sweden and comprises of 41 residential units or districts. Out of those 41 districts, 14 districts are known as "urban districts"[3]. In this paper, an attempt has been made to analyze the inter-district crimes based on the police reports published in the newspaper from January 2014 to November 2014.

In this paper, the definition of crime and their categorization is out of the scope therefore, crime types and categories described by official police websites and Crime Prevention Council are used; according the sources there are roughly 15-22 crime types ranging from shoplifting to human trafficking.

Data Acquisition

Västerås tidning, a city newspaper publishes twice every week used as main source of data because no machine-readable data was found. Police reports are published in the newspapers about the major incidents reported and registered in the police register in textual form with some description, location and date. As we are only interested in crime related incidents therefore reports about "fire", "traffic accidents", and "injury at workplace" has been excluded. Apart from the excluded crime types, 13 types of crimes have been found in the 45 police reports.

1. Drunk driving
2. Burglary
3. Drug driving
4. Molestation
5. Assault
6. Fraud
7. Mugging
8. Shoplifting
9. Drug offence
10. Theft
11. Unlawful threats
12. Vandalism
13. Environmental crime

Data Construction & Pre-processing

Textual descriptions of each report were written in spreadsheet format (comma separated values) with the intention to load into Gephi later for analysis. Two spreadsheets were created, edge table and node table. Node table contains the number of residential units (denoted as Id in the node table) and totals crime counts recorded that against each district (denoted as weight in the node table). Edge table comprises of source and target nodes, since it is an undirected graph therefore no distinction between a source and target is of any importance. However, edge weight is in place in the edge table, which increases with the amount of crimes recorded between a source and target node. Initially all the edges have weight "1" and increases with a factor of "1" if both source and target node weight is increased within the same type of crime.

Id	Weight	Labels
Source/Target Node	# of crimes recorded atsource/target	Type of crimes

Table 1: Node table formation

Source	Target	Label	Type	Weight
Source node	Target node	Type of crime	Undirected	# of the same crimes recorded between a source and target node.

Table 2: Edge table formation

Below is the quick look of necessary building blocks of the data.

- **Node**: nodes represent districts/residential units.
- **Node attributes**:
- **Node weight**: number of incident reported for a district.
- **Edge**: an edge between two nodes if they share a common crime type.
- **Edge weight**: number of times the same crime type reported between two districts.

Limitations

- A few residential units have been further divided into sub-districts or sub residential units in order to bring some clarity to data. This is to avoid the confusion about the general perception of the districts. With this limitation we end up with 50 residential units in total.
- The textual description of the published police report does not explicitly mention exactly one crime type; for instance, in some cases both burglary and theft are mentioned. In such cases if both crimes are recorded it would increase the number of edges. An effort has been made to reduce this inappropriate linking but still there could be some chance of error.

Network Construction

Newly assembled data as edge list and node list is imported in to Gephi for visualization and analysis purpose. The resulting network consists of 49 nodes and 44 links. An important point to notice is the number of links that have been reduced to 441 that should have been 554 according to the edge table formation. This is due to the fact that there are multiple connections between two same nodes (source and target) for different crime types. Gephi automatically add up those duplicate connections and different crime types merges together, higher crime type takes over the lower crime type. This a caveat with come up with this formation to avoid extremely lengthy edge formation table.Short summary statistics of the network is given in the table 1 and the initial look of the network with giant component infocus is shown is figure 1.

Measure	Statistic
Number of node	49
Number of edges	441
Average Shortest Path	1.606
Connected Components	4
Diameter	3
Average Clustering Coefficient	0.855

Table 3: Network summary statistics

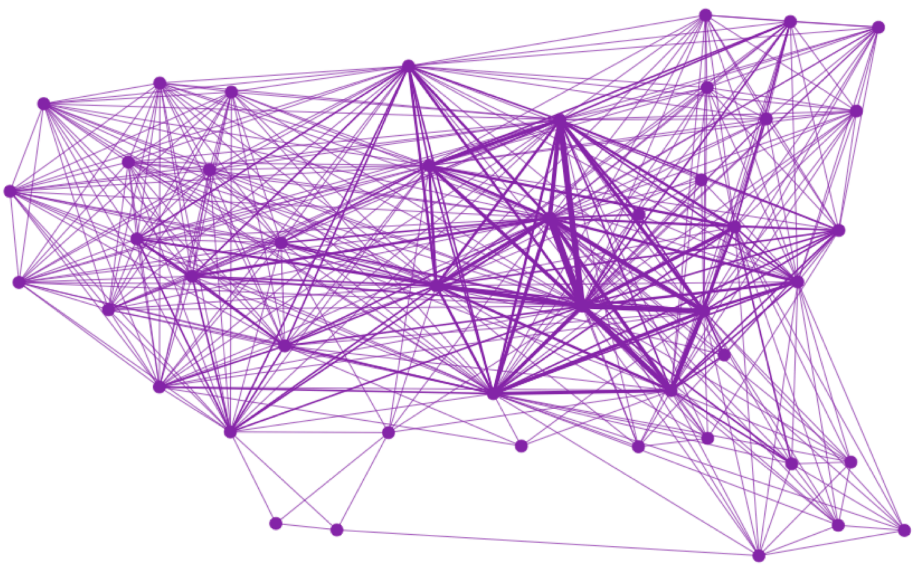


Figure 1: Initial layout of the network using Force Atlas 2 layout algorithm in Gephi with "enable dissuade hubs" and "prevent overlap" options enabled

Network Navigation & Degree Distribution

On average each node is connected with 18 other nodes and degree distribution shows that most of the nodes are well connected. Almost more than a half of the nodes have degrees higher than the average degree (see figure 2 and figure 3). It's interesting to look at some of the statistics about degree.

Measure	Statistic
Average Degree	18
# of node with degree > Avg. degree	26 (53%)
# of node with degree < Avg. degree	23 (47%)

Table 4: Number of nodes falling above or below average degree

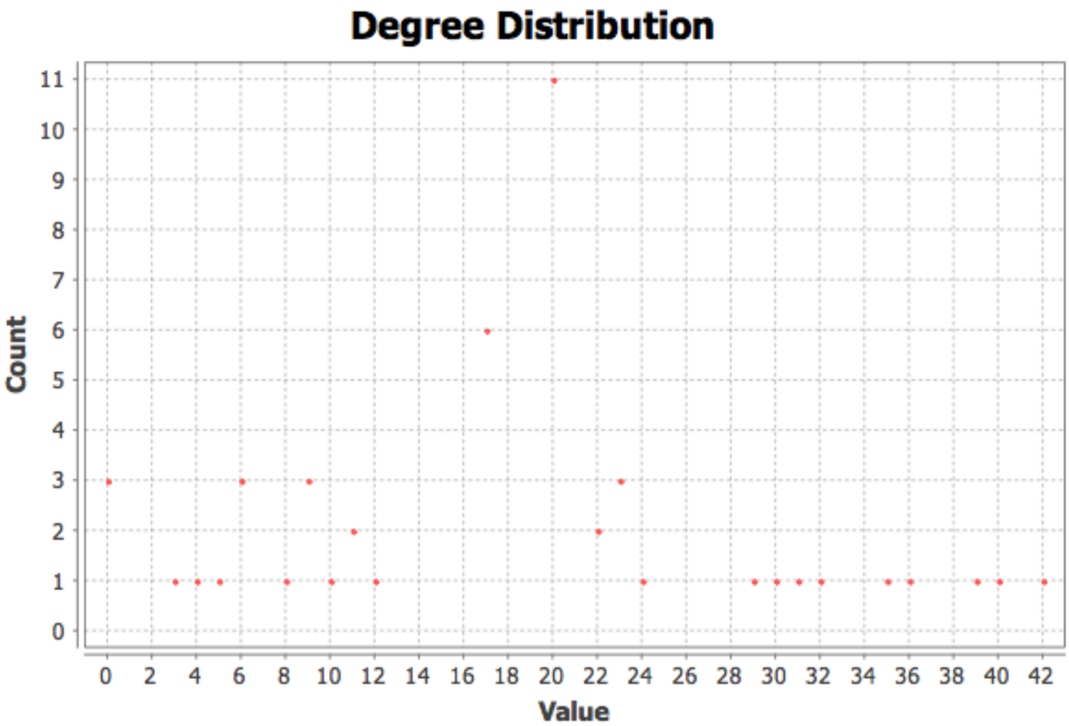


Figure 2: Degree Distribution

From the perspective of degree distribution, in this inter-district crime network, half of the districts in the city are well connected. But it important to ask whether the nodes with higher degree only share same type of crime with all or most of its neighbors or there is smooth distribution in terms of variety of crimes that two places share. To capture this variation we have to bring in the node weight attribute, which is the total count of crimes in a certain district. Take a look at the figure 3 and table 5 that shows the nodes with crime count against its degree.

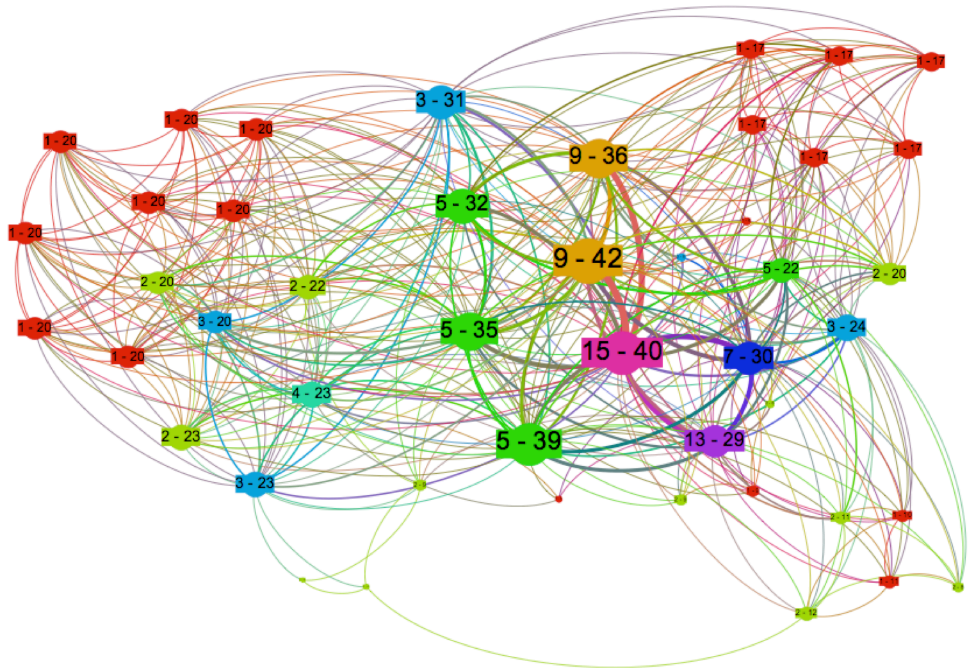


Figure 3: Number on the nodes represents the weight-degree. Nodes are sized according to their degree

Comparison of Centralities

As the measure of centralities are central in any network as they are here. It can be seen from the table that nodes with high weight have degree and betweenness centralities as opposed to closeness centralities that shows the mixed behavior as it can be seen in the case of "Bjurohovda", see table below, which has a both low degree and weight but very high closeness centrality and very low betweenness. We want to extend the centralities in terms of general perception of the residential district if we could infer something. Since weight is most central element which is enabling force behind the dynamics of the network. It's worth taking a look at the centralities as a function of weight also.

Node	**Weight [0, 15]**	**Degree [0, 42]**	**Closeness [0, 2.37]**	**Betweenness [0, 105]**
Vallby	9	42	1.067	105.021
City	15	40	1.111	82.796
Råby	5	39	1.133	76.343
Erikslund	9	36	1.2	52.44
Bäckby	5	35	1.222	42.98
Hammarby	5	32	1.289	26.406
Haga	3	32	1.311	20.406
Centrum	7	30	1.356	43.753
VästeråsO	13	29	1.356	37.582
Gideonberg	5	22	1.533	9.229
Bjurohovda	2	3	2.378	0

Table 5: Weight attribute and centrality comparison for node with high weight

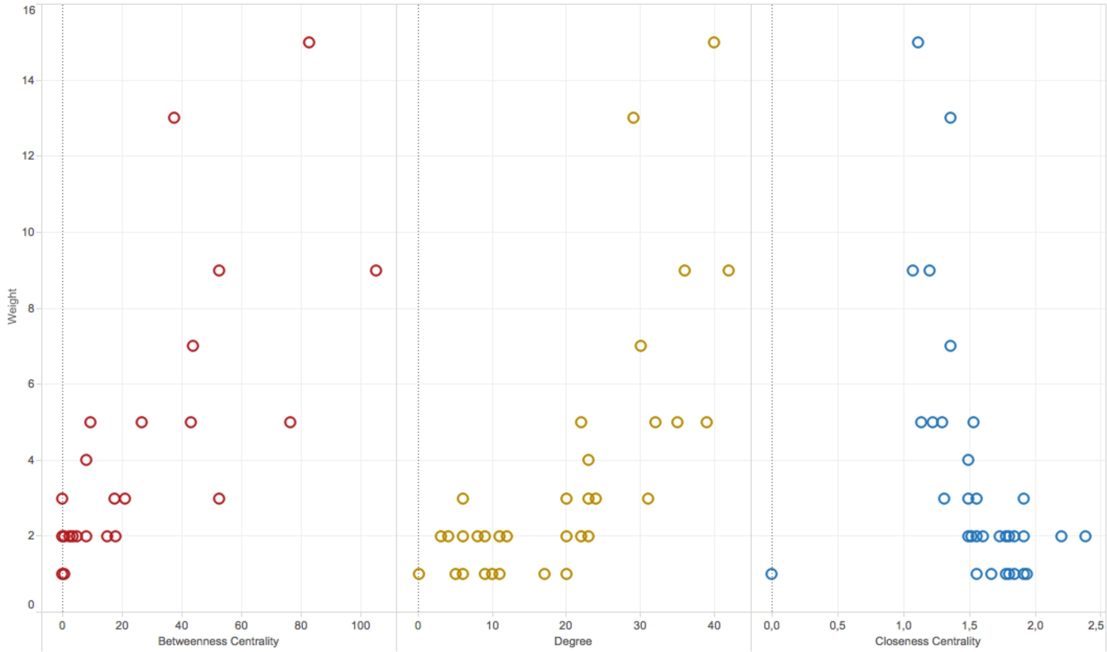


Figure 4: Centralities as a function of weight

The above plot also shows the same intuition as described previously that both degree and betweenness do exhibit some commonalities and closeness also showing some trend but not in the same way as other centralities.

We can argue that the current data is too small to draw any conclusion about the perception of the districts and for exploring the possible properties of this network.

Network Formation Process

At this point it is important to see what is the process behind this network have this small data set. Before continuing the discussion further towards process formation, looking back to figure 3 could reveal some groups of nodes consisting of similar degree. This could probably capture the phenomenon of hybrid model that has properties of preferential attachment and random network formation simultaneously [5]. In which each newborn node is attached at random with some probability and then form links with its neighbors.

In our case, the probability of occurring a certain crime varies because some of the crimes are very unlikely to occur, for example human trafficking and some of the crime are more likely to happen like drunk driving.

So, each newborn node has some probability of getting attached with previously existing nodes, after selecting that node it forms links with all the other nodes within that group. This process formation does not proved here but at least it fulfills the general intuition about the formation of such network.

Community Detection

On running the modularity class feature in Gephi, some community detection has been done. It has been further filtered with partition modularity class setting to get a clear view about the communities. The raw size of communities is 6. Among them there are 3 major communities making up the 94% of the total network (giant component), (**Blue 47%, Green 33%, Red 14%**) and rest of the 3 communities make up a total of 6% with 2% each. Leaving those nodes aside communities on giant component is depicted below.

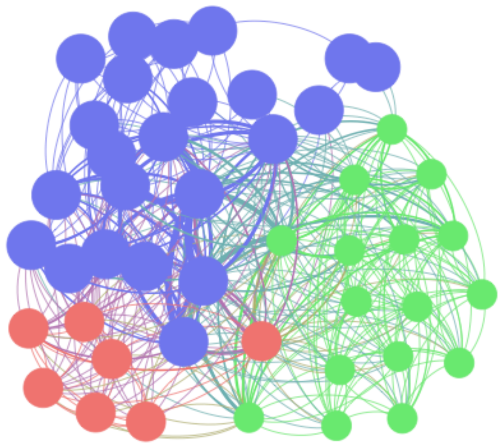


Figure 5: Three largest communities in a giant component

K-cores

Splitting with K-core is another approach used to find the communities at different levels of variation. The giant component was filtered in Gephi with different K-core settings.

Unless and until a criterion is specified for the k-core, it is difficult to say something which k-core setting is the best. However, it has been noted that the certain crime type rapidly falling with different k-core settings. That can be useful to detach nodes associated with certain crime types. With the following k-core settings, "Narcotic Offence" drops rapidly at k=11 and the other useful insight is that only "Burglary" is left with a very negligible portion "Shoplifting" at k=18.

K-cores	Nodes	Edges
5	44	435
8	39	406
11	32	347
18	21	210
21	0	0

Table 6: K-core splitting of giant component

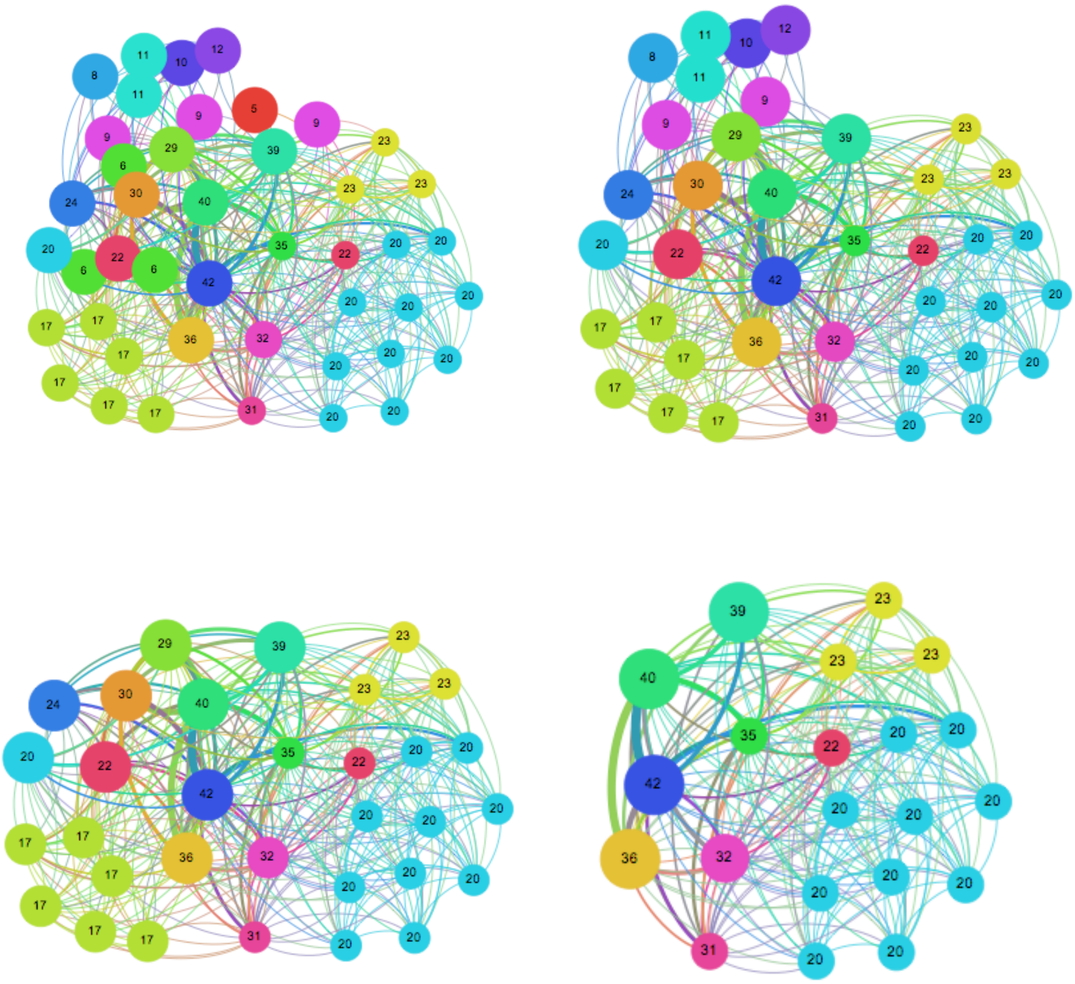


Figure 6: K-core split (k=5, 8, 11, 18 respectively) of giant component consisting of 46 node 441 nodes

Benchmarking with Erdös-Renyi Model & Small-world-ness

The inter district crime network has also been compared against the Erdös-Renyi model to get some insights. One of the take away one can get out of this comparison is to look at the small-world-ness [6]. The small-world-ness takes the clustering coefficient and average path length to see if a given network shows the small world phenomenon. Calculation for the small-world-ness (*S*) is given below.

$$S = \frac{(\text{clustering coefficient of observed network}) \text{ x } (\text{Avg.path length of ER})}{(\text{clustering coefficient of ER}) \text{ x } (\text{Avg.path length of observed network})} > 1$$

Network	**Avg. path length**	**Clustering coefficient**
Erdös-Reyni	1.614	0.399
Observed Network	1.606	0.855

The small-world-ness in the case of criminal network is **2.153531 > 1**. So this network do possesses the property of small world.

Applicability

This network was constructed with the sole purpose of mapping crimes on residential units but navigation through this network and looking at the network in general shows us that such network could be useful in other settings, for instance, diffusion, serial crime investigations, for public awareness campaigns and for real estate experts to check the general perception of city residential units if a strict criteria is given.

Summary and Discussion

In this paper, crime reports published in the local newspaper of a city are mapped to construct a network, in which nodes are residential areas and links between them exist if same type of crime occurs two nodes.

We have generally gone through the network and have seen very short average path length that is exactly equal to the Erdös-Reyni model with same number of nodes with like probability of 0.0376. The small-world-ness property has also been observed in this network.

The measures of centralities are compared and have been plotted against the weight associated with each node, which is the total number of crimes recorded at that node but this network is too small to infer something or to find trend to capture properties like the general perception of a city district. However, a very little similarities and dissimilarities have been found when centralities are compared against weight attribute of a node.

The modularity algorithm and k-core filleting have been used to split the network in communities. Bother community detection algorithms serve different purposes and it has been seen that a criteria is necessary for separating communities, for example, in case of k=18 when the network was not only left with only high degree but also a community of nodes which has same type of connection.

Future Work

In order to achieve the profound results this study needs to be extended in two ways.

- Reconstruct the network in bipartite form.
- Continue current network by accumulating previous data into it.

References

[1] Albert-László Barabási. *Network Science*. Chapter 1 (section 2). 2012

[2] Giovanni Mastrobuoni, Eleonora Patacchini. Organized Crime Networks: an Application of Network Analysis Techniques to the American Mafia. Review of Network Economics Volume 11, Issue 3, 2012

[3] http://vasteras.se/omvasteras/statiskochfakta/omradesfakta/Sidor/omrade sfakta.aspx

[4] http://polisen.se/Utsatt-for-brott/Olika-typer-av-brott/

[5]https://d396qusza40orc.cloudfront.net/networksonline/lecture_slides/Jackson-NetworksOnline-Week3-Slides.pdf

[6] Humphries M.D., Gurney K. "*Network 'small-world-ness'*: A quantitative method for determining canonical network equivalence." PLoS ONE 3: 3: e0002051. (2008).

Data

https://github.com/iraisi/SNA-Project

Appendix

Sample police report in the local newspaper

POLISRAPPORTEN

GIDEONSBERG
Torsdag 18/9, kl 10:10
Narkotikabrott. En 28-årig man fick följa med till polishuset för provtagning och förhör, efter att ha påträffats misstänkt påverkad av narkotika. Han uppehöll sig i anslutning till Nettobutiken på Bangatan tillsammans med en kamrat. Misstänkt för ringa narkotikabrott, eget bruk och innehav.

STENBY
Torsdag 18/9, kl 4:47
Arbetsplatsolycka. En 26-årig kvinna bröt fotleden i en klämolycka på Stenby.
Olyckan inträffade på V-TAB på Stenbygatan. Kvinnan hade kört truck och fått foten i kläm mellan sin truck och en lastpall. Hon fördes med ambulans till sjukhus.

GRYTA
Torsdag 18/9, kl 2:13
Brand. Sex bilar brandskadade varav tre helt utbrända på Drevervägen. Bilarna stod parkerade utomhus på en parkeringsplats. Spridningsrisken till inliggande garagelängor var liten. Det är oklart hur branden uppstått, men det kan inte uteslutas att branden var anlagd.

KÄRRBO
Onsdag 17/9, kl 7:29
Inbrott. På ett lantbruk i Kärrbo öster om Västerås har det varit inbrott i en verkstad. Okänd gärningsman har på okänt sätt tagit sig in i verkstaden och där tillgripit gods för stora värden. Polis har varit på plats och brottsplatsundersökning är gjord.

VÄSTERÅS
Onsdag 17/9, kl 1:02
Stöld. Kolliderade med vildsvin. En 32-årig man kolliderade under morgonen med ett vildsvin på LV543. Vildsvinet linkade iväg och bilen ådrog sig en del skador. Föraren klarade sig helt utan skador. Jägare kallades till platsen för att försöka lokaliera djuret, oklart hur allvarligt skadat det blev.

Local map of the city of Västerås: distribution of city in to residential units