E-COMMERCE FRAUD:
An approach with social network analysis

Analytcs – Experian Brazil

Thaine Clemente de Souza

E-mail:

thaine.souza@br.experian.com

thaineclemente@gmail.com

ABSTRACT

Organized crimes such as fraud, terrorism and drug dealing are performed by multiple attackers collaborating, who can form groups and teams that play different roles and therefore the relevance of introducing the concepts of social networking in the investigative analysis of these crimes. First, we introduce the concept of social networks and criminal behavior. Afterwards, the database used is related to purchases in e-commerce, and the fraud is defined by the chargeback occurrence (refund of the purchase by the card issuer to the holder). The aim of this work is to analyze if the fraudulent and non-fraudulent networks differ in their structural characteristics, using the logistic regression and ego social network methodologies. Among the main results, the structure of the network of fraudulent and non-fraudulent purchases differs in: 1) number of connections, 2) number of betweenness with other purchases and 3) the network density. The ego networks of fraudulent purchases have more connections, less betweenness and higher density.

Keywords: E-commerce; fraud; social network analysis.

INTRODUCTION

The revenue in internet sales in Brazil has grown, about 40% per year and this growth reflects in number of legitimate and fraudulent purchases. Fraud are growing with the expansion of modern technology and global communication, resulting in substantial losses in business (Kou, Lu, & Sinvongwattana, 2004). According to data published by E-Net, the number of fraud attempts on e-commerce involving credit card increased between 2011 and 2012, from 3.6% to 3.8% in the period from January to May. In addition, the most common online fraud are in products like cell phones, electronics and airline tickets (E-NET, 2012).

Fraud with credit cards in e-commerce are cited as a major reason for closing virtual stores in Brazil, mainly due to chargeback (Hitmidia, 2012): refund of the purchase by the card issuer to the holder. In addition, the banks are not accountable for fraudulent activity at the POS, the merchants take on any losses. Because this the merchants need to prevent the fraud to protect their profit.

Beyond the loss generated by fraud, the merchants are the image denigrated with good customers when they do not approve a good transaction. Therefore, having an anti-fraud management is one of the prerequisites for the success of online shops and reinforces the importance of this theme.

Several studies in the literature are related with the detection of fraud in transactions with credit card, mainly methods of analysis patterns, such as artificial neural networks, case-based reasoning and data mining (data mining) (Ghosh & Reilly, 1994; Hanagandi, Dhar, & Buescher, 1996; Aleskerov, Freisleben, & Rao, 1999; Sorronsoro, Ginel, & Cruz, 1997; Richardson, 1997; Chan, Fan, & Stolfo, 1999; Liu & Yao, 1999; Brause, Langsdorf, & Hepp, 1999) .

The authors cited above studied the individual behavior of each purchase, but unlike other types of crimes, often carried out by a single or few offenders, organized crime, such as fraud, are performed by multiple attackers that can form groups and teams that play different roles. Hence the relevance of introducing the concepts of social networks in the analysis (Xu & Chen, 2005) . According to the authors, although the approach through social network analysis (SNA) is not traditionally considered a data mining technique is suitable for mining large volumes of data to uncover hidden structural patterns in criminal networks (McAndrew, 1999).

The objective of this study is to identify which network structure characteristics we can use to predict fraud. For this, the paper analyzes a shopping sample of e-commerce, in which each network node is a purchase and the links among them, the common data. The study unit is each transaction, because in e-commerce often the fraudster has a false identity. Therefore, it is not possible to associate the purchase to the individual. Therefore, the work want to find predictive patterns of fraud in the shopping network structure and not find criminal individuals.

LITERATURE REVIEW

E-COMMERCE AND FRAUD

Sales in the e-commerce began to grow in the United States around 1995, with the emergence of some companies like Amazon.com. Five years later, many merchants have emerged in Brazil and since then sales and revenues have grown about 40% per year (Figure 1). Several factors are attributed to this growth, including the ongoing digital inclusion, the cheapening of computers and the increasing use of smartphones, the ease and convenience, affordable and competitive with the prices of physical stores.

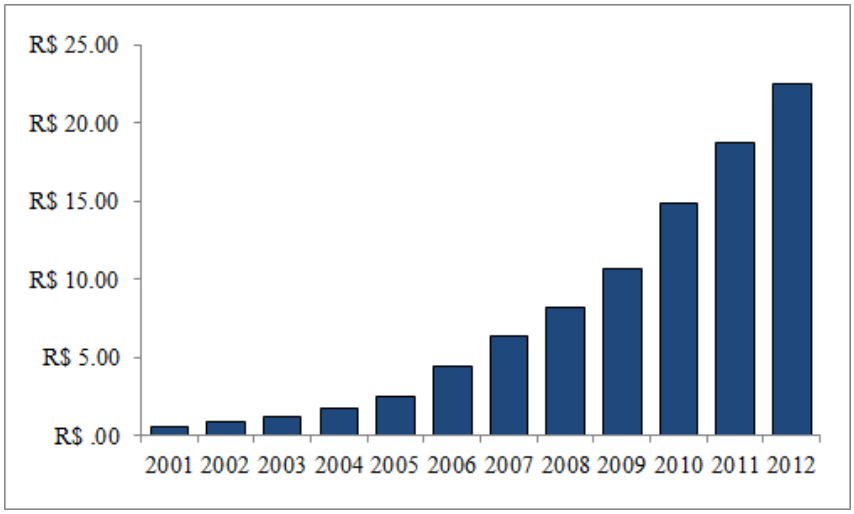

Figure 1: Annual Revenue E-Commerce Brazil (billion)
Source: ebit – http://www.ebitempresa.com.br/

On the other hand, fraud in e-commerce is also growing, fraudulent transactions increased by 26% from 2010 to 2011, from $ 2.7 billion to $ 3.4 billion, according to the Report Online Fraud (2012). The report also brings the largest online stores have the lowest rates of fraud, given the greater capacity to make investments in tools and training people to prevent fraud.

SOCIAL NETWORK ANALYSIS (SNA)

The perspective of social network covers theories, models and applications that are expressed in terms of concepts or relational processes (Wasserman & Faust, 1994). The social network analysis (SNA) is the mapping and measuring of relationships and flows among people, groups, organizations, computers, etc. The network nodes are the

unit studied, while the links show relationships or flows between the nodes. Networks are often presented in a diagram, where nodes are represented by dots and the connections (links) represented by lines.

The SNA has become an important tool for criminologists who seek to understand the connections between patterns of interactions and criminal behavior. The links between criminal activities is used as a way to indicate the relationships between communities, which can potentially identify communities of criminals (Park, Tsang, & Brantingham, 2012).

The results of published studies show that the offender has a strategic position in social network, often occupy central positions and must be removed or controlled (Baker e Faulkner 1993; McAndrew 1999; Sparrow 1991). A central member can play a key role in a network, acting as a leader who issues commands and provides steering mechanisms or serving as a porter to ensure that the information or the flow of goods disperse efficiently between the different parts of the network. The removal of these core members can effectively cut the net and finish with running a criminal enterprise. Krebs (2002) mentioned that the terrorist leaders have the highest centrality scores (degree, betweenness and closeness) ant they can control and have the necessary information and resources. Morselli e Roy (2008) showed that the stolen vehicle operations are centralized and resistant because they have catalysts members (high betweenness) that increase the degree of flexibility to achieve the collective goal.

About the high density of the networks, Finckenauer e Waring (1998) showed that low density is related to criminal networks of money laundering. Density is represents the proportion of possible relationships in a network that are actually present. The value ranges from 0 to 1; the closer the value is to 0, the sparser the network is while the closer the value is to 1, the denser the network is. Secret networks such as criminal or terrorist networks, have low density because fewer contacts between members of the network reduces chance of revelation and the detection of a member of it (Goede, 2012).

The methodology of the studies cited above considered network analysis at the level of the interest group like drugs, fraud or terrorists, because networks are complete (without subnets) and they have easy interpretation / measurement. As purchases are not face to face and do not have the true person who performs the purchase, i.e., when it comes to fraud, the fraudster often used to other identity to make the crime. This work analyzes the networks of each purchase, with a methodology called Ego Network Analysis.

The ego networks, as well as other networks, are formed of social units and their relationships, which connect the pairs of units for some kind of connection. Ego network, however, is built around a particular social unit designated ego and their connections, called alters (Freeman, 1982), as can be seen in Figure 2 and Figure 3 below. In other words, a ego network is a sub-graph to original network where the set of vertices is the ego and its direct neighbors.

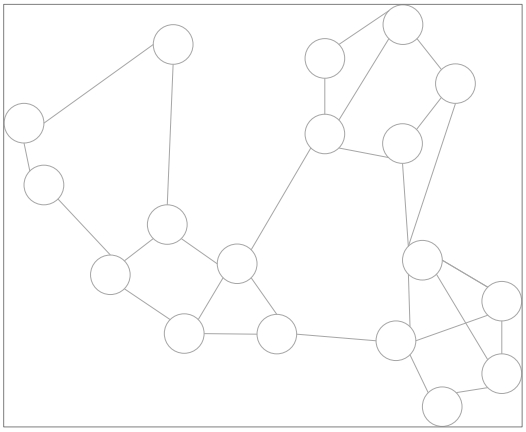

Figure 2: Social Network

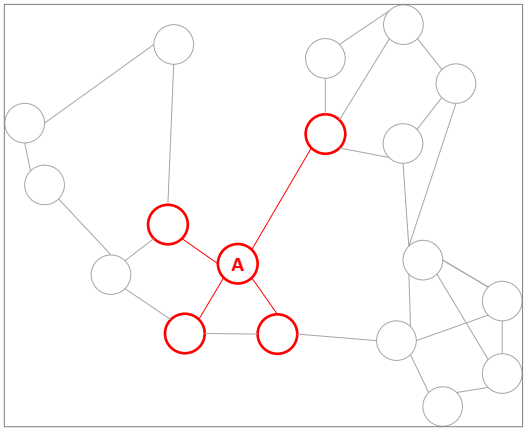

Figure 3: Ego network of the node A

METODOLOGY

EGOS MEASURES

SIZE

Size measures how connections (alters) the ego has in a social relationship. The size of the node i shows its connectivity and is defined by:

$$C_S(i) = \sum_{j=1}^{n} a_{ij}, \quad (1)$$

where $a_{ij} = 1$ when there is a direct link between the i and j and $a_{ij} = 0$ otherwise.

BETWEENNESS

Betweenness centrality is an indicator of a node's centrality in a network. It is equal to the number of shortest paths from all vertices to all others that pass through that node. A node with high betweenness centrality has a large influence on the transfer of items through the network, under the assumption that item transfer follows the shortest paths.

$$C_B(i) = \sum_{k \neq i \neq j \in N} \frac{\sigma_{kj}(i)}{\sigma_{kj}}, \quad (2)$$

onde σ_{kj} is the total number of shortest paths from node v_k to node v_j, e $\sigma_{kj}(i)$ s the number of those paths that pass through v_i.

DENSITY

The density reflects the number of connections compared to the number of possible connections of the ego:

$$C_D(i) = \frac{C_L(i)}{C_P(i)}, \quad (3)$$

where $C_L(i)$ is the number of connections and $C_P(i)$ is the number of ego's connections.

RESULTS

The data analyzed come from a sample of Brazilian e-commerce with at least one connection to previous purchases. The initial sample had 500 purchases: 250 fraud and 250 non fraud and their connections. The criteria for defining a purchase as fraud is the chargeback, contested by the cardholder.

For the construction of the connections between the purchases considered the information: address, phone, e-mail, cardholder or CPF. That is, if two purchases have at least one of the data mentioned in common they are connect. The study did not distinguish between the links, so they were treated with the same weight. Furthermore, the connection of the purchases are indirectly, i.e. if node A connects to node B, then the node B connects to node A.

We extracted 7 connections of the nodes, that is, according to the image below, the direct link from node A is the node B, however, we too extracted the nodes C, D, E, F, G.

VARIABLES

To test the influence of network variables, we used logistic regression and the variable response to the model is a dummy indicating fraud (0) or non-fraud (1). The tested explanatory variables are divided into two categories, they are dependent and control. The dependent variables are: density, betweenness and size.

The control variables can affect the probability of fraud. Three control variables in the study were included:

CEP Risk: Zip Codes classified as high risk according to historic purchases by region;

E-mail Risk: E-mail domain classified as high risk to historic purchases. Examples of e-mail domains: gmail, yahoo, UOL, bol, walla, etc .;

Items: dummies for categories telephony, electronics, watches, beauty and health.

Below the results of logistical models:

Parameter	Model 1	Model 2	Model 3	Model 4
Intercept	2.770 ***	3.1436 ***	2.819 ***	3.295 ***
Value	-0.002 ***	-0.002 ***	-0.002	-0.002 ***
Region with historial of fraud	-1.966 ***	-2.051 ***	-2.0563 ***	-2.007 ***
E-mail with historical of fraud	-0.707 **	-0.700 **	-0.719 **	-0.6762 **
Dummy of phones	-0.494 *	-0.444 *	-0.423 *	-0.5033 *
Dummy of eletronics	-0.346 *	-0.357 *	-0.366 *	-0.3514 *
Dummy of watches	-1.825 *	-1.8959 **	-1.8536 *	-1.9343 **
Dummy of Beauty and health	-2.714 **	-2.9162 **	-2.848 **	-2.8363 **
Ego Size		-0.0006 ***	-0.0006 ***	-0.0002 ***
Ego Betwenness		0.0146 **	0.0178 ***	
Ego Densit		-0.048		-0.00694 **
KS	**58.8**	**58.7**	**60.9**	**59.2**

*** p-value less than l0.001;** p-value less than 0.05; * p-value less than 0.10

The results of the models show when we added the SNA variables the KS increase (the performance increase). In the model 2 the variable Density was not significance, but in the Model 4 it was, this results could happen because the correlation between Density and Betweenness. So, the 3 variables of social network were significance to prevent the fraud and they increase the model results.

In the model 3 we tested Betweenness and Size and in the model 4 Density and Size to remove the effect of correlation between the variables. In these models all variables were significance. These results suggests that the fraud happens when the number of ego connections are high and number of Betweenness is low or value of density is high. These results can be explained by fraudster profile: they make a lot of purchases and they relate to each other – the fraudsters have a restricted data to make the purchases.

The variable density has the signal negative in Model 2 and 4 and this reinforce the theory that high density reflects in a network more coordinate, more opportunities to share information and resources among the nodes (Corteville & Sun, 2009). About the size of the networks, the signal positive reinforce the results that show that the offender has a strategic position in social network, often occupy central positions and must be removed or controlled (Baker e Faulkner 1993; McAndrew 1999; Sparrow 1991).

About the control variables, all of them were significance, as we expected. The fraudsters make purchases with high value, products of easy resale (electronics, watches, beauty and health), and use domain of e-mail and zip code with historical of fraud.

CONSIDERAÇÕES FINAIS

The fraud in e-commerce needs to be fought for the survival of merchants because fraudsters cause considerable damage to stores. They act as professionals in this area to make many purchases and they are so dynamic that traditional patterns of prevent the fraud not detect their purchases. Usually they buy easy resale products and work as "sellers" of products such as TVs, video games, GPSs, among others. Combat this crime is of paramount importance to the merchants, and mostly for good clients, who are the victims generated by it.

This work innovates in the methodology to prevent the fraud and proposes SNA to estimate the fraud. Tree measures were studied: density, betweenness and size. The results show the fraudulent purchases tents to have more connections, less Betweenness and high density. These results confirms the hypothesis of that the centrality measures imply control over the resources because the central actors can choose alternatives to share these ones (Sparrowe, Liden, Wayne, & Kraimer, 2001).

Therefore, it is assumed that fraudsters capture the data necessary to carry out shopping, make many purchases in order to win maximum profit with the data, and they form networks with many connections, few Betweenness and closed.

However, the work includes only three measures of ego networks. Therefore, it is suggested that, for future work, new measures / variables to be explored. Furthermore, the size variable would be easily obtained without the use of network methodology, since it measures only the direct ego. By contrast, the betweenness and density variables consider the relationship of alters and imply the use of the network methodology for the capture these information.

BIBLIOGRAFIA

Aleskerov, E., Freisleben, B., & Rao, B. C. (1999). A neural network based database mining system for credit card fraud detection. *IEEE/IAFE of Computational Intelligence for Financial Engineering*, 220-226.

Anthonisse, J. (1971). *The rush in a directed graph.* Amsterdam: Stichting Mahtematisch Centrum.

Backstrom, L., Huttenlocher, D., Kleinberg, J., & Lan, X. (2006). Group formation in large social networks: Membership,growth, and evolution. *Proc. 12th ACMSIGKDD International.*

Baker, W. E., & Faulkner, R. R. (1993, Dezembro). The Social Organization of Conspiracy: Illegal Networks in the Heavy Electrical Equipment Industry. *58*, pp. 837-860.

Bott, E. (1955). Urban families: conjugal roles and social networks. *Human Relations, 08*, 345-384.

Brause, R., Langsdorf, T., & Hepp, M. (1999). Neural data mining for credit card fraud detection. *IEEE International Conference on Tools for Artificial Intelligence*, 103-106.

Burt, M. R. (1980). Cultural myths and supports for rape. *Journal of Personality and Social Psychology, 33*, pp. 217-230.

Burt, R. (1992). Structural Holes: the Social Structure of Competition. *Harvard University Press*.

Burt, R. S. (1980). Models Of Network Structure. *Annual Review of Sociology*, 79-141.

Burt, R. S. (2000). The Network Structure Of Social Capital. *Research in Organizational Behaviour, 22*, 345-423.

Chan, P. K., Fan, W., & Stolfo, S. J. (1999). Distributed data mining in credit card fraud detection. . *IEEE Intelligent Systems*, 67-74.

Chin, A., & Chignell, M. (2007). Identifying communities in blogs: roles for social network analysis and survey instruments. *International Journal of Web Based Communities, 3*, 345-362.

Conover, W. J. (1999). Practical Nonparametric Statistics. *Wiley*(3).

Cortes, C., Pregibon, D., & Volinsky, C. (2003). Computational methods for dynamics graphs. *Journal of Computational and Graphical Statistics, 12*(4), pp. 950-970.

Corteville, L., & Sun, M. (2009). *An Interorganizational Social Network Analysis of the Michigan Diabetes Outreach Networks.* Michigan: Michigan Department of Community Health.

Crites, S. (2008). Best Practices in Addressing Online Cash Management Security. *Commercial Lending Review*, 21-26.

CyberSource. (2012). *Online Fraud Report: Online Payment Fraud Trends, Merchant Practices and Benchmarks.*

Décary-Hétu, D. (2011). Chit-Hack. Information exchange paths in IRC hacking chatrooms. *Universidade de Montreal*.

E-NET, C. (2012, 7 27). *Câmera E-Net*. Retrieved 9 15, 2012, from Câmera E-Net: http://www.camara-e.net

Everett, M., & Borgatti, S. P. (2005). Ego network betweenness. *Social Networks*, 31-38.

Finckenauer, J., & Waring, E. (1998). Russian Mafia in America: Immigration, Culture, and Crime. *Northeastern University*.

Freeman, L. C. (1977). A set of measuring centrality based on betweenness. *Sociometry, 40*, pp. 35–41.

Freeman, L. C. (1978). Centrality in social networks conceptual clarification. *Social Networks*, 215-239.

Freeman, L. C. (1982). Centered Graphs and the Structure of Ego Networks. *Mathematical Social Sciences*, 291 -304.

Ghosh, S., & Reilly, D. L. (1994). Credit card fraud detection with a neural network. *Annual Hawaii International Conference on System Sciences*, 621-630.

Goede, M. d. (2012). Fighting the network: a critique of the network as a security technology. *Distinktion: Scandinavian Journal of Social Theory* , 215-232.

Hair, J. H., Anderson, R. E., Tatham, R. L., & Black, W. C. (2005). trad. Adonai Schlup Sant'Ana e Anselmo Chaves Neto. Análise Multivariada de Dados. (5).

Hanagandi, V., Dhar, A., & Buescher, K. (1996). Density based clustering and radial basis function modeling to generate credit card fraud scores. . *IEEE/IAFE Conference on Computational Intelligence for Financial Engineering*, 247-251.

Hitmidia. (2012, 8 22). *HitMidia*. Retrieved 9 2012, 24, from HitMidia: http://www.hitmidia.com.br/?post=tipos-de-fraudes-com-cartao-de-credito-no-e-commerce

Hosmer, D., & Lemeshow, S. (2000). Applied Logistic Regression.

Institute, P. (2012). *Second Annual Cost of Cyber Crime Study* . Traverse City: Benchmark Study of U .S .Companies.

Johnson, R., & Wichern, D. W. (1998). Applied Multivariate Statistical Analysis. 642p.

Kerr, K. (2000, 1 24). *Accepting Credit Card Payments on the Internet*. Retrieved from Gartner: http://gartner11.gartnerweb.com/public/static/hotc/hc00085970

Knoke, D., & Burt, R. (1983). Applied network analysis: A methodological introduction. *Prominence*, 195-222.

Knoke, D., & Kuklinski, J. (1982). Network Analysis. Sage University Paper Series on Quantitative Applications in the Social Sciences. 07–028.

Kou, Y., Lu, C., & Sinvongwattana, S. (2004). Survey of Fraud Detection Techniques. 749-754.

Krebs, V. E. (2002). Mapping Networks of Terrorist Cells. *Connections, 24*(3), 43-52.

Liu, Y., & Yao, X. (1999). Simultaneous training of negatively correlated neural networks in an ensemble. *IEEE Transactions on Systems, Man, and Cybernetics — Part B: Cybernetics*, 716-725.

Marsden, P. V. (1987). Core discussion networks of Americans. *American Sociological Review*, 122-131.

Marsden, P. V. (2002). Egocentric and sociocentric measures of network centrality. *Social Networks*, 24, 407-422.

Matusitz, J. (2008). Similarities Between Terrorist Networks in Antiquity and Present-Day Cyberterrorist Networks. *Trends in Organized Crime, 11*, pp. 83-199.

McAndrew, D. (1999). The structural analysis of criminal networks. The Social Psychology of Crime: Groups, Teams, and Networks. *Offender Profiling Series*.

Moreno, J. L. (1934). Who Shall Survive? Washington, DC: Nervous and Mental Disease. *Publishing Company*.

Morselli, C., & Roy, J. (2008). Brokerage Qualifications In Ringing Operations. *CRIMINOLOGY, 46*, 71-98.

Morselli, C., Giguere, C., & Petit, K. (2007). The efficiency/security trade-off in criminal networks. *Social Networks*, pp. 143-153.

Pandit, S., Chau, D. H., Wang, S., & Faloutsos, C. (2007). NetProbe: A Fast and Scalable System for Fraud Detection in Online Auction Networks. *Computer Science Department*, p. Paper 531.

Papachristos, A. V., Braga, A. A., & Hureau, D. (2011, Fevereiro 26). *Six-degree of violent victimization: Social networks and the risk of gunshot injury*. Retrieved 05 22, 2013, from Social Science Research Network Website: http://papers.ssrn.com/sol3/papers.cfm?abstract_id=1772772

Papachristos, A. V., Meares, T. L., & Fagan, J. (2012). Why do Criminals Obey the Law? The Influence Of Legitimacy And Social Networks on Active Gun Offenders. *The Journal Of Criminal Law & Criminology, 102*, 397-440.

Park, A. J., Tsang, H. H., & Brantingham, P. L. (2012). Dynalink: A Framework for Dynamic Criminal Network Visualization. *European Intelligence and Security Informatics Conference*, (pp. 217-224).

Raab, J., & Milward, H. B. (2003). Dark Networks as Problems. *Journal Public Administration Research and Theory, 13*(4), pp. 413-439.

Richardson, R. (1997). Neural networks compared to statistical techniques for Financial Engineering. *IEEE/IAFE Computational Intelligence*, 89-95.

Roman, C. G., Cahill, M., Lachman, P., Lowry, S., Orosco, C., & McCarty, C. (2012). *SOCIAL NETWORKS, DELINQUENCY, AND GANG MEMBERSHIP: USING A NEIGHBORHOOD FRAMEWORK TO EXAMINE THE INFLUENCE OF NETWORK COMPOSITION AND STRUCTURE IN A LATINO COMMUNITY.* Washington, D.C.: The Urban Institute.

Rothenberg, R. B., Potteratb, J. J., Woodhouseb, D. E., Darrowc, W. W., Muthb, S. Q., & Klovdahld, A. S. (1995). Choosing a centrality measure: Epidemiologic correlates in the Colorado Springs study of social networks. *Social Networks*, 273–297.

Serrat, O. (2009). Social Network Analysis. *DigitalCommons@ILR*.

Sorronsoro, J. R., Ginel, F. S., & Cruz, C. S. (1997). Neural fraud detection in credit card operations. . *IEEE Transactions on Neural Networks*, 827-834.

Sparrow, M. (1991). The application of network analysis to criminal intelligence:An assessment of the prospects. *Social Networks*, 251-274.

Sparrowe, R. T., Liden, R. C., Wayne, S. J., & Kraimer, M. L. (2001). Social Networks and the Performance of Individuals and Groups. *Academy of Management Journal, 44*(2), 316-325.

Valente, T. W., Coronges, K., Lakon, C., & Costenbader, E. (2008). How Correlated Are Network Centrality Measures? *Connect (Tor)*, 16-26.

Wasserman, S., & Faust, K. (1994). *Social network analysis: Methods and applications.* Cambridge: Cambridge University Press.

Watts, R., & Witham, A. (2012). *Social Network Analysis of Sustainable Transportation Organizations.* Burlington: A Report from the University of Vermont Transportation Research Center.

Wetherell, C., Plakans, A., & Wellman, B. (1994). Social networks,kinship, and community in Eastern Europe. *Journal of Interdisciplinary History*(24), pp. 639–663.

Xu, J., & Chen, H. (2005). Criminal Network Analysis and Visualization. *Communications of the acm, 48*(06), 101-107.

Xu, J., & Chen, H. (2008). The Topology of Dark Networks. *communications of the acm, 51*(10).