

SOCIAL NETWORKING IS HERE TO STAY

Supervisor: Professor Cecilio Mar-Molinero

September 2013

Universitat Autònoma de Barcelona

Aziz Smagulov



**Universitat Autònoma
de Barcelona**

Abstract

With the introduction of online networking systems the social networks have become even more important and integrated part of our lives. Their increased significance has gained the growing attention from researchers as well. Many different studies proved that information obtained from Social Network Analyses (SNA) can be useful in conducting wide variety of research works and for the sake of this, the ability to analyze and deeply understand the social networking structure is of higher importance as well. This study attempts to summarize the SNA based on personal social network example and data available on particular empirical case study in order to show how social network information can be organized and used in explaining the behavioral influences among different networks and groups.

Keywords: social networks, social network analysis (SNA), correlation analysis, human capital

Aknoledgements

I would like to express my gratitude to Professor Cecilio Mar-Molinero, Master Thesis Supervisor, and my wife, Dinara, for their help and support during the work on this study.

Table of Contents

List of Tables and Figures	3
1. Introduction	4
1.1 Topic Justification.....	4
1.2 Research Questions.....	4
1.3 Structure Outline	5
2. Literature Review	6
2.1 Literature Describing Theoretical Background on Social Networking.....	6
2.1.1 Social Capital	9
2.1.2 Structural Hole Theory	10
2.1.3 Social Closure Theory.....	11
2.1.4 Milgram's Small World Experiments.....	12
3. Social Network Analysis Based on Personal Data Examples.....	14
3.1 Methodology and Data	14
3.2 Facebook Social Network	14
3.3 LinkedIn Professional Network.....	20
3.4 SNA Summary	21
4. Empirical Analysis of Social Networks based on Case Study Examples.....	22
4.1 Methodology and Data	22
4.2 Measured Variables.....	23
4.3 Findings	24
5. Conclusions	28
5.1 Concluding Remarks.....	28
5.2 Limitations and Further Research Recommendations	29
References.....	30

List of Tables and Figures

		page
Figure 2.1	Strong and weak ties in a social network	8
Figure 2.2	Example of ego network	9
Figure 2.3	Example of sparse network	11
Figure 2.4	Example of dense network	12
Figure 3.1	The author's Facebook network	15
Figure 3.2	Mutual friend clusters	16
Figure 3.3	Color-coded friend network	17
Figure 3.4	Highlighted friends	17
Figure 3.5	Color-coded friend network by gender	18
Figure 3.6	Friends' locations	19
Figure 3.7	The author's LinkedIn professional network	21
Figure 4.1	Frequency of smoking between students in the last three months	24
Figure 4.2	Descriptive Statistics	24
Figure 4.3	Correlation between smoking ego and smoking network	25
Figure 4.4	Estimated Results Table	25

1. Introduction

1.1 Topic Justification

Social Networks became an integral part of everyday life for millions of people worldwide. Recently, the new class of informational networks called “online social networking” revolutionized the World Wide Web in popularity, after which the Internet is never going to be the way it was before. Social networking sites such as Facebook (over 1.4 billion users), Vkontakte (over 290 million users), Instagram (over 150 million users), and LinkedIn (over 230 million “professionals”) are examples of extremely popular networks used to find and organize contacts and to share multimedia content. Since the introduction of these websites, millions of users have been attracted and at the same time they integrated social networking into their daily lives to communicate with friends, organize events, learn new information and spread ideas, find jobs, and so on. Online social networks are used to maintain and strengthen existing social ties, or make new social connections. The sites allow users to “articulate and make visible their social networks”, thereby “communicating with people who are already a part of their extended social network” (Boyd and Ellison, 2007). All these issues make social networking an extremely interesting topic of nowadays. This study serves as an example of so called Social Network Analysis (SNA) and how this analysis can be used in empirically explaining particular interrelations in people’s behavior. The research is based on particular examples and case studies due to impossibility to capture the detailed information on large social networks in one analysis. The topic is quite wide, so the proof of existence or absence of many other important interrelations can be conducted within any other social network examples. This emphasizes how useful the SNA and information they contain they can be for researchers.

1.2 Research Questions

The study is designed to answer the following research questions:

Question 1: How social networks, their structure and interrelations can be analysed?

Question 2: Do the social networks affect individual's behaviour?

Based on data availability constrain this question later will be narrowed to smaller subquestion, followed by subsequent hypothesis testing: Do social networks affect the frequency of smoking among students in secondary school?

Hypothesis: The frequency of smoking among secondary school students is highly correlated to the frequency of smoking among his/her friends.

1.3 Structure Outline

First chapter introduces the context of the study with brief justification of chosen topic and brief outline of the work. The following section is designed to give the reader general review of the literature on the subject together with indicating main theories used in the study. The next two sections are designed to answer two main research questions. Chapter Three is made to show how the social networks, their structure and interrelations can be analysed based on particular examples from author's social networks. The next section indicates how useful the social networks can be in explaining other people's behaviour based on school social network sample data variables with methodology used for empirical framework. Finally, the last chapter will contain the main concluding remarks on the previous two parts of analysis and list the limitations and further research recommendations.

2. Literature Review

2.1 Literature Describing Theoretical Background on Social Networking

Social links between individuals in many social systems play an important role in dictating their behavior. Social influence is the phenomenon that the individual's actions can cause his/her friends to behave in analogous way. Understanding and recognizing social influence is of immense interest from analysis and design points of view.

There has been some empirical research on how a user's behavior can be correlated to his/her social affiliations. Backstrom et al. (2006) studied the membership problem in an online community. They observed correlation between a user joining an online community and the number of friends who are members of that community. Marlow et al. (2006) considered the tagging issue in Flickr and examined the set of tags placed by a user and those by user's friends. They found a correlation between tag vocabulary and social connectivity.

Networks are extensively used to reflect the data on connections between associated actors or nodes. In social networks, each actor indicates a person or social group, and each arc, link or tie indicates the existence or intensity of bonds between two actors. Nodes can be used to represent large social units, objects or abstract entities. An attribute of most social networks is transitivity of relations, which means that two nodes that have links to a third node are more likely to be linked than nodes that do not. Transitivity has been broadly studied both empirically and theoretically (White *et al.*, 1976). Transitivity can lead to some clustering of communication within the network. The probability of a tie usually depends on features of the nodes. For instance, the likelihood of a relationship in most social networks depends on the age, gender, geography, race and status of the individuals. In addition, ties are often more likely to occur between actors that have similar attributes than between those who do not, a tendency that is called homophily by attributes (Lazarsfeld and Merton, 1954; Freeman, 1996; McPherson *et al.*,

2001). This means that there is a tendency to connect with people “like” you. For instance, two individuals who play football are more likely to become friends due to the common interest. Many social networks expose clustering beyond what can be explained by transitivity and homophily on observed attributes. This can be driven by homophily on unobserved attributes or on endogenous attributes such as position in the network (Wasserman and Faust, 1994), ‘self-organization’ into groups or a preference for popular actors. Often the key questions in a social network analysis revolve around the identification of clusters, but conclusions about clustering are usually drawn by informal visual examinations of the network rather than by more formal inference methods (Liotta, 2004).

The idea of introducing a social network by allocating positions in a continuous space to the nodes was presented in the 1970s; see, for example, McFarland and Brown (1973), Faust (1988) and Breiger *et al.* (1975), who used multidimensional scaling and this approach has been broadly used since (Wasserman and Faust, 1994). A strong point of this approach is that it considers transitivity automatically in a natural way. On the other hand dissimilarity measure must be supplied to the algorithm for each dyad, and a lot of various dissimilarity measures are possible, so the outcome depends on a choice for which there is no clear theoretical reference.

Consider any two randomly selected individuals A and B, and the set, $S = C, D, E, \dots$, of all persons with connections to either or both of them. The hypothesis which allows us to refer dyadic ties to bigger framework is: the stronger the arc between A and B, the larger the proportion of individuals in S to whom they will both be connected, that is, connected by a weak or strong arc. This overlap in their friendship circles is expected to be least when their connection is absent, most when it is strong, and intermediate when it is weak.

In the figure 2.1 we can observe an example of weak tie in a social network. Consider individuals A, B, C, representing the vertices x, y, z in a graph D . Then, A and B are linked via a strong tie, if and only if $xy \in E(D)$, which corresponds to the existence of a direct link between

A and B. On the contrary, there will be weak tie between B and C that is not directly but indirectly present.

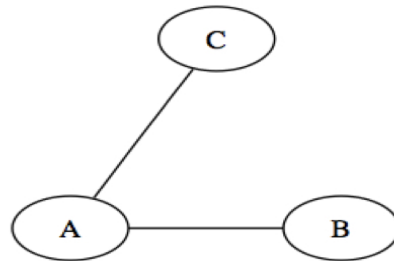


Figure 2.1: Strong and weak ties in a social network

If A is strongly tied to both B and C, there must exist a connection between B and C, be it a weak or a strong one. If there is a strong connection between B and C, then $yz \in E(D)$ must be true, otherwise B and C are linked indirectly via A. The absence of this connection, Mark S. Granovetter in his article “The Strength of Weak Ties” called the forbidden triad. In the same article he also proves, that weak ties are often much more important for the length of the chain between two randomly chosen vertices in the graph, since strong ties primarily exist inside very small subsets of close to each other persons in the net and therefore do not contribute to short paths between persons in these cliques and others outside them. Weak ties instead, form the “bridges” between the cliques and thus contribute to shorter paths between arbitrarily chosen individuals. Social network analysis takes into account different attributes such as statistical analysis of pathways via networks with changing degrees of the structure, the qualitative description of the structure of networks in terms of local (clustering) and nonlocal (weak ties) features, readjustment of highly clustered complicated networks into equivalent sub networks.

The social network theory reviews the different benefits of social capital that emerge from different social network structures. The two main theories in this field are Burt’s (1992) structural hole theory and Coleman’s (1990, 1988) social closure theory. Both theories propose that social capital is a product of the relationships between people in social networks; however, each theory takes an opposite argument on which social network characteristics are engaged in

the formation of social capital.

Using the terminology of network sociology, each individual is an actor or node in the network. When referring to a social network from a particular actor's frame of reference, that actor is referred to as the ego, and that social network is his or her ego network. In an ego network, any other actor connected to the ego is called an alter. The edges between nodes represent links or relationships between actors. These links can characterize any type of relationship between actors within a social network. The frequency of interaction between actors defines the strength of the relationship (Wasserman and Faust, 1994).

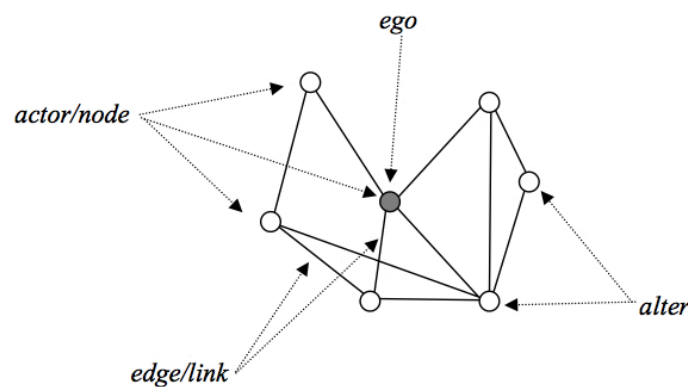


Figure 2.2: Example of ego network

2.1.1 Social Capital

Social capital is one of the key concepts in the theory of social network. It is one of the productive outcomes of social networks (Gabby and Leenders, 2001). Although there is no single definition of social capital (Burt, 2000, Gabby and Leenders, 2001), pioneering researchers, such as Coleman (1990), Burt (1992), and Putnam (1993), have developed several key descriptions and theories of social capital. According to Coleman (1990, p.203), social capital is defined as “some aspect of social structure, facilitating certain actions in individuals who are within the structure”. Putnam (1995, p.67) defines social capital as “the features of social organization, such as networks, norms and social trust that facilitate coordination and

cooperation for mutual benefit”. Similarly, Bourdieu (1997, p.49) states that “social capital is the sum of the resources, actual or virtual, that accrue to an individual or group by virtue of possessing a durable network of more or less institutionalized relationships of mutual acquaintance and recognition”.

There are two proposed perspectives on the nature of the formation of social capital (Gabby and Leender, 2001); the first views the formation of social capital as a natural or developmental process, and the second views its creation as a deliberate process. The first one occurs due to the normal social reciprocity between persons who may enter or leave these informal social networks. The second proposes that individuals display strategic behavior by looking for connections to create social capital for their own benefit (Coleman, 1990; Burt, 1992; Gabby and Leender, 1999).

2.1.2 Structural Hole Theory

There are two clearly conflicting theories about how social networks create social capital, despite the fact that it is obvious with the social capital being a productive output of social networks (Gabby and Leenders, 2001).

The first theory is Burt’s (1992) structural hole theory, which is also referred to as the instrumental approach by Johanson (2001). Burt (1992) proposed that structural holes in the network act as bridges that connect people in the network to new information, knowledge and resources. According to the author the structural hole is a relationship of non-redundancy between two nodes in a network. Since these nodes are linked to each other by only one path, each of their resources can provide advantages to the other that are additive. When the relationship between nodes is redundant, they are linked by multiple actors within the network, and can access the same resources via these multiple paths, therefore provide overlapping resource benefits.

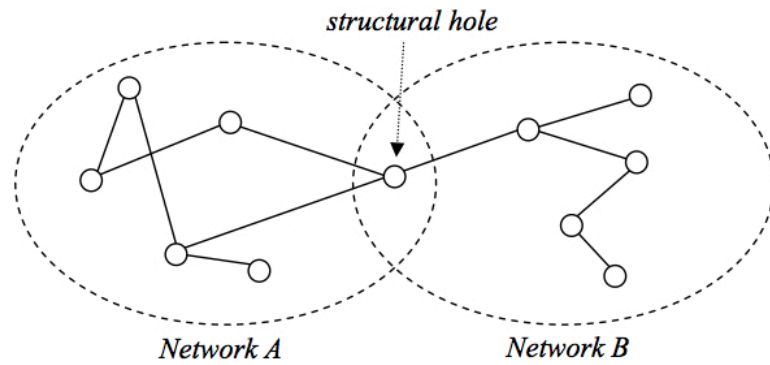


Figure 2.3: Example of sparse network

According to Burt (1992), certain actors within the network serve as connecting bridges that provide access to non-redundant resources and information to other members of the network that would otherwise be inaccessible to these members. This fact provides these members the opportunity to gain control in a competitive environment. The actor that fills the position of a structural hole is the only point of contact between two otherwise disjoint networks, as shown in Figure 2.1, and is therefore in an advantageous position to use these opportunities.

2.1.3 Social Closure Theory

There is an opposite theory by Coleman (1998, 1990), who proposes the social closure theory, which Johanson (2001) refers to as the expressive approach to social capital. The social closure theory proposes that cohesive networks can provide a greater value because the members are more tightly bound by a level of trust that leads to the assumption that members of the group will help each other by sharing knowledge and resources (Figure 2.4). The key elements of Coleman's social closure theory are trust, expectation, and reciprocal obligation.

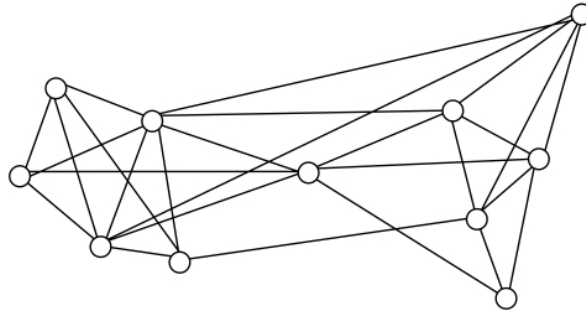


Figure 2.4: Example of dense network

2.1.4 Milgram's Small World Experiments

One of the first major achievements in the study of social networks occurred with the innovative experimental approach that was developed by Stanley Milgram in the 1960's [126, 170]. His goal, more or less, was to experimentally determine the length of the shortest chain of friends between a typical pair of people in the United States. In the best-known small-world experiment, Milgram selected a target person and a larger number of starting persons. All starting persons were informed of the identity of the target person, and asked to contact someone who might know that person, or know someone who might. They were strictly informed not to contact the target person directly, unless they knew that person on a personal basis.

The original Milgram experiment was plagued by some serious methodological problems, most seriously with many incomplete chains and the possible problem of non-symmetry of the chains, but the findings are widely cited nonetheless. The primary finding of that experiment was that the median number of intermediaries between starting and the target person was 5 (mean of 5.2) varying from two to ten (Travers & Milgram, 1969).

This was not, however, the most significant finding, even though the primary point of the five intermediaries has evolved into the academic equivalent of an urban myth under the popular heading of six degrees of separation, spawning websites, games and influencing popular culture.

Milgram (1967) points to the following: (i) In those chains, participants were three times more likely to contact people of the same gender. (ii) Many chains passed through the same few people – stars – as the last step before reaching the target person. (iii) Although physical distance is important in the chains of intermediaries, social distance seems even more important.

3. Social Network Analysis Based on Personal Data Examples

As was already mentioned the main part of the research is divided into two separate, but interrelated analyses.

3.1 Methodology and Data

Now after the social-network models were defined, I will analyze the structure of my own social network based on the popular social networking websites as Facebook, LinkedIn and Wolfram Alpha. Mathematically, a social network is just a graph $\langle V, E \rangle$, where the nodes represent people, and an edge $\langle u, v \rangle \in E$ denotes some type of social relationship between the people u and v . A social network with a single type of connection between actors can be represented using an adjacency matrix A whose elements A_{ij} give the weight of the link between nodes i and j . The Facebook and LinkedIn networks analyzed here are weighted, so $A_{ij} \in \{0,1\}$, where the value is 1 if a tie exists and 0 if it does not.

3.2 Facebook Social Network

With over 1.4 billion users Facebook is the largest online social network representing a huge business. To examine my own network represented in Figure 3.1, I used a tool from Wolfram Alpha to get an extensive and personalized analytics report of my profile on Facebook. One can analyze its own data using different tools or apps, but I used Wolfram Alpha due to its reliability and popularity. Wolfram Alpha is a computational search engine that is known for, among other things, being part of the brains that power Apple's voice-assistant, Siri. After computing my data, I had a detailed and interactive report of my Facebook life. There were plenty of informative graphs with interesting details of my network.

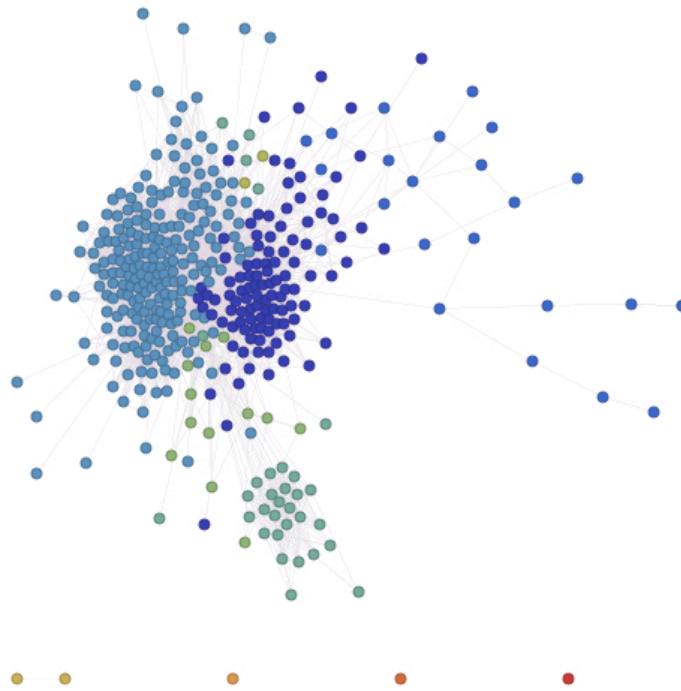


Figure 3.1: The author's Facebook network. A statistical analysis of the link pattern (grey lines) reveals information on the Facebook friends (circles). The algorithm colors the circles, examination of the coloring reveals that the clusters are easily interpretable.

From the figure 3.1 we can clearly observe that the statistical model places the friends in the social space. Those individuals who are located close together tend to connect, and friends far apart tend not to. The model groups into subsets of highly connected friends together. From the figure we can also observe that friends in groups assigned to their own subset color so that they most likely belong to some probabilistic model.

Even though I haven't mentioned any additional data, the model was able to pick up six main friend clusters. The description of the groups according to the color is shown on the figure 3.2.

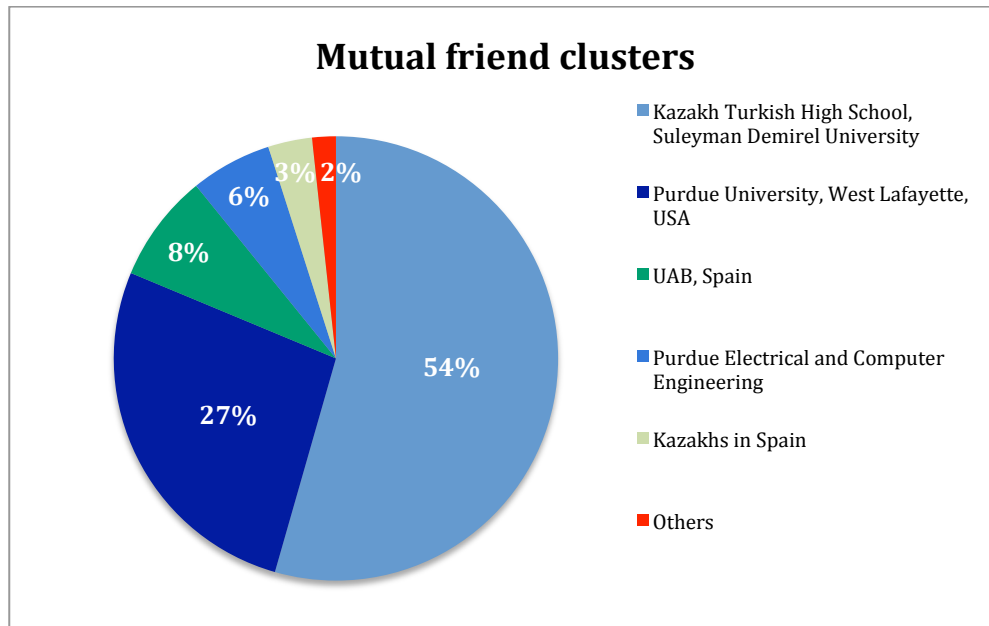


Figure 3.2: Mutual friend clusters

The groups in my network are those friends who studied with me or related to me via Kazakh Turkish High School and Suleyman Demirel University. This is the largest group in my Facebook network and they form 54%, which are 221 friends. The next largest cluster contains friends from Purdue University and other friends from the city of West Lafayette and USA. They form 27% (109) of my Facebook connections. Then my UAB and Spanish friends follow with 8% (32). The rest are friends from Electrical and Computer Engineering department at Purdue (they were not included into the USA cluster) and Kazakhstani friends in Spain with 6% (24) and 3% (13) respectively. I should note that Wolfram Alpha could analyze only 406 out of 474 my Facebook friends. The reason for this is that not all of my friends' Facebook profiles allow analyzing their data.

Besides this mutual friends cluster information, we can analyze a visualization that highlights friends based on the way that they fit into my network. On the figure 3.3 we can observe a color-coded friend network based on their «network roles». There are five different roles Wolfram Alpha identifies, such as: social insiders, social outsiders, social neighbours,

social gateways and social connectors. On the figure 3.4 you can observe the top friends of each of the roles described.

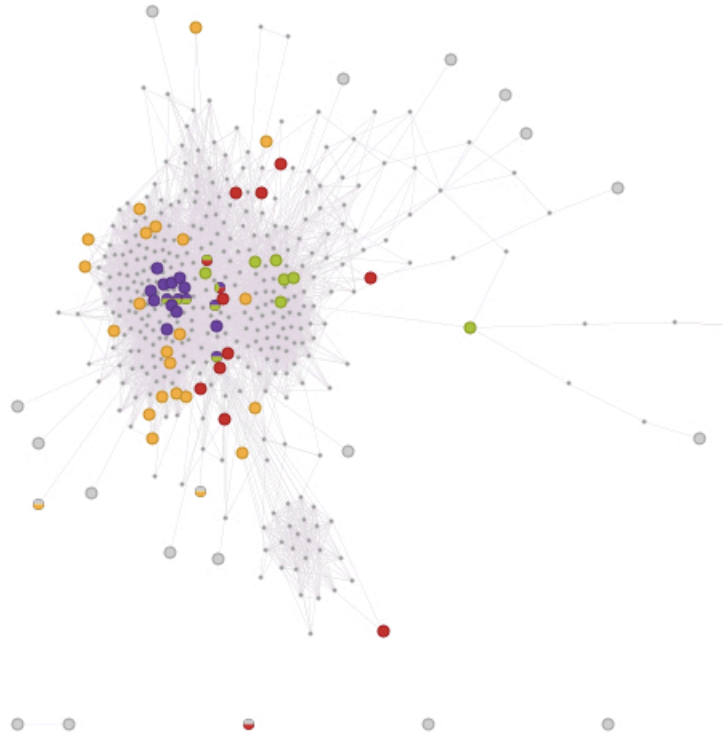


Figure 3.3: Color-coded friend network

■ top "social insiders":
Rinat Tainov Yerlan Sapargaliyev Shyngys Khan Danday Amokachi Azat Bora Beibit Bashabayev Anuar Adilkhan ...
■ "social outsiders":
Tyler Mantel Mark Powell Sergey Kurdyukov Jay Chen Sam Lahti Federico Parker Infantozzi Renat Orazbay ...
■ top "social connectors":
Yerlan Sapargaliyev Temirlan Dauletaliyev Azat Bora Askar Almaty Arman Batayev Lucas Sun Sayasat Nurbek ...
■ top "social neighbors":
Ержан Конырбаев Stanislav Sabiev Elaman Muhtarhanov Didar Kassenov Bauyrzhan Kurmangaliyev Dias Mukashev Данияр Рамазанов ...
■ top "social gateways":
Amirkhan S Omarov Aizhan Hamit Igor Filevski Actual Positions Aruzhan Sain Asset Abdualiyeu Leila Rakhimjanova ...
(based on data from 406 of 474 friends)

Figure 3.4: Highlighted friends

Social insiders and outsiders are opposites: a social insider has a lot of friends in common with me (since I am analysing my own network); conversely, a social outsider is someone with

whom I have few or no mutual friends. Social gateways and neighbours are also opposites: a social gateway contact has a lot of friends that are outside your network, whereas a social neighbour has few friends outside your network.

Social insiders and social neighbours sometimes overlap, as do gateways and outsiders—but not always. For example, a social outsider simply may not have enough outside friends to be a social gateway. Social connectors are those who are also connected to many separate groups of your networks. The good example of social connector in my case is my friend Yerlan Sapargaliyev, who was my classmate in high school, studied with me at Purdue University and Universitat Autònoma de Barcelona – the main three clusters in my social network.

On the figure 3.5 we can observe a graph of my Facebook friends according to the gender. Blue dots correspond to male friends while red ones correspond to female friends. The structure of the disposition of clusters remains the same on all of the graphs.

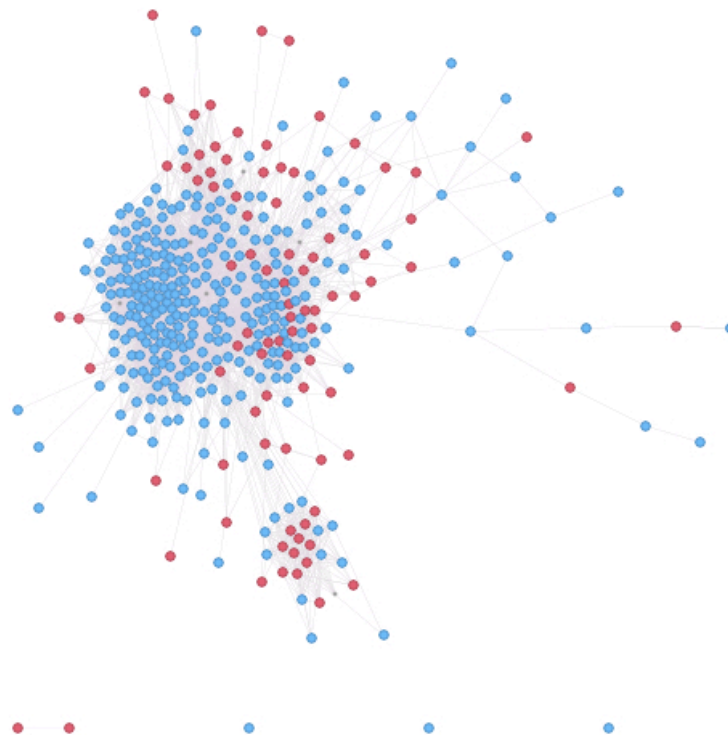
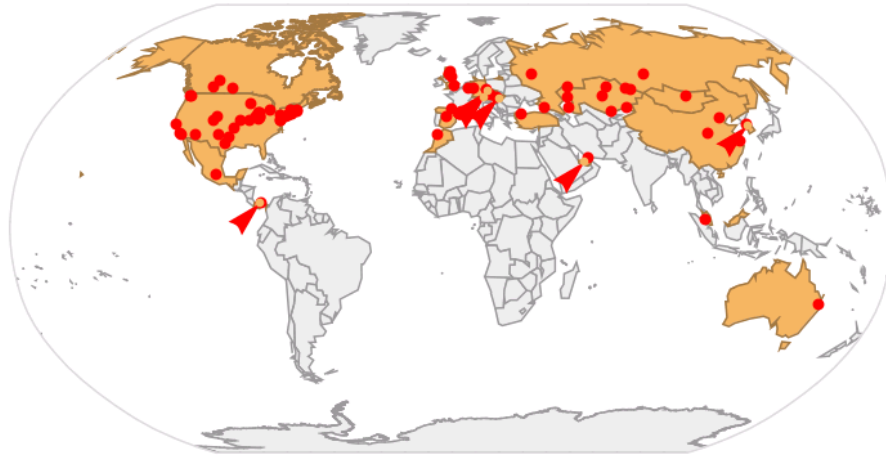







Figure 3.5: Color-coded friend network by gender

Also on figure 3.6 I could analyse the data according to geographical relationship to me, based on their hometown or current locations reported.



country	number of friends
 Kazakhstan	128 (45.1%)
 United States	101 (35.6%)
 Spain	21 (7.4%)
 Russia	6 (2.1%)
 United Kingdom	6 (2.1%)
other	22 (8.4%)

(77 cities , 57 states , 20 countries)

(based on data from 284 of 466 friends)

Figure 3.6: Friends' locations

3.3 LinkedIn Professional Network

With over 230 million users in over 200 countries and territories around the globe LinkedIn is the largest online platform for professional networking. The mission of this social network is to connect the world's professionals to make them more productive and successful by providing access to people, jobs, news, updates, etc. Using the <http://inmaps.linkedinlabs.com> I could visualize representation of my own professional universe, based on the relationships between the connections. LinkedIn uses information about how people in my network are connected to me and each other to create my personalized map. Groups like colleagues, people you went at school with, or friends are separated into color-coded clusters, as people within these groups are also interconnected with each other. With the InMap, it is easy to visually understand how to better leverage your professional network to help pass along job opportunities, seek professional advice, or gather insights. The result is shown in figure 3.7.

Again, as in Facebook example, here we can observe different separated clusters of professionals. The large blue group is colleagues from high school and Suleyman Demirel University. The orange group is college friends and colleagues from USA. The green group on the right is friends from UAB and Spain. The purple group is ECE professionals. The small grey group is random isolates, not connected to anyone but me.

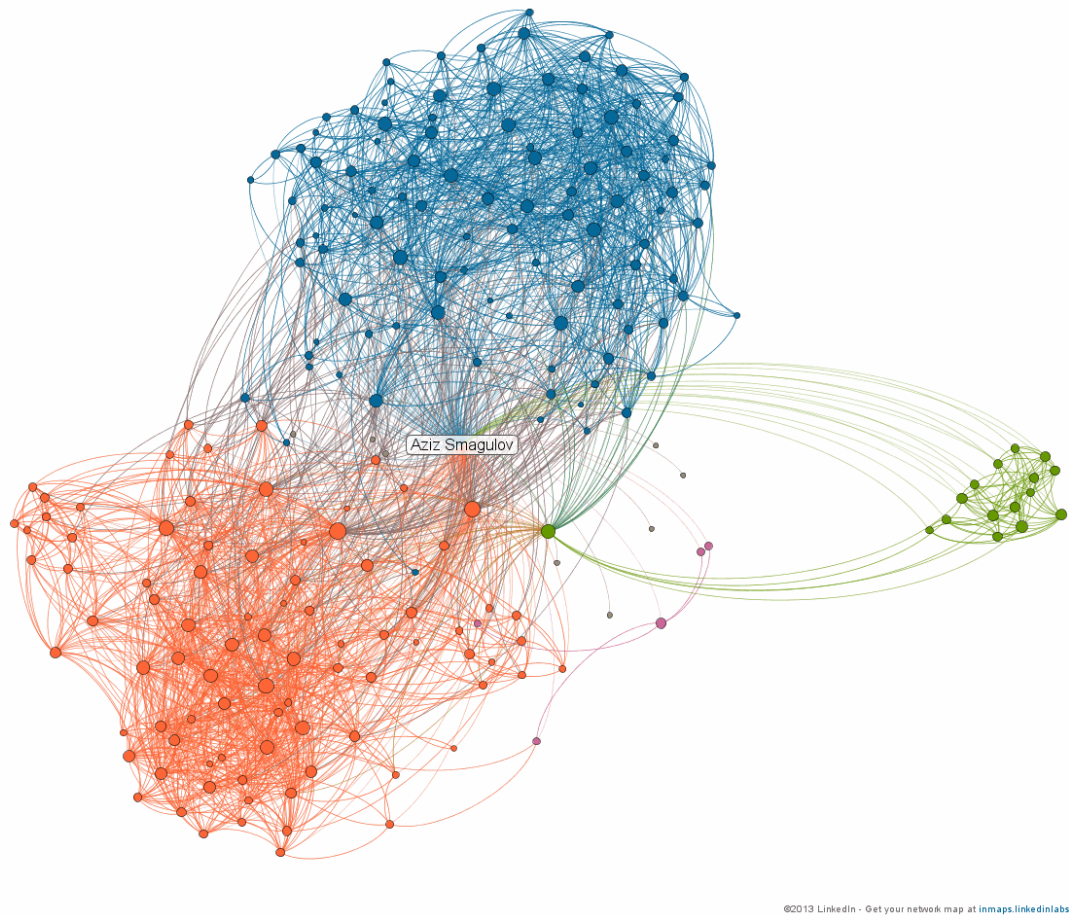


Figure 3.7: The author's LinkedIn professional network. The pattern of linkage is quite different from the Facebook network in figure 3.1. A different algorithm is used, but sensible groupings are found.

3.4 SNA Summary

Using some of the basic tools of social network analysis I was able to analyse the structure of author's social networks and visualize one example of how personal network of friends and colleagues look like. This analysis of any individual network might look very interesting, and at the same time this example works as a reminder of how much information we willingly share with social networks. It could inspire some people to clamp down on the amount of information they share online, and others to record even more and embrace all the interesting possibilities of tracking personal data.

4. Empirical Analysis of Social Networks based on Case Study Examples

To make example of how the information from SNA can be used for researchers and answer the next research question: “Do social networks affect individual’s behaviour?” the following sub question and hypothesis was studied based on data availability constrain:

Do social networks affect the frequency of smoking among students in secondary school?

Hypothesis: The frequency of smoking among secondary school students is highly correlated to the frequency of smoking among his/her friends.

4.1 Methodology and Data

There is a common idea that our social network may positively or negatively affect our behaviour. Moreover, parents have been always concerned about their kids' environment, doubting whether the bad behaviour of their children were copied from their peers or adapted from some other sources. Peer influence, peer selection, and peers being subject to the same contextual influences are three main types of processes, which may explain similarity between youngsters.

In this research I decided to concentrate on secondary school students based on the data retrieved from Data Archiving and Networked Services website for 2003-2004 in Dutch secondary schools. There are several reasons for choosing secondary-school first-graders as a study population. First, there were not many relevant data on SNA available for my research. Second, in early age, peers are very important for the development of character (e.g., Giordano, 2003; Steinberg & Sheffield Morris, 2001), and it is very interesting to analyze their influence on each other. Third, in many European countries and in this case as well, the school system is designed that way as the majority of students who enter secondary school at the age of 12 lose many of their former classmates and meet their new peers for the first time. A strong activity in new friendship formation is observed. Fourth, classmates spend most of their time at school

together and the classroom activities play important role in behavior, socialization and identity formation (Kassenberg, 2002).

So, the found data was the secondary source data collected from 3,171 students in 126 first-grade classrooms in 14 secondary schools in the Netherlands. The data collection was performed using a standardized questionnaire. Six of the 126 classes did not participate in all waves or a high number of students were missing on the day of the survey, resulting in the end with 120 classes and 2932 students. Correlation analysis and regression results from SPSS software we conducted later on and presented in the next section.

4.2 Measured Variables

As mentioned above, self-reported questionnaires were used as the instrument for collecting the data. The response rate was about 94%.

- *Friendship.* Students were asked to indicate up to 12 best friends in class. Therefore the information about the entire friendship network was obtained.
- *Control variables.* 49% of the students were girls. Dutch students were in the majority (83%). Being Dutch is defined as having at least one parent who was born in the Netherlands.
- *Background variables.* The students were on average 12 years and 1 month old at the first data collection. Socioeconomic status is assigned based on the father's and mother's job. Most of the students have most of their friends and the most important friends in their class.
- *Smoking.* The information about frequency of smoking was based on the question "How often did you smoke with friends in the last three months?" There were five answer categories: "never", "once", "two to four times", "five to ten times" and "more than 10 times".

4.3 Findings

From the output table of frequency of smoking between students in secondary school in the figure we can observe that the majority of students (82%) answered that they never smoked in the last three months. 6.5% indicated that they had smoked once, 3.6% smoked from two to four times, 1.4% smoked from five to ten times and 5.3% smoked more than ten times. The response rate for this particular question was 99.4% with missing 18 students out of 2932.

r. smoked in the last three months

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	never	2423	82,6	83,2	83,2
	one time	192	6,5	6,6	89,7
	two to four times	105	3,6	3,6	93,3
	five to ten times	40	1,4	1,4	94,7
	more than ten times	154	5,3	5,3	100,0
	Total	2914	99,4	100,0	
Missing	9	2	,1		
	System	16	,5		
	Total	18	,6		
Total		2932	100,0		

Figure 4.1: Frequency of smoking between students in the last three months

	Coding	Mean or %	Std. Deviation	Missing in %
<i>Dependent variables</i>				
Friends of respondent	1-12	4,06	2,51	7,2
Smoking	1: never - 5: > 10 times	1,76	1,23	7,8
<i>Control variables</i>				
Gender	1: female 2: male	49%		0,0
Ethnicity	1 native 2: non-native	83%		0,0
<i>Background variables</i>				
Age	10 - 15 years	12,11	0,49	5,6
Socio-economic status	1: low - 4: high	2,56	0,93	31,2

Figure 4.2: Descriptive statistics

After playing with data and adjusting it to a right way, we were able to answer the following question: is there a correlation between smoking ego and smoking network?

```
*compute average smoking.
comp av_smoke=mean(alter_smoke.1 to alter_smoke.12).
match files/file=* /keep id actsmokd alter_smoke.1 to alter_smoke.12 av_smoke
all.

*correlation between smoking ego and smoking network?
corr actsmokd av_smoke.
```

Correlations

Correlations		r. smoked in the last three months	av_smoke
r. smoked in the last three months	Pearson Correlation	1	,326
	Sig. (2-tailed)		,000
	N	2914	2760
av_smoke	Pearson Correlation	,326	1
	Sig. (2-tailed)	,000	
	N	2760	2774

Figure 4.3: Correlation between smoking ego and smoking network

From the figure we can clearly see that there is highly significant, strong and positive correlation of 33% between frequency of smoking of individual and frequency of smoking among his/her friends. Therefore our hypothesis is supported.

	Estimated mean parameter	Standard error of mean parameter	Estimated true std. deviation	p-value of test that variance of parameter is 0
<i>Selection Part</i>				
Shared level of smoking	0,057***	0,012	<0,001	0,981
Smoking ego	-0,019	0,013	<0,001	0,087
Smoking alter	-0,015	0,013	<0,001	0,372
Outdegree	-1,965***	0,035	0,22	<0,001
Reciprocity	0,960***	0,043	0,27	<0,001
Transitivity	0,156***	0,005	0,02	0,020
Gender similarity	0,759***	0,059	0,43	<0,001
Ethnicity similarity	0,148***	0,031	0,03	0,567
Friends in primary school	0,569***	0,035	<0,001	0,031
<i>Influence part</i>				
Average smoking of friends	0,135	0,078	<0,001	1,000
Tendency smoking	-0,822***	0,044	<0,001	<0,001
Tendency smoking squared	0,270***	0,019	<0,001	0,790
Male	0,047	0,069	<0,001	0,999
Non-native	-0,049	0,106	<0,001	0,999

Significance * $p < 0,05$; ** $p < 0,01$; *** $p < 0,001$

Figure 4.4: Estimated Results Table

Analyzing results on Figure 4.4, we can say that estimated mean parameters with standard error, estimated standard deviations of the parameters corrected for the standard errors of the estimates per class, and *p-values* of the tests that the parameter variance is 0 are reported. Most of the effects in this model are significant, as most parameter estimates are more than 1.96 times their standard errors, indicating significance at the 5% level.

In the selection part of the model we tested the hypothesis that teenagers in secondary school with high frequency of smoking to be friends with peers being a high level smokers and vice versa. The *shared level of smoking use* parameter that represents this hypothesis has a positive value (0.057). The combination of personal and others' level of smoking barely counts for friendship choices. The effects of *smoking ego* and *smoking alter* are not significant. The *outdegree* effect is negative (-1.965), representing that friendship ties do not develop casually. The other effects in the selection model are all significant and in the predicted way. Secondary school students have a tendency to form and maintain reciprocated (0.960) and transitive (0.156) friendship ties. Resemblance in ethnicity and gender contribute friendship creation and retention. Male students tend to become friends with each other, the same with female students; Same ethnicities tend to choose friends among their own ethnicity and vice versa. The effect of *ethnicity similarity* (0.148) is weaker than the effect of *gender similarity* (0.759). The friendship bonds originating in primary school are likely to be retained in secondary school (0.569).

In the influence part, we tested the hypothesis that the level of smoking dominating among the peers affects students' own level of smoking. This hypothesis is not confirmed; the coefficient has the predicted, positive direction but falls short of significance ($p=0.08$, two-sided). We found a negative *tendency smoking* parameter (-0.822). Due to the fact that the variable is centered, based on the observation period, students are more likely in the long run to score on the low end of the smoking scale than on the high end of it. A strong significant and positive *tendency smoking squared* effect (0.270) indicates a positive feedback effect. Those

students who scored high on the smoking variable tend to score even higher. Smoking has an accelerating effect. Those who previously did not smoke tend to hold back from it, but students who previously smoked incline to smoke increasingly more. The effects of *male* and *non-native* in the influence model are not supported by our sample.

As shown by the *p-value* ($p < .05$), effects vary significantly between classes for *outdegree*, *reciprocity*, *transitivity*, *gender similarity*, *friends in primary school*, and the *tendency smoking* parameter. The standard deviations are typically much lower than the mean estimate, so there is no proof that the direction of the estimate varies between classes. The shared *level of smoking* and influence (*average smoking of friends*) do not change between classes. This indicates that the relation between smoking and friendship choices is not affected by classroom characteristics such as educational level or neighborhood.

In this research we explored the effects of smoking on friendship selection and the influence effects of friends' smoking on secondary school students' smoking. We proposed that similar students with same interests are more likely to become friends. For the influence process we hypothesized that young students' level of smoking is positively correlated by the average level of their friends' smoking.

Our hypothesis that positively related levels of teenagers' smoking directs to friendship formation and maintenance was confirmed. Adolescents with high levels of smoking tend to have friends with high levels of smoking as well. Students with low levels tend to have friends with low levels. Further, we found no support that level of smoking affects the number of friendship nominations. Similarity smoking is not the only factor affecting friendship selection: structural effects such as reciprocity and transitivity, interpersonal effects like similarity in gender and ethnicity, and having been friends in primary school influence friendship choices positively. Also there is weak evidence ($p < 0.10$) that students adjust their level of smoking to their friends' average level. The positive effect is as expected. Our results show that students

have tendency to have low levels of smoking and that student's earlier smoking has a reinforcing effect on latter smoking behavior. Our data did not support the effect of gender and of ethnicity on level of smoking.

5. Conclusions

5.1 Concluding Remarks

Apart from describing the Social Network Analysis in the first part of the research, in this study we explored the effects of smoking on friendship selection and the influence effects of friends' smoking on secondary school students' smoking. We proposed that similar students with same interests are more likely to become friends. For the influence process we hypothesized that young students' level of smoking is positively correlated by the average level of their friends' smoking.

Our hypothesis that positively related levels of teenagers' smoking directs to friendship formation and maintenance was confirmed. Adolescents with high levels of smoking tend to have friends with high levels of smoking as well. Students with low levels tend to have friends with low levels. Further, we found no support that level of smoking affects the number of friendship nominations. Similarity smoking is not the only factor affecting friendship selection: structural effects such as reciprocity and transitivity, interpersonal effects like similarity in gender and ethnicity, and having been friends in primary school influence friendship choices positively. Also there is weak evidence ($p < 0.10$) that students adjust their level of smoking to their friends' average level. The positive effect is as expected. Our results show that students have tendency to have low levels of smoking and that student's earlier smoking has a reinforcing effect on latter smoking behavior. Our data did not support the effect of gender and of ethnicity on level of smoking. The results are somehow similar to the research made by Blansky et al (2013) based on Spanish high schools students sample. They also proved that social network influentially correlates with the behavior and academic achievements, as another example, of K-

12 students.

So, together with describing the structural analysis of social networks and how do they work in terms of nodes and ties, the empirical case study helps to understand the importance of SNA to each individual and the use of social network information in analyzing their behavior , choices, and influences within the network.

5.2 Limitations and Further Research Recommendations

This study has some limitations as well. First, in this particular data we focused only on friends within school classes, but it is not possible to analyze the large samples of social networks in detail. Although relationships within a classroom play an important role for secondary school students, they do have other peers outside of the school, who should not be disregarded. Second, the narrowing research question to subquestion is also one kind of a limitation, and the data on the level of smoking is based on frequency of smoking with friends and not on the amount of smoking without friends. Further research might provide more insight into the different types of cigarettes and include the cases when students smoke cannabis or drink alcoholic beverages and many other minor and major influences within different social networks that can be used for various subjects.

References

- Backstrom, L., Huttenlocher, D., Kleinberg, J., and X. Lan. Group formation in large social networks: Membership, growth, and evolution. In 12th KDD, pages 44–54, 2006.
- Boyd, D. and Nicole B. Ellison. Social network sites: Definition, history, and scholarship. *Journal of Computer-Mediated Communication*, 13(1), 2007.
- Breiger, R. L. (1974) The duality of persons and groups. *Soc Forces*, **53**, 181–190.
- Burt, R.S. (1992). *Structural Holes: The Social Structure of Competition*, Cambridge, MA: Harvard University Press.
- Burt, R.S. (2000). “The Network Structure of Social Capital”, *Research in Organizational Behaviour*, Vol. 22, Elsevier Science.
- Coleman, J.S. (1988). “Social Capital in the Creation of Human Capital”, *American Journal of Sociology*, Vol. 94, pp. S95-S120.
- Coleman, J.S. (1990). *Foundations of Social Theory*, Cambridge, MA: Harvard University Press.
- Faust, K. (1988) Comparison of methods for positional analysis: structural and general equivalence. *Soc Networks*, **10**, 313–341.
- Freeman, L. C. (1996) Some antecedents of social network analysis. *Connections*, **19**, 39–42.
- Gabbay, S.M., & Leenders, R.A.J. (2001). “Social Capital of Organizations: From Social Structure to the Management of Corporate Social Capital”, in Gabbay, S.M. and Leenders, R.A.J. (Ed.), *Research in the Sociology of Organizations*, Vol. 18, pp. 1-20, Stamford, CT: JAI Press.
- Giordano, P. C. (2003). "Relationships in adolescence." *Annual Review of Sociology* **29**: 257-281.
- Hartup, W. W. (1996). "The company they keep: Friendships and their developmental
- Granovetter, M. (1973) *The strength of weak ties*. *The American Journal of Sociology* Volume 78, Issue 6 (May, 1973), 1360-1380.
- Handcock, M.S., Raftery, A.E., and Tantrum, J.M. (2007) Model based clustering for social

networks. *Journal of the Royal Statistical Society, Series A*, 170 (2), 1-22.

Johanson, J. (2001). "The Balance of Corporate Social Capital", in Gabbay, S.M. and Leenders, R.A.J. (Ed.), *Research in the Sociology of Organizations*, Vol. 18, pp. 1-20, Stamford, CT: JAI Press.

Kassenberg, A. (2002). *Wat scholieren bindt. Sociale gemeenschap in scholen [What binds pupils? Social communities in schools]*. University of Groningen, Groningen.

Knecht, A. (2006), Network and actor attributes in early adolescence [2003/04]. Utrecht, The Netherlands Research School ICS, Department of Sociology, Utrecht University. (ICS-Codebook no. 61).

Lazarsfeld, P. and Merton, R. (1954) Friendship as social process: a substantive and methodological analysis. In *Freedom and Control in Modern Society* (eds M. Berger, T. Abel and C. Page), pp. 18–66. New York: Van Nostrand.

Liotta, G. (ed.) (2004) Graph drawing. *Lect. Notes Comput. Sci.*, **2912**.

Marlow, C., M. Naaman, D. Boyd, and M. Davis. Ht06, tagging paper, taxonomy, Flickr, academic article, to read. In 17th HYPERTEXT, pages 31–40, 2006.

McFarland, D. D. and Brown, D. J. (1973) Social distance as a metric: a systematic introduction to smallest space analysis. In *Bonds of Pluralism: the Form and Substance of Urban Social Networks* (ed. E. O. Laumann), pp. 213–253. New York: Wiley.

McPherson, M., Smith-Lovin, L. and Cook, J. M. (2001) Birds of a feather: homophily in social networks. *A. Rev. Sociol.*, **27**, 415–444.

Milgram, S. (1967). "The small-world problem". *Psychology Today*, 2, 60-67.

Travers, J., & Milgram, S. (1969). "An Experimental Study of the Small World Problem". *Sociometry*, 32(4), 425-443.

Salter-Townshend, M. (2012) Analysing my facebook friends. *Significance*, August 2012.

Steinberg, L., & Sheffield Morris, A. (2001). Adolescent development. *Annual Review of Psychology*, 52, 83–110.

Traud, A.L.; Kelsic, E.D., Mucha, P.J., and Porter, M.A. (2011) comparing community structure to characteristics in online collegiate social networks. *SIAM Review*, 53 (3), 526-543.

Wasserman, S. S. and Faust, K. (1994) *Social Network Analysis: Methods and Applications*. Cambridge: Cambridge University Press.

White, H. C., Boorman, S. A. and Breiger, R. L. (1976) Social structure from multiple networks: I, Blockmodels of roles and positions. *Am. J. Sociol.*, **81**, 730–780.

Software Resources

Data Archiving and Networked Services Database

LinkedIn Labs (2013) Available from <http://inmaps.linkedinlabs.com>

Wolfram Alpha

IBM SPSS