

Piela, Jonas

**Working Paper**

## The social network's influences on individual performance

Discussion Papers, No. 12/2011

**Provided in Cooperation with:**

Witten/Herdecke University, Faculty of Management and Economics

*Suggested Citation:* Piela, Jonas (2011) : The social network's influences on individual performance, Discussion Papers, No. 12/2011, Universität Witten/Herdecke, Fakultät für Wirtschaftswissenschaft, Witten

This Version is available at:

<https://hdl.handle.net/10419/49951>

**Standard-Nutzungsbedingungen:**

Die Dokumente auf EconStor dürfen zu eigenen wissenschaftlichen Zwecken und zum Privatgebrauch gespeichert und kopiert werden.

Sie dürfen die Dokumente nicht für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, öffentlich zugänglich machen, vertreiben oder anderweitig nutzen.

Sofern die Verfasser die Dokumente unter Open-Content-Lizenzen (insbesondere CC-Lizenzen) zur Verfügung gestellt haben sollten, gelten abweichend von diesen Nutzungsbedingungen die in der dort genannten Lizenz gewährten Nutzungsrechte.

**Terms of use:**

*Documents in EconStor may be saved and copied for your personal and scholarly purposes.*

*You are not to copy documents for public or commercial purposes, to exhibit the documents publicly, to make them publicly available on the internet, or to distribute or otherwise use the documents in public.*

*If the documents have been made available under an Open Content Licence (especially Creative Commons Licences), you may exercise further usage rights as specified in the indicated licence.*

**discussion papers**  
Fakultät für Wirtschaftswissenschaft  
Universität Witten/Herdecke

Neue Serie 2010 ff.  
Nr. 12 / 2011

**The Social Network's Influences On  
Individual Performance**

**Jonas Piela**

**discussion papers**  
Fakultät für Wirtschaftswissenschaft  
Universität Witten/Herdecke  
[www.uni-wh.de/wirtschaft/discussion-papers](http://www.uni-wh.de/wirtschaft/discussion-papers)

Adresse des Verfassers:  
Jonas Piela [jonas.piela@gmail.com]

Redakteure  
für die Fakultät für Wirtschaftswissenschaft

Prof. Dr. Michèle Morner / Prof. Dr. Birger P. Priddat

Für den Inhalt der Papiere sind die jeweiligen Autoren  
verantwortlich.

# The Social Network's Influences On Individual Performance.

**Are your friends responsible for your marks at university?**

## 1 Introduction

In the past century, social network analysis has gained significant importance not only in but also beyond the field of sociology. According to Freeman in his 2006 overview of social network research “social network analysis is currently very ‘hot’”. The most significant increase in publications can be seen in the last 30 years starting off when the first Journal “Social Networks” was published in 1978. Social network analysis “is finally succeeding in providing an alternative to the traditional individualism of most mainstream social research”. ([Freeman 2006](#)).

The concept of social networks was initially used in the early days of the 19<sup>th</sup> century with the first scholar “that proposed a way of looking at society in terms of the interconnections among social actors” ([Freeman 2006](#)). Émile Durkheim and Ferdinand Tönnies continued in the late 19<sup>th</sup> century with the idea that groups are interlinked through personal and direct ties and therefore do not consist of individual properties but exist in their own reality. Throughout the early years of the 20<sup>th</sup> century and especially around the 1950s it has been relatively quiet in social network research. Only Milgram's small world experiment could disrupt the silence in 1967. With this experiment on the social separation of men his article gained worldwide acceptance. In its basic idea the experiment consisted of several letters that were sent out to random people in the United States. Each person was randomly assigned another person unknown to that first person. The first person was instructed to try to send this letter to a friend that might know the target person. Finally, the chain of friends the letter took to reach the target person was examined. As a result Milgram found that each person on the planet is separated by *six degrees of separation* ([Milgram 1969](#)). In 1973 Mark Granovetter published his widely cited work “The strength of weak ties” examining the information flow between actors in a social networks. Weak ties are ties that connect actors that rarely see or speak to each other but still know each other. He found out that especially between weak ties the flow of new and valuable information is higher as there is little overlapping between information and information sources. Granovetter was graduate student of Harrison White who is also an important contributor to social network theory ([Azarian 2005](#); [Freeman 2006](#)). White created a number of mathematical models of social structure. Among them are block models, vacancy chains and models of social structure that use pattern of relations to define an actor within a network as opposed to attributes and attitudes.

### 1.1 Aim & Motivation

This paper is finally motivated by the work of Nicholas A. Christakis and James H. Fowler on influence within social networks. Nicholas A. Christakis is Professor for Medicine at Harvard Medical School, Professor for Sociology at Harvard Faculty of

Arts and Science and Professor for Medical Sociology in the Department of Health Care Policy. He is an internist and social scientist who conducts research on social factors that affect health, health care, and longevity. James H. Fowler is a social scientist specializing in social networks, cooperation, political participation, and genopolitics. He is currently Professor of Medical Genetics in the School of Medicine and Professor of Political Science in the Division of Social Science at the University of California, San Diego.

Christakis and Fowler used a variety of sets of data to show the contagious effect of happiness, obesity, smoking and many more on actors within social networks. In their work they showed that these aspects of personal behavior that are featured mainly by personality traits are under constant influence by the people around us. These people might be part of our family, our neighbors or our friends.

The aim of this paper is to drive the number of aspects under influence further and extend the field of research to students and their grade point average. I will show that our own social network that is around us almost twenty-four hours a day has a deep impact on our performance in life. That is not just graduation marks at university but also work performance and ultimately general performance every day. I will use the results of a study of university students ( $n=63$ ), that were asked to report their grade point average (GPA) in both high school and university and a number of friends they spent significant time with, and apply social network analysis methods to identify influence between students on their respective performance during their studies. By doing so I will show that close friends will have an impact on how well an individual is doing in his or her life. The analysis is not from an information flow perspective as Granovetter sees it, where the constant flow of new and valuable information results in the cutting edge. The analysis rather takes into account the fact that specific personality traits are responsible for how well we are doing in life. These personality traits, as Christakis and Fowler have showed, tend to spread from actor to actor and therefore influence others. I will prove that this also applies to personality traits responsible for individual performance in both university and working environments.

Finally, I will elaborate on the strength and limitations of this method and on how the concept can be used to create a large-scale analysis to tighten the results. I will also evaluate the study's results and derive suggestions for further action in the domain of team performance within professional work environments.

## **1.2 Concepts**

Let me now examine a number of definitions and concepts that embody specific parts of a social network. I will reference to those in this paper. Most definitions will be common knowledge in the field of social network analysis. Some definitions are altered to serve the specific purpose of this paper.

### **1.2.1 Network**

A network or social network is the whole set of egos, alters and the relations between them. It can be plotted in graphs and statistical methods can be applied to find repeating pattern or specific properties. The smallest social network consists of two actors A and B that are connected to each other. Furthermore, there can be different types of social networks depending on the linking element. There can be social networks that consist of friendship relations, of work relations or of family relations. Hence, "different types of relations identify different networks" ([Knoke and Yang 2008](#)). A typical example of a social network is the personal network of a person's friends: One person knows several others and those again might now some of the people the person knows but also know individuals beyond that.

### **1.2.2 Actor**

Social network analysis is concerned with interlinking different subjects in network structures. An actor is a single, discrete entity that can be either an individual or a social unit such as a family or a company. In each person's individual friendship network that person is an actor as well as his or her friends are actors. In a political network, different actors might be the parties involved in a decision as well as the companies and the citizens that are affected thereof. I will use the term actor to refer to students that have participated in the study. Different actors can be interlinked in direction or non-directional relations.

### **1.2.3 Ego & Alter**

The concept of "ego" relates to a single actor within a network. It might be a natural person such as a student or a friend but can also stand for a unit of a company or a company itself. In this study we will talk about "Ego" as a single student. The difference to the general term actor is the fact that ego can only appear in combination with a perspective. When talking about an actor as ego then everything that applies to ego applies to the actor from his or her point of view.

Each ego will have zero to unlimited alters. Alters exist only with relation to an ego to which they are directly connected. Such as egos, alters are actors that have a perspective attached. I will talk about alters as participants that were nominated by an ego as being a friend.

### **1.2.4 Relations**

A relation is what connects ego to alter. They can be directional and non-directional, indicating whether there is a connection from ego to alter and whether this connection is reciprocal or not. I will talk in this study about directional relations. What is transported via relations within a social network can be different. For instance, Granovetter talks about information that are forwarded through relations ([Granovetter 1973](#); [Granovetter 2005](#)). In one of his examples he refers to information as information about job opportunities. A job seeking person gets a valuable tip from a close friend about a job offering that then leads to new employment for this specific person. We can easily think of other examples such as information about events happening in the surrounding area or novel information about a family member's feelings. In contrast to that, I am talking about influence that is forwarded through relations. One could argue that influence is exerted through information as information may lead a person to behave in a way that would have not happened without that information. I totally agree, but I would go a step further and say that influence is a meta category to information. Influence can not only be exerted by novel information but also by means of behaving, acting and responding. It is not uncommon that close friends tend to have similar verbal expressions or similar way of behaving after a while. Someone who is impressively eager in learning or focused while working may lead close others to be more eager and more focused themselves. In a nutshell, a relation is a connection between two actors through which influence is exerted from alter to ego.

### **1.2.5 Group & Clique**

Groups are subparts of networks and therefore can be described by the same structural properties as networks. If we look at the world's population, we will see that there is

just one social network that connects all people on the planet<sup>1</sup>. All networks that are studied are groups of this global network. But in contrast to a social network, within a study there can be many groups that are created to serve a specific purpose and discarded when calculations have been executed on them. Most the time this purpose is to compare two or more groups with each other in order to gain new knowledge. For example, I will compare groups of friends with groups of randomly assigned actors. Thus, groups appear in combination with measurements and are determined by them. Furthermore, members of a group have one or many attributes in common that make them a group. This can be same age, same gender or same high school grade. Finally, a group is a finite set of members of a social network with a definite number of members. Once calculations start members are neither added to nor removed from a group.

A clique within a social network is a special type of a group. Only those groups are cliques where all members are adjacent to each other. It is a “maximal complete sub graph of three or more nodes” ([Wasserman and Faust 1995](#)). If there is a group of four friends and all friends know each other, this group is a clique. A clique’s density is therefore always 1.00.

### 1.2.6 Centrality

Centrality is a way of defining the most important actor within a network. The concept of degree centrality takes into account the number of ties an actor has to other actors. The more the actor is connected to others the higher the centrality index is. Two methods to calculate an actor’s degree of centrality are InDegree centrality and eigenvector centrality. InDegree centrality in a directed social network considers only incoming ties from alter to ego. Eigenvector centrality is a more advanced approach to InDegree centrality. It is a recursive function of InDegree centrality, weighting each incoming connection from alter to ego with alters’ eigenvector score.

### 1.2.7 Density

For each network one can compute a specific density. The density indicates how closely-knit a network is. This is done by counting all existing ties and dividing this number by all possible ties. The higher this fraction is, the higher the network’s density is. Densities can also be applied to groups. Cliques always have a density of 1.00.

For example, Granovetter concluded that density has an impact on the stability of social norms. Social problems as the free-rider problem are less likely to happen in small, dense networks since “actors in such networks typically internalize norms that discourage free riding and emphasize trust” ([Granovetter 2005](#)).

## 1.3 Current Research

In one of their papers, the two researchers Nicholas A. Christakis and James H. Fowler measure the impact of social networks on obesity. To answer this question they used data from the Framingham Heart Study. This study is a long-term, ongoing cardiovascular study on residents of the town of Framingham, Massachusetts, USA. It began in 1948 with 5,209 adult subjects and is still ongoing on its third generation of participants ([Dawber 1980](#)). A total of 12,067 subjects from the period between 1971 and 2003 were examined. Christakis and Fowler found out that chance of becoming

---

<sup>1</sup> One could argue that there are tribes that live independently from civilization. But to know about these tribes, they need to be discovered and those who discovered them are without any doubt socially connected to them and bridge them to the civilized world. By being “connected” I talk about being aware of each other or having knowledge of each other’s existence.

obese increased by 57%, if that specific person had a friend who became obese during that period. Among pairs of adult siblings, if one sibling became obese, the chance the other would become obese still increased by 40%. Same is valid for spouses with a probability of 37%. Their paper shows that obesity, as a result of specific behavior, is influenced by a person's social network ([Christakis and Fowler 2007](#)).

In another paper Christakis and Fowler used the same Framingham Heart Study data to examine the spread of happiness in social networks. In this paper they drew on a sample of 4,739 individuals from the period of 1983 to 2003. By applying methods of social network analysis they found out that a friend living within a mile radius becoming happier increases the chance for others to become happier by 25%. This also holds true for co resident spouses and siblings who live within a mile and next door neighbors. They showed that happiness depends on others with whom one is connected ([Fowler and Christakis 2008](#)).

In a third paper they conducted research on "The Collective Dynamics of Smoking in a Large Social Network". Christakis and Fowler measured the extent of the person-to-person spread of smoking behavior in a large social network. Their data was again the Framingham Heart Study and again they used the same 12,067 subjects that were used to study the spread of obesity. The results were that smoking cessation of a spouse decreased a person's chances of smoking by 67%. Smoking cessation by a sibling decreased the chances by 25%. Smoking cessation by a friend decreased the chances by 36%. Among persons working in a small firm, smoking cessation by a coworker decreased chances of others smoking by 34%. The higher the education of friends is, the higher is the influence between them. For neighbors in the immediate geographic area these effects could not be seen. With these findings Christakis and Fowler could proof that smoking behavior spreads through close and distant social ties ([Christakis and Fowler 2008](#)).

In a different study Ann L. Shattuck, Emily White and Alan R. Kristal examined the effect of a wife's low-fat diet on the husband's fat consumption. As a follow-up to the Women's Health Trial the team sent out dietary and health questionnaires to a randomly selected sample of participants. They found an absolute difference in fat intake between groups of 4 percent. The "wife's attitude and fat intake were among the most important predictors of her husband's fat intake". Thus, the specific and individual traits towards food consumption spreads through social networks ([Shattuck, White et al. 1992](#)).

In a 2006 controlled field experiment Armin Falk and Andrea Ichino studied peer effects on the output of filling letters into envelopes. In order to gain their results the team recruited 24 subjects and allocated them to their task either alone or in pairs. Eight subjects were given the task to fill letters into envelopes alone. 16 subjects, forming eight teams, were given the instruction to work in pairs. They found evidence of peer effects in pair treatment because the standard deviations of output were smaller within pairs than between pairs and average output was higher in the pair treatment. Thus, peers effect each other not only in the spread of personality traits but also in terms of motivation and productivity ([Falk and Ichino 2006](#)).

In 1996 Gerhard Blickle analyzed the relationships between personality traits, learning strategies, and performance. He conducted two multivariate studies that yielded two factors as influential on academic performance: *Consciousness* and *openness to experience* ([Blickle 1996](#)).

## 2 Do peers influence each other?

The question I am trying to answer is: Does a person's performance in his or her everyday life, whether at work or in private, depend on with whom he or she is befriended or not? The dictionary puts performance equal to accomplishment, execution and fulfillment "usually with regard to effectiveness" ([Neufeldt and Guralnik 1994](#)). Performance in this context means the extent to which a person does something good or bad. In university performance is reflected by the student's grades. The higher the grades are the higher the performance. In a working environment we could measure performance in terms of monetary outcome. In sport performance could be measured in time a runner needs to complete a given track or in the number of goals a soccer player scores. In academic research people tend to measure performance in the number of publications in specific journals. In one's everyday life performance is the sum of all achievements in different fields. Fields range from sports through university over work to hobbies. Thus, I will talk about performance as the level of achievements over the course of life.

But in order to conduct a study where we can compare performance, we need a quantifiable measure. Performance in the way just described is not quantifiable. One could measure the number of achievements per person in his or her life but we cannot guarantee that all these achievements are equally easy to achieve. Moreover, there is doubt about what exactly counts as achievement and in order to take all achievements of a person into account we would have to follow that person throughout his or her life. It follows that we need some sort of proxy to performance that is both quantifiable and can reliably reflect a person's true performance. As before, we could use a person's monthly income to calculate performance. But even though monetary figures are quantifiable there is too much bias involved. The salary is not solely set by the employee's performance and is also under influence of supply and demand. Additionally, benefits as a corporate car, discounts on products or other job specific adjustments have an impact on the salary. We could also rely on the ability of people to evaluate the performance of their peers. Since 2006 this is called crowd sourcing ([Howe 2006](#)). The idea behind crowd sourcing is to draw knowledge or information from a large number of novice individuals rather than from a small group of experts to solve problems or help form inventions. Mathematics ensures that small deviations in the data are flattened by the sheer number of contributors. With each new contributor the impact of a single contributor diminishes reducing the impact each mistakes has. In small expert groups, each mistake counts more as there is just as limited number of experts in this group. Considering the fact that there are only a limited number of peers being able to evaluate someone else correctly we cannot count on mathematics to correctly measure an individual's performance. Furthermore, performance is a vague definition and it is difficult to ensure that all evaluating respondents have the same understanding of performance.

It is hence necessary to find a proxy to measure one's performance that is quantifiable and ideally objectively set by an independent authority. In addition, each individual under examination should have been evaluated against this method. A proxy that meets these criteria is the GPA. The GPA is the average of all grades achieved in either high school or university. Because the GPA is a number, it is easily quantifiable and comparable. Furthermore, when we pick university students as participants, all of them went through high school as well as university examination and grades of both institutions are therefore available. The criterion in question is the objectivity of grades. Still, grades achieved in both high school and university are always influenced by the teacher in charge ([Ickes-Dunbar 2004](#)). But there are also authors saying that

“performance is objectified by grades” ([Bernstein 2000](#)). And just like it behaves with crowd sourcing, if we assume that we always have an existing yet always different deviation of the given grade from the real performance, we can say that this deviation is flattened by the large number of grades that lead to the GPA. Furthermore, when we consider that we will take a large number of individual GPAs into account, the impact of the deviation of a single grade is diminishing small given the large number of grades that are hence used for calculation.

How do we finally measure the impact of friends on a person’s performance? Duckworth and Seligman found that “self-discipline predicted academic performance more robustly than did IQ” ([Duckworth and Seligman 2005](#)) and in another study conducted by Chamorro-Premuzic et. al. there was “modest but significant correlations between conscientiousness and academic performance” ([Chamorro-Premuzic and Furnham 2003](#)). Additionally, Blickle found that the two character traits *conscientiousness* and *openness to experience* drive academic performance ([Blickle 1996](#)). We hence know that the GPA is under strong influence of specific personality traits. We therefore need a way to show that these personality-traits are under constant influence by close friends. The Studies of Christakis and Fowler on the spread of happiness, smoking or obesity, etc. in social networks deliver us with this proof ([Christakis and Fowler 2007](#); [Christakis and Fowler 2008](#); [Fowler and Christakis 2008](#); [Fowler and Christakis 2008](#); [Nicholas A. Christakis 2010](#)). Thus, the underlying magic of this study is to build a social network reflecting the relationships between the participants to show in which ways they influence each other. Necessary data are the names of close friends as well as high school and university GPA of each participant and the respective data of the named friends. We can then plot a graph of the actual social network and use social network analysis methods to identify correlations between grades and network positions of individuals or whole sub-networks with special properties.

In this chapter I will first elaborate in detail on which data was collected and in what ways the data collection was realized. Then I will explain the methods applied to interpret this set of data and how these methods can be improved. Finally I will present the unique results that can be derived from the set of data.

## 2.1 Data

In my study I relied on two measures that can both reflect individual performance, include a temporal perspective and are objective.

Changes in individual performance are a process that does not happen from one day to another. Hence, in order to conduct a study that measures the impact of anything on individual performance it is necessary to apply longitudinal analysis that can catch the changes in individual behavior over a long period of time. Large studies like the ones conducted by Christakis and Fowler (2007, 2008) therefore take a large body of data collected over two decades into consideration. Ideally, that is research that has been done by others in a similar field that can be taken to extract valuable network and performance data. Also Baldwin, Bedell and Johnson (1997) mention the limitation of the one-time questionnaire aspect of their study on network effects on student satisfaction and performance within a team-based M.B.A. program. They state that their study design “precludes ironclad causal inferences”([Timothy T. Baldwin 1997](#)). Unfortunately, it is extremely difficult to find reliable data that suits the needs and has the ability to reflect changes over a long period of time. If such data is not available it increases the efforts for research many times. When time limitations are one restriction it is important to collect data that is both available and has the ability to reflect long term changes.

Besides that the measures also had to cover a period that can be clearly differentiated in terms of one's social network. In order to measure the effects of an individual's social network on that one's performance it was necessary to take a period of time where this social network remained stable. In my terms a network remains stable when important network actors do not change. In other words, it remains stable when one's friends remain the same over the whole period of research. Finally, these two measures had to be as objective as possible to rule out any bias.

I identified the individual high school GPA and a person's university degree GPA suitable figures that meet these requirements. The temporal perspective of both measures holds because on average there is a time difference of three years between high school graduation and graduation from university in Germany. Furthermore, the network remains stable because during a bachelor degree a student meets new people that accompany him or her throughout the whole course. The aspect of objectivity has been discussed above. Many papers use the GPA to assess performance ([Timothy T. Baldwin 1997](#); [Duckworth and Seligman 2005](#)) and Johnson (1997) says that the "Grade point average, or GPA, is the most widely used summary of undergraduate student performance in our educational system".

Besides the individual performance, it was important to build a social network around the participants. But there is not just one social network, which leads us to the main question: Which social network is the network that leads to the most influence in individual performance? McPherson mentions "marriage, friendship, work, advice, support, information transfer, exchange, co-membership" and other types of social networks that exist between people ([Miller McPherson 2001](#)). Christakis and Fowler even build up social networks that consist of sexual intercourse ([Nicholas A. Christakis 2010](#)). In order to reliably give proof of the influence of the GPA of one student on another it was important to choose the network with the highest probability of influence. Christakis et. al. showed that friendship networks are among all social networks those with an high influence of individual actors on each other<sup>2</sup> ([Christakis and Fowler 2007](#); [Christakis and Fowler 2008](#)). This argument is emphasized by Marsden et. al.'s (1993) conclusion that "influence tends to occur among actors with consistent interests ([Peter V. Marsden 1993](#)). I therefore chose to create a network of friendships.

The question was what properties a friend needs to have to be a suitable network actor in this social network. Despite Granovetter arguing that the most novel information flows between weak ties ([Granovetter 1973](#)) the highest influence happens between strong ties ([Shattuck, White et al. 1992](#); [Christakis and Fowler 2007](#); [Christakis and Fowler 2008](#); [Fowler and Christakis 2008](#); [Nicholas A. Christakis 2010](#)). The aim was therefore to build a friendship network out of close friends. Webster's Dictionary translates friends as "intimate associate" or "close acquaintance" ([Neufeldt and Guralnik 1994](#)). Hence, in this definition close friends are friends that spend a lot of time together in their private life, who discuss personal matters and are loyal towards each other.

After having defined the target social network it was important to collect all relevant information. In an online questionnaire each participant was asked to name five friends. To ask for five friends instead of more or less follows two good reasons: The first reason follows a more technical aim. When collecting research data the response rate is of central importance in order to collect a suitably large body of data in a reasonable amount of time. With internet-based surveys, the smaller the questionnaire

---

<sup>2</sup> Christakis et. al (2010) also use networks of sexual intercourse to get a network of even higher "friendship" but because these connections are rather rare and also happen in most cases only between people of different sex the use of this type of network was inappropriate in this case.

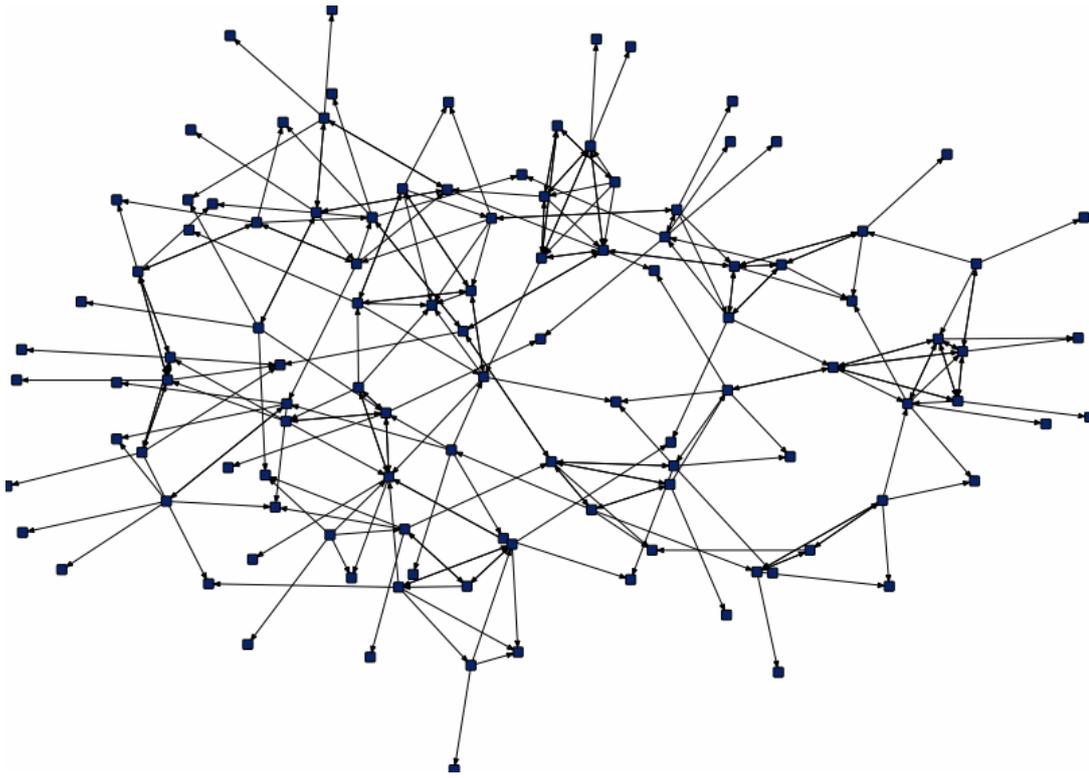
the higher the response rate ([Deutskens, de Ruyter et al. 2004](#)). A number of five friends is therefore short enough to yield a high response rate and long enough to not cut off important nominations. For 34.9% of the respondents five friends were still too many as they did not enter all five friends. The second reason is about prioritization. The aim was to ask for the closest and therefore most important of all friends a participant has. A maximum of five friends forces the respondents to fully concentrate on the five most important friends in their lives because they had only five “votes” to give away. Thus, reducing the number of possible nominations lets user focus on their most important friends.

To maximize the use of doing a survey, I also asked for the respondent’s perception whether the named friend did better or worse in their university degree. Although it was not the first aim to test for the effect of perception it was a welcomed way to test for one more metric. In order to simplify the questionnaire and increase response rate the survey asked for each friend whether he or she did better, as good as oneself, or worse in his or her studies.

The scope of students was limited to two important variables: Students of Bachelor of Arts in Business Economics and students at Witten/Herdecke University. In order to create a social network it is important that each actor that is named by a respondent can be matched to the right set of data the named person has filled in. Every record containing the personal data of a random person is associated with a name. In order to fully match named students and individual records the survey had to be free of spelling mistakes. This can only be accomplished by giving the respondents a valid list of possible friends they could fill in. Furthermore, in order to notify students to fill in their data it was necessary that for each named person valid contact details were available. Hence, it was important to predefine a set of possible students which was only possible by limiting the whole set of students to Witten/Herdecke University.

The limitation to students of Bachelor of Science in Business Economics comes from the fact that they undergo the same degree and therefore attend same or at least similar courses. As a result from that the same course’s examinations are the same and thus any grading biases can be ruled out. Second, the bachelor degree takes only three years of completion, which fits in the temporal horizon of the study. In addition to that the group of students of Bachelor of Science is the largest group at Witten/Herdecke University that graduates with a Bachelor degree, yielding the highest probability of a large sample size.

The survey itself was delivered through a special website. Each respondent got a unique URL that guided him or her to a page that already contained some basic information. The contact data for each student was retrieved from the university’s address book. Because the survey could only be opened on exclusive invitation it was ensured that only students from Witten/Herdecke University were able to fill in the form. Once a respondent filled in and submitted the form the system automatically informed the named friends. They were told that they have been mentioned in the survey and are hence important parts of the overall process followed by a request to fill in the survey themselves. In case this person then followed the request and submitted the survey as well a new set of friends was informed about the survey. From a researchers perspective the snow-ball-characteristics of the system helped to reduce efforts to find the right respondents as the social network was build almost on its own. There was only a need to trigger the process by manually asking possible respondents to fill in the form and to sometimes fill structural holes to ensure sound network data.



**Fig. 1: The graph of the resulting social network**

In summary, I collected 63 qualified answers with each up to five friends and both high school GPA and university degree GPA. All of these 63 answers were interlinked in a large directed social network.

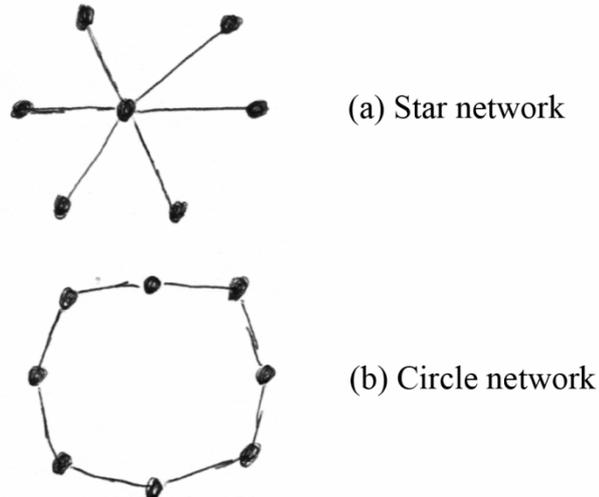
## **2.2 Methods applied**

The question whether friends influence each other in terms of academic performance is clearly a matter in the field of behavioral and social science. But in order to quantify influences and measure the impact that is done by one friend on the other, statistical methods are necessary. Also, the study makes extensive use of the relations among friends and the influence that arises from there. These fields are combined in social network analysis and as Wassermann, et. al. put it:

*“Many researchers have realized that the network perspective allows new leverage for answering standard social and behavioral science research questions by giving precise formal definition to aspects of the political, economic, or social structural environment. From the view of social network analysis, the social environment can be expressed as patterns or regularities on relationships among interacting units.”*

Hence, I will use measures of social network analysis to find and interpret correlations in metrics between different actors within the given social network.

The analysis is broadly divided into two parts: Network-wide analysis and Ego-centered analysis. The main difference between both is the perspective from which we will look at the data. Network-wide analysis deals with significant patterns between individual actors, groups of actors or the whole network. We will have a broad overview of what is happening in the network as a whole. For example, how does an individual GPA relate to the overall GPA average of the network? How are GPAs



**Fig. 2: Examples of a star network and a circle network (modeled after Wassermann and Faust, 1995)**

spread within the network compared to a specific sub network? The network-wide analysis is more a macro level analysis than looking at individuals. The ego-centered approach is the micro level analysis starting with the single actor and expanding the research to comparison groups that are related to this actor. There we look at how a single actor's properties relate to properties of friends. This might be the question whether an individual's GPA is connected to his or her friend's GPA or what type of friends is closely related to what GPA?

### **2.2.1 Network-wide analysis**

In this section I will examine regular patterns on a network-wide macro level. I will first investigate, whether an actor that has been nominated by a large number of other actors has any influence on a networks average GPA. After that I will try to find out if GPAs among friends are closer to each other than throughout the whole network.

#### ***2.2.1.1 Influence of actors with high degree of Centrality on their surrounding network***

The question is whether a specific position within a network yields the possibility of certain influence on the network's GPA. We know that friends are influential towards each other. We also know that centrality is the structural property that is most often associated with influence in a network ([Friedkin 1993](#); [Raymond T. Sparrowe 2001](#)). We may thus derive the assumption that a friend who holds a central position within a network influences the network's performance. There are many measures of centrality ([Wasserman and Faust 1995](#)). We will concentrate here on two measures: InDegree and eigenvector.

The concept of degree centrality takes into account the number of ties an actor has to other actors. The more the actor is connected to others the higher the centrality index is. To illustrate, this imagine a star network (a) and a circle network (b). A star network consists of one central actor who has  $g - 1$  ties to other actors. The remaining  $g - 1$  actors have each one tie to the central actor. It is obvious that the central actor is the one with the highest centrality index. In contrast to the star network resides the circle

network where each actor has two ties in total: one to each neighbor. Here it is obvious that there is no actor who is more central than anyone of the others. InDegree in a directed social network then considers only incoming ties. In our study a tie is only relevant for ego's InDegree index when alter nominated ego as a friend.

But why may an actor with a high InDegree index influence his or her network? We know that each ego is influenced by his or her friends. These are the friends each participant nominated that influence him or her most. We count a nomination of alter by ego as a directed connection from ego to alter. For each nomination alter receives an incoming tie because that person seems to have had an influence on ego. The more incoming connections an alter or an actor receives, the more influential he or she is. Because the more incoming ties an actor has the higher the InDegree index is we can thus take InDegree centrality for an indicator for influence.

In order to find proof of a correlation between the most influential actor's GPA and the network-wide GPA we needed a number of actor-network-GPA combinations. Just one single occurrence does not allow concluding to the basic population. Instead of creating more social networks of the same type I divided the existing network into subgroups according to their respective density. The density gives us an idea about how closely-knit the network is. Each network has a different density and within networks there are also sub networks with different density. Because the higher the density the tighter a network is, I used the density to identify subgroups. To identify subgroups I computed three sub networks with each 20 actors that have the highest possible density meaning that these are the three subgroups that are connected most closely<sup>3</sup>. For each of the three subgroups I could calculate the most central actor in terms of InDegree centrality. In the next step I could then compute the average GPA within each subgroup. To find any relevant influences I took the average GPA and determined its correlation with the most central actor's GPA.

I also tested for correlation between the subgroup's GPA and the most central actor in terms of eigenvector. The concept of eigenvector takes the InDegree concept to the next level by weighting incoming ties with the emitting actor's eigenvector. It is therefore a recursive function that assigns each actor a centrality index equal to the sum of each incoming tie that is multiplied with the emitting actor's centrality index. In the first iteration each actor is a centrality index of one assigned. The function then repeats itself a given number of times and returns the centrality index for each actor.

The reasoning why this procedure is a valid way to measure influence within networks is the same as for the InDegree centrality. It is only more precise, because it is a recursive function that repeats itself and hence the influential power of other actors is also taken into account ([Wasserman and Faust 1995](#)).

### ***2.2.1.2 Standard deviation within close friends and within network***

The underlying assumption of this paper is that over time a friend's academic performance will be automatically aligned to his or her friend's performance. Provine calls this "synchronizing the state of a group" ([Provine 2005](#)). The gap between both friends' performance will be smaller and grades will be closer to each other. The more two friend's performances are in line the closer the grades lie to the mean. The more different two friend's performance is, the further away the grades lie from the mean. In statistics the degree to which the values of an independent random variable are in line with each other is represented by the variable's standard deviation. Technically the

---

<sup>3</sup> Having three subgroups is still not enough to compute a significant correlation, but on the other hand the smaller a group the less reliable the average GPA is. It remains that to solve this dilemma a larger sample of participants is necessary.

standard deviation of a statistical population is the square root of its variance. The variance again is the sum of all squared differences between each value and the mean divided by the number of values.

When we assume that friends influence each other and positive influence is represented by grades that lie closer to each other than on average we can use the standard deviation of an actor's friend's grades and compare it to a control group to find proof for influence in social networks. In order to find proof the standard deviation of an actor's friend's grades needs to be smaller than in the control group indicating that grades lie closer to each other.

I used two different ways of grouping friends together. The first approach is pretty straight forward. Because we are looking for actors within a network that are closely connected we can rely on the subgroups of first-level-friends within the main network. These subgroups each consist of one ego and zero to five alters that were nominated by ego when answering to the survey. The structure of this network is therefore a star having ego in the center with outgoing ties to different alters. The advantage of this method is the easy access to the network information as there is no need for further computation of subgroups. The subgroups are already available as data sets in the database. The only requirement is to match the nominated alters to their respective GPA. The disadvantage of this method is that it relies heavily on the information provided by a single respondent. If the respondent did not enter any alters a standard deviation could not be computed as there need to be more than a single value to calculate standard deviations. A second disadvantage is that this method does only pay attention to outgoing ties from ego to alter and not incoming ties to ego from other alters. The fact that ego could also be nominated by other respondents is ignored. The reliability of the data is thus limited. In order to eliminate this inaccuracy I used an algorithm by Borgatti et. al. (2002) that optimizes a cost function to find a predefined number of clique-like structures within a network using a tabu search method. The algorithm by Borgatti et. al. (2002) then tries to find the desired number of structures that are most like cliques. In other words the algorithm tries to find  $n$  subgroups that have the highest density within the given network. The structure of these subgroups is not star-like because now alters can also be tied together. Important in this approach is the overall density of ties between the actors. In order to find a comparable index for the influence within these groups I first computed twelve non-overlapping and independent groups. For each group I calculated the standard deviation of all actor's BA GPAs and created the average of all standard deviations to end up with a single, comparable number.

This algorithm is more precise than selecting star-like cases of ties from ego to alter. Because the closer actors are, the higher their influence is and the fact that this algorithm uses measures of density to determine subgroups and it selects those groups with the highest density, the influence among members of these groups must be higher than within star-like subgroups.

In terms of finding a control group, I also used two different approaches, because difficult is, what a valid control group is and how a valid control group is determined ([Peter V. Marsden 1993](#)). One way to find a control group was to use the same logic of computing group-related individual standard deviations as used before. One reason for doing so is the advantage that in this case the input data's structure between the test group and the control group is relatively the same. In both cases groups are used to determine a group-related standard deviation and then compute the average over the whole set of groups. To create such control groups each respondent was

assigned a random number between one and twelve<sup>4</sup>. By doing so I made sure that there was no real-life metric like geodesic distances or other structural properties responsible for forming the group that could have influenced the outcome. Disadvantageous is that by using random numbers it may still happen that actors are grouped together that are in each other's neighborhood and thus are influential towards to each other. This problem may be solved by generating a large number of group samplings and then again computing the average of all these samples. Instead of doing many iterative calculations, I decided to verify these results by applying a second way of defining a control group. This approach is way much easier as I consider the whole network as a control group. This is applicable, because the whole network is less interconnected or less dense (0.0478) than the test groups (0.5325 on average). Furthermore, the whole network has not been chosen according to any network specific properties or structures.

## 2.2.2 Ego-centered analysis

In this chapter, I will present my findings on influence among close friends on each other's GPA. In contrast to the previous chapter the emphasis will be on small subgroups of friends. I will not talk about the influence of specific network position on an actor's performance but will investigate the importance of an actor's friend's performance. It is more the micro perspective of the individual actor and his or her surroundings than the macro perspective of the whole network. Does one's close friends' performing better in high school have an influence on that person's own performance? Do actors with friends of a higher academic aptitude perform better?

In my research I used two different yet closely related approaches. Both approaches measure the impact of an actor's friends on his or her performance by looking at the friend's performance compared to the actor. One approach looks at the number of friends that an actor has that performed better in high school than the actor him or herself. The other approach takes the actor's friend's average GPA into account. The question here was whether the fact that an actor's friends as a whole are performing better or worse has an impact on the actor's performance. The second approach is broader in scope than the first. It considers all friends to have an influence whereas the first approach considers only those friends who did better as influential.

### 2.2.2.1 *The impact of better performing friends*

There is a significant influence among friends regarding smoking behavior, the feeling of happiness or loneliness and obesity ([Christakis and Fowler 2007](#); [Christakis and Fowler 2008](#); [Fowler and Christakis 2008](#); [Cacioppo, Fowler et al. 2009](#)). In this chapter I will suggest that there is also an existing influence regarding academic performance. The assumption is that friends tend to drag each other in their directions. A friend, who has higher achievements than oneself, causes us to improve. A friend, who has lower achievements, hence causes us to drop in performance. Because attitudes and specific character traits are drivers for individual performance, this assumption is in line with the fact that "attitudes are confirmed and reinforced when they are shared with the comparison group but altered when they are discrepant" ([Erickson 1988](#); [Peter V. Marsden 1993](#)).

In order to conduct this analysis it is important to quantify the influence each actor is exposed to. This can be done by introducing a new metric: The *Relative Friend Performance Score*. This score indicates to which degree an actor's friends as a whole

---

<sup>4</sup> The number twelve resembles twelve groups with each five members allowing for both large enough groups to calculate the GPA standard deviation as well as enough samples to calculate averages thereof.

perform better or worse than him or herself. To calculate the score each friend that performed better in high school is assigned the value 1. Each friend that performed equally is assigned the number 0 and each friend that performed worse is assigned the number -1. The *Relative Friend Performance Score* is then the sum of these friend's values. Let us assume an actor has five friends. Two of these friends did better in high school and three did worse than the actor. The *Relative Friend Performance Score* for this set of five friends relative to the actor is therefore  $-1^5$ . The higher the score is the higher is the influence towards better grades. The lower the score is the higher is the influence towards lower grades. A score of zero results in no influence.

A second metric that is needed to measure the influence of friends is the individual change in performance. The individual improvement needs to be quantified. We already know that we can use an actor's GPA to measure his or her performance. But because the change in behavior and the effects on an actor's performance are a time consuming process we need to pay attention to the longitudinal aspect. We will accomplish that by comparing the high school results of each actor with his or her university results. The fraction of both, university GPA divided by high school GPA, gives us the relative performance improvement.

The idea is then to test for correlations between the *Relative Friend Performance Score* and individual improvements from high school GPA to university GPA.

### 2.2.2.2 *The impact of the whole set of friends*

As mentioned before we can also look at an actor's set of friends as a whole. In this section, I ignored the detailed structure of how many friends did better, how many were as good as the actor and how many did worse and took a rather broad approach. Instead of quantifying the influence using the *Relative Friend Performance Score* I used the friend's average high school GPA to determine, whether they all together performed better or worse than the actor. This change in the way the friend's performance is measured and quantified is profound and implies a different way of thinking. The most critical aspect lies in the way the performance is calculated. The *Relative Friend Performance Score* is calculated by subtracting those friends that performed worse than the actor from the friends that performed better. This is different from the method applied in this case. Here the friend performance is determined by calculating the average of all friend's GPAs. That is as simple as adding all GPAs together and dividing them by the number of friends. This resulting average GPA can be either better or worse than the actor's GPA. In case it is better, the actor's friends are considered to be performing better than the actor. In case the average GPA is lower than the actor's GPA, the actor's friends are considered to be performing worse than the actor. This bears one mathematical implication. One high performing friend can outweigh slightly underperforming others resulting in a positive friend performance measure where the *Relative Friend Performance Score* might have indicated a negative friend performance. Let us assume the actor having scored any random number of 5. His or her five friends have scored four times 4 and one time 10. The average score of the friends is 5.2 which is higher than 5, resulting in the friends to be considered better. The *Relative Friend Performance Score* instead would have resulted in a strong negative friend performance value of  $-3$ .<sup>6</sup> By using the average GPA of an actor's friend, the friends are given a different weighting according to their GPA. Hence, it is assumed that those friends, who

---

<sup>5</sup>  $2 * 1 + 3 * (-1) = -1$

<sup>6</sup> Relative Friend Performance Score:  $1 * 1 + (4 * -1) = -3$

did either extraordinarily well or performed extraordinarily poor, have a greater impact on the actor's performance than friends that were close to the mean.

### **2.3 Methodological Review**

The basic idea behind this study was the desire to apply the concepts of social network influence measurement to a new field of research. The driving question was whether actors have an influence on each other's academic performance when they are socially closely connected. In this section I will elaborate on the positive and negative key aspects during conceptualization, execution and evaluation of this study.

As this research highly involves connections between actors and behavior in a social context it was obvious that the theory of social network analysis was of fundamental importance to this project. Social network theory is at its essence a statistical framework that draws on graph theory and graphs of social connections and relationships to gain insights into social behavior. One requirement of statistical analysis is sufficient data. Thus, in order to conduct the study it was necessary to either find or collect relevant data. I decided for collecting exactly the data that I needed to conduct the study. One way to collect relevant data is to create a survey that covers the needed information which can be evaluated later on. Unfortunately, research surveys can be a time consuming process. As mentioned above, I created a survey that featured a special type of snow ball system to help speed up the process. Nominated actors were emailed a message informing them about them being mentioned and asking them to fill in the survey themselves. The snowball system worked astonishing well. My first aim was to get about 50 qualified responses within four weeks. I ended up having 63 qualified responses within three weeks. Though the system reinforced itself it was important to jump start it by calling or mailing about 20 to 30 people and asking them to participate in the study. After that the answers were rolling in. Most of the answers came during the first week. But because the process was self-triggered the results depended highly on the people who replied and who did not. As a result of that, the social network that was created emerged around those people who answered leaving holes of incomplete data sets every now and then. These holes came from the fact that some actors were mentioned but never filled in their respective high school and university GPA. In order to fill these holes it was important to contact these specific actors which took most of the second and third week. The snow ball system proved helpful and was of good help in collection responses. Needless to mention that this is only possible where all the relevant contact data is available as it was in this case.

Difficult besides the hardly steerable self-iterating process were privacy fears that sometimes created reluctance. Respondents hesitated to fill in their perception of their friend's academic performance. Others were not sure whether their name will show up in the study later on. In order to reduce privacy concerns as much as possible I thought about privacy issues from the very first moment. The whole survey was only accessible to those having received an invitation link drastically reducing the possibility of unauthorized persons seeing private data. Names of respondents were only used to map different actors to others in order to create the necessary social network. Wherever calculations were done, names were deleted from the records. The nature of statistical models ensured that single sets of data were combined to new figures making it impossible to reengineer and to recreate the original data records. The trust into the study and into the way private data were handled was established by demonstrating full transparency in the processes. Throughout the survey website it was stated what data was used, why I am asking for it, in which way I will be using the data and for how long. This procedure proved valuable in most cases resulting in a small number of five occasions where privacy concerns were raised. In three of five cases these concerns

were resolved by giving further and more detailed information. Only two cases turned out to be unsolvable. Asking for certain information in a more subtle way could have reduced this number of concerns from the very beginning.

Of higher importance and higher impact were connectional biases that influenced the results. Regarding the GPA as the ultimate way to objectify academic performance, I have to mention a strong discrepancy in grading practices among high schools and universities. Because the respondents attended different schools in different areas and the 16 regions in Germany each have their own academic system, comparability of high school GPAs is thus limited. The academic differences are as big as regions having twelve years of compulsory school attendance compared to others still having 13 years of compulsory school attendance. Also comparability is limited as there is no standardized graduation examination in Germany, resulting in many different ways of assessing academic performance. A second aspect comes from each person's individual plans for his or her life. It is clear that different professions require different grades in university. Though this bias was addressed by concentrating on students of business economics, still within this group students have different aims in life that require different academic grades. As aims change and evolve throughout life, motivation changes and thus performance is influenced. Additionally there is a clear motivation that comes from the high school GPA itself. Data has shown that a low high school grade correlates positively with a high improvement from high school to university suggesting a significant impact of low high school grades on performance.

Regarding the execution we can state positive as well as negative aspects. Mainly the execution speed was a significant advantage in this study. The use of the already mentioned snowball system increased return rates and made it possible to collect more data with less effort. Instead of four weeks that were calculated for this segment it took just about three weeks to get to the same number of participants. This is an increase by 25%. Helpful to this success was the fact that detailed contact information was available. I could draw on a database of e-mail addresses and names that is only available to members of this university in order to inform the relevant participants of about their importance to the study. Furthermore, a lean design of the questionnaire and clear communication of reasons and impact of the study as well as underlining the importance of every single participant additionally drove numbers of respondents significantly. There are three aspects that I consider as the most important drivers concerning the e-mail content: The first driver is a personalized greeting. Each and every e-mail started with the real first name of the recipient, differentiating the e-mail from regular mailing list e-mails from the very beginning. Recipients feel personally addressed when opening the message resulting in an increase in response rate ([Levinson, Levinson et al. 2007](#)). The second driver was directly mentioning urgency and connecting this urgency with a social aspect. Since social networking websites like Facebook we know about the power of social contacts and how much friends can influence our behavior ([Kirkpatrick 2010](#)). Mentioning that a person has been nominated by one of his or her friends plays an important role: Mentioning that a known person has already answered to the survey lends credibility to the study hence reducing the personal threshold of replying. The third driver was again building on social relationships. I gave the messages a layout, writing style and tone that was far away from what is known from mass mailings and that concentrated on real life communications. I created a message template by picturing writing such an e-mail to the recipient in real thus giving it a look and feel of everyday mails. The personal look further drove response rates. Despite all optimization efforts, it was still a relatively difficult and frustrating task to convince students to respond to the survey. Just as with every other mass mailing, people either read mails too late, are too busy to act upon

them or simply ignore them. Sometimes, when the nominated people did not respond to e-mails, it was necessary to follow up on automatically generated invitation e-mails by calling that particular person or writing additional e-mails. I used even other communication methods like Facebook messages and instant messaging to get people to fill in the survey. This proved especially important when the person in question was sitting at a from a social networking perspective central position. Because the whole system of inviting participants was more or less self-driven it happened a lot that individuals that got many nominations and that were linking salient subparts of the network did not participate in the survey thus leaving this spot blank. Because drawing conclusions from social networks relies per definition on the relations between actors, blank spots within a network can lead to incomplete datasets, thus resulting in invalid results. Trying to fill these blank spots by reaching out to those individuals emerged as a difficult to sometimes impossible task.

In the phase of evaluation the software tools used to calculate networks, plot network graphs and find correlations as well as dependencies were crucial. I used a combination of Microsoft Excel, UCINET 6 with NetDraw by Analytic Technologies as well as IBM's SPSS and my own software that I developed as needed. UCINET 6 was helpful when running standard calculations from network theory to find central actors, calculate densities and divide the network in sub networks upon these parameters. NetDraw extended these functions by plotting graphical examples of the network in order to visualize and visually confirm results. I used Microsoft Excel and my own software to further extract data that could not be extracted by UCINET 6 or NetDraw or to transcribe existing data into a standardized format that could then be handed over to SPSS. In SPSS I finally combined the existing sets of data in order to apply statistical models such as correlation, partial correlation and regression analysis models. Most the time the different sets of data could be transcribed to fit together.

My approach had one downside. During the phase of evaluation and while putting numbers together for calculations, it became evident that the underlying models were defined too vaguely. It happened that in order to successfully compute one of the models, further information about the participants would have been necessary. But these information pieces were missing, because the questionnaire used to collect the data was not designed to collect exactly this type of information. One of the main pieces of information missing was the question, whether the participants and their friends knew each other throughout the whole time at university. Or did they just get to know each other at a time where one of them was already done with his or her degree leaving no possibility that he or she was ever influenced by his or her friend.

## **2.4 Methodological advancements**

Besides actual research data that hints at significant influence among actors regarding individual performance, the study conducted serves as a first methodological field test. With every new approach there are many lessons learned which allow us to derive suggestions for further research. In this chapter I will present ideas and concepts for future studies. I will draw from experiences gained during the current study in terms of conceptualization, execution and evaluation. First, I will suggest improvements regarding the early conceptual period of the research study where foundations are laid for later. After that, I will turn the focus to the actual execution of the study and will identify ways to avoid difficulties in the future. Finally, I will elaborate on how to improve the statistical models used to identify and verify influence among actors.

In what ways can we improve the conceptual stage and how do we address biases? One suggestion is to rely on data that comes from respondents that completed their education in similar environments. In terms of university this might be making

sure that all respondents attended a university and studied a subject that includes a high amount of objectively measured data. This can be a subject where the same exams are taken independently from the issuing university. Degrees under governmental supervision such as law, pedagogies and medicine fall within this category. In medicine, for example, each year all students take the same exam and results are fully comparable. That would be one objectively comparable measure. A second one would be relying on students from the same school or, if there is not sufficient data available, from the same region where exams are taken together. In Germany, 15 out of 16 regions have introduced centralized high school examination. Final exams are set up and administered by a central authority increasing comparability among grades. Unfortunately central examination accounts for roughly 25% percent of the final high school GPA, leaving the majority of the assessment to individual decisions by teachers in charge. Nonetheless, this bias can be addressed by selecting respondents from the same school in the same region because as each region is responsible for their own educational system, we can ensure that students went through the same system of examination and are therefore more comparable than students from different regions. A new approach to this topic that reduces biases drastically would be then to choose respondents from a group that fulfills both the criteria of centralized high school examination and the criteria of centralized university examination.

Another way of improving results is to increase granularity of measured points in time. This study used two measures at two specific points in time to measure changes in performance. One point was at high school graduation and the second one was at graduation from university. This gives us an idea on how performance changed from one point to another but does not reveal how performance changed between these two points. Whether performance changed in a linear function or if there were peaks and valleys that correlate with specific happenings in an actor's social network remains hidden. For example, this study does not capture the possibility that friends change during the course of an actor's time at university. When individual performance is influenced by friends and friendships change the performance must hence change accordingly. There are several ways to improve the study and gain more insights. It is easy to just ask for more details in the survey itself. The survey can ask for more grades at different points in time or for more information about who was friend with whom at a given time. But too many questions in a survey slow down the process of asking respondents and this again decreases response rates as participants are less motivated to take part in the survey. We therefore need a way to increase amount of information but keep the level of time consumption and efforts the respondent has to take as little as possible. For example, we can ask each participant for permission to use the university's records on grades. This does not need any further efforts for the participant but gives us a lot more data bringing in a higher degree of precision. Despite that we still need data on who was a friend with whom and when but because friendships change not as much as exams are taken and grades are assigned, there is a likely chance that we can gain these records through questionnaires. To measure influence we could then compare ups and downs in performance among friends. When there is significant influence we might see a similar history of peaks and valleys.

A second approach to a longitudinal analysis is to follow students throughout their time at university. Over a period of three years at the end of each semester we could ask students about their grades and their friends during the last semester. This might be the highest degree of time and efforts to invest but will definitely unveil the most precise information.

There are other ways of improving the study. Besides increasing degree of detail over time we can ask for more details about how friends are truly related. Just being a

friend does not necessarily mean the same for everyone. For instance, we can ask for how much time spent per week with the friend in question. We can also ask for the various activities one was involved with a friend. Was it just having some beers in a bar, was it studying together for an exam or was it a more the sort of friend you meet at parties? The questionnaire can also ask for other external effects that could have an impact on grades. Students are involved in activities that do not count for their GPA but still take a lot of time and are valuable for their education. But these activities can have an impact on the GPA in such as they take up time used for studying thus generating lower marks due to poor preparation. But we have to keep in mind that these types of data tend to be hardly quantifiable and hard to define precisely. In order to create comparable data it is necessary that respondents have a clear understanding of what is asked for. If that is not full ensured, it's not possible to get homogeneous data. But in order to isolate effects homogeneous data is necessary. If one student understands friend as someone he or she is extremely close with and another respondent understands friend as someone he or she meets randomly at university, it is obvious that the data is biased as both "friends" have a different influence on the person in question.

Something that is definitely important is to simply increase the sample size. This study relies on a sample size of 63 qualified answers. In some cases where the whole group was taken into consideration that was enough to gain significant results but in other cases where only subgroups were compared, the size proved too small to derive any new and relevant information. The study that examined the influence among students of MBA programs used a sample of about 250 students and 62 teams to obtain significant results ([Timothy T. Baldwin 1997](#)). Another team of researchers conducted two multivariate studies of  $n=139$  and  $n=92$  cases ([Blickle 1996](#)). The most promising case that suffered from a sample size that was too small is the test for correlation between friends that performed better in high school than the actor him or herself and that actor's performance by selecting a subgroup of respondents that had a high school GPA of 1.9 in common. Testing this group revealed a moderate positive correlation between performance of friends and own improvement of 0.221 or 0.224 and additionally this group proved to have a significance level of about 0.650 or 0.594 depending on whether the friend's performance was measured using the *Relative Friend Performance Indicator* or the second approach. Compared to other test where the significance level was above 0.9 easily the smaller significance considering the small number of  $n=8$  ( $n=7$  respectively) test records qualifies these tests for further investigation. Considering the 250 cases from the MBA study and the 63 cases from this study we can conclude that an appropriate number of records starts at 100 and lays ideally around 200 and 300. Obviously the more records one has the better the results are but the available records show that this sample size can yield useful results.

Another suggestion questions the general way this study was conducted. What needs to be scrutinized is the idea that the correlation needs to be found between friends GPA and an actor's GPA. That the GPA reflects individual performance is not a question but it seems as if the GPA is under influence of a sheer uncountable number of external effects. This study shows that there is also a correlation between high school GPA and university GPA and one can easily think of other factors as personal motivation that change as personal aims for life changed. Maybe people outside the friend domain enter or leave a person's circle of influential people. This change in social network influence is not captured by this study. A solution could be to step back a step and look at what is in between the causal chain of social influence and GPA. As mentioned before there are special character traits that are influenced by one's social network and that have a significant influence on one's GPA. Instead of examining the change of a person's GPA we should look deeper as before at exactly these individual personality traits that define

the success or failure in university. We should first expand the domain of analyzing the social networks influence on character traits towards more parameters and maybe then we can suggest that social networks influence a single part of the large number of drivers which, as a unit, exert the final and important influence that affects a person's GPA. Blickle (1996), for example, have found that there are two specific personality traits that have a strong and significant influence on a person's academic performance. In their study on personality traits, learning strategies and performance they examined college students in order to find personality traits that drive academic performance. They suggest that conscientiousness correlates highly significantly with learning discipline (0.57 and 0.48) and, additionally, they say that openness to experience correlates significantly with elaboration (0.49 and 0.39). The suggestion is thus that instead of testing a social networks influence on a person's GPA it is more useful to test a social networks influence on both conscientiousness and openness to experiences. In case there is a significant correlation it is easier to reason a significant influence of social networks on academic performance. What could such a study look like and how can we measure one's social network's influence on these two traits?

In order to measure the degree to which a specific target person is either open to experience or conscientious Blickle (1996) relied on the NEO Personality Inventory (NEO-PI-R) as developed by Costa and McCrae in 1992 and updated to NEO PI-3 in 2010 (PAR 1992; PAR 2010). This framework enables researches to assess a person's personality according to the big five personality traits as described by McCrae and John (1992). Those five personality traits are *Extraversion*, *Agreeableness*, *Conscientiousness*, *Neuroticism* and *Openness*. These five factors can be further specified by various adjectives. For the two for us relevant dimensions Conscientiousness and Openness these adjectives are *efficient*, *organized*, *planful*, *reliable*, *responsible* and *thorough* respectively *artistic*, *curious*, *imaginative*, *insightful*, *original* and *wide interest*.(McCrae and John 1992).

Once we have assured that we can assess an individual's personality traits we can continue building our study around that. Just as measuring the impact of one's social network on that person's GPA we can examine a social network's influence on specific personality traits. We will use a longitudinal-analysis of changes of a person's personality and link this data to that person's social network. In order to accomplish that, we need to have at least two points in time where we assess a person's personality. The first assessment will be at start and the following assessments when we can assume that an influence has taken place. The differences between both or all assessments will be determined. Then, as mentioned before we will create a social network around the participants to see who is connected with whom. This will give us the opportunity to trace changes through the network. We can then look for correlations between specific positions within a network, between actors, specific subgroups of the network or within the whole network and changes in personality traits both in the domains of conscientiousness and openness.

The last suggestion for conducting future studies focuses on the group of influencing people a specific person has. So far this study concentrated on one's friends in university as the main drivers for influence. The question is whether this assumption is sufficient or whether this range of influential people needs to be extended. If we look around, it is obvious that one's personality is not only shaped by those friends one relates to during his or her studies. There are more people in everyone's live throughout each and every day that can have a significant influence. Evidence comes from other studies. For example, besides examining the influence of friends the studies that used the Framingham Heart Study to identify spread of behavior through social networks also investigated the influence of family members like spouses and siblings as well as

neighbors ([Christakis and Fowler 2007](#); [Christakis and Fowler 2008](#); [Fowler and Christakis 2008](#); [Nicholas A. Christakis 2010](#)). In order to include all influences of one's social network the number of alters needs to be increased and diversified. The question is: What is the number of alters needed to conduct a comprehensive study? What is the size of a person's social network? To which degree do individuals in one's social circle influence an actor and do all of them have the same amount of influence? Independently from what is practically possible the size of an actor's social network is unlikely to exceed the maximum number of social contacts one individual can maintain. There is a number referred to as Dunbar's number ([Dunbar 2010](#)). Dunbar claims to have found a maximum number of social contacts one can maintain and that this number is about 150. Still, researchers other than Dunbar are in doubts about the exact number but studies have shown that it resides between 150 and 300 ([Christopher McCarty 2001](#); [A. Hernando 2009](#)).

Despite the uncertainties regarding the degree of influence of individual actors, it is important to increase the domain of influences beyond friends for a single reason: We need to catch the whole set of influencing actor's of an individual to be fully able to include all relevant changes in his or her environment. In reality this might be close to impossible: If we believe Edward Lorenz, inventor of the butterfly effect, catching the whole set of influences an individual is exposed to, is almost impossible. The term "butterfly effect" is based in chaos theory and sensitive dependence on initial conditions. A common example to describe the butterfly effect is that of a ball placed at the crest of a hill. The ball, once let go, can roll in a number of different valleys depending on slight differences in its initial position. Likewise the slightest occasion in an individual's life can have a severe impact on his or her future ([Lorenz 1993](#)). Still, increasing the diversity of influencing people is one step towards more reliable results.

From an exaction perspective, what could be improved when conducting future studies? We can assume that getting participants to respond to surveys is and will be a difficult task that will not change in the future. Unless any social context is present that allows for social punishment, respondents will not change their behavior and respond to inquiries faster, more often or more reliably. It will stay a difficult and labor-intensive task to collect reliable data in a quantity that enables researchers to draw significant conclusions. What could have been improved to drive participation numbers are both the scope of participants and the way and frequency new snowball-effects were triggered. So far the scope was limited to students of Witten/Herdecke University. Because of the university's relatively small number of students the study's circle of potential respondents was already limited. Moving the study to a larger university with a multitude of students and respective environmental conditions such as type of degree, interconnectivity and density of the student body can supply researchers with more significant data. Also of importance is in which way the actual, already existing, network is tabbed. Because the snowball-system repeats itself and the growth and therefore the respondents that participate are fully depending on the underlying social network's structure, it is important if not even result-changing in which way the snowball-process is started. If initially two students are asked to participate that reside close to each other in terms of the social network's structure, it is highly likely that there will be a lot of redundancy of nominations and thus redundancy of potentially next participants, slowing down the process of iteration. It is hence recommendable to trigger the initial sequence of asking participants to respond by making sure that those participants are unlikely to have many common network contacts or are otherwise closely related to each other. For example, this could be achieved by triggering the process within the same faculty of a university but in different years of study. Or if we stick to the same year it could be advisable to initiate the system in two different

faculties. On the contrary it is important to not include students that are too far away from each other. It is important to have a complete network to draw any conclusions. Despite the fact that there is just one social network and that we can always only determine a sub network which then will eventually connect to a larger, yet still sub network, it may take too much time to connect all sub networks together in case they were triggered at spots that are not close or closely knit enough.

This design problem during the phase of evaluation can be addressed by first developing relevant models for calculation followed by in-depth testing of these models with sample data. Then it is important to refine the models until they return the desired results. Only when the statistical models are created and tested, one can continue with designing the appropriate questionnaire that is able to collect the data required. Approaching the task in this way ensures that we do not lack important pieces of information when computing the models with real data later on.

## 2.5 Results

The introductory question was whether a person's performance in life depends on with whom he or she is befriended or not. After having examined different cases involving the data from the online survey we can conclude that there is indeed proof that this dependency exists.

I found a vague correlation between an actors' GPA and his or her surrounding networks' GPA average. As expected due to the small sample the results are statistically not significant, but still give an idea on the direction the research goes. I found a positive correlation of 0.436 (0.713 significance,  $n=3$ ) for the variables most-central-actor's GPA and subgroup GPA average.

When using eigenvector as the method of determining the most central actor one could assume the correlation to be stronger than for InDegree centrality. My research does not follow this assumption revealing a smaller positive correlation of 0.132 with less significance of 0.916 considering the same  $n=3$ .

In the second test run I found that the different GPAs of friends lie closer to each other than the GPAs that can be found in a full social network. The average standard deviation within first-level-friendships is 0.2507 and 0.2492 within the twelve high-density test groups. The random control group's standard deviation is 0.2730 and the whole network shows a standard deviation of 0.2867.

From an ego-centered analysis we can deduct that the sum of all friends being better or worse in their high school GPA than the actor (represented by the *Relative Friend Performance Score*) are a strong indicator to the actor's performance. The test data revealed a highly significant (0.001) medium to high correlation of 0.684 ( $n=63$ ) suggesting that actors with many better performing friends improved more than actors with less better performing friends.

Unfortunately these results are highly controversial and need to be questioned. There is a logical flaw in the argumentation. The flaw resides in the fact that there is a highly negative correlation of -0.873 (0.001 significance,  $n=63$ ) between low high school grades and improvement from high school to university. Students that performed poorly in high school tend to improve more than students that were high achievers in high school. Once thought about, the flaw is obvious: Having many friends that performed better in high school is a significant indicator (0.764 correlation, 0.001 significance and  $n=63$ ) for having performed poorly. It is easy to see and it is simple logic that the worst student will only have friends that did better and the best student will only have friends that did worse than him or her. That implies that there is a higher probability of having lots of better friends if your own performance was poor. The *Relative Friend Performance Score* is therefore, as shown above, a reversed proxy for

high school grade. If we now look at the high correlation between high school grades and improvement alone we can assume that we can find the reason for the individual improvement in performance within the high school grade and that the *Relative Friend Performance Score* is less reliable when predicting improvements and influence than assumed. When we control for high school GPA it reveals that the correlation between *Relative Friend Performance Score* and individual improvement is almost nonexistent (0.056 correlation, 0.668 significance, n=60). But instead of mathematically ignoring the influence of the individual high school performance in

	<b>Standard deviation</b>	<b>Correlation</b>	<b>Significance</b>	<b>N</b>
<b>Most-central-actors vs. sub group GPA Average (InDegree)</b>		0.436	0.713	3
<b>Most-central-actors vs. sub group GPA Average (eigenvector)</b>		0.132	0.916	
<b>First-level-friendships</b>	0.2507			
<b>High-density test groups</b>	0.2492			
<b>Random control groups</b>	0.2730			
<b>Whole network</b>	0.2867			
<b>Individual improvement vs. Relative Friend Performance Score</b>		0.521	0.186	8
<b>Individual Improvement vs. the friend's performance as a whole</b>		0.221	0.650	7

**Tbl. 1: Results of statistical analysis**

order to remove that bias we can also remove the bias by only selecting cases that include the same high school GPA. In the data sample the GPA represented by the most cases was 1.9 which was mentioned in eight cases. If we calculate the correlation between *Relative Friend Performance Score* and performance improvement for these  $n=8$  cases we are presented with a 0.521 correlation at a low significance of 0.186. Because of the low significance we cannot assume the correlation to be statistically correct but considering the small sample size we can clearly see the tendency.

Despite the bias in the overall correlation between *Relative Friend Performance Score* and individual performance we can still conclude that there is a small shadow of light regarding an existing correlation when considering cases having the same underlying preconditions. Yet, this correlation needs proper research and confirmation.

And finally, I found that just relying on the friends' average GPA's as the relative performance indicator is less significant than the *Relative Friend Performance Score*.

I calculated a medium positive correlation of 0.579 (0.001 significance,  $n=59$ ) between an actor's improvement in performance and having either friends that performed better or worse. We have to keep in mind that the same bias applies as mentioned before: Actors also improve when having performed poorly in high school. Controlling for high school GPA reveals no existing correlation (-0.004 correlation, 0.974 significance,  $n=56$ ). But using cases with the same high school GPA and thus the same preconditions I could find a slightly positive correlation of 0.221 (0.650 significance,  $n=7$ ). As already mentioned before, this needs strong verification and confirmation by using a larger sample of data.

Still, it is difficult to constitute a definite answer. In most cases the underlying data was not large enough to return any statistically significant results. The influence of network-wide InDegree and eigenvector centrality was calculated from a sample base of three samples. The ego-centered influences of friend's GPAs on own performance was lastly calculated from a sample of eight valid data sets. Despite the fact that the methods and models applied are able to deliver relevant results, the number of sample records needs to be increased in order to allow deep and meaningful insights.

With regard to GPA standard deviation among friends and within the whole network it is necessary to keep in mind the biases that follow self-selection. Self-selection biases are a common problem in sociology. Generally, these are biases that evolve when groups of research participants are formed in a way that implies imbalanced, non-randomly appearances of specific attributes. For example, in case we start a similar self-repeating, social network based study and the first participant is student at a girls-only high school we obviously end up with an imbalance of participants towards girls. Because it is a widely known form of bias, many scholars have dedicated significant time researching this topic ([Kleinbaum, Kupper et al. 1982](#); [Bergstrand, Vedin et al. 1983](#); [Benfante, Reed et al. 1989](#); [Sogaard, Selmer et al. 2004](#)). In our situation the self-selection bias might occur because friendships are not set up randomly but friends rather find themselves due to common features. Moreover friends tend to be similar towards "their school attitudes, their educational aspirations, and their actual achievement" ([Berndt 1982](#)). The question that needs to be clarified is therefore, what direction of cause is dominant. Is the standard deviation of GPAs among friends lower because individuals tend to look out for similar individuals to be friends with or is it rather the case that friends tend to influence each other towards their achievements?

Concerning the influence of the most-central actor on the sub network we should keep in mind why most researchers consider centrality to be influential. Here, in this study, the argumentation is that being central equals being influential because centrality comes from the number of nominations an actor receives from others. The more actors vote on someone as being influential the higher that person's centrality index is. In common literature on social network analysis one of the fundamental explanatory tenets is that centrality implies influence because centrality implies access to relevant and valued resources ([Timothy T. Baldwin 1997](#);

[Raymond T. Sparrowe 2001](#); [Granovetter 2005](#)). There is a discrepancy in the backgrounds on why centrality is influential. Both explanations are valid because they ground on different types and different functions of the underlying network. In this study we do not examine the flow of novel information or information as such. Instead, we try to find out what influences arise from actors that are directly nominated as influential. The indicator here is the nomination of an actor by someone else and not the mere fact of friendship. At its base centrality is the number of connections one actor has to others<sup>7</sup>. This holds true independently of what a connection symbolizes. If it symbolizes the flow of information, the most central actors have access to the most information. If it symbolizes direction of influence, the most central actors are those of the highest influence.

### 3 Discussion

With every study in every field of research it is important to evaluate the results towards their relevance with respect to existing research and practical application. In this chapter I will elaborate on implications for business practices, the studies' limitations and strength and I will give suggestions for further research. I will close my work with a conclusion. The part regarding implications for business practices is intended to put the results in a professional context. I will give suggestions on how team management and team building should be done in order to improve team performance. I will then continue writing about the limitations and strength of this work from a general perspective. The questions here are to what extent we can trust the results and how far we can go to ground future action on these results. The section on future research is then intended to refer to implications for researches conducting similar studies in the future as well as researchers in the same field of social network and performance analysis. As the last part I will give a wrap-up on the study, its results and its future implications.

#### 3.1 Implications for business practices

Practical recommendations on team management and team building can be found constantly in research literature. In current research there is a relatively large body of research on team performance. Paris et. al. (2000) found about eight concepts of teamwork and team performance. Cohen et. al. 1997 selected 200 articles and focused on 54 studies that included measures of team effectiveness between January 1990 and April 1996. In their review they present "a heuristic framework for team effectiveness that illustrates recent trends in the literature". Yang et. al. (2004) evaluated team performance of information system development teams and the respective social network structure to find parallels.

Keeping in mind the results of studies by Christakis and Fowler, we can draw recommendations for business practices from two main aspects of the study: First, there is the introduction of the *Relative Friend Performance Score* in combination with the standard deviation model and second, there is the concept of self-selection.

The *Relative Friend Performance Score* is a measure indicating to which extent any individual is under influence of his or her peers. Peers that perform better, have a positive influence on the individual's performance and peers that perform less have a negative influence on the individuals' performance. In other words, when working in a team, the top-performers tend to drop in performance whereas the low-performers tend to improve. Consequently, the team member's individual performances will eventually be brought into

---

<sup>7</sup> In reality centrality is more complex enabling researchers to differentiate between incoming and outgoing connections as well as to do recursive calculations of connections with individual weighting of the sender or recipient (Wasserman and Faust 1995).

line and will center on the average performance of all team members. When we look at the standard deviation model we can see that friends tend to be closer in performance than on average, thus supporting this assumption. This finding has an influence on how teams should be assigned. We can say that teams should be assigned with the required output later on. Top-performers can increase the output; low-performers can decrease the output. But eventually the performances of each team member will be aligned and reside on an average level. We can say that team performance is a null sum game. A low-performer can make the added value of a top-performance irrelevant. Hence, when setting up project teams the expected outcome will be the same as if all team members were average-performers. A side question that is imposed here is that of the lasting of performance. Do performance decreases of top performers that happen within a team last even after the team formation has been resolved? Or do changes in performance only happen while the team is active, having no impact on further performance?

Attention should also be paid to the overall setting of the team. Just putting together members according to their performance to get a desired output is not enough. There might be the case where top performers are not motivated enough to work with low performers and thus reduce the effort of their work or limit access to information and resources. In a different setting, low performing team members may be hesitating putting forth their ideas or requesting information and resources from top performing team members due to a lack of credibility. Another reason might be that “high-ability groups will exceed the additive capacities of individual team members” (Bowers, Pharmer et al. 2000) again resulting in a drop of performance. This can be due to various extrinsic and intrinsic motivations. Researchers say that “interpersonal dimensions may restrict the performance of diverse teams” (Bowers, Pharmer et al. 2000). It is therefore necessary to keep a certain level of equality within a team to prevent drops in performance caused by the team setting yet allowing enough space for team members to benefit from their differences.

### **3.2 Limitation, strength and further research**

When we look at this study and this paper from a more distant perspective we can see limitations and strength the lie herein. With every work it is important to measure and define the extent to which it has an impact on current and future research. With regard to this paper we can conclude the following. By applying methods of social network analysis to deduct a study on influence among students on performance, this paper further fosters awareness for the field of social network influence research. It emphasizes the importance and relevance of conducting studies on the question to which extend friendship is an indicator for influence when it comes to individual academic performance. By doing so, it opens up new ideas on how to approach social network influence studies that are focused on performance evaluation. Especially, it extends the theme of social influence research into the domain of students.

A first field test gives suggestions that have been tested in a real environment. By conducting a real study as the foundation for this paper, relevant information, knowledge and key experiences can be shared that have been made and acquired during the phases of conceptualization, execution and evaluation. By reaching out to use online questionnaire and web technology to apply statistical models and social network analysis to students and academic performance, this paper is able to give recommendations and suggestions on how to further improve the methods and models applied. Future research will profit from these experiences as basic and fundamental mistakes during the design phase can be prevented. This study does not provide full and comprehensive research results but is a clear indicator for future research. It gives profound reason to conduct further and more fundamental research projects in the field of social network influence and social network analysis. It serves as a first introduction of future researches into the field of social network influence analysis, explaining a specific way of conducting a study and what results one can expect.

This paper raises a significant number of questions that remain unanswered. These are possible cases where future research can hook in and further foster the examination of social network influence. Research should therefore pay attention to and consider the key outcomes of this study. Basic issues are a way to decrease the creation of biases during the phase of conceptualization. What are main drivers of biases and how they can be eliminated? Are private versus public universities an important factor? How do we reduce the inaccuracy that is brought in by too subjective means of grading high school students? To which extent are the results distorted by human teachers evaluating students' performance? Also of importance is the degree to which influence on academic performance is transported through the network. Christakis et. al. discovered that most the time influence keeps within a distance of three degrees (Christakis and Fowler 2007; Christakis and Fowler 2008; Fowler and Christakis 2008; Nicholas A. Christakis 2010). A friend of a friend of a friend is the last one that has an influence on oneself. Does this rule also apply to the subject of academic performance? Also of importance is the question, to which degree the findings of this paper are valid in a working environment. Do people behave the same way in a professional environment? Is professional performance as much influenced by the own social network as academic performance among students? If yes, what else are recommendations and suggestions that businesses should consider in their day-to-day operations? If not, what are the main drivers for professional performance from a social network point of view? Is it possible to apply these findings to form recommendations for governments on educational politics? Should school kids be separated in school to allow the higher performing students to keep their level or should all pupils attend one school allowing the low performing children to improve further?

### 3.3 Conclusion

This paper extended the idea of influence due to relations within a social network to the new field of academic performance measurement among students. The underlying question was, whether academic performance is under influence of one's social network of friends. In order to answer these questions this paper explained the conducted study and its findings and gave suggestions and recommendations to improve studies of this kind in the future. I first introduced the GPA as a quantifiable and objective way to measure academic performance. The advantage was that all students underwent the same evaluation through high school and GPA and when comparing high school GPA and university GPA a longitudinal aspect came into play. In order to measure the influence that is evident I conducted a web-based survey that collected information of GPAs and friendships of each participant and created a social network from the resulting data.

Statistical models were used to extract relevant information from the existing data. The results can be divided on either a network-wide or an ego-centered perspective. Regarding the network-wide perspective I found a positive but highly insignificant correlation (0.437 correlation, 0.713 significance,  $n=3$ ) between the grades of the most central actor of a social network and the social network itself. When looking at how far away GPAs of close friends lie, I could find that GPAs of friends lie closer to each other than on average of a random control group. Still, we have to keep in mind that these results might be biased by self-selection as people tend to be friends with those who are similar to themselves. On an ego-centered view I introduced the *Relative Friend Performance Score* that allows quantification of social impact on academic performance. The score itself is calculated by increasing the value by one for each friend that does better than the actor him or herself. For each friend that does worse than the actor the value is decreased by one. In the end the correlation between the *Relative Friend Performance Score* and the degree to which a specific actor increased his academic performance lies at 0.521 (0.186 significance,  $n=8$ ). In a second ego-centered test I used the group of friends as a whole in comparison to the actor to draw conclusions on to which extent the group of friends is influential towards the actor's academic

performance. Here the average GPA of the group was compared to the actor's GPA indicating whether the friends performed better or worse in high school than the actor him or herself. In case the friends did perform better, an increase in performance was expected. In case the friends did perform worse, a decrease was expected. The results for this test featured a correlation of 0.221 (0.650 significance,  $n=7$ ).

Interpreting the results I found that the model itself is usable and applicable but the number of records used needs to be increased in order to allow deep and meaningful insights. Also self-selection biases can be a significant source of errors in calculations as the participant collection process is self-triggered and therefore depending on where the process was started. Also the definition and application of centrality is different in this study. Researchers generally refer to centrality as a specific access to information rather than being influential. In this study the concept of centrality is applied to influence still following the definition of centrality.

Regarding recommendations for future research the paper refers to the phases of conceptualization, execution and evaluation. The snow ball system that helped driving the number of participants proved useful in this study. Privacy issues were an aspect that was important to think about from the very beginning. Important with regards to biases is, to further increase the degree of objectivity. In order to accomplish this goal it is necessary to examine students that underwent centralized examinations in both high school and university. Further, to improve results it is important to increase granularity of measured points in time. This could be realized by using transcripts of record of each participant provided by the university and further details about friends in the questionnaires. Another way to accomplish that is to follow students throughout their time at university observing the evolution of friendships and grades as they happen. Of fundamental importance is to increase sample size. Most the time results lacked significance due to a small number of participants. Increasing that number can help receiving more relevant results. The study also reveals that it might be useful not to directly examine changes in the GPA but to look at changes in specific character traits that have the most impact on the GPA. These character traits are *conscientiousness* and *openness to experience*. The advantage of this approach is that it offers a way to better isolate the influential drivers for academic performance. Following that, it is important to increase the diversity of influencing people to gain more reliable results. Reason for that is the fact that a person's individual academic performance is not just influenced by that person's friends but also by other people in his or her environment.

With respect to the phase of execution among important positive factors was the availability of contact details for possible respondents that simplified execution. A lean design of the questionnaire and a thoughtful personalization of invitation e-mails as well as mentioning urgency and building on social relationships drove response rates to the maximum. In order to fill gaps in the social network that were the result of the self-driven selection process of participants it was necessary to follow up on invitation e-mails manually. This proved especially important when the person in question was sitting at a from a social networking perspective central position. This task emerged as the most difficult and time consuming aspect of the data collection but is one of the difficulties that remain unsolvable as habits of potential participants are not changeable. In order to increase participation numbers it is also valuable and useful to increase the scope of participants and the way and frequency new snow ball effects are triggered. Methods to succeed here are to either include larger universities where student numbers within a single subject are higher or include more students vertically by letting different years and generations participate.

Lastly, during evaluation it can be stated that underlying models were defined too vaguely resulting in missing data when doing calculations and comparisons. This design problem can be addressed by developing and testing relevant models in advance and then proceeding towards questionnaire design.

The results of this study are also applicable to business practices. Regarding team building and team management we can assume a similar influence from peers on actors. Consequently, team member's individual performance within a team will be brought into line with other team members resulting in an overall team performance that is the same as if all team members were performing on average. We can therefore state that team performance is a null sum game where low performing members take up the value added by high performing members. But even though team performance zeros in on an average performance, too heterogeneous teams are underperforming as there are social conflicts that prevent individual team members to exert their full potential. It is therefore necessary to maintain a certain level of equality within a team yet allowing enough space for team members to benefit from their differences.

Beyond recommendations for business practices this paper is a field test that has been tested in a real environment and therefore functions as a model for future research. It does not provide full and comprehensive research results but is a clear indicator to conduct further and more fundamental research in the field of social network influence. Among issues for further research there is a way or method to decrease the creation of biases during the phase of conceptualization. There are also the questions whether the many times postulated three-degree-of-influence rule also applies to academic performance and whether professional performance is as much influenced by social networks as is academic performance. Lastly, the question for applicability to governmental and public issues such as education politics is raised.

Social network analysis is regarded as a powerful way to examine social behavior and it is hoped that this research could contribute to and inspire future research.

## 4 References

- A. Hernando, D. V., C. Vesperinas, M. Abad, A. Plastino (2009) "Unravelling the size distribution of social groups with information theory on complex networks."
- Azarian, G. (2005). The general sociology of Harrison C. White: chaos and order in networks, Palgrave Macmillan.
- Benfante, R., D. Reed, et al. (1989). "Response bias in the Honolulu Heart Program." Am J Epidemiol **130**: 1088 - 1100.
- Bergstrand, R., A. Vedin, et al. (1983). "Bias due to non-participation and heterogenous sub-groups in population surveys." J Chronic Dis **36**: 725 - 728.
- Berndt, T. J. (1982). "The Features and Effects of Friendship in Early Adolescence." Child Development **53**(6): 1447-1460.
- Bernstein, B. (2000). Pedagogy, symbolic control, and identity: theory, research, critique, Rowman & Littlefield.
- Blickle, G. (1996). "Personality traits, learning strategies, and performance." European Journal of Personality **10**(5): 337-352.
- Bowers, C. A., J. A. Pharmed, et al. (2000). "When Member Homogeneity is Needed in Work Teams." Small Group Research **31**(3): 305-327.
- Cacioppo, J. T., J. H. Fowler, et al. (2009). "Alone in the crowd: The structure and spread of loneliness in a large social network." Journal of Personality and Social Psychology **97**(6): 977-991.
- CAROL R. PARIS, E. S. a. J. A. C.-B. (2000). "Teamwork in multi-person systems: a review and analysis." ERGONOMICS **43**(8): 1052 1075.
- Chamorro-Premuzic, T. and A. Furnham (2003). "Personality traits and academic examination performance." European Journal of Personality **17**(3): 237-250.
- Christakis, N. A. and J. H. Fowler (2007). "The Spread of Obesity in a Large Social Network over 32 Years." New England Journal of Medicine **357**(4): 370-379.
- Christakis, N. A. and J. H. Fowler (2008). "The Collective Dynamics of Smoking in a Large Social Network." New England Journal of Medicine **358**(21): 2249-2258.
- Christopher McCarty, P. D. K., H. Russell Bernard, Eugene C. Johnsen, Gene A. Shelley (2001). "Comparing Two Methods for Estimating Network Size." Human Organization **60**(1): 28-39.
- Dawber, T. (1980). The Framingham study: the epidemiology of atherosclerotic disease, Harvard University Press.
- Deutskens, E., K. de Ruyter, et al. (2004). "Response Rate and Response Quality of Internet-Based Surveys: An Experimental Study." Marketing Letters **15**(1): 21-36.
- Duckworth, A. L. and M. E. P. Seligman (2005). "Self-Discipline Outdoes IQ in Predicting Academic Performance of Adolescents." Psychological Science **16**(12): 939-944.
- Dunbar, R. (2010). How Many Friends Does One Person Need?: Dunbar's Number and Other Evolutionary Quirks, Faber and Faber.
- Erickson, B. H. (1988). The Relational Basis of Attitudes. Social structures: a network approach. S. D. B. Barry Wellmann. New York, Cambridge University Press: 99-121.
- Falk, A. and A. Ichino (2006). "Clean Evidence on Peer Effects." Journal of Labor Economics **24**(1): 39-58.
- Fowler, J. and N. Christakis (2008). "Estimating peer effects on health in social networks: A response to Cohen-Cole and Fletcher; and Trogdon, Nonnemaker, and Pais☆." Journal of Health Economics **27**(5): 1400-1405.

- Fowler, J. H. and N. A. Christakis (2008). "Dynamic spread of happiness in a large social network: longitudinal analysis over 20 years in the Framingham Heart Study." Bmj **337**(dec04 2): a2338-a2338.
- Freeman, L. (2006). The Development of Social Network Analysis. Vancouver, Empirical Pres.
- Friedkin, N. E. (1993). "Structural Bases of Interpersonal Influence in Groups: A Longitudinal Case Study." American Sociological Review **58**(6): 12.
- Granovetter, M. S. (1973). "The Strength of Weak Ties." The American Journal of Sociology **78**(6): 1360-1380.
- Granovetter, M. S. (2005). "The Impact of Social Structure on Economic Outcomes." The Journal of Economic Perspectives **19**: 33-50.
- Howe, J. (2006). "The Rise of Crowdsourcing " Wired (14.06) Retrieved 18.11.2010, 2010, from <http://www.wired.com/wired/archive/14.06/crowds.html>.
- Ickes-Dunbar, A. (2004). "Beyond the Dunce Cap: "What's My Grade?"" Phi Kappa Phi Forum **84**(4): 3-7.
- Kirkpatrick, D. (2010). The Facebook Effect: The Inside Story of the Company That Is Connecting the World, Simon & Schuster.
- Kleinbaum, D., L. Kupper, et al. (1982). "Selection Bias." In: Epidemiologic Research: 194 - 219.
- Knoke, D. and S. Yang (2008). Social network analysis, Sage Publications.
- Levinson, J., J. Levinson, et al. (2007). Guerrilla marketing: easy and inexpensive strategies for making big profits from your small business, Houghton Mifflin.
- Lorenz, E. (1993). The essence of chaos, University of Washington Press.
- McCrae, R. R. and O. P. John (1992). "An Introduction to the Five-Factor Model and Its Applications." Journal of Personality **60**(2): 175-215.
- Milgram, J. T. a. S. (1969). "An Experimental Study of the Small World Problem." Sociometry **32**(4): 425-443.
- Miller McPherson, L. S.-L. a. J. M. C. (2001). "Birds of a Feather: Homophily in Social Networks." Annual Review of Sociology **27**: 30.
- Neufeldt, V. and D. Guralnik (1994). Webster's New World dictionary of American English, Prentice Hall.
- Nicholas A. Christakis, J. H. F. (2010). Connected: The Surprising Power of Our Social Networks and How They Shape Our Lives: The Surprising Power of Social Networks and How They Shape Our Lives, Little, Brown and Company.
- PAR, I. (1992). Revised NEO Personality Inventory (NEO PI-R) and NEO Five Factor Inventory. Professional Manual. Lutz, FL, Paul T. Costa, Jr., Robert R. McCrae.
- PAR, I. (2010). The NEO™ Inventories: NEO™ Personality Inventory-3 (NEO™-PI-3) Professional Handbook. Lutz, FL, Paul T. Costa, Jr., Robert R. McCrae.
- Peter V. Marsden, N. E. F. (1993). "Network Studies of Social Influence." Sociological Methods & Research **22**(1): 127-151.
- Provine, R. R. (2005). "Contagious yawning and laughing: Everyday imitation- and mirror-like behavior." Behavioral and Brain Sciences **28**(02): 142-142.
- Raymond T. Sparrowe, R. C. L., Sandy J. Wayne and Maria L. Kraimer (2001). "Social Networks and the Performance of Individuals and Groups." The Academy of Management Journal **44**(2): 10.
- Shattuck, A. L., E. White, et al. (1992). "How women's adopted low-fat diets affect their husbands." Am J Public Health **82**(9): 1244-1250.
- Sogaard, A., R. Selmer, et al. (2004). "The Oslo Health Study: The impact of self-selection in a large, population-based survey." International Journal for Equity in Health **3**(1): 3.

- Timothy T. Baldwin, M. D. B. a. J. L. J. (1997). "The Social Fabric Of A Team Based M.B.A. Program: Network Effects On Student Satisfaction And Performance." Academy of Management Journal **40**(6): 1369 - 1397.
- Wasserman, S. and K. Faust (1995). Social network analysis: methods and applications, Cambridge University Press.