UNIVERSITÄT PADERBORN

Betriebswirtschaftslehre, insb. Corporate Governance

HEINZ NIXDORF INSTITUT
UNIVERSITÄT PADERBORN

Determinants and outcomes of the clustering coefficient in social networks

–

Insights from an empirical, longitudinal study

Dissertation im Fach Corporate Governance

Themensteller: Prof. Dr. René Fahr

von:

Diplomkaufmann Philipe Gerlach

Paderborn 2025

Für meine Eltern

Danksagung

Ohne den Austausch und die Unterstützung zahlreicher Personen hätte ich diese Arbeit nicht abschließen können. Wie bei jedem Projekt ist es die Summe der Einzelteile, welche zum Erfolg führen, unabhängig von der Wichtigkeit der einzelnen Aufgabe. Den folgenden Personen gilt besonderer Dank.

Der größte Dank gilt meinen Eltern, die mir eine außergewöhnliche universitäre Ausbildung ermöglichten, welche die Basis für diese Arbeit darstellte. Ohne sie hätte ich nicht eine bis heute andauernden Begeisterung für Forschung entwickelt und nicht den Mut gefasst, ein solches Projekt anzugehen. Ihre ständige Bestärkung, trotz Widerständen auf dem richtigen Weg zu sein und Chancen zu nutzen, sind ein wesentlicher Baustein für den Abschluss dieser Arbeit.

Mein Doktorvater René Fahr gab mir die Chance, mich mit einem Thema zu beschäftigen, das sowohl fordernd als auch erfüllend ist. Ohne mich unter Druck zu setzen, hat er mich über einen langen Zeitraum konstruktiv unterstützt, sodass ich stets ein Licht am Ende des Tunnels sehen konnte.

Meinem Zweitgutachter Matthias Trier gilt besonderer Dank, da er aufgrund seiner Expertise den Wert der Arbeit für das Forschungsfeld herausgestellt und die Arbeit innerhalb seines Themengebietes eingeordnet hat. Dies hat deutlich zum Abschluss und zu meinem Selbstverständnis der Arbeit beigetragen.

Ich danke Behnud Djawadi und Philipp Herrmann für die Beantwortung der vielen Fragen rund um Stata, den Datensatz und die winzigen analytischen mikroökonomischen Details der Auswertungen.

Ich danke Daniel Böhmer für die Programmierung der Extraktions- und Aggregationssoftware. An diesem Thema sind viele gescheitert, er hat einen wesentlichen Beitrag zum Gelingen der Datenaufbereitung geleistet.

Melanie Fearn danke ich für das unermüdliche Organisieren von Abstimmungsterminen. Ebenso danke ich Charlotte Neuhäuser für das Überprüfen der Literatur.

Der wichtigste Dank aber gilt meiner Frau und meinen Kindern. Meine beruflichen Aktivitäten zwangen mich, meine Tätigkeit an dieser Arbeit in die Freizeit zu verlagern und dafür andere wichtige Dinge zurückzustellen. Meine Frau Petra Gerlach hat nie Zweifel daran gezeigt, dass ich diesen Weg gehen kann und ihn sogar durch die Herstellung des Kontaktes zu meinem Doktorvater initiiert. Die Diskussionen Ihrer eigenen Forschung haben mich erst in Tiefe für das Themengebiet begeistert und mich mit dem notwendigen Respekt an meine Forschung gehen lassen.

Table of Contents

I. List of Figures

II. List of Tables

III. List of Abbreviations

MIT Massachusetts Institute of Technology

MMPORG Massively Multiplayer Online Role-Playing Games

IV. List of Definitions

V. List of Appendices

1 Introduction

1.1 Motivation

Networks are a fundamental constituent of human life. Our everyday lives are embedded in and facilitated by both technical and social networks - at home or on the road, at work or at play, when socializing or travelling. This thesis is concerned with social networks. Social networks consist of actors and their actions and communications (Burt, 2005; Easley and Kleinberg, 2010; Jansen, 2006; Wasserman and Faust, 1994). Referring to social networks usually implies that someone benefits from greater access to relevant information, which in turn leads to better career chances or improved efficiency. In social network research the improvement potential of networks has been termed *social capital* which, in this context, is defined as

"The advantage created by a person's location in a structure of relationships" (Burt, 2005, p. 4)

Definition 1: Definition of social capital by Ronald Burt

We examine a work-related network as one possible expression of social networks and try to deeper understand the drivers of performance in it.

Put differently, we examine if an actor's position in a work-related social network becomes a potential contribution to his[1] performance, career development or even bonus payments. To evaluate an actor's position within a network, social network research has determined and developed a set of specific measures (Jansen, 2006; Wasserman and Faust, 1994). One group of measures aims to evaluate different degrees of density of a network surrounding an actor, the so called ego-centered network (Jansen, 2006, pp. 105–110). More specifically, the so-called clustering coefficient describes the number of realized relations between an actor's direct friends in relation to the potential number of relations between these friends. If a person has 4 friends, of which

[1] For improved readability we chose the masculine form for all networks actors. This form can adequately be exchanged against the feminine form.

2 are friends themselves, the clustering coefficient would be 1/3[2] (Watts and Strogatz, 1998, p. 441). This measure can vary from 0 to 1, which equals a situation with no relations up to each possible relation being established. Recent research tries to find answers to the question of whether situations of low or high clustering coefficients result in higher performances of individuals and/or groups.

The classical opposing views of Ronald Burt and James Coleman who see either the absence of closure, structural holes (Burt, 1992), or dense networks with closure (Coleman, 1988) as a source of social capital, become more integrated and differentiated in recent research, as the explanatory value of each single concept is limited. Scholars like Gargiulo, Benassi, Ertug, Galunic and even Burt himself, try to compare or integrate both views or strive to show that one of the constructs is a greater source of social capital in a specific context. We believe that the clustering coefficient is a valuable predictor for the performance of network members in knowledge networks – applied to a suitable set of data. The search of a suitable dataset was one of the major challenges we had to solve for this piece of research. We set out to analyze a social network which meets two different criteria. The network had to be large enough to create significant and robust results and enable the monitoring of the assumed effects not only (Ronald S Burt, 2001; Galunic et al., 2012; Gargiulo et al., 2009; Gargiulo and Benassi, 2000) at a single moment in time. As most of the related research refers only to a very limited set of data, from single small companies to several teams, we were looking for a dataset that consists of a large social network. The motivation was mainly driven by the intention to reduce specific, single effects tied to a specific dataset or group. Often these effects are additionally diluted by the questionnaires that are the basis of most of these empiric evaluations. Furthermore, a majority of the quoted studies focuses on contexts that have a specific framing of cooperation or competition, as investment banks, research networks, neighborhoods or high schools. We believe that social capital is built in different ways if the context of a network changes. We assume that there is a difference between planning a neighborhood barbecue or running an investment

---

[2] 4 friends do have a maximum of number of possible friends of 6 ($\frac{friends*(friends-1))}{2}$). Divided by the number of realized friendships between friends this results in a clustering coefficient of 1/3 ($\frac{friendsfriends}{possiblefriends}$).

bank. We categorize one context as being less competitive compared to another. As a consequence the characteristics of a favorable network position also changes due to the context of the network itself.

Another dataset restriction refers to the assumption that we believe that hierarchical level and membership duration both have an effect on social capital, which is built on and interacts with the clustering coefficient. As a result, not every actor will act in the same way and may – over time – change his behavior. For us one of the most significant downsides of the research we evaluated were the limited timespan of observations, the deduction of the dependent variable and the relatively small sample, which was examined. Our dataset enables us to examine a large network over time while the network itself generates the dependent variables, as the common performance or activity measure. The same is valid for the network data. We can raise original network data, instead of generating network data from questionnaires or secondary data as e-mails or project memberships. In a next step this data can be analyzed to evaluate the effect of different degrees of embeddedness (high or low clustering coefficients) on the performance and activity of network members.

This piece of research adds additional empirical evidence to existing social network research and gives insights into the mechanisms that can lead to the creation of social capital in certain contexts. We try to add new insights to the mode of action of the clustering coefficient, especially concerning its direction of action, the impact on the activity and performance of the network members and its interaction with other variables. We are convinced that the mode of action of the clustering coefficient is not describable in simple terms as “the higher the clustering coefficient, the higher the performance” and we thereby strive for a better understanding of the mode of action of this measure on the one hand and a better understanding of the restrictions, constraints or interactions that can occur, on the other. We are convinced that the requirements of network positions change over time as careers evolve and duration of membership extends. Consequently, we consider the clustering coefficient to be more a dynamic than a stable predictor of success in the knowledge network we are examining. While we assume that a high clustering coefficient is favorable at the start of a new network member, we also assume that a low clustering coefficient is more favorable at later career stages or for senior network members.

1.2 Recent social network research fields and focus of own research

1.2.1 Social capital research in the context of social networks research

Besides the overall appearance of social networks, it can easily be stated that social network research is *en vogue*. Three major aspects mainly drive this fact:

1. The collection of network data has never been easier, Internet-based platforms and large databases generate large amounts of network data, to cover a large variety of research topics.
2. The possibilities of computer-aided analysis of social networks have never been better than now. Established tools[3] such as SNAP, UCINET, Pajek or recent methods developed by institutions like the MIT and Harvard allow the analysis of large networks and providing robust results.
3. Interest in human behavior and the role of social networks to reflect on and influence decision making and the action of individuals, groups and whole societies is growing continuously.

Overall, this leads to a number of publications which grew exponentially (Borgatti et al., 2009; Borgatti and Foster, 2003; Borgatti and Halgin, 2011, p. 2; Wasserman and Faust, 1994). A variety of different approaches exist in the research field of social network research. On a high aggregation level, we see two main analytical dimensions of social network analysis. We will refer to these dimensions as the mathematical/structural and the sociological-cum-psychological dimension. We give a short overview of both dimensions and explain the arguments for our chosen dimension in the context of our dataset and focus of research. In a second step, we discuss our focus within the field of social capital driven research. We follow the argumentation of Borgatti and Halgin (Borgatti and Halgin, 2011) who show a linking theory between the

---

[3] SNAP is a tool published at Stanford University enabling researchers to handle large scale network data with some hundred million knots, MIT and Harvard built a joint initiative to analyze large scale networks: https://www.hmdc.harvard.edu/services.html. UCINET as an established and proven tool can be downloaded at: https://sites.google.com/site/ucinetsoftware/home and Pajek can be downloaded via: http://mrvar.fdv.uni-lj.si/pajek/. Both tools are very established and can load, calculate and output data in various formats. In spite of other solutions, which might have advantages concerning multi-platform compatibility or calculation speed, these two solutions have reliable and assured algorithms. Besides freeware, there is a number professional applications, which offer the possibility to analyze huge amounts of data on the one hand but are very cost intensive on the other hand.

viewpoints of Granovetters' concept of weak ties, and Burts' approach to structural holes, to illustrate why we are of the opinion that both authors describe two sides of one coin and can build the backbone of our argumentation and research questions. The concepts of Granovetter and Coleman that can be contrasted with the approach of Burt will be discussed in detail in Chapter 3.

Structural and mathematical network analysis

The structural and mathematical trait of network research is quite developed. It is applied to a number of younger research fields, which still continue to grow. The rather technical and mathematical aspects of network research are covered by authors such as Wasserman & Faust, Dorogovtsev & Mendes, Newman, Schank & Wagner or Watts & Strogatz (Dorogovtsev and Mendes, 2002; Newman, 2003; Schank and Wagner, 2005; Wasserman and Faust, 1994; Watts and Strogatz, 1998). Their work provides essential contributions to increase today's potential to analyze the huge virtual networks of the World Wide Web and enable large data analysis to be undertaken. The measures and instruments of the graphs theory (Wasserman and Faust, 1994) are applied to research fields such as economics, sociology, game theory or information technology (Easley and Kleinberg, 2010, pp. 8–11) and allow the technical and mathematical analysis of research questions in these fields. As a consequence, this leads to a rather fragmented research environment, which does not homogeneously strive towards a common goal or shares a common focus. Still, a comparable set of instruments is used to describe social networks mathematically. As this aspect has a rather subordinated importance for this piece of research, we will not specify the technical methods of network research at this point. The analytical tools used for this study will be discussed in the specific context of our research questions.

Sociologically-driven network analysis

Today the research map of social network analysis is dominated by a dozen of North American authors[4] (Jansen, 2006, p. 48) who focus on the sociology driven aspects of

[4] The main protagonists are Stephen Berkowitz, Ronald Breiger, Ronalds Burt, J.A. Davis, Joseph Galaskiewicz, Mark Granovetter, David Knoke, Edward Laumann, Samuel Leinhardt, Peter Marsden, Berry Wellmann and Harrison White.

social network analysis. The different research areas are far from being clearly separated and have the tendency to overlap. This is why we will refer to a limited selection of topics and researchers with a focus on recent events (Easley and Kleinberg, 2010; Scott and Carrington, 2011).[5] Two main fields are quite vivid compared to others – game theory and economics. The research questions and research angles of these fields are again broad and fragmented. Robert Gilles for example combines social network theory with research in game theory which focused on directed communication networks and hierarchical organizations (Gilles, 2010). Easley and Kleinberg mention the combination of network theory and game theory to solve the *Braess's Paradox* (Braess, 1968), which shows that additional capacity in networks can also slow down traffic (Easley and Kleinberg, 2010, pp. 8–9). Lothar Krempel published on international trade using the methods of social network analysis (Krempel and Plümper, 2003), an approach which had already been tested by Krempel and Plümper since they published their research on the international trade activities for automobiles in 1999 (Krempel and Pluemper, 1999). Economical problems, especially trade relations, are very suitable to be analyzed with the methods of social network analysis as trade relations can usefully be expressed in terms of knots and ties. Sanjeev Goyal shows that social models have an impact on economics and individuals and can significantly influence economic institutions (Goyal, 2009). Kick, McKinney et al. present a network analysis approach to the *"World System of Nations"* (Kick et al., 2011).

Across different research fields, social network research follows a variety of goals depending on the outcomes that are the focus of the investigation (Borgatti and Halgin, 2011). Borgatti and Halgin propose to distinguish between two types of outcomes, choice and success. Choice-related research, focusing on behaviors, attitudes, beliefs and internal structural characteristics, tries to explain why actors have similarities in their choice making, due to the network environment they act in. The second research tradition deals with success-related outcomes that describe the performance of individuals, groups or entire networks. This research deals with social capital and its origins and occurrence in social networks. Both outcome types can further be distin-

---

[5] Biological, physical, neural and other natural science related network research account for a large amount of network research but will not explicitly be part of this part of research.

guished according to their character of being the result of an optimized flow of information or of optimized coordination, originating in the structure of the network. As our research focuses on the positive and negative effects of network structures on the performance of knowledge network participants, we locate our focus in the area of *Capitalization*, following Borgatti and Halgin (Borgatti and Halgin, 2011).

1.2.2 Common ground of "weak ties" and "structural holes"

Mark Granovetter aimed to relate the micro and the macro level of social network analysis. From his point of view, the majority of research describes the dynamics of large networks and single actors while less effort has been invested in showing how group behavior influences whole networks. Granovetter introduces the "strength of a tie" as a term to describe the intensity, frequency or quality of a relation, or a combination of these factors (Granovetter, 1973). Based on this work, Granovetter has developed a model of so-called "weak ties". This term refers to lower frequented connections of minor quality or intensity, although acknowledging that weak ties offer significant advantages for actors if they connect to areas within a social network that are not otherwise connected (Granovetter, 1983) As a result, an actor could access information sources that other actors do not have access to and thereby be able to influence future outcomes from which he could derive a competitive advantage.

Ronald Burt bases his work on Granovetter but focuses less on the intensity of the connections between actors and more on the absence of ties in general. To him, the term "structural hole" defines the gap between actors, or group of actors within a social network, whose direct friends have no common friendship (Burt, 1992, p. 18). Burt sees the ownership of these brokerage positions as a chance of "*participation in, and control of information diffusion*" (Burt, 2000, p. 8) and as a source of social capital. For Burt these holes create a competitive advantage, similar to the advantage Granovetter sees within a weak tie. Burt refers to a number of empirical studies and lab experiments to find evidence for his theory (Cook et al., 1983; Cook and Emerson, 1978; Granovetter, 2005, 1995, 1983; Lin, 2002; Markovsky et al., 1988) and conducted numbers of studies himself (Burt, 2005). These studies have shown that structural holes have positive effects on organizational learning, making it easier to find a (new) job or improve the chance of winning the Eurovision Songs Contest, for example (Yair and Maman, 1996). It is plausible that positive effects from occupying these brokerage positions can occur, because they offer chances to access additional information, which other actors do not have in this exclusive way. The research conducted in this area focuses on access to information or organizational learning and rarely considers outcomes like the performance of team members or organizations.

James Coleman has developed a similar concept, but sees the occurrence of social relations, the closure of ties, as a source of social capital (Coleman, 1988). In contrast to Burt, Coleman argues that dense, tightly knit networks are more favorable as they encourage shared norms and the building of trust between network members (Adler and Kwon, 2002, p. 24).

While a positive effect of building a network with independent but helpful stakeholders sounds quite intuitive, it is also possible that an internal competition for these bridges could harm the performance of organizations. Sales staff could, for example, compete for access to organizations that open a larger customer group and thereby lose focus on alternative sales deals. In our opinion the brokerage of structural holes is rather a question of holding power over information channels than about collaboration within a team or a sub-network. In this discussion, some authors focus on the differences between both approaches and argue that they are quite distinct (Kilduff and Brass, 2010; Moran, 2005). They see the contrast of either the one or the either concept as being advantageous. Kilduff and Brass see also both viewpoints and note that these may easily lead to neglection of common understandings (Kilduff and Brass, 2010). Other authors like Borgatti and Halgin focus on the theoretical similarities between both concepts and try to integrate the viewpoints of Granovetter and Burt based on shared concepts (Borgatti and Halgin, 2011). They argue that each concept stresses the importance of both structure and position. Granovetter and Coleman argue that dense networks are favorable and Burt argues that the holes make it favorable. It is therefore plausible that both concepts can be compared on the level of the ego-networks structure (Wasserman and Faust, 1994, pp. 41–43) and ego's position in the network and do not have to be compared on the level of ties being either weak or strong. Borgatti and Halgin see differences in the outcome perspective of both concepts as Granovetter focuses on job placement and Burt deals with promotions, but they are convinced that "[b]oth theories are based on the same underlying model of how networks work" (Borgatti and Halgin, 2011, p. 5).

1.3 Further demand for research and own focus

Having the intention to analyze an information network from the ego perspective we focus on activity and performance as dependent variable. We start with a short overview of this topic to identify possible research gaps. The single viewpoints of this overview will be further detailed in chapter 2 where our hypotheses are deduced.

An observable trend within social network science is the growing number of studies dealing with the influence of network structure on the performance of actors, groups and networks. Even though performance was one of the observed outcomes in early research, we still see a focus of some authors in this field. Machado and Ipiranga (Machado and Ipiranga, 2013) examine the impact of network structures on the performance of biotech innovation networks and find a strong need for integration strategies to increase the number of patents. Galunic, Ertrug and Gargiulo (Galunic et al., 2012) examine the positive spillover effects of higher ranked employees on the average rating an employee receives. In another study, Gargiulo (Gargiulo et al., 2009) examines the effect of network closure on the individual performance of knowledge workers. A very interesting aspect of this work is the differentiation between different levels of hierarchy and the distinction made between the role of an information acquirer and an information provider. Other authors focus on the impact of centrality[6] on the performance of network members and groups (Ahuja et al., 2003; Cross and Cummings, 2004; Sparrowe et al., 2001).

As mentioned before, our research tries to build a bridge between two – originally opposing – viewpoints on the mode of action of dense ego-networks. Burt sees structural holes as a source of social capital as they offer brokerage opportunities and give actors who hold brokerage positions better access and control over information (Burt, 2005), which results in rather sparse networks. Coleman on the other hand sees dense networks with high clustering coefficients as a source of social capital as they support trust and social bindings (Coleman, 1988). Recent research has either endeavored to

[6] Network analysis considers different measures of centrality. The most common one in the context of ego-centered networks is the Betweenness Centrality (C_B) that measures the number of shortest relations from each actor to each other actor that need to pass an actor as intermediate. The actors of a network with the highest C_B have a very prominent role in the network as they can for example control information flow (Wasserman and Faust, 1994, pp. 188–191).

bring both perspectives together in a combined approach (Ronald S Burt, 2001), or tried to find evidence for one of the hypotheses or developed a differentiated view.

Our research focuses on a better understanding of the mechanisms of structural holes and network density for ego-centered network in different situations. We follow the arguments of Gargiulo (Gargiulo et al., 2009) and try to find evidence for a more differentiated view of density and structural holes. We are also convinced that not only the motivation of the actors within the network has an influence on the development of social capital but also the network context itself. In our view, different network contexts rely on different network structures to allow social capital to grow. This is an aspect which has not been the focus of many of the quoted studies. From our point of view a competitive environment, like an investment bank for example, provides completely different mechanisms for building personal social capital than a collaborative environment such as a neighborhood or a school class.

We conclude that the analysis of a large knowledge network over a period of time can add additional knowledge to the process of building social capital. In addition, it can allow us to gain deeper insight into the circumstances that are either favorable or unfavorable for the buildup of social capital, especially concerning dense and broad ego-centered networks for network members at different stages of seniority.

2 Performance in online knowledge networks

2.1 Differentiation criteria of social networks

As already mentioned, the term social network is in widespread use nowadays. Having outlined recent work on the topic in the introduction, we now give a short summary of the historical development of social network science, before describing our approach to the classification of social networks, and the criteria we have adopted.

A social network can generally be defined as any group of individuals that is interconnected by social relations. The sociologist Georg Simmel (1858-1918) was one of the first authors to consider groups from a sociological perspective (Simmel, 2009, pp. 47–133). To Simmel, the study of the interaction between individuals constitutes a main object of sociology (Jansen, 2006; Simmel, 2009) and hence he placed relational attributes at the center of his research. For example, Simmel researched dyads and triads, as the basic building blocks of social networks (Simmel, 2009). His work is widely recognized as the cornerstone for quantitative research on individuals and relations of individuals in sociology. In the 1940s and 1950s, social network research was developed in the US in two main streams, the first dominated by a quantitative, mathematical and graph-oriented approach, and the other by a qualitative approach which focuses on social and psychological aspects of social networks (Borgatti et al., 2009). Quantitative studies have included investigations of communities (Hollingshead, 1949, 1948), of the division of labor, and have addressed theoretical issues arising in this context. Several publications on the subject were published between 1960 and 1980, in particular the “small world problem” coined by Milgram (Milgram, 1967; Pool et al., 1989), which revealed the surprisingly short average path lengths between members of large groups and even societies. Social networks can be as diverse as neighborhoods, school classes, companies, terrorists’ cells, gaming communities or knowledge networks – even secret organizations like the freemasons.

This research field has been growing at great speed and intensity (Borgatti et al., 2009), and has contributed not only numerous criteria with which to describe and analyze social networks, but also identified the different purposes of networks. Different ways of distinguishing between social networks have been proposed, along with different sets of criteria. These include: the type of members (Jansen, 2006, pp. 51–67), the characteristics of their relations (Jansen, 2006, pp. 69–90), a network's size or scope – from ego-networks to groups to whole societies – and whether they operate in a virtual or physical environment, as well as their context, and whether they are primarily cooperative or competitive in nature (Kilduff and Brass, 2010). Emerging from this, there are a number of different network types and network environments. To answer our research questions, we selected two criteria with which to distinguish social networks. First, we differentiate between virtual and physical networks and further distinguish them by their inherent purpose whether they create a rather competitive or a rather cooperative environment. We conclude this section with the classification of the knowledge network that provides the focus of our research, according to these criteria.

2.1.1 Physical versus virtual networks

A common distinction between types of networks is whether they are virtual or physical (Blanchard and Horan, 1998). The common language for these terms can sometimes be misleading, as it is often implied that a network operating on a virtual (online) platform automatically creates a virtual network. From our point of view, however, this assumption is not reliable: Setting up a network via a virtual infrastructure, with its members coming together or communicating online, does not in itself determine or change the context or purpose of a given network. If it were the case, it would mean that a group of scientists changes the nature of its network from physical to virtual as soon as they start using a virtual meeting room, infrastructure and knowledge management system. We do not negate that going online can affect such a workgroup, but we believe that this step alone does not significantly change the network's original task, structure, incentive systems, and the way social capital is built in the network.

For this reason, our definition of a virtual social network focuses on a network's mode of action and task structure. We define a social network as being only virtual if the actors within the network are separated from their true identity and act in an environment that is designed for a virtual identity rather than for a real person. Examples of virtual networks include MMPORGs (Massively multiplayer online role-playing games) which create their own virtual world, with its own rules and physical laws, and in which real-world players take on a completely new identity and role that has little in common with the real person behind that player. Most of these networks are based on computer-generated environments. The actions an actor performs in this context are by and large only related to this computer-generated environment and have no impact in the real world. The online nature of their activity is valuable for research, however, as the behavior of actors becomes more visible and suitable for research than the behavior of real network actors. While the relevant network data of virtual networks will be recorded by the platform, data on real networks has to be gathered from participants (e.g. through questionnaires or interviews). On a virtual platform, virtual identities are enacted on the platform itself. Nevertheless, virtual networks tend to share many common characteristics of real life social networks. This turns virtual networks such as *World of Warcraft*, for example (among many others) into an interesting environment in which to analyze human behavior (Castronova, 2006; Castronova et al., 2009; Ellison et al., 2011; Thurau and Bauckhage, 2010). On the other hand, such virtual platforms often entail two major drawbacks for research. The first is the potential use of operant conditioning on players as part of the “career-system” of these games (Burgun, 2012; Elliott et al., 2012), which risks introducing a bias into the data and thereby making it less suitable for social capital research. The second problem concerns the evolution of gaming rules over time. With every new release, the rules of these games need to be adapted to avoid an imbalance within the community, such as the introduction of a new player classes which requires adjustments to be made to the gaming rules. Such adjustments tend to correlate to the size and complexity of the game. It is also hard to observe the effect of the network structure on the actions of its members, as platform owners constantly try to manipulate the players into keep them playing, even though they may no longer be interested. At the same time, it is exactly these frequent adjustments to the rules and environmental conditions of the network that make virtual networks interesting for research by providing an ongoing experimental context for studying the effects of these adjustments on social networks (Castillo, 2020, 2019; Snodgrass, 2016).

To sum up, virtual networks tend to be set in a non-real-world context and consist of members that are decoupled from their real-life identity. These networks often undergo frequent adjustments to their environment and operating conditions, which has negative implications for the potential of analyzing behavior in this context, but might on the other hand offer opportunities to run and analyze sequential experiments.

Real life networks, in contrast, are those that are omnipresent in daily life and their existence often remains unnoticed by their members. Families, schools, neighborhoods or companies are the most common social networks almost everybody is part of at one point or other in their life. These networks also offer a vivid ground for research, since the beginning of social network research in the late 1940s (Hollingshead, 1949, 1948; Moreno, 1946) to today (Backstrom and Kleinberg, 2014; Hanifan, 1916; Rice and Yoshioka-Maxwell, 2015). There is a great variety of real-life social networks, from criminal investigation networks (Carrington, 2011; McGloin and Kirk, 2010; McIllwain, 1999) to terrorist networks (van der Hulst, 2014) or citation networks (Dorogovtsev and Mendes, 2002, pp. 1085–1086; Easley and Kleinberg, 2010, pp. 336–337), or networks of scientists (White, 2011). Even though the boundaries are not clearly defined, we can assume that real networks deal with actors who, by and large, are tied to their real identity and act primarily in the environment of the real world. It becomes more difficult to draw the line if the network structure is set up in an IT framework, especially when there is the possibility to adopt nicknames or aliases, but all the same we would consider these networks to be real networks. A virtual identity, which is defined by a nickname instead of a person's real name, will act more like the real person it represents than a virtual character in a MMPORG, for example. Casting a spell on somebody is quite different to talking badly about another person while hiding behind a nickname. From an analytical point of view, and judging by the exponential growth in publications, real world social networks constitute a very vivid field of research that deals with a great number of research questions. It deals either with structural or mathematical questions that can be verified by real life data, or addresses a number of practical questions such as, for example, group dynamics that lead to suicide (Bearman and Moody, 2004), trending and prognosis of outcomes in technical communities (Vernet et al., 2013), the competitive advantage of companies (Chai et al., 2011) or the improvement of knowledge work.

To sum up, in contrast with virtual network research, the knowledge network at the center of our analysis is a real-world network. It consists of real people interacting on the basis of their very own knowledge capabilities. They capitalize on these to earn status points on the network, which constitute the social capital of the network.

2.1.2 Network purpose and task motivation

Our contention is that both network purpose and task motivation have an impact on the creation of social capital within a social network. We believe that both aspects not only influence how social capital is built, but also how some network constellations are more favorable for its creation than others. In the context of our research questions the purpose of a social network is highly relevant. Even decades after Burt's work on structural holes, the upsides of sparse networks are discussed in terms of their context being either mainly cooperative or competitive (Kilduff and Brass, 2010). In the following we give a short overview of the different purposes of networks and the different task motivations for the network members that are associated with this purpose. We found that many of the quoted studies which focus on the benefits of structural holes have targeted a network purpose linked to either professional work, economics or academia (Burt, 2004, 1987; Bush, 1995; Galunic et al., 2012; Gargiulo and Benassi, 2000; Kleinberg, 2007; Kossinets et al., 2008; Vernet et al., 2013). In contrast, studies which aimed to find evidence for the benefits of dense, closed networks are often set in surroundings such as families, communities or neighborhoods (Bearman and Moody, 2004; Ronald S. Burt, 2001; Coleman, 1990, 1988; Mangino, 2009; Morgan and Sørensen, 1999; Vernet et al., 2013). Both clusters have in common that they see the structural prerequisites of the social network as a source of social capital and thereby share a common point of view on a meta-level. They differ, however, in the nature of the structure and in the network purpose and motivation of their members. Competitive networks are based on their members' motivation to optimize an individual goal, whilst cooperative networks (families, schools, communities) focus on a common goal. It is evident that members who are in competition with each other in the same social network will try to safeguard their personal information advantages, whilst those in cooperative networks will favor the dissemination of information. A competitive network will benefit from being sparse, whereas a cooperative network works better when it is dense. The downsides of both concepts are also obvious, with sparse networks being vulnerable to false information, as it has high impact or might lack new information input. Dense networks risk to contain redundant information or lack of transparency. Even if we were to try to distinguish between both network purposes by being either competitive or cooperative, we cannot exclude the possibility that a single network member or subgroup acts in their own interest and not in that of the group. The network purpose, nevertheless, leads directly to the task motivation of a single network member, and in the case of our knowledge network, whether they are motivated to answer questions in the community. As the output for activity and knowledge contribution in the network is not monetary and the members do not receive any benefits outside of the network itself, we assume that the motivation is mainly intrinsic. Members will receive recognition not in form of a financial outcome or other assets but only by progressing through the career levels of the platform and an increase in their reputation in the social network. This reputation is expressed by a platform-specific currency, status points. As each member collects status points in reward for certain activities, resulting in a higher status levels within the network, we see status as the main driver for their participation and development of a career in the network, apart from the more altruistic goal of enhancing their own and the collective knowledge of the community.

2.1.3 Knowledge networks as social networks

One currently underrepresented network type in research is the knowledge network, despite the fact that they are also widely found in organizational contexts. Indeed, we found very few publications dealing with research questions that undertake social network analysis in the context of knowledge networks. Knowledge networks can be either public or private. They tend to be virtual platforms that directly assign actions to their members and allow to raise the relations between them. As a consequence, most of the data can be found on the network itself and is often already stored digitally. This allows to differentiate between knots and ties, along with their attributes, and to run analyses at different levels such at the level of the individual, of subgroups, or of the whole network. In addition to this, the information of these networks is often stored together with a time variable for each action or information that takes place on the network. The network information is held in a relational database, which makes it possible to analyze information network data over longer periods of time. Lada Adamic, for example, conducted research on political blogs before the 2004 US presidential election (Adamic and Glance, 2005). He also researched the tracking of digital traces through blogs and avatars using the tools and methods provided by social network analysis to track opinion formation processes within communities in networks.[7] This research also includes the analysis of information diffusion in networks like *Second Life*[8].

Within knowledge platforms, members can pose questions on a variety of topics to accumulate or distribute knowledge among themselves or with others. In most cases the platform is searchable and holds the knowledge of its members to make it accessible to anybody searching the internet, as they are public platforms. In contrast, private knowledge platforms, such as company knowledge databases or research databases, tend to have incentive systems that are built into the employment contract and job objective of the individual employee or scientist. Public platforms on the other hand tend to have incentive systems that are only based on the platform itself and on its own rules. Usually, a point system is used to encourage network members to contribute knowledge and establish a gamification culture within the social network.[9] Users are rewarded for the quality and quantity of their contributions in the form of points, credits or a certain amount of a platform-specific currency, which reflects their ability to broker knowledge within the network. The reward for participation in a public knowledge network uses two incentives. The first are responses to questions posed by members, the second is the status, usually measured in points and earned by answering questions posed by other members. Therefore, the motivation to contribute to a public knowledge network draws on the motivation of their members, rather than being imposed by the platform rules. Whenever a member poses a question that is relevant to many members, the question will trigger more answers and greater network activity than a question that is of lower interest to the community. In contrast to social networks found in companies, the task environment of knowledge networks tends to stay constant over time. In a company, a C-level position requires a different skillset than an entry position. Within a knowledge network this skillset tends to stay more constant as it has a constant need for knowledge acquisition and brokerage. In addition,

---

[7] For a detailed overview on the research project see: http://www.si.umich.edu/node/1072.
[8] http://secondlife.com.
[9] The gamification of a platform is not to be confounded with the changing rules of online MMPORGs, which keep players in the game. Rather it is a way of breaking down long-term goals into small, actionable items that can be achieved over a shorter period of time.

knowledge networks assign each distinct knowledge item (i.e., a question or an answer) directly to the actors involved, compared to splitting the responsibility and task assignment between employees that is typical of hierarchical organizations.

2.2 Knowledge transfer in social networks

After having categorized social networks as being either physical or virtual and having pointed out how networks differ in terms of network tasks and motivation, we introduced the knowledge network as a special type of social network. In the following chapter we focus on the mechanisms that occur when knowledge transfer takes place in social networks in general and on online knowledge platforms in particular.

2.2.1 Acquisition and provision as types of knowledge transfer

Information seeking, learning and knowledge management are core concepts associated with knowledge transfer within social contexts (Borgatti and Cross, 2003; Borgatti and Foster, 2003). Knowledge transfer, defined as "one [actor] being affected by the experience of another" (Argote and Ingram, 2000, p. 15) is a key element for the success of many social networks. Benefits of knowledge transfer can also be found at organizational, team or single-actor level, or even between organizations or societies (Brass et al., 2004; Wei et al., 2011). As different network constellations can influence the development potential and performance of their members, the impact of specific network structures on the process of knowledge transfer is one of the most vivid fields of social network research (Brass et al., 2004; Fritsch and Kauffeld-Monz, 2010; Norman and Huerta, 2006). In this context, specific network constellations are seen as beneficial to the transfer and exchange of knowledge. Because of its more complex and tacit nature, knowledge transfer relies on more specific and task-related structures than the mere exchange of information which can take place under less restrictive conditions (Fritsch and Kauffeld-Monz, 2010). The position of an actor within a knowledge network is of significant importance, as the exchange of knowledge requires access to actors holding this knowledge. This means that an actor can hold more or less favorable network positions depending on his specific goal within the network. Thus the two opposing points of view – network closure and network brokerage – are broadly discussed in the context of social network research, whereas their influence on knowledge transfer is rarely investigated (Tortoriello et al., 2012). The results of these studies show no clear dominance of a single network measure to explain this influence, but suggest that different network aspects are drivers of knowledge transfer. Tortoriello, Reagans and McEvily see tie strength, network cohesion and network range as attributes that are crucial to the success of gatekeepers (Tortoriello et al., 2012), while other authors see ego-network density and tie strength as factors influencing knowledge transfer (Fritsch and Kauffeld-Monz, 2010). In this context the bridging of structural holes (Tortoriello et al., 2012) is seen as an important source of knowledge transfer and its relevance can even increase over time (Conklin et al., 2013).

As some of these results seem contradictory, and the bridging of structural holes and dense networks are not consistent in the same situation, we aim to identify reasons for these differences to occur. A differentiated view of two main types of knowledge network members – those primarily providing knowledge and those primarily acquiring knowledge - allows to distinguish between these contradictory views. Both types of knowledge transfer impose different network requirements for their actors (Gargiulo et al., 2009) and provide the contextual frame for knowledge transfer in the given situation (Conklin et al., 2013). Similarly, Gargiulo, Ertug and Galunic differentiate between the needs of knowledge workers by identifying the role of information acquirer, on the one hand, and of information provider, on the other (Gargiulo et al., 2009, pp. 304–307). We follow these roles, as they may potentially influence the creation of social capital. An information acquirer benefits from a dense network, as closure increases his chances to receive the desired information. An information provider or broker decreases the value of his knowledge if it is distributed too easily without him controlling this process. Gargiulo, Ertug and Galunic see this contextual frame of either being an acquirer or provider of knowledge as a central point of differentiation that even changes the way in which knowledge transfer works (Gargiulo et al., 2009). The authors examine the influence of tie density on bonus payments of knowledge workers and differentiate their hypothesis by knowledge transfer type and the hierarchical rank of the actor. The results show that dense networks, because of network closure, have positive effects for information acquirers but are negative for information providers. The findings also show that career level has a moderating effect on this correlation.

Our research focuses on one type of knowledge transfer, the provision of knowledge. Our aim is to look for evidence that specific network constellations are favorable for information providers that try to take advantage of *selling* their knowledge to their surrounding network. We are also aware of the fact that the role of a member of a knowledge network can change over time. Our dependent variable nevertheless focuses on the information provider role rather than on the acquirer role. In turn it can also mean that network constellations that are favorable for information providers are less favorable for information acquirers, and vice versa.

2.2.2 The specifics of public online communities with regards to knowledge transfer

Faraj et al. define knowledge collaboration as the "sharing, transfer, accumulation, transformation and cocreation of knowledge" (Faraj et al., 2011, p. 1224), and knowledge collaborations in online communities as an act of "adding to, recombining, modifying and integrating knowledge that others have contributed" (Faraj et al., 2011, p. 1224). The authors further distinguish between knowledge work occurring in online communities and knowledge work occurring in an organizational context by pointing out that online communities often lack the typical hierarchical structures, membership restrictions and ties that come with an organizational context. Additionally, Faraj et al. suggest that, an environment of lower organizational boundaries and the lack of presence of existing social relationships does not harm knowledge exchange in open online communities (Faraj et al., 2011). The combination of knowledge is even enhanced (Hughes and Lang, 2006), as long as the platform is equipped with appropriate technical prerequisites. Following in this line of thought we also distinguish between private and public knowledge network platforms, as already stated in chapter 2.1.3. While private platforms tend to restrict access and are mostly run by organizations as a way of hosting and managing internal knowledge, public platforms do not have entry barriers and allow the free development and exchange of individual knowledge. Even if "online community" is a definition based on the technical structure provided by the platform, the term is widely used to describe a network which is not set in a single organizational context but available to the public. Knowledge transfer research has so far primarily focused on knowledge networks in organizations, while little work has been done on public knowledge platforms (Faraj et al., 2011; Faraj and Johnson, 2011; Nahapiet and Ghoshal, 1998). The majority of the (few) studies that exist on online knowledge communities have examined the structures that encourage the active participation of network members, or the techniques deployed to sustain their long-term involvement (Faraj et al., 2011, p. 1225; Faraj and Johnson, 2011; Wasko and Faraj, 2005, p. 53).

Rather than examining the process of knowledge sharing in online communities, our research focuses on the network structures that allow members of online communities to broker their knowledge, and specifically, whether we can find a link between the type of network structure and the effectiveness of its members. Consequently, we must examine the similarities and differences between online communities and other types of social networks respectively, to distinguish between open knowledge communities and organizational knowledge platforms.

It is easier to collect data on online communities than on real life networks, which often tends to be questionnaire-based. From our point of view, this aspect is one of the strongest drivers behind the recent popularity of online community research. Online communities not only offer the advantage of having available data but also usually offer easy access to dependent variables. The performance of their members and their different levels of success are captured either in a point-related system, status, an internal currency or similar systems, or a combination of these. Online gaming communities spend a significant amount of their development work on developing and reviewing their point and career systems in order to secure the long-term stability of the platform and reduce churn rates. Online communities, by definition, do not require their members to be in one specific physical location nor do they necessarily require a simultaneous working mode – especially not knowledge platforms. In this way they allow their members to deliver input according to individual schedule preferences.

To summarize, online communities offer the chance to collect a more holistic and precise set of data to describe social network dependencies and behavior. They allow to overcome local and time-related boundaries that keep social networks from growing and sustaining. An obvious downside of online data gathering, and analytics can be the amount of noise and aggregated data that needs to be cleared, prepared and structured, to suit the study's specific analytical aims. The knowledge network dataset we examined meets the criteria of the two highlighted aspects: it is an online platform and has no organizational context, which means that it should allow us to monitor effects that are less biased by membership restrictions of professional organizations. Hence, the above-mentioned attributes of our dataset give us the opportunity to examine both the structural circumstances and the dependent variables that describe the ability of networks members to broker their knowledge.

2.3 Performance indicators in online communities

2.3.1 Types of performance indicators in online communities

To measure the success of network participants, existing research utilizes a broad variety of performance indicators, from measuring the number of citations in citation networks, up to complex point systems in MMPORGs like *World of Warcraft* (Ahuja et al., 2003; Castillo, 2019; Conklin et al., 2013; Faraj et al., 2011; Faraj and Johnson, 2011; Freeman, 2004; Fritsch and Kauffeld-Monz, 2010; Gargiulo et al., 2009; Mehra et al., 2001; Thurau and Bauckhage, 2010; Trier and Bobrik, 2007; Vargas et al., 2018; Zaheer and Bell, 2005). Most of these performance measurement systems tend to share similar characteristics as they reward different kinds of behaviors. One widely accepted performance indicator in this context is activity, as it is an obvious – and quantifiable – measure of network participation. Whether playing a game, publishing posts in knowledge networks or developing products leads, in most cases, to a positive output and thus can be used as an indicator for performance. Researchers can easily get hold of the number of publications in research, of quotations of a certain publication, or a mixture of both, weighted by their quality (Ahuja et al., 2003). Another dependent variable can be the quality of a network member's contribution, evaluated either by a predetermined system that is linked to specific tasks, like in online games, or as the result of a third-party evaluation rating a network contribution, or clicking a *like* button in Facebook. This type of evaluation tends to be rather subjective, compared to one that is performed automatically by gaming platforms, for example, which distribute the same number of points or gaming currency for each platform participant performing a specific task or reaching a goal.

In addition to these two most common indicators – quantity and quality – another group of indicators can be used to evaluate the performance of social network members: Career status or hierarchical levels, and membership duration. Career status provides both an indication of performance in an organizational context as well as acting as a moderator[10] of performance (Gargiulo et al., 2009). The length or duration of membership, especially in a high-performance organizational context, usually indicates that an individual is a valued member of staff, and is often associated with other positive effects such as promotions or pay rises. Yet, to our knowledge, membership duration as a predictor of performance in a social network has not yet been investigated in the existing research, although we think it could be an important factor.

There are advantages in using a combination of the four above-mentioned indicators as variables with which to predict and evaluate the performance of members in our knowledge network. In the following sections we present the explanatory model adopted for our research.

2.3.2 Measuring performance in online knowledge communities

Considering the afore-mentioned performance indicators activity level, quality of contributions, career status and membership duration – it would appear that it is possible to measure all of these quite easily for an online knowledge community. Indeed, a knowledge community's database stores data such as members' contributions (questions and answers), the quality of their contributions as rated by other members, their status level and length of membership on the platform.

As explained, there is great potential in including all these variables as they allow to develop a far more detailed assessment of the performance of network members than any single variable would on its own.

A differentiated view of the qualitative and quantitative aspects of performance, and the inclusion of moderators such as hierarchy and membership duration, may allow us to gain valuable insights into the mechanisms that drive the performance of members of knowledge networks. Despite this potential, we found only very few publications that cover more than one of these performance variables. Amongst them are Gargiulo et al. who measure moderator effects of employee rankings on the performance of knowledge workers (Cross and Cummings, 2004; Gargiulo et al., 2009). However, we only found research that has measured performance variables in online knowledge

---

[10] For a detailed view on the concept of moderators see (Baron and Kenny, 1986; Brambor et al., 2006; Kilduff, 1992)

communities by using data generated by questionnaires or which considered monetary outcomes such as bonus payments, market shares, or profit and loss positions (Cross and Cummings, 2004; Gargiulo et al., 2009; Kogut, 2000). One of our major challenges was to identify a performance indicator that directly indicates the ability of a member to broker knowledge in a virtual knowledge network. Another challenge was to find a network with a sufficiently large number of evaluators to judge the quality of the knowledge generated, thus mitigating against the subjectivity of a single or a small number of evaluators, which otherwise could bias this performance measure. The ideal measure should be generated by the knowledge platform system or by the users that benefit from the generated knowledge. Apart from the qualitative aspect, we also aim to measure the quantitative aspect in terms of members' activity levels. Both aspects matter to a knowledge community. While the qualitative aspect considers the quality of the content contributed to the network, and therefore the value of the knowledge shared in the community, the quantitative aspect is an indicator for the relevance and popularity of the social network, and therefore, its long-term viability as a knowledge community.

It is plausible to argue that 'young' members of knowledge networks initially must gain acceptance in the community before they can effectively broker their knowledge and gain a certain level of independence. We therefore consider membership duration and the status acquired by a member as moderators that influence their performance once they have started to broker knowledge effectively in the community.

3 Embeddedness as determinant of performance in social networks

As indicated previously, social network research is a very lively field of research and mainly driven by two streams, one mathematical and quantitative, and the other social science-oriented one. While our research takes a social science approach to network research, we will also draw on quantitative methods and tools where appropriate. The social science approach to network research has its roots in the beginning of the 20th century with authors like Georg Simmel, Wolfgang Köhler and Fritz Heider, but is currently dominated by a dozen Anglo-American authors. Meanwhile the research field has been growing continuously in depth and breadth and has a very fragmented and broadly structured shape. Reviewing the different research streams, Borgatti and Foster propose nine specific fields to organize social network research (Borgatti and Foster, 2003, pp. 993–994). These categories are social capital, embeddedness, network organizations, board interlocks, joint ventures, inter-firm alliances, knowledge-management, social cognition, and group processes. Aiming to investigate the effect of embeddedness on the building of social capital in knowledge networks, our main interest lies in connecting three of these categories. In chapter 2, we introduced our view of knowledge networks as social networks, as a conceptual starting point and therefore first category. We will now define social capital and embeddedness for the context of our research and will then describe the interaction and importance of the three afore-mentioned research fields, to formulate our hypothesis and to explain their mode of interaction for our study.

3.1 Embeddedness as a source of social capital

The term social capital is in widespread use across academic disciplines, so much so that it has been variously defined and hence lacks a clear definition (Easley and Kleinberg, 2010, pp. 61–62). Given its many definitions and the variety of research foci, established authors like Ronald Burt have come to dub social capital research the "*Wild West of academic work*" (Burt, 2005, p. 5). A lot of authors see that at the core of social capital lies *"the ability of actors to secure benefits by virtue of membership in social* *networks or other social structures"* (Portes, 1998, p. 6) and use this definition for their argumentation (Easley and Kleinberg, 2010, p. 61). Even though social capital has been a subject of much recent research, the first studies go back to the beginning of the 20th century, and specifically to the work of Lydia Hanifan and her studies of rural schools (Hanifan, 1916). One of the first economically driven analysis has been conducted by Glenn Loury who saw social capital as a resource which results in the positioning of an individual within a social network (Loury, 1976). James Coleman considers social capital as a category of human capital, in contrast with physical capital. Just as human capital is created when actors change their abilities or qualifications, social capital is altered when relations of actors are involved in a way that enables productivity (Coleman, 1990, p. 304). Ronald Burt defines social capital as the benefits of a single actor's location within the context of a social network situation (Burt, 2005, p. 4). Nowadays social capital is seen increasingly relevant to career development, besides the influence of human capital (Coleman, 1988; Iseke, 2009; Iseke et al., 2011). It can be seen as the advantage a person derives from being better positioned in a network, compared to other network members. Put differently, social capital definitions hone in on the benefit that actors accrue through their social interaction. Based on these notions we formulate our working definition of social capital as follows:

> *Social capital is the ability and success of an actor within a social network that exceeds his personal abilities and qualifications. It is thereby rather a quality of the network position of an actor than of the actor himself.*

Definition 2: Working definition of social capital

Even though *ability* and *success* are common terms, as they can be seen in a cause-and-effect relation, the *quality of a network position* is a rather vague definition of a complex phenomenon. The authors mentioned above share a common understanding of the benefits of social capital, such as success or performance of an individual, but vary in their point of view concerning how it is created and the factors that drive its development. The discussion surrounding these aspects seems to drive recent research endeavors more than its definition. Along with the quoted authors, we argue that social capital in knowledge networks arises due to certain influence variables that impact positively or negatively on its occurrence. In other words, certain variables can either enhance or inhibit the amount of social capital in a network. We posit that an actor's position within the social network is a significant factor which drives the development of social capital. Our research, then, implies that social capital occurs in varying intensity and is dependent on the characteristics of an actor's social network environment.

Besides the definition of social capital, researchers try to find answers to the question of the origin of social capital, and in so doing distinguish different sources of social capital. The authors mention that social capital is either derived from, the sum of, or provides access to actual and potential resources. Though it seems to be a common point of view that networks contribute to building social capital, the view on the underlying mechanism varies. Adler & Kwon name opportunity, motivation and ability as main sources of social capital and build a "conceptual model of social capital" which also shows the influence of social structures such as market relations, social relations and hierarchical relations on social capital (Adler and Kwon, 2002). Other authors, like Coleman, only see a limited set of sources for social capital – tightly knit and dense networks – and do not strive for a complete set of possible sources (Coleman, 1988). A more current overview by Dorothea Jansen suggests six different sources of social capital: solidarity of families and groups, trust in socially accepted norms, information, power in the form of structural autonomy, the ability of collectives to organize themselves, and power in the form of social influence (Jansen, 2006, pp. 28–32). Each of these sources focuses on the implications of structural aspects and ties within groups (Robins et al., 2009), as a foundation for trust and power (Burt, 2005; Coleman, 1990; Easley and Kleinberg, 2010). Jansen further argues that a balance between social embeddedness and autonomy is important for long-term success in corporations to leverage power through social influence as a source of social capital. Bridging structural holes can be positive as it allows to discover business opportunities, while the ability to cooperate can foster long-term success in corporations by leveraging power through social influence. When Adler & Kwon formulated their *Conceptual Model of Social Capital* (Adler and Kwon, 2002) they introduced a source of social capital called *Opportunity*. Opportunity reflects a combined view of these classical viewpoints that are discussed in the context of social capital creation – network closure and structural holes. At this point we can state that social capital is related to different sources of which embeddedness is one that is mentioned by a vast number of authors. Granovetter was one of the first to use the concept of embeddedness in the context of social capital and to explain its occurrence (Granovetter, 1985). He introduces the term as behavior being embedded in a social network of relations (Granovetter, 1985) and transferred it even to markets and economic activity. This contrasted with existing research which had focused on an "atomized" view, omitting the behavior of other groups and the history of own relations (Granovetter, 1985). We will now focus on two main concepts, brokerage and closure, to further differentiate the concept of embeddedness before setting out the operationalization of this concept in our work.

3.2 Embeddedness and its antagonists: brokerage and closure

Brokerage, network closure and embeddedness are concepts that share a common conceptual meaning, even though they refer to slightly different aspects of network theory. Mark Granovetter's theory of *weak ties* (Borgatti and Foster, 2003, p. 994)[11] was developed with the intention to analyze the influence of social relations on behavior in institutions (Granovetter, 1985, p. 481). He sees embeddedness as a way of explaining this relation. Easley & Kleinberg define the embeddedness of an edge as "the number of common neighbors shared by the two endpoints" (Easley and Kleinberg, 2010, p. 58). Sociologists like Coleman see Granovetter's view of embeddedness as *"a structure with history and continuity that give it an independent effect on the functioning of economic systems"* (Coleman, 1988, p. 97). The authors agree that embeddedness influences the behavior of actors within social networks. Granovetter explains the effect of embeddedness with the simple example whereby it is less damaging to cheat on a friend with no friends in common than to cheat on a friend with whom we share lots of common friends (Easley and Kleinberg, 2010, p. 59; Granovetter, 1992). From this point of view, embeddedness can be seen as an internal fraud protection. As we have seen in the cited studies, some of the authors use the term of embeddedness as a measure of edges and not of nodes. It can therefore only be applied as an attribute of a relation and not as an attribute of a person. For our research this is an important aspect, as we will refer to an independent variable that clearly refers to an actor and not to ties.

---

[11] Both points of views look contradictory at first sight but were on the one hand differentiated by Adler & Kwon and integrated by Burt while separating an internal and external perspective that can complement each other (Adler and Kwon, 2002; Ronald S. Burt, 2001).

Coleman's approach to network closure offers such a concept. In relation to embeddedness, closure can be seen as a degree of operationalization of embeddedness. The higher the degree of closure, the higher an actor is embedded in the surrounding network. For Coleman, dense networks that allow social norms to form are a main source of social capital (Coleman, 1988). The term "closure" means that a triad with two existing connections is closed. Network closure can therefore be seen as one possible way to quantify the embeddedness of an actor within a network. The advantages of dense networks that Coleman mentions are also plausible in that within a dense network each member of the network can be reached very easily. Even if the information flow of some members is disturbed, there will always be a number of alternative paths that can be used to share information.

When we compare Coleman's point of view to Burt's thesis of brokerage, we must note that both authors share one important aspect. They both see social structures as the source of social capital, but they vary in their viewpoint of the shape and quality of certain structural conditions (Ronald S Burt, 2001). The opposing character of these concepts can be represented by a network measure, the clustering coefficient. We chose our definition of the clustering coefficient in relation to Dorogovtsev and Mendes, Newman and Easly/Kleinberg (Dorogovtsev and Mendes, 2002; Easley and Kleinberg, 2010, pp. 44–45; Newman, 2003):

> *The clustering coefficient measures the number of closed triads in relation to the number of potentially closed triads, from the perspective of a single actor or a whole network structure.*

Definition 3: Definition of the clustering coefficient

Given this definition, the clustering coefficient can vary from 0 to 1. When the value equals zero, we can say that the actor is bridging at least one structural whole. When it equals one, each of the actor's friends are friends themselves and we are looking at a highly closed, dense network. In other words, the actor is deeply embedded in his social environment. Coleman argues that a position in a network is better when it is well connected. This means that due to this restriction, an optimal position in the network would be one where each friend of an actor has a connection with each other friend of the actor and a clustering coefficient of 1. Coleman sees various advantages in dense structures. Social capital, originating in closure, protects from negative externalities as it helps to establish trust and norms. It is noticeable that the circumstances, which Coleman reviews and analyzes, often build on a common focus or goal that is shared, and less on competitive situations. Coleman analyzes unions, communal housing projects, neighborhoods and families. These networks are characterized by a common goal that can be achieved more easily if the single actors share information and trust each other. It is also noticeable that these situations tend to have an internal focus as they reflect a group striving towards a common goal.

The opposite situation of a deeply embedded actor can be an actor who does not or only partially participate in triadic closure and thereby functions as a local bridge. From what we have outlined before, these actors have the attribute of being rather externally oriented as they bridge from internal and the external environment of a group or sub network. Referring to the possible sources of social capital, the source of this kind of social capital will then be its structural autonomy, which is the inherent attribute of a bridge between two parts of a network. In contrast to a dense, closed network, structural holes are defined by missing links within the network and not by the connections that are made.

Mark Granovetter's work can be seen as the foundation of Burt's theory of structural holes. From his point of view a great amount of research was done to describe dynamics of large networks and single actors but only little effort had gone into showing how group behavior links up and influences whole networks. Granovetter introduces the "strength of a tie" as a term that describes the intensity, frequency or quality of a relation, or a combination of these factors (Granovetter, 1973). Based on this, Granovetter developed a model of so-called "weak ties", which are rather lower frequented connections of minor quality and intensity but offer significant advantages for an actor if they connect to areas within a social network that are not otherwise connected (Granovetter, 1983). A good real-life example could be a salesperson who has advantages by not only having strong ties to his colleagues but a weak tie to somebody in product development. As a result, he would have access to information the others do not have and could maybe influence future products in a way that is favorable for him.

Ronald Burt's work builds on the work of Granovetter and focuses less on the intensity of the connection between actors and more on the absence of ties in general. The term "structural hole" is a definition of an actor within a social network whose direct friends have no common friendship[12]. Burt sees the ownership of these brokerage positions as a chance of "*[p]articipation in, and control of information diffusion*" (Burt, 2000, p. 8) and as a source of social capital. For Burt these holes create a competitive advantage, similar to the advantage Granovetter sees within a weak tie. Burt refers to a number of empirical studies and lab experiments to find evidence for his theory (Cook et al., 1983; Cook and Emerson, 1978; Granovetter, 1983; Lin, 2002; Markovsky et al., 1988) and conducted several studies himself. These studies show that structural holes have positive effects on organizational learning, make it easier to find a (new) job or could even improve the chance of winning the Eurovision Songs Contest (Yair and Maman, 1996). It is plausible that having these positions as an actor can have positive effects too, because they offer additional chances of access information, which other actors do not have. In our view, the research he conducted has a greater focus on access to information or organizational learning and is not as often related to topics like the performance of team members or organizations.

In contrast to these situations, Burt focuses more on situations that have a competitive character and favors actors that can generate a competitive advantage compared to others, like corporate managers. If we pick up the starting definitions of social capital and consider Coleman's approach as rather internally oriented, we could say that Burt's viewpoint reflects rather externally oriented situations. Consequently, neither Coleman's nor Burt's approach can be proven right or wrong without considering the network set-up and whether their focus is an internal or external orientation of the mechanism that builds social capital. Adler and Kwon recognize this and identify the need to structure and categorize the social capital-related research according to these criteria and align the different scholars and viewpoints on social capital to their applying an internal or external orientation (Adler and Kwon, 2002). The difference between both perspectives is their definition of network boundaries. A tie of an organization's employee to a customer represents an external orientation, while a tie to a colleague within the same department would rather represent an internal orientation. Therefore, closure is seen as a source of social capital in internally oriented networks and bridges

[12] Later we will see that a structural whole can thereby have a clustering coefficient of 0.

while structural holes are a source in externally orientated networks (see Appendix 1 for an overview of different definitions of social capital)

These opposing viewpoints of a network's internal and external orientation can be explained by comparing the definitions of social capital given by the quoted authors. In addition to the external or internal orientation, we notice that studies which see closure as a source of social capital tend to focus on cooperative environments while studies focusing on competitive environments tend to see the bridging of structural holes as a source of social capital. In sum we can say that authors who try to find evidence for situations of closed networks, with a high degree of embeddedness, tend to examine environments with an internal, non-competitive focus. On the other hand, authors focusing on an external orientation tend to examine competitive situations that favor the appearance of structural bridges and less dense networks. The quoted studies imply that network environments can be differentiated by being rather externally or internally orientated. We are of the opinion that such a strict division of the two situations is not always possible and does not reflect real life situations. To deeper understand the mode of closure we will examine the possible positive and negative aspects of closure and mirror them with the characteristics of local bridges to see the full spectrum of highly and lowly embedded situations.

To conclude with our own operationalization of embeddedness we can state that:

> *Embeddedness, measured by the clustering coefficient, is the degree of closed network connections between the friends of an actor, in relation to the maximum possible number of these connections. A clustering coefficient of 0 would mean that the actor is not embedded at all, a degree of 1 would mean that the actor is not bridging any local bridge and is – consequently – highly embedded.*

Definition 4: Definition of embeddedness

3.3 Integrative concepts for embeddedness

Practical examples of network closure are given by social network platforms like Facebook, LinkedIn or XING. When these platforms suggest a new friend, we often look at a request for triadic closure. In other words, the suggested third tie closes a triple with two existing relations. The platform's system suggests a new friend or contact based on the simple rule that the likelihood of two friends of an actor becoming friends themselves is higher than becoming anybody else's friend (Rapoport, 1954). From Granovetter's perspective, triadic closure eliminates a forbidden triad in the social network (Granovetter, 1973, p. 1363). A broad variety of research has been conducted to examine the influence of network closure on individuals, groups and organizations. Bearman found evidence that female high school students with dense personal networks had significantly lower suicide rates than those with sparse networks (Bearman and Moody, 2004). Easley & Kleinberg argue that the closure of a triad will lead to more stable relations and better information flow in between all triad members. In addition to that, it is likely to lower the level of stress in the group as it works as a source of trust in a relationship (Easley and Kleinberg, 2010, p. 45). Consequently, closure leads to higher social capital.

In the field of knowledge work and social networks dealing with knowledge management, we can see a clear focus on the positive aspects of closure towards the performance of knowledge workers (Ahuja et al., 2003; Brass et al., 2004; Cross and Cummings, 2004; Ichniowski and Shaw, 2009; Machado and Ipiranga, 2013; Mehra et al., 2001; Vernet et al., 2013; Wei et al., 2011) and innovation (Schilling and Phelps, 2007; Wasko and Faraj, 2005). In the context of these studies, closure has a positive effect on the outcome of groups, enabling information flow (it is more probable), and social control (the information about unsocial behavior is harder to hide). Coleman argues that closed networks induce social trust, cohesiveness and cooperative behavior within the network (Coleman, 1988).

While some authors focus on the advantages of closure, another set of authors sees significant downsides of the concept. They see the effect of a closed network as a barrier that blocks the sub-network from external information input. As a possible consequence, closure thereby may reduce diversity (Burt, 1992). Burt sees dense networks rather as a barrier to the coordination of networks. He argues that the internal information exchange forms out of a culture that prevents teams from positive external input and innovation (Burt, 1992). In addition, actors in dense networks have a significantly lower degree of independence than actors who are building local bridges. Burt is convinced that this independence is an important force, which allows organizational coordination (Ronald S Burt, 2001). His viewpoint of brokerage being an advantageous position is based on the hypothesis that the bridging of structural holes allows structural autonomy and access to exclusive information. As a consequence, dense networks can have a lower ability to adapt to changing environments as they have a strong internal determination (Gargiulo and Benassi, 2000). Furthermore, it is possible that a negative image of a group marks off to a single actor, making it difficult for him to form new external ties (Raub and Weesie, 1990). Mangino finds evidence that African boys are less delinquent when they are not a member of only one dense group (Mangino, 2009) but hold brokerage positions between several dense networks.

As mentioned, the discussion of structural holes and network closure as oppositional viewpoints may lead to the conclusion that both angles are contradictory. Thus, we want to emphasize that they both deal with the same source of social capital. Each of them sees the ego-centered network as a source of social capital that adds value to the actor (Gargiulo and Benassi, 2000) but their effects may vary in different frameworks, with different motivational backgrounds and situations. The difference exists in the mode of action that is assumed. While the rather externally oriented view of bridging structural holes supports the hypothesis of benefitting from structural autonomy (Burt, 2005, pp. 139–141), the authors supporting the hypothesis of closure see the benefits of a free flow of information and protection from externalities (Coleman, 1988). Coming back to the starting point of the positive and negative effects of network closure we can summarize that researchers tend to a bipolar view of the subject. The disadvantages of closure are the advantages of structural holes, and it is difficult to examine one construct without at least bearing the other in mind. We are convinced that neither perspective can be judged as right or wrong. The studies and publications are partially dealing with very different contexts and purposes of networks and with an individual focus on either an external or internal orientation, as mentioned in the previous chapter.

This circumstance may have driven some researchers to try to integrate both viewpoints in one concept. Portes and Sensenbrenner distinguish clearly between the up- and downsides of embeddedness (Portes and Sensenbrenner, 1993) and point out its suitability as an umbrella concept. The authors name explicitly the negative effects of freeriding, the limitation of individual freedom and leveling pressure – keeping group members away from chances of advancement. On the other hand the authors name a number of positive effects of closure leading to solidarity and trust (Portes and Sensenbrenner, 1993).

The study by Gargiulo, Ertug and Galunic contributes to the relevance of different environments of social networks and shows, in which contexts actors benefit from structural holes and in which contexts they benefit from network closure (Gargiulo et al., 2009). We think this study is ahead of other approaches by uniting the angles of closure and structural holes compared to Adler and Kwon, and Burt, who focus rather on a theoretical concept than on an empirical approach. Adler and Kwon (Adler and Kwon, 2002) seek the integration of both concepts into one model and argue that:

> *"Closure provides social capital's cohesiveness benefits within an organization or community; structural holes in the focal actor's external linkages provide cost-effective resources for competitive action. But even when we focus on external ties for competitive goals, both closure and sparse networks can yield benefits. Which is more valuable depends on the state of the other sources of social capital and on the task and symbolic environment confronting the actor."* (Adler and Kwon, 2002, p. 25)

They argue that "*opportunity, motivation and ability"* (Adler and Kwon, 2002, pp. 23–27) are influenced by "*market relations, hierarchical relations and social relations"* (Adler and Kwon, 2002, pp. 18–19) and are the main influence factors on the mode of social capital's action (*benefits and risks* (Adler and Kwon, 2002, pp. 28–32)). This model puts basically every factor together and integrates closure and brokerage into a situative model of coexistence within a network. Neither the one or the others needs to be the only source of social capital and both approaches do not need to exclude another. When we reduce this model to a basic idea, we can say that network context (school class, neighborhood, investment bank or knowledge platform), competence (hierarchical position, seniority, access to information) and personal motivation (need of information, ambitions to develop) are important influence factors. From our point of view these factors determine whether ego-centered networks with structural holes or with high closure are favorable sources of social capital. In contrast, Burt tries to integrate the two concepts and argues that both orientations, internal and external, need to be respected. He argues that "*Brokerage across structural holes is the source of value added, but closure can be critical to realizing the value buried in structural* *holes"* (Ronald S Burt, 2001, p. 31). Thus, he respects both sources of social capital but integrates them into a construct of a necessary and a sufficient condition to realize the full potential of social capital. Even though Burt shows a wide range of empirical evidence for his hypothesis we favor the arguments made by Adler and Kwon who name more specific influence factors for the creation of social capital.

Coming back to the study of Gargiulo, Ertug and Galunic we find a very differentiated viewpoint. The authors distinguish between the roles of an information acquirer and an information provider and take the rank of the information worker into account. Their sample consists of 2000 employees of an "*equities division of a major financial services firm"* (Gargiulo et al., 2009, p. 309) spread over 41 international business units. The dependent variable is the annual bonus payment. They summarize their findings as follows:

- *"Network closure, however, not only enhances trust, it also enhances mutual control. […] Though trust is likely to benefit both parties to an exchange, mutual control may not."* (Gargiulo et al., 2009, p. 326)
- *"Our results suggest that, in asymmetric exchanges, closure may compel providers to put more time and energy into the exchange than they might have preferred."* (Gargiulo et al., 2009, p. 326)
- *"Our analysis of the knowledge exchange network and individual performance in an investment bank shows that network closure in bankers' acquirer role increases their bonus, but closure in their provider's role decreases it."* (Gargiulo et al., 2009, p. 326)
- *"Our results have theoretical and methodological consequences for research on the relationship between network structures and outcomes. Theoretically, they highlight the importance of paying attention to the control effects of network structures on individual outcomes, as well as to the different roles actors can play in relationships."* (Gargiulo et al., 2009, p. 326)

We see the approach of Gargiulo, Ertug and Galunic as the more progressive and differentiated among the publications examined, covering the conflicting discussion of either structural holes or closure as a source of social capital. We also follow the idea that the authors do not integrate both phenomena into one approach but identify situations and variables that make only one of the concepts favorable for certain actors. In addition, the authors implicitly say that both approaches can be of advantage to different actors in different situations within the same network. In our research, we try to narrow these different situations down to a more concrete model that follows not only the career level of an actor but also his membership duration within the network he acts in. We believe that for each network participant the influence of the clustering coefficient on performance is changing over the duration of their membership.

3.4 The impact of network closure on performance

In the preceding chapters, we pointed out that most of the different authors adopt either an internal or external focus, or a more cooperative or competitive task environment in their definition of social capital. We also stated that the quoted authors had a different viewpoint on the network environment characteristics – dense networks or bridging of structural holes – that favor the development of social capital.

When we look at the characteristics of our dataset[13] we can find arguments for both viewpoints. The ambition of newly joined network members will probably be to quickly establish a number of social ties and to build up a personal network and reputation. As a result, good opportunities to identify questions and participate in other network activities will occur. This is a plausible ambition of newly joined network members, disregarding the specific context of knowledge networks. Many new joiners in companies, clubs or organizations follow this objective and prioritize their activities accordingly. In contrast, there will also be a group of members in the network that focus on selling their abilities, their knowledge or goods, built up over time, and which are a valuable asset now. This group could potentially suffer from communication being too easy and the network being too dense as this limits their own opportunities to broker and facilitate information diffusion and drain.

Generally, we can state that every member will try to broker his knowledge to the community in the expectation of gaining status and acceptance within the network. In addition, a continuous flow of new members joins the network every day. They try to find their way around and strive to build up reputation while collecting status points. When we tempt to analyze the impact of network closure on the performance of social

[13] Chapter 4 follows with the detailed description of the dataset.

network members, we must be aware that we potentially look at both of these groups at the same time. This is mainly driven by the fact that we look at participants with different levels of experience at different points in time. Therefore, the dataset is an overlapping composition of members with different objectives, who are active simultaneously.

To better understand the motivation and working mode of these groups, we must look at the social capital that is produced in our network – status points. Network members receive status points as a reward for their personal activity and effort. These status points are distributed for participation in form of answering questions and due to the number of people competing in that endeavor. Consequently, the network actors compete for pieces of the same pie. To better understand this mechanism, we will examine in more detail the underlying characteristics of network ties or friendships for both groups. When we argue that our dataset rather reflects a competitive environment, we must ask ourselves why members of this network engage in triadic closure and become friends anyhow. A possible answer could be that they do not really become friends. We can say that we rather look at groups that gather around certain topics rather than with a view to establishing friendships. It could even be a disadvantage to be embedded too deeply in the network, as this would increase the risk of information loss to other network members. In other words, having too many friends in common reduces the chances of brokering own knowledge. In the end, the knowledge of network members is the basic asset for their own ability to acquire status via status points and become well-known members of the community. However, this asset needs to be developed in a first step and must be protected later, to reach and maintain a long-term successful network position. Nevertheless, we can find positive effects of being friends in the network. Topics can be shared; a knowledge group can become known for certain topics in the network or topics can be brought up and discussed jointly. But even though friends work together, share information and answer questions, the basic relation and setup remains competitive. Any additional friend who also provides an additional helpful answer to a question will reduce an expert's share of points.[14] This effect may even

[14] The exact point system will be explained in a detailed manner in chapter 4.1.3.

get stronger with the density of the personal network as the information flow improves – and more people will be able to find questions they can answer.

After having pointed out what the influence of the network environment for different groups of the network can be, we want to better understand the dependent performance variable. Individual performance, as a dependent variable, is a measure that has increasingly come to the fore in recent publications. Cross and Cummings examine individual performance as a result of dense networks (Cross and Cummings, 2004) and find full support for greater in-betweenness centrality as a reason for higher individual performance in information and awareness networks. Galunic, Ertug and Gargiulo find evidence for positive spillover effects of social capital and measure the performance of employees as dependent variable via the evaluator ratings of 2,200 bankers of an investment bank (Galunic et al., 2012). Younger brokers thereby benefit from contacts to senior brokers with sparse networks. Ahuja, Galetta and Carley develop a model that explains the effect of both the structural position and the individual role influencing the individual performance (Ahuja et al., 2003) of knowledge workers. The examined sample consists of two virtual work groups both of which showing centrality as being a stronger predictor for performance than individual characteristics. The analysis is based on the email correspondence of team members. A study by Mehra, Kilduff and Brass examines the impact of self-monitoring personalities on the performance of employees in a high-tech company (Mehra et al., 2001). Various other publications discuss the impact of network structures on the performance of companies, teams or individuals in different contexts (Fritsch and Kauffeld-Monz, 2010; Machado and Ipiranga, 2013; Schilling and Phelps, 2007; Vernet et al., 2013). Most of the quoted studies see either dense networks, the occurrence of structural holes, or centrality as independent variables or sources of social capital. We also recognize that the examined networks have the tendency of being rather small and monitored only at a single point in time.[15] In the context of these studies the data on the performance measures is generated in very different ways. We see a range from third party performance evaluation to amount of bonus payments or the number of issued patents. From our perspective,

[15] Ahuja, Galletta and Carley compared the email correspondence at two points in time with a four-year gap. As the sample was a virtual workgroup we would rather treat those two networks as independent observations than a different status of the same network.

all of these indicators are somehow plausible, but we find also notable downsides for these measures. One possible downside could be the risk that qualitative performance evaluations by team members or supervisors may be strongly biased. Typical biases can be aspects that do not have their origin in the network structure and can be an expression of another person's agenda or reasons out of the network context (Hargittai, 2015; Kumbasar et al., 1994). In addition, the quality of the network data itself is not optimal either, as it is partially generated via questionnaires and limited to a certain number of contacts. This data can be strongly biased by self-selection and preferences of the interviewee or may not really reflect the person's real network (Choi and Pak, 2005; Lee et al., 2019). It can also be a sign of socially desired answers (Paulhus, 1991, 1986, 1984; Watkins and Cheung, 1995). The analysis of emails for example may exclude other ways of communication, which may even be of higher importance for certain team members, especially within virtual teams, where they can constitute a more personal and intimate way of communication. Whilst the quoted studies undeniably have their justification and deliver a contribution to social capital and social network research, they are a good example of how difficult the analysis of performance as dependent variable is. To start with, the selection of network and performance variables can significantly influence the results. For our own study this implies that we must find a performance variable that is as little biased as possible and, additionally, we aim to make our network data as comprehensive as possible. As argued, we consider the influence of an online knowledge platform's network structure on the performance of different network participants, and formulate our first hypothesis as:

> *H1: The performance of a network member is correlated with his clustering coefficient.*

Apart from the content-driven aspect of performance we also consider the quantitative aspect of performance, i.e., the quantity of activity a network member contributes to the network. In a knowledge network not only does the qualitative aspect of an answer to a question influence a member's contribution but so does the frequency and number of answers. While we found quite a few publications on the relationship between embeddedness and performance, we sense that it is harder to identify research related to the impact of embeddedness on the activity of actors. We think that activity level can have a positive influence on performance too, and can be defined as a quantitative attribute of performance. Together with a qualitative aspect of the evaluation of work results or a project contribution, activity can create a more holistic view of an actor's performance. Martinez, Arredondo and Roesch studied the relationship of activity on neighborhood cohesion but focused on the physical aspect of the activity (Martinez et al., 2013). We believe that different degrees of embeddedness influence the activity of network actors, such as their level of participation in projects, voluntary tasks or the frequency of their task performance. We argue that actors who are more embedded in their social networks will show different activity patterns to those that are less embedded. Even if the direction of action is not clearly specified – closure being either a positive or negative driver of performance in knowledge networks –we can assume that closure has some influence. As the performance indicators of the quoted studies do not consistently measure quantitative performance, we add a more general performance indicator to our analysis, the degree of activity. Believing that activity itself is evidence of performance, we test the following hypothesis:

> *H2: The activity level of a network member is correlated with his clustering coefficient.*

3.5 Status and duration of membership as moderators

Whilst we were unable to identify social capital research that examines the activity level of network members, we have found a few studies that examine the influence of embeddedness on different levels of hierarchy. As mentioned before, we argue that the effect of embeddedness on performance is influenced by the career status of a network member. We argue that different degrees of embeddedness influence the performance of network members in relation to their career level or, put differently, that for new joiners of networks, dense networks have an edge over structural holes. Having close ties to the surrounding networks reduces integration time and allows quicker knowledge transfer. During career development it becomes more important to leverage the acquired knowledge and, thus, by becoming an independent actor, he becomes increasingly essential to the network. To protect this position, a certain degree of structural autonomy is beneficial. Gargiulo and Benassi find evidence for such an effect when they analyze the adaptability of managerial networks to change (Gargiulo and

Benassi, 2000). They show that the cohesiveness of a manager's employee network has a negative influence on his ability to coordinate the network within a new task environment. In other words, we can argue that dense networks show a certain resistance to change. Their results show that a lack of structural holes raises the number of coordination failures in these networks (Gargiulo and Benassi, 2000, p. 191). Another publication, by Galunic, Ertug and Gargiulo, shows spillover effects from social capital. From their point of view the performance of network members increase when they are connected to brokers of a higher rank, via a sparse (not dense) network (Galunic et al., 2012).

At this point we must introduce the concept of interacting variables. Two main concepts of interaction exist: moderators and mediators. While mediators affect the dependent variable at least partially by having an effect on another independent variable, moderators change the mode of influence of an independent variable along with their specification (Müller, 2009). In other words, the moderator variable influences the direction or intensity of the relation between an independent variable and its dependent variable.

Social sciences describe the concept (Baron and Kenny, 1986; Brambor et al., 2006) which we will introduce in more detail before applying it to our hypothesis. As our data and hypothesis only support the use of moderators, we will focus on this concept. As mentioned, a moderator variable changes the mode of action of another independent variable in a regression model (Figure 1). Regarding the regression model, both terms, the independent and the moderator variable, are part of the function as well as a multiplicative version of both variables.[16]

---

[16] $Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_1 X_2 + e$

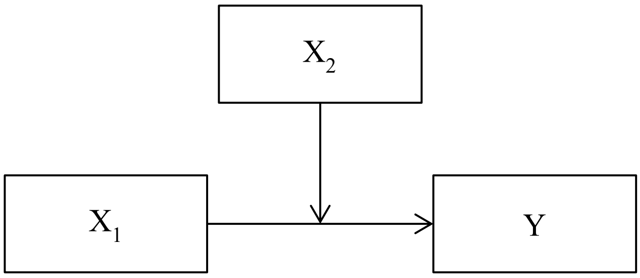


Figure 1: The concept of moderators

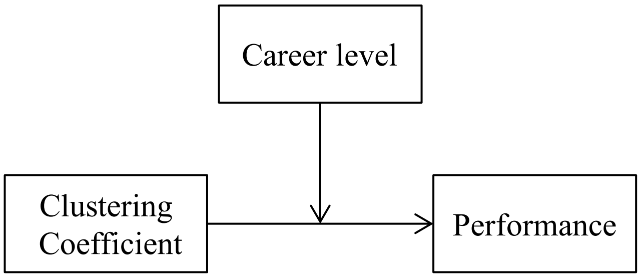


Figure 2: Career level as a moderator

We will consider this aspect in our analysis to examine the effect of two moderators on the relation between hypothesis 1 and 2. Following Galunic, Ertug and Gargiulo, we consider the career level of a knowledge network member to influence the mode of action of the clustering coefficient (Figure 2).

> *H3: The clustering coefficient correlates negatively with the career level of the network actor.*

As we believe that this effect also occurs independently from the career level, we will test whether the mode of action of the closure changes with increasing length of membership duration (Figure 3). We believe that new network joiners need to strengthen their network position by developing a dense network from the very beginning. This allows them to learn the culture and mode of action of the network quicker than if they were only poorly connected. After a period of "onboarding" it is necessary to achieve a certain amount of independence in the knowledge network.

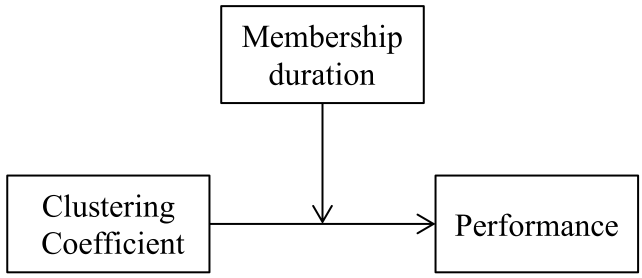


Figure 3: Membership duration as a moderator

Their acquired knowledge or specialization needs to be sold and brokered in an effective way, which is easier to achieve in a structurally autonomous network environment than in one with greater density and closure.

We therefore hypothesize that:

> *H4: The clustering coefficient correlates negatively with the timespan of network membership*

As not all members of a network follow a strict career path but undergo development on an individual career level, we deem career level and membership duration to be complementary variables. Some may develop a career over time, while others may stay on one career level but will also undergo a specific development on this career stage.

4 Dataset Description and Extraction

4.1 Data selection and dataset structure

4.1.1 Introduction to the dataset

The examined dataset was extracted from a public online knowledge platform. The platform went live at the beginning of 2006 and was online until 2018 – although its name has changed in the meantime as well as some of the framework conditions. The platform has undergone a couple of different evolutionary steps in the past, but the basic idea stayed the same: people can pose questions and receive points in return for their answers. Most of the time these questions are simple questions dealing with everyday problems such as, for example, "which 3rd party battery fits my MacBook best?", but they vary greatly in terms of content and level of complexity. According to the answer quality, points are distributed via an algorithm. The aggregation of these points leads to a hierarchical system of career positions that represent a network member's prestige and power. Our data set is a subset of the complete network dataset.

The data preparation of the dataset consists of two main steps, the selection of members and the selection of timespan. After having described the complete data of the network from the first action in the network in December 2005 to June 2008[17], we cut down the dataset to the relevant size of full months after the go-live point up until the end of the available data. In addition, we limit the members to those that are normal network members and do not have the role of a moderator. All the actions that are taking place in the network are written into a SQL-database, marked with the UNIX timestamp of the action time. This database is the starting point of our analysis. The dataset[18] consists of 28 tables that hold data on:

- Users' registration details
- Question details
- Answer details

[17] The actual „going-live" time of the network was on 12th of January 2006. As we are analyzing all connections of the network up to 2008, we needed to start our data extraction from the very start, to get a holistic view on all connections that were established between the actors. We reduced the data in a second step to a time period that consists only of full months after the go-live moment of the network, meaning that we analyze full months from February 2006 to May 2008.
[18] The total database has a size of 5.7GB of data.

- Details on the points distributed
- Details on the different kinds of points
- Status information
- Information about friendship status between members
- Lists of links used in the dataset

Our hypotheses tests refer to two main tables of this database, one holding information about the connections between the actors, and the other holding the information concerning the points that are distributed. The user information and the information about his network of friends are held in a single table giving the information of who requested a contact to whom, with the time of acceptance. This event allows us to define each slice of the networks we look at. Each time a new friendship is established, the shape of our network changes and therefore, after each of these events, the network will be treated as a new network. As a result, we have a potential of 29,622 single networks to analyze, as 29,622 friendships form during our observation period.[19]

We were faced with a very broad field of analytical options with which to approach the dataset. We could extract a great variety of information, from the number of network members, the number of friends of each actor, to the number of friends of those friends – at any given point in time. As a dependent variable we were able to consider both the quantitative and the qualitative aspect of outcomes. For the quantitative aspects, we measured the activity level of an actor, and for the qualitative aspect we measure the points that the actor received for the quality of his answer(s). Both the quantitative and the qualitative dimension are based on the same kind of data: we used the number of events of point distribution as a measure of activity for the quantitative data and the sum of these points as a qualitative indicator for a member's output.

4.1.2 The ego-centered network perspective of the dataset

Our dataset contains a total number of 93,261 members that registered during the mentioned period of two and a half years in our dataset. Of these, 75,972 received some kind of status or credit points for their actions within the network, and 7,877 acquired

[19] We will later argue why we decided to analyze monthly slices of the dataset.

one friend or more during this period. As members can only be connected to the network when they have at least one friend, only this subset is of interest for us. Each friend request from one member to another is recorded as long as they received a positive answer to that request. For our period of time, 29,622 friend requests were accepted, which equals the number of friendships we analyze. In social network analysis terms – at the end of the observation period – we look at a network containing 29,622 ties between 7,877 nodes. As we want to investigate the development of social capital in relation to specific aspects of network structure, we have a very interesting opportunity here in the form of information about each relation in the network in combination with the time when the relation was set up. The first friendship can then be seen as the first, and the last friend can be seen as the last network we analyze. If we look at all ego-centered networks we have a range from 1 to a maximum of 1,020 friends per network member, with a mean of 8.5 friends and a standard deviation of 27.1[20]. A friendship is in this case an undirected edge between two network members. In addition to the friendship perspective, we used the dataset to create an additional variable, the number of friendships between the friends of a network member. As the methodology of this calculation is more complex, we will give a detailed overview about all relevant variables in chapter 4.2.

4.1.3 The point system of the data set

As outlined, the dataset represents a limited life span of a question & answer network. The prerequisite to joining the network is a free registration. After having registered the user will become part of the network's point system and will be able to gain points for different actions he performs within the network. In general, points are collected for a valid and useful reply to questions and the provision, or sharing of, knowledge. The point system consists of 61 different point types that are clustered into two main subgroups, status points and credit points.

[20] These values refer to the adjusted dataset, which has been cleared of the observation with experts having no friends. These observations occurred in the original dataset due to the extraction process and needed to be cut out as the values would influence the result otherwise. The mentioned values do also refer to the final, monthly dataset.

4.1.3.1 Status points

Status points are the backbone of the point system. On the one hand they are an expression of the activity of an expert and on the other hand they are an indication of the value of the provided content. In total there are 34 types of status points, of which 29 were actually distributed in the dataset. The sum of status points an expert collects through his network activity determines his rank and status. Status points are not transferable between experts and are the most distributed point type within the platform. One type of status points provides a good indicator for the quality of a given answer: the points received for an answer voted as helpful (AQ_HELPFUL). The distribution of this type of points can only be initiated by the questioner or the system and might be an especially good indicator for the performance of a network member. There are some other point types that can also be seen as performance indicators but were not selected due to lacking other qualities. Not all these points were distributed equally during the whole timespan of the dataset and some other points show high correlation with the selected kind of status points.

So, whenever a network member answers a question in the network, and the answer quality is evaluated to be helpful, the member receives status points. The number of points that are donated for a helpful answer varies with the importance of a question. The importance again is symbolized by the number of credit points[21] invested by the questioner (Table 1).[22]

Credit points investment	Status points reward
0-9	20
10-24	25
25-49	30
50-74	40

[21] Credit points are the second type of points distributed in the network; we will refer to these points at the end of this chapter.

[22] The maximum number of status points for AQ_HELPFUL in our dataset was 25, the descriptive statistics of all status and credit points can be found in Appendix 4 and Appendix 5.

>100	50

Table 1: Credit points investment versus status points rewards

When the questioner evaluates the quality of the answer, the rewarded status points are split into different quality clusters, namely, not helpful, helpful and top. Due to this cluster system, the points are split (Table 2).

Answer quality	Share of status points
not helpful	0
helpful	1 part
top	3 parts

Table 2: Split of status points due to answer quality

If an answer quality is not classified for a certain period, the platform classifies it automatically. The answer quality will then be ranked as "helpful".

The choice and evaluation of the most suitable variable for performance and activity was challenging. We needed to match two conflicting goals, a representative measure for success and activity within a Q&A network and a sleek approach to data processing. As a result, we focused on a single type of status points that would be a suitable indicator for the building of social capital. When we evaluated the suitability of the different kinds of status points, we considered both quantitative and qualitative criteria. We finally picked the points that were, respectively, both an indicator for answer quality and activity level in the network, the most distributed type of points within the network, and that were distributed during our whole period of the examined dataset.

4.1.3.2 Credit points

In addition to status points, the network uses credit points. Credit points can be considered as a type of currency for the network. As they do not really have an exchange value, the real purpose of credit points is not entirely clear. In later phases of the network's life, credit points could be donated to aid organizations. As a base budget, each network member received 1,000 credit points to start with and to invest into posing questions. We mentioned before that the number of credit points awarded to a question (how the quality or relevance of the question is rated) influences the number of status points that can be received when the question is answered, but in the end, credit points do not add up to a value or indicator that should be further used or evaluated, nor can they be used as an indicator of performance for the purpose of our analysis.

4.2 Data Extraction of the dataset

Due to the structure of the dataset, different tools and software packages had to be used to extract the data. Commonly accepted and proven tools to compute network KPIs like Pajek or UCINET have the drawback by not allowing batch- or panel data-processing. As we needed to compute network KPIs for about 233 million observations, batch processing was crucial for the analysis. Consequently, we decided to compute the input factors for our network KPI – the clustering coefficient – by using a proprietary program based on c-sharp. The rest of the data transformation, calculation and analysis were run in Stata. The complex and large structure of the data made it necessary to use different programs and methods of extraction, consolidation, and analysis, during the extraction and preparation processes. The limitation of standard software packages was the main reason to develop a dedicated software module to extract parts of the data.

The extraction follows the steps from the extraction of the relevant data from the original SQL-database to the preparation of this data using a self-programmed software. After the transformation of the final data into a format that Stata is able to process, the data can be prepared for analysis. We describe this process in the following and provide the level of detail that is needed to understand the approach, and attach the original data, source code and procedures in the appendix.

One of the main differences of this analysis compared to existing studies is the integration of two different observation perspectives. The data will be examined over time, representing the linear development of the dataset, and the individual membership duration of each member. The second perspective – the *lifetime* perspective – will enable us to see the data for each member in the first month of his membership up to the last month of the slice of data. Taking this perspective will allow us to monitor and analyze effects that are caused by the seniority of members in a knowledge network and declutter the data from the fact that members enter the dataset at different points in time and while the average *lifetime* of the dataset changes.

4.2.1 Extraction of relevant tables and pre-cut of data

The first step of the data extraction was performed partly by using SQL-queries on a SQL server and partly by applying a dedicated c-sharp program to the processed data, based on the SQL-server infrastructure. We already noted that we only used a limited set of tables of the available database. The data extraction process is displayed in Figure 4 and consists of four main steps:

1. Import of the base data into a SQL server
2. Selection of the relevant data
3. Calculation of the desired output
4. Extraction to .csv files

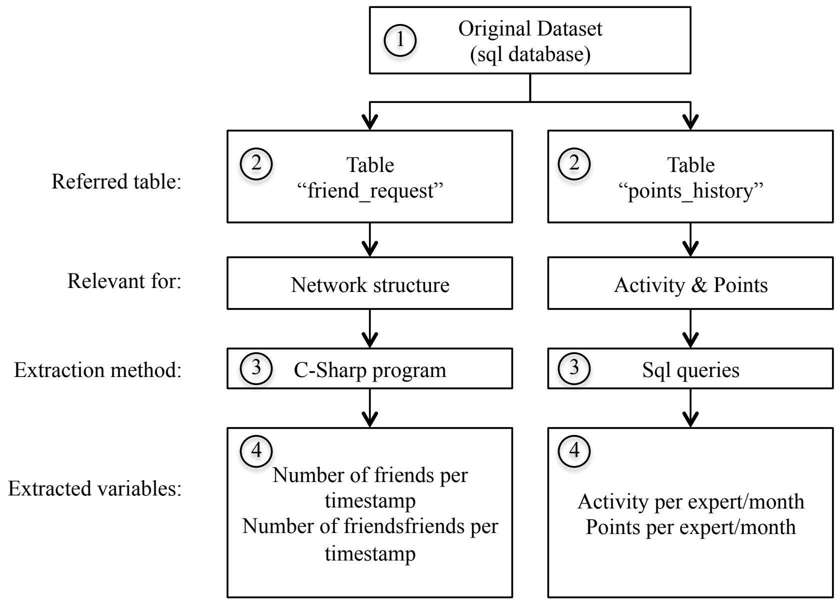


Figure 4: Data extraction process

All the steps of the "friend_request"-related output were managed by a c-sharp program, addressing the different objects, fulfilling the calculation process and providing the output of the data to be further processed within Stata. The second table we extracted data from was the table holding the relevant point- and thereby also the activity-information.

We used a .net framework to set up a SQL-server to hold our relevant SQL tables. The first table we extracted, the friendship registry, is the "*friend_request*" table (Appendix 2). The information of the "*friend_request*" table holds the information about the ID of the network member requesting a friendship, the target ID, the status of the request, the create time, the answer time and possible messages. To test our hypotheses, the relevant data will be accepted requests with both IDs, including the timestamp of the formation of the friendship.

The second relevant table holds the information about all distributed points and is named "*points_history*" (Appendix 3). We use the table to extract two different types of information, the number of status points received per expert and month, and the number of events of the point distribution. The number of events will be our indicator for activity level. We kept only the points which were allocated to helpful answers (AQ_helpful), collating both the number of points and the number of activities.

The table "*friend_request*" had an additional function besides the information about network members. It provides the cutting off points of the network due to the network timestamps. Whenever a new friendship is accepted, the network – and the conditions for social capital – change. We therefore decided to define these 29,622 timestamps as the first measurement points of our analysis. Everything that happens up to each of these timestamps is part of this network.

In the next step we generated .csv files based on four different variables:

1. Number of accepted friend requests per expert for all 29,622 timestamps
2. Number of accepted friend requests between the friends of the experts of the first variable for all 29,622 timestamps
3. Sum of points of each expert, with his status points, on a monthly basis
4. Activity of each expert, based on status points, on a monthly basis

We will now run through the steps of the program code to explain how the dataset is processed. The data we extracted from the original SQL database was written into a new database holding only the table "friend_request". In addition to this table, other supporting tables had to be written to allow the calculation process. All these tables and the referring program code were written into a single SQL file.[23] In the next step the SQL database was opened within a standard database program and the c-sharp program was opened and connected to the database by using the form of Figure 5.

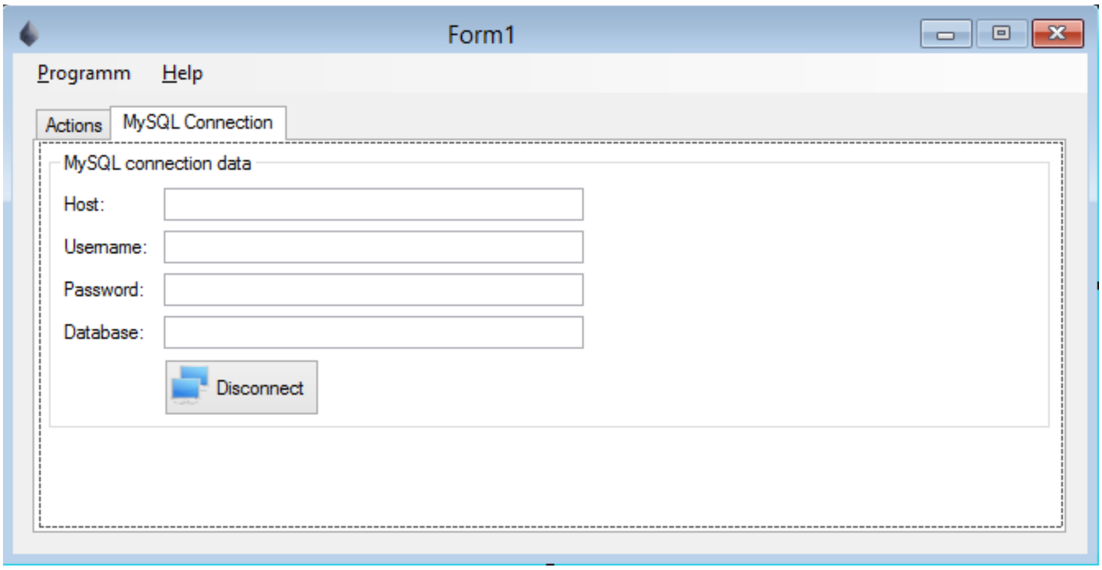


Figure 5: SQL server connection sheet

[23] The program code and the SQL database is provided in the digital appendix (Appendix 16). Further information regarding the digital appendix can be requested from the author.

After having established the data connection, each of the three variables can be extracted using the extraction mask displayed in Figure 6.

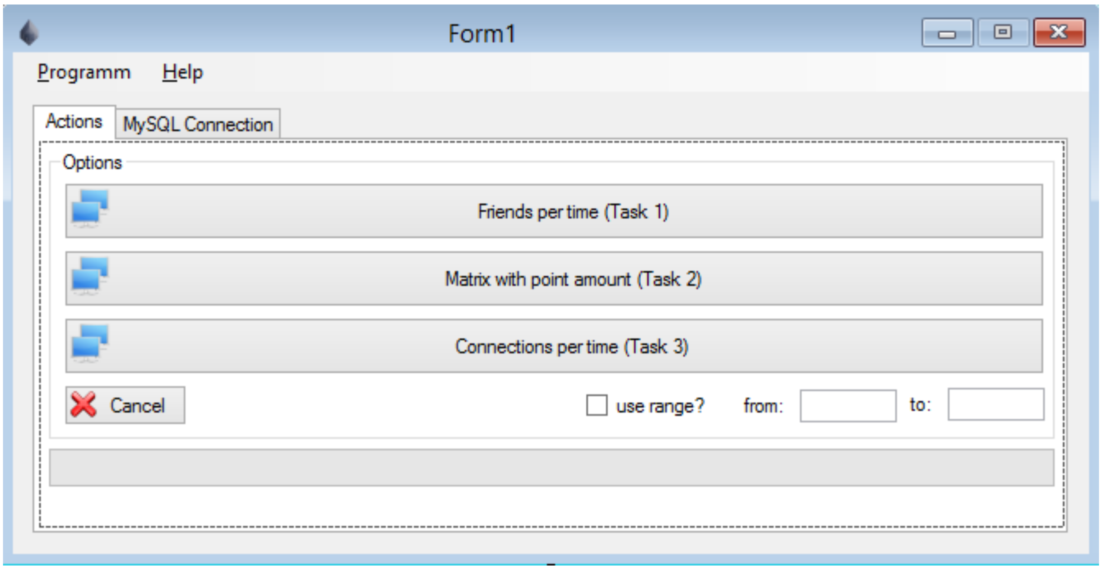


Figure 6: Data extraction sheet

As first variable, we computed the number of friends each network member has (Task 1). We limited the analysis to network members that have ≥ 1 friends. Only these can be considered to be part of the network and thereby can benefit from social capital of the structure (n= 7,877).

These individuals represent one dimension of the matrix holding the different variables. The second dimension is represented by the moments when a friendship request is accepted. These points, 29,622[24] in total, represent the observations from the timeline perspective. As stated before, each of these represent a single network slice. If we combine the two perspectives, we can say that we have a matrix of 233,332,494 observations per variable. The first variable, which we call "friendspertime", shows the number of accepted friend requests per user up to the timestamp it is computed for. The program computes the number of accepted friend requests for each user, disregarding the direction of the request. This means that both active friend request from the user to others (which were accepted) and from others to the user (which were also accepted) are counted. The result will show a "1" to indicate the acceptance of the

[24] The timestamps and the individuals are both distinct

friendship. When the calculation is done, a csv file is generated, named "friendspertime", which holds the information of the expert ID in the first column and the 29,622 timestamps in the second column.

The computing of the next variable was far more complicated and the most challenging part of the data extraction (Task 3). When we computed the variable "friendsfriendspertime" we wanted to know how many of the direct friends of a network member are friends themselves. This means that the criteria of an accepted friend request needs to be met, either from the one or the other side, between two friends of an expert. As the whole program, the program code and the results are part of the appendix of this research, we only summarize the procedure to extract these values. For each of the 7,877 experts we had to generate a table of all their friends at each of the 29,622 points in time. We then created a sociomatrix (Wasserman and Faust, 1994, pp. 80–83) for these friends. We counted the friendships, and divided them by two, as the matrix shows each relation twice. The results were saved as csv files. This calculation was by far the most complex one, as we did not compute values in the first step but had to iterate over an array of 29,622 sociomatrices to calculate our results.

The last two variables, the points and the activity were computed by SQL queries (Appendix 6). In the appendix we provide the source code as well as the whole program and the extracted files and the SQL queries for each step of the extraction process.[25]

Due to the data extraction process, we had to verify the quality of the data in different ways. A dedicated program computed the number of friends and friendships between friends. This led to the need to verify the data in an alternative way and not only by the program, as it would produce the same results repeatedly. We chose to compare the results of 50 random combinations of the 7,877 experts and 29,622 timestamps

[25] As we developed the code of the extraction software in the beginning, we computed the number of status points for each expert at each of the 29,622 timestamps, too. We decided to add activity as a variable as well as the monthly perspective of the dataset. Later we decided to create the variables *points* and *activity* by using my-SQL and drop the data generated by the c-sharp program. The main reason to do so was the extensive work that needed to be done to cut and reshape the data to a proper panel dataset as well as the long computing time. The results of both methods are the same. The C-sharp program is designed to work in a .net framework and therefore requires a Windows-based machine. All other calculations and analyses can be performed both in a Windows and in a Mac OSX environment. Figure 4: Data extraction process, still shows the full capability to calculate the data by using the c-sharp program we generated.

based on SQL queries. These queries were made on the original dataset without any transformation and tested against the final dataset in Stata (Appendix 7). One variable, the number of friendships between the friends of an actor, could not be verified by using simple SQL-queries. The reason for the complex computing of the second variable, via c-sharp, was the inability to process this calculation via SQL-queries. Consequently, the quality check needed to be set up in an iterative way. A program code-based quality check would always use the same algorithm as the original one and come to the same results. To compensate for this, we started by verifying the program code, observing general standards for code review. During the coding of the program, we evaluated the single iteration steps and found no evidence of miscalculation. The most robust proof for the correctness of the variable "friendsfriendspertime" was the final variable we computed using Stata. Based on the number of friends, we calculated the number of possible connections between these friends and used this together with the "friendsfriendspertime" variable to calculate the clustering coefficient. None of the 233,332,494 clustering coefficients were calculated out of their range from 0 to 1 and the number of zeros was greater than the number of non-existing friendships for the first variable.[26]

4.2.2 Data preparation for Stata

To test our hypotheses, we needed to create a panel data set. This meant that we had to transform the single variables we computed into a format which Stata could import and transform into panel data. As the number of variables was limited in Stata, the dataset had to be divided into 8 parts per variable, which then needed to be transposed and put back together again (Figure 7). This procedure had to be done for both variables, the friends per expert variable (friends), and the friendships between the friends per expert (friendsfriends).

---

[26] For completeness, we calculated the number of friendships of zero in the dataset. Since the network built up over time and the selection criteria of the network members was to have one or more friends at any one time in the network, this observation is quite normal. When somebody started as a member of the network in 2008 and connected to others, a few days later he will still have zero friends and status points. As a consequence his „friendsfriends" and clustering coefficient will be zero, too. This data is eliminated in a later step as the panel would otherwise have been wrongfully balanced.

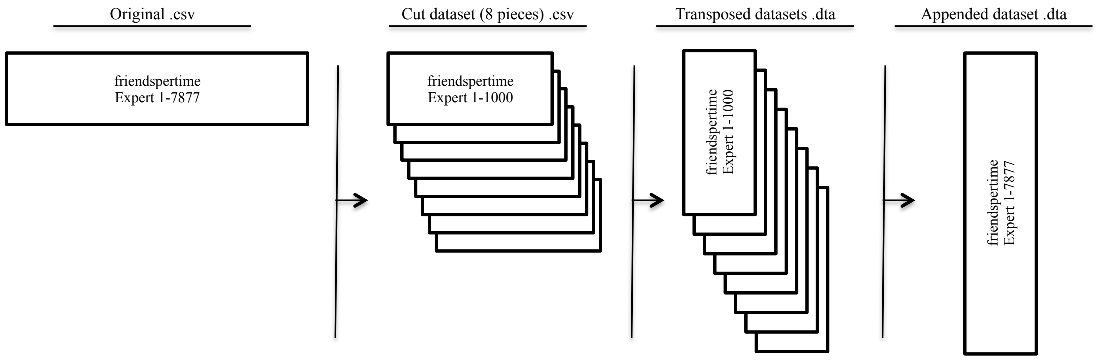


Figure 7: Data transformation process to panel (example friendspertime)

The datasets per variable were then merged into the final dataset. The final dataset held the expert ID as panel variable and the timestamps as time variable. In total, the dataset contained 233,332,494 observations.

4.2.3 Data preparation in Stata

In this part, we run through the steps involved in our data preparation and generation of variables within Stata. Although in the previous section we described the first steps of the data preparation, which already involved Stata, we see the real starting point of data management with the program now. We executed the following steps to prepare the data for analysis:

1. Generation of a real time variable that can be assigned to real dates
2. Reduction of the dataset to the last data-point of each single month
3. Calculation and creation of the new variables
4. Appending of the activity and point data generated by SQL queries
5. Generation of additional variables and setting to panel dataset

1. Generation of the real time variable

The original data only had timestamps as time variable and was therefore not easy to read and process for analysis, so we decided to transform the timestamp in real time (Appendix 8). In so doing we could change the format of the time from the original timestamp point in time to days, weeks or months and test the hypothesis in each examination level. As the timestamps are distributed unevenly throughout days, month and years we can also smoothen the data by aggregating the results and filling the gaps with the data of the previous period.

2. Reduction of the dataset to the last data-point of each single month

In a next step we created a variable which reflects the month of the observation and for each expert dropped all observations but the last of each month for the *friendspertime* and *friendsfriendspertime*. We chose to reduce the dataset to monthly slices for several reasons. First, the timestamp-based dataset was distributed quite unevenly, and many gaps needed to be filled with data from previous periods. Second, both a daily and a weekly dataset was from our point of view too granular and would have required an evaluation of the time-delayed effects from one day or week to another. Third, the monthly perspective allowed us to monitor the cause and effect of the clustering coefficient and the other independent variables on the dependent variable within one period and still represented a long time series of 31 periods, or 2.5 years.

In addition to the reduction to the monthly dataset, we renamed the variables *friendspertime* and *friendsfriendspertime* to *friends* and *friendsfriends* and renamed the variable *expertid* to *expertid_continous*, as it does not reflect the original expert ID from the dataset but a newly assigned ID due to the order of appearance in our dataset.

3. Calculation and creation of new variables

We prepared the calculation of the clustering coefficient by generating the maximum possible number of closed friendships between each expert's friends for every month. This variable *possiblefriends*[27] was further used to compute the clustering coefficient by dividing the number of actual realized friendships by it (*friendsfriends*)[28].

4. Appending of the activity and point data generated by SQL queries

As we computed the status points for answers of good quality and the activities per month by SQL queries, we had to create Stata datasets from this data. In a next

---

[27] $possiblefriends = \frac{friends*(friends-1))}{2}$

[28] $clusteringcoefficient = \frac{friendsfriends}{possiblefriends}$

step we merged the three datasets holding all variables – *friends*, *friendsfriends*, *points* and *activities* – for each expert and month.

5. Generation of additional variables and setting to panel dataset

In a next step, we dropped all data for observations with no friends. This was necessary, as Stata would otherwise have handled the dataset as a balanced panel even though it is unbalanced. We added a *lifetime* variable that counted for each expert and active month a lifetime. We also added and the number of experts that were part of the network for each month (*amountmember*). The variables *points* and *activities* were calculated for each month, and we added a cumulative perspective for both variables. We calculated both the *points* and *activities* for the status points, but only for helpful answers, and the sum of *points* and *activities* for all status points. This was a prerequisite for the calculation of the status *ranking*. According to Kundisch and Mutter (Mutter and Kundisch, 2014) we added the *ranking* of each expert to enable us to distinguish between the different status levels, which are an indicator for the career level of network users. To enable a selection of users that are normal users of the network and do not have a special role as a moderator, we added the *role* and *status* of the experts.[29] In a last step we transformed the dataset into a panel dataset, using the expert ID as panel variable and the month as time variable. At the end of this procedure, we had generated the following variables of Table 3.

[29] The *role* and the *status* of the experts were extracted from a table holding the expert basic information and then matched with the point, activity and friendship data.

Variable Name	Value Content
expertid_continous	A continuous ID for each expert. To reduce calculation time we replaced the original ID by a number from 1 to 7,877.
expertid_original	Original number provided for the experts' ID to allow data consistency verification with original data
expert_role	The role of the user in the network. Only "users" were selected
expert_status	The status can have different, non-active attributes; only users with status "user_ok" were selected
month	Month of observation in the format %tm
month1	Continuous number for each month
friends	The number of accepted friend requests either from or to each single expert for each timestamp.
friendsfriends	Accepted friend requests between two friends of an expert
possiblefriends	Maximum number of possible friendships between the friends of an expert $$\frac{friends * (friends - 1))}{2}$$

clusteringcoefficient	Share of closed friendships of maximum number of possible friendships between the friends of an expert $$\frac{friendsfriends}{possiblefriends}$$
aq_helpful_points	Number of points received for answers with good quality within the month of observation
aq_helpful_count	Number of single actions that led to a distribution of status points for good answer quality
aq_helpful_points_sum	Total number of points from first month of dataset membership including month of observation per expert
aq_helpful_count_sum	Total number of activities from first month of dataset membership including month of observation per expert
grandtotal_points_sum	Sum of all points for answer quality_helpful for all experts of the dataset
grandtotal_count_sum	Sum of all activities for answer quality_helpful for all experts of the dataset
entrymonth	First month of dataset membership due to *month1* variable
lifetime	Number of dataset membership month including observed months
amountmember	Number of experts with >= 1 accepted friendrequest for each month of the dataset
ranking_by_system	For each expert and month, the status level due to the ranking system of the dataset was assigned

Table 3: Overview variables final dataset

4.2.4 Discussion of final dataset

Despite the limited amount of information and number of variables that we were able to extract directly from the original data set (i.e., the friendships between the members of the Q&A platform and the main success indicator, the point system) we were able to build a dataset with sufficient variables required for our analysis.

The process of data extraction turned out to be a very complex and time-consuming part of this research. Both the path and processes were the result of several iterations. When we tried to extract the dataset by using different methods in the beginning, we faced two main issues. The first issue was related to the software and the second with hardware. None of the software packages we tested, e.g., UCINET or Pajek, were capable of processing batch data without programming of additional functionality. In addition, we could not reach acceptable computing times using mySQL queries for the given tasks, and the second task (friendsfriendspertime) was not at all computable using SQL. Even after having solved the issue with our own software, the total calculation time, for the extraction of the first .csv files alone, took several days on a high-performance machine. Another issue was the reproducibility of the dataset. All the methods and software we used are well established in the literature and commonly used. The number of different process steps, from extraction to the final dataset, is minimal and, for most steps, the process is automated. As a result, we were able to reduce the error potential and verify the data at several stages during the process. Only two files and a few SQL queries were needed to reproduce the final data.

In total, we computed basic data for 7,877 individuals at 29,622 different points in time, each of which represents a single social network as the points in time represent the moments when a tie forms between two nodes. We extracted and calculated three variables: the friends of an expert, the number of friendships between the friends of an expert, and the number of points the expert received. In total this comes to 699,997,482 data points, or 233,332,494 points per variable.[30] The final dataset can be used to generate all further variables and is reduced to monthly chunks of panels over a period of 2.5 years for 7,877 experts, with the constraint that the number of experts increases during the period of observation.

[30] In the beginning we did not know, which aggregation level we would choose for the data. As a consequence, we extracted the number of points per expert also with the use of a c-sharp program and later replaced the data by the monthly data of the SQL-query.

4.3 Metrics for Hypothesis testing

When we look at our dataset, we can take different perspectives. The first intuitively chosen one will be the perspective of variable distribution over the timespan of the dataset, as we look at the data over a certain period. As our third and fourth hypothesis do refer to the lifetime of an expert and the rank, we will include those perspectives and focus also on them. The perspective of a month has the downside of mixing data of newer and older network members and neglecting their overall performance. This again does not allow us to examine the expert perspective but forces us into the perspective of the total network. In the following paragraphs, we will therefore focus on the lifetime and career-level orientated perspective of the dataset and compare it to the monthly perspective if suitable.

4.3.1 Amount of experts and average lifetime development

The expert perspective is the main angle point and panel id of our dataset. In the last observation of the dataset, we have 7.877 experts with at least one accepted friend request. The experts build up over time and simultaneously the friends and friends-friends of these experts build up. As we can see in Figure 8, the number of experts builds up exponentially and slows down at the end of the observation period.

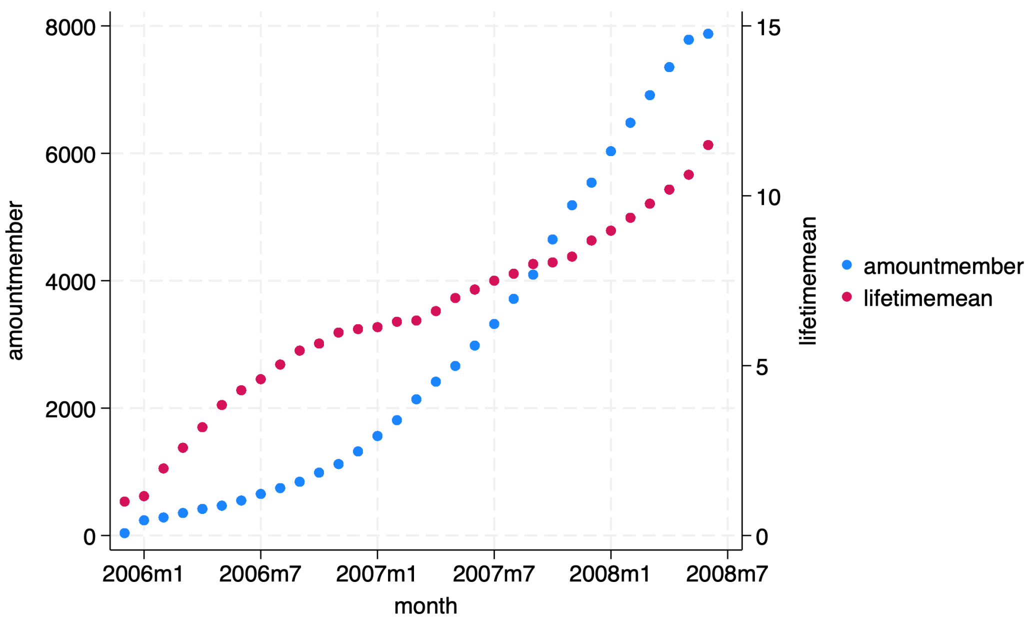


Figure 8: Expert build up and average lifetime per month

When we look at the data from the perspective of the average network membership duration per expert and month, we can see the average age is declining as the exponential growth of the members slows down the average aging of the membership duration. Only at the end of the observation period we can see that the average age is growing exponentially again. In general, we can say that we look at a population which is growing in the beginning, stable over a certain period of time and slowing down to grow at the end. As a result, the average membership lifetime is the mirrored function of the amount of member development over time.

For our analytic purpose we have the advantage to have a good mix of experts that are long-term members of the dataset and others that have joined the network recently (Figure 9). We can thereby state that we mitigate the growth effect of the network and that our analysis is not biased by it.

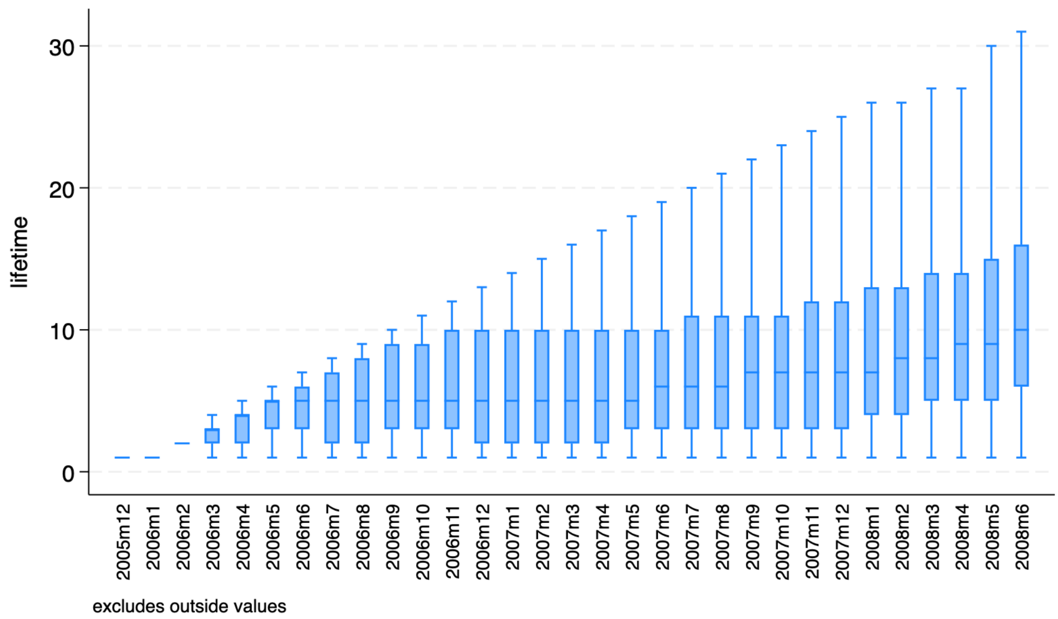


Figure 9: Boxplots of membership lifetime per month

This allows us to examine a population that is not dominated by one type of member, showing only a specific lifetime. In addition to that we can use the lifetime of a network member as independent variable.

4.3.2 Friends, friendsfriends and clustering coefficient

The second set of variables we extracted from the database, reflects the number of friends of an expert and the closed friendships between these friends. Based on this data, we computed the resulting clustering coefficient.

From the monthly perspective, the average number of friends per expert starts at a minimum of 2,5 and runs up to a maximum of 10,2 with a total mean of 8,5 and a standard deviation of 0,7 (Figure 16). From the lifetime perspective the average expert has 2,6 friends in his first month of the membership and a maximum number of 21,9 friends in the last month of the membership. The total average is 8,5 for all month of lifetime, with a standard deviation of 4,0 (Figure 16). When we take a look at the respective boxplots, we can see a wider variance of the values in the upper quartiles than below the median (Figure 10).

The “typical” network member seems to have less friends than the upper quartile members, which could be more focused on the acquisition of new relationships in the network. Members that are extreme “friend collectors” (the maximum number of friends per expert is 1020) are not reflected graphically in the boxplots, as they exclude outside values, but are included in Figure 16.

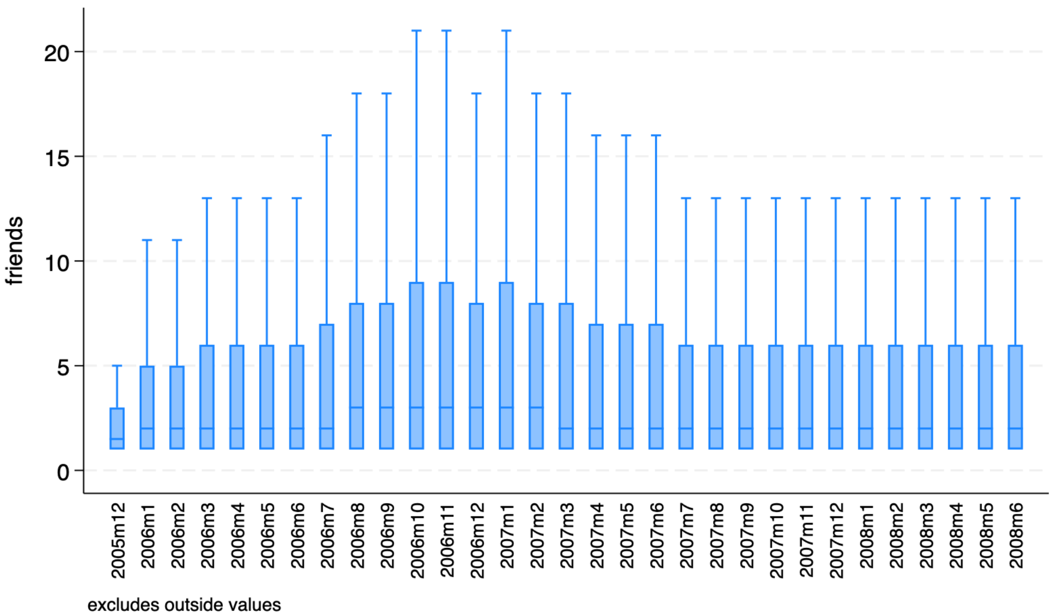


Figure 10: Friends per month box

Compared to our reflection on the experts in the previous section, we can again state, that the monthly perspective of the dataset does not allow straight assumptions, as there might be a number of overlapping effects in the network (Figure 10, Figure 11). The lifetime perspective on the other hand implies a “career” of friendships and friendships between the friends of an expert. This career seems to follow a certain pattern and comes to a climax of as well friends and friendsfriends (Figure 12, Figure 13).

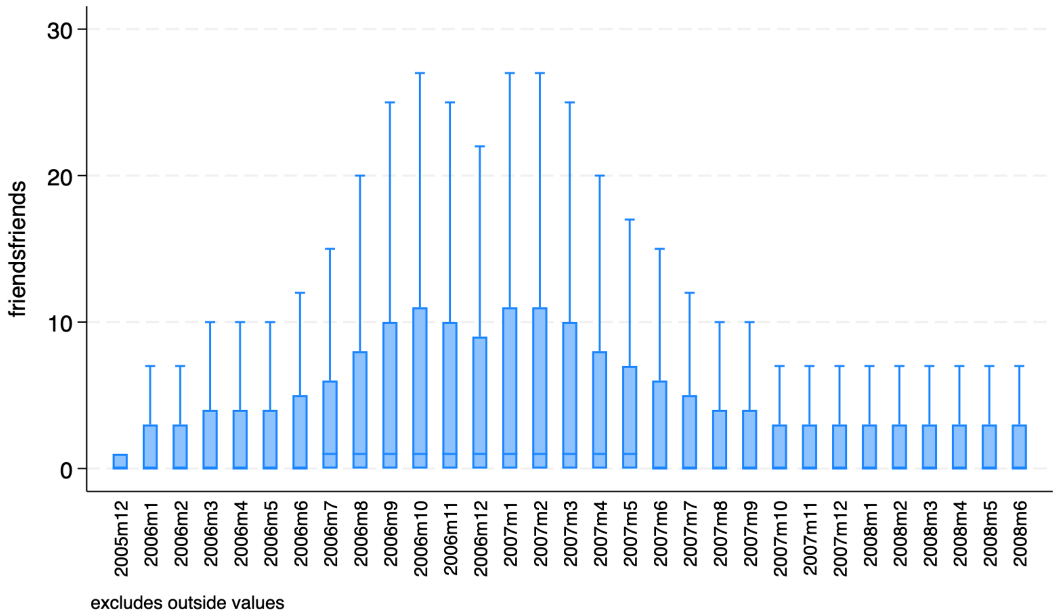


Figure 11: Friendsfriends per month box

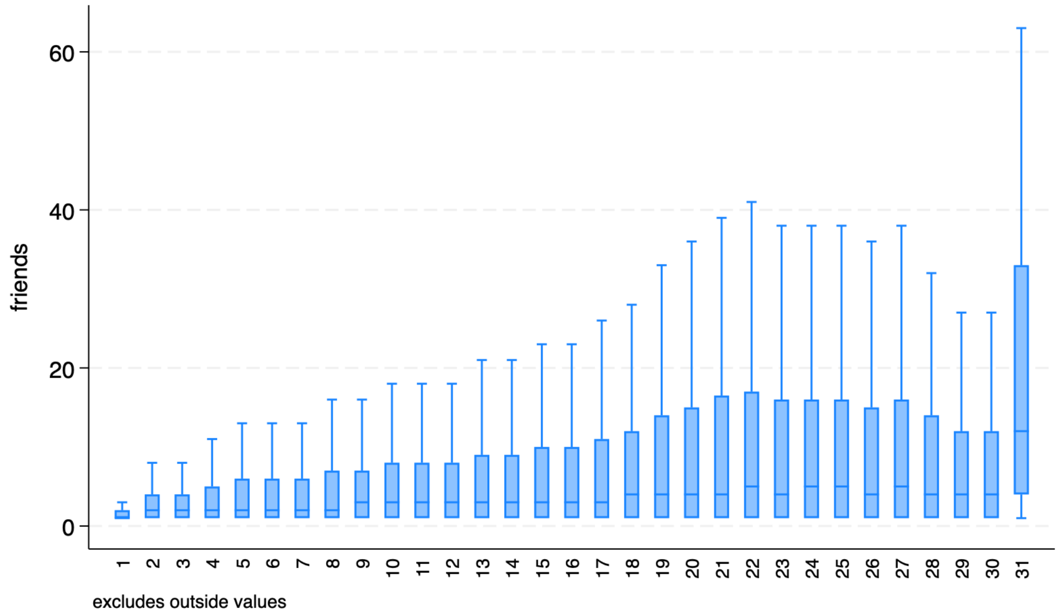


Figure 12: Friends per lifetime box

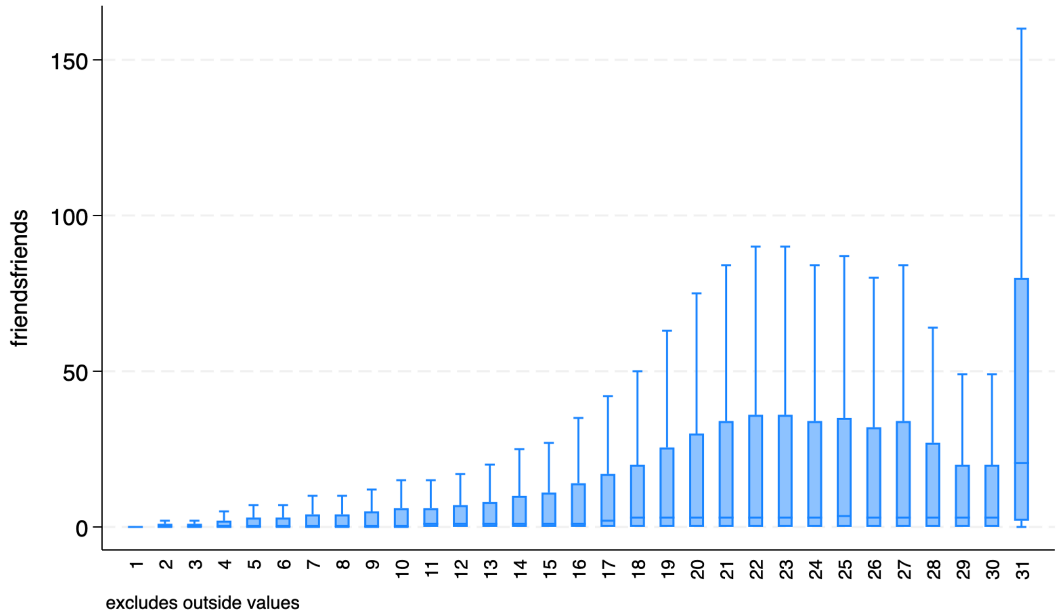


Figure 13: Friendsfriends per lifetime box

A similar pattern can be seen when the clustering coefficient is plotted. The monthly perspective does not follow an as stringent pattern (Figure 14) as the lifetime perspective which seems to follow the track of building up a clustering coefficient which fades out in the later month of the dataset membership (Figure 15).

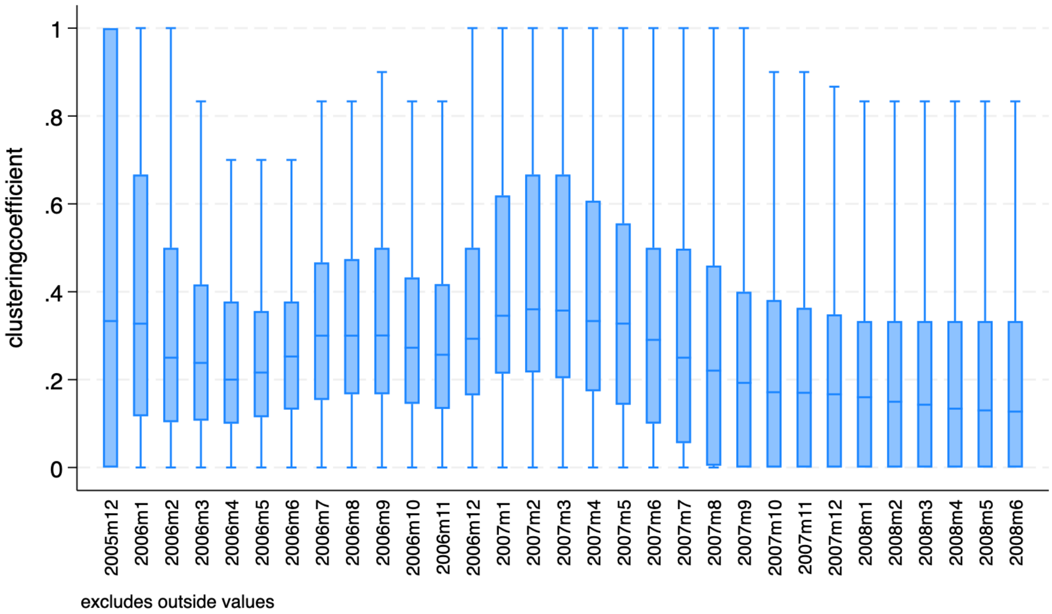


Figure 14: Clustering coefficient per month box

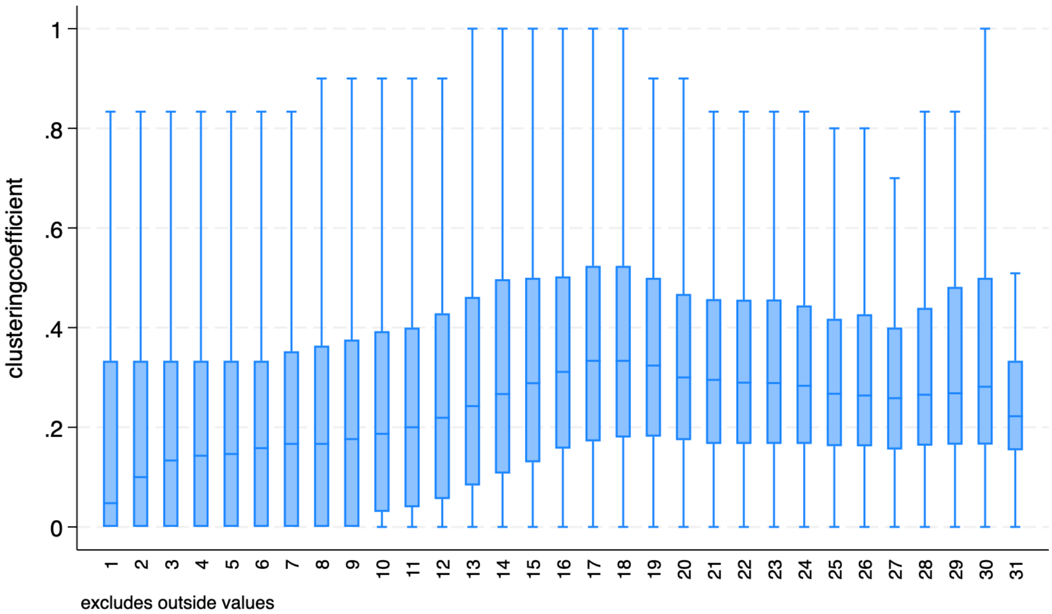


Figure 15: Clustering coefficient per lifetime box

These effects become even more obvious when we look at the means of these variables. The values follow a much more stringent pattern of a "career" when we look at it from a lifetime perspective then from the monthly perspective (Figure 16).

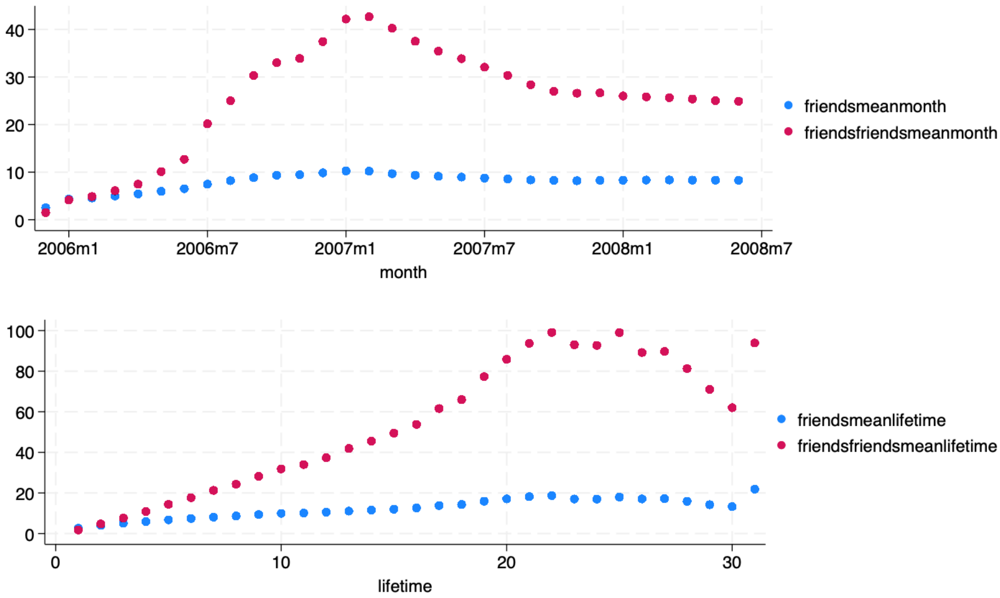


Figure 16: Mean friends and friendsfriends per month and lifetime

When we examine the clustering coefficient from a monthly perspective, we find a random walk trend, which flattens out at the end of the observation period (Figure 17). In contrast, the clustering coefficient shows a gradual increase in the beginning of experts' membership, decreasing after reaching its peak after 18 months, but with a slight increase in the latter stages of membership duration (*lifetime* variable).[31]

[31] We will later cut the dataset and drop the data of the first months of the dataset as it had not been online at this point of time, so this effect will not occur in the same intensity anymore.

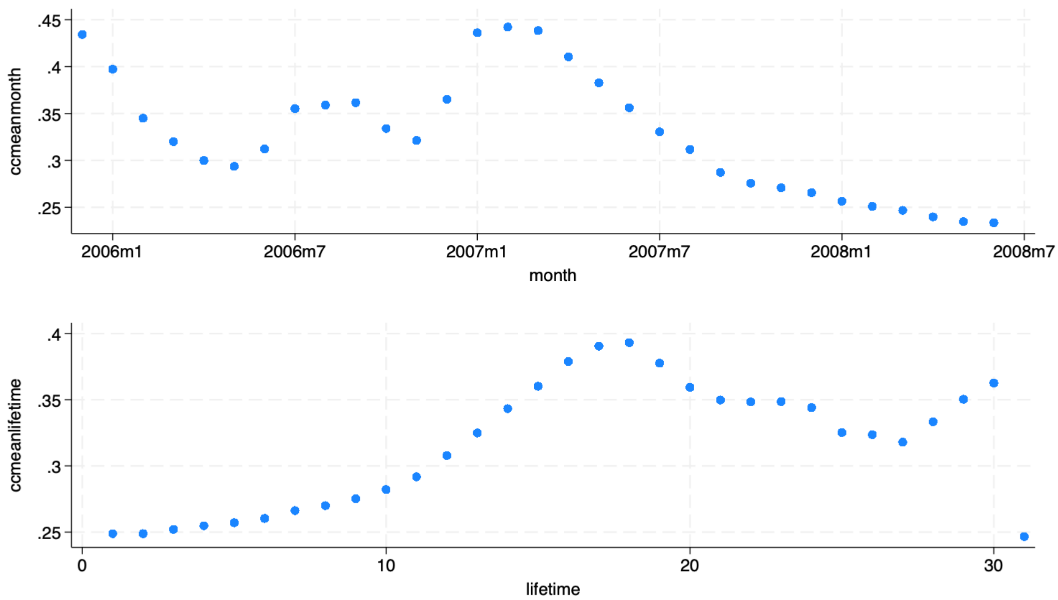


Figure 17: Mean clustering coefficient per month and over lifetime

4.3.3 Performance

When we look at the distribution of the performance variable over the months and lifetime, we can also see a differentiated picture.

The boxplots show a similar pattern. The average points per activity have a higher median and quite more variation in the first 18 months of the dataset, flattening out afterwards (Figure 18, Figure 19). The average number of status points rises and falls quite a few times, on a monthly basis, only to flatten out in the second half of 2007 (Figure 20). Seen over the whole lifetime, the performance stays quite stable. The means of these values are about the same, 3.67 points per activity for the lifetime perspective and 3.62 points per activity for the monthly perspective[32]. The standard deviation of the monthly perspective is 0.63 compared to 0.12 points in the lifetime perspective. A possible explanation for the slight differences between monthly and lifetime perspective could be that new members tend to be more active, both from a quan-

[32] Technically they should be identical; the difference is caused by a number of inactive experts in their last month on the platform. This incomplete month will be cut from the dataset for our hypothesis test.

titative and a qualitative perspective. Given the fact that the dataset is growing exponentially in the beginning, this could lead to a peak in the average points from the monthly perspective.

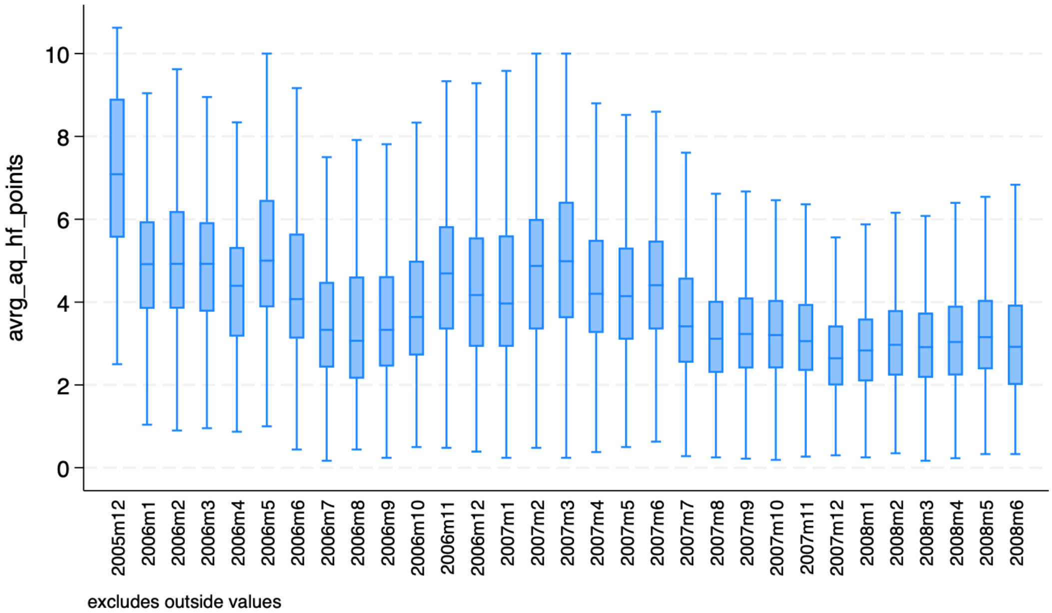


Figure 18: Boxplot average points per activity by month

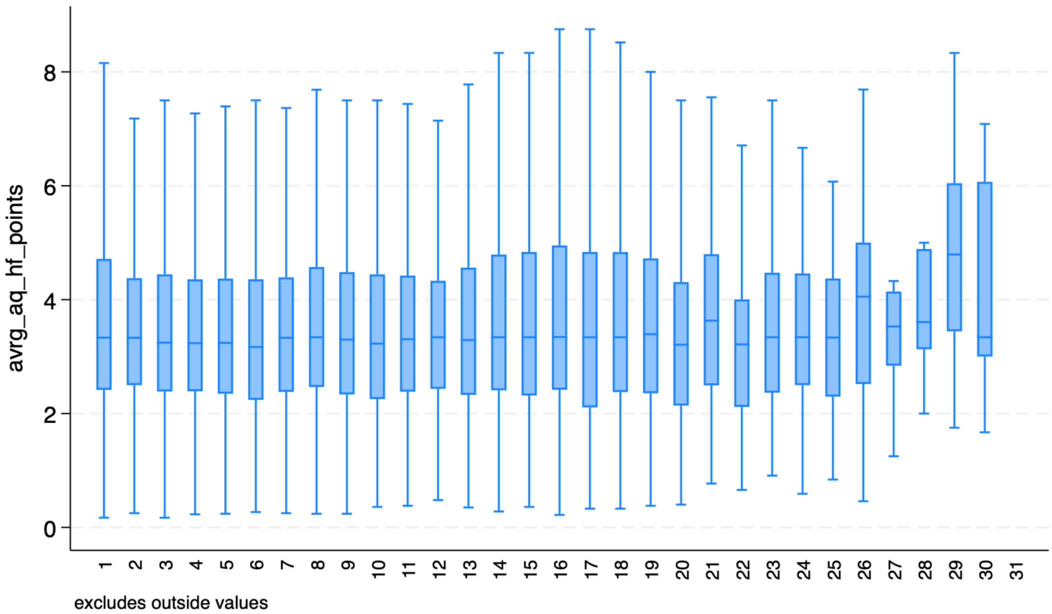


Figure 19: Boxplot average points per activity by lifetime

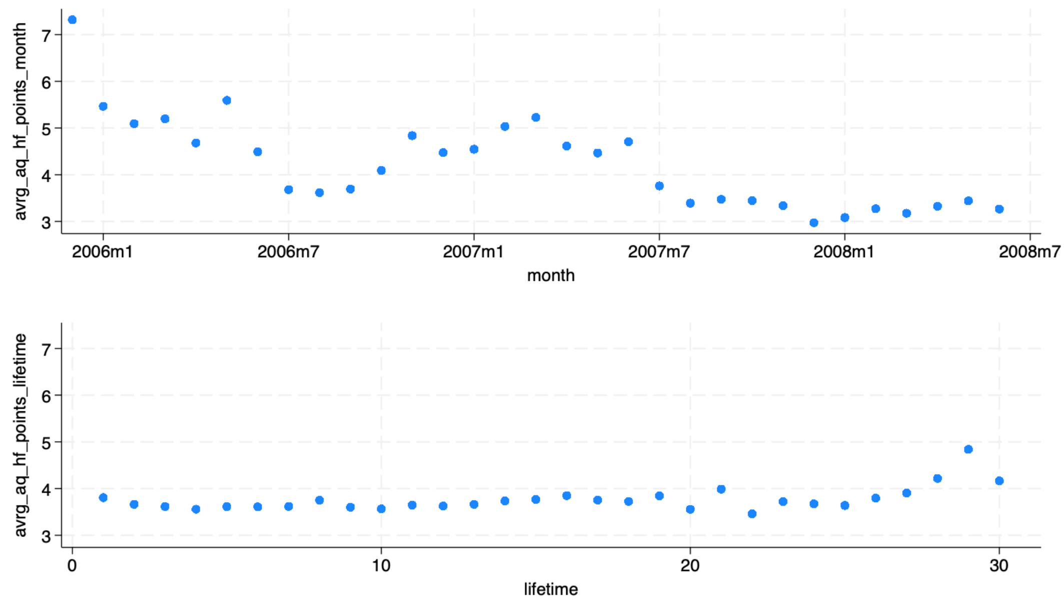


Figure 20: Average points per month and over the lifetime

4.3.4 Activity

Looking at the activity variable we find a different picture again. While the activities do not seem to follow any distinct pattern or trend on a monthly perspective (Figure 21), we can observe an obvious decline in activity during the lifetime of an expert's membership (Figure 22).

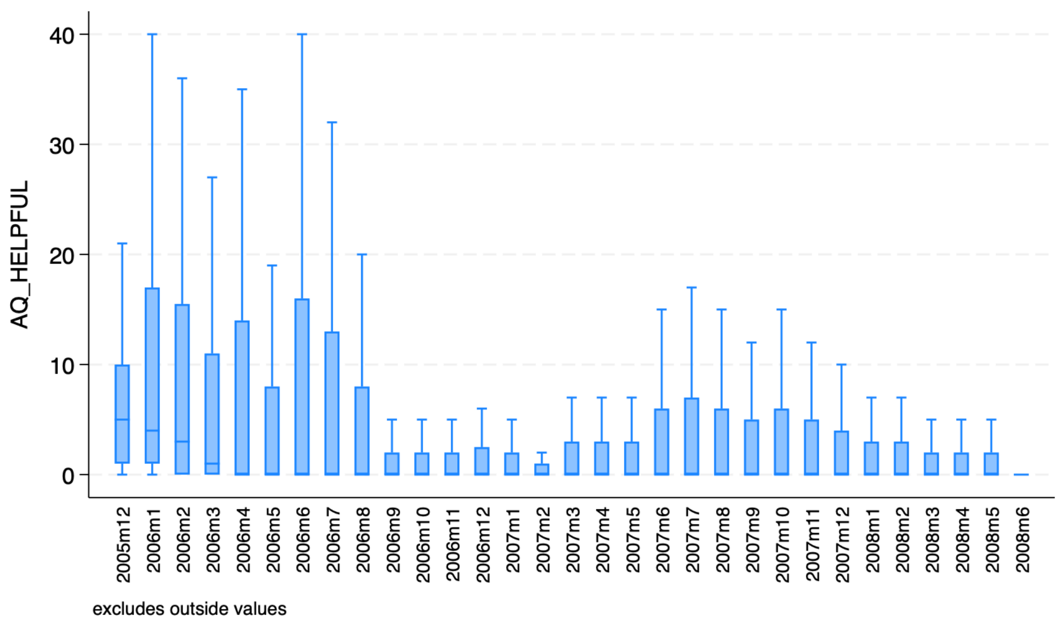


Figure 21: Monthly activities box

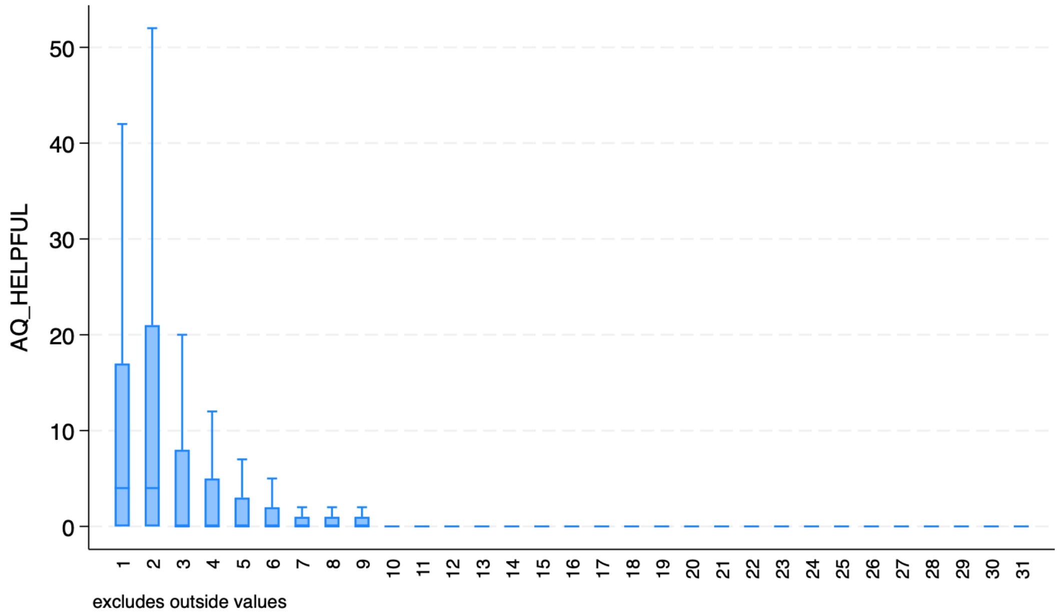


Figure 22: Lifetime activities box

The same picture emerges when we look at monthly and lifetime means (Figure 23). While there is no general trend visible in the monthly graph, we can see a clearly decreasing trend in activity level over the lifetime of the average expert. The average number of activities is 11.21, the standard deviation for the monthly perspective is 4.64, and for the lifetime perspective it is 6.77. We assume that the explanation for the high variation is similar to the explanation for the variation of points per activity. From the monthly perspective, we see a mixture of experts who have only just joined the network and those who are more senior. From the lifetime perspective, we see a decline of activity after the first month of membership.

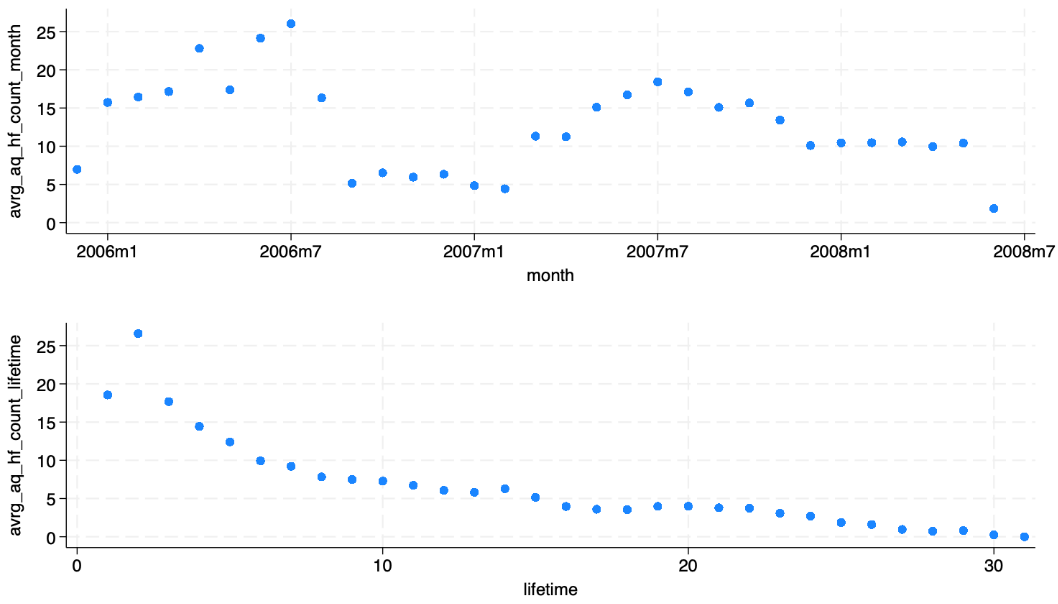


Figure 23: Mean activities per month and over the lifetime

4.3.5 Hierarchy

The dataset contains a dedicated career system in which experts can develop individual careers based on the number of status points they earn. Even though the points received for good quality answers are the major type of points distributed, other types of activity rewarded by points drive the career level too (Appendix 4). The career system was revised in February 2007 and changed from an 18-stage model to a 20-stage model.[33]We considered this change in our dataset and assigned the respective career level at each point of the dataset. The career model follows an exponential path, as it requires more points to achieve the next level compared to the previous one.

[33] For a dedicated analysis of the career system of the dataset see (Mutter and Kundisch, 2014)

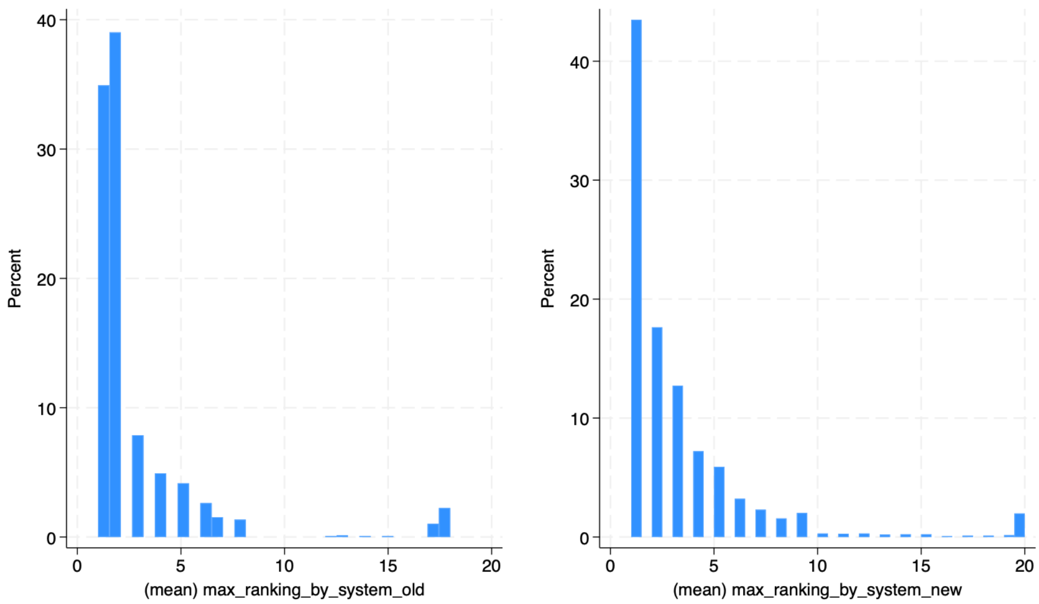


Figure 24: Maximum rank by before and after the ranking system reformation

When we look at the maximum rank per expert before and after the reformation of the point system, we find over 60% of users with a rank of 1 or 2, followed by declining percentages to a rank of level 8 or 9. Afterwards the highest ranks (17, 18 and 20) have higher percentages and the assumption is that these users reflect power users and moderators (Figure 24).

Even though all the mentioned levels were assigned to experts at least once, the average rankings are quite low, as we can see in Figure 25. It is also noticeable that the first ranking system leads to an exponential, or inflationary rise of the average rank. After a significant consolidation in February 2007, this trend could be transformed into a more linear development for the second part of the dataset. However, we are convinced that the hierarchy level might have an influence on the direction taken by the clustering coefficient and will test this in our third hypothesis.

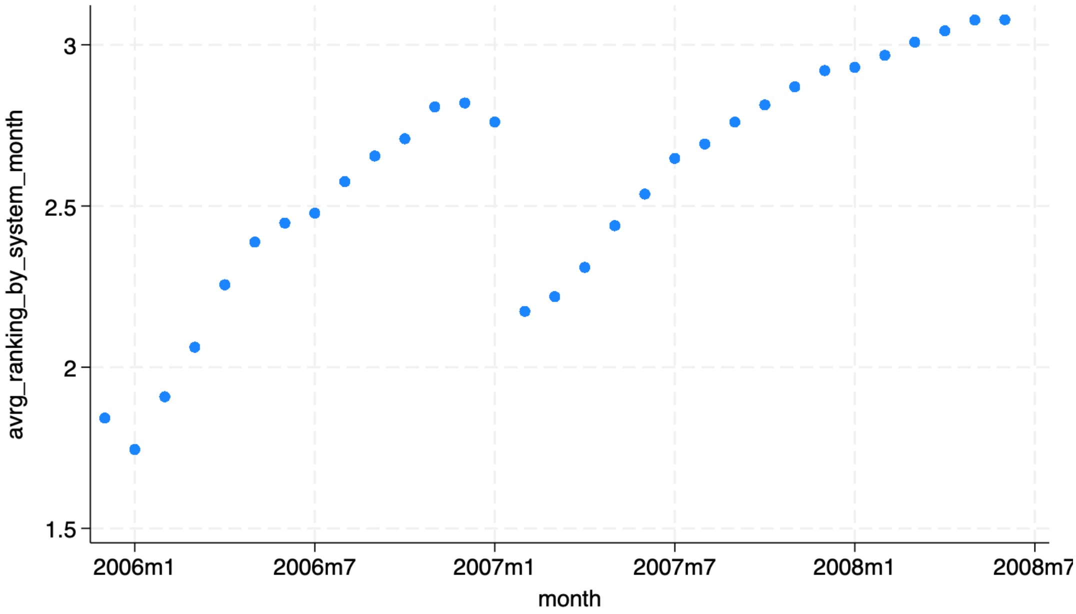


Figure 25: Mean hierarchy by month

4.3.6 Summary of metrics

In total, we generated about 20 variables, including those used to create our figures or to compute variables. To test our hypotheses, we focused on the number of an expert, his lifetime, the observation month and the points he received as well as his activity and status level.

The remaining variables were needed to describe the dataset and to create a comprehensive view of the setup of our study. This chapter has shown that the visualization of the data, as a first step of the analysis, has given first indications on how to approach the dataset adequately. The first, intuitive way to approach the dataset was to plot the variables along the timeline of the slice of data we want to analyze. We soon recognized and assumed that for all variables we were looking at mixed effects and behavior of network members with shorter and longer membership duration. Especially with regard to the work of Galunic, Ertug and Gargiulo (Galunic et al., 2012), we assumed that the structuration of the dataset along membership duration would show clearer effects in differences in behaviors and, therefore, would serve as a valid predictor of the performance of network members in knowledge networks. The lifetime perspective therefore can be seen as a variable representing the career level or experience of an actor within the context of a knowledge network or social network in general. Taking this perspective would allow to see the data free of effects that occur due to an alternating stream of new members entering the database during our observation period. If we were to only plot and analyze the entire data during the observation period regardless of membership duration, we would risk combining the effects of younger with that of more senior members.

The structure of the data shows some irregularities, however. For some variables we see deviations from the general trend and could assume that the data is somehow polluted by members with different attributes or a finishing month that has characteristics that do not match those of the rest of the sample.

In the next chapter, we will therefore reduce the dataset to allow a clean analysis to answer our hypothesis. The summary statistics of all variables that have been created are shown in Table 4, and the corresponding Stata do-file used to create the graphics and the related variables can be found in Appendix 16.

Variable	Observations	Mean	Standard deviation	Min	Max
expertid_continous	90,545	2,655.45	2,034.62	1.00	7,877.00
expertid_original	90,545	116,101.40	118,822.10	1.00	460,283.00
expert_role	0				
expert_status	0				
month	90,545	573.45	6.40	551.00	581.00
month1	90,545	573.45	6.40	551.00	581.00
friends	90,545	8.49	27.08	1.00	1,020.00
friendsfriends	90,545	28.04	156.59	-	4,454.00
possiblefriends	90,545	398.44	8,222.57	-	519,690.00
clusteringcoefficient	55,025	0.29	0.31	-	1.00
aq_helpful_points	90,545	42.96	189.43	-	8,432.49
aq_helpful_count	90,545	11.21	44.17	-	1,622.00
aq_helpful_points_sum	90,545	441.53	1,563.81	-	35,323.37
aq_helpful_count_sum	90,545	107.98	344.67	-	8,292.00
grandtotal_points_sum	90,545	1,170.69	3,266.60	-183.40	76,211.01
grandtotal_count_sum	90,545	383.26	990.27	-	19,857.00
entrymonth	90,545	565.90	7.64	551.00	581.00
lifetime	90,545	8.55	6.40	1.00	31.00
amountmember	90,545	5,107.26	2,278.83	38.00	7,877.00
ranking_by_system	90,545	2.82	3.16	1.00	20.00

Table 4: Summary of generated variables

4.4 Trimming of the dataset

4.4.1 Expert limitations

As mentioned before, the users of the dataset, the so-called *experts*, can have different *roles*. Some of them function as *moderators* (n=44), while the vast majority (n=7,833) are normal *users*. As we focus our research on the average (normal) user of a knowledge network we will drop the moderator data and keep only the user data for our hypotheses tests. In Figure 26 we see that the *gmod* and *mod* data of friends shows a significantly different distribution along their membership duration to that of the average *user*. We see this alternating data structure for all variables measured for moderators.[34] In addition, the *moderator* data shows significant variations over time that cannot be explained by the data itself and seems to be the result of changes in the platform rules.

Consequently, we decided to exclude the moderator data (*gmod* and *mod*) from the dataset, to focus on the average user of the network.

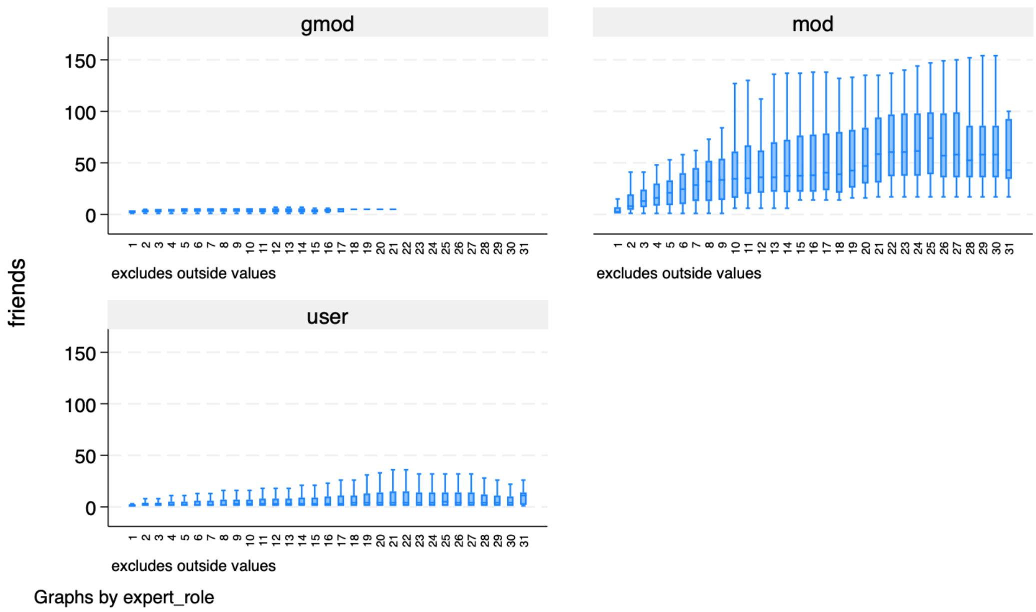


Figure 26: Boxplots by lifetime over role

[34] See Appendix 14 and Appendix 15 for variables friendsfriends and aq_helpful_points_sum

4.4.2 Time limitations

While describing the dataset based on the total available set of data, we have discussed the need to cut the data from the time perspective. The knowledge network went life on the 12th of January in 2006 and our dataset lasts until (and including) the 6th June 2008. During this period both major and minor changes to the point system were implemented. Luckily, the status points we examine during our analysis – the points for answers of good quality – were not affected by these changes, so we do not see the need to clean the data due to concerns that this would influence our dependent variable. Nevertheless, there are other reasons to limit the data. The go-live of the dataset and the recording time differ, so we need to start our analysis from the first full month of the dataset going public as a knowledge network platform, i.e. in February 2006. As June 2008 data did not cover a full month, we will also exclude that month.

As a result, we have observations for 28 months instead of 31 months, reflecting the go live time in point of the dataset and the need for comparable observations of full months. In addition to that, we had to exclude the number of experts that either do not have normal user status or have entered the dataset before or after the observation period from the dataset. This reduces the dataset to a residual panel of 7,514 users.

5 Empirical Results

5.1 General tests for the reduced dataset

All the following tests and regression estimations will run based on the reduced dataset consisting of 7.514 individuals, 28 month and a total of 75.183 observations. The panel is unbalanced, observations are not available for all points in time for each member of the network, as not all members are part of the network in all months. Thus, everyone has at least one observation and meets the criteria of being connected to at least one other network member during its observation period. The analyzed panel meets the criteria of being both, long and wide, as it holds data of a huge number of subjects over large number of points in time. Spoken form a dataset view, the panel is shaped in a long format, as the experts are the leading variable, and each data row represents a point in time. We are aware that technically not all experts of the network are connected to each other via a connected path. We do not see this as a disadvantage, as each possible sub-group has the same restrictions as the rest of the network and can show each analyzed variable in the same variety. The dataset in the monthly format is easier to handle than the original dataset, representing 29.622 points in time. The additional advantage of the compressed dataset is the better emphasized view on the data and has a higher relevance for real life situations.

5.1.1 Testing for fixed and random effects

Our hypotheses are based on the influence of network circumstances, embeddedness, duration of membership and career level, on the performance and activity of the actors. Even though we are convinced that these variables influence the dependent variables we are aware that the characteristics of the dataset's members will explain most of the variance of the estimated regressions. We see the individual motivation, the education and the intelligence, among other factors, as the main influence factor of success in the network. These factors are not likely to change during our period of observation. As a consequence we have to use a methodology to estimate our regressions that can cope with time-constant, unobserved effects (Wooldridge, 2010, pp. 481–483). We assume that the time-constant effect does also correlate with the explanatory variable and are convinced that we cannot meet the restrictions of a random-effects model (Wooldridge, 2010, pp. 489–491) but see the restrictions of a fixed-effects model as met (Wooldridge, 2010, p. 481). The fixed-effects model allows *for arbitrary dependence between the unobserved effect [...] and the observed explanatory variables* (Wooldridge, 2010, p. 286). Therefore, we will apply a fixed-effects model that is able to estimate the explanatory value of the dependent variables if they correlate with the individual heterogeneity that stays constant over time.

Even though inter-panel relations might exist as special network constellations, as subgroups or cliques may occur, we have included these effects into the independent variables as the clustering coefficient and the number of friends or friendsfriends. In the end, we are controlling for an effect that is caused by the individual and not by the interaction of the different panels. When we look at the results of chapter 4, we can already see that we do not have an indication for clusters of minimal within-cluster variation as the data varies a lot, over time and over the lifetime of all panels.

We performed the Hausman test for all applied models[35] for all hypotheses (Table 5). For Hypothesis 1, 3 and 4 we see that the results support our rejection of the 0 hypothesis, while we see Prob > chi^2 values under 0.05. Even though the values do not meet the criteria to reject the 0 hypothesis (Prob > chi^2 = 0.0812), we do not see the criteria of a random effects model as met and will also use the fixed effects model for our estimations of hypothesis 2 and we can assume that the unique errors are correlated with the regressors.

[35] For our work we estimate xtreg models to substract the group effects out of the model. The areg models are added in Appendix 9 and Appendix 13.

Hypothesis	Model	Regression	Prob > *chi*^{2}
1	1	xtreg aq_helpful_points_sum_ln clusteringcoefficient clusteringcoefficientsqrd	0.0050
2	2	xtreg aq_helpful_count_sum_ln clusteringcoefficient clusteringcoefficientsqrd	0.0812
3	3	xtreg aq_helpful_points_sum_ln clusteringcoefficient clusteringcoefficientsqrd ranking_by_system	0.0000
	4	xtreg aq_helpful_points_sum_ln clusteringcoefficient c.clusteringcoefficientsqrd##c.ranking_by_system	0.0000
4	5	xtreg aq_helpful_points_sum_ln clusteringcoefficient clusteringcoefficientsqrd lifetime	0.0000
	6	xtreg aq_helpful_points_sum_ln clusteringcoefficient c.clusteringcoefficientsqrd##c.lifetime	0.0000

Table 5: Results Hausman tests

5.1.2 Testing for stationarity

To be able to use our regression models we had to find evidence that the panels do not follow a unit root and are stationary. Stationarity of our data would mean that the joint distribution of the panels would stay the same over the variation of time and that we do not look at more than one trend in the series. We have strong arguments to assume that our panels are stationary. First, the dataset is a well-shuffled sample that consists of panels that are relatively new joiners of the network and more experienced panels at the same time. The mean lifetime of the cut dataset is 7,6 month throughout the whole length of the dataset (28 month). In addition to that we assume that the dataset is more influenced by the characteristics of the single actor and thereby use a fixed effects model to estimate the explanation of the variance. We do also not find any plausible reasons for a unit root, as a general underlying trend, conditional careers (the panel members do not have to fulfill any prerequisites to gain points during their career) or a special dynamic in the point system that could allow members to gain more points in the future if the earned higher amounts in the present. Summing up we did not find any reason for which our dataset is not stationary. Nevertheless, we conducted a unit root test to reject the null hypothesis that our panels contain unit roots. The structure of our data does not permit to execute more than one unit root test that is offered by Stata17 (Breitung and Das, 2005; Choi, 2001; Hadri, 2000; Harris and Tzavalis, 1999; Im et al., 2003; Levin et al., 2002), the Fischer-type unit root test (Choi, 2001). This test is the only test that allows panels to be unbalanced and as we do not have observations for all points in time and panels, our dataset meets this criterion and fits the right number of observations. We executed the Fisher-type unit root test in different variations[36] and for up to 3 lags (see example for 0 lags and aq_helpful_points_sum in Figure 27). We found full support to reject the null-hypothesis with values under 0,000 for all p-values of the variables aq_helpful_points_sum, aq_helpful_count_sum, aq_helpful_points_sum_ln, helpful_count_sum_ln and clusteringscoefficient. The variable ranking_by_system shows the same P statistics with zero lags but cannot reject the null-hypothesis for 1-3 lags. For the variable lifetime we cannot reject the null hypothesis for all tests, so at least one of the panels is stationary. We can explain the results due to the structure of these variables. The lifetime variable will count upwards for each month of membership in the dataset, so the variable follows a continuous n+1 pattern. The variable ranking_by_system follows a similar pattern as the ranking will always grow from one level to the next, while the timespan between the promotion may differ in relation to the activity and success of the network member.

---

[36] Test was conducted for the variables: aq_helpful_points_sum, aq_helpful_count_sum, aq_helpful_points_sum_ln, helpful_count_sum_ln, lifetime, ranking_by_system and clusteringcoefficient.

```
Fisher-type unit-root test for aq_helpful_points_sum
Based on augmented Dickey-Fuller tests
--------------------------------------------------
Ho: All panels contain unit roots           Number of panels       =   7514
Ha: At least one panel is stationary        Avg. number of periods =  10.01

AR parameter: Panel-specific                Asymptotics: T -> Infinity
Panel means:  Included
Time trend:   Included
Drift term:   Not included                  ADF regressions: 0 lags
------------------------------------------------------------------------------
                                  Statistic      p-value
------------------------------------------------------------------------------
 Inverse chi-squared(13286)P       7.83e+04       0.0000
 Inverse normal            Z      -133.8538       0.0000
 Inverse logit t(17944)    L*     -324.2865       0.0000
 Modified inv. chi-squared Pm      398.9526       0.0000
------------------------------------------------------------------------------
 P statistic requires number of panels to be finite.
 Other statistics are suitable for finite or infinite number of panels.
------------------------------------------------------------------------------
```

Figure 27: Results of Fisher-type unit-root test

The Fischer-type unit root test allows us to control for Ho “all panels contain unit roots” and for the alternative Ha “at least one panel is stationary”. After performing the test we can state that not all panels follow a unit root for the variables (aq_helpful_points_sum, aq_helpful_count_sum, aq_helpful_points_sum_ln, helpful_count_sum_ln and clusteringscoefficient) and at least one panel is stationary for the variables lifetime and ranking_by_system.

5.1.3 Variable correlation

As a starting point for our analysis, we calculated a correlation matrix with the relevant set of variables we were planning to use for our model estimation. Our first two hypotheses target a correlation of both, the points, and the activities with the clustering coefficient. Even though we see only a slight correlation, both dependent variables correlate with the clustering coefficient. We can also find first indications on the effects of both additional variables we mention in our hypothesis, the membership duration of a network member (lifetime) and the career level he achieves (ranking_by_system).

Variables	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
(1) aq_helpful_poi~m	1.000							
(2) aq_helpful_cou~m	0.962	1.000						
(3) clusteringcoef~t	-0.101	-0.124	1.000					
(4) lifetime	0.157	0.155	0.127	1.000				
(5) ranking_by_sys~m	0.766	0.804	-0.184	0.148	1.000			
(6) friends	0.263	0.295	-0.123	0.153	0.356	1.000		
(7) friendsfriends	0.303	0.334	-0.082	0.179	0.344	0.925	1.000	
(8) possiblefriends	0.047	0.056	-0.044	0.044	0.087	0.823	0.700	1.000

Figure 28: Correlation matrix

Figure 28 shows the detailed values. Even though the values are highly correlated themselves, the points do not correlate as strong with the clustering coefficient as the activities. This might be an indication that the influence of the clustering coefficient is more directed toward the quantitative than the qualitative aspect of performance. Additionally, we can say that the number of friends an expert has, does correlate with the points and activities he earns in the network, with a slight favor on the quantitative aspects, the activities. In addition to that we see a relatively high correlation of the number of friends and the career level (*ranking_by_system*).

5.2 Hypothesis testing

5.2.1 Clustering coefficients as performance driver

5.2.1.1 Data analysis

Our first hypothesis is based on the assumption that the performance of an actor is influenced by the number of realized 2nd degree friendships in relation to the possible number of these friendships. In other words, if one individual does have two friends and these are not friends themselves, the clustering coefficient is 0. If they are friends themselves the clustering coefficient is 1.

H1: The performance of a network member is correlated with his clustering coefficient.

This clustering coefficient (see 4.2.3) is the network metric reflecting this relationship and can take on a value from 0 to 1, while 0 means that none and 1 mean that all possible friendships are made. We will start our analysis with a scatterplot based on the reduced dataset (Chapter 0).

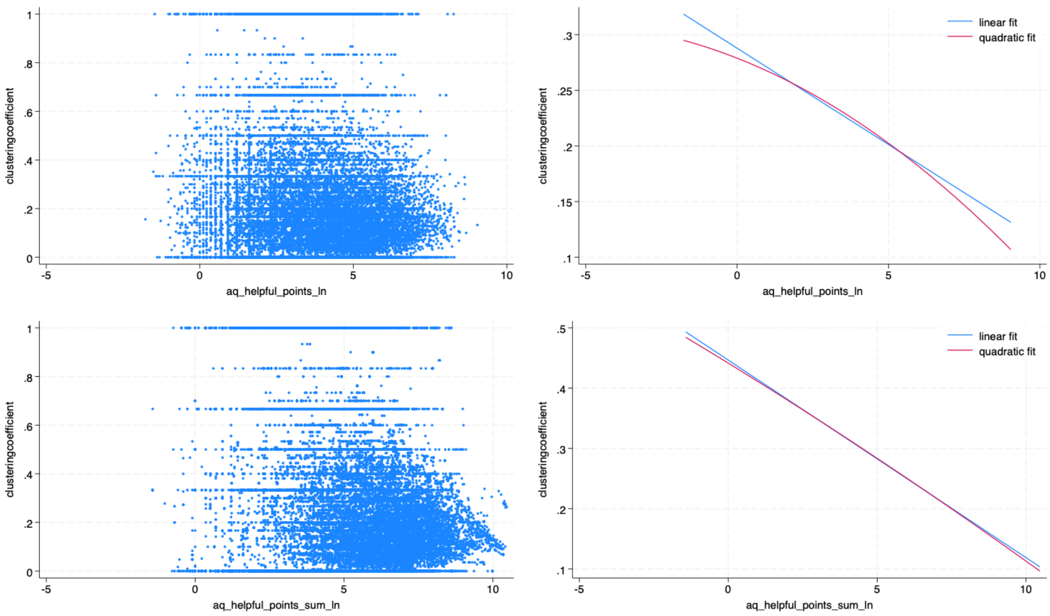


Figure 29: Plots and fitted models aq_helpful_points and aq_helpful_-points_sum

The explanatory performance variable of the first hypothesis is the quantity of monthly points that is earned when questions are answered with content being categorized as helpful.[37]

Figure 29 shows two scatterplots with two fitted models for the natural logarithm of the monthly points and the natural logarithm of the aggregated sum of monthly points an expert earns during his participation in the monitored period. Independently of the aggregation level we see the same pattern for scatterplots and fitted models. The aggregated view of the points contains implicitly the information of the single point view;

[37] To correct for the right skewness of the data we used a logarithm transformation for the explanatory variables.

the observations are only stretched over a greater number of total values. The linear models imply a negative relation of points and clustering coefficient equal to the exponential models, which are slightly curved towards the x-axis.

Before we estimate the regressions for our first hypothesis, can state that the negative correlation of the clustering coefficient and the qualitative performance of an actor seems to develop from a higher level for lower number of distributed points to a lower level for higher number of distributed points. The first assumption of low clustering coefficient being the driver of high status points seems not to be correct, as high clustering coefficients seem to drive the sum of points at early career stages of the network members, too.

5.2.1.2 Empirical model, results and robustness check

Our empirical model for the first hypothesis will be estimated as:

(1) $$\log(\text{aq\_helpful\_points\_sum}) = \beta_{it} + clusteringcoefficient_{it} + clusteringcoefficientsqrd_{it} + u_{it} + \varepsilon_{it}$$

We will estimate a fixed effects model, using the xtreg model fit. This model will exclude group effects of the panels which can be included by using the areg model. The results of the xtreg model are provided in the empirical models and the results of the areg models are provided in Appendix 9. The intercept will be estimated as β_{it} for each expert and week, the effect of the clusteringcoefficent will be estimated for each expert and week. We do also include the user specific fixed effects as u_{it} for each expert and week and the error term as ε_{it}well.

We estimate the model for our first hypothesis as shown in Figure 30.

VARIABLES	(1) aq_helpful_points_sum_ln
clusteringcoefficient	1.993***
	(0.0853)
clusteringcoefficientsqrd	-3.291***
	(0.0832)
Constant	5.045***
	(0.0123)
Observations	42,197
Number of expertid_continous	4,318
R-squared	0.070

Standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Figure 30: Panel regressions hypothesis 1

We find full support for our first hypothesis, as the model is highly significant with p-values <=0.01.

The estimated regression for our first hypothesis is:

(1) $$\log(\text{aq\_helpful\_points\_sum}) = 5.045_{it} + 1.993\ clusteringcoefficient_{it} - 3.291\ clusteringcoefficientsqrd_{it} + u_{it} + \varepsilon_{it}$$

5.2.2 Clustering coefficient as activity driver

5.2.2.1 Data analysis

We defined the points an expert collects during his membership in the network as a rather qualitative and the number of events that results in points as a rather quantitative measure of network performance. We are aware that there is a high correlation of 0,96 for both measures (see Figure 28) but still want to estimate a model to be able to compare both outcomes in relation to the clustering coefficient. So we formulated our second hypothesis as:

> *H2: The activity level of a network member is correlated with his clustering coefficient.*

If we plot the observations for both, the monthly and aggregated activities and let Stata estimate two basic models analogous to the hypothesis 1 we get the results shown in Figure 31.

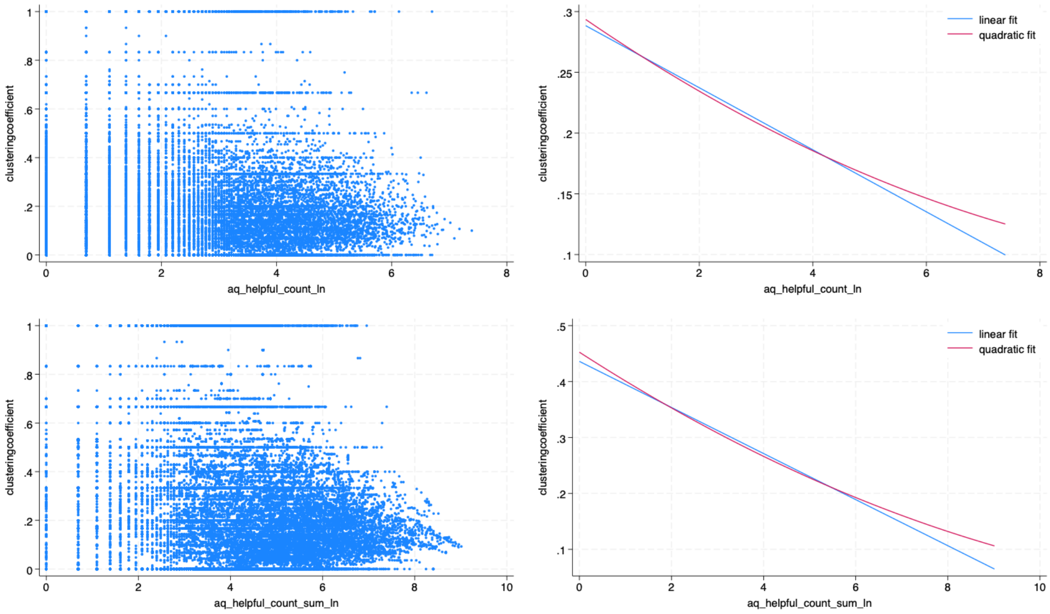


Figure 31: Plots and fitted models activities and activities_sum

According to the points perspective, the linear models imply a negative relation of activities and the clustering coefficient while the exponential models imply a slightly u-shaped relation.

For the aggregated view of the *aq_helpful_count_sum_ln* we see the clustering coefficient rather converging towards 0 than towards 0,10 as in the *aq_helpful_count_ln* model. Analog to our findings concerning the points we can state that we cannot generally say that the activities of an actor rise with a falling clustering coefficient, we need to further investigate the relation.

5.2.2.2 Empirical model, results and robustness check

Our empirical model for the second hypothesis will be estimated as:

(2) $\log(\text{aq\_helpful\_count\_sum}) = \beta_{it} + clusteringcoefficient_{it} - clusteringcoefficientsqrd_{it} + u_{it} + \varepsilon_{it}$

We will estimate a fixed effects model, using the xtreg model fit as shown in Figure 32. This model will exclude group effects of the panels which can be included by using the areg model. The results of the xtreg model are provided in the empirical models and the results of the areg models are provided in Appendix 10. The intercept will be estimated as β_{it} for each expert and week, the effect of the clusteringcoefficent will be estimated for each expert and week. We do also include the user specific fixed effects as u_{it} for each expert and week and the error term as ε_{it}well.

VARIABLES	(2) aq_helpful_count_sum_ln
clusteringcoefficient	1.971***
	(0.0831)
clusteringcoefficientsqrd	-3.330***
	(0.0811)
Constant	3.774***
	(0.0120)
Observations	42,197
Number of expertid_continous	4,318
R-squared	0.077

Standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Figure 32: Panel regressions hypothesis 2

According to the first hypothesis we find full support for our second hypothesis, as the model is highly significant with p-values <=0.01.

The estimated regression for our second hypothesis is:

$$(2)\ \log(\text{aq\_helpful\_count\_sum}) = 3.774_{it} + 1.971\ clusteringcoefficient_{it} - 3.330\ clusteringcoefficientsqrd_{it} + u_{it} + \varepsilon_{it}$$

5.2.3 Hierarchy as an indicator for strategy selection

5.2.3.1 Data analysis

In chapter 3.5 we argued that the career level of an actor is likely to interact with the direction of action of the clustering coefficient. We see the career level of a network member interact with the clustering coefficient, meaning that the hierarchical level of an actor influences the direction or intensity of action of the clustering coefficient. When we visualize the data, using the clustering coefficient and the ranking of the experts we can see the non-linear character of the relation (Figure 33). According to Gargiulo, Ertug and Galunic, we assume an influence of rank on the effect of density on the performance of knowledge workers (Gargiulo et al., 2009). As a consequence, we are convinced that the hierarchical situation of the member of a knowledge network influences the direction of action of the clustering coefficient. Our third hypothesis is:

> *H3: The clustering coefficient correlates negatively with the career level of the network actor.*

Lower hierarchical ranks have a higher need to be interwoven within the network they operate in. This has several plausible reasons. In the beginning of a career the experts need to understand the working mode of the network and will try to connect to other people to find questions that can be answered with the intention to receive status points in return. To maximize possible contact points and thereby access to information, a high clustering coefficient is favorable. Higher career levels on the other hand do not depend on this information flow as their authority raises it. In addition to this, higher ranks do have the incentive to protect and broker their knowledge, as, in knowledge networks, it is their personal asset. We thereby expect the clustering coefficient to be higher for lower ranks and lower for higher ranks.

The visualization of the data supports our hypothesis, showing a general trend of declining clustering coefficients for higher ranks. Both, linear and exponential models show the same trend. The exponential model shows a turning trend for the highest ranks, however.

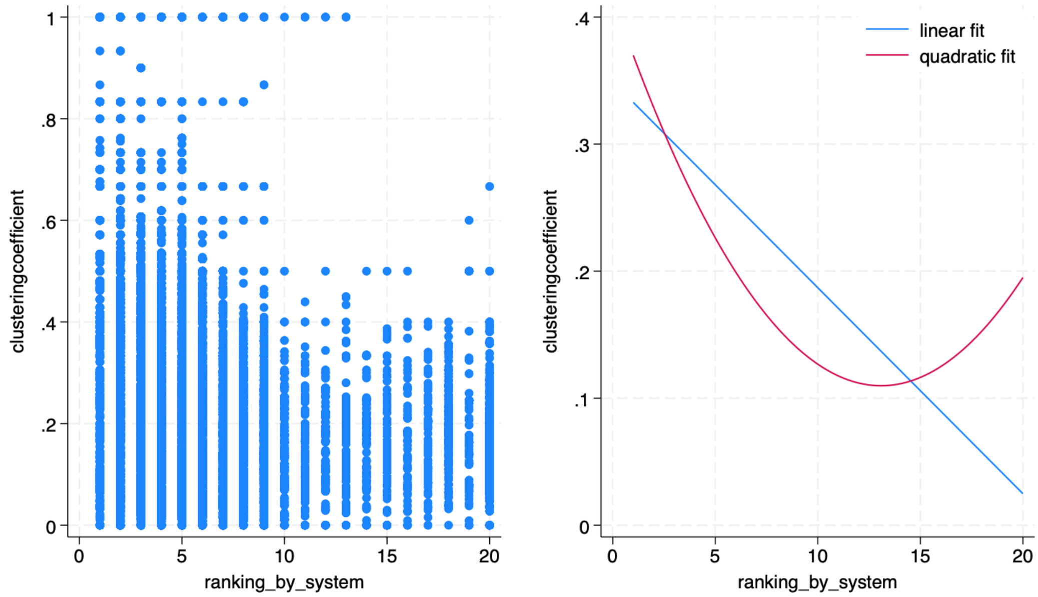


Figure 33: Plots and fitted models ranking_by_system

5.2.3.2 Empirical model, results and robustness check

Our empirical models for the third hypothesis will be estimated as:

(3) log (aq_helpful_points_sum) = $\beta_{it} + clusteringcoefficient_{it} + clusteringcoefficientsqrd_{it} + ranking\_by\_system_{it} + u_{it} + \varepsilon_{it}$

and

(4) log (aq_helpful_points_sum) = $\beta_{it} + clusteringcoefficient_{it} + clusteringcoefficientsqrd_{it} + ranking\_by\_system_{it} + ranking\_by\_system_{it}\ clusteringcoefficient_{it} + ranking\_by\_system_{it}\ clusteringcoefficientsqrd_{it} + u_{it} + \varepsilon_{it}$

We will estimate fixed effects models, using the xtreg model fit. These models will exclude group effects of the panels which can be included by using the areg model. The results of the xtreg model are provided in the empirical models and the results of the areg models are provided in Appendix 11. The intercepts will be estimated as β_{0it} for each expert and week, the effects of the clustering coefficient and the ranking will be estimated for each expert and week. We do also include the user specific fixed effects as u_{it} for each expert and week and the error term as ε_{it} well.

VARIABLES	(3) aq_helpful_points_sum_ln	(4) aq_helpful_points_sum_ln
clusteringcoefficient	1.938***	2.738***
	(0.0655)	(0.0887)
ranking_by_system	0.246***	0.276***
	(0.00151)	(0.00276)
c.clusteringcoefficient#c.ranking_by_system		-0.337***
		(0.0198)
clusteringcoefficientsqrd	-2.528***	-4.058***
	(0.0641)	(0.0917)
o.ranking_by_system		-
c.clusteringcoefficientsqrd#c.ranking_by_system		0.644***
		(0.0241)
Constant	3.971***	3.919***
	(0.0115)	(0.0132)
Observations	42,197	42,197
R-squared	0.451	0.464
Number of expertid_continous	4,318	4,318

Standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Figure 34: Panel regression hypothesis 3

Our regression models consist of two different models, a basic model (3) showing the influence of the clustering coefficient and the ranking of the experts on the distribution of the points for helpful answers and a moderator model (4) showing the influence of a moderator of the ranking and the clustering coefficient in addition (Figure 34).

The estimated regressions for our third hypothesis are:

$$(3) \quad \log(\text{aq\_helpful\_points\_sum}) = 3.971_{it} + 1.938\, clusteringcoefficient_{it} - 2.528\, clusteringcoefficientsqrd_{it} + 0.246\, ranking\_by\_system_{it} + u_{it} + \varepsilon_{it}$$

and

$$(4) \quad \log(\text{aq\_helpful\_points\_sum}) = 3.919_{it} + 2.738\, clusteringcoefficient_{it} - 4.058\, clusteringcoefficientsqrd_{it} + 0.276\, ranking_{by_{system}}{}_{it} - 0.337\, ranking\_by\_system_{it}\, clusteringcoefficient_{it} + 0.644\, ranking\_by\_system_{it}\, clusteringcoefficientsqrd_{it} + u_{it} + \varepsilon_{it}$$

According to the first and second hypothesis we find full support for our third and fourth hypothesis, as the models are highly significant with p-values <= 0.01.

5.2.4 Membership duration as moderator of the clustering coefficient

5.2.4.1 Data analysis

In chapter 3.5 we argued that the membership duration of an actor is likely to interact with the direction of action of the clustering coefficient, meaning that the time an actor spends for his career, influences the direction or intensity of action of the clustering coefficient. When we visualize the data, using the product of the clustering coefficient and the lifetime, we can see the non-linear character of the relation. Gargiulo, Ertug and Galunic assumed an influence of rank on the effect of density on the performance of knowledge workers (Gargiulo et al., 2009). We also assumed that not only the rank but also the membership durance has an influence on the performance of network members. Our fourth hypothesis thereby is:

> *H4: The clustering coefficient correlates negatively with the timespan of network membership*

During the first month of membership in the dataset, we assume that the experts try to find their orientation within the network and simultaneously try to build up reputation by answering questions. We are convinced that experts that have a higher clustering coefficient in the beginning of the career are more successful and earn more status points, for answers with helpful quality, then those who don't. After having built up a network and some reputation, it is likely that the requirements for being successful in the network, change. A high clustering coefficient has the downside that information drain is possible and that the own position can be endangered. After a certain timespan of membership within the network we assume that experts become more successful when they protect their knowledge and control information drain. This again implies that the clustering coefficient should fall.

The data visualization supports our hypothesis as the clustering coefficient seems to raise at first hand, to fall again after a certain period (Figure 35).

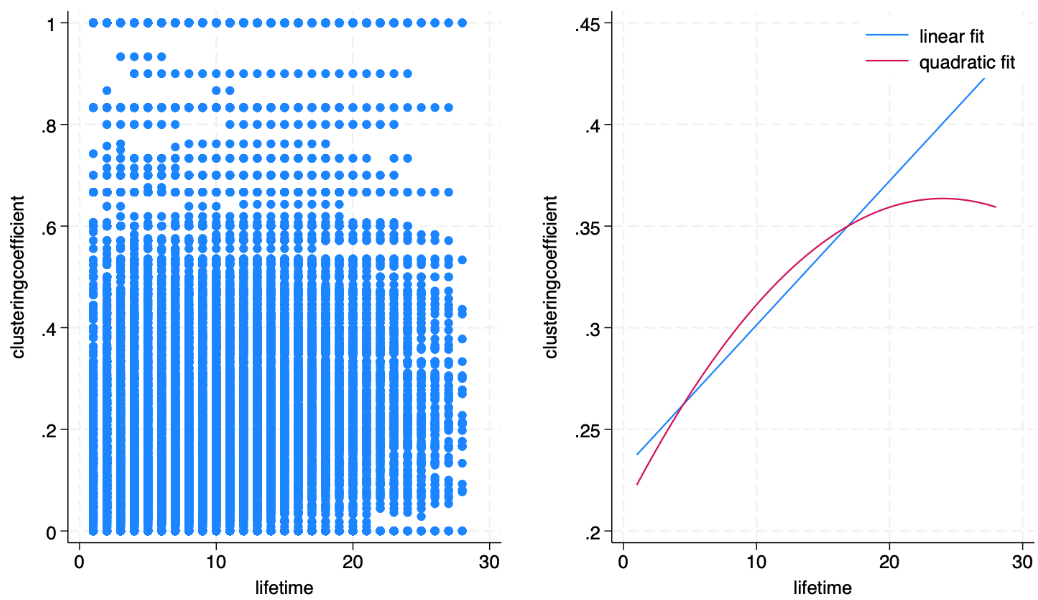


Figure 35: Plots and fitted models lifetime

5.2.4.2 Empirical model, results and robustness check

Alike before, we built up the models starting with the basic model and the predictor clustering coefficient and lifetime followed by lifetime as a moderator of the clustering coefficient in a second step.

Our empirical models for the fourth hypothesis will be estimated as:

(5) $\log(\text{aq\_helpful\_points\_sum}) = \beta_{it} + clusteringcoefficient_{it} + clusteringcoefficientsqrd_{it} + lifetime_{it} + u_{it} + \varepsilon_{it}$

and

(6) $\log(\text{aq\_helpful\_points\_sum}) = \beta_{it} + clusteringcoefficient_{it} + clusteringcoefficientsqrd_{it} + lifetime_{it} + lifetime_{it}\ clusteringcoefficient_{it} + lifetime_{it}\ clusteringcoefficientsqrd_{it} + u_{it} + \varepsilon_{it}$

We will estimate fixed effects models, using the xtreg model fit. These models will exclude the effects of the panels which can be included by using the areg model. The results of the xtreg model are provided in the empirical models and the results of the areg models are provided in Appendix 12. The intercepts will be estimated as β_{it} for each expert and week, the effects of the clusteringcoefficent and the lifetime will be estimated for each expert and week. We do also include the user specific fixed effects as u_{it} for each expert and week and the error term as ε_{it}well.

Our regression models consist of two different models, a basic model (5) showing the influence of the clustering coefficient and the lifetime of the experts on the distribution of the points for helpful answers and a moderator model (6) showing the influence of a moderator of the lifetime and the clustering coefficient in addition (Figure 36).

VARIABLES	(5) aq_helpful_points_sum_ln	(6) aq_helpful_points_sum_ln
clusteringcoefficient	0.814***	1.684***
	(0.0756)	(0.0820)
lifetime	0.0633***	0.0954***
	(0.000593)	(0.00120)
c.clusteringcoefficient#c.lifetime		-0.150***
		(0.00642)
clusteringcoefficientsqrd	-1.828***	-2.446***
	(0.0742)	(0.0806)
o.lifetime		-
c.clusteringcoefficientsqrd#c.lifetime		0.0911***
		(0.00630)
Constant	4.602***	4.445***
	(0.0116)	(0.0124)
Observations	42,197	42,197
R-squared	0.285	0.306
Number of expertid_continous	4,318	4,318

Standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Figure 36: Panel regression hypothesis 4

The estimated regressions for our fourth hypothesis are:

(5) $\log(\text{aq\_helpful\_points\_sum}) = 4.602_{it} + 0.814\, clusteringcoefficient_{it} - 1.828\, clusteringcoefficientsqrd_{it} + 0.0633\, lifetime_{it} + u_{it} + \varepsilon_{it}$

and

(6) $\log(\text{aq\_helpful\_points\_sum}) = 4.445_{it} + 1.684\, clusteringcoefficient_{it} - 2.446\, clusteringcoefficientsqrd_{it} + 0.0954\, lifetime_{it} - 0.150\, lifetime_{it}\, clusteringcoefficient_{it} + 0.0911\, lifetime_{it}\, clusteringcoefficientsqrd_{it} + u_{it} + \varepsilon_{it}$

Model number 5 shows a highly significant influence of lifetime on the status points. Model number 6, with lifetime as a moderator of the clustering coefficient has an even higher R^2 than model number 5, with same level of significance and equal standard errors. Consequently, we find full support for our fifth and sixth hypothesis, models being highly significant with p-values <=0.01.

6 Discussion of Results

6.1 Reflections of the regressions

6.1.1 First hypothesis

In addition to the support of the hypothesis we see that the effect of the clustering coefficient on the points distributed for helpful answers is slightly negative. The higher the clustering coefficient, the lower the distributed points. Respecting only this hypothesis and model we could deduce, that our results rather support Burt's point of view of structural holes being a source of social capital, rather than dense networks as they are suggested by Coleman. But when we look at the regression plot (Figure 37) we do also recognize that we can make a second statement, as the distributed points rise from lowest levels of clustering coefficients to fall after a climax at a clustering coefficient of 0.25. We therefore have a first indication for the clustering coefficient not only being a different driver in different situation, but also reaching a turning point. As all panels at all points in lifetime, career level and time of the dataset are mixed in this perspective of the first hypothesis, we need to look into the plots of the following hypothesis to gain a more differentiated view of this observation.

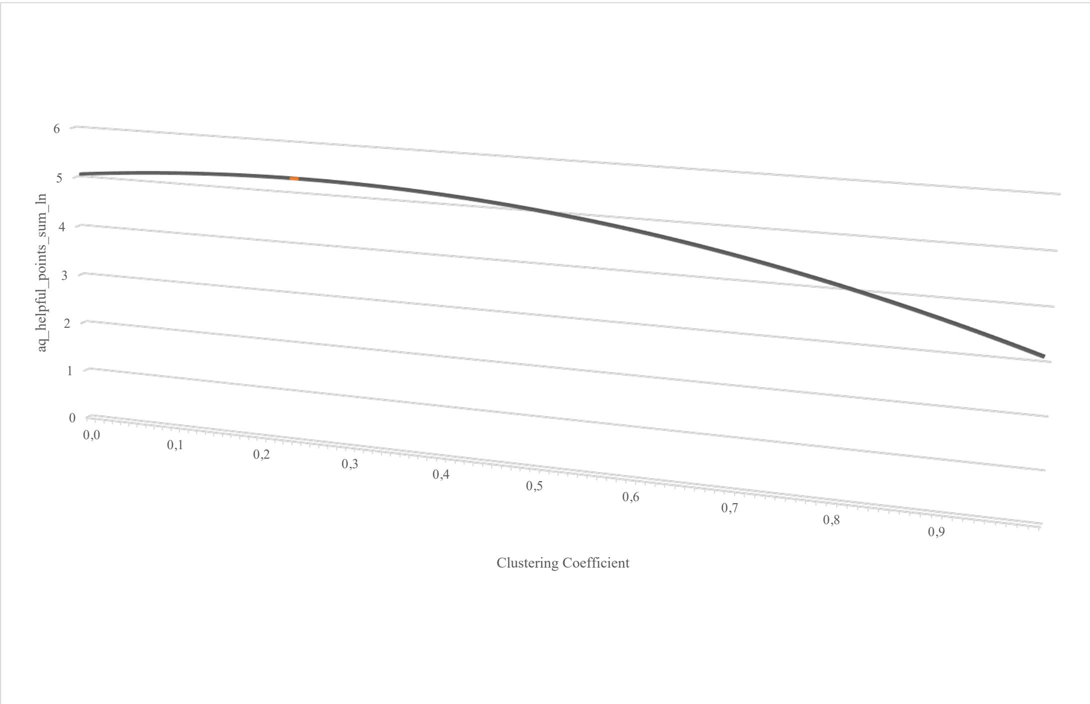


Figure 37: Regression plot hypothesis 1 (model 1)

6.1.2 Second hypothesis

In addition to the support of the hypothesis we see that the effect of the clustering coefficient on the activity represented by the number of events that cause the distribution of points for helpful answers, is slightly negative. The higher the clustering coefficient, the lower the number of events for distributed points. Respecting only this hypothesis and model we could deduce, that our results also rather support Burt's point of view of structural holes being a source of social capital, than dense networks as they are suggested by Coleman. Again, graphically (Figure 38) a clearer view becomes obvious, as activities seem to be driven by lower clustering coefficients before they rise to a climax at 0.3, to fall under the level of the lowest clustering coefficients afterwards. When we add our observation to the preceding one, we can state that the clustering coefficient seems to have not only a changing effect on the quality (the amount of points distributed), but also on the activity level (the number of events when points were distributed) of the network members.

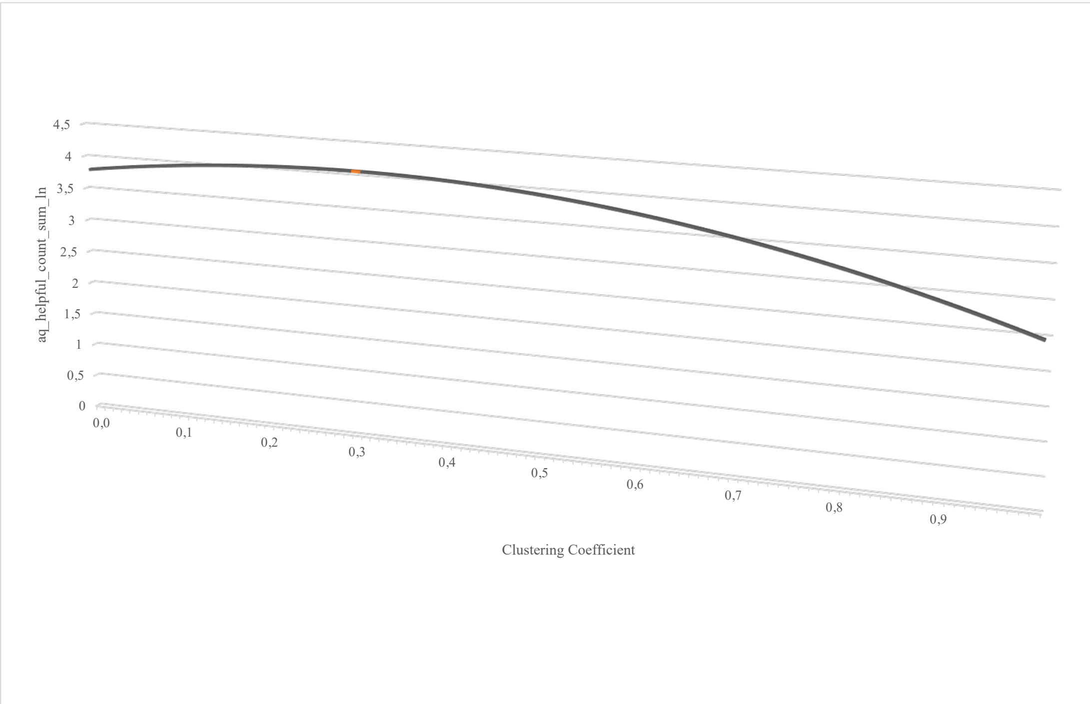


Figure 38: Regression plot hypothesis 2 (model 2)

6.1.3 Third hypothesis

As the ranking correlates strongly with the number of distributed status points, we can expect a higher R-squared for model 3 and 4. Model 4, containing the moderator term is an even better fit compared to the normal model. Therefore, we can state that the clustering coefficient correlates with the career level of the experts of our network, and we find support for the hypothesis, that it is influence by it in terms of a moderator variable. The direction of action seems to be influenced by the career level of the experts. The plots (Figure 39) show different pictures while model 3 shows a very similar development as the plots of the previous models, with lower outputs for lower clustering coefficients, coming to a climax at a clustering coefficient of 0.38 and falling again below entry level. We see parallel effects for growing career levels only on higher levels of points.

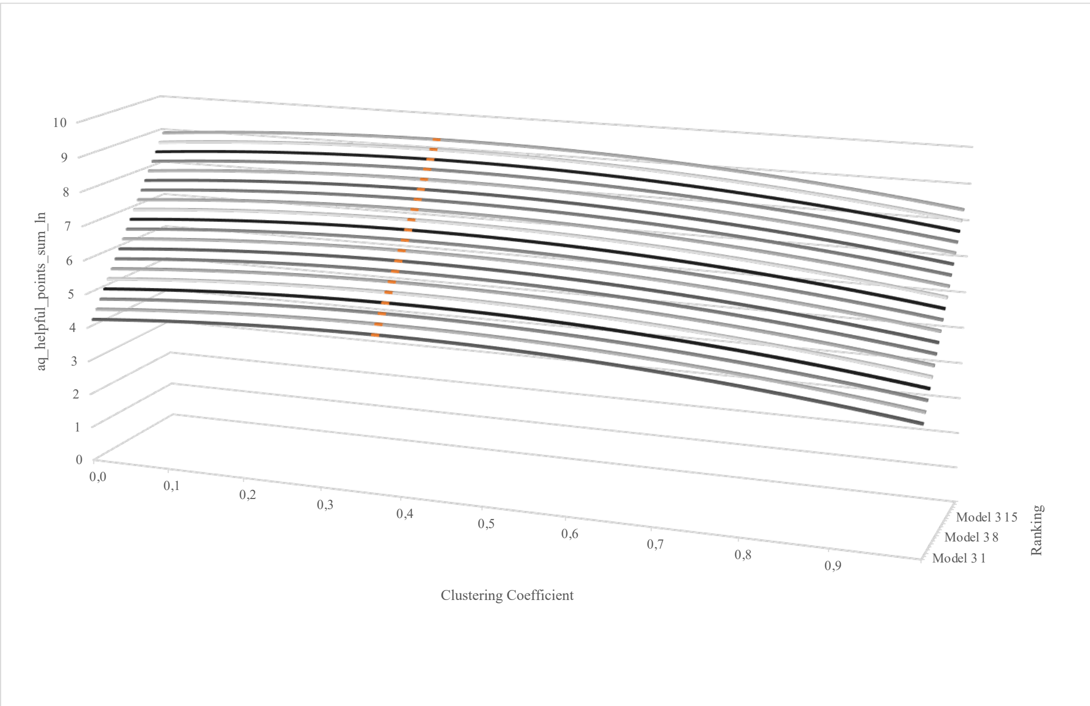


Figure 39: Regression plot hypothesis 3 (model 3)

For model 4, we see these effects also for the first five career levels of the interaction model (highest points for clustering coefficients at 0.35, 0.37, 0.41, 0.47 and 0.63) but a completely different development for higher career levels (Figure 40). To be precise, the estimations start on moderate levels of the clustering coefficient and fall to a minimum to rise again to a climax at a clustering coefficient of 1. The career level though is the only variable that underlies a change during the considered dataset. We therefore cannot say, if this effect is caused by the change in the career system or if different career levels benefit in different ways of higher and lower clustering coefficients. We can still support both hypotheses as the effect of the clustering coefficient is neither linear, in a way that it is at first favorable and later unfavorable a vice versa, nor not influenced by the career level. Even if we cannot state to have isolated the effect, we can confirm a moderating effect of the career level on the direction of action of the clustering coefficient. We could also state, that for the given dataset, there could be a tipping point for a model that is beneficial when building up, followed by lowering clustering coefficients, followed by a model that is beneficial for low and high, but not for medium clustering coefficients.

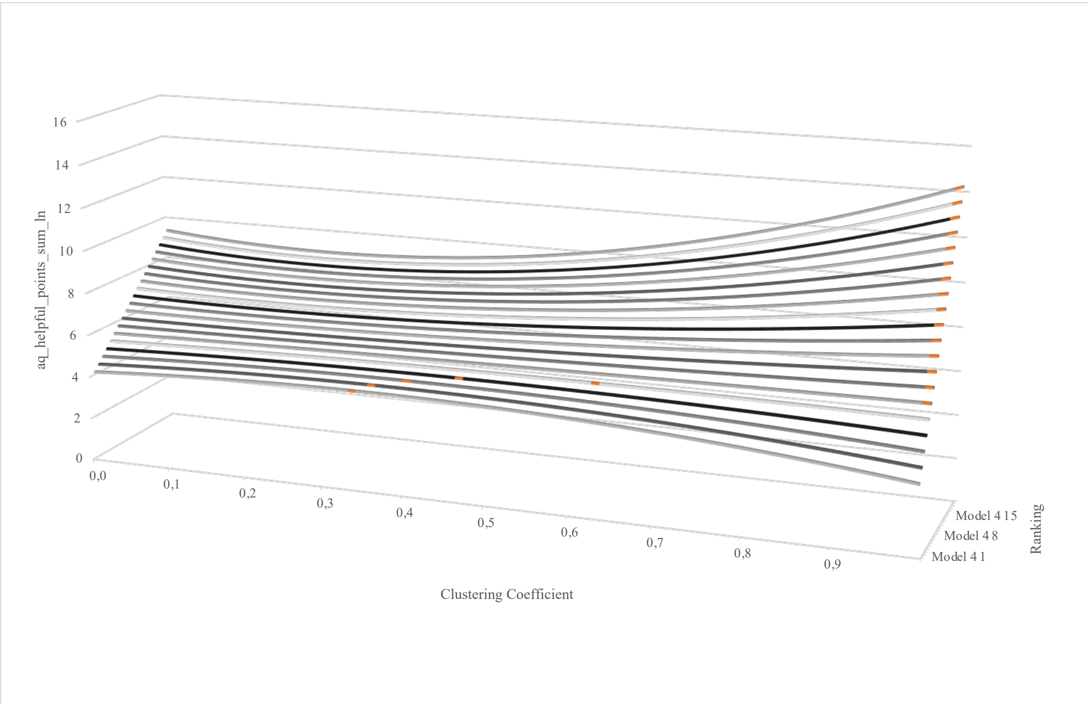


Figure 40: Regression plot hypothesis 3 (model 4)

6.1.4 Fourth hypothesis

As the lifetime is only weakly correlated with the distributed points, we expected a lower R^2 than for model 3 and 4. In the xtreg models we can see a rise auf the R^2 when we apply the moderator model. We can state that we found strong evidence for a moderator effect of the membership duration on the mode of action of the clustering coefficient. The regression plot (Figure 41) for model 5 shows a prediction of the distributed points that is comparable to the one for model 1 to 3. The clustering coefficients drive the points to a climax of 0.22. For rising levels of lifetime, we see this effect identical on higher levels of points.

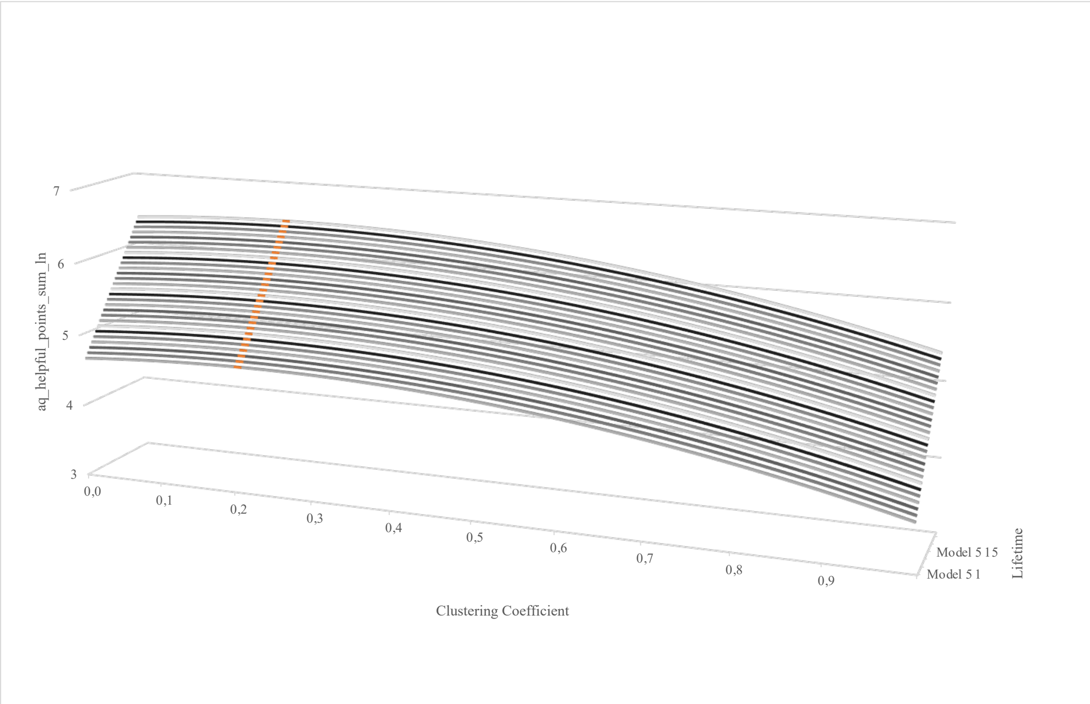


Figure 41: Regression plot hypothesis 4 (model 5)

Model 6, the interaction model of model 5, shows different effects for different levels of lifetime (Figure 42). While the first ten models show a falling climax from 0.33 to 0.06, the remaining models fall from lowest to highest clustering coefficients. Again we can state, that we can also graphically identify a moderating effect of lifetime on the direction of action of the clustering coefficient. While younger network members with lower membership durance still benefit more from higher clustering coefficients in a way that those with longer durance do not. After ten month we could state that there is a tipping point towards lower clustering coefficients always being more beneficial than higher ones.

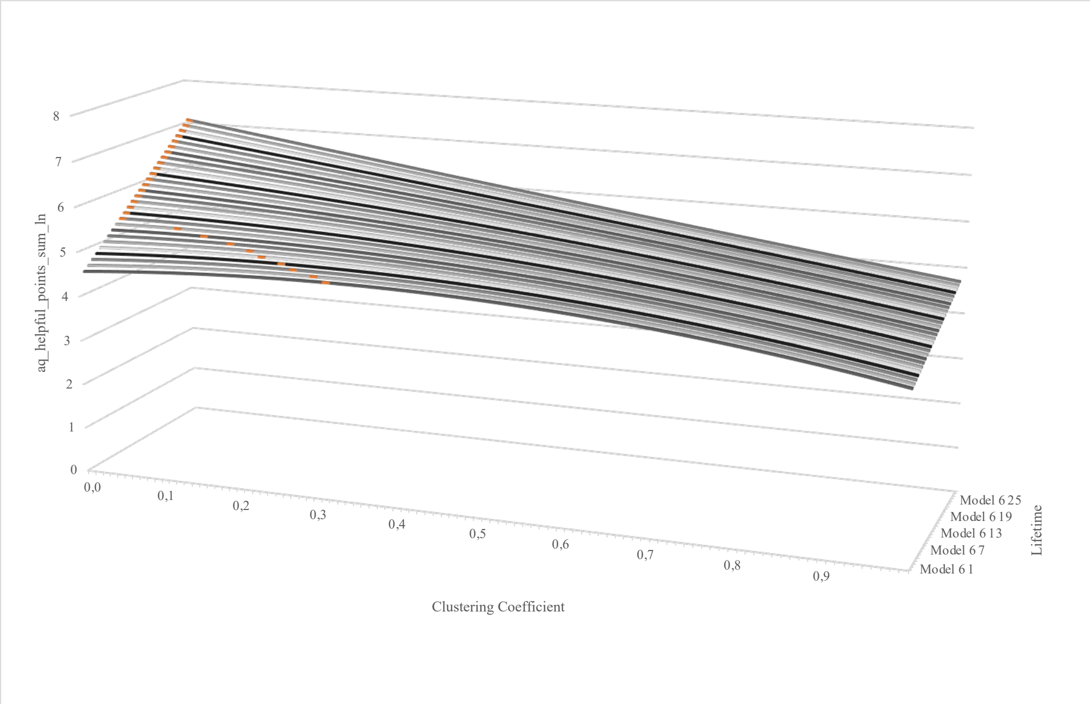


Figure 42: Regression plot hypothesis 4 (model 6)

6.2 Relevance of the results

When we look at our results displayed in Figure 43, we can say that we found significant proof for the influence of network density on the performance and activity of network members, with low standard errors. Hypothesis 1 & 2 can thereby be fully supported. Both, the quantity of activity and the qualitative output, the points for answers with good quality are thereby influenced by the clustering coefficient, following a quadratic development. The relation is negative for both models, the clustering coefficient falls with the rise of total points and activities, meaning experts with higher total points and activities have lower clustering coefficients than those with low values.

Proceeding to Hypothesis 3, we integrate the first model of the first hypothesis, the squared clustering coefficient, with the ranking of an expert as additional independent variable. We estimate two models, a simple additive one for the ranking and a moderator model that adds a multiplicative variable of clustering coefficient and ranking to the model. The moderator variable is added to examine a possible effect of the ranking on the mode of action of the clustering coefficient. The results are again highly significant with low standard errors and show a significant rise of the R-squared[38], compared to the models of hypothesis 1 & 2. The fourth model, which included the moderator variable is both, qualitative and from the perspective of the statistical model the preferred.

Hypothesis 4 finds also full support. The model consists of the model of Hypothesis 1 and two additional models according to Hypothesis 3. The additional independent variable of Hypothesis 4 is the lifetime, the number of months an expert is member of the dataset – meeting the criteria of one or more friends. The results are all highly significant with low standard errors and the model including the moderator variable is again preferred, from a qualitative and from a quantitative perspective.

We conclude that both, the ranking of an expert and the lifetime of an expert work as moderators on the mode of action of the clustering coefficient which itself has highly significant influence on the performance of members of knowledge networks.

We were also able to identify possible tipping points for all four hypotheses, which was one of the most interesting findings of our study, as we rather were expecting the effects rather being divided into constantly falling or rising effects. We now found first evidence for a turning point in network careers and membership durance that could lead network members into a more strategic and differentiated behavior, to reach their network goals.

[38] We are aware that the ranking of the experts correlates with the total number of points and thereby drive the R-squared to a larger extend.

	Hypothesis 1	Hypothesis 2	Hypothesis 3		Hypothesis 4	
	(1)	(2)	(3)	(4)	(5)	(6)
VARIABLES	aq_helpful_points_sum_ln	aq_helpful_count_sum_ln	aq_helpful_points_sum_ln	aq_helpful_points_sum_ln	aq_helpful_points_sum_ln	aq_helpful_points_sum_ln
clusteringcoefficient	1.993***	1.971***	1.938***	2.738***	0.814***	1.684***
	(0.0853)	(0.0831)	(0.0655)	(0.0887)	(0.0756)	(0.0820)
ranking_by_system			0.246***	0.276***		
			(0.00151)	(0.00276)		
c.clusteringcoefficient#c.ranking_by_system				-0.337***		
				(0.0198)		
clusteringcoefficientsqrd	-3.291***	-3.330***	-2.528***	-4.058***	-1.828***	-2.446***
	(0.0832)	(0.0811)	(0.0641)	(0.0917)	(0.0742)	(0.0806)
o.ranking_by_system				-		
c.clusteringcoefficientsqrd#c.ranking_by_system				0.644***		
				(0.0241)		
lifetime					0.0633***	0.0954***
					(0.000593)	(0.00120)
c.clusteringcoefficient#c.lifetime						-0.150***
						(0.00642)
o.lifetime						-
c.clusteringcoefficientsqrd#c.lifetime						0.0911***
						(0.00630)
Constant	5.045***	3.774***	3.971***	3.919***	4.602***	4.445***
	(0.0123)	(0.0120)	(0.0115)	(0.0132)	(0.0116)	(0.0124)
Observations	42,197	42,197	42,197	42,197	42,197	42,197
R-squared	0.070	0.077	0.451	0.464	0.285	0.306
Number of expertid_continous	4,318	4,318	4,318	4,318	4,318	4,318

Standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Figure 43: Panel regressions combined

6.3 Reflections from a theoretical perspective

Our hypotheses were based on the existing research in the field of social network science, dealing with knowledge-based networks. We were arguing that, to be favorable for the performance of their actors, the networks in our research field would have to be either dense networks or networks with local bridges. Even though we mostly agreed to the arguments made by both Coleman (Coleman, 1988) and Burt (Burt, 2000), we could not endorse either as right or wrong but instead postulated that it is also a network's contextual aspect that drives their results: indeed, whether the social network is a neighborhood community that is planning a barbecue area or whether it is managers competing for bonuses, makes a difference. Whilst we are aware that the authors explain the individual context, we were, missing a critical reflection on the relationship between network context and results. In addition to the context, we saw a situational aspect that was bound to the individual member of the network. Consequently, we set out to test membership duration and career level as moderators on the direction of action of the clustering coefficient, while trying to consider situational aspects as part of our analysis. Our contention is that being tightly knit into a network has a different effect at the beginning of a career or a membership lifetime than at later stages.

Attempts to integrate both concepts were made by Portes and Sensenbrenner, Adler and Kwon or Burt (Adler and Kwon, 2002; Ronald S Burt, 2001; Portes and Sensenbrenner, 1993). Whilst we acknowledge these as providing valuable starting points in this endeavor, we do believe that the concepts would benefit from greater specification, including the mechanism that links them. We particularly endorse two main arguments. First, Burt's, who stated that brokerage is a source of social capital that can be accessed by closure (Ronald S Burt, 2001, p. 31), and second the arguments made by Gargiulo, Ertug and Galunic, who found the career level and the seniority of an actor as highly significant influence factors, which means that either dense or bridged networks are favorable for the members of knowledge networks (Gargiulo et al., 2009).

Summarizing our results in this theoretical context we can say that we found evidence for each quoted concept but have a different view on their mode of action. From our perspective we do not see to have found evidence for or against either concept, nor for a future concept which is a combination of both. Rather, we found significant evidence that both concepts, holes and closure, cannot really be differentiated as being favorable or not, but have their advantages and shortcomings depending on the network context and situation of network members and their evolution within the network. It is important to mention that both concepts are relevant to the same person at different points in time and at different career levels. In the beginning of a career and at lower hierarchical levels it is more favorable to be tightly knit into at network than it is at later stages of one's career or participation in a network. In later stages it seems that holding bridge-like positions and being a gatekeeper for others is more crucial for success in knowledge networks than it is in earlier stages of careers and membership.

From this point we see our research not as a mean to judge one concept right or wrong, but as a possibility to integrate both concepts in a new way. Despite the empirical proof both concepts alone are very plausible. Tightly knit networks offer multiple chances for information exchange but also risk information churn. Local bridges have the risk of restricting information access but have better possibilities to be controlled. But both concepts are rarely found isolated in real world situations. Therefore, we tried to find a way to measure both concepts, by using the clustering coefficient, on an individual level. Besides the hypothesis tests we see this as a main success factor of this work.

6.4 Limitations of the work and need for further research

We are aware that our dataset and empirical evidence in general has limited validity. Our dataset is placed in a specific cultural context (Germany) and underlies many effects that are not respected in the analysis. The effect of changes in the career system or the point system might have influences on the performance of the network participants, these were not respected in our analysis. There might also be a variety of environmental and economic factors that might drive to data, to mention only a few. Nevertheless, we are convinced that the results of our research will be confirmed by further research, also in alternating contexts. The effects that we have measured were raised based on a very large dataset with nearly a billion of different datapoints and a population of nearly 8.000 individuals. Even if some of the conditions of the dataset varied over time, essential metrics, as average number of friends per expert or the average quantity of points per answer, stayed almost constant over time.

When we did our research on existing work dealing with the clustering coefficient and the different arguments for and again closure and local bridges, we were surprised that we did not find many studies, which tried to combine both concepts into one, while it obvious that each concept refers to alternating environments and contexts. In addition to that, we did not find research that brought these contexts together into a meta-concept, which allowed examining both concepts with one approach.

The major difficulty of our work was the combination of different fields of research and methods, while we applied methods of behavioral economics to the field of social network science and had to develop small software tools ourselves. These difficulties may be an entry barrier for many researchers to push social network science and especially the research of social capital in social networks further.

We encourage applying our concept of lifetime in the network and career levels as moderators of the clustering coefficient to different networks to find further proof for our hypothesis. The methodology used can be seen as a first blueprint to analyze other networks and challenge our results.

6.5 Practical implications of the research

Our research has both, theoretical and practical implications, as we were striving to integrate two opposing concepts and found also empirical evidence for our hypothesis by using real life data. As knowledge networks are a common type of social networks in our time, if not the most common one, we see a high relevance of our research for practical improvements of those networks. These improvements head for different purposes. First of all, the knowledge about career and membership duration related influences enables the platform operators to anticipate changes in behavior of the members and react proactively. These effects may also keep platform operators from only holding to the activity of network members as an indicator for their participation. In a first phase of network membership, the participants of knowledge networks are striving to interact and network themselves, to be more successful in the beginning of their career, before they turn to capitalize on their status and achievements and form local bridges.

Applied to the field of crime control, for larger criminal organization the communication structures of more mature and more influential members may change. Therefore, it might be possible to derive future influence and career development of key actors in criminal organization due to their current behavior. Consequently, to number of relevant observation targets can be reduced to a more relevant set of actors. For career or knowledge networks it might be the best suggestion to just suggest now contacts to mature networks members but to introduce them rather into fields they are not active in. Different career levels in network may also be an indicator for the need of different incentive structures in networks, especially knowledge networks. As in the beginning of a career there should be a focus on linking members, in later stages members should be incited to provide high value content. As social network platforms allow to implement changes only to a subset of actors, new approaches and changes in the algorithms can be seen as both, an experimental and a development approach.

7 List of Internet Ressources

http://www.analytictech.com/products.htm, quoted 2023/06/15

https://www.facebook.com/index.php?stype=lo&lh=Ac_Wvnoof6dmLSHP&aik=8ywpeQaVa5qedxgNU4rt3g, quoted 2023/06/15

http://www.linkedin.com/static?key=what_is_linkedin&trk=hb_what, quoted 2023/06/15

https://ocw.mit.edu/courses/14-15j-networks-spring-2018/ , quoted 2023/06/15

http://secondlife.com/whatis/?lang=en-US, quoted 2023/06/15

https://sites.google.com/site/ucinetsoftware/home, quoted 2023/06/15

http://vlado.fmf.uni-lj.si/pub/networks/pajek/, quoted 2023/06/15

http://en.wikipedia.org/wiki/Social_network_analysis_software, quoted 2023/06/15

http://en.wikipedia.org/wiki/Unix_time#Representing_the_number, quoted 2023/06/15

https://wowclassic.blizzard.com/de-de/, quoted 2023/06/15

https://www.xing.com, quoted 2023/06/15

8 List of References

Adamic, L.A., Glance, N., 2005. The Political Blogosphere and the 2004 U.S. Election: Divided They Blog. In: Proceedings of the 3rd International Workshop on Link Discovery, LinkKDD ’05. Association for Computing Machinery, New York, NY, USA, pp. 36–43.

Adler, P.S., Kwon, S.W., 2002. Social capital: Prospects for a new concept. Academy of management review 27, 17–40.

Ahuja, M.K., Galletta, D.F., Carley, K.M., 2003. Individual centrality and performance in virtual R&D groups: An empirical study. Management Science 49, 21–38.

Argote, L., Ingram, P., 2000. Knowledge Transfer: A Basis for Competitive Advantage in Firms. Organizational Behavior and Human Decision Processes 82, 150–169.

Backstrom, L., Kleinberg, J., 2014. Romantic Partnerships and the Dispersion of Social Ties: A Network Analysis of Relationship Status on Facebook. In: Proceedings of the 17th ACM Conference on Computer Supported Cooperative Work & Social Computing, CSCW ’14. Association for Computing Machinery, New York, NY, USA, pp. 831–841.

Baker, W.E., 1990. Market Networks and Corporate Behavior. American Journal of Sociology 96, 589–625.

Baron, R.M., Kenny, D.A., 1986. The moderator–mediator variable distinction in social psychological research: Conceptual, strategic, and statistical considerations. Journal of Personality and Social Psychology 51, 1173–1182.

Bearman, P.S., Moody, J., 2004. Suicide and friendships among American adolescents. American Journal of Public Health 94, 89–95.

Belliveau, M.A., Belliveau, M.A., suffix, C.A.O. suffix I., suffix, C.A.O. suffix I., Wade, J.B., Wade, J.B., 1996. Social Capital at the Top: Effects of Social Similarity and Status on CEO Compensation. The Academy of Management Journal 39, 1568–1593.

Blanchard, A., Horan, T., 1998. Virtual Communities and Social Capital. Social Science Computer Review 16, 293–307.

Borgatti, S.P., Cross, R., 2003. A Relational View of Information Seeking and Learning in Social Networks. Management Science 49, 432–445.

Borgatti, S.P., Foster, P.C., 2003. The Network Paradigm in Organizational Research: A Review and Typology. Journal of management 29, 991–1013.

Borgatti, S.P., Halgin, D.S., 2011. On Network Theory. Organ Sci 22, 1168–1181.

Borgatti, S.P., Mehra, A., Brass, D.J., Labianca, G., 2009. Network Analysis in the Social Sciences. Science 323, 892–895.

Bourdieu, P., 1986. The Forms of Capital. In: Richardson, J.G. (Ed.), Handbook of Theory and Research for the Sociology of Education. Greenwood Press, New York, pp. 241–258.

Bourdieu, P., Wacquant, L.J.D., Wacquant, L.J.D., 1992. An Invitation to Reflexive Sociology.

Boxman, E.A.W., De Graaf, P.M., Flap, H.D., 1991. The impact of social and human capital on the income attainment of Dutch managers. Social Networks 13, 51–73.

Braess, D., 1968. Über ein Paradoxon aus der Verkehrsplanung. Unternehmensforschung Operations Research - Recherche Opérationnelle 12, 258–268.

Brambor, T., Clark, W.R., Golder, M., 2006. Understanding Interaction Models: Improving Empirical Analyses. Political Analysis 14, 63–82.

Brass, D., Galaskiewicz, J., Greve, H., Tsai, W., 2004. Taking stock of networks and organizations: A multilevel perspective. The Academy of Management Journal 47, 795–817.

Brehm, J., Rahn, W., 1997. Individual-Level Evidence for the Causes and Consequences of Social Capital. American Journal of Political Science 41, 999–1023.

Breitung, J., Das, S., 2005. Panel unit root tests under cross-sectional dependence. Statistica Neerlandica 59, 414–433.

Burgun, K., 2012. Game Design Theory, A K Peters Limited.

Burt, R.S., 1987. Social Contagion and Innovation: Cohesion versus Structural Equivalence. American Journal of Sociology 92, 1287–1335.

Burt, R.S., 1992. Structural Holes : The Social Structure of Competition, Cambridge, Mass. : Harvard University Press.

Burt, R.S., 2000. The Network Structure of Social Capital. Research in Organizational Behavior 22, 345–423.

Burt, Ronald S, 2001. Structural Holes versus Network Closure as Social Capital, Social Capital: Theory and research.

Burt, Ronald S., 2001. Attachment, decay, and social network. Journal of Organizational Behavior 22, 619–643.

Burt, R.S., 2004. Structural Holes and Good Ideas. American Journal of Sociology 110, 349–399.

Burt, R.S., 2005. BROKERAGE & CLOSURE, Oxford University Press.

Bush, V., 1995. As We May Think (an article that appeared in The Atlantic Monthly in 1945 predicting the electronic revolution). The Journal of Electronic Publishing 1.

Carrington, P.J., 2011. Crime and social network analysis. In: Scott, J., Carrington, P.J. (Eds.), The SAGE Handbook of Social Network Analysis. SAGE Publications Ltd., London, pp. 236–255.

Castillo, R.P., 2019. Exploring the differential effects of social and individualistic gameplay motivations on bridging social capital for users of a massively multiplayer online game. Computers in Human Behavior 91, 263–270.

Castillo, R.P., 2020. "What's on Your Mind?" Examining the Influence of Facebook User Motivations, Usage Intensity, and Public Self-Disclosure on Perceived Social Capital. Sociological Focus 53, 271–292.

Castronova, E., 2006. On the Research Value of Large Games : Natural Experiments in Norrath and Camelot. Games and Culture 1, 163–186.

Castronova, E., Williams, D., Shen, C., Ratan, R., Xiong, L., Huang, Y., Keegan, B., 2009. As real as real? Macroeconomic behavior in a large-scale virtual world. New Media & Society 11, 685–707.

Chai, K.-H., Yap, C.-M., Wang, X., 2011. Network closure's impact on firms' competitive advantage: The mediating roles of knowledge processes. Journal of Engineering and Technology Management 28, 2–22.

Choi, B.C.K., Pak, A.W.P., 2005. A catalog of biases in questionnaires. Preventing Chronic Disease 2, A13.

Choi, I., 2001. Unit root tests for panel data. Journal of international money and Finance 20, 249–272.

Coleman, J.S., 1988. Social Capital in the Creation of Human Capital. American Journal of Sociology 94, 95–120.

Coleman, J.S., 1990. Foundations of social theory, Cambridge, Mass. : Belknap Press of Harvard University Press.

Conklin, J., Lusk, E., Harris, M., Stolee, P., 2013. Knowledge brokers in a knowledge network: the case of Seniors Health Research Transfer Network knowledge brokers. Implementation Science 8, 1–10.

Cook, K.S., Emerson, R.M., 1978. Power, Equity and Commitment in Exchange Networks. American Sociological Review 43, 721–739.

Cook, K.S., Emerson, R.M., Gillmore, M.R., Yamagishi, T., 1983. The Distribution of Power in Exchange Networks: Theory and Experimental Results. American Journal of Sociology 89, 275–305.

Cross, R., Cummings, J.N., 2004. TIE AND NETWORK CORRELATES OF INDIVIDUAL PERFORMANCE IN KNOWLEDGE-INTENSIVE WORK. The Academy of Management Journal 47, 928–937.

Dorogovtsev, S.N., Mendes, J.F.F., 2002. Evolution of networks. Advances in physics 51, 1079–1187.

Easley, D., Kleinberg, J., 2010. Networks, Crowds, and Markets: Reasoning about a Highly Connected World. Cambridge University Press, New York.

Elliott, L., Ream, G., McGinsky, E., Dunlap, E., 2012. The Contribution of Game Genre and Other Use Patterns to Problem Video Game Play among Adult Video Gamers. International Journal of Mental Health and Addiction 10, 948–969.

Ellison, N.B., Steinfield, C., Lampe, C., 2011. Connection strategies: Social capital implications of Facebook-enabled communication practices. New Media & Society 13, 873–892.

Faraj, S., Jarvenpaa, S.L., Majchrzak, A., 2011. Knowledge Collaboration in Online Communities. Organization Science 22, 1224–1239.

Faraj, S., Johnson, S.L., 2011. Network Exchange Patterns in Online Communities. Organization Science 22, 1464–1480.

Freeman, L., 2004. The development of social network analysis. A Study in the Sociology of Science.

Fritsch, M., Kauffeld-Monz, M., 2010. The impact of network structure on knowledge transfer: an application of social network analysis in the context of regional innovation networks. The Annals of Regional Science 44, 21–38.

Fukuyama, F., 1995. Trust: The social virtues and the creation of prosperity. The Free Press, New York.

Fukuyama, F., 1997. Social capital and the modern capitalist economy: Creating a high trust workplace. Stern Business Magazine 4, 1–16.

Galunic, C., Ertug, G., Gargiulo, M., 2012. The Positive Externalities of Social Capital: Benefiting from Senior Brokers. Academy of Management Journal 55, 1213–1231.

Gargiulo, M., Benassi, M., 2000. Trapped in Your Own Net? Network Cohesion, Structural Holes, and the Adaptation of Social Capital. Organization Science 11, 183–196.

Gargiulo, M., Ertug, G., Galunic, C., 2009. The Two Faces of Control: Network Closure and Individual Performance among Knowledge Workers. Administrative Science Quarterly 54, 299–333.

Gilles, R.P., 2010. The Cooperative Game Theory of Networks and Hierarchies, Springer Science & Business Media.

Goyal, S., 2009. Social Networks in Economics. The SAGE Handbook of Social Network Analysis 67–79.

Granovetter, M., 1973. The Strength of Weak Ties. American Journal of Sociology 78, 1360–1380.

Granovetter, M., 1983. THE STRENGTH OF WEAK TIES: A NETWORK THEORY REVISITED. Sociological Theory 1, 201–233.

Granovetter, M., 1985. Economic Action and Social Structure: The Problem of Embeddedness. American Journal of Sociology 91, 481–510.

Granovetter, M., 1992. Problems of explanation in economic sociology. Networks and organizations: Structure, form, and action 25–56.

Granovetter, M., 1995. Getting a job: a study of contacts and careers, University of Chicago Press.

Granovetter, M., 2005. The impact of social structure on economic outcomes. Journal of Economic Perspectives 19, 33–50.

Hadri, K., 2000. Testing for stationarity in heterogeneous panel data. The Econometrics Journal 3, 148–161.

Hanifan, L.J., 1916. The Rural School Community Center, Annals of the American Academy of Political and Social Science 67: 130.

Hargittai, E., 2015. Is Bigger Always Better? Potential Biases of Big Data Derived from Social Network Sites. The ANNALS of the American Academy of Political and Social Science 659, 63–76.

Harris, R.D.F., Tzavalis, E., 1999. Inference for unit roots in dynamic panels where the time dimension is fixed. J. Econ., Journal of Econometrics 91, 201–226.

Hollingshead, A.B., 1948. Community Research: Development and Present Condition. American Sociological Review 13, 136–156.

Hollingshead, A.B., 1949. Class and Kinship in a Middle Western Community. American Sociological Review 14, 469–475.

Hughes, J., Lang, K., 2006. Transmutability: Digital Decontextualization, Manipulation, and Recontextualization as a New Source of Value in the Production and Consumption of Culture Products. In: Proceedings of the 39th Annual Hawaii International Conference on System Sciences (HICSS'06), Proceedings of the 39th Annual Hawaii International Conference on System Sciences (HICSS'06). Kauai, HI, pp. 165a–165a.

Ichniowski, C., Shaw, K.L., 2009. Connective Capital as Social Capital: The Value of Problem-Solving Networks for Team Players in Firms.

Im, K.S., Pesaran, M.H., Shin, Y., 2003. Testing for unit roots in heterogeneous panels. Journal of Econometrics 115, 53–74.

Inglehart, R., 1997. Modernization and Postmodernization, Canadian Journal of Political Science.

Iseke, A., 2009. Humankapital und Sozialkapital: Gibt es einen Matthäus-Effekt bezüglich der Sozialkapitalbildung von Nachwuchsführungskräften. ZFBF: Schmalenbachs Zeitschrift für Betriebswirtschaftliche Forschung 61, 186–211.

Iseke, A., Plaßmann, B., Schneider, M.R., 2011. Organizational Divide, Demographic Divide, and Performance of R&D Teams: A Fuzzy-set Qualitative Comparative Analysis 1–39.

Jansen, D., 2006. Einführung in die Netzwerkanalyse, VS Verlag fnr Sozialw.

Kick, E.L., McKinney, L.A., McDonald, S., Jorgenson, A., 2011. A multiple-network analysis of the World System of Nations, 1995-1999. The SAGE Handbook of Social Network Analysis 311–328.

Kilduff, M., 1992. The friendship network as a decision-making resource: Dispositional moderators of social influences on organizational choice. Journal of Personality and Social Psychology 62, 168–180.

Kilduff, M., Brass, D.J., 2010. Organizational Social Network Research: Core Ideas and Key Debates. The Academy of Management Annals 4, 317–357.

Kleinberg, J., 2007. The wireless epidemic. Nature 449, 287–288.

Knoke, D., 1999. Organizational Networks and Corporate Social Capital. In: Leenders, R.Th.A.J., Gabbay, S.M. (Eds.), Corporate Social Capital and Liability. Springer US, Boston, MA, pp. 17–42.

Kogut, B., 2000. The network as knowledge: Generative rules and the emergence of structure. Strategic Management Journal 21, 405–425.

Kossinets, G., Kleinberg, J., Watts, D., 2008. The Structure of Information Pathways in a Social Communication Network. In: Proceedings of the 14th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, KDD '08. Association for Computing Machinery, New York, NY, USA, pp. 435–443.

Krempel, L., Pluemper, T., 1999. International division of labor and global economic processes: an analysis of the international trade in automobiles. Journal of World-Systems Research 486–499.

Krempel, L., Plümper, T., 2003. Exploring the dynamics of international trade by combining the comparative advantages of multivariate statistics and network visualizations. Journal of social structure 4, 1–22.

Kumbasar, E., Rommey, A.K., Batchelder, W.H., 1994. Systematic Biases in Social Perception. American Journal of Sociology, American Journal of Sociology 100, 477–505.

Lee, E., Karimi, F., Wagner, C., Jo, H.-H., Strohmaier, M., Galesic, M., 2019. Homophily and minority-group size explain perception biases in social networks. Nature Human Behaviour, Nature Human Behaviour 3, 1078–1087.

Levin, A., Lin, C.-F., Chu, C.-S.J., 2002. Unit root tests in panel data: asymptotic and finite-sample properties. J. Econ., Journal of Econometrics 108, 1–24.

Lin, N., 2002. Social capital: A theory of social structure and action.

Loury, G., 1992. The economics of discrimination: Getting to the core of the problem. Harvard Journal for African American Public Policy 91–110.

Loury, G.C., 1976. A Dynamic Theory of Racial Income Differences. Discussion Papers.

Machado, D. de Q., Ipiranga, A.S.R., 2013. Characteristics and performance of knowledge networks in the biotechnology sector. Revista de Administração Contemporânea 17, 350–367.

Mangino, W., 2009. The Downside of Social Closure: Brokerage, Parental Influence, and Delinquency Among African American Boys. Sociology of Education 82, 147–172.

Markovsky, B., Willer, D., Patton, T., 1988. Power Relations in Exchange Networks. American Sociological Review 53, 220–236.

Martinez, S.M., Arredondo, E.M., Roesch, S., 2013. Physical activity promotion among churchgoing Latinas in San Diego, California: Does neighborhood cohesion matter? Journal of Health Psychology 18, 1319–1329.

McGloin, J.M., Kirk, D.S., 2010. An Overview of Social Network Analysis. Journal of Criminal Justice Education 21, 169–181.

McIllwain, J.S., 1999. Organized crime: A social network approach. Crime, Law and Social Change 32, 301–323.

Mehra, A., Kilduff, M., Brass, D.J., 2001. The Social Networks of High and Low Self-Monitors: Implications for Workplace Performance. Administrative Science Quarterly 46, 121–146.

Milgram, S., 1967. The small world problem. Psychology today 2, 60–67.

Moran, P., 2005. Structural vs. relational embeddedness: social capital and managerial performance. Strategic Management Journal 26, 1129–1151.

Moreno, J.L., 1946. The Sociometric View of the Community. Journal of Educational Sociology 19, 540–545.

Morgan, S.L., Sørensen, A.B., 1999. Parental Networks, Social Closure, and Mathematics Learning: A Test of Coleman's Social Capital Explanation of School Effects. American Sociological Review 64, 661–681.

Müller, D., 2009. Methodik der empirischen Forschung, Springer.

Mutter, T., Kundisch, D., 2014. Don't take away my status! – Evidence from the restructuring of a virtual reward system. Computer Networks 75, 477–490.

Nahapiet, J., Ghoshal, S., 1998. Social Capital, Intellectual Capital, and the Organizational Advantage. The Academy of Management Review 23, 242–266.

Newman, M.E.J., 2003. The structure and function of complex networks. SIAM review 45, 167–256.

Norman, C.D., Huerta, T., 2006. Knowledge transfer & exchange through social networks: building foundations for a community of practice within tobacco control. Implementation Science 1, 20.

Paulhus, D.L., 1984. Two-component models of socially desirable responding. Journal of Personality and Social Psychology 46, 598–609.

Paulhus, D.L., 1986. Self-Deception and Impression Management in Test Responses. In: Angleitner, A., Wiggins, jerry S. (Eds.), Personality Assessment via Questionnaires: Current Issues in Theory and Measurement. Springer Berlin/Heidelberg, pp. 143–165.

Paulhus, D.L., 1991. Measurement and control of response bias., Academic Press.

Pennar, K., 1997. The ties that lead to prosperity: the economic value of social bonds is only beginning to be measured. Business Week 153–155.

Pool, I. de S., Milgram, S., Newcomb, T., Kochen, M., 1989. The Small World, Communication and information science, Ablex Publishing Corporation. Norwood, N.J.

Portes, A., 1998. Social capital: Its origins and applications in modern sociology. Annual Review of Sociology 24, 1–24.

Portes, A., Sensenbrenner, J., 1993. Embeddedness and Immigration: Notes on the Social Determinants of Economic Action. American Journal of Sociology 98, 1320–1350.

Putnam, R.D., 1995. Bowling Alone: America's Declining Social Capital. Journal of Democracy 6, 65–78.

Rapoport, A., 1954. Spread of information through a population with socio-structural bias: III. Suggested experimental procedures. The bulletin of mathematical biophysics 16, 75–81.

Raub, W., Weesie, J., 1990. Reputation and Efficiency in Social Interactions: An Example of Network Effects. American Journal of Sociology 96, 626–654.

Rice, E., Yoshioka-Maxwell, A., 2015. Social Network Analysis as a Toolkit for the Science of Social Work. Journal of the Society for Social Work and Research 6, 369–383.

Robins, G., Pattison, P., Wang, P., 2009. Closure, connectivity and degree distributions: Exponential random graph (p*) models for directed social networks. Social Networks 31, 105–117.

Schank, T., Wagner, D., 2005. Approximating Clustering Coefficient and Transitivity. Journal of Graph Algorithms and Applications 9, 265–275.

Schiff, M., 1992. Social Capital, Labor Mobility, and Welfare: The Impact of Uniting States. Rationality and Society 4, 157–175.

Schilling, M.A., Phelps, C.C., 2007. Interfirm Collaboration Networks: The Impact of Large-Scale Network Structure on Firm Innovation. Management Science 53, 1113–1126.

Scott, J., Carrington, P.J. (Eds.), 2011. The SAGE Handbook of Social Network Analysis. SAGE Publications Limited, London.

Simmel, G., 2009. Soziologie. Untersuchungen über die Formen der Vergesellschaftung (German Edition), BiblioBazaar.

Snodgrass, J.G., 2016. Online virtual worlds as anthropological field sites: Ethnographic methods training via collaborative research of Internet gaming cultures. Annals of Anthropological Practice 40, 134–147.

Sparrowe, R.T., Liden, R.C., Wayne, S.J., Kraimer, M.L., 2001. Social networks and the performance of individuals and groups. Academy of Management Journal 44, 316–325.

Thomas, C.Y., 1996. CAPITAL MARKETS, FINANCIAL MARKETS AND SOCIAL CAPITAL (An Essay on Economic Theory and Economic Ideas). Social and Economic Studies 45, 1–23.

Thurau, C., Bauckhage, C., 2010. Analyzing the Evolution of Social Groups in World of Warcraft®. In: Proceedings of the 2010 IEEE Conference on Computational Intelligence and Games, Presented at the Proceedings of the 2010 IEEE Conference on Computational Intelligence and Games. Copenhagen, pp. 170–177.

Tortoriello, M., Reagans, R., McEvily, B., 2012. Bridging the Knowledge Gap: The Influence of Strong Ties, Network Cohesion, and Network Range on the Transfer of Knowledge Between Organizational Units. Organization Science 23, 1024–1039.

Trier, M., Bobrik, A., 2007. Analyzing the Dynamics of Community Formation Using Brokering Activities. In: Communities and Technologies 2007. pp. 463–477.

van der Hulst, R.C., 2014. Terrorist Networks: The Threat of Connectivity. The SAGE Handbook of Social Network Analysis 256–270.

Vargas, D.L., Bridgeman, A.M., Schmidt, D.R., Kohl, P.B., Wilcox, B.R., Carr, L.D., 2018. Correlation between student collaboration network centrality and academic performance. Physical Review Physics Education Research 14, 020112.

Vernet, A., Kilduff, M.J., Salter, A., 2013. Binoculars and Blinders: Anticipating Trends and Breakthroughs in Communities. Academy of Management Proceedings 1.

Wasko, M.M., Faraj, S., 2005. Why should I share? Examining social capital and knowledge contribution in electronic networks of practice. MIS quarterly 35–57.

Wasserman, S., Faust, K., 1994. Social Network Analysis: Methods and Applications (Structural Analysis in the Social Sciences). Cambridge University Press, Cambridge/New York.

Watkins, D., Cheung, S., 1995. Culture, Gender, and Response Bias: An Analysis of Responses to the Self-Description Questionnaire. Journal of Cross-Cultural Psychology 26, 490–504.

Watts, D.J., Strogatz, S.H., 1998. Collective dynamics of ‘small-world’networks. Nature 393, 440–442.

Wei, J., Zheng, W., Zhang, M., 2011. Social capital and knowledge transfer: A multilevel analysis. Human Relations 64, 1401–1423.

White, H.D., 2011. Scientific and scholarly networks. The SAGE Handbook of Social Network Analysis 271–285.

Woolcock, M., 1998. Social Capital and Economic Development: Toward a Theoretical Synthesis and Policy Framework. Theory and Society 27, 151–208.

Wooldridge, J.M., 2010. Econometric Analysis of Cross Section and Panel Data.

Yair, G., Maman, D., 1996. The Persistent Structure of Hegemony in the Eurovision Song Contest. Acta Sociologica 39, 309–325.

Zaheer, A., Bell, G.G., 2005. Benefiting from network position: firm capabilities, structural holes, and performance. Strategic Management Journal 26, 809–825.

9 Appendix

Internal vs. External Focus	Author	Definitions of social capital
External	Baker	"[a] resource that actors derive from specific social structures and then use to pursue their interests; it is created by changes in the relationship among actors" (1990, p. 619)
	Belliveau, O'Reilly, & Wade	"an individual's personal network and elite institutional affiliations" (1996, p. 1572)
	Bourdieu	"the aggregate of the actual or potential resources which are linked to possession of a durable network of more or less institutionalized relationships of mutual acquaintance or recognition" (1986, p. 248). "made up of social obligations ('connections'), which is convertible, in certain conditions, into economic capital and may be institutionalized in the form of a title of nobility" (1986, p. 243).
	Bourdieu & Wacquant	"the sum of the resources, actual or virtual, that accrue to an individual or a group by virtue of possessing a durable network of more or less institutionalized relationships of mutual acquaintance and recognition" (1992, p. 119)

	Boxman, De Graaf. & Flap	"the number of people who can be expected to provide support and the resources those people have at their disposal" (1991, p. 52)
	Burt	"friends, colleagues, and more general contacts through whom you receive opportunities to use your financial and human capital" (1992, p. 9)
	Knoke	"the process by which social actors create and mobilize their network connections within and between organizations to gain access to other social actors' resources" (1999, p. 18)
	Portes	"the ability of actors to secure benefits by virtue of membership in social networks or other social structures" (1998, p. 6)
Internal	Brehm & Rahn	"the web of cooperative relationships between citizens that facilitate resolution of collective action problems" (1997, p. 999)
	Coleman	"Social capital is defined by its function. It is not a single entity, but a variety of different entities having two characteristics in common: They all consist of some aspect of social structure, and they facilitate certain actions of individuals who are within the structure" (1990, p. 302)
	Fukuyama	"the ability of people to work together for common purposes in groups and organizations" (1995, p. 10) "Social capital can be defined simply as the existence of a certain set of informal values

		or norms shared among members of a group that permit cooperation among them" (1997)
	Inglehart	"a culture of trust and tolerance, in which extensive networks of voluntary associations emerge" (1997, p. 188)
	Portes & Sensenbrenner	"those expectations for action within a collectivity that affect the economic goals and goal-seeking behavior of its members, even if these expectations are not oriented toward the economic sphere" (1993, p. 1323)
	Putnam	"features of social organization such as networks, norms, and social trust that facilitate coordination and cooperation for mutual benefit" (1995, p. 67)
	Thomas	"those voluntary means and processes developed within civil society which promote development for the collective whole" (1996, p. 11)
Both	Loury	"naturally occurring social relationships among persons which promote or assist the acquisition of skills and traits valued in the marketplace ... an asset which may be as significant as financial bequests in accounting for the maintenance of inequality in our society" (1992, p. 100).
	Nahapiet & Ghosha	"the sum of the actual and potential resources embedded within, available through, and derived from the network of relationships possessed by an individual or social unit. Social capital thus comprises both the network and

		the assets that may be mobilized through that network" (1998, p. 243).
	Pennar	"the web of social relationships that influences individual behavior and thereby affects economic growth" (1997, p. 154).
	Schiff	"the set of elements of the social structure that affects relations among people and are inputs or arguments of the production and/or utility function" (1992, p. 160)
	Woolcock	"the information, trust, and norms of reciprocity inhering in one's social networks" (1998, p. 153)

Appendix 1: Definitions of social capital (Adler and Kwon, 2002)

Table structure of friend_request

Users can request each other to be friends. Any request with a certain request message is stored in this table.

Field	Type	Null	Standard
requester_id	int(10)	Yes	0
target_id	int(10)	Yes	0
status	enum('pending', 'accepted', 'declined')	Yes	pending
create_time	int(10)	Yes	0
answer_time	int(10)	Yes	0
message	varchar(255)	Yes	

Appendix 2: Table structure friend request

Table structure of points_history

Any earning or loosing of points (status and credit) is logged in this history table. There is a huge amount of reasons wrapped in types and certain reasons are stored addionally.

Actual available types are:

TRESHOLD_PQ	(status points: threshold reached for status points gained from asking questions) -> obsolete after point system revamp in 2007
TRESHOLD_PQ_RHF	(status points: threshold reached for helpful question votes, first)
PQ_HELPFUL	(status points: helpful question votes, first)
PQ_ANSWERED	(status points: first answer to a question, later N_QA_ANSW)
PQ_RATING	(status points: vote for answers)
AQ_HELPFUL	(status points: status points for an answer that was rated helpful by the questioner)
AQ_TOP	(status points: status points for an answer that was rated top by the questioner)
AQ_HELPFUL_BOTH	(status points: status points for an answer that was rated helpful by the questioner [and got also some credit points])
AQ_TOP_BOTH	(status points: status points for an answer that was rated top by the questioner [and got also some credit points])
TRESHOLD_AQ	(status points: threshold for status points gained from answering questions)
OBSERV	(status points: user is observed)
TRESHOLD_BO	(status points: threshold for being observed)

OBSERV_FR	(status points: user is observed because somebody wants to be his friend)
TRESHOLD_BO_FR	(status points: threshold for being observed because of friend requests)
OBSERV_FR_OK	(status points: user is observed because somebody confirmed his friend request)
TRESHOLD_BO_FR_OK	(status points: threshold for being observed because of a confirmed friend request)
UN_OBSERVE	(status points: user is no longer observed)
AU_UPRANK	(status points: inviter gains status points because an invited member has reached a new rank)
TRESHOLD_AU_UPR	(status points: threshold for points from uprank of invited members)
AU_INVITE	(status points: invited person registers for Lycos iQ [full registration])
TRESHOLD_AU_INV	(status points: threshold for registration of invited persons)
LINK_ADDED	(status points: adding a public link to the link library)
TRESHOLD_AL	(status points: threshold for adding public links to the link library)
TRESHOLD_AL_PUB	(status points: make a private link public but threshold is reached)
TRESHOLD_AL_TAG	(status points: adding tags to an official link but the threshold is reached)
LINK_COPIED	(status points: my link was copied / bookmarked from other users)
TRESHOLD_CL	(status points: threshold for copied links [by other users])
BM_TAGS_CHANGED	(status points: tag changed for a bookmark)
BM_VISIBILITY_CHANGED	(status points: public->private or private->public)
BM_TAGS_REMOVED	(status points: tags removed for a bookmark [because bookmark was removed])
ADMIN_ADD	(status points: admin ads status points)
ADMIN_SUB	(status points: admin subducts status points)
UPDATED_USER_TAGS	(status points: changed own tags in profile)
AQ_VOTE	(status points: gained 20th vote for own answer)
UPRANK	(credit points: reached a new level)
DONATEDTO	(credit points: give away credit points)
DONATIONFROM	(credit points: receive credit points)
ASK_QUESTION	(credit points: subtract credit points for asking a question)
AQ_HELPFUL	(credit points: an answer was rated helpful by the questioner or the system)
AQ_TOP	(credit points: an answer was rated top by the questioner or the system)

AQ_BACK	(credit points: retrieve credit points because no one answered the question)
ADMIN_ADD	(credit points: admin ads credit points)
ADMIN_SUB	(credit points: admin subducts credit points)
N_QA_ANSW	(get an answer for an own question)
N_ANSW_AMNDM	(get an answer amendment for an own question)
N_QA_CMNT	(get a comment for an own question)
N_QA_OPIN	(get a comment for an answer of an own question)
N_QA_CLOSE	(question closed)
N_QA_CLOSE_NA	(question closed, without answers)
N_QA_OBS_ANSW	(answer for an observed question)
N_QA_OBS_CMNT	(comment for an observed question)
N_QA_OBS_AMND	(answer amendment for an observed question)
N_ANSW_OPIN	(get a comment to an own answer)
N_QA_DIRECT	(get a direct question)
N_QA_DISAB	(own question got disabled)
N_QA_ENAB	(own gestion got enabled)
N_FREQ	(new friend request without new observing)
N_AUTO_FREQ	(auto friend function, both are observing the other, watch/watch)
N_MOD_MSG	(message from moderator)
N_MSG	(message from other member)
N_QA_ANSW_AMND	(got an amendment for a question that was answered by the user)

Field	Type	Null	Standard
id	int(10)	Yes	NULL
ptype	enum('sp', 'cp', 'event')	Yes	sp
type	enum('', 'TRESHOLD_PQ', 'TRESHOLD_PQ_RHF', 'PQ_HELPFUL', 'PQ_ANSWERED', 'PQ_RATING', 'AQ_HELPFUL', 'AQ_TOP', 'AQ_HELPFUL_BOTH', 'AQ_TOP_BOTH', 'TRESHOLD_AQ', 'OBSERV', 'OBSERV_FR', 'OBSERV_FR_OK', 'TRESHOLD_BO', 'TRESHOLD_BO_FR', 'TRESHOLD_BO_FR_OK', 'UN_OBSERVE', 'AU_UPRANK', 'TRESHOLD_AU_UPR', 'AU_INVITE', 'TRESHOLD_AU_INV', 'LINK_ADDED', 'TRESHOLD_AL', 'TRESHOLD_AL_PUB', 'TRESHOLD_AL_TAG', 'LINK_COPIED', 'TRESHOLD_CL', 'BM_TAGS_CHANGED', 'BM_VISIBILITY_CHANGED', 'BM_TAGS_REMOVED', 'ADMIN_ADD', 'ADMIN_SUB', 'UPDATED_USER_TAGS', 'UPRANK', 'DONATEDTO', 'DONATIONFROM', 'ASK_QUESTION', 'AQ_BACK', 'N_QA_ANSW', 'N_ANSW_AMNDM', 'N_QA_CMNT', 'N_QA_OPIN', 'N_QA_CLOSE', 'N_QA_CLOSE_NA', 'N_QA_OBS_ANSW', 'N_QA_OBS_CMNT', 'N_QA_OBS_AMND', 'N_ANSW_OPIN', 'N_QA_DIRECT', 'N_QA_DISAB', 'N_QA_ENAB', 'N_FREQ', 'N_AUTO_FREQ', 'N_MOD_MSG', 'N_MSG', 'N_QA_ANSW_AMND', 'AQ_VOTE')	Yes	
expert_id	int(10)	Yes	0
target_id	int(10)	Yes	0
note_id	int(10)	Yes	0
amount	int(10)	Yes	0
total	int(10)	Yes	0
reason_temp	varchar(255)	Yes	
data	varchar(30)	Yes	
create_time	int(10)	Yes	0
flag	set('', 'seen', 'dismissed', 'nomsg')	Yes	

Appendix 3: Table structure points history

type	minamount	maxamount	countamount	sumamount	avgamount
TRESHOLD_PQ	-	-	3.305	-	-
TRESHOLD_PQ_RHF	-	-	129	-	-
PQ_HELPFUL	200	400	110.743	22.592.200	204,01
PQ_ANSWERED	100	400	569.996	213.953.440	375,36
PQ_RATING	100	200	353.617	69.002.060	195,13
AQ_HELPFUL	8	2.500	1.237.116	478.610.848	386,88
AQ_TOP	23	1.000	178.397	65.861.219	369,18
AQ_HELPFUL_BOTH	3	2.500	862.972	333.670.750	386,65
AQ_TOP_BOTH	7	2.500	434.301	195.954.671	451,20
OBSERV	-	2.000	16.190	10.969.400	677,54
OBSERV_FR	-	2.000	42.438	7.710.400	181,69
OBSERV_FR_OK	-	2.000	29.966	3.829.000	127,78
TRESHOLD_BO	-	-	525	-	-
TRESHOLD_BO_FR	-	-	444	-	-
TRESHOLD_BO_FR_OK	-	-	1.567	-	-
UN_OBSERVE	- 2.000	-	13.122	- 2.032.300	- 154,88
AU_UPRANK	100	8.000	2.151	1.608.800	747,93
AU_INVITE	500	2.000	347	270.500	779,54
LINK_ADDED	100	200	154.554	22.470.470	145,39
TRESHOLD_AL	-	-	5.852	-	-
TRESHOLD_AL_TAG	-	-	24	-	-
LINK_COPIED	5	200	140.600	7.401.725	52,64
BM_TAGS_CHANGED	- 200	200	2.058	62.540	30,39
BM_VISIBILITY_CHANGED	- 200	200	1.203	101.470	84,35
BM_TAGS_REMOVED	- 200	- 100	5.457	- 656.460	- 120,30
ADMIN_ADD	1	999.900	342	2.142.784	6.265,45
ADMIN_SUB	- 1.300.000	- 9	2.984	- 38.485.594	- 12.897,32
UPDATED_USER_TAGS	- 1.000	1.000	68.391	18.073.700	264,27
AQ_VOTE	300	300	450	135.000	300,00

Appendix 4: Descriptive statistics of status points[39]

type	minamount	maxamount	countamount	sumamount	avgamount
AQ_HELPFUL	1	1.685.184	924.690	1.491.642.491	1.613,13
AQ_TOP	3	5.055.550	434.301	863.514.289	1.988,29
ADMIN_ADD	14	5.000.000	14.616	1.332.204.604	91.147,00
ADMIN_SUB	- 207.533.151	- 16	877	- 433.246.580	-494.009,78
UPRANK	-	300.000	214.987	583.310.000	2.713,23
DONATEDTO	- 7.000.000	- 100	39.035	- 923.222.800	- 23.651,15
DONATIONFROM	100	7.000.000	39.035	923.222.800	23.651,15
ASK_QUESTION	- 10.091.100	- 2	405.089	- 2.694.366.003	- 6.651,29
AQ_BACK	-	2.000.000	28.877	212.962.674	7.374,82

Appendix 5: Descriptive statistics of credit points[40]

[39] Points in the database need to be divided by 100 to represent actual points
[40] Points in the database need to be divided by 100 to represent actual points

Sum points per status point type

SELECT FROM_UNIXTIME(create_time, '%m-%Y') , expert_id, SUM(amount) AS sum, type

FROM points_history

WHERE ptype = "sp"

GROUP BY FROM_UNIXTIME(create_time, '%m-%Y') , expert_id, type

ORDER BY expert_id, create_time, type

Count Activities per status point type

SELECT FROM_UNIXTIME(create_time, '%m-%Y') , expert_id, COUNT(amount) AS count, type

FROM points_history

WHERE ptype = „sp"

GROUP BY FROM_UNIXTIME(create_time, '%m-%Y') , expert_id, type

ORDER BY expert_id, create_time, type

Appendix 6: SQL queries points and activities

			Results Software		Manually proven results
timestamp	expert_id (Original)	expert_id (fortlaufend)	friendspertime	amountpertime	timestamp expertid amount~e f~frie~e friend..
1137316921	673	173	0	3150	5095211. 1137316921 173 3150 0 0
1144057131	2690	364	0	3575	10753578. 1144057131 364 3575 0 0
1153147785	6296	512	2	21085	15138685. 1153147785 512 21085 0 2
1154455991	11683	670	47	1413269	19819260. 1154455991 670 1413269 276 47
1155057977	19654	925	1	2527	27373008. 1155057977 925 2527 0 1
1157987005	19765	934	1	2209	27640187. 1157987005 934 2209 0 1
1159126632	28543	1303	0	0	38570961. 1159126632 1303 0 0 0
1159809777	49849	2002	0	0	59276903. 1159809777 2002 0 0 0
1160406255	59003	2186	0	0	64727476. 1160406255 2186 0 0 0
1161085837	59856	2197	0	0	65053591. 1161085837 2197 0 0 0
1167767528	71416	2347	0	0	69498915. 1167767528 2347 0 0 0
1168889182	86472	2544	0	0	75335140. 1168889182 2544 0 0 0
1171789930	97654	2733	0	0	80934868. 1171789930 2733 0 0 0
1174744536	112775	2949	0	0	87334436. 1174744536 2949 0 0 0
1175096172	141824	3265	0	0	96695117. 1175096172 3265 0 0 0
1176145117	146212	3321	0	0	98354300. 1176145117 3321 0 0 0
1179773692	157140	3445	0	0	102028514. 1179773692 3445 0 0 0
1180965818	173694	3633	0	0	107597867. 1180965818 3633 0 0 0
1182981518	178193	3707	0	0	109790755. 1182981518 3707 0 0 0
1189688695	184545	3801	1	45953	112578105. 1189688695 3801 45953 0 1
1191405163	201947	4017	10	36872	118977256. 1191405163 4017 36872 3 10
1192268521	211042	4165	1	22031	123361867. 1192268521 4165 22031 0 1
1193148698	220798	4304	4	121295	127479902. 1193148698 4304 121295 2 4
1193348358	238208	4582	16	326716	135714966. 1193348358 4582 326716 8 16
1194891916	240517	4623	6	77739	136930948. 1194891916 4623 77739 2 6
1195137515	240814	4629	0	2165	137108850. 1195137515 4629 2165 0 0
1195486772	244068	4682	2	25308	138679036. 1195486772 4682 25308 1 2
1195500087	261240	4954	14	470799	146736231. 1195500087 4954 470799 20 14
1195535322	264165	5024	0	21953	148809788. 1195535322 5024 21953 0 0
1198052152	265444	5051	1	6522	149611155. 1198052152 5051 6522 0 1
1198415556	266890	5081	0	38726	150500027. 1198415556 5081 38726 0 0
1199077085	267077	5083	0	122087	150559580. 1199077085 5083 122087 0 0
1199243415	269528	5130	1	30007	151951898. 1199243415 5130 30007 0 1
1200162311	269858	5140	7	79623	152248731. 1200162311 5140 79623 0 7
1200956386	284013	5383	7	29109	159447482. 1200956386 5383 29109 1 7
1201872886	290068	5497	3	29346	162825124. 1201872886 5497 29346 0 3
1201986645	292577	5545	25	223777	164247071. 1201986645 5545 223777 4 25
1202893899	296208	5612	8	78572	166232389. 1202893899 5612 78572 0 8
1202980164	322772	6159	1	56291	182435672. 1202980164 6159 56291 0 1
1203108259	326862	6247	52	140697	185042526. 1203108259 6247 140697 76 52
1203196319	332612	6344	1	7558	187915950. 1203196319 6344 7558 0 1
1203801715	333318	6357	0	1867	188301424. 1203801715 6357 1867 0 0
1205953485	364087	6824	0	5656	202136310. 1205953485 6824 5656 0 0
1206477134	365703	6856	1	0	203084559. 1206477134 6856 0 0 1
1206714445	366740	6871	1	38863	203529033. 1206714445 6871 38863 0 1
1208524117	393057	7225	10	170244	214016427. 1208524117 7225 170244 2 10
1209672191	408353	7387	13	242006	218815862. 1209672191 7387 242006 10 13
1210445562	415028	7478	1	0	221511929. 1210445562 7478 0 0 1
1210585032	415673	7488	0	1000	221808245. 1210585032 7488 1000 0 0
1211201227	443159	7749	0	0	229539994. 1211201227 7749 0 0 0

Appendix 7: Dataset verification of 50 datapoints

```
replace datetime = datetime*1000 + msofhours(24)*3653 - msofhours(5)
```

Appendix 8: Unixtime conversion Stata code

VARIABLES	(1) aq_helpful_points_sum_ln
clusteringcoefficient	1.993***
	(0.0853)
clusteringcoefficientsqrd	-3.291***
	(0.0832)
Constant	5.045***
	(0.0123)
Observations	42,197
R-squared	0.904

Standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Appendix 9: Hypothesis 1 areg model

VARIABLES	(2) aq_helpful_count_sum_ln
clusteringcoefficient	1.971***
	(0.0831)
clusteringcoefficientsqrd	-3.330***
	(0.0811)
Constant	3.774***
	(0.0120)
Observations	42,197
R-squared	0.902

Standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Appendix 10: Hypothesis 2 areg model

VARIABLES	(3) aq_helpful_points_sum_ln	(4) aq_helpful_points_sum_ln
clusteringcoefficient	1.938***	2.738***
	(0.0655)	(0.0887)
ranking_by_system	0.246***	0.276***
	(0.00151)	(0.00276)
c.clusteringcoefficient#c.ranking_by_system		-0.337***
		(0.0198)
clusteringcoefficientsqrd	-2.528***	-4.058***
	(0.0641)	(0.0917)
o.ranking_by_system		-
c.clusteringcoefficientsqrd#c.ranking_by_system		0.644***
		(0.0241)
Constant	3.971***	3.919***
	(0.0115)	(0.0132)
Observations	42,197	42,197
R-squared	0.943	0.944

Standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Appendix 11: Hypothesis 3 areg models

VARIABLES	(5) aq_helpful_points_sum_ln	(6) aq_helpful_points_sum_ln
clusteringcoefficient	0.814***	1.684***
	(0.0756)	(0.0820)
lifetime	0.0633***	0.0954***
	(0.000593)	(0.00120)
c.clusteringcoefficient#c.lifetime		-0.150***
		(0.00642)
clusteringcoefficientsqrd	-1.828***	-2.446***
	(0.0742)	(0.0806)
o.lifetime		-
c.clusteringcoefficientsqrd#c.lifetime		0.0911***
		(0.00630)
Constant	4.602***	4.445***
	(0.0116)	(0.0124)
Observations	42,197	42,197
R-squared	0.926	0.928

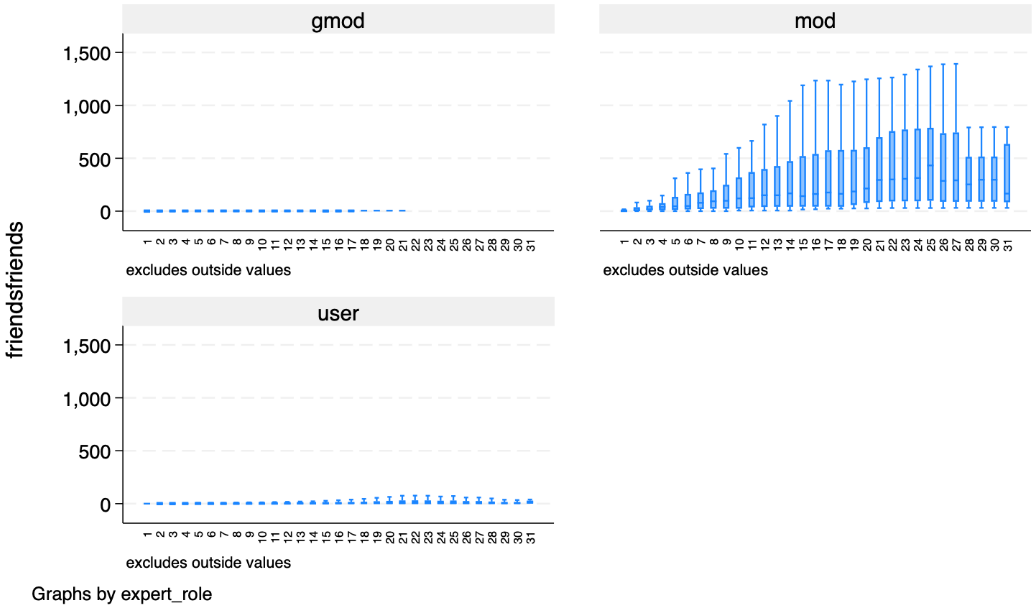
Standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Appendix 12: Hypothesis 4 areg models

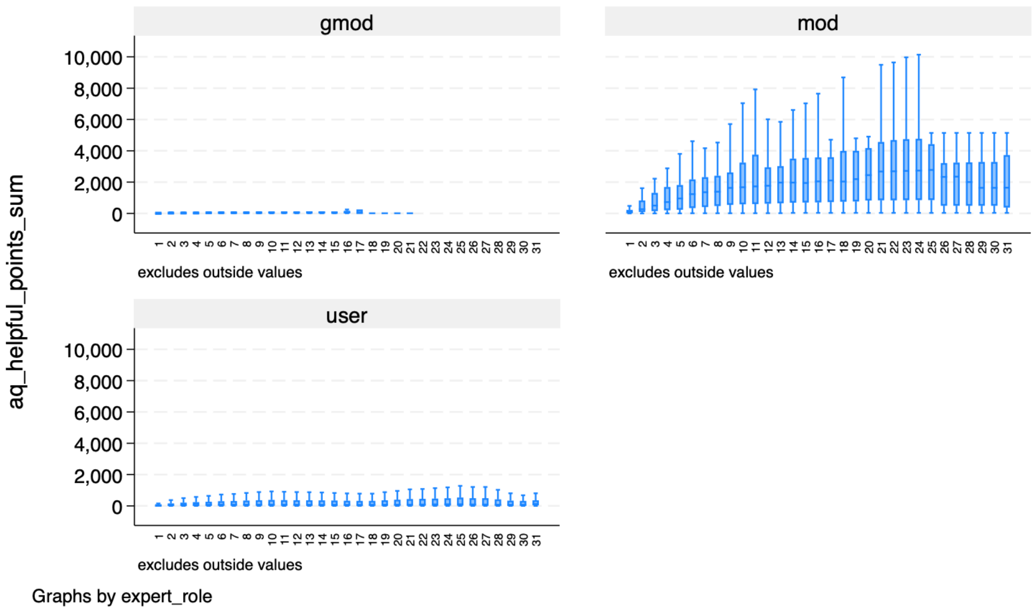
Appendix 13: Combined areg models

	(1)	(2)	(3)	(4)	(5)	(6)
VARIABLES	aq_helpful_points_sum_ln	aq_helpful_count_sum_ln	aq_helpful_points_sum_ln	aq_helpful_points_sum_ln	aq_helpful_points_sum_ln	aq_helpful_points_sum_ln
clusteringcoefficient	1.993***	1.971***	1.938***	2.738***	0.814***	1.684***
	(0.0853)	(0.0831)	(0.0655)	(0.0887)	(0.0756)	(0.0820)
o.clusteringcoefficient				-		-
ranking_by_system			0.246***	0.276***		
			(0.00151)	(0.00276)		
c.clusteringcoefficient#c.ranking_by_system				-0.337***		
				(0.0198)		
clusteringcoefficientsqrd	-3.291***	-3.330***	-2.528***	-4.058***	-1.828***	-2.446***
	(0.0832)	(0.0811)	(0.0641)	(0.0917)	(0.0742)	(0.0806)
o.ranking_by_system				-		
c.clusteringcoefficientsqrd#c.ranking_by_system				0.644***		
				(0.0241)		
lifetime					0.0633***	0.0954***
					(0.000593)	(0.00120)
c.clusteringcoefficient#c.lifetime						-0.150***
						(0.00642)
o.lifetime						-
c.clusteringcoefficientsqrd#c.lifetime						0.0911***
						(0.00630)
Constant	5.045***	3.774***	3.971***	3.919***	4.602***	4.445***
	(0.0123)	(0.0120)	(0.0115)	(0.0132)	(0.0116)	(0.0124)
Observations	42,197	42,197	42,197	42,197	42,197	42,197
R-squared	0.904	0.902	0.943	0.944	0.926	0.928

Standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1



Appendix 14: Boxplots by friendfriends over role



Appendix 15: Boxplots by aq_helpful_points_sum over role

Appendix number	Folder name	Content Description
1	Database scheme	Description of tables and variables of the complete dataset
2	Database zip	Copy of the complete dataset, including all tables and variables
3	Input C-Sharp and SQL	Reduced dataset for research need (points and friends history tables)
4	C-Sharp Program	Dedicated program to solve data extraction
5	Output C-Sharp and SQL	Output of dedicated program und SQL queries
6	Preparation final dataset STATA	Buildup of complete panel dataset, based on previous output
7	STATA do files	Final set of STATA do files to proceed all necessary steps in STATA

Appendix 16: Structure of digital appendix

10 Eidesstattliche Erklärung

Hiermit versichere ich, Philipe Gerlach, die vorliegende Arbeit selbstständig und unter ausschließlicher Verwendung der angegebenen Literatur und Hilfsmittel erstellt zu haben. Alle Stellen, die wörtlich oder sinngemäß veröffentlichtem oder unveröffentlichtem Schrifttum entnommen sind, habe ich als solche kenntlich gemacht. Die Arbeit wurde bisher in gleicher oder ähnlicher Form keiner anderen Prüfungsbehörde vorgelegt und auch nicht veröffentlicht.

Willich, 02/09/2024