

Gutenberg School of Management and Economics  
& Research Unit “Interdisciplinary Public Policy”

Discussion Paper Series

*How individual characteristics shape the  
structure of social networks*

Yann Girard, Florian Hett and Daniel Schunk

November 2014

Discussion paper number 1414

Johannes Gutenberg University Mainz  
Gutenberg School of Management and Economics  
Jakob-Welder-Weg 9  
55128 Mainz  
Germany  
[wiwi.uni-mainz.de](http://wiwi.uni-mainz.de)

## Contact details

Yann Girard  
Goethe University Frankfurt  
GSEFM  
[yann.girard@web.de](mailto:yann.girard@web.de)

Florian Hett  
Goethe University Frankfurt  
GSEFM  
[hett@econ.uni-frankfurt.de](mailto:hett@econ.uni-frankfurt.de)

Daniel Schunk  
Department of Economics  
Johannes-Gutenberg-Universität Mainz  
Jakob-Welder-Weg 9  
55128 Mainz  
  
[daniel.schunk@uni-mainz.de](mailto:daniel.schunk@uni-mainz.de)

# How individual characteristics shape the structure of social networks

Yann Girard, Florian Hett and Daniel Schunk\*

November 2014

## Abstract

We study how students' social networks emerge by documenting systematic patterns in the process of friendship formation of incoming students; these students all start out in a new environment and thus jointly create a new social network. As a specific novelty, we consider cooperativeness, time and risk preferences - elicited experimentally - together with factors like socioeconomic and personality characteristics. We find a number of robust predictors of link formation and of the position within the social network (local and global network centrality). In particular, cooperativeness has a complex association with link formation. We also find evidence for homophily along several dimensions. Finally, our results show that despite these systematic patterns, social network structures can be exogenously manipulated, as we find that random assignments of students to groups on the first two days of university impacts the students' friendship formation process.

**Keywords:** Social networks, education, link formation, homophily, cooperation, field and lab data.

**JEL Classification Codes:** C93, D85, I25, J24.

---

\*Girard: Goethe University Frankfurt, GSEFM, yann.girard@web.de. Hett: GSEFM, Goethe University Frankfurt, hett@econ.uni-frankfurt.de. Schunk: Department of Economics, University of Mainz, daniel.schunk@uni-mainz.de. We would like to thank several colleagues for their help in conducting the experiments, as well as Felix Schmidt, the editor and three anonymous referees for helpful comments.

# 1 Introduction

A large literature in the social sciences studies the effect of social networks on educational outcomes. The majority of empirical studies in this area document these *peer effects* to be not only statistically significant but also of substantial magnitude.<sup>1</sup> Given this potentially important role, a natural question is whether measures to improve educational outcomes via influencing the network can and should be taken. This, however, requires a deeper understanding of how networks form in educational contexts and which individual characteristics matter in their emergence.

To address this issue, this paper empirically analyzes the formation of a social network within an educational environment. We elicit data on social ties and combine this with a broad array of individual characteristics that have been measured prior to the start of the network formation process. We then take an exploratory approach to identify individual characteristics that predict emerging social ties and corresponding aggregated network structures by documenting associations between certain individual characteristics and social network outcomes. In addition, we test the possibility to affect social network structures by intervening in the process of network formation. To do so, we exploit exogenous variation in the social environment individuals have been exposed to.

Our study uses data from incoming university students. This is beneficial for analyzing the general patterns of network formation as the observed individuals all start out in a new environment and thus jointly create a new social network, as compared to a situation in which a stable social network already exists and only some new individuals enter. As a second novelty of our study, we are able to measure students' time and risk preferences and cooperativeness with experimental methods that have recently been shown to have strong predictive power for economic outcomes (Borghans et al., 2008; Heckman, 2011),<sup>2</sup> but are so far missing in existing social network data sets like AddHealth (Harris et al., 2009).

Our analysis consists of two parts. In the first part, we consider *local* network measures and individual link formation, i.e. how well different individual characteristics of two students serve as predictors of a link between these students. Individual characteristics enter

---

<sup>1</sup>Epple and Romano (2011) and Sacerdote (2011) provide overviews on the theory and evidence of peer effects in education. Further recent related studies include Burke and Sass (2013), Calvó-Armengol et al. (2009), Kremer et al. (2011), Patacchini et al. (2014).

<sup>2</sup>For example, recent studies show that individuals' time-preferences (Ventura, 2003; Kirby and Petry, 2004; Eckel et al., 2005) and cooperativeness (Rustagi et al., 2010; Fehr and Leibbrandt, 2011; Lamba and Mace, 2011) are important predictors for economic outcomes.

the analysis in three ways: (i) We look at the level of specific characteristics of “senders”, i.e. students who *nominate* the other student to be a friend. (ii) We look at the level of specific characteristics of “receivers”, i.e. students who *are nominated* to be a friend by the other student. (iii) We look at the difference of specific characteristics between the two respective students. This last set of explanatory variables allows investigating the extent of *homophily*, i.e. the tendency to form links among individuals with similar characteristics. As outcomes, we consider different aspects of link formation: For one, we distinguish between “out-degree” and “in-degree”. While the former refers to the number of friends an individual nominates, the latter refers to the number of nominations an individual receives (the individual’s *popularity* (Conti et al., 2013)). We also consider the difference between these two measures. Furthermore, we distinguish between the *extensive* and the *intensive* margin of friendship formation. The *extensive* margin refers to the relationship between different characteristics and the *existence* of links, while the *intensive* margin refers to the effect of these characteristics on the *intensity* of *existing* links. In the second part of our analysis, we examine the aggregated *global* network structure that results from these individual link formation patterns. Here we look at several measures that describe how central an individual is in the overall network and we analyze how this relates to its characteristics.

Our results can be summarized into three main findings. First, we find strong evidence for homophily along several dimensions. We complement the existing evidence by documenting homophily also for time preferences and cooperativeness. These measures from the experimental economics literature have so far been neglected in the analysis of social networks. In addition, we show that homophily is also present with respect to link intensity. Second, we are able to show that cooperativeness – as measured by behavior in the public goods game – plays a decisive role in the network formation process, as it predicts a variety of individual as well as aggregated network characteristics. Contrary to what one might expect, it turns out that more cooperative behavior is generally associated with less centrality and a lower number of social ties and popularity. However, there are some opposite effects of cooperativeness on link intensity. Third, we show that students who belong to the same (randomly composed) introductory groups during the introductory days have a strong tendency to form friendships. This implies that despite the robust and systematic patterns in network formation, social networks can still be exogenously affected by suitable - and potentially not even extensive - interventions. However, we also show that common group membership during the introductory days has no effect on link *intensity*.

From a policy perspective and given the existing evidence on the role of the social network for educational outcomes, our results are valuable in at least two ways. First, our findings on how social networks form improve our understanding of the sources of variation of academic performance across students: if specific characteristics affect study outcomes of connected students while at the same time increasing the likelihood to have friends with similar attributes, then this represents an additional indirect mechanism through which specific characteristics affect outcomes. Second, the fact that mere membership in the same introductory group has a strong impact on the probability to form social ties reveals that social networks are manipulable and that their structures are not at all predetermined. This suggests that peer effects can actually be exploited as a tool to improve outcomes by changing group compositions. Carrell et al. (2013), for example, assign freshmen to peer groups to maximize academic performance of the lowest ability students. However, they find that the “optimally designed” peer groups have a negative and significant treatment effect, because the intended to be treated students do not interact with the treating students. This highlights the importance of a deeper understanding of the actual friendship formation process for design of public policies.

Apart from its implications with respect to link formation and peer effects in education, our results provide further contributions. First, we extend the growing empirical literature on social network formation, confirming the phenomenon of homophily. Beyond simply replicating the existing evidence, our paper shows that the phenomenon of homophily also extends to behavioral measures elicited in the lab. Second, while link formation in long-lasting networks may be affected by, e.g., geographical restrictions and the sequentiality of the arrival of new individuals, our setup allows for an unconfounded view on link formation, because it analyzes a situation where a social network emerges “from the start”. Third, the combination of social network information with behavioral measures from the lab connects two recent strands of the literature that have, so far, been analyzed mainly in isolation. Thereby our paper also adds to the literature on the predictive power of behavioral lab measures for field behavior (Camerer, 2011), thus motivating their use in other field settings as well.

There are several papers which are closely related to ours. Currarini et al. (2009) provide empirical evidence for racial homophily in friendship formation. They find inbreeding homophily among racial groups at 84 US high schools. Members of relatively large groups have more friends per capita and inbreeding is strongest for middle sized groups. In their paper,

geographical proximity is also an important determinant of friendship formation. Marmaros and Sacerdote (2006) exploit a random allocation procedure to dormitories and roommates for freshmen in Dartmouth. They show that students are three times more likely to communicate with another student if that student lives in the same dorm and three times less likely to talk to students of another ethnical background given the same geographical distance. Mayer and Puller (2008) and Lewis et al. (2012) use data from Facebook to analyze network formation. While the former find that friendship formation depends on race and educational background the latter find that students who share certain tastes in music and movies are significantly more likely to befriend one another. Two early papers on homophily are Lazarsfeld and Merton (1954) as well as McPherson et al. (2001). The former discusses homophily with respect to social status, i.e. that the probability of mutual association is higher for individuals with similar social status than it would be if links were formed on a purely random basis. The latter describes many different dimensions and contexts in which homophily matters. Other dimensions of homophily in friendship formation are cognitive ability (Foster, 2005), educational background (Sacerdote, 2001), political affiliation (Huckfeldt and Sprague, 1995) and taste (Lewis et al., 2008).

Another strand of the literature investigates the impact of peer effects on economic behavior. Gandal et al. (2009) find that betweenness centrality is stronger correlated with revenues than any demographic or human capital factor. Bandiera et al. (2010, 2013) analyze the role of friendship in the workplace and document that sorting into working teams is strongly driven by friendships. Within these teams individual levels of productivity of friends depend on each other. However, introducing incentive schemes highlighting the importance of team production reduces the role of friendship for team formation. The role of social networks for behavior has further been documented for obesity (Christakis and Fowler, 2007), smoking (Christakis and Fowler, 2008), suicide attempts (Bearman and Moody, 2004), academic achievements (Sacerdote, 2001) and criminal activities (Glaeser et al., 1996). A third strand of literature that emerged recently consists of studies exploring the genetic influence on and the heritability of network characteristics like in-degree, transitivity and centrality (Fowler et al., 2009; Jackson, 2009). The authors infer that natural selection might have played a role for the evolution of social networks. In a follow-up study Fowler et al. (2011) analyze genotypic clustering in social networks.

The structure of our paper is as follows: in Section 2 we explain our setting as well as the measurement of individual and social network characteristics and describe the data

collection procedure. Section 3 presents our empirical strategy and the respective results. Section 4 summarizes and discusses the main results. Finally, Section 5 concludes.

## 2 Measurement and data collection

Our study is based on a sample of incoming students in the field of management and economics at the University of Mainz (JGU) in Germany in the winter term 2011/12. This setup has various advantages for our analysis: First, freshmen meet each other on a regular basis during the course of the first year and thus have ample opportunities to create links among each other. Second, the number of common links existing already prior to our study should be rather low, thus limiting potential endogeneity problems that are due to a shared history. Third, as it is typically not possible to elicit complete social networks including all in- and outgoing links of a group of individuals, the focus on the pool of incoming students and links within the university provides a reasonable boundary of a relevant sub-network to analyze. Finally, our setup allowed us to run experimental sessions at the very first meeting of freshmen students in October 2011 as well as nine months later in July 2012. As a result, we obtain data on two points in time: prior to the formation of any link and after the network has developed over the first two terms.<sup>3</sup>

The institutional environment in which we collected our data was as follows: a week before the start of courses, the department of management and economics at the University of Mainz (JGU) organizes an introductory event to welcome incoming students. This event involves speeches of the faculty and marks the beginning of two introductory days, organized by students from older cohorts. In these introductory days incoming students receive information about the department, their schedule and curriculum and, most importantly, are introduced to each other. These introductory days are carried out in smaller groups of up to thirty students to which students are randomly assigned by drawing lots. Each lot displays a different animal which then corresponds to the group the student belongs to, e.g. the “rabbit” group or the “monkey” group. We monitored this process to ensure that students are gathering in the groups prescribed by their respective lots. To confirm that this process was carried out properly, we additionally run a set of regressions to test whether there are significant between-group differences in member characteristics. Table 1 shows the

---

<sup>3</sup>For those students who only participated in the second experiment (approx. 20% of the sample) we used the personal data elicited in July 2012.



respective results and confirms that there are no substantial between-group differences.<sup>4</sup>

The experimental sessions took place right after the group assignment procedure was finished and before social interaction within groups took place. Data on social networks of students were elicited approximately 9 months after the introductory days within a second set of lab sessions. To minimize attrition, we incentivized participation in the second round by delaying payments from the first round to this second round (which subjects were informed about at the beginning of the first round) and an additional show-up fee of 10 Euros.

---

<sup>4</sup>Of 13 dependent variables we look at only for the regression of gender on group dummies we can reject the null hypothesis that all slope coefficients are jointly zero at the 5 percent level.

Table 1: Randomization check

	Male (1)	Age (2)	Partner (3)	Ravens IQ (4)	Alcohol (5)	Smoking (6)	Time Preferences (7)	Risk (8)	Contributor (9)	Free Rider (10)	Altruist (11)	Siblings Number (12)	Friends Number (13)
Group1	0.311* (0.158)	1.103* (0.614)	-0.297* (0.163)	-0.348 (0.351)	-0.057 (0.128)	-0.094 (0.124)	-0.909 (0.995)	2.242** (0.864)	0.015 (0.101)	0.015 (0.124)	-0.015 (0.033)	-0.113 (0.314)	6.919*** (2.016)
Group2	-0.107 (0.139)	0.097 (0.540)	0.031 (0.143)	0.148 (0.308)	-0.118 (0.112)	-0.118 (0.109)	-0.780 (0.874)	0.533 (0.759)	0.124 (0.088)	0.100 (0.109)	-0.015 (0.029)	-0.144 (0.276)	-0.221 (1.771)
Group3	-0.238** (0.112)	-0.209 (0.443)	0.111 (0.118)	-0.027 (0.257)	0.095 (0.092)	0.015 (0.090)	-1.258* (0.728)	0.208 (0.632)	-0.034 (0.074)	-0.042 (0.091)	0.027 (0.024)	-0.197 (0.227)	0.406 (1.455)
Group4	0.262* (0.147)	0.215 (0.573)	-0.108 (0.152)	0.169 (0.327)	0.138 (0.119)	0.354*** (0.116)	0.268 (0.927)	0.179 (0.805)	0.001 (0.094)	-0.013 (0.115)	-0.015 (0.031)	-0.400 (0.293)	2.338 (1.878)
Group5	0.293* (0.165)	-0.669 (0.640)	0.031 (0.170)	0.315 (0.365)	0.215 (0.133)	0.015 (0.130)	-2.591** (1.037)	-0.167 (0.900)	-0.076 (0.105)	-0.067 (0.129)	-0.015 (0.035)	-0.377 (0.327)	2.846 (2.100)
Group6	-0.174 (0.173)	-0.280 (0.670)	-0.014 (0.178)	-0.652* (0.382)	0.104 (0.139)	0.038 (0.136)	0.279 (1.085)	0.667 (0.943)	0.146 (0.110)	-0.167 (0.135)	-0.015 (0.036)	-0.144 (0.342)	-1.354 (2.199)
Group7	-0.144 (0.158)	0.376 (0.614)	0.158 (0.163)	0.015 (0.351)	0.124 (0.128)	-0.094 (0.124)	-0.121 (0.995)	0.879 (0.864)	-0.076 (0.101)	0.015 (0.124)	-0.015 (0.033)	-0.659** (0.314)	0.373 (2.016)
Group8	-0.150 (0.143)	0.188 (0.555)	-0.141 (0.147)	-0.199 (0.317)	0.073 (0.116)	-0.042 (0.112)	0.100 (0.899)	0.119 (0.780)	0.139 (0.091)	0.048 (0.112)	-0.015 (0.030)	-0.691** (0.284)	1.003 (1.822)
Group9	-0.007 (0.165)	0.231 (0.640)	0.031 (0.170)	-0.085 (0.365)	-0.085 (0.133)	-0.185 (0.130)	0.942 (1.037)	-0.767 (0.900)	0.124 (0.105)	0.133 (0.129)	-0.015 (0.035)	-0.277 (0.327)	-0.854 (2.100)
Group10	-0.046 (0.147)	-1.092* (0.573)	-0.031 (0.152)	-0.369 (0.327)	-0.092 (0.119)	0.046 (0.116)	0.499 (0.927)	-0.359 (0.805)	0.155* (0.094)	-0.013 (0.115)	-0.015 (0.031)	-0.400 (0.293)	0.415 (1.878)
Constant	0.507*** (0.059)	21.169*** (0.234)	0.569*** (0.062)	1.985*** (0.132)	0.785*** (0.049)	0.185*** (0.047)	6.091*** (0.376)	6.667*** (0.326)	0.076** (0.038)	0.167*** (0.047)	0.015 (0.013)	1.477*** (0.119)	6.354*** (0.767)
Observation	199	196	196	196	196	196	196	196	196	196	196	196	196
R <sup>2</sup>	0.105	0.055	0.042	0.039	0.052	0.083	0.066	0.052	0.060	0.026	0.018	0.054	0.080
F	2.206	1.083	0.804	0.747	1.022	1.678	1.302	1.023	1.173	0.487	0.347	1.054	1.618
Prob > F	0.0191	0.3772	0.6248	0.6798	0.4267	0.0887	0.2322	0.4255	0.3116	0.8969	0.9668	0.4001	0.1043

Notes: Robust standard errors; \*\*\* Significant at the 1 percent level. \*\* Significant at the 5 percent level. \* Significant at the 10 percent level.

## 2.1 Individual data

In our experimental sessions we collected a broad variety of individual data. Table 14 in the appendix provides a detailed variable description. On the one hand, we designed an extensive questionnaire to elicit socioeconomic characteristics, environmental factors and the big five personality traits (Costa and McCrae, 1992; Dehne and Schupp, 2007). On the other hand, we measured students' cognitive abilities, their level of cooperativeness and competitiveness as well as time and risk preferences using established methods from experimental economics. For every subject, one of the experiments was randomly determined to be payoff relevant. The corresponding measures will be discussed in the following.

Subjects' cognitive abilities were measured on the basis of the Raven's advanced progressive matrices test (Bors and Stokes, 1998). This IQ test is considered to be cultural-free as it is independent of language, reading or writing skills or other knowledge explicitly taught during formal education. The test consists of 8 items, each of which presents a full pattern of abstract figures. For each item, the subject must choose the one out of 8 presented options that correctly complete an incomplete pattern. Students were given two sample questions before they had to complete the 8 items. They were paid EUR 0.50 for each correct answer.

The individual's cooperativeness level was measured by a one-shot four person public goods game. Each subject was given an endowment of EUR 10.00 and had to decide how much to contribute to a common project that yields a return rate of 200 percent. Afterwards, total returns of the common project were equally split among the four group members. Thus, individual  $i$ 's payoff function was  $\Pi_i = 10 - c_i + 0.5 \sum_{j=1}^4 c_j$ . Subjects' cooperativeness was measured by their unconditional contribution ( $c_i$ ). Furthermore, using the strategy method (i.e. subjects had to decide how much to contribute depending on the level of the other group members), we were also able to classify subjects according to four contribution types, as in Fischbacher et al. (2001): conditional cooperators (contributing more the more the group members contribute), free-riders (contributing nothing irrespectively of the contribution level of the other group members), unconditional cooperators (always contributing the complete endowment) and non-classifiable subjects.

Our measure of competitiveness is closely related to the established Niederle and Vesterlund (2007) procedure. Participants had to do a real effort task which involved adding five two-digit numbers (Bartling et al., 2009).<sup>5</sup> Before the task, participants could choose be-

---

<sup>5</sup>Similar mathematical real effort tasks were used by Niederle and Vesterlund (2007) and Dohmen and Falk (2011).

tween a piece-rate and a tournament scheme as the payment scheme for their performance. The subjects' were classified as competitive when choosing the tournament scheme. Under the standard procedure, the student who chooses the tournament scheme, competes against another randomly matched anonymous student. The student with a higher (lower) performance in the real effort task wins a given prize of EUR 20.00 (EUR 0.00). In case of a tie the prize is split equally between students. If a student chooses the piece-rate scheme, he is paid EUR 0.50 for each correct answer. We slightly adjusted the standard procedure to get more variation in our competitiveness measure. Instead of offering a single piece-rate, we used a price-list-method where participants face a series of decisions of a similar structure. While the tournament prize remained constant over the course of the 10 decisions, we added a fixed payment to the piece-rate which increased from EUR 3.00 to 12.00. The level of this fixed payment at which an individual switched from the tournament to the piece-rate scheme was then our measure of competitiveness.

Risk preferences were elicited by asking subjects to make 15 binary decisions under uncertainty (Holt and Laury, 2002; Andersen et al., 2008). In each decision subjects had to choose between a lottery (winning EUR 20.00 or EUR 0.00 with probability 0.5) and a certain payment of EUR  $X$  with  $X \in \{3, 4, 5, \dots, 17\}$ . The level of  $X$  at which subjects switched from the certain payment to the lottery reflects subjects' certainty equivalent associated with the lottery (Dohmen et al., 2010).

Finally we measured time preferences. Participants had to make 15 binary choices between EUR 40.00 today and some larger amount ranging from EUR 40.00 to EUR 70.00 in 6 months (Frederick et al., 2002). One out of 24 students and one of the 15 binary decisions was randomly selected and paid accordingly. To avoid transaction costs associated with delayed payments, delayed payments were implemented via scheduled bank transfers. The reservation interest rate can be derived from the payment level at which a subject chooses the later payment.

## 2.2 Network data

We applied a questionnaire method to collect information on links between individuals.<sup>6</sup> There are several ways to elicit information on friendships and social networks through questionnaires: classifying individuals as friends can be based either on a roster or a free recall

---

<sup>6</sup>There are several ways to gather social network data (Prell, 2012). Advantages and disadvantages of questionnaires are discussed in Carrington et al. (2005).

procedure, respondents can be forced to name a specific number of friends, and nominations can either be binary or value-weighted. A roster, i.e. a list of relevant participants in the previously defined network, is useful if the network boundaries are well defined and the number of actors is processable for the respondents. As in our case this would refer to all incoming students, i.e. a list of several hundred names, a roster approach seems unfeasible and we rely on free recall.

A specific trade-off exists with respect to the question of whether to allow respondents to list as many names as they want or to impose a fixed upper limit. The argument against the former is that with an increasing number of potential nominations participants feel obliged to list more friends, potentially including subjects that do not fit the typical definition of a friend (Coleman et al., 1966). On the contrary, a fixed upper limit of nominations potentially causes measurement error due to the possibility of incomplete or inaccurate lists (Wasserman and Faust, 1994). As a compromise, we decided to run a fixed upper limit procedure with a rather high limit of ten friendship nominees. As our results show, no subject named more than nine matchable friends, indicating that our design had no negative effect on data quality. In order to also distinguish between differences in the strength, the frequency of interaction or the duration of a social interaction we asked each student “*Whom of the other students do you consider as your friend?*” (binary) in combination with a 4-point Likert scale asking the students to indicate whether the nominated person is *considered as a 'university-acquaintance' or as a close friend.*<sup>7</sup> Our questionnaire can be found in the appendix.

## 2.3 Network centrality measures

### Local network measures

In social network analysis the subject of interest is called a node and a connection between two nodes is called either a link or a tie. A tie  $(g_{i,j})$  is assumed to be present if subject  $i$  names subject  $j$  as a friend. Otherwise, it is assumed that there is no tie between the two

---

<sup>7</sup>Other studies, analyzing centrality of different agents within a network, have shown that aggregated network characteristics are stable over different interview rounds and the variation in individual's nominations is small (Burt et al., 1985).

nodes in this direction<sup>8</sup>

$$g_{i,j} = \begin{cases} 1 & \text{if } i \rightarrow j, \\ 0 & \text{otherwise.} \end{cases}$$

A natural starting point for the analysis of a social network is the direct neighbourhood (1<sup>st</sup> degree links) of every individual. However, even for this simple environment there exist already three different local measures. With subject  $i$  as the person of interest, out-degree measures how many friends are named by subject  $i$ , in-degree counts how many subjects name subject  $i$  as a friend (which will later be interpreted as “popularity”) and two-sided-degree represents the number of reciprocated links, where both sides name each other as a friend.

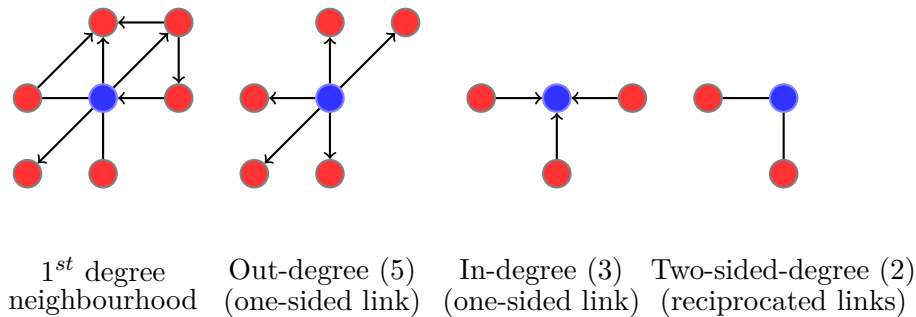


Figure 1: Illustration of degree centrality measures

Notes: Blue node is the subject of interest. Red nodes represent the corresponding 1<sup>st</sup>-degree (direct) neighbors.

Figure 1 illustrates the differences between these measures. The left network graph displays all connections of the blue node, without considering the different nature of the respective tie. Reciprocated (undirected) links are represented by a straight line. An arrow indicates a directed link and points from the nominator to the nominee. If  $i$  names  $j$  as a friend then the arrow points from  $i$  to  $j$ . The two network graphs in the middle show only one-sided links. These are connections between nodes that only consider one direction. In the application to friendship data, this implies that only one of the two respective individuals considered their relationship to be a friendship. The right network graph shows only links

<sup>8</sup>A connection in the other direction, i.e. from individual  $j$  to individual  $i$  would be denoted as  $g_{j,i}$ .

that are reciprocated, i.e. where both individuals consider the link to exist (two-sided-degree).<sup>9</sup>

## Global network measures

Local network measures only consider the direct social environment of an individual and ignore the aggregated network structure, i.e. links of second and higher orders. As a result, the role and position of an individual with respect to the entire social network remains unclear. Therefore, the empirical analysis of social networks is complemented by *global* network centrality measures. These measures focus on the centrality of an individual.

There are three prominent global network (centrality) measures that can be interpreted in an economically meaningful way (Jackson, 2008):<sup>10</sup>

- Closeness centrality

$$C_i^C = \left( \sum_{i \neq j} \frac{l(i, j)}{n-1} \right)^{-1}$$

$l(i, j)$  measures the minimal number of links required for  $i$  to reach  $j$ . Closeness centrality is thus defined as the inverse of the average distance to another node in the network. Thus,  $C_i^C$  captures the ease of a node to reach others.

- Betweenness centrality

$$C_i^B = \sum_{k \neq j, i \notin \{k, j\}} \frac{P_i(kj)/P(kj)}{(n-1)(n-2)/2}$$

$P_i(kj)$  indicates whether  $i$  lies on the shortest path between node  $k$  and  $j$ . If there are several shortest paths in terms of number of links between  $k$  and  $j$  then  $P_i(kj)$  is weighted by the total number of shortest paths between  $k$  and  $j$ , denoted by  $P(kj)$ . Betweenness centrality thus captures how important a given node is as an intermediary.

- Eigenvector centrality

$$\lambda C_i^E(g) = \sum_j g_{ij} C_j^E(g)$$

---

<sup>9</sup>This classification is in line with, for example, the design of the AddHealth Data set (Moody, 2001; Jackson, 2009), which is commonly used in the social network literature.

<sup>10</sup>A fourth prominent centrality measure is the Bonacich centrality measure. It is omitted in our analysis because the Eigenvector centrality measure is special case of the Bonacich centrality measure.

Table 2: Centrality measure differences

	Node 1	Node 3	Node 4
Degree	.33	.50	.33
Closeness	.40	.55	.60
Betweenness	.00	.50	.60
Eigenvector	.47	.63	.54

The idea is that the centrality of a node  $i$ ,  $C_i^E(g)$ , is proportional to the sum of the centrality of its neighbors  $j$  in a network  $g$ . In matrix notation this becomes  $\lambda C^E(g) = gC^E(g)$  with  $C^E(g)$  being the eigenvector and  $\lambda$  the corresponding eigenvalue. Following Katz’s (1953) definition, eigenvector centrality proxies  $i$ ’s influence by measuring the centrality of  $i$ ’s neighbors. Thus, it captures the ideas that “it is not important what you know, but who you know” (Bush and Hattery, 1951, p. 456).

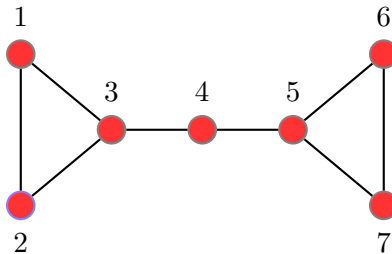


Figure 2: Illustration of global centrality measures

Notes: Adapted network from Jackson (2008, p. 38).

To illustrate the differences in centrality measures, a stylized network with the corresponding measures is shown in Figure 2 and Table 2. Node 3 is most central in terms of degree centrality. The degree of node 1 and 4 is identical although their position in the network seems to differ substantially. These differences are captured by global centrality measures. Node 4 lies most often on the shortest path between any two nodes and therefore has the highest betweenness measure. On average it is also closest to other nodes. However, in terms of prestige, node 3 has the highest eigenvector centrality score.

In the remainder of the analysis of our data, we will focus on the “largest” component of our network when we consider global centrality measures.<sup>11</sup> There are two reasons for

<sup>11</sup>A component refers to a subgraph where every node can reach and be reached through paths in the



this: first, as the distance between two unconnected nodes or any components is not defined, network centrality measures like closeness and eigenvector cannot be calculated. Second, the limited comparability of network measures across components can lead to biases (Wasserman and Faust, 1994). For example, if we consider the betweenness measure of nodes in two disconnected components, differences in the overall size of the component should directly influence the average value of betweenness. It is therefore common to focus on the largest component of a network. However, doing so reduces our network sample from 199 to 144 subjects in the analysis of global centrality measures.

### 3 Empirical analysis

#### 3.1 The microstructure of social networks: empirical approach

We start by investigating the determinants of individual link formation, i.e. we look at the microstructure of the social network. In doing so, we use a variety of dependent variables to distinguish different dimensions of link formation.

As a first step, we generate a binary variable for all possible combinations of  $i$  and  $j$  in our sample in both directions and thus all *possible* links.<sup>12</sup> The variable takes the value of one if a link from individual  $i$  to  $j$  exists, i.e. individual  $i$  has nominated individual  $j$  as a friend. We refer to this as the *extensive margin* of link formation. The conditional probability of an existing link can now be modeled as

$$p_i = Pr(g_{i,j} = 1|X) = F(X'\beta)$$

where  $X$  is the matrix of individual characteristics of subjects in our sample. Theoretically, an individual's decision to invest into a relationship depends on the utility that stems from forming this link. If this utility  $u_{i,j}$  exceeds a certain threshold  $u^*$  the link is formed:

$$g_{i,j} = \begin{cases} 1 & \text{if } u_{i,j} > u^*, \\ 0 & \text{if } u_{i,j} \leq u^*. \end{cases}$$

---

network. The largest component refers to the component with the maximal number of nodes. See also Figure 8 in the appendix.

<sup>12</sup>This means that the combination of individual  $i$  and  $j$  shows up two times, once for either possible direction of a link.

Table 3: Descriptive statistics of friendship valuation (link measures)

	Obs.	Mean	Std. Dev.	Min,	Max.
Binary Link	39402	0.01	0.12	0.00	1.00
Link Value	39402	0.04	0.34	0.00	4.00
Link Value 2	39402	0.02	0.20	0.00	2.00

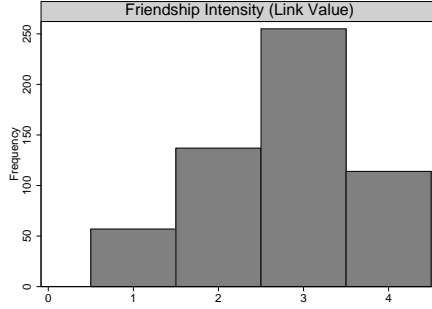


Figure 3: Friendship valuation distribution (link value)

Thus, we model the utility of individual  $i$  from forming a link with individual  $j$  as the latent variable  $u_{i,j}$  and assume it to be linear in individual characteristics, i.e.  $u_{i,j} = X' \beta + \epsilon_i$ . Given the dichotomous nature of the dependent variable we apply logit and probit regressions to estimate our relationship of interest. We also employ the linear probability model.<sup>13</sup>

In addition, we also use the stated intensity of links as a dependent variable. Here, 0 refers to a non existing link, while the values 1-4 describe the intensity of existing links where a higher value corresponds to a stronger link. In our analyses, we refer to the intensity of links conditional on these links being formed as the *intensive margin* of link formation. We use ordered probit models as well as OLS.<sup>14</sup>

Individual characteristics of potential friends enter our regressions in three ways: first, we include sender characteristics ( $X_i$ ), i.e. information about the student who nominates the other one as a friend. Second, we include receiver characteristics ( $X_j$ ), i.e. information about the nominated student. Finally, we include the absolute difference of a specific characteristic between two potential friends as a measure of their similarity ( $|\Delta X_{ij}|$ ). Descriptive statistics on individual characteristics are provided in Table 4 . Overall, the general model looks as

<sup>13</sup>In addition, we use a rare event logistic regression model developed by King and Zeng (2001). This controls for the relatively low number of actual friendships compared to the number of observations in the data set. We do not explicitly report these results as they do not differ from our baseline results.

<sup>14</sup>See Table 16 and Table 17 in the appendix for the respective results.

follows:

$$g_{ij} = \beta_0 + \beta_1 X_i + \beta_2 X_j + \beta_3 |\Delta X_{ij}| + \epsilon_i.$$

Table 4: Summary statistics of individual measures

	Obs.	Mean	Std. Dev.	Min.	Max.
<i>Behavioral Measures</i>					
TimePreferences	196	5.81	3.08	0.00	15.00
RiskPreferences	196	6.89	2.65	0.00	15.00
Cooperativeness	196	4.77	2.96	0.00	10.00
Competitiveness	196	4.02	2.73	0.00	10.00
<i>Big Five</i>					
Agreeableness	196	50.34	9.61	21.53	69.92
Conscientiousness	196	50.39	9.31	13.43	68.91
Extraversion	196	50.07	9.98	25.32	77.78
Neuroticism	196	50.14	9.95	19.18	79.64
Openness	196	50.54	8.71	20.25	68.01
<i>Personal Characteristics</i>					
Male (d)	199	0.49	0.50	0.00	1.00
Age	196	21.15	1.89	18.00	30.00
RavensIQ	196	1.93	1.07	0.00	7.00
Partner (d)	196	0.56	0.50	0.00	1.00
Smoking (d)	196	0.18	0.39	0.00	1.00
Alcohol (d)	196	0.81	0.39	0.00	1.00
LivingAlone (d)	199	0.26	0.44	0.00	1.00
LivingSharedFlat (d)	199	0.35	0.48	0.00	1.00
SiblingsNo.	196	1.26	0.96	0.00	6.00
SiblingRank	196	1.60	0.81	0.00	6.00
RelativesNo.	196	10.13	9.02	0.00	80.00
FriendsNo.	196	7.09	6.28	1.00	50.00
ParentsKM100	177	1.41	3.88	0.00	38.00
Relocation (d)	196	0.57	0.50	0.00	1.00
Participate (d)	199	0.66	0.47	0.00	1.00

Notes: The number of observations reduce by three whenever using experimental measures, because there was one session with an insufficient number of participants to run the experiment. Therefore we collected only network information from these four students.

After these analyses, we turn to local centrality measures, thus switching from student pairs to individual students as observational units. Here we use in-degree, out-degree and two-sided-degree as dependent variables. We estimate these models using OLS.<sup>15</sup>

<sup>15</sup>Given the nature and empirical distribution of the dependent variables (see Figure 4), count data models might be better suited for the analysis of degree centrality. Hence, we also run Poisson and negative binomial regression models. However, given that our results are similar to those in our baseline regressions, we report these results only in the Appendix (see Appendix Table 18).

Table 5: Correlation of characteristics among friends

Variable	Correlation	p-value
Gender	0.3807	0.0000
Nationality	0.2507	0.0000
TimePreferences	0.1791	0.0000
RiskPreferences	0.0891	0.0346
Cooperativeness	0.1090	0.0096

## 3.2 The microstructure of social networks: results

Before we turn to regression analysis, we first check whether simple descriptive statistics already indicate the existence of homophily with respect to behavioral measures taken from the lab. In Table 5 we thus simply look at the correlation of characteristics within the sample of existing links. Confirming results from the literature, we see a strong correlation with respect to gender and nationality. In addition, we see that behavioral measures from the lab also correlated significantly.

Now we turn to regression analysis, Table 16 shows the respective results. First, columns (1) and (2) use the intensity of links as a dependent variable, ranging from 0 (no link) to 4 (strongest possible link). This measure combines the extensive and intensive margin, thus providing a natural starting point of the analysis.

Looking at differences in characteristics of students, note that a negative coefficient implies homophily: the higher the absolute difference between two individuals in a specific characteristic, the less intense a link is. Accordingly, we see that friendship formation in our setting is strongly characterized by homophily. We observe homophily with respect to gender, time preferences, and cooperativeness. In addition, we see a strong effect of common group membership during the introductory days.<sup>16</sup> When we look at how sender and receiver characteristics affect friendship formation, we find that time preferences (being more patient) as well as cooperativeness have a negative effect on link intensity for both, senders and receivers.

### 3.2.1 Distinguishing the extensive and intensive margin of link formation

Now we analyze to what extent these aggregated effects of individual characteristics work through the extensive or the intensive margin. Regarding the extensive margin, we compare

---

<sup>16</sup>The variable *Group* is 1 if the respective individuals were in different introductory groups and 0 if they were in the same group.

Table 6: Friendship formation (OLS)

	Link Value [0,...,4]		Binary Link [0;1] (Extensive Margin)		Link Value [1,...,4] (Intensive Margin)	
	(1)	(2)	(3)	(4)	(5)	(6)
<i>Absolute Differences (<math> X_i - X_j </math>)</i>						
Gender	-0.0316*** (0.0035)	-0.0294*** (0.0034)	-0.0110*** (0.0012)	-0.0103*** (0.0012)	-0.1270 (0.0854)	-0.1489* (0.0874)
Participate	-0.0069** (0.0033)	0.0852*** (0.0103)	-0.0028** (0.0011)	0.0301*** (0.0035)	0.1014 (0.0907)	0.1552 (0.1271)
RiskPreferences	0.0003 (0.0007)	-0.0004 (0.0007)	0.0002 (0.0002)	-0.0000 (0.0003)	-0.0274* (0.0145)	-0.0284* (0.0159)
TimePreferences	-0.0027*** (0.0006)	-0.0026*** (0.0006)	-0.0010*** (0.0002)	-0.0010*** (0.0002)	0.0048 (0.0154)	0.0010 (0.0156)
RavensIQ	-0.0015 (0.0016)	-0.0027 (0.0018)	-0.0007 (0.0005)	-0.0013** (0.0006)	0.0428 (0.0364)	0.0315 (0.0438)
Cooperativeness	-0.0024*** (0.0006)	-0.0024*** (0.0006)	-0.0009*** (0.0002)	-0.0009*** (0.0002)	-0.0075 (0.0161)	-0.0190 (0.0160)
Group	-0.0622*** (0.0072)	-0.1924*** (0.0190)	-0.0230*** (0.0025)	-0.0695*** (0.0065)	0.0574 (0.0921)	-0.0310 (0.1034)
<i>Sender Characteristics (<math>X_i</math>)</i>						
Male		-0.0021 (0.0042)		0.0016 (0.0015)		-0.4355*** (0.1078)
Participate		0.0913*** (0.0099)		0.0332*** (0.0034)		-0.0545 (0.1079)
RiskPreferences		-0.0018* (0.0009)		-0.0006* (0.0003)		-0.0108 (0.0220)
TimePreferences		-0.0018*** (0.0006)		-0.0006*** (0.0002)		-0.0045 (0.0125)
Cooperativeness		-0.0019*** (0.0006)		-0.0004** (0.0002)		-0.0573*** (0.0165)
<i>Receiver Characteristics (<math>X_j</math>)</i>						
Male		-0.0024 (0.0042)		0.0001 (0.0015)		-0.0220 (0.0982)
Participate		0.0879*** (0.0100)		0.0310*** (0.0034)		0.1922* (0.1027)
RiskPreferences		0.0000 (0.0009)		-0.0000 (0.0003)		0.0103 (0.0226)
TimePreferences		-0.0011* (0.0006)		-0.0003* (0.0002)		-0.0032 (0.0127)
Cooperativeness		-0.0025*** (0.0006)		-0.0009*** (0.0002)		0.0114 (0.0170)
Further Controls	yes	yes	yes	yes	yes	yes
Observations	39402	39402	39402	39402	563	563
R <sup>2</sup>	0.0088	0.0214	0.0099	0.0234	0.0273	0.1406

Notes: Robust standard errors in parentheses. \*\*\* Significant at the 1 percent level. \*\* Significant at the 5 percent level. \* Significant at the 10 percent level.

results from using the link intensity (from 0-4) as a dependent variable to results from using only a binary variable. Columns (3) and (4) show how results differ across these specifications. While the size of these effects is not directly comparable, we at least see that all effects go into the same direction and show similar significance levels, no matter which of the two dependent variables is used. To see whether the effect of individual characteristics on the intensive margin is also similar, we now restrict our sample by dropping all observations where link intensity is zero, i.e. where no link exists, and use the link intensity, now ranging from 1-4, as a dependent variable (see columns (5) and (6)). The results from these regressions can now be interpreted as the effect of individual characteristics on pure link *intensity* as it conditions on a link being formed.

While some effects carry over, this analysis also reveals some striking differences in the effects specific individual characteristics have on the intensive and the extensive margin. To start with similarities, we see that differences in gender as well as in risk preferences are, albeit just weakly, significantly associated with link intensity. This indicates that homophily also applies with respect to intensity and not merely to the existence of links. This neither holds true for the effect of common group membership during the introductory days, nor for time preferences or cooperativeness. While these characteristics have a substantial effect on the formation of links, it is completely unrelated to the subsequent intensity of these links, i.e. it purely affects the extensive margin of link formation. Moving on to sender characteristics, we see that while being male was not significantly associated with forming a link, it has a large and statistically significant negative association with link intensity. This means that women at least perceive their links to be more intense than men. Arguably the most interesting results refer to the role of cooperativeness. While cooperativeness of senders as well as receivers is negatively associated with the existence of links, there is an asymmetric effect when it comes to the mere intensity of these links. Here we see that while cooperative students claim their friendships to be on average less intense, this perception is not mirrored by receivers. In contrast, if anything, students perceive links to more cooperative individuals to be stronger, albeit not statistically significant.

### 3.2.2 Popularity and other local degree centrality measures

Now, we use local centrality measures as dependent variables. These regressions allow identifying characteristics associated with *popularity* in a group, i.e. a high number of other students naming a particular individual as a friend (in-degree), but also what drives the

number of friends an individual claims to have (out-degree). Finally, we also look at the difference between out-degree and in-degree. Figure 4 and Table 7 show the corresponding distribution and descriptive statistics of different (local) degree centrality measures. All three degree centrality measures exhibit a positive skewness, which is a common property of count data sets. Average in- and out-degree are both 2.55 and average two-sided-degree is 1.78.<sup>17</sup>

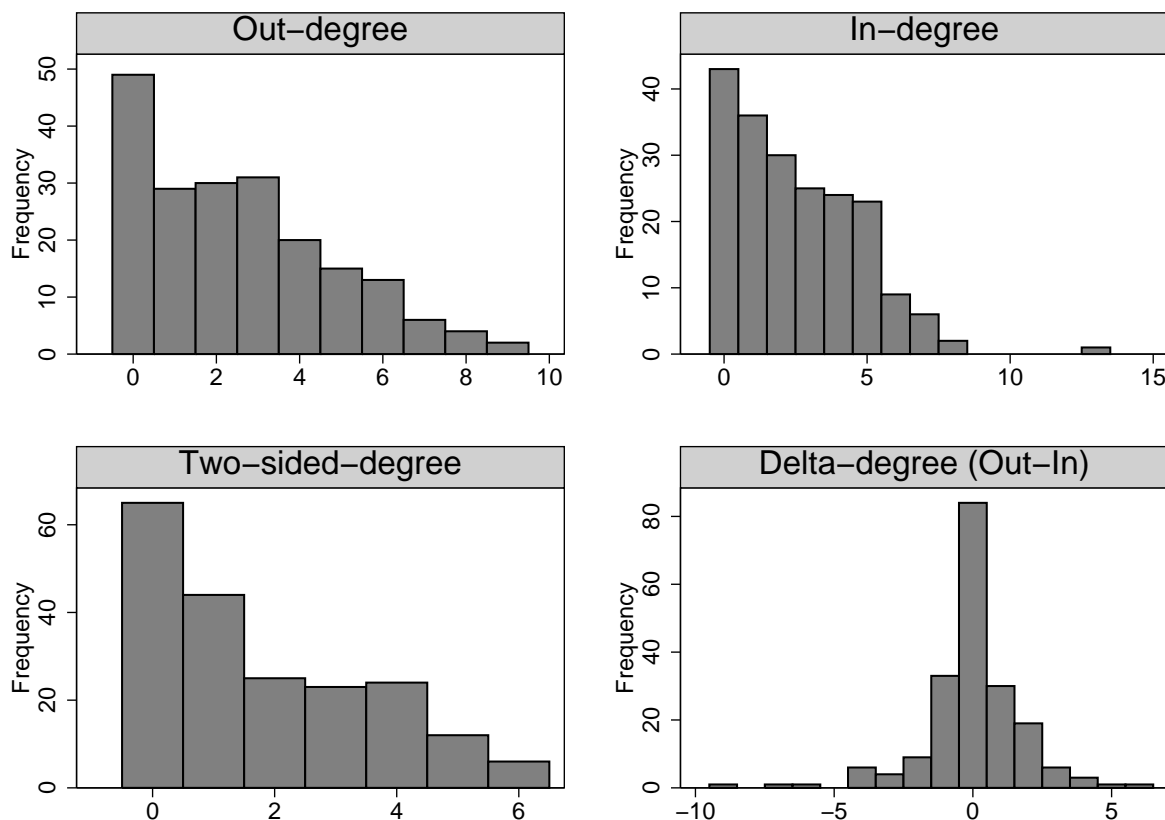


Figure 4: Distribution of Local Centrality Measures

Table 10 depicts the respective regression results. We see a significant effect of participating in the introductory days. This result is as expected: first, given that this event is meant to let students get to know each other, it seems intuitive that these students are also more often mentioned as friends of other students and, second, showing up at this event might

<sup>17</sup>Individual differences between out- and in-degree (delta-degree, not shown) are approximately normally distributed with a mean of zero, indicating that the individuals' perceptions of their connectedness (degree) is on average unbiased.

Table 7: Summary statistics of local centrality measures

	Obs.	Mean	Std. Dev.	Min.	Max.
Out-degree	199	2.55	2.26	0.00	9.00
In-degree	199	2.55	2.21	0.00	13.00
Two-sided-degree	199	1.78	1.77	0.00	6.00
Delta-degree	199	0.00	1.79	-9.00	6.00

generally reflect a higher level of “sociability” and thus represent a selection effect. More importantly, we see a statistically highly significant association of cooperativeness with the probability to be nominated as a friend. However, maybe contrary to what one might initially expect, this effect is negative, implying more cooperative individuals being nominated less often. We also find that being older and being more patient (time preferences) reduce the number of incoming friendship nominees. In turn, being in a relationship, agreeableness and conscientiousness are associated with higher numbers of friendship nominees and thus popularity. Looking at out-degree as well as two-sided-degree, we see a similar picture. Only the negative effect of cooperativeness on out-degree turns statistically insignificant, although remaining negative. Considering delta-degree we do not find strong effects.

### 3.3 The macrostructure of social networks

#### 3.3.1 Aggregated characteristics of the social network

We now turn to the analysis of the relation between individual characteristics and the *macrostructure* of the social network. Figure 5 depicts the structure of the social network in our data, based on an energy algorithm.<sup>18</sup>

Table 9 and Figure 6 illustrate the respective distributions of global network centrality measures. Average closeness, betweenness and eigenvector centrality measures are 0.018, 0.03 and 0.03. An average closeness of 0.2 indicates that every two students are approximately five links apart from each other. The exact degree of separation in the network is 5.26. Due to the right skewed distribution, average betweenness is rather low as most people are at the edge of a component and have only few links. In addition, subjects with only one link have

<sup>18</sup>There are two classes of network representation: circular and energy. While the circular algorithm positions vertices on a circle, the energy algorithm has a main advantage in revealing systematic patterns, because it positions nodes according to two objectives: (i) edges are of more or less equal length and (ii) there are as few crossing edges as possible. The sometimes called spring algorithm for drawing graphs is developed by Kamada and Kawai (1989) (Kamada-Kawai algorithm). A similar spring algorithm is also developed by Fruchterman and Reingold (1991).



Table 8: Local network centrality measures (OLS)

	Out-degree (1)	In-degree (2)	Two-sided-degree (3)	Delta-degree (4)
<i>Personal Characteristics</i>				
Male	-0.182 (0.346)	-0.293 (0.368)	-0.463 (0.295)	0.111 (0.286)
Age	-0.148* (0.076)	-0.188** (0.081)	-0.089 (0.062)	0.040 (0.066)
Partner	1.095*** (0.297)	0.647** (0.308)	0.945*** (0.228)	0.448* (0.268)
RavensIQ	-0.072 (0.131)	-0.083 (0.134)	-0.036 (0.120)	0.011 (0.103)
Alcohol	-0.017 (0.349)	0.214 (0.335)	0.140 (0.298)	-0.231 (0.284)
Smoking	-0.485 (0.383)	0.508 (0.502)	-0.470 (0.300)	-0.994** (0.482)
<i>Behavioral Measures</i>				
TimePreferences	-0.120** (0.049)	-0.112** (0.050)	-0.097** (0.040)	-0.008 (0.045)
RiskPreferences	0.025 (0.053)	-0.000 (0.085)	0.034 (0.043)	0.025 (0.072)
Cooperativeness	-0.080 (0.053)	-0.167*** (0.053)	-0.096** (0.042)	0.087* (0.047)
Competitiveness	0.152 (0.164)	0.156 (0.151)	0.110 (0.124)	-0.005 (0.131)
<i>Big Five</i>				
Agreeableness	0.050*** (0.016)	0.049*** (0.015)	0.032** (0.013)	0.001 (0.013)
Conscientiousness	0.036** (0.017)	0.032** (0.016)	0.036*** (0.013)	0.004 (0.013)
Extraversion	-0.012 (0.015)	-0.028* (0.015)	-0.008 (0.012)	0.016 (0.013)
Neuroticism	-0.020 (0.015)	-0.013 (0.016)	-0.020 (0.012)	-0.007 (0.012)
Openness	-0.031 (0.020)	-0.041* (0.022)	-0.037** (0.018)	0.010 (0.019)
<i>Contextual Factors</i>				
LivingSharedFlat	0.247 (0.379)	0.739** (0.366)	0.268 (0.286)	-0.492 (0.341)
FriendsNo.	-0.004 (0.030)	-0.005 (0.021)	-0.012 (0.018)	0.001 (0.026)
Participate	1.516*** (0.277)	0.964*** (0.302)	0.843*** (0.241)	0.552** (0.262)
Further controls	yes	yes	yes	yes
Observations	195	195	195	195
$R^2$	0.347	0.329	0.357	0.147

Notes: Robust standard errors in parentheses; \*\*\* Significant at the 1 percent level. \*\* Significant at the 5 percent level. \* Significant at the 10 percent level.

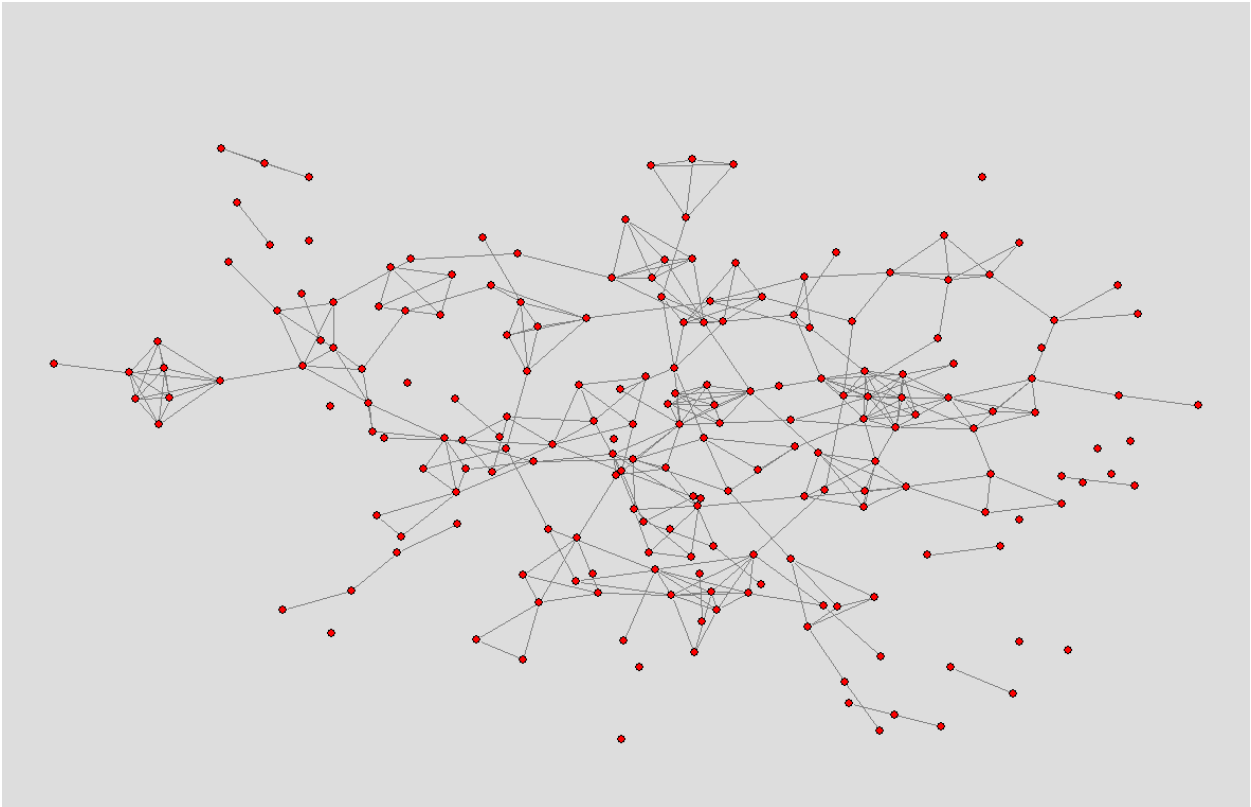


Figure 5: Global network of JGU business and economics student

Notes: The figure was created by *Pajek* (de Nooy et al., 2005).

a betweenness centrality of 0. A comparison of the most central individual to the median of betweenness (0.019) reveals that he or she is almost ten times more often on that shortest path (0.18).

Table 9: Summary statistics of global centrality measures

	Obs.	Mean	Std. Dev.	Min.	Max.
Closeness	144	0.18	0.03	0.11	0.25
Betweenness	144	0.03	0.03	0.00	0.18
Eigenvector	144	0.03	0.07	0.00	0.35

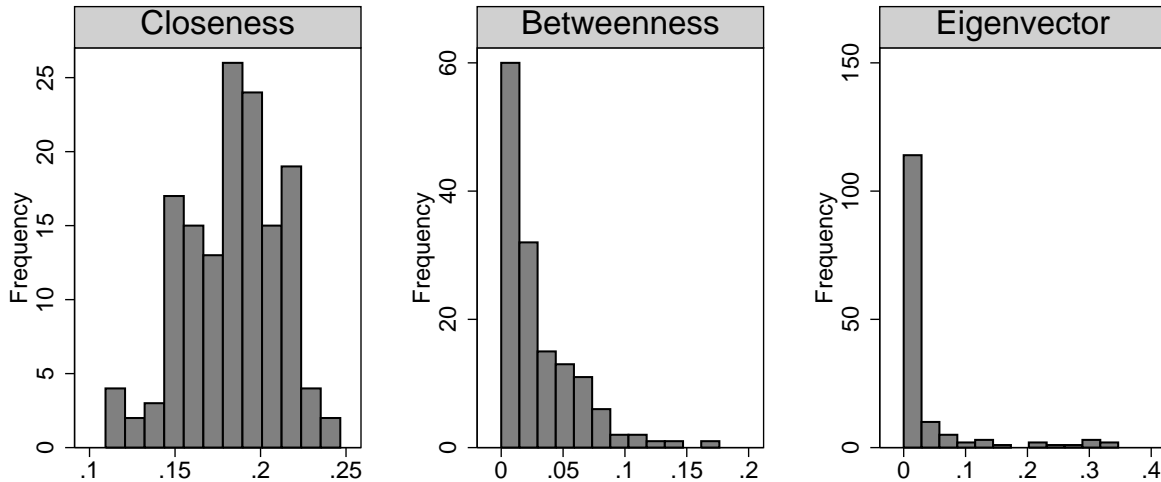


Figure 6: Global degree centrality measures

### 3.3.2 Individual characteristics and global network centrality

We now analyze how global network centrality measures correspond to individual characteristics. As discussed previously, we restrict the analysis to the largest component of the network. Table 19 in the appendix shows which characteristics predict being in the largest component. On average a younger male student who participated during the introductory days with lower patience and more confidence is more likely to be part of the biggest component. We control for personal characteristics (male, age, partner, Ravens IQ, alcohol- and cigarette consumption), behavioral measures (risk-, time preferences, cooperativeness and competitiveness), big five personality traits and contextual factors (living situation, number of family members, siblings and non-university friends, whether a student participated in the introductory days, the corresponding group size and two proxies for the distance to student's previous home with relocation and ParentsKM100).<sup>19 20</sup>

<sup>19</sup>In order to control for the current living situation, we asked a student if he (still) lives at the parent's home (default), alone (*LivingAlone*) or in a shared flat (*LivingSharedFlat*).

<sup>20</sup>For representational purposes, the table does not report cooperation types, a proxy for financial constraints, geographical distance to their family (ParentsKM100), the self-reported number of family members

Table 10 shows the empirical results from OLS regressions of different global centrality measures (closeness, betweenness and eigenvector centrality) on characteristics of individuals.<sup>21</sup> As the different centrality measures are very heterogeneous in scales, we report standardized coefficients to ease comparability. Coefficients shown refer to the number of standard deviations the dependent variable changes due to a one standard deviation change of the regressor.

Overall, individual characteristics do not serve as particularly good predictors of global centrality. Agreeableness is the only personality trait that has a positive and significant effect on all three global centrality measures, suggesting that increases in agreeableness-scores are associated with higher centrality in terms of closeness and betweenness as well as the tendency to be more influential (eigenvector). Students scoring higher on *RavensIQ* and competitiveness also seem to be more important as a connector between other students (betweenness). Smokers tend to be more central and cooperative students tend to be less influential. The dummy variables *LivingAlone* and *LivingSharedFlat* indicate whether a student lives alone or in a shared flat with living at home as the default living situation. Students living in a shared flat are more (less) central in terms of closeness (eigenvector) centrality. Finally, reporting a larger non-university friendship network is positively associated with being closer to other students and being more likely to act as a broker between students (Burt, 2005).

### 3.4 A closer look at cooperativeness

The results obtained so far document cooperativeness to be a key predictor for friendship formation in various dimensions. However, the previous analysis considers cooperative behavior only with respect to contribution levels in the unconditional public goods game. A second dimension of cooperative behavior that has been stressed in the literature (Fischbacher et al., 2001) is the conditionality of cooperation, i.e. if and how individuals condition their behavior on others players' behavior. We now look at how these differences in cooperation types are associated with friendship formation.

Three types are commonly distinguished: Free-riders, who never cooperate no matter what other players do, unconditional cooperators (altruists), who always contribute no matter what other players do, and conditional cooperators, who only contribute if other players

---

and the group size during the introductory days.

<sup>21</sup>Again, we run count data models as a robustness check and confirm the results.

Table 10: Global network centrality measures (OLS)

	Closeness (1)	Betweenness (2)	Eigenvector (3)
<i>Personal Characteristics</i>			
Male	0.005 (0.006)	0.002 (0.006)	0.011 (0.019)
Age	-0.001 (0.002)	-0.001 (0.002)	0.003 (0.004)
Partner	0.007 (0.005)	-0.004 (0.006)	0.012 (0.014)
RavensIQ	-0.001 (0.002)	0.005* (0.003)	-0.006 (0.004)
Alcohol	0.000 (0.007)	-0.004 (0.008)	0.027* (0.015)
Smoking	0.010 (0.006)	0.017* (0.009)	0.016 (0.016)
<i>Behavioral Measures</i>			
TimePreferences	-0.001 (0.001)	0.000 (0.001)	-0.001 (0.002)
RiskPreferences	-0.001 (0.001)	-0.000 (0.001)	0.001 (0.003)
Cooperativeness	-0.000 (0.001)	0.001 (0.001)	-0.005** (0.002)
Competitiveness	0.004 (0.003)	0.006** (0.003)	-0.000 (0.010)
<i>Big Five</i>			
Agreeableness	0.001** (0.000)	0.001** (0.000)	0.002** (0.001)
Conscientiousness	-0.000 (0.000)	-0.001 (0.000)	0.000 (0.001)
Extraversion	0.000 (0.000)	0.000 (0.000)	-0.000 (0.001)
Neuroticism	-0.000 (0.000)	0.000 (0.000)	0.001 (0.001)
Openness	-0.000 (0.000)	-0.000 (0.000)	-0.001 (0.001)
<i>Contextual Factors</i>			
LivingSharedFlat	0.011* (0.006)	0.008 (0.007)	-0.027 (0.017)
FriendsNo.	0.001** (0.000)	0.001 (0.000)	0.001 (0.001)
Participate	0.007 (0.006)	0.008 (0.006)	0.015 (0.013)
Further controls	yes	yes	yes
Observations	141	141	141
$R^2$	0.286	0.217	0.222

Notes: Robust standard errors in parentheses; \*\*\* Significant at the 1 percent level. \*\* Significant at the 5 percent level. \* Significant at the 10 percent level.

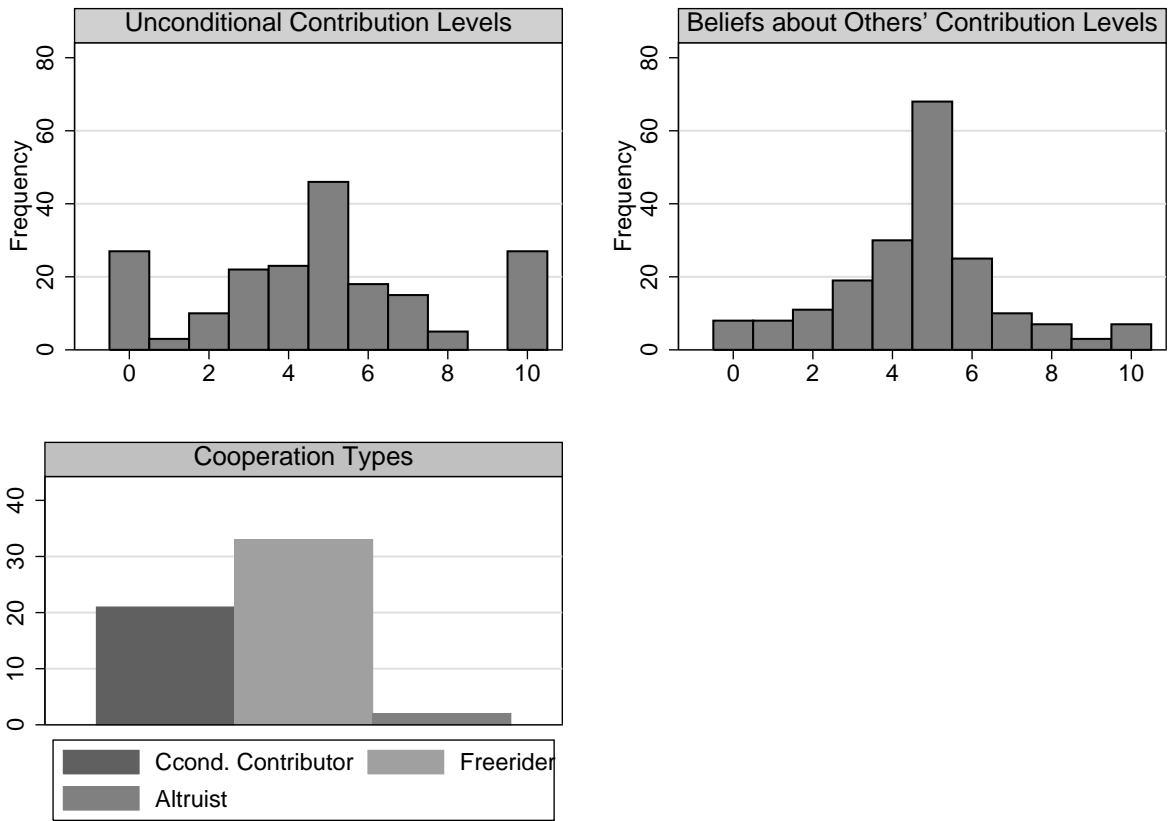


Figure 7: Public good game distributions and cooperation types

do so as well. Figure 7 shows the distribution of these types in our sample.

Table 11 shows how our set of local and global centrality measures varies with these different cooperation types. We see that especially altruists, i.e. individuals who contribute to a public good even if others do not, have on average lower values in local and global centrality measures. In contrast, free-riders tend to have more friends and to be more central.

Table 12 and Table 13 present results from regressions where link intensity is used as a dependent variable. In the unrestricted sample (Table 12) we see that free-riders tend to have more value-weighted links, while altruists have the fewest. To isolate the intensive margin, we move to the restricted sample (Table 13) and here the picture changes: Compared to freeriders, both altruists and – albeit to a smaller extent – conditional cooperators report

Table 11: Network centrality and cooperation types (OLS)

	Local centrality measures			Global centrality measures		
	Out-degree (1)	In-degree (2)	Two-sided-degree (3)	Closeness (4)	Betweenness (5)	Eigenvector (6)
Conditional	-0.248	-0.218	-0.106	-0.003	-0.010	-0.023**
Cooperator	(0.537)	(0.496)	(0.427)	(0.006)	(0.007)	(0.010)
Freerider	0.289	-0.281	-0.023	0.014**	0.003	0.011
	(0.435)	(0.402)	(0.332)	(0.006)	(0.008)	(0.020)
Altruist	-1.121***	-1.308***	-0.905***	-0.006*	-0.021***	-0.012
	(0.346)	(0.299)	(0.186)	(0.003)	(0.003)	(0.010)
Further Controls	yes	yes	yes	yes	yes	yes
Observations	195	195	195	141	141	141
$R^2$	0.347	0.329	0.357	0.286	0.217	0.222

Notes: Further controls refers to the set of individual characteristics already used in previous regressions. Robust standard errors in parentheses; \*\*\* Significant at the 1 percent level. \*\* Significant at the 5 percent level. \* Significant at the 10 percent level.

their friendships to be particularly strong, while free-riders have comparably weak links.

Finally, there is also a finding that is of interest from a methodological perspective: While students' level of agreeableness – as measured in a self-reported questionnaire – is positively associated with network centrality, cooperativeness – as measured in an incentivized experiment – tend to be rather negatively associated with network centrality. At first glance, this seems puzzling as these measures have similar interpretations: agreeableness, reflects an individuals' *self-image* about cooperation and social harmony concerns. In contrast, our measure of cooperativeness captures individuals' actual cooperativeness *behavior* when actions have real payoff consequences.<sup>22</sup> The difference in the association of these measures with network centrality as well as their very low correlation in the data (approx. 0.005) discourages the use of questionnaire-based personality measures as a substitute for behavioral measures from incentivized laboratory experiments.

## 4 Summary and discussion of results

The paper has presented a detailed explorative analysis of a social network that has developed after students were randomly assigned to small groups on the first two days of university. This analysis has revealed a rich picture about the relationship between individual characteristics and link formation as well as the manipulability of social networks. From our perspective

<sup>22</sup>“Agreeable individuals are considerate, friendly, generous, helpful, and willing to compromise their interests with others”(Costa and McCrae, 1987).

Table 12: Friendship formation and cooperation types (OLS, full sample)

	Link Value (1)	Link Value (2)	Link Value (3)	Link Value (4)
Conditional Cooperator	0.0014 (0.0057)	-0.0007 (0.0057)		
Freerider	0.0121** (0.0053)		0.0120** (0.0052)	
Altruist	-0.0121 (0.0148)			-0.0143 (0.0148)
Constant	0.0374*** (0.0020)	0.0395*** (0.0018)	0.0374*** (0.0018)	0.0395*** (0.0017)
Cond.C. - Freerider p-value	-0.0107 0.1401			
Cond.C. - Altruist p-value	0.0135 0.3884			
Freerider - Altruist p-value	0.0242 0.1180			
Observations	39402	39402	39402	39402
$R^2$	0.0002	0.0000	0.0002	0.0000
Adj. $R^2$	0.0001	0.0001	0.0001	0.0000

Notes: Robust standard errors in parentheses. \*\*\* Significant at the 1 percent level. \*\* Significant at the 5 percent level. \* Significant at the 10 percent level.



Table 13: Friendship formation and cooperation types (OLS, restricted sample)

	Link Value (1)	Link Value (2)	Link Value (3)	Link Value (4)
Conditional Cooperator	0.2283** (0.1085)	0.1891* (0.1071)		
Freerider	0.1594 (0.1072)		0.1273 (0.1057)	
Altruist	0.6344** (0.2766)			0.5798** (0.2752)
Constant	2.6990*** (0.0435)	2.7382*** (0.0401)	2.7311*** (0.0400)	2.7536*** (0.0377)
Cond.C. - Freerider p-value	0.0689 0.6220			
Cond.C. - Altruist p-value	-0.4061 0.1630			
Freerider - Altruist p-value	-0.475 0.1023			
Observations	563	563	563	563
$R^2$	0.0114	0.0040	0.0033	0.0023
Adj. $R^2$	0.0061	0.0022	0.0015	0.0005

Notes: Robust standard errors in parentheses.. \*\*\* Significant at the 1 percent level. \*\* Significant at the 5 percent level. \* Significant at the 10 percent level.

three results are particularly worthwhile to be summarized and highlighted.

### **Homophily with respect to behavioral measures**

First, our results confirm and substantially expand previous results on the prevalence of homophily in social network formation. Similarity positively affects link formation not only for known factors like gender and nationality, but also with respect to determinants of economic behavior that have not been analyzed in social networks so far, like time preferences and cooperativeness. We find homophily not only in link formation (extensive margin) but also in link intensity (intensive margin).

### **Manipulability of social networks**

Second, another key finding of our study is the positive impact of common group membership during the introductory days on friendship formation. The fact, that this variable has explanatory power although it was randomly determined, documents the path dependence of social relationships and thus network structures. It also indicates that social networks are – at least to some extent – malleable and thus potentially subject to policy interventions. However, our analysis of the intensive margin of link formation also reveals that link intensity is unrelated to common group membership. Here, individual characteristics appear to be the relevant determinants. This suggests that while weak ties can rather easily be manipulated during the formation of social networks, the emergence of particularly strong ties again depends on the characteristics of individuals and thus might not be manipulable as easily. It also implies that the effectiveness of policies aiming at exploiting peer effects depends on the underlying mechanisms these peer effects operate through, e.g. whether they require strong ties or already unfold through weak links among individuals.

### **The role of cooperativeness in link formation**

Our study identifies cooperativeness as a particularly powerful and multi-faceted predictor of link formation. Its most direct effect is its negative association with local centrality measures like in-degree (popularity), out-degree and the extensive margin of link formation. In line with this, the more cooperative students behave in the unconditional public goods game the more they tend to claim their links to be less intense and – considering the aggregate network structure – they are also less influential in the sense of a lower eigenvector centrality.

Even our analysis of cooperator types reveals that free-riders seem to be most central in the network. However, this picture of an overall negative effect of cooperativeness on link formation becomes more complex once we move on to other results: Links to cooperative students (receiver characteristics) are not perceived to be less intense but, if anything, rather the opposite.<sup>23</sup> When we look at the effect of different cooperation types on the intensive margin of link formation, the sign of the effect of cooperativeness even flips, as altruists report their friendships to be particularly strong, while free-riders have comparably weak links.

We can only speculate about the underlying mechanisms driving these results, which - at first glance - seem surprising: typically, one might expect cooperative individuals to be rather popular because cooperation, especially as measured by the public good game, benefits those interacting with the cooperator. However, it is important to notice that reported friendships as well as cooperation levels in our data should be understood as characteristics of a “stable social equilibrium”. As such, even if other individuals like to befriend cooperative individuals, those might simply reject other individuals that are non-cooperators, thereby ultimately becoming less central and less popular. In turn, one could also imagine that individuals with many friends are less cooperative, as in a wider and potentially more anonymous network, free-riding behavior could be more common, as the famous group-size paradox suggests (Olson, 1965). Our strong evidence for homophily with respect to cooperativeness is in line with these arguments.

However, this does not yet explain the positive effect of cooperative types on link intensity. Also the fact that cooperative senders claim a lower link intensity while this does not hold for links to cooperative receivers, i.e. the asymmetry in the perception of link intensity, still lacks an explanation.

A potential explanation is that the building up and sustaining of a cooperative relationship requires an investment, as it might require more effort and maintenance. This should then lead to a smaller overall number of links for individuals who like to cooperate. However, the links these cooperative individuals form might then be particularly strong. In contrast, social connections not featuring cooperation might be loosely held without much effort, such that free-riders and generally less cooperative students might end up with a higher number of links and a more central network position. The negative effect of time preferences on

---

<sup>23</sup>The sign of the coefficient becomes positive once we restrict the sample to existing links, although the effect is not statistically significant.

local centrality is also consistent with this explanation, as especially individuals with more patience should be willing to undertake such investments.<sup>24</sup>

The strong association of cooperation with different dimensions of link formation in our data is a robust empirical finding. Our explanations are probably just one of many explanations that are able to rationalize these intriguing empirical patterns and they indicate that it might be worthwhile to pursue further research to shed more light on the role of cooperation behavior for link and network development.

## 5 Conclusion

This paper studies how social networks emerge in an educational context. We analyze the relationship between individual characteristics of students and the process of individual friendship formation. Further, we study the shape of local and global network structures resulting from this process. Our study benefits from specific characteristics of the setting at hand: First, as it considers the formation of social networks of *university students*, our results are informative for *educational* contexts. Second, as our study is conducted during the start of students' university careers, our results refer to the *emergence* of new social networks – in contrast to changes of already existing networks. Third, we do not restrict the analysis to socioeconomic and personality factors but additionally consider behavioral measures from the laboratory like cooperativeness, time and risk preferences that are elicited using established experimental methods.

Our main findings can be summarized into three results. First, we extend existing knowledge on homophily in two dimensions: On the one hand, we document that homophily not only refers to characteristics like nationality and gender but also to variables so far not present in social network data sets, namely behavioral measures like cooperativeness and time preferences. On the other hand we show that homophily matters not only for link *formation* (extensive margin) but also for link *intensity* (intensive margin). Second, we find a robust effect of joint group membership during the introductory days. The fact that this random assignment has a substantial effect on network formation implies that educational policy might exploit group composition in order to affect social network structures and the

---

<sup>24</sup>It also provides an explanation for another puzzling result: While cooperativeness of senders is negatively associated with stated link intensity, links *to* cooperative students are not proclaimed to be less intense. This asymmetry in the *perception* of link intensities between more and less cooperative students might simply reflect differences in average intensities of their respective links.

resulting peer spillovers. However, as we also find that the predictive power of common group membership does not carry over to link *intensity*, the effectiveness of such policies might depend on the required strength of links. Finally, we find a complex relationship between cooperativeness and link formation. While overall the direct association seems to be negative in the sense of cooperative individuals being less central with respect to local and global centrality measures, our results also show that links from altruists (unconditional cooperators) are particularly strong. We speculate that the evidence on the role of cooperativeness for link and network development might be understood by taking an investment perspective on friendship formation.

Testing this conjecture could serve as a potential avenue for further research. More generally, it seems worthwhile to take our results as a starting point for analyzing – theoretically and empirically – specific behavioral mechanisms that underlie the process of link formation and thus generate the patterns documented in our data. Finally, our results could be of further use in follow-up studies examining the role of peer effects for skill and preference formation: The analysis showed that common group membership during introductory days fulfills the requirements for a valid instrument in a regression of changes in preferences and characteristics on social network characteristics.

## References

- ANDERSEN, S., G. HARRISON, M. LAU, AND E. RUTSTRÖEM (2008): “Eliciting Risk and Time Preferences,” *Econometrica*, 76, 583–618.
- BANDIERA, O., I. BARANKAY, AND I. RASUL (2010): “Social Incentives in the Workplace,” *Review of Economic Studies*, 77, 417–458.
- (2013): “Team Incentives: Evidence from a Firm Level Experiment,” *Journal of the European Economic Association*, 11, 1079–1114.
- BARTLING, B., E. FEHR, M. MARECHAL, AND D. SCHUNK (2009): “Egalitarianism and Competitiveness,” *American Economic Review*, 99, 93–98.
- BEARMAN, P. AND J. MOODY (2004): “Suicide and Friendships Among American Adolescents,” *American Journal of Public Health*, 94, 89–95.
- BORGHANS, L., A. DUCKWORTH, J. HECKMAN, AND B. TER WEEL (2008): “The Economics and Psychology of Personality Traits,” *Journal of Human Resources*, 43, 972–1059.
- BORS, D. AND T. STOKES (1998): “Raven’s Advanced Progressive Matrices: Norms for First-Year University Students and the Development of a Short Form,” *Educational and Psychological Measurement*, 58, 382–398.
- BURKE, M. A. AND T. R. SASS (2013): “Classroom Peer Effects and Student Achievement,” *Journal of Labor Economics*, 31, 51 – 82.
- BURT, R., P. MARSDEN, AND P. ROSSI (1985): “A Research Agenda for Survey Network Data,” *North*, 1–72.
- BURT, R. S. (2005): *Brokerage and Closure: An Introduction to Social Capital*, Oxford University Press, USA.
- BUSH, G. AND L. HATTERY (1951): “Federal Recruitment of Junior Engineers,” *Science*, 114, 455–458.
- CALVÓ-ARMENGOL, A., E. PATACCHINI, AND Y. ZENOU (2009): “Peer Effects and Social Networks in Education,” *The Review of Economic Studies*, 76, 1239–1267.

- CAMERER, C. (2011): “The Promise and Success of Lab-Field Generalizability in Experimental Economics: A Critical Reply to Levitt and List,” *SSRN eLibrary*.
- CARRELL, S., B. SACERDOTE, AND J. WEST (2013): “From Natural Variation to Optimal Policy? The Importance of Endogenous Peer Group Formation,” *Econometrica*, 81, 855–882.
- CARRINGTON, P., J. SCOTT, AND S. WASSERMAN (2005): *Models and methods in social network analysis*, Cambridge University Press.
- CHRISTAKIS, N. AND J. FOWLER (2007): “The Spread of Obesity in a Large Social Network over 32 Years,” *New England Journal of Medicine*, 357, 370–379.
- (2008): “The Collective Dynamics of Smoking in a Large Social Network,” *New England Journal of Medicine*, 358, 2249–2258.
- COLEMAN, J., E. KATZ, AND H. MENZEL (1966): *Medical Innovation: A Diffusion Study*, Bobbs-Merrill Company.
- CONTI, G., A. GALEOTTI, G. MÜLLER, AND S. PUDNEY (2013): “Popularity,” *Journal of Human Resources*, 48, 1072–1094.
- COSTA, P. AND R. MCCRAE (1987): “Validation of the five-factor model of personality across instruments and observers.” *Journal of personality and social psychology*, 52, 81–90.
- (1992): *Revised NEO Personality Inventory (NEO PI-R) and NEO Five-Factor Inventory (NEO-FFI)*, Psychological Assessment Resources.
- CURRARINI, S., M. JACKSON, AND P. PIN (2009): “An Economic Model of Friendship: Homophily, Minorities, and Segregation,” *Econometrica*, 77, 1003–1045.
- DE NOOY, W., A. MRVAR, AND V. BATAGELJ (2005): *Exploratory Social Network Analysis with Pajek*, Structural Analysis in the Social Sciences, Cambridge University Press.
- DEHNE, M. AND J. SCHUPP (2007): “Persönlichkeitsmerkmale in Sozio-oekonomischen Panel (SOEP) - Konzepte, Umsetzung und empirische Eigenschaften,” DIW Research Notes 26, Berlin.

- DOHMEN, T. AND A. FALK (2011): “Performance Pay and Multidimensional Sorting: Productivity, Preferences, and Gender,” *American Economic Review*, 101, 556–590.
- DOHMEN, T., A. FALK, D. HUFFMAN, AND U. SUNDE (2010): “Are Risk Aversion and Impatience Related to Cognitive Ability?” *American Economic Review*, 100, 1238–1260.
- ECKEL, C., C. JOHNSON, AND C. MONTMARQUETTE (2005): “Saving Decisions of the Working Poor: Short- and Long-Term Horizons,” in *Field Experiments in Economics (Research in Experimental Economics)*, ed. by G. Harrison, J. Carpenter, and J. List, Amsterdam: Emerald Group Publishing Limited, 219–260.
- EPPLE, D. AND R. ROMANO (2011): “Chapter 20 - Peer Effects in Education: A Survey of the Theory and Evidence,” North-Holland, vol. 1 of *Handbook of Social Economics*, 1053 – 1163.
- FEHR, E. AND A. LEIBBRANDT (2011): “A Field Study on Cooperativeness and Impatience in the Tragedy of the Commons,” *Journal of Public Economics*, 95, 1144–1155.
- FISCHBACHER, U., S. GÄCHTER, AND E. FEHR (2001): “Are People Conditionally Cooperative? Evidence from a Public Goods Experiment,” *Economics Letters*, 71, 397–404.
- FOSTER, G. (2005): “Making Friends: A Non-Experimental Analysis of Social Pair Formation,” *Human Relations*, 58, 1443–1465.
- FOWLER, J., C. DAWES, AND N. CHRISTAKIS (2009): “Model of Genetic Variation in Human Social Networks,” *Proceedings of the National Academy of Sciences*, 106, 1720–1724.
- FOWLER, J., J. SETTLE, AND N. CHRISTAKIS (2011): “Correlated Genotypes in Friendship Networks,” *Proceedings of the National Academy of Sciences*, 108, 1993–1997.
- FREDERICK, S., G. LOEWENSTEIN, AND T. O’DONOGHUE (2002): “Time Discounting and Time Preference: A Critical Review,” *Journal of Economic Literature*, 40, 351–401.
- FRUCHTERMAN, T. AND E. REINGOLD (1991): “Graph Drawing by Force-Directed Placement,” *Software: Practice and Experience*, 21, 1129–1164.



- GANDAL, N., C. KING, AND M. ALSTYNE (2009): “The Social Network within a Management Recruiting Firm: Network Structure and Output,” *Review of Network Economics*, 8, 302–324.
- GLAESER, E., B. SACERDOTE, AND J. SCHEINKMAN (1996): “Crime and Social Interactions,” *The Quarterly Journal of Economics*, 111, 507–548.
- HARRIS, K., C. HALPERN, E. WHITSEL, J. HUSSEY, J. TABOR, P. ENTZEL, AND J. UDRY (2009): “The National Longitudinal Study of Adolescent to Adult Health: Research Design,” Www document, URL: <http://www.cpc.unc.edu/projects/addhealth/design>.
- HECKMAN, J. (2011): “Integrating Personality Psychology into Economics,” NBER Working Papers 17378, National Bureau of Economic Research, Inc.
- HOLT, C. AND S. LAURY (2002): “Risk Aversion and Incentive Effects,” *American Economic Review*, 92, 1644–1655.
- HUCKFELDT, R. AND J. SPRAGUE (1995): *Citizens, Politics, and Social Communication: Information and Influence in an Election Campaign*, Cambridge University Press.
- JACKSON, M. (2008): *Social and Economic Networks*, Princeton University Press.
- (2009): “Genetic Influences on Social Network Characteristics,” *Proceedings of the National Academy of Sciences*, 106, 1687–1688.
- KAMADA, T. AND S. KAWAI (1989): “An Algorithm for Drawing General Undirected Graphs,” *Information Processing Letters*, 31, 7–15.
- KATZ, L. (1953): “A New Status Index Derived from Sociometric Analysis,” *Psychometrika*, 18, 39–43.
- KING, G. AND L. ZENG (2001): “Logistic Regression in Rare Events Data,” *Political Analysis*, 9, 137–163.
- KIRBY, K. AND N. PETRY (2004): “Heroin and Cocaine Abusers Have Higher Discount Rates for Delayed Rewards than Alcoholics or Non-Drug-Using Controls,” *Addiction*, 99, 461–471.

- KREMER, M., E. DUFLO, AND P. DUPAS (2011): “Peer Effects, Teacher Incentives, and the Impact of Tracking,” *American Economic Review*, 101, 1739–1774, this file is an earlier version of the published paper.
- LAMBA, S. AND R. MACE (2011): “Demography and Ecology Drive Variation in Cooperation Across Human Populations,” *Proceedings of the National Academy of Sciences*, 108, 14426–14430.
- LAZARSFELD, P. AND R. MERTON (1954): “Friendship as a Social Process: A Substantive and Methodological Analysis,” in *Freedom and Control in Modern Society*, ed. by M. Berger, T. Abel, and C. Page, New York: Van Nostrand, 18–66.
- LEWIS, K., M. GONZALEZ, AND J. KAUFMAN (2012): “Social Selection and Peer Influence in an Online Social Network,” *Proceedings of the National Academy of Sciences*, 109, 68–72.
- LEWIS, K., J. KAUFMAN, M. GONZALEZ, A. WIMMER, AND N. CHRISTAKIS (2008): “Tastes, Ties, and Time: A New Social Network Dataset Using Facebook.com,” *Social Networks*, 30, 330–342.
- MARMAROS, D. AND B. SACERDOTE (2006): “How Do Friendships Form?” *The Quarterly Journal of Economics*, 121, 79–119.
- MAYER, A. AND S. PULLER (2008): “The Old Boy (and Girl) Network: Social Network Formation on University Campuses,” *Journal of Public Economics*, 92, 329–347.
- MCPHERSON, M., L. SMITH-LOVIN, AND J. COOK (2001): “Birds of a Feather: Homophily in Social Networks,” *Annual Review of Sociology*, 27, 415–444.
- MOODY, J. (2001): “Race, School Integration, and Friendship Segregation in America,” *American Journal of Sociology*, 107, 679–716.
- NIEDERLE, M. AND L. VESTERLUND (2007): “Do Women Shy Away from Competition? Do Men Compete Too Much?” *The Quarterly Journal of Economics*, 122, 1067–1101.
- OLSON, M. (1965): *The Logic of Collective Action: Public Goods and the Theory of Groups*, Harvard University Press.

- PATACCHINI, E., E. RAINONE, AND Y. ZENOU (2014): “Heterogeneous Peer Effects in Education,” CEPR Discussion Papers 9804, C.E.P.R. Discussion Papers.
- PRELL, C. (2012): *Social Network Analysis; History, Theory and Methodolgy*, Sage Publication Ltd.
- RUSTAGI, D., S. ENGEL, AND M. KOSFELD (2010): “Conditional Cooperation and Costly Monitoring Explain Success in Forest Commons Management,” *Science*, 330, 961–965.
- SACERDOTE, B. (2001): “Peer Effects With Random Assignment: Results For Dartmouth Roommates,” *The Quarterly Journal of Economics*, 116, 681–704.
- (2011): *Peer Effects in Education: How Might They Work, How Big Are They and How Much Do We Know Thus Far?*, Elsevier, vol. 3 of *Handbook of the Economics of Education*, chap. 4, 249–277.
- VENTURA, L. (2003): “Direct Measures of Time Preference,” *The Economic and Social Review*, 34, 293–310.
- WASSERMAN, S. AND K. FAUST (1994): *Social Network Analysis. Methods and Applications*, Cambridge University Press.

# APPENDIX

## A Empirical information

### A.1 Additional figures and tables

Table 14: Description of variables

Variable	Description	Coding
TimePreferences	Participants make 15 binary choices between EUR 40.00 today and some larger amount ranging from EUR 40.00 to EUR 70.00 in 6 months (Frederick et al., 2002)	Number of late choices (0 to 15)
RiskPreferences	Participants make 15 binary decisions choosing between a lottery (winning EUR 20.00 or EUR 0.00 with probability 0.5 each) and a certain payment of EUR X were X increases from 3 to 17 (Dohmen et al., 2010)	Number of lottery choices (0 to 15)
Cooperativeness	Participants had to decide how much of an endowment of EUR 10.00 they contribute to a public good that yields a group return of 200 percent and was subsequently split among all 4 group members	EUR invested into the public good (0 to 10)
Competitiveness	Participants had to make 10 binary decisions on whether to be compensated by a piece rate or a tournament scheme (Niederle and Vesterlund, 2007), where the decisions vary by the attractiveness of the piece rate scheme	Number of tournament choices (0 to 10)
Agreeableness	As described and calculated in Costa and McCrae, 1992 and Dehne and Schupp, 2007	(0 to 100)
Conscientiousness	As described and calculated in Costa and McCrae, 1992 and Dehne and Schupp, 2007	(0 to 100)
Extraversion	As described and calculated in Costa and McCrae, 1992 and Dehne and Schupp, 2007	(0 to 100)
Neuroticism	As described and calculated in Costa and McCrae, 1992 and Dehne and Schupp, 2007	(0 to 100)
Openness	As described and calculated in Costa and McCrae, 1992 and Dehne and Schupp, 2007	(0 to 100)
Male	Gender (male or female)	1 if participant is male, 0 if female
Age	Age of participant	in years
RavensIQ	Participants face 8 abstract figures, where for each they have to choose one out of 8 possible objects that fits the current screen.	Number of correct answers (0 to 8)
Partner	Survey question whether the participant is in a relationship	1 if in a relationship, 0 if not
Smoking	Survey question whether the participant is a smoker	1 if yes, 0 if no
Alcohol	Survey question whether the participant drinks alcohol	1 if yes, 0 if no
LivingAlone	Survey question whether the participant lives alone in his flat	1 if yes, 0 if no
LivingSharedFlat	Survey question whether the participant shares his flat	1 if yes, 0 if no
SiblingsNo.	Survey question on the number of siblings	(0 to no upper bound specified)
SiblingRank	Survey question on number of older siblings	(0 to no upper bound specified)
RelativesNo.	Survey question on number of known relatives	(0 to no upper bound specified)
FriendsNo.	Survey question on number of friends outside the university	(0 to no upper bound specified)
ParentsKM100	Survey question on distance to parents' home	0 to no upper bound specified; in 100 kilometers
Relocation	Survey question on whether the participant moved prior to starting university	1 if yes, 0 if no
Participate	Dummy variable on whether student participated in the introductory days	1 if yes, 0 if no
Nationality	Survey question on nationality of the participant	0 if foreigner, 1 if not

Table 15: Cross-correlation table

Variables	Male	Age	Partner	Ravens IQ	Alcohol	Smoking	Time	Risk	PG Con.	Con. Con.	Free-Rider	Altruist Altruist	Comp. etiv..	Siblings No.	Friendse No.	Participate
Male	1.000															
Age	0.046	1.000														
Partner	-0.089	0.017	1.000													
RavensIQ	-0.001	0.084	-0.026	1.000												
Alcohol	0.207	0.018	0.099	-0.079	1.000											
Smoking	0.067	-0.046	0.101	-0.020	0.128	1.000										
TimeP	-0.138	0.008	-0.039	-0.172	-0.246	-0.155	1.000									
RiskP	0.176	0.018	-0.094	0.053	0.121	-0.108	-0.085	1.000								
PG Contribution	-0.126	0.246	0.007	0.051	0.014	0.012	0.101	0.059	1.000							
Cond. C.	-0.009	-0.003	-0.025	0.116	-0.043	-0.080	0.015	-0.029	-0.051	1.000						
FreeRider	0.186	-0.111	-0.040	-0.098	-0.061	-0.003	0.086	-0.027	-0.153	-0.156	1.000					
Altruist	-0.099	-0.063	0.090	0.007	0.049	-0.048	0.095	-0.015	-0.026	-0.035	-0.046	1.000				
Competitiveness	0.240	0.174	0.035	-0.028	0.242	-0.046	-0.176	0.245	0.171	0.010	-0.013	-0.075	1.000			
SiblingsNo	0.019	-0.007	0.021	-0.023	-0.089	0.011	0.073	0.012	-0.057	0.045	0.036	0.026	-0.085	1.000		
FriendsClose	0.152	0.023	0.036	0.009	0.136	-0.003	0.001	0.086	-0.007	-0.019	-0.012	-0.042	-0.074	-0.051	1.000	
Participate	-0.029	-0.006	-0.011	-0.038	0.048	-0.002	-0.066	0.060	-0.053	0.072	0.003	-0.035	0.013	-0.162	0.083	1.000

Table 16: Friendship formation (alternative specifications)

	Probit		Logit		OLS	
	(1)	(2)	(3)	(4)	(5)	(6)
<i>Absolute Differences (<math> X_i - X_j </math>)</i>						
Gender (d)	-0.147*** (0.0012)	-0.143*** (0.0010)	-0.205*** (0.0011)	-0.193*** (0.0010)	-0.046*** (0.0012)	-0.044*** (0.0012)
RavensIQ	-0.024 (0.0006)	-0.046** (0.0005)	-0.030 (0.0005)	0.055** (0.0005)	-0.007 (0.0005)	-0.011** (0.0006)
Participate (d)	-0.055** (0.0016)	0.128*** (0.0022)	-0.073* (0.0016)	0.123** (0.0020)	-0.012** (0.0011)	0.126*** (0.0035)
RiskPreferences	0.015 (0.0002)	0.009 (0.0002)	0.024 (0.0002)	0.008 (0.0002)	0.005 (0.0002)	-0.000 (0.0003)
TimePreferences	-0.081*** (0.0002)	-0.080*** (0.0002)	-0.111*** (0.0002)	-0.109*** (0.0002)	-0.024*** (0.0002)	-0.024*** (0.0002)
Cooperativeness	-0.063*** (0.0002)	-0.060*** (0.0002)	-0.090*** (0.0002)	-0.083*** (0.0002)	-0.019*** (0.0002)	-0.020*** (0.0002)
Group (d)	-0.156*** (0.0035)	-0.294*** (0.0059)	-0.205*** (0.0034)	-0.349*** (0.0054)	-0.071*** (0.0025)	-0.214*** (0.0065)
<i>Sender Characteristics (<math>X_i</math>)</i>						
Male (d)		0.023 (0.0014)		0.042 (0.0012)		0.007 (0.0015)
Participate (d)		0.181*** (0.0011)		0.212*** (0.0010)		0.132*** (0.0034)
RiskPreferences		-0.049* (0.0003)		-0.060* (0.0002)		-0.014* (0.0003)
TimePreferences		-0.058*** (0.0002)		-0.077*** (0.0001)		-0.017*** (0.0002)
RavensIQ		0.017 (0.0005)		0.018 (0.0005)		0.002 (0.0006)
Cooperativeness		-0.032 (0.0002)		-0.035 (0.0002)		-0.011** (0.0002)
<i>Receiver Characteristics (<math>X_j</math>)</i>						
Male (d)		-0.006 (0.0013)		-0.002 (0.0012)		0.000 (0.0015)
RavensIQ		0.012 (0.0005)		0.012 (0.0004)		0.001 (0.0006)
Participate (d)		0.138*** (0.0014)		0.152*** (0.0013)		0.123*** (0.0034)
RiskPreferences		0.001 (0.0002)		-0.004 (0.0002)		-0.001 (0.0003)
TimePreferences		-0.033 (0.0002)		-0.038 (0.0002)		-0.010* (0.0002)
Cooperativeness		-0.081*** (0.0002)		-0.097*** (0.0002)		-0.024*** (0.0002)
Further Controls	yes	yes	yes	yes	yes	yes
Observations	39402	39402	39402	39402	39402	39402
(Pseudo-) R <sup>2</sup>	0.0581	0.1052	0.0601	0.1056	0.0099	0.0234

Notes: Probit and Logit estimations show marginal effects evaluated at the mean. Standard errors clustered at the individual level in parentheses; (d) for discrete change of dummy variable from 0 to 1. \*\*\* Significant at the 1 percent level. \*\* Significant at the 5 percent level. \* Significant at the 10 percent level.

Table 17: Friendship formation (alternative specifications)

	Link Value (ordered probit) [0,...,4] (1)	Link Value (ordered probit) [1,...,4] (Intensive margin) (2)	Binary Link (probit) [0;1] (Extensive Margin) (3)
<i>Absolute Differences</i> ( $ X_i - X_j $ )			
Gender	-0.3065*** (0.0354)	-0.1507 (0.1020)	-0.3066*** (0.0357)
Participate	-0.1108*** (0.0410)	0.1226 (0.1096)	-0.1151*** (0.0412)
RiskPreferences	0.0046 (0.0068)	-0.0325* (0.0175)	0.0060 (0.0069)
TimePreferences	-0.0283*** (0.0063)	0.0064 (0.0188)	-0.0292*** (0.0063)
RavensIQ	-0.0231 (0.0166)	0.0468 (0.0442)	-0.0241 (0.0167)
Cooperativeness	-0.0253*** (0.0069)	-0.0097 (0.0193)	-0.0257*** (0.0071)
Group	-0.4385*** (0.0415)	0.0556 (0.1097)	-0.4462*** (0.0421)
Agreeableness	-0.0008 (0.0021)	0.0031 (0.0063)	-0.0012 (0.0021)
Conscientiousness	-0.0017 (0.0020)	0.0030 (0.0057)	-0.0018 (0.0020)
Extraversion	-0.0022 (0.0020)	0.0029 (0.0060)	-0.0024 (0.0020)
Neuroticism	0.0030 (0.0019)	0.0045 (0.0055)	0.0029 (0.0019)
Openness	0.0047** (0.0022)	-0.0027 (0.0069)	0.0048** (0.0023)
Further Controls	yes	yes	yes
Observations	39402	563	39402
(Pseudo) R <sup>2</sup>	0.0463	0.0111	0.0581

Notes: Marginal effects evaluated at the mean. Standard errors clustered at the individual level in parentheses. \*\*\* Significant at the 1 percent level. \*\* Significant at the 5 percent level. \* Significant at the 10 percent level.



Table 18: Local network centrality measures (alternative specifications)

	Out-degree			In-degree			Two-sided-degree		
	OLS (1)	Poisson (2)	NB (3)	OLS (4)	Poisson (5)	NB (6)	OLS (7)	Poisson (8)	NB (9)
<i>Personal Characteristics</i>									
Male (d)	-0.121 (0.337)	-0.056 (0.290)	-0.112 (0.297)	-0.185 (0.348)	-0.191 (0.299)	-0.173 (0.306)	-0.436 (0.278)	-0.384 (0.238)	-0.396* (0.239)
Age	-0.158** (0.078)	-0.167** (0.077)	-0.182** (0.078)	-0.181** (0.082)	-0.226*** (0.083)	-0.232*** (0.082)	-0.098 (0.062)	-0.085 (0.063)	-0.088 (0.063)
Partner (d)	1.082*** (0.298)	0.943*** (0.243)	0.944*** (0.249)	0.643** (0.311)	0.551** (0.272)	0.563** (0.271)	0.928*** (0.229)	0.841*** (0.184)	0.841*** (0.187)
RavensIQ	-0.085 (0.133)	-0.039 (0.108)	-0.062 (0.113)	-0.086 (0.138)	-0.014 (0.117)	-0.029 (0.114)	-0.039 (0.121)	0.008 (0.092)	0.002 (0.092)
<i>Behavioral Measures</i>									
TimePreferences	-0.116** (0.050)	-0.102** (0.042)	-0.104** (0.044)	-0.115** (0.050)	-0.102** (0.042)	-0.110** (0.044)	-0.096** (0.040)	-0.076** (0.032)	-0.077** (0.033)
RiskPreferences	0.036 (0.053)	0.028 (0.045)	0.024 (0.046)	0.019 (0.085)	0.029 (0.069)	0.021 (0.064)	0.047 (0.044)	0.040 (0.037)	0.038 (0.038)
Cooperativeness	-0.090* (0.051)	-0.068 (0.045)	-0.066 (0.046)	-0.167*** (0.052)	-0.148*** (0.048)	-0.148*** (0.050)	-0.106*** (0.040)	-0.096** (0.039)	-0.094** (0.040)
Competitiveness	-0.079 (0.323)	-0.098 (0.266)	-0.112 (0.267)	-0.068 (0.297)	-0.188 (0.244)	-0.207 (0.251)	-0.194 (0.262)	-0.160 (0.205)	-0.178 (0.206)
<i>Big Five</i>									
Agreeableness	0.049*** (0.017)	0.044*** (0.012)	0.044*** (0.013)	0.049*** (0.015)	0.043*** (0.012)	0.042*** (0.012)	0.031** (0.013)	0.028*** (0.009)	0.028*** (0.009)
Conscientiousness	0.038** (0.017)	0.039*** (0.015)	0.040*** (0.015)	0.032* (0.016)	0.029** (0.014)	0.030** (0.014)	0.037*** (0.014)	0.033*** (0.011)	0.033*** (0.012)
Extraversion	-0.011 (0.015)	-0.007 (0.015)	-0.010 (0.015)	-0.028* (0.015)	-0.022 (0.015)	-0.023 (0.015)	-0.008 (0.012)	-0.003 (0.012)	-0.004 (0.012)
Neuroticism	-0.019 (0.015)	-0.012 (0.012)	-0.012 (0.013)	-0.014 (0.016)	-0.010 (0.013)	-0.011 (0.013)	-0.020 (0.012)	-0.011 (0.010)	-0.012 (0.010)
Openness	-0.032 (0.021)	-0.032* (0.017)	-0.033* (0.018)	-0.040* (0.023)	-0.034* (0.019)	-0.033* (0.018)	-0.036* (0.018)	-0.035** (0.014)	-0.035** (0.014)
<i>Contextual Factors</i>									
Participate (d)	1.168* (0.626)	1.248*** (0.394)	1.314*** (0.385)	0.785 (0.630)	0.881* (0.474)	0.904** (0.456)	0.790* (0.451)	0.740** (0.308)	0.763** (0.306)
Further Controls	yes	yes	yes	yes	yes	yes	yes	yes	yes
Observations	195	195	195	195	195	195	195	195	195
(Pseudo) R <sup>2</sup>	0.346	0.162	0.100	0.328	0.146	0.098	0.355	0.168	0.119

Notes: Marginal effects are evaluated at the mean; Robust standard errors; (d) for discrete change of dummy variable from 0 to 1; \*\*\* Significant at the 1 percent level. \*\* Significant at the 5 percent level. \* Significant at the 10 percent level.

## A.2 (Non-random) Sample selection: Largest component

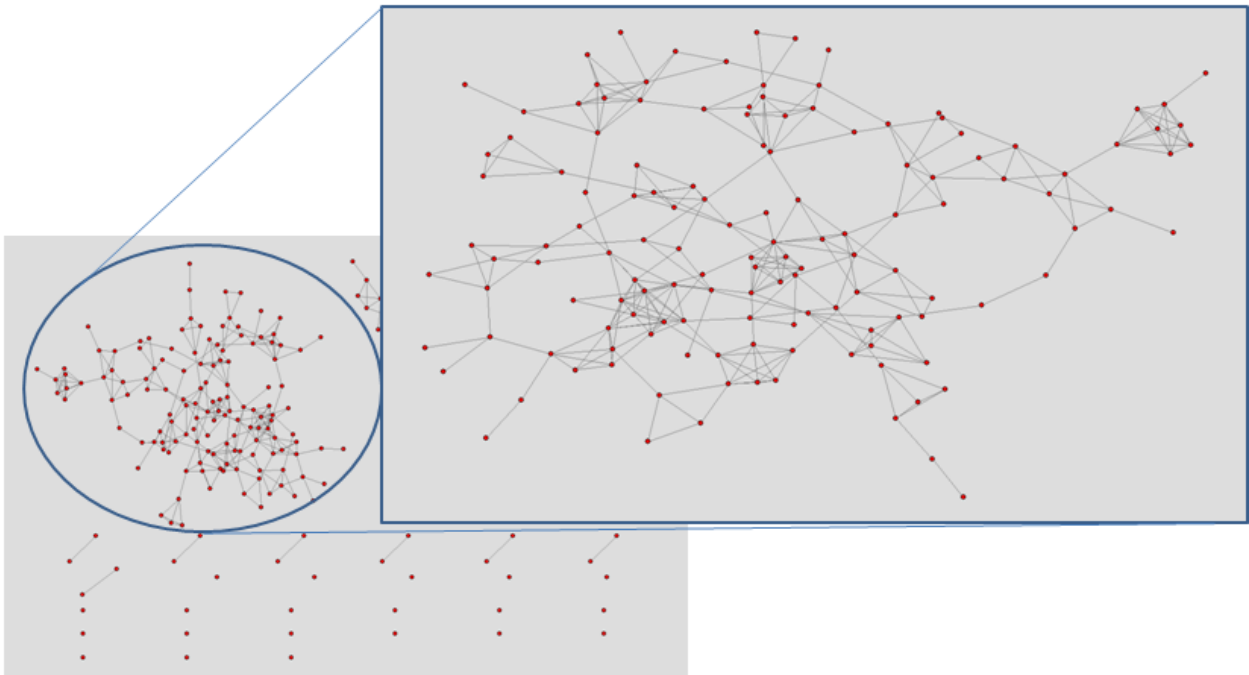


Figure 8: JGU network - Largest component

Notes: The figure was created by *Pajek* (de Nooy et al., 2005).

Table 19: Selection: Largest component

Dependent variable: Largest component		
	Probit (1)	OLS (2)
<i>Personal Characteristics</i>		
Male (d)	0.142** (0.071)	0.121* (0.068)
Age	-0.058*** (0.018)	-0.054*** (0.015)
Partner (d)	0.153** (0.069)	0.098 (0.060)
RavensIQ	-0.024 (0.028)	-0.021 (0.028)
<i>Behavioral Measures</i>		
TimePreferences	-0.019* (0.010)	-0.013 (0.010)
RiskPreferences	0.000 (0.013)	0.003 (0.010)
Cooperativeness	-0.017* (0.011)	-0.010 (0.010)
Competitiveness (d)	0.100 (0.071)	0.059 (0.062)
<i>Big Five</i>		
Agreeableness	0.005 (0.004)	0.004 (0.003)
Conscientiousness	-0.002 (0.004)	-0.001 (0.003)
Extraversion	-0.009** (0.004)	-0.006* (0.003)
Neuroticism	-0.009** (0.004)	-0.007** (0.003)
Openness	-0.002 (0.005)	-0.002 (0.004)
<i>Contextual Factors</i>		
Participate (d)	0.479*** (0.142)	0.452*** (0.106)
RelativesNo.	-0.003 (0.004)	-0.001 (0.004)
FriendsNo.	0.001 (0.006)	0.000 (0.004)
Further Controls	yes	yes
Observations	195	195
(Pseudo) $R^2$	0.338	0.350

Notes: Marginal effects are evaluated at the mean; Robust standard errors in parentheses; (d) for discrete change of dummy variable from 0 to 1. \*\*\* Significant at the 1 percent level. \*\* Significant at the 5 percent level. \* Significant at the 10 percent level.

## B Experimental information

### B.1 Experimental stages

Table 20: Experimental stages

Stage 1	Stage 2	Stage 3	Stage 4	Stage 5	Stage 6
Eliciting cognitive abilities with Ravens IQ	Eliciting risk preferences with lottery choice based method	Eliciting time preferences with lottery choice based method	Eliciting cooperativeness with a public good game	Eliciting competitiveness with a real effort task	Computer-based questionnaire, network questionnaire and payoff

### B.2 Network questionnaire (paper format)

#### Fragebogen

Vorname: \_\_\_\_\_

Nachname: \_\_\_\_\_

*Nennen Sie bitte bis zu 10 StudentInnen aus Ihrem Semester die Sie als Ihre(n) engsten Freund(e) erachten.*

Nr.	Vorname	Nachname	Frage 1: <i>Betrachten Sie die Freundschaft als eine reine Studienfreundschaft oder geht die Freundschaft zu dieser Person Ihrer Meinung nach darüber hinaus?</i>				Frage 2: <i>Kennen Sie die Person schon aus der Zeit vor dem Studium?</i>
			„Uni-Bekannschaft“....Enge Freundschaft				Ja
1.			1	2	3	4	<input type="checkbox"/>
2.			1	2	3	4	<input type="checkbox"/>
3.			1	2	3	4	<input type="checkbox"/>
4.			1	2	3	4	<input type="checkbox"/>
5.			1	2	3	4	<input type="checkbox"/>
6.			1	2	3	4	<input type="checkbox"/>
7.			1	2	3	4	<input type="checkbox"/>
8.			1	2	3	4	<input type="checkbox"/>
9.			1	2	3	4	<input type="checkbox"/>
10.			1	2	3	4	<input type="checkbox"/>