Eidgenössische Technische Hochschule Zürich
Swiss Federal Institute of Technology Zurich

ETH Zurich Sociology Working Paper No. 7

Personality on Social Network Sites: An Application of the Five Factor Model

Stefan Wehrli

September 2008

ETH Zurich, Chair of Sociology

SEW E 21, Scheuchzerstrasse 70
8092 Zurich, Switzerland

Tel. +41 44 632 55 56
Fax +41 44 632 10 54
info@soz.gess.ethz.ch
www.socio.ethz.ch

Personality on Social Network Sites: An Application of the Five Factor Model *

Stefan Wehrli
ETH Zürich
wehrlist@ethz.ch

September 5, 2008

Abstract: In this paper we explore how individual personality characteristics influence online social networking behavior. We use data from an online survey with 1560 respondents from a major Swiss technical university and their corresponding online profiles and friendship networks on a popular Social Network Site (SNS). Apart from sociodemographic variables and questions about SNS usage, we collected survey data on personality traits with a short question inventory of the Five Factor Personality Model (BFI-15). We show how these psychological network antecedents influence participation, adoption time, nodal degree and ego-network growth over a period of 4 months on the networking platform. Statistical analysis with overdispersed degree distribution models identifies extraversion as a major driving force in the tie formation process. We find a counter-intuitive positive effect for neuroticism, a negative influence for conscientiousness and no effects for openness and agreeableness.

Keywords: Online social networks, personality, Big Five, degree distributions

1 Introduction

How does an individual set of attributes like personality affect the formation of relationships and individuals' positions in social networks? Some people seem to accumulate a large number of acquaintances while others have close circles. High variability in the number of contacts is a frequently observed feature of social networks. Especially in the recent network literature, empirical estimation of the exact functional form of such skewed degree distributions have received considerable attention (cf. Newman et al. 2001, Newman 2003, Hancock & Jones 2004).

*We thank Elisabeth Coutts, Debra Hevenstone, Ben Jann and Wojtek Przepiorka for helpful comments and several students for their participation in the data collection process, especially Marcel Breu, Michael Müri, Marc Osswald, and Florian Schlapbach.

The sources of this excess variability is best illustrated with violations of the independence assumption of the underlying random tie formation process. A random network exhibits a poisson degree distribution with the imposed restriction that the conditional variance equals the conditional mean. Higher variance arises in statistical terms either due to true or spurious contagion (Cameron & Trivedi 1986). In the first case, all individuals initially have the same probability of forming ties, but the probability of successive ties depends on prior occurrence. Cumulative advantage, for example, is a well known related candidate of true (but pure) contagion exhibiting heaviest forms of overdispersion. On the other hand, according to the spurious contagion model individuals are assumed to have constant but unequal probabilities of forming ties. In this case, the major source of overdispersion would be simply population heterogeneity. Unfortunately, it has been shown that both violations can result in the same outcome distribution, therefore making them impossible to distinguish with cross-sectional data. In the following, we will explore the second route – individual heterogeneity in the propensity to form ties – by incorporating a potentially stable and exogenous source of variation: personality. Note that this approach is similar in design to the intrinsic fitness model of Bianconi & Barabási (2001) and to certain threshold models from classical diffusion research (Coleman 1964). Apart from the degree distribution, social networks are claimed to share at least three other empirical regularities: short average path lengths, high transitivity (clustering) and positive assortative mixing. The appearance of the first two are also known as small-world effect (Watts 1999). We will briefly discuss these metrics and show that our network also has these characteristics.

Traditionally, network theorists have devoted much of their attention to the consequences of networks and how the behavior of individuals depends on their location in the network. Individuals occupying central positions and having denser or wider reaching networks may gain faster access to information and assistance (Borgatti & Foster 2003). Another view on the causal ordering asks about individuals' characteristics influence social structure. Scholars have mainly relied on two micro-level predictors of network tie formation: spatial proximity and similarity in terms of personal attributes like age, race, gender, values and education (McPherson et al. 2001). Only recently, a special interest in interaction between personality traits and network positions has emerged. Personality traits that predispose people to socialize, such as extraversion or openness to experience might foster and accelerate tie formation in social networks while others like neuroticism constrain individuals from creating ties.

Mehra et al. (2001), for example, found that high self-monitors – people who are concerned about how they are perceived by others – occupied more central positions in the friendship network of a high-tech company. Burt et al. (1998) showed how entrepreneurial personality characteristics are correlated with network constraint and bridging structural holes. While these authors applied very specific and narrow conceptions of personality, others rely on a more comprehensive instrument: the five factor model (FFM, Goldberg 1990). The FFM now seems to be the de facto standard in psychometrics. Klein et al. (2004) found negative effects on centrality from neuroticism and openness, and a positive effect of agreeableness in friendship networks of work group members. Surprisingly, extraversion had no effect on friendship centrality. Vodosek (2003) reported positive effects from extraversion in a friendship network of an undergraduate business class. Asendorpf & Wilpers (1998) found in a very careful assessment of friendship formation processes mainly positive effects from extraversion on contact frequency. See Schafer et al. (2008) for a survey of recent developments.

In our study we follow the strategy of Klein et al. (2004) and regress exogenous personality traits on behavioral and structural indicators, i.e. on the probability of network membership, on the time until students adopt the social networking technology, the number of contacts (degree) and on several well known node level centrality and clustering measures. We conducted an online survey at a major Swiss technical university (ETH Zürich) to collect personality scores, sociodemographic variables, and usage patterns from 1560 students. At the same time and on two subsequent time points we collected profiles and contact lists from a very popular social networking site (StudiVZ) by means of a snowball crawling approach. The two sources were linked, giving us the possibility to study ego-centered networks in the context of a complete network data set.

The remainder of the article is organized as follows. In the next section we will briefly discuss validity concerns, and the advantages and disadvantages of data acquired on social network sites. Personality traits and hypotheses will be presented in section 3, followed by some methodological considerations in section 4. In the last two sections results are presented and discussed. We will show that extraversion plays a pivotal role in the tie formation process, exhibiting significant effects on degree and centrality measures.

2 Social Network Sites

Social Network Sites (SNS) like Facebook, MySpace or LinkedIn, as well as the German Facebook clone StudiVZ, have shown dramatic growth over the last few years. At the same time they started to attract the attention of scholars due to the availability of ready-made, process generated complete network data. Boyd and Ellison (2007) characterize Social Network Sites (SNS) as web-based services that allow individuals to (1) construct a web presence usually including a photo and descriptors like location, age, study concentration and interests, (2) publicly display a list of other users with whom they share a connection, and (3) to traverse those list of connections to view the profiles of others within the system. SNSs particularly evolve around special interests or shared contexts like students populations at universities. Although they now represent a typical form of computer-mediated communication, Facebook and StudiVZ are well known to be rooted in the offline world (Lampe et al. 2007), i.e. students use them typically to stay in contact, communicate with and "spy" on their offline friends. Luckily, this makes this type of data source compatible to classic class room sociometry (Coleman 1964, Hallinen 1978, Moody 2002), where scientists have collected network data of classes or entire schools by means of survey name generators. The major methodological limitation of these studies was the number of contacts they were able to process, normally concentrating on the 3 to 5 strongest ties. On SNSs students collect their acquaintances by themselves, leading to a much wider and more complete picture of the social structure.

Unfortunately, SNS data has obvious limitations as well. Not all students are members of SNSs, and often members are scattered across different platforms. Nevertheless, we found that about 70% of all ETH students maintain a StudiVZ account, and StudiVZ was, according to our survey results, definitely the most popular site to join. A major limitation of the StudiVZ data is that we cannot see who initiates a tie or when. Although acquaintance and friendship are thought to be symmetric relationships, directed network data would allow this study to measure whether individuals with high extraversion scores initiate or attract SNS-relations. As already mentioned, the degree distribution has high variability, with some students having far more than 100 friends within ETH. The mean is at 24.5 contacts with a standard deviation of 20.5. Note that this is only a quarter of the numbers that undergraduates from Michigan State have accumulated on Facebook (Lampe et al. 2007). These high numbers suggest very low tie formation and virtually no tie maintenance costs in contrast to kinship, exchange or support networks. Although strong ties are present in these networks, most of the contacts will be part of a wider circle with a low propensity of being resourceful or helpful. However, survey respondents reported that they know a substantial proportion of their contacts from face-to-face interaction. They meet with about 75% of their friends every once a while, suggesting a considerable overlap between online and offline social networks. Only 20% of the respondents added unknown people to their lists, and in 70% of these cases less than three persons were added. Unfortunately, we know little about real returns to social capital investment and maintenance stemming from this sort of organized friendship. According to Granovetter (1973) we should expect users in more central positions to find better summer jobs, access class material or residences easier, and they should be generally better informed about the latest rumors on campus. Ellison et al. (2007) came to a promising conclusion: "Our findings demonstrate a robust connection between Facebook usage and indicators of social capital, especially the bridging type. Internet use alone did not predict social capital, but intensive use of Facebook did."

3 Personality and Social Networks

In the last few decades a broad consensus has emerged that the structure of the personality trait domain can be encompassed by five major dimensions. The Five Factor Model ("Big 5") received considerable empirical support and is now the standard taxonomy to organize and measure personality traits (Goldberg 1990, Costa & McCrae 1992). The model is based on early trait research and lexical assumptions that socially relevant personality differences become encoded in the language of a population. The scientific term "personality" is conceptualized as the entire mental organization of a person's traits, where traits are defined as a cross-situational and temporally stable set of individual attributes. The five factors are extraversion, agreeableness, conscientiousness, neuroticism (emotional stability), and openness to experience (intellect). These traits are claimed to be, at least to some extent, heritable on the basis of the neurotransmitter regime (Jang et al. 1998), unaffected by external influences, and very stable after 30 years of age (McCrae & Costa 1990, Soldz & Vaillant 1999). Unfortunately, individuals tend to be less stable between 18 and 30 years of age when most people experience major life course events and transitions, such as leaving their family of origin, entering university or the job market (Asendorf & Wilpers 1998). In this period people are presumed to select an environment congruent with their dispositions, preferences and attitudes while, on the other hand, environment reinforces individual characteristics and behavior. This makes it hard to disentangle the endogenous relationship between personality and the environment. The five personality factors have been shown to relate to people's behavior in a broad variety of social contexts. It is likely that they predispose people's propensity to form more or fewer social ties, or they may be related to the extent which others form and maintain relationships with the focal actor. In the following, we will briefly discuss each personality trait and it's possible association with behavior on social network sites.

Openness. McCrae (1996) suggested that openness to experience may have the strongest influence on social and interpersonal phenomena among all of the five factors. Survey questions on the openness dimension measure the propensity of individuals to display imagination, curiosity, originality and open-mindedness. In contrast, low openness scores indicate people who are practical, traditional and down-to-earth. According to Vodosek (2003) the literature offers very little empirical evidence for an association with tie formation, but it suggests a positive relation to satisfaction and stability once the relationship is formed. In the context of SNSs and members coming from a technical university, we expect individuals with high scores on openness to be more likely to try, to use and to keep up with new social networking technologies.

Extraversion. Extraversion refers to the extent to which individuals are outgoing, active, assertive and talkative. Extraverts are expected to approach others more easily and engage in more social interaction. In contrast, individuals with low levels of extraversion tend to be "introverted," reserved, serious, and prefer to be alone or stay within close circles. Hardly surprising, extraverted individuals have been found to have larger networks and show higher contact frequencies (Russel et al. 1997, Anderson et al. 2001). For example, Asendorf & Wilpers (1998) found that extraversion was highly associated with students' interaction rates and their peer-relationship formation. Extraversion is the least controversial dimension in this context and is expected to exhibit obvious and strong effects.

Conscientiousness. Conscientiousness refers to the extent that an individual is dependable, careful, responsible, organized, and has a high will to achieve. It has been shown to be associated with performance in the workplace and is known to be the most prominent dimension in the context of education and learning, exhibiting substantial correlations with grade point average, educational performance and persistence (De Raad & Schouwenburg 1996). Conscientious students have been found to have more frequent contact with family members (Asendorf & Wilpers 1998). Furthermore, one might expect this trait to be the central source of strategic network formation. Wanberg et al. (2000) reported in their study on job-seekers higher levels of contacting and asking in association with high levels of conscientiousness. However, SNSs don't promise fast and obvious returns to social networking. We expect that high scores on Conscientiousness will lead to lower numbers of contacts in this specific context. Conscientious individuals will refrain from high investments in SNS profiles, they will stick to their main goals and try to avoid such sources of distraction.

Agreeableness. Agreeable persons tend to be courteous, kind, flexible, trusting, forgiving, are inclined to cooperate but known to avoid conflict. Agreeableness is associated with positive relations to alters, and has been shown to foster peer acceptance and friendship among children from middle and junior high school (Jensen-Campbell et al. 2002). McCarty & Green (2005) report that agreeableness and conscientiousness were most highly correlated with personal network structure. At the same time they only found a small influence from extraversion. Generally, agreeableness is said to have favorable influence to social interactions and their perceived quality. Although the effect on tie formation stays unclear, we can at least presume that agreeable individuals will not reject an offer of friendship.

Neuroticism. Relatively few studies report on the relationship between neuroticism and network characteristics. It refers to the extent to which individuals experience and display negative affects like anxiety, sadness, embarrassment, depression, guilt, and is tied to the ability to cope with stress. It is negatively related with status in male social groups (measured by indegree) and the number of peer relationship (Anderson et al. 2001). Individuals with high levels on the factor neuroticism are expected to impose higher tie maintenance costs on their alters, although these negative consequences are less plausible in a computer-mediated environment with a very low tie formation threshold and, on average, low tie strength. On the other hand, people with high scores on neuroticism tend to believe that they are not attractive to others and are fearful of rejection. Therefore an intensified desire for an unstained self-presentation could result – counter-intuitively – in higher activities on SNSs. Despite the meager empirical evidence, neuroticism is generally assumed to be negatively associated with social relationships (Wanberg et al. 2000, Klein et al. 2004).

Table 1: Hypotheses for Personality Effects on Outcome Variables

	SNS Use	Adoption Time	Number of Friends	Network Centrality	Local Clustering
Openness	+	+	+	+	-
Conscientiousness	-	-	-	-	+
Extraversion	++	++	++	++	- -
Agreeableness	+	+	+	+	-
Neuroticism	-	-	-	-	+

Table 1 shows an overview of the hypotheses and the direction of the expected effects. Unfortunately, the current literature on associations between personality and network positions is characterized by mixed evidence and inconsistencies. We treat personality as exogenous and stable and try to replicate previous results. Extraversion is thought to be an intrinsic quality of an individual exhibiting evident and strong effects on networking behavior. For the other four personality traits, there are several rivaling hypotheses and potential stories to tell. We invite the reader to interpret the effects with precaution, since we were not able to derive all hypotheses from a well established theory. Additionally, we assume a homogenous influence, i.e. a trait is expected to show the same sign and effect strength on all models, except in the clustering models, where effect signs should be reversed. The reason for this will be explained below. The Five Factor Model is typically measured with extensive question inventories, sometimes with more than 100 items. Only recently short inventories have been developed and tested which are especially suited for survey research. We applied the BFI-15, an inventory developed for the German Socioeconomic Panel (Gerlitz & Schuep 2005). The inventory consists of 15 questions, measured on 7-point scales, where always 3 questions map on one of the 5 dimensions. We conducted a Principal Component Analysis (factor analysis) and were able to replicate the 5 factor nature of the FFM. Table 2 displays the factor loadings. The results are as expected. Every item loads on the predetermined dimension, although the factor loadings are in some cases unsatisfactory. The same holds for the scale reliability measures Cronbach's α. A low reliability of survey questions in an additive index indicates that the underlying dimension is not properly captured. Only the factor extraversion exhibits sufficient reliability, agreeableness is with an $\alpha = 0.54$ much too low; the other three factors are with $\alpha \approx 0.65$ at the lower end of what is commonly accepted in survey research. All personality scores are correlated with gender. Apart from neuroticism, female respondents show on all other four dimensions higher scores but similar standard deviations. Furthermore, conscientiousness seems to be at least to some extent correlated with age (r = 0.09). All other dimensions show virtually zero correlation with age.

Table 2: Factor Analysis of Personality Scores

Item / Wording in German "Ich bin jemand, der ..."	O Factor 1	C Factor 2	E Factor 3	A Factor 4	N Factor 5
04) "originell ist, neue Ideen einbring"	0.726				
09) "künstlerische Erfahrungen schätzt"	0.439				
14) "eine lebhafte Phantasie, Vorstellungen hat"	0.631				
01) "gründlich arbeitet"		0.807			
07) "eher faul ist" [-]		0.534			
11) "Aufgaben wirksam und effizient erledigt"		0.550			
02) "kommunikativ, gesprächig ist"			0.780		
08) "aus sich herausgehen kann, gesellig ist"			0.725		
12) "zurückhaltend ist" [-]			0.674		
03) "manchmal etwas grob zu anderen ist" [-]				0.554	
06) "verzeihen kann"				0.346	
13) "rücksichtsvoll und freundlich mit anderen umgeht"				0.756	
05) "sich oft Sorgen macht"					0.585
10) "leicht nervös wird"					0.591
15) "entspannt ist, mit Stress gut umgehen kann" [-]					0.722
MIC	0.376	0.388	0.550	0.287	0.389
Cronbach's α	0.643	0.655	0.786	0.548	0.657
Mean	4.828	4.937	4.520	5.305	3.756
SD	1.119	1.098	1.224	1.224	1.173
N	1540	1541	1539	1540	1539

Note: Iterated principal component analysis of personality items results in a five-factor solution. Reported are the varimax rotated factor loadings of the normalized matrix, they are omitted when |loading| < 0.3. Items 3, 7, 12 and 15 are reversed. We build additive (unstandardized) indices for later use as exogenous predictors. MIC shows the average interitem correlation for the personality indices, α is a scale reliability measure.

4 Data and Methods

Data. The data of this study comes from two distinct sources: an online survey and observations of SNS profiles. The survey was implemented using the online survey facility at the Chair of Sociology at ETH Zurich and was active between November 22 and December 24 2007. Students were randomly recruited from the official university address list, i.e. we randomly invited 66% of the student population (13'000 students) by e-mail and received 1560 completed questionnaires. This corresponds to a response rate of 17.5%. Female respondents are over-represented by about 10% in comparison with official enrollment count (40% in survey vs. 30% in student body). Students from management and social sciences are over-represented while the proportion of engineering students is too low in our sample. As observed in other online surveys, the response rate decreases with age. The survey consisted of questions about SNS usage intensity and preferences and also confirmed those indicators we collected simultaneously from the SNS site. The correlation between self-reported and observed indicators are very high (e.g. number of friends: 0.968, sex: 0.995), suggesting that most students fill out their online profiles truthfully.

At the same time, on November 22, we collected all StudiVZ profiles of ETH students with an automated web crawling approach. We started with a highly active user and followed the lists of friends until no new ETH member was found. Students are impossible to reach with this procedure if they are not connected to someone in the giant component of the graph. However, according to our survey, we expect less than 2% of all students to be isolates in the friendship network collected by crawling. This snowballing procedure resulted in 7318 observed profiles. Compared with the enrollment statistics, we estimate that 65% of ETH students participate in StudiVZ (excluding graduate students). The direct question in the survey indicates a StudiVZ participation rate of 67.5%. Ten percent of students maintain a membership at another platform such as Facebook, MySpace, or Xing. The two data sources were matched using first and last name. In cases where students have identical names, we used study subject and birthday. Record linkage was successful in 97.9% of all cases. Comparing our survey respondents with the other platform members, we find our survey respondents are more active, with a mean of 24.5 friends versus 21.1 and a median of 19 friends versus 16.

Analysis. We apply maximum likelihood estimation and simple OLS regressions, in total eleven statistical models (M1-M11). For explaining participation on social network sites we perform a logistic regression on the dependent variable of having at least one membership at a SNS (M1: 0=none, 1=at least one). For model 2 (M2) The adoption rate is modeled with a Cox proportional hazards model as expressed in equation (1),

$$h(t|X) = \lambda(t)\exp(\beta' X) \tag{1}$$

where $\lambda(t)$ is the time-dependent baseline hazard function, X are vectors of independent variables and β are regression coefficients. The dependent variable is the duration in days from the opening day of StudiVZ (Oct 31 2005) to the time when respondents sign up at StudiVZ. We control for different exposure times with cohort dummies. Positive regression coefficients indicate a higher sign-up rate, i.e. theses respondents adopt faster.

For the degree distributions and changes in the number of friends we use count models (M3-M5). The number of friends Y_i of student i is assumed to follow a negative binomial distribution with expected value λ_i and variance σ^2 (Negbin II, Cameron & Trivedi 1986).

$$E(Y_i|X) = \lambda_i = \exp(\beta' X) \quad \text{and} \quad \sigma^2 = \lambda_i + \alpha\lambda_i^2 \tag{2}$$

The expected event occurrence λ equals that of a poisson regression, but is itself a random variable drawn from a Gamma distribution. The continuous gamma mixture of Poisson distributions results in the negative binomial distribution where the conditional variance exceeds the conditional mean. The dispersion parameter α serves to estimate the level of overdispersion independently of the mean. Different exposure times are normalized by introducing the log of the duration form the sign-up date at StudiVz to the survey response date with a regression coefficient constrained to equal 1.

The last set of models in table 4 are OLS regressions where network node indices of the friendship network are regressed on independent variables. We report standardized beta coefficients and bootstrapped standard errors. The first set of node level indices are the centrality measures degree, closeness, betweenness (Freeman 1979), and eigenvector centrality (Bonacich 1972). Centrality measures capture the relative position of an individual within the network. Degree-based measures focus on the level of communication activity, while betweenness captures stress control and the capacity to interrupt communication. Closeness refers to the freedom from such control in measuring how close/far apart an individual is from everybody else in the network. Eigenvector centrality reflects the fact that individuals might profit from well connected friends; it is the positive multiple of the sum of adjacent centralities recursively solved for the entire network. The second set of two indices are in essence densities of the ego-networks. Network constraint measures the extent to which ego is invested in people who are invested in other of ego's alters. It is typically applied in valued networks, but well suited for our setting as well. Burt et al. (1998) used this indicator to show that individuals with "entrepreneurial personalities" avoid redundant investment. Transitivity, last but not least, measures the probability that an individual's neighbors are connected; it's sometimes also called the clustering coefficient. Local clustering coefficients for each node were computed, measuring the ratio of triangles connected to the vertex and the triples centered on the vertex. See Wasserman & Faust (1994) for detailed graph theoretic definitions. We will not discuss the centrality and clustering models in detail. The aim of the analysis is to show that an external source of variation like personality will be associated with a wide array of node indices, given it is in strong relation with the degree of a node. All models use the same independent variables. Apart from the five personality scores, we include gender, age in years, and a set of 8 dummies for cohort-groups (entry cohorts by year) with the freshmen as the reference category. Estimates for the cohorts are omitted in the tables. The cohort dummies exhibit always the same pattern: a parabolic influence with a maximum at the end of the undergraduate study time.

5 Personality Effects on Social Networks

In Table 3 and 4, personality effects on several indicators of networking behavior and network position are reported. Three of the personality scores, conscientiousness, extraversion and neuroticism, have a significant effect in the majority of models. There is no support from openness and mixed results for agreeableness. First, we discuss the models capturing networking activity and then we will report on the associations with centrality and clustering and discuss some further structural properties of the ETH/StudiVZ friendship network.

Activity. In M1 we model the determinants of SNS participation. Estimates from a logistic regression indicate that extraversion is the most prominent predictor of joining and, as expected, people with high scores in conscientiousness refrain from social networking. A standard deviation increase in extraversion alters the odds of being a member by 49%, while a standard deviation increase in conscientiousness leads to a 13% decrease in the odds of a registration, holding all other independent variables constant at their mean. Age is the most important expla-

Table 3: MLE of Personality Effects on Network Activity

	M1	M2	M3	M4	M5
			Number and Growth of Friends		
	Participation	Adoption	Friends T_2	Friends T_4	Growth Δ_{2-4}
Openness	-0.007	-0.002	0.018	0.014	0.055
	(0.064)	(0.031)	(0.020)	(0.021)	(0.045)
Conscientiousness	-0.129*	-0.060*	-0.048*	-0.039*	-0.024
	(0.063)	(0.029)	(0.020)	(0.019)	(0.046)
Extraversion	0.327***	0.163***	0.211***	0.222***	0.143***
	(0.061)	(0.028)	(0.019)	(0.019)	(0.043)
Agreeableness	0.086	0.102**	-0.008	0.006	-0.016
	(0.069)	(0.033)	(0.025)	(0.023)	(0.056)
Neuroticism	0.042	0.042	0.051*	0.050*	0.101*
	(0.055)	(0.030)	(0.021)	(0.021)	(0.047)
Female	-0.056	-0.090	0.065	0.049	0.023
	(0.141)	(0.071)	(0.045)	(0.045)	(0.097)
Age (years)	-0.139***	-0.118***	-0.024	-0.024	-0.014
	(0.023)	(0.015)	(0.018)	(0.017)	(0.030)
Constant	3.014***	–	0.711^{+}	0.776*	1.398^{+}
	(0.755)		(0.404)	(0.365)	(0.733)
Cohort Dummies	Yes	Yes	Yes	Yes	Yes
Exposure			Yes	Yes	No
ln(α)			-1.011***	-1.083***	0.097
			(0.055)	(0.051)	(0.074)
N	1520	1520	922	942	906

Note: Maximum likelihood estimates of personality effects. Bootstrapped standard errors in parentheses. $^{+}$ $p < 0.10$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$. Model 1 is a logistic regression on having at least one membership at a social networking site. Model 2 is a Cox proportional hazards rate model of the duration until the start of the StudiVZ membership. Models 3, 4 and 5 are negative binomial regressions with overdispersion and normalized exposure time (Negbin II). Personality scores are unstandardized, cohort dummies are not reported.

nation for being a member at a network site. Students who are one standard deviation older are 8.2% less likely to be members. Given that other studies (eg. Lampe et al. 2007, Tufekci 2008) have found that women have higher participation and activity rates, it is surprising we find no gender effect at all for ETH students. M2 is similar to M1 but with the major difference that we measure the time until registration takes place. Non-members are treated as censored and are kept in the sample for an adequate estimation of the set of people at risk of joining StudiVZ. The Logit and Cox models show very similar coefficients and marginal effects, with the only difference between the two models being a positive significant association of agreeableness with adoption time. We didn't expect any positive results due to its low scale reliability. Presumably agreeable individuals don't easily reject invitations from other users.

Models M3-M5 are the count models, where we estimate effects on the number of friends a user has acquired. The dependent variable of M3 is the number of friends the individual has had at the time of the survey, while the dependent variable in M4 is the number of friends 4 month later. In

M5 we regress independent variables on the difference between the two. Again, extraversion has the largest influence on friendship. For a standard deviation increase in extraversion, a student's expected mean number of friends increases by 29%. Conscientiousness significantly reduces the number of friends in M3 and M4, while neuroticism significantly increases the counts in all three models. Neuroticism's effect was unanticipated. While all evidence from team and work group studies indicate debilitating influences from emotional instability, we find evidence for the opposite. One possible clarification comes from the time students spend on the networking site. A path analysis revealed that neuroticism and conscientiousness are moderated by time investment. While conscientious students tend to visit the site less frequent and for shorter periods of time, the opposite holds true for individuals with high scores on neuroticism. Extraversion and conscientiousness show direct and moderated effects, while neuroticism only shows a moderated effect. In essence, a high level of emotional instability is not a direct cause for having more friends, it's rather the cause for intensified online networking behavior. Further analysis is needed to clarify this puzzle.

Node level indices. We now turn to the analysis of associations between personality and indices of the complete network, i.e. for every survey respondent node level indices were computed from the entire network collected by the snowballing procedure. Before looking at the OLS estimates, some of the global structural characteristics of the friendship network are briefly reviewed. The network consists of 7318 nodes and 78631 edges at the reference time when the online survey starts (T_2). Two month earlier (T_1, September 2007) and shortly before new freshmen arrived for the autumn term, the network was populated with 5975 students. Only within 2 months the number of nodes in the network rose by 20%. On the two subsequent periods up to wave 4, we observed further increases by 6% and 2% which resulted in 7930 members in March 2008. In the same period the number of edges rose by 30%, while the density of the network slightly decreased from 0.0035 to 0.0029. The following structural measures of the friendship network are computed for the reference time point T_2. Hardly surprising, ETH is a small world. The network has a diameter of 10 and an average shortest path length of 3.64. Small Worlds are defined as having short average path length but high clustering, both requirements are met. The global clustering coefficients for the friendship network is 0.28, far beyond from what we would expect from a random network with equal density and size. This indicates that the network exhibits typical triadic closure and cliquishness regularly observed for social networks. Although the network lacks extreme events and huge hubs, the global structural measures strongly resemble several other social networks described in Newman (2002). We will come back to a last structural metric of social networks – assortativity – at the end of the section.

Table 4 presents personality effects on the network positions of our survey respondents. In model M6, where the logarithm of the degree is the dependent variable, we simply replicate the results from the count model M3 by means of linear regression. In essence, M3 and M6 are identical. All centrality models indicate the same three personality traits at work, which were also significant in the activity models: extraversion, conscientiousness and neuroticism. Extraversion regularly displays effect sizes that are three times as strong as the other two factors, making it the most relevant trait for studying large social networks. Once more, we find no evidence for a gender bias. Age is strongly negatively associated with degree and closeness, but has no in-

Table 4: OLS Estimates of Personality Effects on Node Level Indices

	Centrality Measures				Local Clustering	
	M6 Degree	M7 Closeness	M8 Betweenness	M9 Eigenvector	M10 Constraint	M11 Transitivity
Openness	0.036	0.030	0.048	0.014	-0.043	0.006
	(0.025)	(0.001)	(0.049)	(0.042)	(0.022)	(0.006)
Conscientiousness	-0.055^{+}	-0.075**	-0.081*	-0.067*	0.061*	0.054
	(0.025)	(0.001)	(0.055)	(0.043)	(0.022)	(0.006)
Extraversion	0.326***	0.296***	0.304***	0.248***	-0.322***	-0.198***
	(0.023)	(0.001)	(0.047)	(0.040)	(0.020)	(0.005)
Agreeableness	-0.002	-0.004	-0.001	-0.001	0.009	0.018
	(0.027)	(0.001)	(0.055)	(0.051)	(0.025)	(0.006)
Neuroticism	0.084**	0.070*	0.106**	0.067*	-0.070*	0.017
	(0.024)	(0.001)	(0.052)	(0.043)	(0.022)	(0.006)
Female	-0.025	-0.003	0.004	-0.008	-0.011	0.068^{+}
	(0.057)	(0.002)	(0.116)	(0.094)	(0.050)	(0.013)
Age	-0.163**	-0.093*	-0.070	-0.026	0.164**	0.181**
	(0.021)	(0.001)	(0.039)	(0.034)	(0.018)	(0.005)
Cohort Dummies	Yes	Yes	Yes	Yes	Yes	Yes
Adjusted R^2	0.227	0.348	0.122	0.412	0.245	0.046
N	958	958	921	958	958	943

Note: Standardized beta coefficients of OLS regressions on node level indices, bootstrapped standard errors (of the unstandardized coefficients) in parentheses, $^{+}$ $p < 0.10$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$. Network measures were computed on the entire friendship network consisting of 7'318 nodes and 78'631 edges with R(Igraph). Due to skewness we take logarithms of the degree, betweenness, eigenvector and constraint scores. Cohort dummies are not reported.

fluence on betweenness and eigenvector centrality. The similarity of the four centrality models is of no surprise. It is well known that all measures of centrality in graphs tend to be correlated with degree. In this study, the number of friends is the overruling characteristic of the network. Attributes strongly associated with degree will be handed over to all other local graph metrics. Having a high degree inevitably results in higher centralities, and is at the same time source of lower local clustering around the focal actor. Nodes with very high degrees tend to have low clustering rates since the many nodes that are connected to them are relatively unlikely to be linked to each other. Jackson and Rogers (2007) claim this to be a key empirical regularity for social networks: "The clustering among the neighbors of a given node, in at least some social networks, is inversely related to the node's degree." In our study, the correlation between degree and the local clustering measure is r = -0.35. Accordingly, the coefficients flip sign between the centrality and clustering models. High scores in extraversion results in the fact that the first neighbors of this individual are less integrated on average (M11). This lowers at the same time the probability of investment in redundant ties, captured by M10 with network constraint as the dependent variable. In the study of Burt et al. (1998) an "entrepreneurial personality" was responsible for low levels of network constraint. This coincided with a high number of structural holes in the networks of those MBA students who claim the personality of the "entrepreneurial outsider." To what extent extra edges are in fact bridges, cannot be answered at this place and needs further investigation. Surprisingly, we find only week evidence for a gender difference in transitivity. Women are known to have a higher proportion of kin in their ego-centered networks and are presumed to invest more time in the near graph neighborhood, resulting in a higher number of closed triads around them. The gender indicator variable in model M11 is only significant on the 10% level, questioning a strong association with gender.

Finally, we have to address the questions whether similar personalities meet. Friendship choices based on similarity in attributes – better known under the term homophily – leads us to the last prominent characteristic of social networks: assortativity. Nodes of the same kind tend to cluster. We found no indication of assortative personality mixing in the literature, suggesting that no specific preferences for people of the same or different personality type exist. Assortative mixing is most easily calculated by a Pearson's correlation coefficient r for the attributes at either side of an edge (Newman 2002). Unfortunately, the subgraph, spanned over the survey respondents by the relations they share among one another, is only a tiny and potentially biased fraction of the whole friendship network (883 nodes, 4182 edges, and a tenfold density of 0.01). Regardless of that fact, we calculated the node-node attribute correlations (with bootstrap resampling) and found only negligible deviations from a perfectly mixed random graph. Openness-Openness showed a correlation coefficient of 0.076, all other possible combinations of personality traits are well below 0.05. This is radically different for a set of other mixing patterns. The degree-degree correlations has a coefficient of 0.20, gender mixing is at the level of r=0.15. The strongest deviations arise from age and cohort. Assortativity by age exhibits a correlation of 0.5, cohort is even higher, indicating that friendship choices most often take place within and not between the age groups. The heavy constraining nature of age on the tie formation process in friendship networks is a well known fact in the literature on sociometric choices at school. However, we find it surprising that this pattern is so persistent even on the university level. To summarize, personality seems to be predictive in how many persons someone meets, what position the person takes in the network, but not in the type of personality the person's interaction partners will have.

6 Conclusions

The results of this study show that extraversion, one dimension of the Big Five Personality Model, plays an important role in the formation of network ties. We conducted an online survey to collect personality scores of 1560 students from a major Swiss university. At the same time, we observed students' online profiles on the popular social network site StudiVZ. Extraverts show a higher probability in joining StudiVZ, they adopt the technology faster and accumulate more friends on their contact lists. Accordingly, individuals with high scores on extraversion take more central positions in the friendship network. All centrality measures applied in this study were highly correlated with degree. So it comes with no great surprise that all centrality models exhibit about the same association with extraversion. Conversely, the clustering among the neighbors of a given node is found to be inversely related to the node's degree. In the models with network constraint and local transitivity as their dependent variable, most of the independent variables changed sign, therefore confirming this inverse relationship. Apart from extraversion, we found retardant effects from conscientiousness. Highly conscientious people tend to refrain from participation on social networking sites, suggesting that they successfully evade this popular source of distraction in student's everyday life. Surprisingly, we found activating, positive effects from neuroticism which stands in sharp contradiction to theory and the majority of empirical findings reported in the literature. One possible explanation is that people exhibiting high levels of emotional instability tend to spend more time on social network sites. In being fearful of rejection, they might try harder to present themselves well in an unstained and attractive manner. There is virtually no support for an association between network characteristics and the last two factors openness and agreeableness. Although agreeableness seems to accelerate the adoption of social networking technology, we found no associations in the count and centrality models. In all models, extraversion exhibits effects three to six times stronger than any other personality trait, making it the most important personality factor in large social friendship networks. Extraversion is responsible for about 10% of the variance in the number of friends, while age and entry cohort are responsible for another 10%. We found no evidence for gender effects in general, and only a week indication that women tend to have higher levels of triadic closure in their ego-networks. Clearly, personality does not tell the whole story. While extraversion fits perfectly into the wider picture, effects from conscientiousness and neuroticism seem to be particularities of online social networks or just the data at hand.

The degree distribution of the number of friends shows high variability and is best described with the negative binomial distribution. By introducing personality as an exogenous and temporally stable source of variation into count models, we were able to show that the skew of the distribution is not entirely dominated by contagious processes in the network. However, it is very likely that people who have a lot of friends at time point 1 will profit from their visibility and might acquire disproportionately more contacts in a subsequent period. Unfortunately, the distinction between heterogeneity and contagion is not an easy undertaking, since both sources of variation give rise to the same distribution. Distributional evidence is not enough to demonstrate the presence of a specific tie formation process. Further analysis beyond the cross-sectional framework and the application of dynamic panel models is needed to discriminate the two potential sources of excess variability.

Apart from the direct effects of personality on the probability of a SNS membership, on the adoption rate, on the number of friends, and on network centrality, we found no evidence that personality traits are responsible for similarity effects. Compared with assortativity based on cohort, age, degree, and gender, personality will not be of great importance to tie formation models based on attribute similarity. Results from the survey indicate a substantial overlap between offline and online circle of acquaintances. Social network sites are just another place where students nowadays meet, opening up the possibility to observe their behavioral traces. The ETH friendship network is characterized by a moderately skewed degree distribution, an average shortest path length of 3.6, a high average clustering coefficient of 0.28, positive assortativity and an inverse relation between clustering and degree, therefore exhibiting all key empirical regularities shared by socially generated networks. This makes these easy to collect SNS data sources an interesting alternative to study processes of friendship formation on a larger scale basis.

7 References

Anderson, C., O.P. John, D. Keltner, and A.M. Kring (2001): Who Attains Social Status? Effects of Personality and Physical Attractiveness in Three Social Groups. *Journal of Personality and Social Psychology* 81:116-132.

Asendorpf, J.B., and S. Wilpers. 1998. Personality Effects on Social Relationship. *Journal of Personality and Social Psychology* 74(6): 1531-1544.

Bianconi, G., and A.-L. Barabási (2001): Competition and multiscaling in evolving networks. *Europhysics Letters* 54: 436-442.

Bonacich, P. (1972): Factoring and weighting approaches to status scores and clique identification. *Journal of Mathematical Sociology* 2, 113-120.

Boyd, D.M., and N.B. Ellison. 2007. Social Network Sites: Definition, History, and Scholarship. *Journal of Computer-Mediated Communication* 13(1).

Burt, R.S., E.J. Janotta, and J.T. Mahoney (1998): Personality Correlates of Structural Holes. *Social Networks* 20:63-87.

Cameron, A.C., and P.K. Trivedi (1986): Econometric Models Based on Count Data: Comparisons and Applications of Some Estimators and Tests. *Journal of Applied Econometrics* 1(1): 29-53.

Costa, P.T., and R.R. McCrae (1992): Revised NEO Personality Inventory NEO PI-R and NEO Five-Factor Inventory NEO-FFI: Professional Manual. Odessa, FL: Psychological Assesment Resources.

Coleman, J.S. (1964): Introduction to Mathematical Sociology. New York: Free Press.

De Raad, B., and H.C. Schouwenburg (1996): Personality in Learning and Education: A Review. *European Journal of Personality* 10, 303-336.

Ellison, N.B., C. Steinfield, and C. Lampe (2007): The Benefits of Facebook "Friends": Social Capital and College Student's Use of Online Social Network Sites. *Journal of Computer-Mediated communication* 12(4).

Freeman, L.C. (1979): Centrality in Social Networks: Conceptual Clarification. *Social Networks* 1: 215-239.

Gerliz, J-Y., and J. Schupp (2005): Zur Erhebung der Big-Five-basierten Persönlichkeitsmerkmale im SOEP. Berlin: DIW Research Note.

Goldberg, L.R. (1990): An Alternative Description of Personality: The Big-Five Factor Structure. *Journal of Personality and Social Psychology* 70:820-835.

Granovetter, M.S. (1973): The Strength of Weak Ties. *American Journal of Sociology* 78(6): 1360-1380.

Hallinan, M.T. (1978): The Process of Friendship Formation. *Social Networks* 1: 193-210.

Handcock, M.S., and J.H. Jones (2004): Likelihood-based inference for stochastic models of sexual network formation. *Theoretical Population Biology* 65: 413-422.

Jackson, M.O., and B.W. Rogers (2007): Meeting Stranger and Friends of Friends: How Random Are Social Networks? *American Economic Review* 97(3): 890-914.

Jang, K.L., R.R. McCrae, A. Angleitner, R. Riemann, and W.J. Livesly (1998): Heritability of Facet-Level Traits in a Cross-Cultural Twin Sample: Support for a Hierarchical Model of Personality. *Journal of Personality and Social Psychology* 74(6): 1556-1565.

Jensen-Campbell, L.A., R. Adams, D.G. Perry, K.A. Workman, J.Q. Furdella, and S.K. Egan (2002): Agreeableness, Extraversion, and Peer Relations in Early Adolescence: Winning Friends and Deflecting Aggression. *Journal of Research in Personality* 36:224-251.

Klein, K.J., B. Lim, J.L. Saltz, and D. M. Mayer (2004): How do they get there? An Examination of the Antecedents of Centrality in Team Networks. *Academy of Management Journal* 47(6), 952-963.

Lampe, C., N. Ellison, and C. Steinfield (2007): A Familiar Face(book): Profile Elements as Signals in an Online Social Network.*CHI 2007 Proceedings*, San Jose, CA.

McCrae, R.R., and P.T. Costa (1990): *Personality in Adulthood*. New York: Guilford Press.

McPherson M., L. Smith-Lovin, and J.M Cook (2001): Birds of a Feather: Homophily in Social Networks. *Annual Review of Sociology* 27: 415-444.

Mehra, A., M. Kilduf, and D.J. Brass (2001): The Social Networks of High and Low Self-Monitors: Implications for Workplace Performance. *Administrative Science Quarterly* 46(1): 121-146.

Moody, J. (2001): Race, School Integration, and Friendship Segregation in America. *American Journal of Sociology* 107(3): 679-716.

McCarty, C., and H.D. Green (2005): Personality and Personal Networks. *Sunbelt XXV*, Conference Contribution.

Newman, M.E.J., S.H. Strogatz, and D.J. Watts (2001): Random graphs with arbitrary degree distributions and their applications. *Physical Review E* 64(2): 026118.

Newman, M.E.J. (2002): Assortatitive Mixing in Networks. *Physical Review Letters* 89(20): 208701.

Newman, M.E.J. (2003): The structure and function of complex networks. *Siam Review* 45(2): 167–256.

Russel, D.W., B. Booth, D. Reed, and P.R. Laughlin (1997): Personality, Social Networks, and Perceived Social Support among Alcoholics: A Structural Equation Analysis. *Journal of Personality* 63(3): 649-692.
Schaefer, C., A. Geyer-Schulz, and S. Berninghaus (2008): Personality in Social Networks: A Theoretical Overview. In: Dreier, T., R. Studer, and C. Weinhardt (Eds.): Information Management and Market Engineering. Karlsruhe: Universitätsverlag.
Soldz, S., and G.E. Vaillant (1999): The Big Five Personality Traits and the Life Course: A 45-Year Longitudinal Study. *Journal of Research in Personality* 33: 208-232.
Tufekci, Z. (2008): Grooming, Gossip, Facebook and MySpace: What can we learn about these sites from those who won't assimilate? *Information Communication & Society* 11(4): 544-564.
Vodosek, M. (2003): Personality and the Formation of Social Networks. Conference paper, ASA Annual Meeting.
Wanberg, C.R., R. Kanfer, and J.T. Banas (2000): Peredictors and Outcomes of Networking Intensity Among Unemployed Job Seekers. *Journal of Applied Psychology* 85:491-503.
Wasserman, S., and K. Faust (1994): Social Network Analysis: Methods and Applications. Cambridge: Cambridge University Press.
Watts, D.J. (1999): Networks, Dynamics, and the Small-World Phenomenon. *American Journal of Sociology* 105(2): 493-527.

8 Appendix

Table 5: Descriptive Statistics

Dependent Variables				Independent Variables			
	Mean	SD	N		Mean	SD	N
Activity				Openness	4.828	1.119	1540
Participation	0.676	–	1560	Conscientiousness	4.937	1.098	1541
Adoption Time	556.05	141.37	1054	Extraversion	4.521	1.224	1539
Friends T_2	24.597	20.516	946	Agreeableness	5.305	0.947	1540
Friends T_4	29.013	21.745	966	Neuroticism	3.756	1.173	1539
Growth Δ_{2-4}	5.954	9.242	999	Female	0.396	–	1560
				Age (years)	22.96	3.778	1560
Node Indices				Cohort 1	0.281	–	1535
Degree	24.617	20.452	975	Cohort 2	0.161	–	1535
Closeness	0.283	0.028	975	Cohort 3	0.141	–	1535
Betweenness	11488.6	18891.72	975	Cohort 4	0.112	–	1535
Eigenvector	0.029	0.069	975	Cohort 5	0.122	–	1535
Constraint	0.123	0.156	975	Cohort 6	0.102	–	1535
Transitivity	0.283	0.182	959	Cohort 7	0.049	–	1535
				Cohort 8	0.031	–	1535