Tweeting Alone?

A Computational Analysis of Twitter Users' Social Capital and their Expressed Well-Being

Student Name: Timo Stöttner
Student Number: 383447

Supervisor: Dr. Ericka Menchen-Trevino

Media, Culture and Society
Erasmus School of History, Culture and Communication
Erasmus University Rotterdam

Master Thesis
June 20th, 2014

Table of Contents

Abstract

Much scholarly research has been devoted to the relation of social capital and happiness (for a review see Helliwell, 2001). This study builds on such previous research by examining whether social capital can be used to predict the expressed well-being of Twitter users. Apart from its focus on Twitter users, the contributions of this study are twofold. First, it distinguishes between bonding, bridging and overall social capital, a differentiation that has been neglected in most previous studies examining the relationship between social capital and happiness although most theoretical discussions of social capital make this distinction.

Second, I take the perspective of the rather new discipline of Computational Social Science, which aims to study societal processes through the analysis of large volumes of data. I used a computational framework for the determination of social capital developed by Smith, Giraud-Carrier, Ventura et al. (2011) to determine the social capital of Twitter users and I calculated their expressed well-being based on a sentiment analysis of their tweets. To my knowledge, the framework for the determination of social capital has not been applied to Twitter before to this extent and I expanded on it in several ways to make it applicable to this study.

Since research suggests that the positivity of tweets is indicative of their author's happiness (Kramer, 2010; Bollen et al., 2011) and much research shows that social capital is related to happiness (for a review see Helliwell, 2001), I expected to find that social capital can be used to predict the expressed well-being of Twitter users.

I analyzed the social capital and expressed well-being of 214 Twitter users that were determined randomly according to specific criteria which made them suitable for the analysis. In order to determine their social capital, I collected a dataset consisting of 43,670,346 user profiles and 79,579,346 tweets**.** Contrary to the expectations, neither bonding, bridging nor overall social capital were found to be significant predictors of Twitter users' expressed well-being.

Keywords: Social capital, happiness, Twitter, computational social science, sentiment analysis, social network analysis

1. Introduction

Scholarly interest in the topic of 'happiness' has recently experienced a significant rise (Gauntlett, 2011). An increasing number of scholars seem to acknowledge that the question of what makes people happy is linked to many political dilemmas and countless social issues. Layard (2006) argues in his influential book *Happiness* that happiness research has the potential to directly improve people's lives since, according to him, the quality of people's lives is mainly dependent on their happiness. While Layard's claims might seem bold to some (e.g. Graham, 2008), there is increasing consensus among economists that happiness research can be of use for the improvement of public policies (Graham, 2008).

Scholars conducting research into this new 'science of happiness' (Dutt & Radcliff, 2009, p. 1) come from various disciplines, including psychology, economics, philosophy, sociology and political science. Among this large number of researchers is the political scientist Robert Putnam who showed in his influential book *Bowling Alone* (Putnam, 2000) that people's 'social capital' is closely related to many forms of individual well-being.

What is social capital and why is there reason to assume that it promotes happiness? The overarching idea of theories of social capital is that social ties are a resource (Putnam, 2000), just like economic, human and intellectual capital are resources people can draw on to achieve certain goals. These social ties are of value because people we know might be ready to help us when we need them, depending on the quality of our relationship and the kind of help we are looking for. However, what social capital is about more exactly is highly disputed among scholars, as I will further discuss below.

A large amount of research has been done examining how social capital is related to happiness (see Helliwell, 2001). Since social capital is about relationships between people and interpersonal relationships contribute to our happiness, there is much reason to assume that social capital makes us happy. Indeed, the vast majority of studies examining the link between social capital and happiness have found this link to be positive (for reviews see Pacek, 2009 and Helliwell, 2001).

Most research studying social capital measures it through surveys asking people about the nature of their social connections, their participation in civic organizations and their perceptions of trust, among other things (Szreter & Woolcock, 2004). However, since the internet has changed the nature of social interactions and social capital is about social relationships that are formed through interactions, it has been argued that new measures need to be developed for studying social capital on the internet (Williams, 2006). Smith, Giraud-Carrier, Ventura et al. (2011) addressed this issue by developing a computational framework for the determination of social capital in online communities which infers people's social capital from their connections with others. In this paper I make use of Smith et al.'s framework, expand on it and adapt it to my case. Specifically, I apply the framework to Twitter and examine whether Twitter users' social capital determined in this way can be used to predict the relative positivity of their tweets.This paper therefore follows much scholarly research examining the link between social capital and happiness (for an extensive review see Helliwell, 2001), however from the perspective of the rather new discipline of Computational Social Science which aims to study societal phenomena through the analysis of large volumes of data (Conte et al., 2012). The overarching research question is: *Does the social capital of Twitter users predict their expressed well-being?*

By 'expressed well-being' I refer to a relative measure of the positive and negative tweets a Twitter user publishes. The larger the amount of positive tweets someone publishes in relation to negative ones, the higher his or her expressed well-being. Research shows that Twitter users who are more satisfied with their lives publish more positive tweets (Kramer, 2010) and it has been argued that tweets also reflect the current mood of their authors (Bollen, Mao, & Pepe, 2010). Thus, I view expressed well-being as an indicator of happiness, as I will further discuss below. Since much research shows that social capital is related to happiness (Helliwell & Putnam, 2004; Helliwell, 2001), I expect that social capital is also predictive of expressed well-being.

Like many other discussions of social capital, the framework of Smith, Giraud-Carrier, Ventura et al. (2011) differentiates between ‘bonding’ and ‘bridging’ social capital. Bonding social capital usually refers to homogeneous groups with strong ties among their members, and bridging social capital refers to heterogeneous groups with weaker ties among their members (Putnam, Leonardi, & Nanetti, 1993).

Bonding social capital is sometimes considered as contributing to the 'dark side of social capital' that can reduce the general well-being of a society, while bridging social capital is mostly seen as unrestrictedly desirable (Putnam, 2000). In his book *Bowling Alone,* Putnam (2000) advocated the need for policies that promote the formation of bridging social capital, arguing that bridging social capital is essential for an effective public debate that is needed to sustain a democracy.

However, it is not clear whether bonding social capital and bridging social capital differ in their effects on people's happiness. Despite the large amount of literature discussing the relation of social capital and happiness, only very little research has been found clearly differentiating between the effects of bonding and bridging social capital on people’s well-being. The only piece of research found differentiating between the relations of bridging social capital and bonding social capital with measures of happiness was relatively small-scale and restricted to U.S. college students (Burke, Marlow, & Lento, 2010). Generally, while most theoretical discussions differentiate between bonding and bridging social capital, most empirical research neglects this distinction (Patulny & Svendsen, 2007).

In line with much current research that aims to derive suggestions for public policies from happiness indicators (Layard, 2006), however, it is a relevant question to ask which one makes people happier. Should governments promote policies that aim to foster bridging social capital, as argued by Putnam? And should they also be attentive of bonding social capital due to undesirable effects from the perspective of happiness research? To address these issues, I will pursue the following two subquestions:

RQ 1: Is bonding social capital of Twitter users a significant predictor of their expressed well-being?

RQ 2: Is bridging social capital of Twitter users a significant predictor of their expressed well-being?

In order to be able to answer the overarching question whether social capital per se is predictive of expressed well-being, the third subquestion asks:

RQ 3: Is the overall social capital of Twitter users a significant predictor of their expressed well-being?

While these questions do not address whether social capital promotes happiness, there is reason to assume that social capital in fact causes people to be happier (Helliwell, 2001), as I will further discuss below. Assuming that social capital makes people happy, these questions are relevant to ask in light of many current efforts to improve public policies through happiness research (e.g. Diener & Seligman, 2004; Layard, 2006).

While there is little research examining how bridging social capital per se is linked to happiness, there is some research looking at related variables suggesting that the link is positive. For instance, research has shown that generalized trust in others, which is considered an outcome of bridging social capital (Putnam, 2000), is associated with greater satisfaction with life (Helliwell, 2003) and better education (La Porta, 1997), which in turn is positively linked to happiness (Putnam, 2000). Therefore, I expect to find that bridging social capital is predictive of Twitter users' expressed well-being and that the correlation is positive.

However, the literature is discordant about the effects of bonding social capital on happiness. On the one hand, it has been argued that the high levels of trust that bonding social capital tends to promote can enhance people's happiness (Helliwell, 2001), like in the case of close family members. On the other hand, exclusive groups can also have negative effects on people's well-being when they segregate people from each other or hinder people from doing what they value (Putnam, 2000), like for example in the case of terrorist organizations. Due to these apparently discordant effects of bonding social capital on individual well-being, bonding social capital might not be a significant predictor of happiness.

As mentioned above and as I will discuss more thoroughly below, there is much research suggesting that social capital per se promotes happiness. Therefore, I expect to find that overall social capital is a significant predictor of happiness and that they are positively correlated.

The contributions of this paper are both theoretical and methodological. As mentioned above, its main theoretical contribution is the differentiation between bridging and bonding social capital in terms of their relation to Twitter users' happiness. There exists much research into whether or not the internet makes people happy (for a review see Amichai-Hamburger & Barak, 2009), and a significant part of this research addresses social networking sites such as Twitter. While I do not directly address whether or not the use of Twitter makes us happy, this paper contributes to an understanding of how homogeneity and heterogeneity in online social networks are related to people's happiness. Thereby it might improve the understanding of the circumstances under which online interactions either promote or reduce people's happiness.

Additionally to this theoretical contribution, the paper contributes to the development of methods for the computational determination of social capital. The framework developed by Smith, Giraud-Carrier, Ventura et al. (2011) has so far only been partially applied to Twitter in a study conducted by two of the authors themselves (Smith & Giraud-Carrier, 2010). I will extend the methods developed by Smith, Giraud-Carrier, Ventura et al. (2011) and thereby make the framework applicable to a wider range of studies than it is with the methods applied in Smith and Giraud-Carrier's experimental study. Thereby, as mentioned above, this paper contributes to the discipline of Computational Social Science by developing methods that can be used to study societal phenomena unobtrusively through big data analysis. Moreover, the paper further tests the validity of the computational framework of social capital by checking whether it conforms with research that has shown a positive association between social capital and happiness. While the authors claim to have partially validated their framework themselves in case studies (Smith, Giraud-Carrier, Ventura et al., 2011, pp. 91), they acknowledge that more research is needed that tests it against theories of the social sciences.

In the following I will first briefly explain the basics of Twitter in order to set the scene. Then I will give an overview of the scientific literature about happiness and social capital that is relevant to this paper, provide definitions of both concepts and discuss ways to measure them. In this course, I will also explain in more detail why research into happiness is useful and give a review of studies showing a link between happiness and social capital. After this theoretical part, I will explain in more detail how I conducted the research and measured expressed well-being and the different dimensions of social capital of Twitter users. Finally, I will describe the results of the analysis and discuss them in light of the theory developed earlier.

2. Theory and Previous Research

2.1 A Brief Overview of Twitter

Twitter was launched in 2006 and its user base grew to more than 200 million monthly active users in 2013 (SEC, 2013). The ‘microblogging’ platform allows its users to publish ‘tweets’ with a maximal length of 140 characters. By default, all tweets are public and can be viewed by anyone with or without a Twitter account.

Twitter users can ‘follow’ others, which lets them see the most recent tweets of the followed users on their start page. Besides following each other, users can interact with one another mainly in three different ways: *mentions*, *retweets* and *favorites*. Mentions are used to directly address a user or to talk about another user in the third person. Mentions consist of the @-symbol followed by the other user’s name. Retweets constitute a way of rebroadcasting a tweet that was created by another user and thereby increase its visibility (Conover, Goncalves, Ratkiewicz, Flammini, & Menczer, 2011). By ‘favoriting’ a tweet, users can express that they like a tweet or that they agree with it.

Moreover, users can add ‘hashtags’ to their tweets that are prefixed with a ‘#’, like for example #Obama, #Ukraine or #music. Hashtags were initially introduced by Twitter users as a means to categorize tweets, indicating that tweets are relevant to a specific topic or theme, and were in the beginning not supported by Twitter itself (Bruns & Burgess, 2011). Now, Twitter allows its users to search for hashtags and thereby makes it possible not only to follow other accounts, but also to look at what others write about certain topics.

Additionally to the mere categorization of tweets, hashtags are also used for other purposes. As Bruns and Burgess (2011) point out, hashtags are commonly used as a simple means of emphasis. For example, a hashtag like #USA is usually not used to address people who are following the hashtag, but rather as a sort of text decoration. Similarly, hashtags can be used as an equivalent to emoticons. For instance, the hashtag #tired indicates that the author is sleepy and it is usually not used to address people searching for the hashtag.

As indicated above, users have the possibility to protect their accounts, in which case their tweets can only be viewed by their followers. If a user wants to follow a protected account, he or she needs to be approved by the owner of the account first. However, only relatively few Twitter users make use of the possibility to protect their 'timelines' (i.e., all the tweets they have published that can be viewed on their Twitter page). As of 2012, only 12 per cent of Twitter users made use of the possibility to make their tweets private (Beevolve, 2012).

In order to allow developers to create new applications making use of Twitter functionalities, the company has developed the Twitter API, a programming interface that allows third party applications to access accounts and retrieve data from Twitter. The Twitter API has been used in a variety of scientific and commercial studies in order to mine data from Twitter (Kouloumpis, Wilson, & Moore, 2011). It allows one to collect a large amount of metadata about users and tweets, such as information about the location of users and where tweets were published, only some of which are visible on the Twitter website. I used the Twitter API to collect the data needed for this study, as I will describe in more detail below.

2.2 Happiness

2.2.1 Defining Happiness

Over the course of history the term 'happiness' has been given many different meanings. In fact, ever since antiquity, most discussions of happiness have been about what it is constituted of (Veenhoven, 1984).

Definitions of happiness can be classified into two major categories: the ones focussing on the 'subjective' enjoyment of life and the ones focusing on 'objectively' desirable conditions (Veenhoven, 1984). For example, a person might be blind and therefore (objectively) have a low quality of life, while he or she (subjectively) is happy.

This distinction between 'subjective' satisfaction or well-being and 'objective' quality of life is common in the scientific literature (Veenhoven, 1996). Tatarkiewicz (1975) gave an influential definition of happiness that focused on its objective component, defining happiness as "a lasting, complete and justified satisfaction with life" (p. 16). He emphasized that mere enjoyment of life does not constitute true happiness if it is not *justified* on some objective level. According to him, people cannot be happy in a situation they find repugnant, like for example prisoners can never be happy since they are imprisoned against their will.

Veenhoven (1984), on the other hand, considers happiness as "the degree to which an individual judges the overall quality of his life-as-a-whole favorably" (p. 22) and thereby bases his definition solely on the subjective emotions of an individual. According to this definition, prisoners can be happy if they feel like their life is good in prison. Veenhoven thus excludes the objective aspect of happiness: Whether or not someone is happy only depends on *their own* judgment of their life.

As Veenhoven (1984, 2000) argues, problems arise when trying to judge happiness from an 'objective' perspective. First, there is no agreement on what objectively constitutes the 'good life'. Neither ancient philosophers nor modern behavioral scientists have ever found a definite answer to such a question and it is unlikely that one will ever be found. Second, somehow related to the first point, definitions of happiness focusing on 'objective' conditions are either rather empty or arbitrary (Veenhoven, 1984): The arbitrary ones specify a desirable way of life or condition but thereby only reflect the author's ideas of how a life ought to be, since there is no definite answer to the question of what constitutes quality of life. The empty ones refer to a desirable way of life without clearly specifying how such a life should look like nor by whom it should be judged. This problem becomes clear when considering Tatarkewicz's definition of happiness as the "… justified satisfaction with life". What kind of satisfaction with life is justified and who judges it?

Therefore, I will take the view that happiness is based merely on an individual's subjective appreciation of life and independent of the 'objective' evaluation of outsiders. I will adopt a definition of happiness that equates it with subjective well-being and consider it as consisting of *"frequent positive affect, high life satisfaction, and infrequent negative affect"* (Lyubomirsky, Sheldon, & Schkade, 2005). In other words, happy people are those who predominantly experience positive emotions and who are satisfied with their lives in general.

This definition of happiness consists of two components that are common in the scientific literature about subjective well-being (Ong, 2009): the 'affective' and the 'cognitive' components of happiness. The affective component refers to moods and emotions that are rather uncontrollable and that do not require any real thinking. It only requires to distinguish between pleasant and unpleasant affects (Veenhoven, 1984). "Frequent positive affect" and "infrequent negative affect" constitute this affective component of happiness, according to the definition given above.

The cognitive component refers to an individual's evaluation of whether his or her life goals are being met. According to the definition above, "high life satisfaction" constitutes this cognitive component of happiness. In order to come to a conclusion of whether or not we are satisfied with our lives, we need to reflect on what we want to achieve and whether or not we have achieved what we want or are likely to achieve it in the future. Hence it requires more thinking activity than the affective component of happiness. It is also more controllable, since people can control their aspirations, at least to a certain extent (Veenhoven, 1984). For example, many religions and cultures advocate modesty and thereby lower the aspirations of their adherents. This distinction between the affective and the cognitive component of happiness will be of importance later in connection with the operationalization of Twitter users' well-being.

It is not clear how exactly affective and cognitive happiness are related to overall happiness and their exact relationship is likely to vary from person to person (Veenhoven, 1984). However, it has been argued that 'affective inference' is also most likely to be the main way people assess their satisfaction with life (Veenhoven, 2009), which constitutes the cognitive component of happiness: Good and bad feelings inform us whether our needs have been gratified or not. It seems unlikely that we can assess to what extent we are satisfied with our lives without considering how we feel most of the time. This would lead to absurd conclusions like that people can be happy under torture. Therefore, the affective and cognitive components of happiness seem to be closely interlinked. The affective component influences to what extent we are consciously satisfied with our lives: If we experience frequent positive affect, our evaluation of life will be more positive. The cognitive component influences whether we experience positive or negative affect: If our live goals are being met and thus our needs are gratified, we experience positive affect and vice versa. This is supported by a study conducted by Diener et al. (1991), who showed that frequent positive affect and infrequent negative affect are both necessary and sufficient for high scores on commonly used scales for measuring happiness. Therefore, the affective and cognitive components of happiness seem to be closely interlinked.

2.2.2 Why Study Happiness?

The idea of using individuals' well-being as a measure of the quality of a society goes back as far as to ancient ethics. Annas (1993) points out that ancient ethics are primarily concerned with the question of how satisfied an individual is with his or her life and "with the way it has developed and promises to" (p. 28). According to

Helliwell (2003), Aristotle had a particular influence on both ancient and modern discussions of happiness, since he attempted to balance different aspects of happiness that others have regarded as antithetical.

Aristotle argued that the constituent parts of happiness are “good birth, plenty of friends, good friends, wealth, good children, plenty of children, a happy old age, and also such bodily excellences as health, beauty, strength, large stature, athletic powers, together with fame, honour, good luck and excellence” (Aristotle, Rhetoric, p. 1360b). While some parts of this definition might seem antiquated at the present day, current research shows that many of Aristotle’s constituents of happiness indeed are closely related to people’s well-being. For instance, research shows that happiness is linked to health (Berkman & Syme, 1979), the breadth and depth of interpersonal relationships (Putnam, 2000), and socioeconomic status (Lyubomirsky et al., 2005).

While Aristotle acknowledged that a sufficient supply of external goods is necessary in order to live a happy life, he at the same time emphasized that material goods are only a means to an end: “We must not think that the man who is to be happy will need many things or great things, merely because he cannot be blessed without external goods; for self-sufficiency and action do not depend on excess, and we can do noble acts without ruling earth and sea; for even with moderate advantages one can act excellently” (Aristotle, Nicomachean Ethics, Book 10, Chapter 8) **.** Again this is in line with current research showing that money only promotes happiness up to a certain amount, as long as the greater amount of money can be used to cover basic needs (Diener & Seligman, 2004; Helliwell, 2001). At high income levels, the relation between even higher income and happiness seems to disappear.

Modern economics has only recently begun to take this argument by Aristotle and its supporting empirical research seriously. Towards the end of the 20th century, a new ‘economics of happiness’ emerged which challenges the assumption that income and other purely financial indicators can be used as a sole measure of a society’s welfare (Graham, 2008). Proponents of this new field of economics take the view that money is only a means to an end, “and that end is well-being” (Diener & Seligman, 2004, p. 2). Diener and Seligman take the argument one step further by showing that economic indicators often not only omit, but also mislead about what people really value. In a review of a considerable number of studies examining the issue, they show that materialism can negatively affect people’s well-being and that greater wealth is only in some cases associated with higher levels of happiness. In line with this new field of research, Bhutan famously introduced the indicator of 'gross national happiness' to replace the commonly used 'gross national product' (Graham, 2008).

However, happiness is not only desirable as an end in itself, but also because it has positive external effects. Research suggests that individual and societal well-being has various other beneficial outcomes. For example, Inglehart and Klingemann (2000) showed that a decline in a society's well-being can lead to major institutional changes, while high levels of happiness support democratic institutions. Although the direction of causality can also go the other way - democratic institutions can have positive effects on a society's well-being (Frey & Stutzer, 2000) - Inglehart and Klingemann found that high levels of happiness can also *cause* changes that strengthen democratic institutions. In another piece of research, reviewing a large number of longterm and experimental studies, Lyubomirsky et al. (2005) show that happiness preceeds success in the workplace, satisfying relationships and superior health, and that positive affect promotes sociability, altruism, conflict resolution skills and original thinking.

Therefore, research into happiness is not only of academic interest but also of societal relevance. As argued by proponents of the economics of happiness, research into well-being can supplement traditional economic indicators and thereby help evaluate public policies. Furthermore, research into individual happiness is not only relevant because happiness itself is desirable, but also because happiness has desirable external effects on democracy, the economy and society in general. While this paper does not directly address how to promote happiness, it aims to contribute to an understanding of how certain aspects of social networks - namely their homogeneity and heterogeneity - are related to the happiness of their members. It might thereby contribute to an understanding of whether policies should rather promote bonding social capital or bridging social capital, which is frequently the subject of scholarly discussions (e.g. Putnam 2000, pp. 360).

2.2.3 Measuring Happiness

a. Conventional Methods for Measuring Happiness

The most common way to measure happiness is by asking people how happy they feel (Kahneman & Krueger, 2006). Such measures of happiness can either be based on multiple questions in order to reduce measurement errors, or on a single question such as, "All things considered, how satisfied are you with your life as a whole these days?" (World Values Survey, as cited in Kahnemann & Krueger, 2006, p. 6). Possible answers commonly range from 0 (extremely unhappy) to 10 (extremely happy) (Layard, 2010).

There has been much dispute about the validity and reliability of these measures of subjective well-being (Veenhoven 2009; Layard 2010). One common concern is that people might not know well enough how satisfied they are with their lives on a scale from (say) 0 to 10. For instance, it has been shown that current moods can have a large influence on scores on happiness surveys and that people's answers in surveys assessing their own happiness can vary significantly within only short periods of time (Schwarz & Strack, 1991). Another concern is that people might indicate to be happier than they are in reality, for reasons of social desirability or in order to defend their ego (Veenhoven, 1984). Lastly, personality or cultural differences might lead to different answers in surveys (Helliwell, 2006). For example, optimists might answer surveys assessing their happiness more positively than pessimists. However, despite all concerns, there is considerable evidence showing a relation between self-assessed happiness and health outcomes, neurological functioning, and future behavior, among other things, which suggests that the measures are valid (Kahneman & Krueger, 2006; Layard, 2010; Veenhoven, 1984).

Other methods for measuring happiness include monitoring brain activity (Layard, 2010), asking friends, and observing verbal or non-verbal expressions (Veenhoven, 1984). While still in its infancy, monitoring brain activity has been shown to provide valid and reliable results (Layard, 2010). However, due to its obtrusive nature it can only be applied in certain studies. The reliability and validity of the latter two unobtrusive methods are disputed (Veenhoven, 2009). Veenhoven (1984) points out that happy people do not necessarily behave happily and that raters of someone else's happiness rather think about how happy they would be in his or her situation, which rather leads to an evaluation of socio-economic circumstances than to an evaluation of the subjective well-being of an individual.

b. Measuring Happiness on Twitter

The internet and the success of blogs and social networking sites in particular has led to the development of new methods that can also be applied to measure the subjective well-being of both groups and individuals. It has been argued that Twitter is particularly suitable to extract emotions from messages published by its users because of the limited length of tweets and the fact that many tweets reflect current experiences of their author (Dodds, Harris, Kloumann, Bliss, & Danforth, 2011).

A line of research that is useful for examining the emotions of Twitter users is the 'sentiment analysis' of tweets. Sentiment analysis, or ‘opinion mining’, refers to the “computational study of opinions, sentiments and emotions expressed in text” (Liu, 2010, p. 629) and has inspired a large amount of research activity in the past decade (Pang & Lee, 2008). Among other things, sentiment analysis has been applied to Twitter in order to infer the public opinion about political issues (O'Connor, Balasubramanyan, Routledge, & Smith, 2010) and to analyze the effects of expressed consumer opinions on Twitter (Jansen, Zhang, Sobel, & Chowdury, 2009).

Moreover, there are a number of studies that used a form of sentiment analysis to analyze Twitter users' subjective well-being. As mentioned before, (Bollen, Gonzalves, Ruan, & Mao, 2011) developed a measure of Twitter users' happiness based on the sentiment analysis of tweets in order to examine the diffusion of subjective well-being on Twitter. In a widely cited study, Dodds and Danforth (2009) developed a 'hedonometer' aiming to measure the hedonic content of written expressions based on a lexicon containing happiness evaluations of a large number of words. In a later paper, Dodds et al. (2011) improved their hedonometer, focusing on the measurement of the happiness of tweets and developing a larger lexicon. Similarly, Kramer (2010) developed a happiness score that is based on a sentiment analysis of tweets and showed that the score is significantly related to scores on Diener et al. (1985's) satisfaction with life scale. Later on, Quercia et al. (2012) showed that Kramer's measure of happiness also holds at the community level: They found that the metric correlates with census-data about the socio-economic well-being of U.S. communities.

The basic problem sentiment analysis tries to solve is to categorize messages as either conveying positive or negative emotions (Bifet & Frank, 2010). According to the definition of happiness established above, ‘frequent positive affect’ and ‘infrequent negative affect’ form parts of people’s happiness. In other words, if someone is frequently in a good mood and only rarely in a bad mood, I consider this to both contribute to and reflect his or her level of happiness. As discussed above, this constitutes the ‘affective component’ of happiness. Assuming that the sentiments of tweets are indicative of the author’s current mood and these sentiments are observed over a prolonged period of time, their analysis allows to draw conclusions about the author's affective component of happiness.

What about the cognitive component of happiness? Kramer (2010) has shown that people's tweets are more positive if they are more satisfied with their lives, according to Diener et al.'s (1985) widely-used satisfaction-with-life-scale. While the correlation of the positivity of tweets with life-satisfaction is only modest ($r = .17$), it is still indicative of life-satisfaction. As discussed above, satisfaction with life constitutes the cognitive component of happiness. Therefore, empirical research suggests that the positivity of tweets is indicative of the cognitive component of happiness.

While no empirical research has been found that examines the relation of happiness measures based on sentiment analysis with the affective component of happiness, the argument has been made that the polarity of a tweet is also likely to reflect its author's mood (Bollen et al., 2010). Research shows that people generally either use Twitter to blog about themselves or to share information, while most tweets fall into the first category (Naaman, Boase, & Lai, 2010). Bollen et al. point out that both types of content allow us to draw conclusions about mood states. In the case of tweets about the authors themselves, they can explicitly mention their current mood (e.g. “I am feeling happy”). In the case of tweets containing more general information, users can indirectly express their mood (e.g. “Very happy to see the #Obama administration & OSHA backing down on the regulation of small, family farms”).

A problem that needs consideration, however, is that people might not express their honest feelings on Twitter because they might strive to convey a socially desirable image of themselves (Zhao, Grasmuck, & Martin, 2008). Research into impression management has revealed two self-presentation strategies that are relevant in connection with social networking sites (Kim & Lee, 2011): positive and honest self-presentation. On the one hand, users of social networking sites tend to highlight the positive aspects of their lives in order to convey a positive image of themselves (Zhao et al., 2008). On the other hand, research on Facebook (Young & Quan-Haase, 2009) and on online dating sites (Gibbs, Ellison, & Heino, 2006) has shown that information shared in these settings tends to be truthful, because users fear that dishonesty might come to light.

While Twitter is different from social networking sites in the stricter sense such as Facebook, since Twitter profiles and tweets are public by default and users can follow anyone without confirmation of the followed user, it is reasonable to assume that these findings also apply to some extent to Twitter. On the one hand, people can be expected to try to convey a positive image of themselves on Twitter to a similar extent as on Facebook, maybe even to a greater extent, since tweets are public and users have little control over who can see them (unless they make their account private). On the other hand, untruthful statements are even more likely to be revealed for the same reason - everyone with access to the internet can potentially see them.

It should be noted, however, that honest and positive self-presentation strategies are not necessarily mutually exclusive (Zhao et al., 2008). People might ‘stretch the truth’ or simply leave out negative aspects of their lives. Therefore, I expect the moods expressed in tweets to be generally more positive than the authors’ actual moods.

However, while tweets might generally be more positive than Twitter users' actual moods, the *relative* positivity and negativity of tweets can still be expected to be informative. Someone experiencing more negative emotions can be expected to report less positive ones and vice versa (Kramer, 2010). As I will describe in more detail below, I operationalized the well-being of Twitter users as the fractional difference of positive and negative tweets and therefore obtained a relative measure of the positivity of tweets.

c. Expressed Well-Being

As discussed above, the emotions people express in their tweets are related to their happiness but the positivity of tweets cannot be equated with their subjective well-being for several reasons: First, the correlation between the positivity of tweets and their authors' life-satisfaction has been found to be only modest in size (Kramer, 2010). Second, while there is reason to assume that tweets reflect the moods of their authors, the effect size of the correlation between the positivity of tweets and the affective component of happiness is mostly unknown. Therefore, it is not entirely clear to what extent the positivity of tweets is related to the overall happiness of Twitter users. Third, evidence from research into impression management suggests that Twitter users tend to highlight the positive aspects of their lives since people in general strive to convey a positive image of themselves.

To account for these complexities, I will in the following distinguish between the *expressed well-being* and the subjective well-being of Twitter users. With ‘expressed well-being’ I refer to the well-being of Twitter users inferred from the amount of positive and negative tweets they publish and with ‘subjective well-being’ I refer to their actual happiness as defined earlier.

Since Twitter users can be expected to highlight the positive aspects of their lives, as established above, I assume expressed well-being to be higher than subjective well-being. Due to the facts that people who are more satisfied with their lives have been shown to publish more positive tweets (Kramer, 2010) and that there is reason to assume that tweets also reflect their authors' moods (Bollen et al., 2010), expressed well-being is very likely to be indicative of subjective well-being.

2.3 Social Capital

2.3.1 Overview and Definition

As mentioned before, the overarching idea behind social capital theory is that social ties can be considered as a resource (Putnam, 2000). A headhunter you know might help you find a job. Your family and good friends might give you moral support when you need it. The fellow members of your bowling club might help you move apartments. The better your relationships with these people you know are - the stronger the social ties between you - the more likely it is that they will be willing to help.

However, when it comes to what social capital is about more exactly, definitions differ almost from author to author (Daniel, Schwier, & McCalla, 2003). One of the earliest definitions of social capital was provided by Hanifan (1920), who defined social capital as “those tangible assets [that] count for most in the daily lives of people: namely goodwill, fellowship, sympathy and social intercourse among the individuals and families who make up a social unit” (Hanifan, 1920, p. 78). This broad definition highlights the value of social ties and anticipated most elements of later understandings of social capital (Putnam, 2000). However, Hanifan’s notion of social capital did not attract much attention until the late 1980s, when it was continued by Pierre Bourdieu (1986) and James Coleman (1988). Coleman, together with Robert Putnam (2000), who wrote the influential book *Bowling Alone*, are credited with popularizing the notion of social capital (Gauntlett, 2011).

According to Daniel et al. (2003), current definitions of social capital fall into two main categories: Some definitions focus on the *content dimension* of social capital, such as Putnam’s (2000) and Hanifan’s (1920) definitions, while others focus on the *structural dimension* of social capital, such as Bourdieu’s (1986) and

Coleman's (1988) definitions. While I do not completely agree with this strict dichotomy, since the two dimensions clearly overlap as will become clearer below, it makes sense to differentiate between them in order to illustrate the different elements that make up the notion of social capital.

a. Content Dimension

According to Daniel et al. (2003), the *content dimension* of social capital refers to the nature of interactions between members of the same social network. Social networks are central to many understandings of social capital and will in the following be understood as consisting "of a finite set or sets of actors and the relation or relations defined on them" (Wasserman & Faust, 1994, p. 20). Some of the variables of the content dimension that form an essential part of many discussions of social capital are trust, reciprocity and shared understanding (Feldmann & Assaf, 1999). Putnam (2000), for example, emphasizes the importance of trust for the formation of social capital by arguing that social networks almost by definition include mutual obligations that are based on trust and reciprocity. He argues that mere contacts are not interesting; they only become interesting once people within the social network feel some sort of obligation to help each other. And people mainly feel obliged to help each other if they trust others to reciprocate the favor. Without this trust in favors being returned, according to Putnam, people are far less willing to help one another.

Shared understanding as a variable of social capital is closely connected to the notion of trust. Daniel et al. (2003) argue that the relationship of trust and shared understanding is reciprocal: On the one hand, trusting relationships between members of a social network can generate common norms and thereby lead to shared understanding. On the other hand, shared understanding between members of a social network can have positive effects on the level of trust between them.

b. Structural Dimension

The *structural dimension* of social capital, on the other hand, refers to the fundamental elements of a social network, like the types of connections between people and the organization of the community[1] (Daniel et al. 2003). Regarding the

1. Communities are understood as groups of interdependent people who "participate together in discussion and decision making and who share certain practices that both define the community and are restored by it" (Bellah, 2008)

structural dimension, discussions of social capital often consider the strength of ties between individuals and the diversity of a social network.

Articles about social capital usually differentiate between *strong* and *weak ties*. According to Granovetter (1973), tie strength is determined by the emotional intensity, the mutual trust, the reciprocity of the tie and by the amount of time spent maintaining the tie. This illustrates that the structural dimension is closely interlinked with the content dimension of social capital. Strong ties are strong, because people connected by them have a high motivation to share their resources with each other (Haythornthwaite, 2005). However, Granovetter (1973) argued in his much cited 'strength of weak ties theory' that ties that are commonly referred to as weak also have their strengths. People we know only a little commonly move in different social circles and therefore have access to different information and views. Thus, the strength of weak ties is that they can provide us with information, attitudes and resources that we might otherwise not be able to obtain.

Another part of the structural dimension of social capital that is commonly included in discussions of social capital is the differentiation between *bonding* and *bridging social capital.* As mentioned earlier, bonding and bridging social capital mainly refer to differences in the diversity of a social network: While bonding social capital is used to describe homogeneous networks that promote exclusive identities, bridging social capital refers to networks that include people from diverse social and cultural backgrounds (Putnam 2000). However, according to some definitions (e.g., Putnam et al., 1993), bonding and bridging social capital also refer to different tie strengths between the members of the social network. Bonding social capital is usually characterized by strong ties; bridging social capital usually comprises weak ties. An example of groups promoting bonding social capital are church groups - they are homogeneous in the sense that all members are of the same religion and that the ties between individuals are rather strong. An example of bridging social capital are environmental movements - they are heterogeneous since they include people from all over the world with diverse social backgrounds and they (mostly) promote rather weak ties among their members. Table 1 illustrates the elements of the two overarching dimensions of social capital.

Table 1: Dimensions of Social Capital

Content Dimension	Structural Dimension
Trust Shared understanding Norms	*Tie strength*: Weak ties, strong ties, latent ties *Group diversity*: Bonding social capital and bridging social capital

The dimensions outlined above are neither mutually exclusive nor comprehensive. They only illustrate one possible way of categorizing the rather elusive and imprecise notion of social capital. Moreover, it should be noted again that the content dimension and the structural dimension are highly interrelated. The content and nature of social interactions, like for example the levels of trust prevailing in a social network, also directly affect the structural dimension of the network: Levels of trust are closely related to the strength of social ties. Tie strengths, in turn, affect whether social capital is to be classified as bonding or bridging social capital.

c. Definition

The variety of different dimensions of social capital illustrates the difficulty in finding a definition that copes with them all. This has led to a significant number of definitions that focus on different aspects of social capital. For example, some authors equate social capital with social networks (e.g. Franke, 2005; Burt, 2009), others with resources (e.g. Smith, Giraud-Carrier, & Stephens, 2011; Bourdieu, 1986), and still others with civic engagement (e.g. Putnam 2000).

I will adopt a network-based definition that was given by the Canadian Public Research Initiative: *“Social capital refers to the social networks that may provide access to resources and social support.”* (Franke, 2005, p. 9)

This network-centered approach is based on the hypothesis that the structure of social interactions determines which resources people can access (Franke, 2005). It allows to draw conclusions about individuals’ social capital by examining the relationships between them, which seems like a sensible approach to analyzing online social networks like Twitter.

d. Social Capital and the Internet

How is social capital affected by the internet? Just like the definition of social capital itself, the answer to this question is highly disputed among scholars. There is reason to assume that the internet impedes the formation of social capital in some ways and enhances it in others. Most recent literature about the issue, however, considers the internet to have positive effects on social capital (Durst, Viol, & Wickramasinghe, 2013)

In the following, I will assume that ‘offline’ and ‘online’ social capital form a reciprocal relationship. Studies suggest that there generally exists a high overlap between offline and online social networks. For example, research on Facebook and MySpace among teens (Lenhart & Madden, 2007) and college students (Subrahmanyam, Reich, Waechter, & Espinoza, 2008) has shown that the social networking sites are mainly used to articulate existing offline relationships and they are only rarely used to connect with previously unknown people. Similarly, research shows that only 7 per cent of Facebook friends are people that users have never met in person (Hampton, Goulet, Rainie, & Purcell, 2011).

It is not entirely clear to what extent these findings also apply to Twitter and to other age groups. Some use Twitter mainly as an information source (Java, Song, Finin, & Tseng, 2007), in which cases Twitter networks are unlikely to reflect people’s offline social networks. However, since a large amount of Twitter users also use the platform for ‘daily chatter’, conversations or more generally to connect with friends (Java et al., 2007), I assume that Twitter networks tend to reflect people’s offline social networks to some extent or another. Therefore, the relationship between ‘offline’ and ‘online’ social capital is reciprocal, as would also be suggested by the theory of social shaping of technology (Williams & Edge, 1996): Our social ties from the offline world are reflected in our virtual social ties, and social ties that are formed on the internet are reflected in our social contacts in the real world.

While I do not look at how the internet affects people’s social capital in this study, I infer people’s social capital from their use of Twitter, as I will discuss in more detail below. Since I assume that online and offline social capital are related to each other, I argue that people’s overall social capital is to some extent reflected in their online social networks and that the Twitter network is therefore a legitimate source to analyze Twitter users’ social capital.

2.3.2 Social Capital and Happiness

How is social capital related to happiness? There are a considerable number of studies from different fields that suggest a positive relation between social capital and subjective well-being (for a review see Helliwell, 2001). However, due to various methodological issues (see Helliwell & Putnam, 2004), it is generally easier to find correlations between social capital and subjective well-being than actual causalities, as it is often the case in social sciences. Nevertheless, there are some studies suggesting that higher levels of social capital can in fact *cause* a rise in people's happiness (Helliwell, 2001; Putnam, 2000).

In the following I will distinguish between direct and indirect relations between social capital and subjective well-being. Direct relations between social capital and happiness are illustrated by studies showing that different measures of social capital are directly linked to subjective well-being. Indirect relations between social capital and happiness are suggested by studies about public health and education, among others. For example, studies show that people who are better connected with their social environment are more educated (Helliwell, 2001) and that education is positively related to happiness (Putnam, 2000). These different relationships are illustrated in Figure 1.

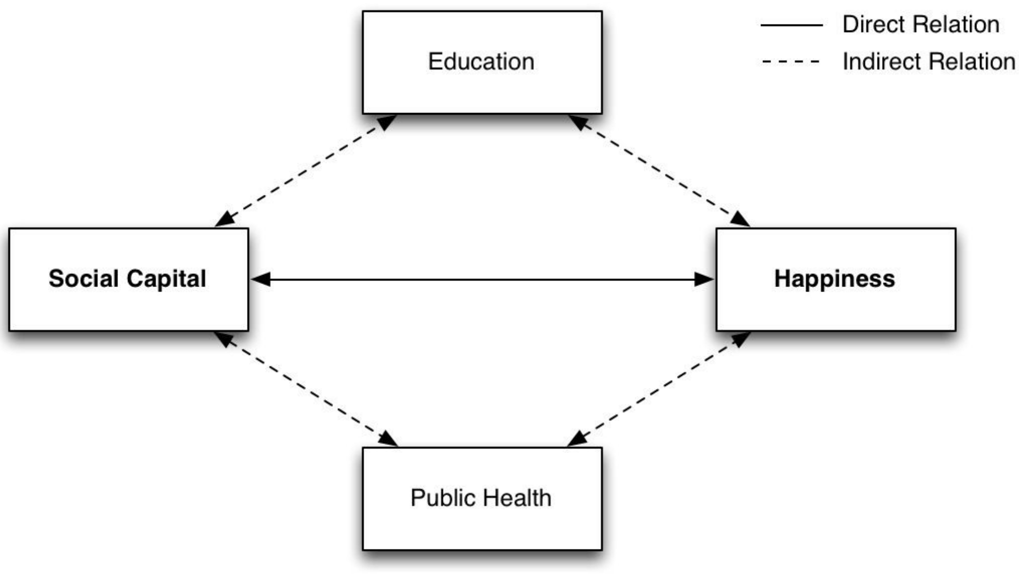

Figure 1: Relation between Social Capital and Happiness

a. Direct Relations

As pointed out above, social capital is mainly about the quality of interpersonal relationships. Therefore, studies showing that personal relationships are linked to happiness illustrate a direct relation between social capital and happiness.

In one of the earliest studies of the relatively new field of scholarly research into happiness, Wilson (1967) showed that marriage is highly related to subjective well-being. In a more recent study, Putnam (2000) confirmed this finding, but also showed that the breadth and depth of personal relationships in general are positively related to people's happiness. In line with this, Layard (2006) identified seven main factors affecting happiness, five of which are "concerned with the quality of our relationships" (p. 42) - namely community and friends, family relations, work, personal freedom and personal values. Moreover, Helliwell and Putnam (2004) have shown that the frequency of interactions with friends is positively related to happiness.

Furthermore, there are studies showing that various sorts of trust, which are part of the content dimension of social capital, are closely linked to life satisfaction and general happiness (Helliwell, 2003; Helliwell, 2006). Also, several studies (Frey & Stutzer, 2000; Helliwell, 2003; Putnam, 2000; Veenhoven, 1996) found that social participation, which constitutes the center of Putnam's (2000) influential definition of social capital, is positively linked to happiness.

In addition, there is much research relating higher levels of social capital directly to greater well-being. For example, Putnam (2001) found that individuals' self-assessed happiness increases both with their own and with their governments' measures of social capital. Correspondingly, Helliwell (2006) showed in an analysis of cross-national samples a large positive relation between various measures of social capital and happiness. In line with this, Bjornskov (2003) found that social capital is an important factor when explaining why some countries are happier than others. Connected to social networking sites, Burke et al. (2010) found that bonding social capital of adolescent Facebook users is positively related to their life satisfaction, both determined through a self-report survey.

From the opposing perspective, much research has shown that social connections prevent depression (Putnam, 2000). While the absence of depression is not to be equated with happiness, since one can be unhappy without feeling depressed, they are clearly related (Helliwell, 2001). Correspondingly, it has been shown that social capital is associated with lower national suicide rates (Helliwell, 2004).

b. Indirect Relations

There is more evidence from different fields of research suggesting an indirect relation between social capital and happiness. Two of these fields with particularly convincing findings concern public health and education.

Studies show that health is closely related to happiness (Helliwell, 2003) and much research shows that social capital is associated with better health (Helliwell, 2001). A particularly convincing piece of research suggesting a relation between social capital and health is a study by Berkman and Syme (1979). The study showed that individuals with more and stronger social ties had lower mortality rates over the nine-year period in which the study was carried out. In line with the findings of Wilson (1967) and Putnam (2000) concerning the effects of social ties on happiness, marriage was found to have the strongest negative effect on mortality rates, followed by close relatives and friends. It could be suspected that the causality's direction might be the other way around, i.e., that less healthy people are also less able to maintain social relationships. However, in the course of the initial survey, Berkman and Syme also collected information about participants' health behavior, like smoking habits, physical activity and obesity, among other things, and they found that the connection between social ties and mortality rates were independent of all these control variables. Therefore, the findings suggest that there exists a causal relationship between the number and quality of social ties and health.

Furthermore, it has been shown that the higher the number of social ties a person maintains, the less susceptible he or she is to the common cold (Cohen, Doyle, Skoner, Rabin, & Gwaltney, 1997). Additionally, Putnam (2000) reviews various other studies showing a link between higher levels of social capital and better public health.

Another line of research suggesting a connection between social capital and happiness concerns education. Evidence shows that social connectedness and trust, elements of social capital, are closely related to better education (Helliwell, 2001) and that education is positively associated with higher levels of subjective well-being. For example, research by Putnam (2000) shows that education benefits subjective well-being even after controlling for the positive effects of education on incomes, health habits and civic participation, which in turn can be expected to increase people's happiness. Interestingly, higher levels of income, which are also connected to social capital (Putnam 2000), only have limited effects on subjective well-being (Diener & Seligman, 2004).

c. The Dark Side of Social Capital

Several scholars have pointed out that there is also a ‘dark side of social capital’ (Putnam, 2000, p. 350) which might have negative effects on a society’s well-being (Putnam, 2000; Woolcock & Narayan, 2000). This dark side of social capital mainly resides in bonding social capital that creates strong, exclusive group identities. To give a stark example, nazis during World War II certainly had high levels of bonding social capital, but this most certainly resulted in lower levels of well-being for society as a whole, and arguably also for convinced national socialists themselves. Even in less radical cases, strong group loyalties can isolate members from information, foster a hostile climate towards socially desirable norms and promote undesirable values (Woolcock & Narayan, 2000), which might have negative impacts on individuals’ subjective well-being. To give an example, community values can raise people’s aspirations and thereby increase the gap between people’s goals and reality (Helliwell, 2001), which has been shown to reduce people’s happiness (Michalos, 1985).

However, these examples illustrate a point made by Putnam (2001), who argued that it is important to distinguish between happiness at the individual and at the community level. Bonding social capital might have harmful effects on society as a whole, while it has positive effects on the individuals who are part of the corresponding group. For example, while the mafia is harmful to people that are not part of it, it might promote the subjective well-being of the mafiosi themselves. Moreover, it should be noted that bonding social capital can also be desirable. Strong ties within a family can also significantly increase individuals’ subjective well-being without having any negative effects on people that are not part of the family. This is in accordance with the findings of Burke et al. (2010) who showed that higher levels of bonding social capital among adolescent Facebook users are associated with greater life-satisfaction.

d. Conclusion

There is much evidence suggesting a positive relation between social capital and the subjective well-being of individuals. The strongest counterarguments to this claim refer to bonding social capital that can have negative impacts on people who are not part of the bonding group, but only to limited extent on the individuals with strong bonding social capital themselves. Since this study measures social capital at the individual level, the literature gives little reason to assume that bonding social capital has negative effects on subjective well-being and instead strongly suggests that social capital in general is positively related to happiness.

Despite the large number of studies examining the effects of social capital on happiness, only one empirical study could be found that differentiated methodologically between bonding and bridging social capital (Burke et al., 2010) and this study was based on the self-assessment of U.S. college students. As I have mentioned above and will further describe below, I differentiated between bridging, bonding and overall social capital when assessing the relation of social capital and well-being and therefore addressed this gap in the scientific literature.

2.3.3 Measuring Social Capital

a. General Methodological Issues

Measuring social capital has been proven to be difficult. Since social capital is such an imprecise concept, almost everyone trying to measure it finds it necessary to provide a definition from their own perspective (Daniel et al., 2003). It has been argued that developing a single ‘true’ measure of social capital is likely to be impossible for various reasons (Woolcock & Narayan, 2000).

First, most definitions of social capital are multidimensional and the dimensions differ almost from author to author. Social capital has been measured, among other things, through perceptions of trust, participation rates in organizations, the nature of social connections, sources of social support, voting rates, access to information, personal efficacy, and political engagement (Szreter & Woolcock, 2004). Furthermore, authors disagree on whether some variables are either elements or indicators of social capital, and some simply assume that they are related to social capital in some way. For example, some equate civic engagement with social capital (e.g., Putnam 2000), others view it as an element of social capital (Shah, McLeod, & Yoon, 2001), and some consider as an outcome of social capital (e.g., Franke, 2005).

Second, some view social capital as a property of communities (e.g., Szreter & Woolcock, 2004), while others consider it as a property of individuals (e.g., Bourdieu, 1986), while still others consider it as both (e.g., Putnam, 2000). However, measuring concepts at the community level is already problematic because scholars have struggled unsuccessfully for decades to agree on a standard definition of community (Daniel et al., 2003). Therefore, and due to the apparent lack of reliable measures of social capital on the group level, the majority of studies measuring social capital target individuals (Szreter & Woolcock, 2004).

Third, most studies measuring social capital are based on secondary indicators like measures of trust or civic engagement and do not attempt to measure social capital as a primary outcome (Daniel et al., 2003).

And fourth, as argued by Williams (2006), the internet has made it necessary to develop new methods for measuring social capital, since it changed the nature of social interactions which made it impossible to measure social capital by simply using existing scales. This illustrates that the methods for measuring social capital must also be adapted to changing contexts.

I will follow the trend of current studies to measure social capital at the individual level. However, I will assume that social capital is not a property of individuals per se, but that it resides in their connections with each other, similar to what has been argued by Szreter and Woolcock (2004) and Putnam (2000). Social capital is based on social ties, and social ties by definition connect people to each other. According to the definition of social capital established above, social capital resides in these connections that together make up a social network. However, since the exact composition of these social networks differ from person to person, social capital also differs between individuals and it is therefore sensible to measure social capital at the individual level.

b. Computational Determination of Social Capital

How can social capital be measured in online social networks more concretely? Smith, Giraud-Carrier, Ventura et al. (2011) developed a framework for the computational analysis of social capital in online networks and therefore addressed the fourth issue pointed out above, namely the necessity to develop new methods for measuring social capital in times of the internet. The computer scientists' expressed goal was to design a framework that provides social scientists with a way to unobtrusively measure social capital in online communities. Among other things, it has been applied to determine social capital in the blogosphere (Smith, Purser, & Giraud-Carrier, 2008) and on Twitter (Smith & Giraud-Carrier, 2010).

The framework measures social capital based on the attributes of individuals, the relationships between them, and the resources that are available to them. However, according to the definition of social capital given above that considers resources as an outcome and not as an element of social capital, I will refrain from analyzing resources. While they might be a useful indicator of social capital, it seems impossible to find a way to measure even a part of all the different kinds of tangible and intangible resources that are made available to Twitter users through the Twitter network, which they might not even be aware of themselves. Smith, Giraud-Carrier, Ventura et al. (2011) also acknowledge that social resources are generally not available to social capital studies.

Smith and Giraud-Carrier (2010, pp. 386-387) distinguish between two types of connections: *explicit connections* and *implicit affinities*.

(a) *Explicit connections* directly link individuals to one another, either based on a well-defined relationship (e.g., “is friend of”) or on some purposive action (e.g., sending a direct message). Individuals are aware of those connections. Social networks based on explicit connections are called *explicit social networks (ESNs).*

(b) *Implicit affinities* connect individuals based on loosely defined affinities or similarities, such as shared interests. In contrast to explicit connections, individuals might not be aware of their similarities nor of each other. Social networks based on implicit affinities are called *implicit affinity networks (IANs).*

Networks that consist of both explicit connections and implicit affinities are so called *hybrid networks.* Figure 2 illustrates such a hybrid network between five individuals.

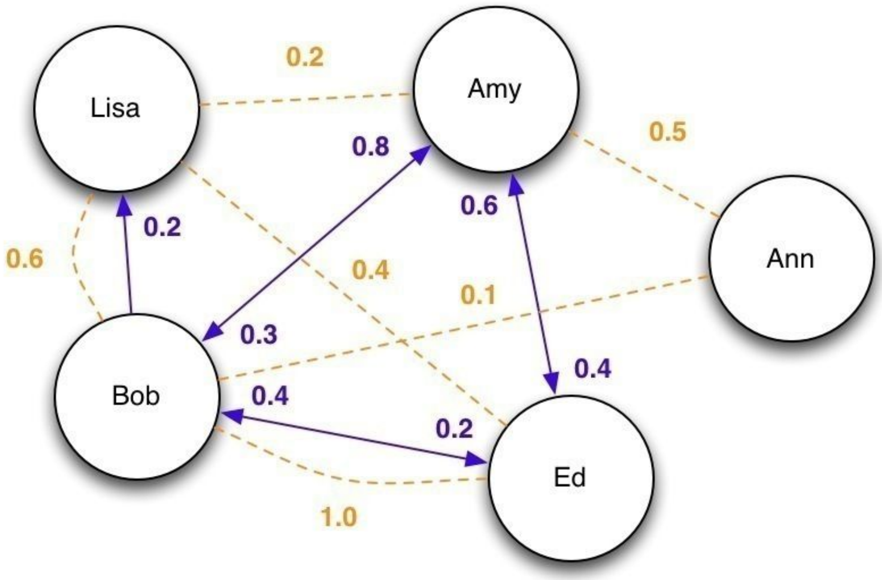

Figure 2: Hybrid Network (adapted from Smith, Giraud-Carrier and Purser, 2009)

The solid lines show the explicit social network, the dotted lines depict the implicit affinity network. The numbers indicate the strength of the social connections, ranging from 0 to 1. Implicit affinities are always undirected, since they are based on the degree of similarity of two individuals. Explicit connections, however, are directed. For example, Bob might follow Lisa on Twitter without Lisa following Bob. Moreover, the strength of an explicit connection between two individuals might vary for each of them. For example, Bob might consider Amy to be his best friend, while Amy thinks of Bob merely as an acquaintance.

How do these different kinds of connections between individuals relate to their social capital? According to Smith, Giraud-Carrier, Ventura et al. (2011), the combination of implicit affinities and explicit connections determines an individual's bridging and bonding social capital. Strong implicit affinities combined with explicit connections lead to bonding social capital - the individual has much in common with others of his social network. Weak implicit affinities combined with explicit connections result in bridging social capital - the individual has little in common with members of his social network. The strength of the explicit connection affects the amount of social capital gained from the connection. They can be seen as a measure of the tie strength which is part of many discussions of social capital.

c. Measuring Explicit Connections and Implicit Affinities in the Case of Twitter

As I will describe in more detail below, I determined the strength of Twitter users' explicit connections based on the number of times they interacted with each other. I assume that the more two users interact with each other, the stronger the social tie between them.

This assumption is based on the argument advanced by Smith, Giraud-Carrier and Stephens (2011) that interactions between individuals form an important part in the evolution of their relationship's strength. However, I do not only assume that the amount of interactions between two Twitter users affect their relationships, but also that it reflects the strength of their relationship. The stronger the tie between two people, the more time they spend interacting with each other (Boase, Horrigan, Wellman, & Rainie, 2006). This is in line with Granovetter's (1973) argument that the strength of social ties is, among other things, based on the amount of time spent maintaining the tie.

However, does the amount of interaction online equally reflect the tie strength for good friends as it does for mere acquaintances or are, for example, good friends more likely to interact in the real world than they are likely to interact online? Research suggests that there are only small differences. A study examining the role of the internet in maintaining people's social networks found that the assumption that people communicate more with strong ties than with weak ties is true both for in-person interactions and for communication via e-mail (Boase, Horrigan, Wellman, & Rainie, 2006). The differences found between strong ties and weak ties when it comes to the amount of in-person versus the amount of online communication were only marginal. While no research could be found examining whether the number of interactions between two Twitter users also reflects the strength of their tie, I assume that these findings also apply to Twitter and that the number of interactions between Twitter users therefore reflects the strength of their relationship.

As mentioned above, the strength of the implicit affinity between two individuals is based on their degree of similarity. The more similar two individuals are, the stronger their implicit affinity. Smith, Giraud-Carrier and Stephens (2011) propose to measure the similarity of two individuals based on the value of attributes people expose to the network. If many of these values overlap, then the users are assumed to be similar to each other.

As I will describe in more detail later, the attribute I chose in order to compare

Twitter users' similarity are the *public users*[2] they follow. I define 'public users' as the top 0.05 per cent of Twitter users with the highest number of followers. These public users include accounts of organizations (e.g., NGOs, political parties, news organizations), public figures (e.g., politicians, actors, musicians, journalists), and private users that I consider opinion leaders due to their high number of followers.

It has been shown that Twitter users tend to follow others with similar interests (Himelboim, McCreery, & Smith, 2013). Therefore, I assume that following public users is an expression of Twitter users' interests. If two private users follow similar public users, I consider them to be similar due to similar interests and they are therefore connected by an implicit affinity. The more similar the set of public users that are followed by these two private users are, the stronger the implicit affinity between them.

To my knowledge the similarity of Twitter users has not been measured in this way before and it therefore constitutes a contribution of this study. In their application of the framework to Twitter, Smith and Giraud-Carrier (2010) measured the similarity of Twitter users by comparing the sets of unigrams and bigrams (phrases consisting of one or two words, respectively) in Twitter users' profile descriptions. While this method might have sufficed for their experiment, it is arguably not very useful when comparing random Twitter users. Most Twitter users do not provide any profile description and if they do, it is questionable whether profile descriptions that tend to be rather short provide valid measures of their author's interests. From the 39.308 random user profiles that I collected in the course of this study (see below), only about 24 per cent provided a profile description, many of which only consisted of one or two words, hashtags or emoticons.

Others have determined the similarity of Twitter users by counting the number of edges between them (Lerman & Ghosh, 2010) or by comparing the topical similarity of the tweets they publish (Weng, Lim, Jiang, & He, 2010). The former method is not applicable to this study, since all individuals whose similarity is compared are exactly one edge apart from each other. Moreover, it does not provide a measure of the homogeneity of social networks. The latter method is much more complex than comparing the public users individuals follow, since it involves machine

2. 'Public users' as defined here are not to be confused with Twitter accounts that are public in the sense that they are unprotected and everyone can access their timelines and user information. When talking about the latter ones, I refer to 'public accounts' instead.

learning techniques to extract topics from tweets. Merely comparing hashtags as indicators of topics, however, would not be very useful since the usage of hashtags is relatively low (Weng et al., 2010) and different hashtags are commonly part of discussions about the same topic (Bruns & Burgess, 2011). Moreover, people might tweet about the same topic but have opposing views, in which case a comparison of the public users followed is likely to be more valid. For example, two Twitter users might be equally interested in presidential elections in the US and take part in discussions about the topic, but support and therefore follow different candidates.

This rather complex design is necessary in order to include as many dimensions of social capital as feasible and therefore to increase the content validity (Cozby, 2011) of the measure. Franke (2005) suggests a range of indicators to measure social capital in networks. Not to complicate the research design even further and due to the fact that some of these indicators cannot be measured by analyzing data obtained from Twitter, the measure employed in this study only includes three of them, namely *network size*, *network diversity* and *relational frequency*. Table **2** describes these indicators and their underlying hypotheses, according to Franke's (2005, pp. 13) discussion, and indicates how I measured them in the course of this study.

Table 2: Social Capital Indicators

Indicator	Description	Hypothesis	Measured through
Network Size	Number of persons with whom a relationship is maintained	The larger the network, the greater the likelihood that a particular resource is available and therefore the more social capital the individual possesses	Number of explicit connections
Network Diversity	Homogeneity of members	Strong social homogeneity creates bonding relations, weak social homogeneity creates bridging relations	Strength of implicit affinities
Relational Intensity	The strength of a relationship	The stronger the ties among the members of a network, the greater the change that they will be disposed to exchanging resources and therefore the higher their amount social capital	Strength of explicit connections

This also illustrates why merely looking at the number of friends and followers of Twitter users in order to measure their social capital would not be sufficient. By looking at the number of friends alone, the network size would be the only indicator considered, which by itself would be a poor measure of social capital. Social capital encompasses more than just the size of a social network. Furthermore, I argue that the number of friends and followers of Twitter users is not even a very good measure of their network size when basing considerations about social capital on it. Social capital as defined above refers to social networks that provide access to resources and social support. However, following or being followed by someone is not necessarily indicative of the other's readiness to provide resources or support. It is possible to follow almost every other Twitter user, but the act of clicking the follow-button does not lead to a gain in social capital. Therefore, I operationalized the strength of explicit connections based on the amount of interactions between Twitter users, which seems like a more useful measure of network size and tie strengths, since it only considers connections between people that interact in some way and therefore are more likely to exchange resources.

3. Research Design

In order to answer the research questions, I conducted a computer-assisted quantitative content analysis of tweets and a social network analysis of Twitter users' online social networks. The sample consists of 214 Twitter users. To determine the social capital and expressed well-being of these users, however, I needed to collect data about their friends, followers and their follower's friends. In total, the collected dataset consists of 43,670,346 user profiles and 79,579,346 tweets[3]. The data were collected between April 15 and May 15, 2014. In order to collect the data, I connected to the Twitter API using the programming language Python. I also used Python to preprocess the data and to compute the expressed well-being and social capital of Twitter users according to the formulas described below.[4]

3.1 Sampling Method

In order to determine the social capital and expressed well-being of the 214 users in the sample, I collected their tweets and their followers' tweets of the previous three months. The analyzed time frame ranges from January 16 to April 16, 2014. The time frame was restricted to tweets from the previous three months to decrease the likelihood that the network had changed considerably in the analyzed period of time.

The sample was chosen as follows: Every Twitter user has a unique user ID between 1 and (currently) 3 billion. I have randomly generated numbers that include the range of all possible IDs and thereby have collected 39,308 random user profiles. To obtain the final sample of 214, I filtered this set of random users according to the following criteria:

(a) I only included Twitter users that had published at least 30 tweets in the three months of interest and that had existed for at least six months. A certain

3. The tweets only include the ones published by the 214 users in the sample and their followers, since the tweets of the users' friends and their followers' friends were not needed for the data analysis.

4. See Appendix for more technical information about the data collection procedure and for a selection of the Python scripts I used.

amount of tweets is necessary to be able to perform a sentiment analysis, which the determination of expressed well-being is based on (see below). I restricted the sample to accounts older than six months in order to decrease the likelihood that the users' networks had changed considerably in the analyzed period of time: Shortly after creating an account, a user's friends and followers are likely to change faster than after a few months, since connections to friends, acquaintances and public users first need to be established. About 90 per cent of Twitter users in the randomly collected set of user profiles have existed for more than six months and 7 per cent have published more than 30 tweets in the previous three months.

(b) Since I employed a sentiment analysis of tweets that is based on an English lexicon, I only included Twitter users whose tweets were primarily in English. About 17 per cent of the random sample published English tweets more than half of the time.

(c) I only included Twitter users that had at least 30 followers. Since I determined Twitter users' social capital based on the connections with their followers, looking at users that mainly use Twitter as an information source (i.e., by following others) but are hardly followed by anyone would most likely not have led to useful results. Of the randomly collected user profiles, 13.6 per cent met this criterium.

(d) Additionally to the lower threshold of 30 followers, I only included Twitter users that are not among the top 0.05 per cent of users with the highest number of followers. According to the random user profiles collected, this upper threshold lies at 12,092 followers. As discussed above, I consider these accounts to belong to 'public users' that include organizations and public figures, following which can be seen as an expression of their followers' interests.

Of the 39,308 random profiles, 245 met these criteria. After collecting the data, I dropped another 4 users that followed less than 3 public users and another 27 who had published less than 10 tweets in the three analyzed months that could be classified as either positive or negative. This left me with a final sample of 214 Twitter users.

In which ways are these Twitter users different from the average user? The set of 39,308 random user profiles shows that most Twitter users seem to be either 'lurkers', publishing hardly any tweets and being followed by hardly any other users, or perhaps only created accounts to try out Twitter. About 40 per cent of Twitter users do not have any followers, and 78 per cent of Twitter users have less than 10 followers. Evidently, these accounts are neither likely to reflect nor to influence the social capital of their creators, so considering them in this research would not be useful. Similarly, 44 per cent of Twitter users have never published a tweet, and 74 per cent have published less than 10 tweets. Basing an analysis of Twitter users' expressed well-being and social capital on such a limited amount of tweets is very unlikely to lead to valid results.

Therefore, according to the criteria established above, the sample consists of Twitter users that have considerably more followers and publish many more tweets than the average Twitter user. Research shows that the younger Twitter users are, the more frequently they use Twitter and the more tweets they publish (Smith & Brenner, 2012). Thus, the sample is likely to be skewed towards younger Twitter users. Moreover, since I only considered users publishing English tweets, most users in the sample are likely to be from countries where English is the official language. Most users from other countries will probably not primarily tweet in English, even if they speak English as a second language.

It should be emphasized, however, that the random set of Twitter users I have just compared the sample of this study to includes all existing accounts, even if they are not used anymore or have never been used. Therefore, it includes accounts people have only created to try out Twitter, duplicate accounts people have forgotten about and accounts that have been created to mine data or for other specific purposes. I do not have access to information about how often or for which purposes these accounts have been used, nor could I find any recent and detailed studies about active Twitter users. Thus, the differences between the average user in the sample and the average *active* Twitter user are not as large as illustrated above. However, since some of the thresholds I defined to filter out users that are suitable for this study are rather high and the final sample only included a small percentage of the set of random Twitter users, I assume that these differences do apply to some limited extent.

As mentioned earlier, Twitter users have the option to protect their timelines, in which cases it is not possible to access their tweets. These users were ignored for the determination of the sample. Moreover, followers of users from the sample with protected accounts were ignored when determining the social capital of the users in the sample. As mentioned above and as described in more detail below, the tweets of a user's followers are needed to determine his or her social capital according to the framework for the determination of social capital I use. Since only 12 per cent of the followers of users in the sample had protected accounts, it is unlikely that this distorted the results significantly, though.

It should also be noted that the Twitter API limits the tweets that can be received from a user to the most recent 3,200 ones. However, this affected the results only marginally, since only about 4 per cent of Twitter users whose tweets were needed had published more than 3,200 tweets in the three months that were subject of this study. In the cases where a user did publish more than 3,200 tweets in the three months of interest, the most recent 3,200 tweets were used to determine expressed well-being and the strength of explicit connections (see below).

3.2 Expressed Well-Being (EWB)

3.2.1 Operationalization

I determined the emotional content of tweets by using the open source software OpinionFinder, whose subjectivity analysis feature is based on research conducted by Wilson et al. (2005). The tool can be used to analyze the subjectivity of texts, indicating whether words have a positive, negative or neutral meaning. It is based on a subjectivity lexicon containing roughly 8,000 English words and employs a method to determine the ‘contextual polarity’ of phrases. In other words, it not only analyzes the word itself but also its context in order to determine a word's polarity, which leads to better results.

OpinionFinder has been used by a large number of studies to determine the subjectivity of texts, among others by Bollen et al. (2011) to determine the subjective well-being of Twitter users and by O’Connor et al. (2010) who used it to show that an analysis of tweets with a lexicon-based sentiment detector was closely related to official poll data.

I adopted the measure for subjective well-being that was suggested by Bollen et al. (2011). However, in contrast to Bollen et al. and as established above, I assume that subjective well-being is not entirely the same as the well-being stated in tweets, and I therefore refer to it as ‘expressed well-being (EWB). Bollen et al. define the well-being of Twitter user *u* as the fractional difference between the number of the user’s tweets that contain positive terms and those that contain negative terms as indicated by OpinionFinder:

$$EWB(u) = \frac{N_p(u) - N_n(u)}{N_p(u) + N_n(u)}$$

Tweets that contain both positive and negative terms are not considered.

According to this operationalization, EWB always assumes a value between −1 and +1. If a user publishes more negative tweets than positive ones, his or her expressed well-being is negative, whereas if the positive tweets prevail, EWB is positive.

3.2.2 Validity

EWB is meant as a measure of the relative positivity of tweets and from this perspective it is valid. In this study it is used as an indicator of Twitter users' happiness. I have discussed in detail above how expressed well-being is likely to be related to subjective well-being. To my knowledge, there is no empirical study testing to what extent this particular measure of the positivity of tweets is correlated with the two components of subjective well-being (i.e, the affective component and the cognitive component). However, as discussed earlier, a similar measure of subjective well-being based on a sentiment analysis of tweets has been found to be significantly correlated with life-satisfaction (Kramer, 2010), which constitutes the cognitive component of happiness, and there is reason to assume that the positivity of tweets also is indicative of the affective component of happiness.

3.3 Social Capital

The measures of social capital are based on the framework for the computational analysis of social capital in online networks that was developed by Smith, Giraud-Carrier, Ventura et al. (2011). As explained above, the framework determines bonding and bridging social capital based on implicit affinities and explicit connections.

3.3.1 Implicit Affinity Networks (IANs)

As discussed above, I determined the implicit affinity between two Twitter users (i.e., their degree of similarity) by comparing which *public users* they follow. I defined public users as the top 0.05 per cent of users with the highest number of followers. If two private user follow similar public users, I consider them to have similar interests and they are therefore connected by an implicit affinity. The more similar the set of public users that are followed by these two private users are, the stronger the implicit affinity between them.

This is in line with the framework developed by Smith, Giraud-Carrier, Ventura et al. (2011), which posits that people are connected by an implicit affinity when they share an attribute whose value sets overlap (pp. 36). They suggest to determine the implicit affinity between two people by the Jaccard Index of these two value sets. I followed this suggestion and determined the implicit affinity between the Twitter users *i* and *j* by determining the Jaccard Index of the sets of public Twitter users A_i and A_j they follow:

$$s_{ij}^{IAN} = \frac{|A_i \cap A_j|}{|A_i \cup A_j|}$$

An example shall illustrate this. Lets assume that Twitter user *i* is Bob and Twitter user *j* is Amy. Bob follows the set of public users A_{Bob} = Unicef, BarackObama, LadyGaga. Amy follows the set of public users A_{Amy} = Unicef, LadyGaga, MittRomney, NewYorkTimes. Then, the absolute value of the intersection of A_{Bob} and A_{Amy} is 2 (Unicef and LadyGaga). The absolute value of the union of A_{Bob} and A_{Amy} is 5 (Unicef, BarackObama, LadyGaga, MittRomney, and NewYorkTimes). Therefore, the strength of the implicit affinity between Bob and Amy is

$$s_{Bob,Amy}^{IAN} = \frac{|A_{Bob} \cap A_{Amy}|}{|A_{Bob} \cup A_{Amy}|} = \frac{2}{5} = 0.4$$

3.3.2 Explicit Social Networks (ESNs)

According to Smith, Giraud-Carrier and Stephens (2011), every interaction affects the strength of the explicit links between the two interacting individuals. However, as discussed earlier, I assume that online ties both affect *and* reflect the offline social ties of an individual. Therefore, I do not merely assume that interactions on Twitter affect the relationship between individuals, but also that the amount of interactions between two individuals *reflects* the strength of the tie between them. Thus, I will measure the strength of *explicit connections* by analyzing the interactions between Twitter users.

As mentioned above, there are three main ways Twitter users can interact with each other - mentions, retweets and favorites. Not to unnecessarily complicate the framework, I only considered mentions and retweets and assumed that they equally reflect the strength of the explicit social connection between the two interacting individuals. Furthermore, for the same reason, I only considered interactions between two Twitter users if they are explicitly connected by a following-relationship.

I then determined the strength of the explicit connection by dividing the total amount of interactions of user *i* with user *j* by the total amount of mentions and retweets that occur in the analyzed tweets of *i*.

$$s_{ij}^{ESN} = \frac{nInteractions_{ij}}{\sum_{x \in N, x \neq i} nInteractions_{ix}}$$

For example, lets assume the tweets Bob has published in the past three months contain 25 mentions and 25 retweets. Within these 50 interactions with other users, Bob mentioned Amy 5 times and retweeted 5 of her tweets. Then, the strength of Bob’s explicit connection with Amy is

$$s_{Bob,Amy}^{ESN} = \frac{10}{50} = 0.2$$

As discussed earlier, explicit connections are directed. In other words, Twitter user *i* might place a different value on the relationship with *j* than the other way around.

Therefore, $s_{ij}^{ESN} \neq s_{ji}^{ESN}$. Applied to the example of Bob and Amy, this means that the strength of the explicit connection of Bob with Amy is not necessarily equal to the strength of the explicit connection of Amy with Bob.

3.3.3 Bonding Social Capital

As Smith and Giraud-Carrier (2010) argue, the amount of social capital that an individual *i* can realize is not based on the value that *i* places on the relationship with *j,* but on the value that *j* places on the relationship with *i.*

For example, if Bob thinks Amy is his best friend, but Amy considers Bob as merely an acquaintance, then Amy will only accommodate Bob based on her perception of their relationship. The fact that Bob values their relationship more highly does not increase the amount of support or resources he receives from Amy.

Therefore, Smith and Giraud-Carrier (2010) define the bonding social capital of user *i* as the product of the strength of the implicit affinity between users *i* and *j* by the strength of the explicit connection of user *j* with user *i.* The total bonding social capital of user *i* is the sum of his or her bonding social capital with all other individuals:

$$bonding(i) = \sum_{j \in N, j \neq i} s_{ij}^{IAN} s_{ji}^{ESN}$$

As pointed out above, s_{ij}^{IAN} and s_{ij}^{ESN} can take on values between 0 and 1 and therefore the bonding social capital between two individual users can also take on values between 0 and 1, according to the formula above. The more similar two users are to each other (the closer s_{ij}^{IAN} is to 1) and the stronger the explicit connection between them (the closer s_{ij}^{ESN} is to 1), the larger the amount of bonding social capital gained from the connection. Since the total amount of bonding social capital of a user constitutes the sum of the bonding social capital gained from every single connection, the total amount of bonding social capital of a user is not restricted and depends on the number of connections the user maintains to others. An example below will illustrate how bonding social capital is determined in practice.

3.3.4 Bridging Social Capital

According to Smith and Giraud-Carrier (2010), bridging and bonding social capital form a reciprocal relationship. They define the total bridging social capital of an individual *i* as:

$$bridging(i) = \sum_{j \in N, j \neq i} (1 - s_{ij}^{IAN}) s_{ji}^{ESN}$$

I modified this formula since it is only sensible in certain scenarios. The formula is based on the assumption that s_{ij}^{IAN} takes on values between 0 and 1, and, as pointed out earlier, in theory it does. I defined s_{ij}^{IAN} as the Jaccard Index of the sets of public users that are followed by the two private users *i* and *j*. The Jaccard Index by definition assumes a value between 0 and 1. However, in the case of large sets that can assume many different values it can be very unlikely that it ever gets close to 1.

This is the case in this study, due to the way s_{ij}^{IAN} was operationalized. There is a very large number of public users[5] and the users in the sample followed 143 public users on average. Consequently, s_{ij}^{IAN} assumed an average value of 0.015 with a standard deviation of 0.03. With that said, it makes little sense to define bridging social capital as defined by Smith and Giraud-Carrier. If s_{ij}^{IAN} tends to take on such small values, bridging social capital as defined by Smith and Giraud-Carrier will be almost equal to the strength of the explicit connection and it would therefore hardly be indicative of the heterogeneity of a social network. Moreover, as I will further explain below, the overall social capital of Twitter users is defined as the sum of their bonding and bridging social capital. According to the formulas established above, if s_{ij}^{IAN} is small, then bonding social capital is also small and bridging social capital is large. Therefore, adopting Smith's and Giraud-Carrier's (2010) formula in this study would have the effect that overall social capital is almost equal to bridging social capital and only marginally affected by bonding social capital.

5. According to estimates, there are roughly 900 Million Twitter accounts (Twopcharts, 2014). Since public users as defined above constitute the top 0.05 per cent, there are approximately 450,000 public users.

This is unlikely to be merely a problem of my operationalization of s_{ij}^{IAN}. According to Smith, Giraud-Carrier, Ventura et al. (2011), implicit affinities are supposed to be indicative of the similarity between two people. They suggest to measure the degree of similarity based on the value of attributes that users expose, more specifically by determining the Jaccard Index of these sets of attributes. However, people can evidently be described based on a very large number of attributes. If they are not to be compared on only a very basic level, the Jaccard Index is likely to be small, since its value decreases with the number of attributes that are not shared by both sets. For example, if I had chosen to use Twitter users' gender as a measure of their similarity, the attribute could only assume two values (male or female) and the Jaccard Index of two sets would either be 1 (same gender) or 0 (different gender). However, this would not be useful since the gender of Twitter users does not characterize them very well. When choosing an attribute that can assume a large number of values, such as the public Twitter users people follow, the Jaccard Index is generally much smaller.

Therefore, I adapted Smith and Giraud-Carrier's formula in order to ensure that the strength of implicit affinities has a recognizable effect on the level of bridging social capital and that the amount of overall social capital is affected by bonding and bridging social capital to a similar extent, which can be seen as a contribution to Smith, Giraud-Carrier, Ventura et al.'s (2011) framework. I define bridging social capital as

$$bridging(i) = \sum_{j \in N, j \neq i} (L - s_{ij}^{IAN}) s_{ji}^{ESN}$$

where

$$L = 3 * std(s^{IAN}) + \overline{s^{IAN}}$$

The limit L therefore constitutes the value that is 3 standard deviations above the mean of s^{IAN}. L is the same for the entire sample and does not differ between users. Thus, for the determination of L the implicit affinities between all users in the sample need to be considered. All values that lie above this limit are considered as outliers. If s_{ij}^{IAN} is higher than L, it is reduced to the value of L. Not to complicate the formula any further, this is not reflected in it above.

This ensures that L corresponds with the values that s_{ij}^{IAN} takes on in reality and not only with the values s_{ij}^{IAN} can take on in theory. Thereby the values of bridging and bonding social capital are closer to each other and affect the overall social capital to a similar extent. In the case of this study, L is 0.11 and therefore considerably lower than in the formula proposed by Smith and Giraud-Carrier. Roughly 99.8 per cent of the values s_{ij}^{IAN} assumed lay below this limit. Therefore, if I had assumed that L is 1 as proposed by Smith and Giraud-Carrier, in 99.8 per cent of the cases bridging social capital had been more than eight times[6] larger than bonding social capital. Determining the limit in the way described above ensures that the values of bonding and bridging social capital are more similar to each other.

According to the formula given above, if the implicit affinity between *i* and *j* is weak (s_{ij}^{IAN} close to 0) and if the explicit connection between *i* and *j* is strong (s_{ij}^{ESN} close to 1), then the amount of bridging social capital gained from the connection is large. Again, an example below will illustrate how bridging social capital is determined in practice.

3.3.5 Overall Social Capital

Finally, the total social capital of user *i* equals the sum of his or her bridging and bonding social capital, which is equal to the sum of the explicit connections of all of user *i*'s followers with him or her:

$$SC(i) = bridging(i) + bonding(i) = \sum_{j \in N, j \neq i} L * s_{ji}^{ESN}$$

Therefore, the total social capital of Twitter users is independent of the homogeneity of their social network and only depends on how frequently their followers interact with them.

6. If s_{ij}^{IAN} is 0.11 and L is 1, then bridging social capital is about 8 times larger than bonding social capital, since bridging = 0.89 * s_{ji}^{ESN} and bonding = 0.11 * s_{ji}^{ESN}.

3.3.6 Determination of Bonding, Bridging and Overall Social Capital: An example

The example of Bob and Amy shall illustrate how to determine the three measures of social capital. Lets assume Bob has two followers - Amy and Ed. Amy has three followers - Bob, Lisa and Ann. The strength of the implicit affinities and explicit connections between them are illustrated in the tables below.

Table 1: Bob's Twitter Network

	Implicit Affinity	**Explicit Connection**
Amy	$s^{IAN}_{Bob,Amy} = 0.4$	$s^{ESN}_{Amy,Bob} = 0.8$
Ed	$s^{IAN}_{Bob,Ed} = 1.0$	$s^{ESN}_{Ed,Bob} = 0.5$

Table 2: Amy's Twitter Network

	Implicit Affinity	**Explicit Connection**
Bob	$s^{IAN}_{Amy,Bob} = 0.4$	$s^{ESN}_{Bob,Amy} = 0.2$
Lisa	$s^{IAN}_{Amy,Lisa} = 0.9$	$s^{ESN}_{Lisa,Amy} = 0.3$
Ann	$s^{IAN}_{Amy,Ann} = 0.1$	$s^{ESN}_{Ann,Amy} = 0.6$

As established above, the implicit affinity between Amy and Bob is the same for both of them - implicit affinities constitute a measure of similarity and are therefore undirected. However, explicit connections are directed and the strength of the explicit connection of Amy with Bob is not equal to the strength of the explicit connection of Bob with Amy: $s^{ESN}_{Bob,Amy} \neq s^{ESN}_{Amy,Bob}$.

We need the strength of the explicit connection of a user's followers with the respective user to determine the user's social capital. Therefore, the tables above depict the strengths of the explicit connections of Bob's and Amy's followers with Bob and Amy instead of the other way around.

Not to overly complicate things in this example, I assume that L is 1. In a real case, L would be determined based on the standard deviation and the mean of all implicit affinities between users in the sample and their followers, according to the formula established above. However, in this example with only four different values of s^{IAN} this would not be very sensible since it is hardly possible to detect outliers based on the standard deviation of such a small sample.[7]

Bob's bridging and bonding social capital are then determined as follows:

$$\begin{aligned}
bonding(Bob) &= \sum_{j \in N, j \neq i} s^{IAN}_{Bob,j} s^{ESN}_{j,Bob} \\
&= s^{IAN}_{Bob,Amy} s^{ESN}_{Amy,Bob} + s^{IAN}_{Bob,Ed} s^{ESN}_{Ed,Bob} \\
&= 0,4 * 0,8 + 1,0 * 0,5 \\
&= 0,82
\end{aligned}$$

$$\begin{aligned}
bridging(Bob) &= \sum_{j \in N, j \neq i} (1 - s^{IAN}_{Bob,j}) s^{ESN}_{j,Bob} \\
&= (1 - s^{IAN}_{Bob,Amy}) s^{ESN}_{Amy,Bob} + (1 - s^{IAN}_{Bob,Ed}) s^{ESN}_{Ed,Bob} \\
&= (1 - 0,4) * 0,8 + (1 - 1,0) * 0,5 \\
&= 0,48
\end{aligned}$$

Bob's total social capital is the sum of his bonding and bridging social capital, which is equal to the sum of all explicit connections of his followers with him times L:

$$\begin{aligned}
SC(Bob) &= bonding(Bob) + bridging(Bob) \\
&= 0,66 + 0,64 \\
&= 1,3
\end{aligned}$$

or,

---

7. When applying framework to a real case, the examined social network are likely to be much larger. If they are this small, they probably do not reflect the social capital of the examined user and the framework should not be applied at all. For this reason I only included Twitter users in this study that are followed by at least 30 other users. Thus, in real cases that are deemed suitable to infer people's social capital from their online social networks *L* should always be determined according to the formula described above instead of assuming that *L* equals 1.

$$
\begin{aligned}
SC(Bob) &= L * (s^{ESN}_{Amy,Bob} + s^{ESN}_{Ed,Bob}) \\
&= 1 * (0{,}8 + 0{,}5) \\
&= 1{,}3
\end{aligned}
$$

Determined in the same way, Amy’s social capital is as follows:

$$bonding(Amy) = 0{,}41$$

$$bridging(Amy) = 0{,}69$$

$$SC(Amy) = 0{,}41 + 0{,}69 = 1{,}1$$

Therefore, the social capital of Bob, inferred from his use of Twitter, is higher than the one of Amy, although Amy has more followers than Bob. Bob’s social capital is higher because the explicit connections of his followers with him are stronger than in the case of Amy, i.e. Bob’s followers interact with him more frequently (in relative terms) than Amy’s followers interact with her.

3.3.7 Validity

Smith, Giraud-Carrier, Ventura et al. (2011) have partially validated their framework through case studies testing whether it conforms with theoretical expectations. They claim that their framework is in accordance with some principles of social theory, namely homophily, reciprocity, and bonding and bridging (pp. 91). However, they acknowledge that tests of additional well-developed theories of the social sciences are needed in order to fully validate their framework.

The method for determining the implicit affinities between Twitter users described above is not part of the framework developed by Smith, Giraud-Carrier, Ventura et al. (2011) and the way of determining the explicit connections between Twitter users was developed according to suggestions of the framework but has not been applied to Twitter before. Therefore, I do not have any empirical evidence for their validity. As I have argued above, however, the face validity of the method for the determination of the strengths of implicit affinities appears higher than in Smith and Giraud-Carrier's (2010) application of the framework to Twitter. A comparison of the sets of public users two individuals follow is likely to be a better measure of their similarity than a comparison of the words occurring in their profile descriptions.

While Smith, Giraud-Carrier and Stephens (2011) suggest to determine the strength of explicit connections based on the interaction frequency of two individuals, they have not validated this approach. In their applications of the framework they only checked whether a connection between two individuals was present or not. Since this study only considers users that are connected by an explicit link anyway, this approach would not make much sense here.

As pointed out above, the strength of explicit connections corresponds with the tie strength in social capital theory. According to Granovetter (1973), the strength of social ties is determined by the emotional intensity, the mutual trust, the reciprocity of the tie and the amount of time spent maintaining the tie, while the different determinants are closely interlinked with each other. The strength of explicit connections was operationalized based on the number of interactions between Twitter users and is therefore indicative of the amount of time spent maintaining the tie, which is in line with Granovetter's definition of the strength of social ties. Thus, the interaction frequency between two Twitter users appears to be a valid measure of the strength of the tie between them.

3.3.8 Practical Implications and Limitations

Since this research design is rather abstract, I will in the following illustrate some of its practical implications and limitations:

1. The determined social capital is not directly related to the number of followers of Twitter users, but to the number of times their followers interact with them

Both bridging and bonding social capital, as determined in the way illustrated above, are weighted by the strength of the explicit connection of a Twitter user's followers with him or her. The strength of these explicit connections, in turn, are determined by the amount of interactions with the respective user. Say Amy never mentions Bob in her tweets nor retweets any of his tweets, then Bob does not gain any social capital from the connection with Amy, since $s^{ESN}_{Amy,Bob} = 0$.

This can be seen as both a strength and a weakness of this design. On the one hand, merely being followed by someone is not necessarily expressive of the value of the social tie. People might follow others merely out of politeness, out of superficial interest in what someone is saying or they might even have been paid to do so. In these cases, being followed by someone does not lead to any gain in social capital. On the other hand, someone might also follow good friends on Twitter, but prefer to interact with them in other ways. In these cases, the framework does not account for strong social ties that would be an important element of a Twitter user's social capital. However, this limitation can not be accounted for in a design that infers people's social capital from their use of Twitter.

Adding a variable that directly relates the number of followers of Twitter users with their social capital would neglect the fact that the strength of social ties have significant implications for social capital. However, social capital determined in the way described above is *indirectly* related to the number of followers a user has, since the more people follow a Twitter user, the more likely he or she generally is to be mentioned or retweeted (Cha, Haddadi, Benevenuto, & Gummadi, 2010).

2. The differentiation between bonding and bridging social capital, which is based on the degree of similarity of two individuals, neglects that the strength of social ties are part of many conceptualizations of social capital (Durst et al., 2013)

According to the methods of determining bonding and bridging social capital described above, the two are only differentiated based on how similar or how different users are from each other. If two users are very different from each other and they are connected by strong explicit connections, then they gain a lot of bridging social capital from this connection. According to many definitions of social capital (e.g., the one given by Putnam, 1993), however, bridging social capital is also characterized by weak ties. When determined in the way suggested by Smith, Giraud-Carrier, Ventura et al. (2011), weak explicit connections result in only little bridging social capital. The framework has been criticized for this (Durst et al., 2013).

However, I argue that this is merely a matter of how bonding and bridging social capital are defined. Just like for the definition of social capital itself, there are no commonly agreed on definitions of bonding and bridging social capital and some neglect the strength of ties (e.g., Daniel et al., 2003; Franke, 2005; Smith & Giraud-Carrier, 2010). According to the way operationalized here, they are independent of the tie strength and are mainly a measure of network homogeneity. Distinguishing between the strength of explicit connections for the determination of bonding and bridging social capital would have made the framework even more complex without providing a clear benefit.

3. The measures of bridging and bonding social capital do not account for whether or not a Twitter user 'bridges' between different networks.

Bridging and bonding social capital, as determined in the way described above, are composite measures of network homogeneity and the number and strength of the social ties of a Twitter user. However, high levels of bridging social capital are not indicative of whether Twitter users in fact 'bridge' between different networks that are in themselves homogeneous (e.g, if they are both interested in republican and democratic politicians), but rather of whether they are 'outsiders' in their social network that have interests different from the ones of many others (e.g., if a user's followers are mainly interested in republicans, while he or she mainly follows democrats).

4. Results

4.1 Descriptives

The average user of the 214 users in the sample has 565 followers (*SD* = 1188) and follows 500 others (*SD* = 686), 147 (*SD* = 197) of which are public users that belong to the top 0.05 per cent with the largest number of followers. Therefore, due to the sampling method, the average number of followers and friends within the sample is substantially higher than within the entire population of Twitter users: The average Twitter user has 67 followers and 66 friends, according to the collected dataset of 39,308 random Twitter users. In the analyzed three months, the 214 users in the sample published 508 tweets on average (*SD* = 759).

As discussed above, the measure of expressed well-being can assume values between -1 and 1. The values of expressed well-being in the sample reach from -0.89 to 1, with a mean of 0.03 (*SD* = 0.26*).* The distribution of the values is fairly close to the normal distribution, as illustrated in Figure 1.

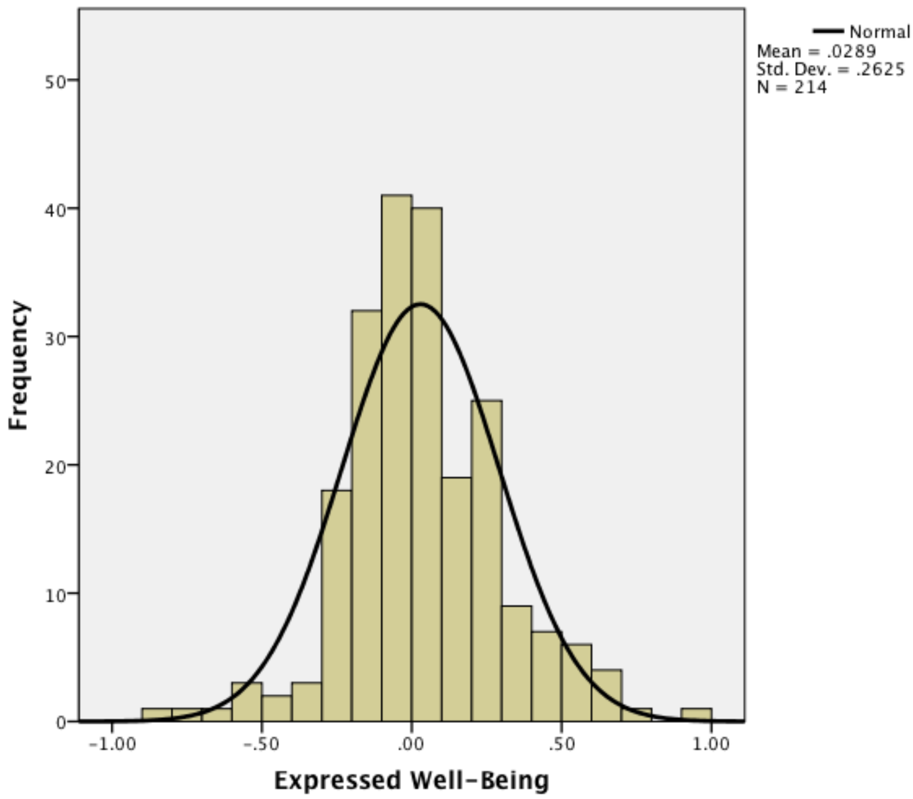

Figure 1: Histogram of Expressed Well-Being

The measures of social capital, however, can assume all values greater than zero. Most of their values assumed in this study are close to zero and their frequency decreases quickly. Figure 2 illustrates the distribution of the values of overall social capital, which is similar to the distribution of bonding and bridging social capital.

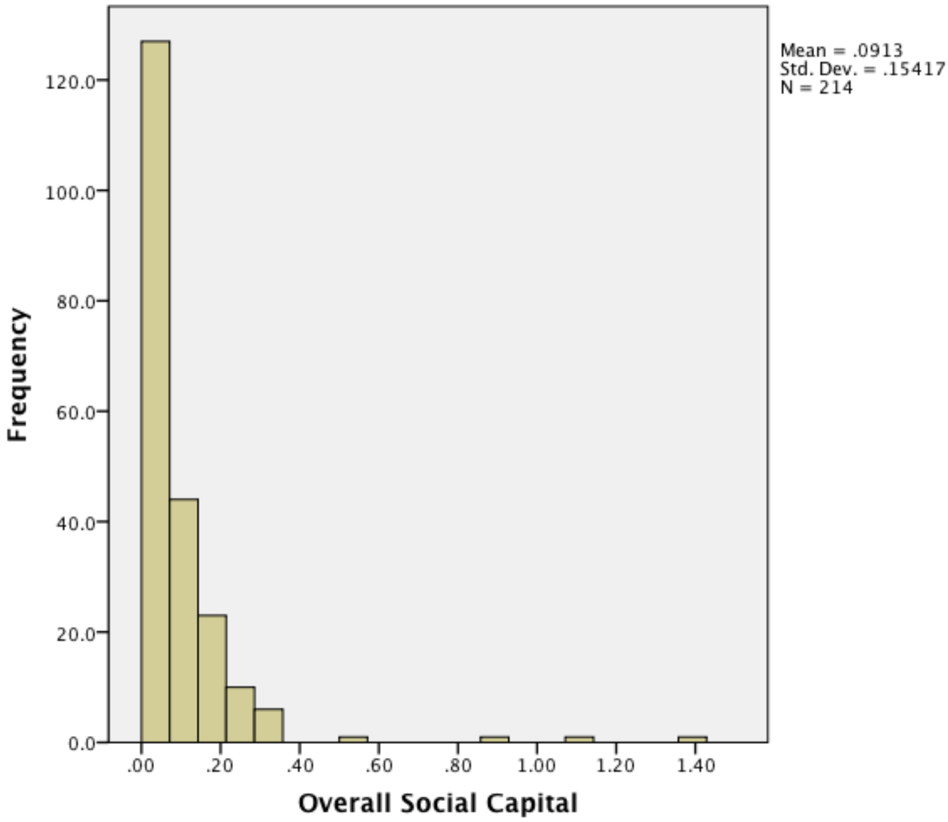

Figure 2: Histogram of Overall Social Capital

Table 1 shows the lowest and highest assumed values of the measures of social capital as well as their means and standard deviations.

Table 1: Social Capital Descriptives (N = 214)

	Minimum	Maximum	Mean	SD
Bridging social capital	.00	1.41	.0679	.1343
Bonding social capital	.00	.45	.0234	.0436
Overall social capital	.00	1.42	.0913	.1542

The bridging social capital of Twitter users is significantly higher than their bonding social capital, $t\ (213) = 4.85$, $p < 0.001$, 95% CI [0.026, 0.063]. According to the operationalization of bridging and bonding social capital, this means that the examined Twitter users tend to follow public users that are very different from the sets of public users followed by the examined Twitter users' followers.[8]

Table 2 shows a Pearson correlation matrix of expressed well-being, the different measures of social capital and some variables illustrating the users' activity on Twitter. Bonding, bridging and overall social capital are all significantly related to the number of followers of a user. As I have pointed out above, I expected this relation since the measures of social capital are based on the amount of times a user's followers interact with him or her and the higher the number of followers of a user, the more likely he or she is to be interacted with (Cha, Haddadi, Benevenuto, & Gummadi, 2010). Moreover, all three measures of social capital are significantly related to the number of times the examined users themselves interact with others by mentioning them or retweeting their tweets. This gives reason to assume that interacting with others also makes it more likely for Twitter users to be interacted with themselves. Interestingly, the number of times Twitter users mentioned or retweeted others is negatively related to their expressed well-being. In other words, the more Twitter users interact with others, the more negative their tweets are in general.

No measure of social capital is significantly related to the expressed well-being of Twitter users. However, in order to adequately address the research questions and to examine the relationships between the different measures of social capital and expressed well-being more closely, I will in the following present the results of regression analyses of the three measures of social capital and expressed well-being.

8. *L* equals roughly 0.11 in this study, determined in the way described above. Therefore, if a Twitter users' bridging social capital is greater than his or her bonding social capital, the strengths of implicit affinities and therefore the Jaccard Indices of the set of public users the examined Twitter user follows and the sets of public users his or her followers follow are lower than 0.055 on average. In other words, if the bridging social capital of Twitter users is greater than their bonding social capital, they follow on average less than 5.5 per cent of the same Twitter users as their followers.

Table 2: Pearson Correlation Matrix (N = 214)

	Expressed Well-Being	Bridging Social Capital	Bonding Social Capital	Overall Social Capital	Number of Followers	Number of times user mentioned or retweeted others
Expressed Well-Being	1	.030	.012	.029	-.029	-.143*
Bridging Social Capital	.030	1	.327**	.964**	.531**	.279**
Bonding Social Capital	.012	.327**	1	.567**	.234**	.367**
Overall Social Capital	.029	.964**	.567**	1	.529**	.346**
Number of Followers	-.029	.531**	.234**	.529**	1	.098
Number of times user mentioned or retweeted others	-.143*	.279**	.367**	.346**	.098	1

*. $p < 0.05$

**. $p < 0.01$

4.2 Regression Analysis of Bonding Social Capital and EWB

I conducted a simple linear regression analysis to determine whether expressed well-being as the dependent variable could be predicted from bonding social capital as the independent variable.

4.2.1 Testing the Assumptions of OLS Regression Analysis

Ordinary least squares regression analysis is based on several assumptions (Berry 1993). In the following I will briefly describe how I tested the ones that apply to simple regression analyses and which measures I took when they were not met:

(a) *Homoscedasticity*: A visual inspection of a graph illustrating the distribution of residuals at different levels of the predictor variable bonding social capital shows that the model violates the assumption of homoscedasticity: There are large differences among the variances of residuals at different levels of bonding social capital. While heteroscedasticity does not affect the OLS coefficient estimators, it leads to biased confidence intervals and distorts the results of t-tests and F-tests when determined with conventional methods (Berry, 1993). Therefore, I used heteroscedasticity-consistent methods implemented in the R package *sandwich* (Zeileis, 2004) to determine standard errors and to conduct the t-test and F-test.

(b) *Non-zero variance of predictors:* As described above, the predictor variable bonding social capital varies in value and thus this assumption is met.

(c) *Lack of autocorrelation:* OLS regression analysis assumes that residuals are uncorrelated. I conducted a Durbin-Watson test that led to the result of 1.85, which is considered acceptable (Durbin & Watson, 1951). The Durbin-Watson test has been found to be only slightly affected by heteroscedasticity (Harrison & McCabe, 1975).

(d) *Normally distributed errors*: A Q-Q plot of studentized residuals suggests that errors are fairly normally distributed. The plot indicates that there are few outliers in the lower and upper quantiles. However, none of the values' Cook's distance exceeds 0.63, therefore they do not exert great influence on the test results. I follow Stevens' (2009) suggestion not to remove any outliers with a Cook's distance of less than 1.

(e) *Linearity:* A scatterplot of bonding social capital and expressed well-being suggests the possibility that there might be a weak linear relationship but does not clearly indicate it. However, a non-linear relationship is not discernible either.

4.2.2 Results

Table 3: Regression model for predicting the expressed well-being of Twitter users from their bonding social capital (N=214)

	Unstandardized Coefficients		Standardized Coefficient	t	p
	B	SE	b*		
Constant	.027	.022		1.185	.237
Bonding Social Capital	.073	.497	.012	.147	.884
R2	0.000				
F	0.022				

Note: Standard errors, t-values and p-values were determined using heteroscedasticity-consistent methods (see Zeileis 2004)

The regression model of expressed well-being as the dependent variable and bonding social capital as the predictor variable is not significant, F (1, 212) = 0.02, p = 0.88. Therefore, the model is not useful for predicting the expressed well-being of Twitter users from their bonding social capital inferred from their Twitter networks. The difference between the slope of the regression model and 0 is not statistically significant either, b^* = 0.01, t = 0.18, p = 0.88, *95% CI* [-0.72, 1.22]. The coefficient of determination *R2* of 0.00 shows that there is no correlation between bonding social capital and expressed well-being of Twitter users in the sample. Figure 3 illustrates a scatterplot of bonding social capital and expressed well-being including the determined regression line.

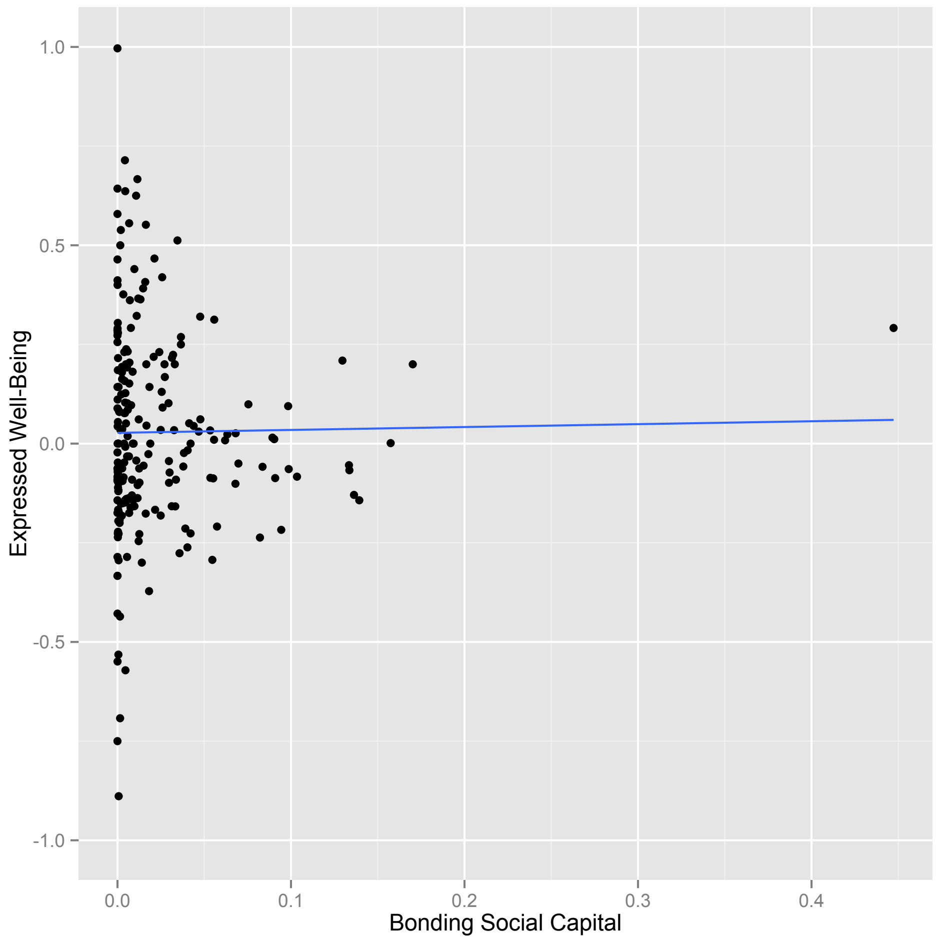

Figure 3: Scatterplot of Bonding Social Capital and Expressed Well-Being with Regression Line

The results of the regression analysis strongly suggest that bonding social capital is not a significant predictor of Twitter users' expressed well-being. If any, the scatterplot of the two variables indicates a linear relationship, however this relation has been found to be non-significant.

4.3 Regression Analysis of Bridging Social Capital and Expressed Well-Being

Similar to above, I conducted an OLS regression analysis to examine the second research question, asking whether Twitter users' expressed well-being can be predicted from their bridging social capital.

4.3.1 Testing the Assumptions of OLS Regression Analysis

(a) *Homoscedasticity*: As for the regression model of bridging social capital and expressed well-being, the residuals are distributed heteroscedastically and the model therefore violates the assumption of homoscedasticity. Thus, I again used heteroscedasticity-consistent methods (see Zeileis 2004) to determine standard errors and to conduct the t-test and F-test.

(b) *Non-zero variance of predictors:* The values of bridging social capital vary and this assumption is therefore met.

(c) *Lack of autocorrelation:* The Durbin-Watson statistic is 1.73, which indicates that the residuals are only slightly correlated with each other (Durbin & Watson 1951).

(d) *Normally distributed errors*: The inspection of the Q-Q plot of studentized residuals suggests that the assumption of normality is met. Again, there exist a few outliers in the lower and upper quantiles, whose maximal Cook's distance is 0.43, however. This indicates that they do not exert a major influence on the test results and they were therefore not removed from the analysis.

(e) *Linearity:* A scatterplot of bridging social capital and expressed well-being does not indicate any linear or non-linear relation between the two variables. This corresponds with the results of the regression analysis described below.

4.3.2 Results

Table 4: Regression model for predicting the expressed well-being of Twitter users from their bridging social capital (N=214)

	Unstandardized Coefficients		Standardized Coefficient	t	p
	B	SE	b*		
Constant	.025	.022		1.141	.255
Bridging Social Capital	.058	.159	.030	.362	.718
R2	0.001				
F	0.131				

Note: Standard errors, t-values and p-values were determined using heteroscedasticity-consistent methods (see Zeileis, 2004)

As for bonding social capital, the linear regression model of expressed well-being as the dependent variable and bridging social capital as the independent variable is not significant, *F* (1, 212) = 0.131, *p* = 0.26. Therefore, the model can neither be used to predict the expressed well-being of Twitter users from their bridging social capital. The regression line is not statistically significantly different from the mean either, *b** = 0.03, *t* = 0.36, *p* = 0.72, *95% CI* [-0.24, 0.32]. The coefficient of determination *R2* of 0.001 shows that there is hardly any correlation between bridging social capital and expressed well-being of the analyzed Twitter users. The regression line is illustrated in Figure *4.*

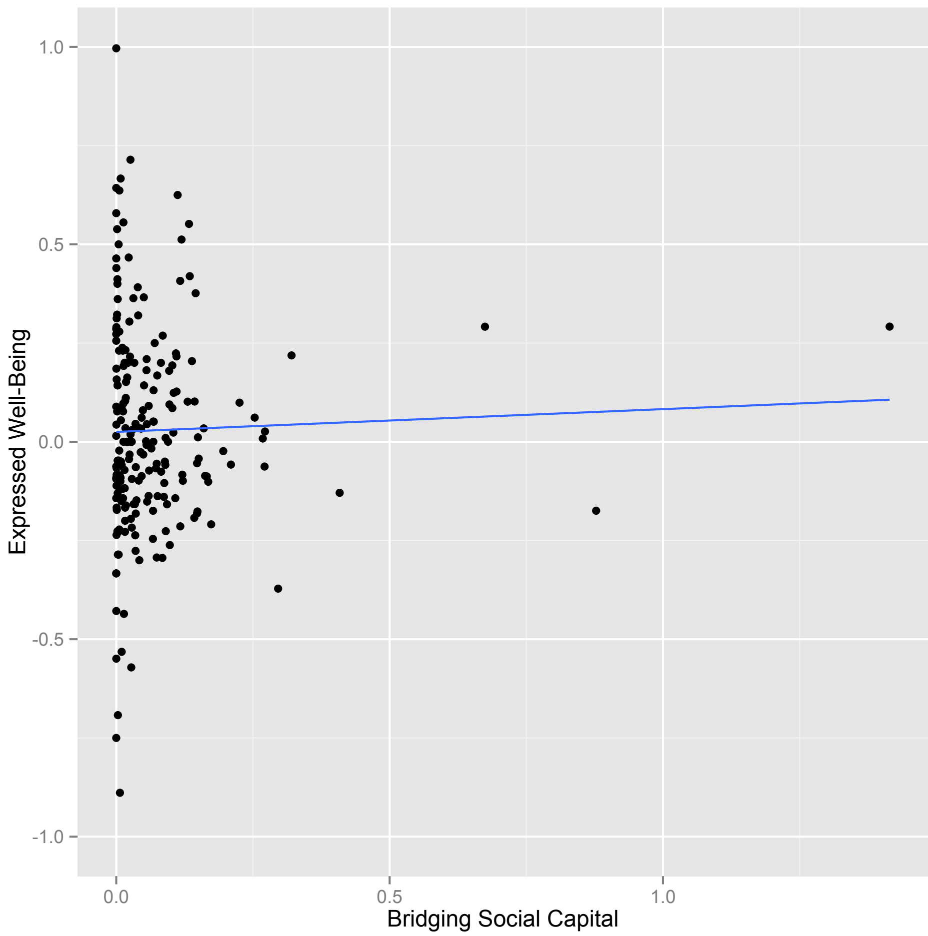

Figure 4: Scatterplot of Bridging Social Capital and Expressed Well-Being with Regression Line

The results of the regression analysis indicate that expressed well-being cannot be predicted from bridging social capital inferred from the use of Twitter. The scatterplot does not suggest any relationship between the two variables and accordingly the linear regression analysis has been found to be not significant.

4.4 Regression Analysis of Overall Social Capital and Expressed Well-Being

To address the third research question, I conducted a simple regression analysis with expressed well-being as the dependent and overall social capital as the independent variable.

4.4.1 Testing the Assumptions of OLS Regression Analysis

(a) *Homoscedasticity*: As for the two regression models above, the residuals are distributed heteroscedastically. To account for the violation of the assumption of homoscedasticity, I again used heteroscedasticity-consistent methods.

(b) *Non-zero variance of predictors:* The values of overall social capital vary and this assumption is therefore met.

(c) *Lack of autocorrelation:* The result of the Durbin-Watson test is 1.73, and the assumption of independent residuals is therefore considered to be met.

(d) *Normally distributed errors*: The Q-Q plot of studentized residuals indicates that the residuals are distributed normally. The maximal Cook's distance of outliers is 0.24, which indicates that they do not exert a large influence on the test results and they were therefore not excluded from the test.

(e) *Linearity:* The scatterplot of overall social capital and expressed well-being suggests the possibility of a weak linear relationship but does not clearly indicate it. No curvi-linear relation is apparent either.

4.4.2 Results

Table 5: Regression model for predicting the expressed well-being of Twitter users from their overall social capital (N=214)

	Unstandardized Coefficients		Standardized Coefficient	t	p
	B	SE	b*		
Constant	.024	.023		1.063	.289
Overall Social Capital	.050	.123	.002	.402	.688
R2	0.001				
F	0.162				

Note: Standard errors, t-values and p-values were determined using heteroscedasticity-consistent methods (see Zeileis 2004)

Just like the two models before, the regression model of expressed well-being as the dependent variable and overall social capital as the independent variable is not significant, *F* (1, 212) = 0.16, *p* = 0.69. Therefore, the model is not useful for predicting expressed well-being from the overall social capital of Twitter users. As above, the slope of the regression line is not statistically significantly different from the one of the mean, *b** = 0.02, *t* = 0.40, *p* = 0.68, *95% CI* [-0.15, 0.33]. The coefficient of determination *R2* equals 0.001, which indicates that there is practically no correlation between the overall social capital and the expressed well-being of Twitter users in the sample. Figure *5* illustrates the regression model.

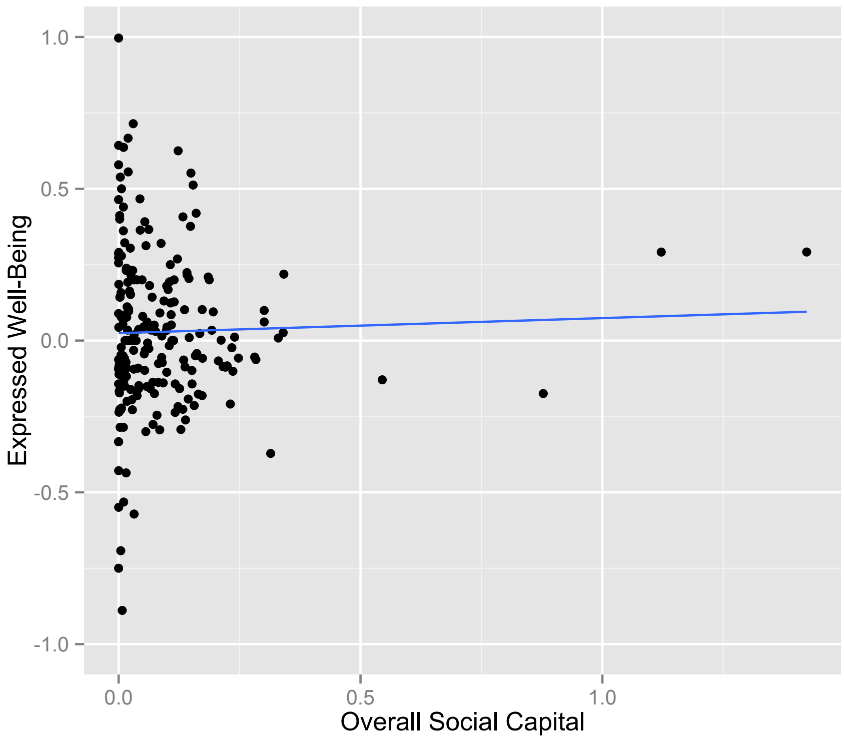

Figure 5: Scatterplot of Overall Social Capital and Expressed Well-Being with Regression Line

Just like for the two regression models above, the results of the regression analysis show that expressed well-being cannot be predicted from the overall social capital of Twitter users either.

5. Discussion and Conclusion

The results suggest that the answer to all three research questions is negative: Neither bonding social capital, bridging social capital nor overall social capital inferred from the online networks of Twitter users is a significant predictor of expressed well-being. The coefficients of determination *R2* of all three regression models are close to zero, which indicates that the three measures of social capital are almost completely unrelated to the expressed well-being of the users in the sample. No non-linear relationship was discernible either.

These results are contrary to the expectations raised by the literature. As established above, there are many studies suggesting a positive relationship between social capital and subjective well-being, and there is research suggesting that expressed well-being is indicative of subjective well-being. Altogether, four major and partially overlapping reasons suggest themselves why social capital might not have been found to be predictive of expressed well-being: (1) The measured dimensions of social capital are not correlated with happiness, (2) the findings of studies about offline social capital do not apply to virtual social capital, (3) expressed well-being is not indicative of subjective well-being, or (4) the measure of social capital is invalid. In the following I will discuss each of these points in light of previous research and theory derived from the literature.

1. The measured dimensions of social capital might not be correlated with happiness

Some studies taking a closer look at how specific dimensions of social capital are related to happiness found that only some indicators of social capital can be used to predict subjective well-being. For instance, Bjornskov (2005) showed that trust is the only component of social capital as defined by Putnam (1993) that underlies the effects on life-satisfaction. Similarly, Miller and Buys (2008) showed in a study examining social capital in Australian communities that only two of seven social capital indicators were predictive of happiness, namely value of life and feelings of trust and safety.

This suggests that trust might be the component of social capital with the highest effect on subjective well-being. As discussed above, trust belongs to the content dimension of social capital. The network-based approach taken in this study, however, focuses on the structural dimension of social capital that primarily considers the strength of social ties and the diversity of social networks. As I have argued above, the content and structural dimensions of social capital are closely interlinked. For instance, the strength of social ties is closely related to the trust among the two individuals connected by the tie. However, how is the strength of ties as operationalized in this study related to trust?

I operationalized the strength of social ties as the strength of explicit connections (s^{ESN}), which in turn was based on the relative amount of times a user interacted with another user. The more often users interacted with each other (relative to the total number of their interactions on Twitter), the stronger I assumed their ties to be. This corresponds with the argument of Smith, Giraud-Carrier and Stephens (2011) that interactions among social agents essentially influence the strength of their relationship and is in line with Granovetter's (1973) discussion of the strength of social ties, who argued that the strength of ties is, among other things, dependent on the amount of time spent maintaining the tie.

However, is the amount of interactions between two Twitter users indicative of trust between them, which in turn can be expected to have positive effects on happiness? This is at least questionable. In fact, I found a statistically significant *negative* relation between the amount of times Twitter users mentioned other users in their tweets and their expressed well-being, $r = -.14$, $p < .05$. While this would warrant some closer consideration in future studies, it might suggest that tweets mentioning other users are more often negative than positive. In this case, interactions between users are more likely to impede trust and social support than to promote it.

Moreover, a Pew study examining the use of social networking sites found that the number of Facebook visits per month was predictive of societal trust and the amount of social support received but that the same was not true for Twitter (Hampton et al., 2011). Social interactions on Facebook might therefore be more likely to promote trust than interactions on Twitter.[9] This might be due to the fact that interactions on Facebook are mostly restricted to users that have been accepted as

9. While the number of visits of a social networking site and the number of interactions with other users on this social networking site are obviously not the same, it can be assumed that, generally speaking, the users visiting Twitter more often also publish more tweets and therefore interact more often with others.

'friends', while on Twitter everyone can interact with any other user, whether they are friends or not. This, in turn, suggests the possibility that social capital determined in the way proposed here could in fact be used to predict the happiness of Facebook users, which should be examined in future studies.

To conclude, it is certainly possible that the examined dimensions of social capital are not correlated with happiness, while other dimensions of social capital are. The large number of studies claiming that social capital per se is related to higher levels of subjective well-being make it hard to decipher which societal phenomena exactly underlie these processes, since there exist many different definitions of social capital and almost as many different ways to measure it.

2. The findings of studies about offline social capital might not apply to virtual social capital

Most of the studies quoted above that show a relation between social capital and subjective well-being are based on 'offline' social capital and real-life social networks. It is possible that when inferring social capital from online social networks other phenomena affect people's subjective well-being that do not apply to offline social networks and that distort the results.

There exists a large amount of research studying the internet's effects on the happiness of its users (see Amichai-Hamburger & Barak, 2009). However, the findings of studies examining the social impacts of the internet are controversial and seem to depend on many different variables (Amichai-Hamburger & Barak, 2009; Huang, 2010; Shklovski, Kiesler, & Kraut, 2006). While this is not the place for reviewing the vast amount of literature discussing the internet's effects on subjective well-being, I will in the following highlight a few studies that illustrate the difficulties in studying the effects of Twitter use on people's happiness.

As mentioned above, a Pew study found that Facebook users were more trusting and receive more social support (Hampton et al., 2011), which is both linked to higher levels of happiness (Helliwell, 2001), but the study did not find similar correlations for Twitter. This suggests that there might be differences between Twitter and Facebook in terms of the effects on their users' subjective well-being. However, it is not entirely clear what the causes of these differences are. One of them might be the different nature of the social networking sites: On Twitter everyone can interact with any other user, while on Facebook users have to be confirmed as friends first. Online interactions with friends might be associated with greater well-being than interactions with strangers, as it has been shown for seventh-graders from a public school in the U.S. (Gross, Juvonen, & Gable, 2002). This, again, suggests the possibility that the results of a study similar to this one examining the relation of Facebook users' social capital and their expressed well-being would have obtained different results, which merits consideration of future research.

Another study conducted by Burke et al. (2010) showed that US-American college students who use Facebook more often to directly communicate with others feel less lonely. However, they also found that increasing levels of consumption of content on Facebook was associated with increased feelings of loneliness. Social capital as operationalized in this study is dependent on the amount of interactions between Twitter users. The more often a user was interacted with, the larger his or her amount of bonding social capital, bridging social capital, and overall social capital. If the findings of Burke et al. (2010) that direct communication is linked to reduced loneliness on Facebook also applied to Twitter, this would suggest that social capital is positively related to expressed well-being. However, this is contrary to the actual findings.

Other studies examining the effect of internet use on happiness in general obtained quite controversial results, but the relation of internet use and subjective well-being seems to be slightly negative according to some meta-analyses. A meta-analysis of 40 studies examining the internet's effects on individuals' well-being found that internet use for social purposes has a weak negative effect on well-being, indicated through slightly higher levels of depression and loneliness and lower levels of self-esteem and life satisfaction (Huang, 2010). Similarly, a meta-analysis conducted by Shklovski (2006) found that internet use seems to have a weak negative effect on the amount of face-to-face interactions, while face-to-face interactions have positive effects on well-being.

If the findings of these two meta-analyses apply to Twitter, this might indicate that higher use of Twitter leads to reduced subjective well-being. Since the amount of social capital of Twitter users as determined here is dependent on the amount of times other users interact with them, the ones with more social capital also tend to be the ones who use Twitter more heavily.[10] Therefore, the positive effects of social

10. Overall social capital of Twitter users is positively related to the number of tweets they publish, $r = .21$, $p < .01$. I assume that the number of tweets a user publishes is associated

capital on subjective well-being might be opposed by negative effects of higher Twitter usage which might distort the results.

Evidently, this is rather speculative and requires more research. It is not clear how exactly the use of Twitter is related to subjective well-being. However, this illustrates that there might be third variables distorting the results which do not apply to studies that focus on offline social capital, since the use of the internet itself affects subjective well-being in many different ways. More research is needed that studies the relation of virtual social capital and subjective well-being and that accounts for other phenomena affecting the subjective well-being of the examined individuals.

3. Expressed well-being might not be indicative of subjective well-being

Another possible reason for the lack of a correlation between social capital and expressed well-being of Twitter users might be that expressed well-being is not a valid indicator of subjective well-being. The only study found that examined whether the positivity of tweets is related to a measure of well-being is the one by Kramer (2010), which showed that the positivity of tweets is statistically significantly related to scores on a satisfaction with life scale, but the correlation has been found to be only modest in size ($r = .17$). While also based on sentiment analysis, Kramer's operationalization of happiness is slightly different from my operationalization of expressed well-being. Therefore, it is not entirely clear to what extent his findings apply to this study. However, since Kramer also used a lexicon-based approach to analyze the positivity of tweets and also computed the happiness score based on the relative positivity of tweets, it is reasonable to assume that they do apply to a similar extent.

As established above, however, life-satisfaction is only one component of subjective well-being and it is not entirely clear to what extent expressed well-being is related to overall happiness. There is no empirical evidence showing how the positivity of tweets correlates with the affective component of happiness, but, as I have discussed in more detail above, there is reason to assume that (a) Twitter users' moods are reflected in their tweets (Bollen et al., 2010) and (b) that the cognitive component of happiness is closely interlinked with the affective component

---

with the amount of time he or she spends using Twitter.

of happiness (Veenhoven, 2009). Therefore, if expressed well-being is correlated with life-satisfaction, which is part of the cognitive component of happiness, it is also likely to be correlated with the affective component of happiness.

In other words, previous research and the theory strongly suggest that expressed well-being as operationalized here is positively related to subjective well-being and that it therefore is a valid indicator of it. Thus, it seems unlikely that no correlation between social capital and expressed well-being was found because the latter one is not indicative of happiness. However, there is a need for research examining the relationship between the positivity of tweets and their authors' happiness more closely.

4. The operationalization of social capital might be invalid

Smith, Giraud-Carrier, Ventura et al. (2011) claim to have validated their framework through showing in case studies that it conforms with well-established theories from the social sciences. However, it is questionable to what extent the case studies show that their measure of social capital is valid. Their paper includes three case studies that measure social capital in a way according to their framework: In the first one (Smith & Giraud-Carrier, 2010) they show that Twitter users are more likely to be followed-back by others, the more similar the users are to each other. They operationalize the similarity of two users by comparing the words occurring in their profile descriptions and use this similarity score to calculate bonding and bridging social capital, as I have done it above. The second case study (Smith et al., 2008) shows that blogs are not only connected explicitly by links but also implicitly by topics. In other words, they used the framework to identify blogs that write about similar topics. The third case study (Smith, Giraud-Carrier, Dewey, Ring, & Gore, 2011) shows that the bridging social capital of students who go abroad predicts the self-assessed language improvement of those students. They found that 'bridgers' acquired more new language skills while abroad than 'bonders'. However, they claim that there has been no previous study examining the relationship between bonding and bridging social capital and language improvement in study abroad programs.

Therefore, only the first one among these case studies tests whether the framework corresponds with expectations that are grounded on previous empirical research. It shows that bonding relationships on Twitter are more likely than bridging relationships and therefore that the framework conforms with the well-established principle of homophily in social networks (McPherson, Smith-Lovin, & Cook, 2001). However, it can be questioned to what extent this validates their measure of social capital. As mentioned above, their operationalization of implicit affinities is based on the similarity of profile descriptions. Thus, in other words, they showed that Twitter users are more likely to be followed-back by users if they indicate to have similar interests in their profile descriptions. However, this does not necessarily show that their measure of social capital is valid, since social capital is a multi-dimensional concept that encompasses much more than the similarity between two people. In other words, it does not provide evidence for the measure's content validity (Cozby, 2011).

Stone and Hughes (2002) suggest to assess the validity of measures of social capital by checking if they correlate with variables social capital has been shown to be correlated with before - they suggest to test the measures' concurrent validity (Cozby, 2011). As established above, there exists much evidence showing that social capital correlates with subjective well-being. However, in this study I found no correlation between social capital determined in the way suggested by Smith, Giraud-Carrier, Ventura et al. (2011) and the expressed well-being of Twitter users. Assuming that expressed well-being is indicative of subjective well-being and that the substantial amount of research showing that social capital is related to subjective well-being can be trusted, this suggests that the framework developed by Smith, Giraud-Carrier, Ventura et al. does not provide valid measures of social capital.

However, as discussed earlier, there exist many different definitions of social capital and just as many different ways to measure it. Scholars do not agree on what exactly constitutes social capital and it is difficult to fully validate a measure of a concept without knowing what its constituents are. Therefore, as pointed out by Smith et al. (2009), validating any approach to measuring social capital is usually done against its underlying assumptions rather than against some commonly accepted truth.

From this perspective, even if social capital measured in the way proposed by Smith, Giraud-Carrier, Ventura et al. (2011) is not found to be a significant predictor of happiness and of other variables that have been shown to be related to social capital in future studies either, whether or not the framework is valid is likely to depend on the point of view. As Smith, Giraud-Carrier, Ventura et al. have shown in their case studies, the framework is useful to study social phenomena on the internet and it conforms with some expectations grounded on well-developed theories of the social sciences. More research is needed that tests if their measures conform with expectations derived from the social capital literature to find out whether or not their framework measures the same construct as the one that is commonly referred to as social capital.

There are a range of limitations of this study that should be noted. I just discussed the major issue that the framework employed in this study might not be valid. A more specific issue related to the measure's validity is that the distinction between bonding and bridging social capital neglects the strength of ties, as I have pointed out earlier. Many definitions of bonding and bridging social capital also differentiate between tie strengths. For instance, Putnam (1993) claims that bonding social capital comprises strong ties among the bonding group while bridging social capital is comprised of weak ties. The framework developed by Smith, Giraud-Carrier, Ventura et al. (2011) has been criticized for neglecting this distinction and only basing the differentiation between bonding and bridging social capital on the degree of similarity of users (Durst et al., 2013). Future studies should add the differentiation between tie strengths to the measures of bridging and bonding social capital.

A limitation of the measure of expressed well-being, besides the aforementioned issue that it might not be indicative of subjective well-being, concerns the method used to analyze the sentiments of tweets: While OpinionFinder has been applied to many studies and it has been found to provide useful results (e.g., Bollen et al., 2011; O'Connor et al., 2010), sentiment analyses generally are subject to inaccuracies. While I could not find research on the accuracy of the exact version of OpinionFinder used in this study, research shows that the accuracy of methods of sentiment analyses usually ranges between 65 and 75 per cent (Kouloumpis et al., 2011). Considering this, the threshold of a minimum of 10 tweets that could be classified as either positive or negative in order to be included in the sample might be too little to analyze the expressed well-being of Twitter users. However, I conducted preliminary analyses with higher thresholds which did not significantly change the outcomes. Moreover, a higher threshold would result in a sample consisting of Twitter users that differ even more from the average user, since they would have had to publish even more tweets in the analyzed three months. Increasing the analyzed period of time in order to increase the amount of tweets that can be analyzed, on the other hand, would also have increased the likelihood that the network of the examined Twitter users changed significantly in this time frame. In this case, the results might have been distorted since the determined social capital could have been significantly different from the users' social capital in the beginning of the analyzed period of time. However, the time frame of three months is also rather arbitrary. It is not clear how fast the expressed well-being can change due to changes in the Twitter network (if it does change at all, which is questionable according to the results of this study). Thus, future studies might experiment with larger or smaller time frames.

This leads to another issue: Many of the thresholds set in the research design are arbitrary due to the lack of previous research suggesting concrete values. For instance, public users, following which I considered an expression of interest, were defined as the top 0.05 per cent of users with the highest number of followers, which resulted in a threshold of 12,092 followers for a user to be considered as a public user. I determined this percentage based on a revision of Twitter users' number of followers and it appears to provide fairly accurate results. However, there are some apparently private individuals with a larger number of followers and some public figures such as politicians with a smaller number of followers. Therefore, the threshold could also have been a little higher or a little lower. Future research should examine more precisely where to set the threshold in order to increase the accuracy of the measure.

The same is true for the thresholds set for the determination of the sample. Setting other thresholds for a user to be included in the sample might have led to different results. For example, the lower threshold of a minimum of 30 followers for a user to be included in the sample could also have been higher. On the one hand, a minimum of 30 followers seemed reasonable to be able to assume that a user's Twitter network includes a considerable number of his or her friends. On the other hand, 30 followers are unlikely to reflect the overall social network of a user which is likely to include many more persons. However, setting a higher threshold would again have resulted in a sample that differs even more from the population of Twitter users. Future research should experiment with different thresholds for the determination of the sample.

Finally, as I have mentioned above, there are some limitations due to the Twitter API that should be noted. The Twitter API only allows to collect the most recent 3,200 tweets of users. About 4 per cent of the users whose tweets were collected[11] had published more than 3,200 tweets in the analyzed three months. However, this is not likely to have distorted the results significantly, since both expressed well-being and the strength of explicit connections, which were the only two variables based on an analysis of tweets, are relative measures and therefore should not be substantially affected by a slightly different number of tweets.

In conclusion, the most probable reasons for the fact that no measure of social capital was found to be predictive of expressed well-being are that (a) the measured dimensions of social capital are not related to people's happiness and that (b) third variables that do not apply to 'offline' social capital distort the results. Due to existing research suggesting otherwise, it seems unlikely that expressed well-being is not indicative of people's happiness. Furthermore, since social capital is such an imprecise concept, whether or not the measure of social capital deployed in this study is valid largely depends on the point of view.

However, it is questionable how useful the concept of social capital is if almost everyone adopts a slightly different understanding of it. Much research concludes that social capital promotes happiness, but this conclusion is only of limited use if it is not clear which characteristics of society are referred to exactly. The concept of social capital has been criticized for subsuming many independent processes under one label and thereby concealing what it is about exactly (Bjornskov, 2005). For similar reasons, Portes (1998) has argued that there is little reason to believe that the notion of social capital will help ameliorate major social problems. With that said, it is crucial to keep working on methods for the measurement of social capital that are commonly accepted in the scientific community for the notion of social capital not to deteriorate to a mere buzzword. If there existed a consensus on which societal processes are studied exactly when measuring social capital, the concept would be of more use for the development and evaluation of public policies.

11. As mentioned above, I collected the tweets of the 214 users in the sample and the ones of their followers.

6. References

Amichai-Hamburger, Y., & Barak, A. (2009). Internet and well-being. In Y. Amichai-Hamburger (Ed.), *Technology and psychological well-being* (pp. 34-76). Cambridge, UK: Cambridge University Press.

Annas, J. (1993). *The morality of happiness.* Oxford, UK: Oxford University Press.

Aristotle. *Nichomachean Ethics* (W. D. Ross, Trans.). Retrieved from http://ebooks.adelaide.edu.au/a/aristotle/nicomachean/index.html

Aristotle. *Rhetoric* (W. R. Roberts, Trans.). Retrieved from http://homepage.cs.uri.edu/courses/fall2007/csc305/Schedule/rhetoric/index.htm

Beevolve. (2012). An exhaustive study of Twitter users around the world. Retrieved from http://www.beevolve.com/twitter-statistics/#c1

Bellah, R. N. (2008). *Habits of the heart. Individualism and commitment in American life.* Berkeley and Los Angeles, CA: Univ. of California Press.

Berkman, L. F., & Syme, S. L. (1979). Social networks, host resistance, and mortality: A nine-year follow-up study of Alameda County residents. *American Journal of Epidemiology*, *109*, 186-204.

Berry, W. D. (1993). *Understanding regression assumptions.* Newbury Park, CA: SAGE Publications.

Bifet, A., & Frank, E. (2010, October). Sentiment knowledge discovery in twitter streaming data. In D. S. C. Proceedings of the 13th international conference, Australia (Ed.), *Discovery science* (pp. 1-15).

Bjornskov, C. (2003). The happy few: cross-country evidence on social capital and life satisfaction. *Kyklos*, *56*(1), 3-16.

Bjornskov, C. (2005). The multiple facets of social capital. *European Journal of Political Economy*, *22*(1), 22-40.

Boase, J., Horrigan, J. B., Wellman, B., & Rainie, L. (2006). The strength of internet ties. Retrieved from http://www.pewinternet.org/files/old-media//Files/Reports/2006/PIP_Internet_ties.pdf.pdf

Bollen, J., Mao, H., & Pepe, A. (2010, July). *Modeling public mood and emotion: Twitter sentiment and socio-economic phenomena*. Proceedings of the fifth international AAAI conference on weblogs and social media.

Bollen, J., Gonzalves, B., Ruan, G., & Mao, H. (2011). Happiness is assortative in online social networks. *Artificial Life*, *17*(3), 237-251.

Bourdieu, P. (1986). The forms of capital. In J. Richardson (Ed.), *Handbook of Theory and Research for the Sociology of Education.* Westport, CT: Greenwood Press.

Bruns, A., & Burgess, J. E. (2011, August). *The use of Twitter hashtags in the formation of ad hoc publics*. Proceedings of the 6th European consortium for political research general conference.

Burke, M., Marlow, C., & Lento, T. (2010, April). *Social network activity and social well-being*. Proceedings of the SIGCHI conference on human factors in computing systems.

Burt, R. S. (2009). *Structural holes: The social structure of competition.* Harvard University Press.

Cha, M., Haddadi, H., Benevenuto, F., & Gummadi, P. K. (2010, May). *Measuring user influence in Twitter: The million follower fallacy*. Proceedings of the Fourth International AAAI Conference on Weblogs and Social Media.

Cohen, S., Doyle, W. J., Skoner, D. P., Rabin, B. S., & Gwaltney, J. M. (1997). Social ties and susceptibility to the common cold. *Jama*, *277*(24), 1940-1944.

Coleman, J. S. (1988). Social capital in the creation of human capital. *American Journal of Sociology*, *94*, 95-120.

Conover, M. D., Goncalves, B., Ratkiewicz, J., Flammini, A., & Menczer, F. (2011). *Predicting the political alignment of Twitter users. Privacy, security, risk and trust*. IEEE third international conference on social computing.

Conte, R., Gilbert, N., Bonelli, G., Cioffi-Revilla, C., Deffuant, G., Kertesz, J., … Helbing, D. (2012). Manifesto of computational social science. *The European Physical Journal Special Topics*, *214*(1), 325-246.

Cozby, P. C. (2011). *Methods in behavioral research* (10 ed.). McGraw-Hill Primis.

Daniel, B., Schwier, R., & McCalla, G. (2003). Social capital in virtual learning communities and distributed communities of practice. *Canadian Journal of Learning and Technology*, *29*(3), 113-139.

Diener, E., Sandvik, E., & Pavot, W. (1991). Happiness is the frequency, not the intensity, of positive versus negative affect. *Subjective Well-Being: An Interdisciplinary Perspective*, *21*, 119-139.

Diener, E., & Seligman, M. E. (2004). Beyond money: Toward an economy of well-being. *Psychological Science in the Public Interest*, *5*(1), 1-31.

Diener, E. D., Emmons, R. A., Larsen, R. J., & Griffin, S. (1985). The satisfaction with life scale. *Journal of Personality Assessment*, *49*(1), 71-75.

Dodds, P. S., & Danforth, C. M. (2009). Measuring the happiness of large-scale written expression: Songs, blogs, and presidents. *Journal of Happiness Studies*, *11*(4), 441-456.

Dodds, P. S., Harris, K. D., Kloumann, I. M., Bliss, C. A., & Danforth, C. M. (2011). Temporal patterns of happiness and information in a global social network: Hedonometrics and Twitter. *PloS One*, *6*(12). Retrieved from http://www.plosone.org/article/info%3Adoi%2F10.1371%2Fjournal.pone.0026752#pone-0026752-g017

Durbin, J., & Watson, G. S. (1951). Testing for serial correlation in least squares regression. II. *Biometrika*, *38*(1/2), 159-177.

Durst, C., Viol, J., & Wickramasinghe, N. (2013). Online social networks, social capital and health-related behaviors: A state-of-the-art analysis. *Communications of the Association for Information Systems*, *32*, 134-158.

Dutt, A. K., & Radcliff, B. (2009). *Happiness, economics and politics.* Cheltenham, Northampton, MA: Edward Elgar.

Feldmann, T. R., & Assaf, S. (1999). Social capital: Conceptual frameworks and empirical evidence. Retrieved from http://siteresources.worldbank.org/

INTSOCIALCAPITAL/Resources/Social-Capital-Initiative-Working-Paper-Series/SCI-WPS-05.pdf

Franke, S. (2005). Social capital: Reference document for public policy research, development and evaluation. Retrieved from http://publications.gc.ca/site/eng/280036/publication.html

Frey, B. S., & Stutzer, A. (2000). Happiness prospers in democracy. *Journal of Happiness Studies*, *1*(1), 79-102.

Gauntlett, D. (2011). *Making is connecting*. Cambridge, UK: Polity Press.

Gibbs, J. L., Ellison, N. B., & Heino, R. D. (2006). Self-presentation in online personals the role of anticipated future interaction, self-disclosure, and perceived success in Internet dating. *Communication Research*, *33*(2), 157-177.

Graham, C. (2008). The economics of happiness. In S. Durlauf & L. Blume (Eds.), *The new palgrave dictionary of economics.* Palgrave Macmillan. Retrieved from http://www.dictionaryofeconomics.com/article?id=pde2008_H000015

Granovetter, M. (1973). The strength of weak ties. *American Journal of Sociology*, *78*(6), 1360-1380.

Gross, E. F., Juvonen, J., & Gable, S. L. (2002). Internet use and well-being in adolescence. *Journal of Social Issues*, *58*(1), 75-90.

Hampton, K., Goulet, L. S., Rainie, L., & Purcell, K. (2011). Social networking sites and our lives. Retrieved from http://www.pewinternet.org/files/old-media/Files/Reports/2011/PIP%20-%20Social%20networking%20sites%20and%20our%20lives.pdf

Hanifan, L. J. (1920). *The community center.* Boston: Silver Burdett.

Harrison, M. J., & McCabe, B. P. M. (1975). Autocorrelation with heteroscedasticity: A note on the robustness of the Durbin-Watson, Geary and Henshaw Tests. *Biometrics*, *62*(1), 214-216.

Haythornthwaite, C. (2005). Social networks and internet connectivity effects. *Information, Community, and Society*, *8*(2), 125-147.

Helliwell, J. F. (2001). Social capital, the economy and well-being. In K. Banting, A. Sharpe, & F. St-Hilaire (Eds.), *The review of economic performance and social progress* (pp. 43-60). Montreal and Ottawa: Institute for research on public policy and centre for the study of living standards.

Helliwell, J. F. (2003). How's life? Combining individual and national variables to explain subjective well-being. *Economic Modelling*, *20*(2), 331-360.

Helliwell, J. F. (2004). Well-being and social capital: Does suicide pose a puzzle? Retrieved from http://faculty.arts.ubc.ca/jhelliwell/papers/w10896.pdf

Helliwell, J. F. (2006). Well-being, social capital and public policy: What›s new? *The Economic Journal*, *116*(510), C34-C45.

Helliwell, J. F., & Putnam, R. D. (2004). The social context of well-being. *Philosophical Transactions of the Royal Society of London B*, *359*(1449), 1435-1446.

Himelboim, I., McCreery, S., & Smith, M. (2013). Birds of a feather tweet together: Integrating network and content analyses to examine cross-ideology exposure on Twitter. *Journal of Computer-Mediated Communication*, *18*(2), 40-60.

Huang, C. (2010). Internet use and psychological well-being: A meta-analysis. *Cyberpsychology, Behavior, and Social Networking*, *13*(3), 241-249.

Inglehart, R., & Klingemann, H. D. (2000). Genes, culture, democracy, and happiness. In E. Diener & M. Suh (Eds.), *Culture and subjective well-being* (pp. 165-183).

Jansen, B. J., Zhang, M., Sobel, K., & Chowdury, A. (2009). Twitter power: Tweets as electronic word of mouth. *Journal of the American Society for Information Science and Technology*, *60*(11), 2169-2188.

Java, A., Song, X., Finin, T., & Tseng, B. (2007, August). *Why we twitter: Understanding microblogging usage and communities*. Proceedings of the 9th WebKDD and 1st SNA-KDD 2007 workshop on web mining and social network analysis.

Kahneman, D., & Krueger, A. B. (2006). Developments in the measurement of subjective well-being. *The journal of economic perspectives*, *20*(1), 3-24.

Kim, J., & Lee, J. E. R. (2011). The Facebook paths to happiness: Effects of the number of Facebook friends and self-presentation on subjective well-being. *Cyberpsychology, behavior, and social networking*, *14*(6), 359-364.

Kouloumpis, E., Wilson, T., & Moore, J. (2011, July). *Twitter sentiment analysis: The good, the bad and the omg!* Proceedings of the fifth international AAAI conference on weblogs and social media.

Kramer, A. D. (2010, April). *An unobtrusive behavioral model of gross national happiness*. Proceedings of the SIGCHI conference on human factors in computing systems.

La Porta, R. (1997). Trust in large organisations. *The American Economic Review*, *87*, 333-338.

Layard, R. (2006). *Happiness: Lessons from a new Science.* London: Penguin UK.

Layard, R. (2010). Measuring subjective well-being. *Science*, *327*(5965), 534-535.

Lenhart, A., & Madden, M. (2007). Social networking websites and teens. Retrieved from http://www.pewinternet.org/files/old-media//Files/Reports/2007/PIP_SNS_Data_Memo_Jan_2007.pdf.pdf

Lerman, K., & Ghosh, R. (2010, May). *Information contagion: An empirical study of the spread of news on Digg and Twitter social networks*. Proceedings of the fourth international AAAI conference on weblogs and social media.

Liu, B. (2010). Sentiment analysis and subjectivity. In N. Indurkhya & F. J. Damerau (Eds.), *Handbook of natural language processing* (pp. 627-666).

Lyubomirsky, S., Sheldon, K. M., & Schkade, D. (2005). Pursuing happiness: The architecture of sustainable change. *Review of General Psychology*, *9*(2), 101-148.

McPherson, M., Smith-Lovin, L., & Cook, J. M. (2001). Birds of a feather: Homophily in social networks. *Annual Review of Sociology*, *27*, 415-444.

Michalos, A. C. (1985). Multiple discrepancies theory (MDT). *Social Indicators Research*, *16*, 347-413.

Miller, E., & Buys, L. (2008). Does social capital predict happiness, health, and life satisfaction in an urban Australian community? *Kotuitui: New Zealand Journal of Social Sciences Online*, *3*(1), 15-20.

Naaman, M., Boase, J., & Lai, C. H. (2010). Is it really about me? Message content in social awareness streams. *Proceedings of the 2010 ACM conference on Computer supported cooperative work,* 189-192.

O'Connor, B., Balasubramanyan, R., Routledge, B. R., & Smith, N. A. (2010, May). *From tweets to polls: Linking text sentiment to public opinion time series*. Proceedings of the fourth international AAAI conference on weblogs and social media.

Ong, A. D. (2009). On the measurement and mismeasurement of happiness: contemporary theories and methodological directions. In A. K. Dutt & B. Radcliff (Eds.), *Happiness, economics and politics* (pp. 33-44). Cheltenham, Northampton, MA: Edward Elgar.

Pacek, A. C. (2009). Politics and happiness: an empirical ledger. In A. K. Dutt & B. Radcliff (Eds.), *Happiness, economics and politics.* Cheltenham,

Northampton, MA: Edward Elgar.

Pang, B., & Lee, L. (2008). Opinion mining and sentiment analysis. *Foundations and Rrends in Information Retrieval*, *2*(1/2), 1-135.

Patulny, R. V., & Svendsen, G. L. H. (2007). Exploring the social capital grid: bonding, bridging, qualitative, quantitative. *International Journal of Sociology and Social Policy*, *27*(1/2), 32-51.

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

Putnam, R. D. (2000). *Bowling alone: The collapse and revival of American community.* New York, NY: Simon and Schuster.

Putnam, R. D. (2001). Social capital: Measurement and consequences. *Canadian Journal of Policy Research*, *2*(1), 41-51.

Putnam, R. D., Leonardi, R., & Nanetti, R. Y. (1993). *Making democracy work: Civic traditions in modern Italy.* Princeton, NJ: Princeton University Press.

Quercia, D., Ellis, J., Capra, L., & Crowcroft, J. (2012, February). *Tracking gross community happiness from tweets*. Proceedings of the ACM 2012 conference on computer supported cooperative work.

Schwarz, N., & Strack, F. (1991). Evaluating one's life: A judgment model of subjective well-being. *Subjective Well-Being: An Interdisciplinary Perspective*, *21*, 27-47.

SEC. (2013). Twitter, Inc. Retrieved from http://www.sec.gov/Archives/edgar/data/1418091/000119312513424260/d564001ds1a.htm

Shah, D. V., McLeod, J. M., & Yoon, S. H. (2001). Communication, context, and community: An exploration of print, broadcast, and internet influences. *Communication Research*, *28*(4), 464.

Shklovski, I., Kiesler, S., & Kraut, R. (2006). The Internet and social interaction: A meta-analysis and critique of studies. In R. Kraut, M. Brynin, & S. Kiesle (Eds.), *Computers, phones, and the Internet: Domesticating information* *technology* (pp. 251-264). New York, NY: Oxford University Press.

Smith, A., & Brenner, J. (2012). Twitter Use 2012. Retrieved from http://www.pewinternet.org/files/old-media//Files/Reports/2012/PIP_Twitter_Use_2012.pdf

Smith, M., Giraud-Carrier, Stephens, S. (2011). Measuring and reasoning about social capital: A computational framework. In M. S. Smith, C. Giraud-Carrier, D. A. Ventura, D. P. Dewey, C. D. Knutson, & D. W. Embley (Eds.), *A computational framework for social capital in online communities. (Doctoral dissertation).* Retrieved from http://site.smithworx.com/publications/d.pdf

Smith, M., Giraud-Carrier, C., Dewey, D. P., Ring, S., & Gore, D. (2011, July). *Social capital and language acquisition during study abroad*. 33rd Annual Conference of the Cognitive Science Society.

Smith, M., Giraud-Carrier, C., & Purser, N. (2009). Implicit affinity networks and social capital. *Information Technology and Management*, *10*(2/3), 123-134.

Smith, M. R., Purser, N., & Giraud-Carrier, C. G. (2008, March). *Social capital in the blogosphere: A case study*. AAAI Spring Symposium: Social Information Processing.

Smith, M. S., & Giraud-Carrier, C. (2010, August). *Bonding vs. bridging social capital: A case study in twitter*. IEEE second international conference on social computing.

Smith, M. S., Giraud-Carrier, C., Ventura, D. A., Dewey, D. P., Knutson, C. D., & Embley, D. W. (2011). *A computational framework for social capital in online communities. (Doctoral dissertation).* Retrieved from http://site.smithworx.com/publications/d.pdf

Stevens, J. P. (2009). *Applied multivariate statistics for the social sciences* (5th ed.). New York, NY: Routledge.

Stone, W., & Hughes, J. Social capital: Empirical meaning and measurement validity. *2002*. Retrieved from http://www.aifs.org.au/institute/pubs/rp27/

Subrahmanyam, K., Reich, S. M., Waechter, N., & Espinoza, G. (2008). Online and offline social networks: Use of social networking sites by emerging adults. *Journal of Applied Developmental Psychology*, *29*(6), 420-433.

Szreter, S., & Woolcock, M. (2004). Health by association? Social capital, social theory, and the political economy of public health. *International Journal of Epidemiology*, *33*(4), 650-667.

Tatarkiewicz, W. (1975). *Analysis of happiness.* Den Haag: Nijhoff.

Twopcharts. (2014). Twitter activity monitor. Retrieved from http://twopcharts.com/twitteractivitymonitor

Veenhoven, R. (1984). *Conditions of happiness. (Doctoral dissertation).* Retrieved from http://www2.eur.nl/fsw/research/veenhoven/Pub1980s/84a-full.pdf

Veenhoven, R. (2009). How do we assess how happy we are? Tenets, implications and tenability of three theories. In A. K. Dutt & B. Radcliff (Eds.), *Happiness, economics and politics* (pp. 45-69). Cheltenham, Northampton, MA: Edward Elgar.

Veenhoven, R. (1996). Developments in satisfaction-research. *37*(1), 1-46.

Wasserman, S., & Faust, K. (1994). *Social network analysis: Methods and applications* (8). Cambridge University Press.

Weng, J., Lim, E. P., Jiang, J., & He, Q. (2010, February). *Twitterrank: finding topic-sensitive influential twitterers*. Proceedings of the third ACM international conference on Web search and data mining.

Williams, D. (2006). On and off the 'Net: Scales for social capital in an online era. *Journal of Computer-Mediated Communication*, *11*(2), 593-628.

Williams, R., & Edge, D. (1996). The social shaping of technology. *Research Policy*, *26*(6), 865-899.

Wilson, T., Hoffmann, P., Somasundaran, S., Kessler, J., Wiebe, J., Choi, Y., & Patwardhan, S. (2005, October). *OpinionFinder: A system for subjectivity analysis*. Proceedings of HLT/EMNLP on interactive demonstrations.

Wilson, W. (1967). Correlates of avowed happiness. *Psychological Bulletin*, *67*, 294-306.

Woolcock, M., & Narayan, D. (2000). Social capital: Implications for development theory, research, and policy. *The World Bank Research Observer*, *15*(2), 225-249.

Young, A. L., & Quan-Haase, A. (2009, June). *Information revelation and internet privacy concerns on social network sites: a case study of facebook*. Proceedings of the fourth international conference on communities and technologies.

Zeileis, A. (2004). Econometric computing with HC and HAC covariance matrix estimators. *Journal of Statistical Software*, *11*(10), 1-17. Retrieved from http://www.jstatsoft.org/v11/i10/

Zhao, S., Grasmuck, S., & Martin, J. (2008). Identity construction on Facebook: Digital empowerment in anchored relationships. *Computers in human behavior*, *24*(5), 1816-1836.

Appendix: Python Code

In the following I present a selection of the Python scripts I wrote to gather, store and preprocess the data. I used some code from Matthew A. Russel's book *Mining the Social Web*[12] to collect the data from Twitter and to store it in a database. I used Python version 2.7 and the database MongoDB (version 2.6.1).

To connect to the Twitter API I used the Python modules *twitter* and *twython.* Using both modules allows to relatively easily authenticate in different ways that are subject to different rate limits and thereby to increase the amount of data that can be mined from Twitter within a certain period of time. Also due to rate limits of the Twitter API, I let some of the scripts run on a web server that was up 24 hours a day.

1. Script used to compute social capital and to store the results in a database

```
import pymongo
from datetime import datetime
import time
import os

def load_from_mongo(mongo_db, mongo_db_coll, return_cursor=False,
                    criteria=None, projection=None, findOne=False,
**mongo_conn_kw):

    client = pymongo.MongoClient(**mongo_conn_kw)
    db = client[mongo_db]
    coll = db[mongo_db_coll]

    if criteria is None:
        criteria = {}

    if findOne == False:
        if projection is None:
            cursor = coll.find(criteria)
        else:
            cursor = coll.find(criteria, projection)
    else:
        if projection is None:
            cursor = coll.find_one(criteria)
        else:
            cursor = coll.find_one(criteria, projection)
        return cursor

    if return_cursor:
```

12. Russell, M. A. (2013). *Mining the Social Web: Data Mining Facebook, Twitter, LinkedIn, Google+, GitHub, and More*. (2nd ed). Sebastopol, CA: O'Reilly Media.

```
        return cursor
    else:
        return [ item for item in cursor ]


def update_sc(bonding, bridging, mongo_db, mongo_db_coll,
mongo_document_id, **mongo_conn_kw):

    client = pymongo.MongoClient(**mongo_conn_kw)

    db = client[mongo_db]

    coll = db[mongo_db_coll]

    return coll.update(
        {
            "_id": mongo_document_id
        },
        {
            "$set":
            {
                "bonding_sc": bonding,
                "bridging_sc": bridging
            }
        }
    )


def jaccard_index(a, b):
    if len(a) == 0 or len(b) == 0:
        return 0

    if not isinstance(a[0], (int, long)):
        a = [ int(x) for x in a ]
    if not isinstance(b[0], (int, long)):
        b = [ int(x) for x in b ]

    set_1 = set(a)
    set_2 = set(b)

    n = len(set_1.intersection(set_2))
    return n / float(len(set_1) + len(set_2) - n)


def update_followers_public_friends(user):

    f = open(os.path.dirname(__file__)+"p_users", "r")
    p_user_ids = set([ line.rstrip() for line in f ])
    f.close()

    client = pymongo.MongoClient()
    db = client["users"]
    degree_2 = db["degree_2"]

    for follower_id in user['followers_ids']:
        for follower in degree_2.find({"id": int(follower_id)}):

            if "public_friends_count" in follower: continue
```

```
            p_friends = [];
            for id in follower['friends_ids']:
                if str(id) in p_user_ids:
                    p_friends.append(id)

            degree_2.update({"_id": follower['_id']},
                            {"$set":{"public_friends_count":
len(p_friends),"public_friends_ids": p_friends}})


date_limit = datetime.strptime("Wed Apr 16 2014", '%a %b %d %Y')

while True:

    degree_1 = load_from_mongo("users", "degree_1",
return_cursor=True)

    try:
        n = 0

        for user in degree_1:
            n+=1

            if "public_friends_ids" not in user:
                continue

            bonding = 0
            bridging = 0

            print "Updating user " + str(user['id']) + "..."

            length = len(user['followers_ids'])
            i=0
            for follower_id in user['followers_ids']:
                follower = load_from_mongo("users", "degree_2",
return_cursor=True, findOne=True, criteria={'id': follower_id})

                if follower is None:
                    print "Follower of " + str(user['id']) + " with
ID " + str(follower_id) + " does not exist"
                    continue

                if "public_friends_ids" not in follower:
                    print "Updating public friends of " +
str(user['id']) + "'s followers..."
                    update_followers_public_friends(user)
                    continue

                IAN = jaccard_index(user['public_friends_ids'],
follower['public_friends_ids'])

                # ESN: Number of of follower's interactions with user
devided by total number of follower's interactions with any user

                # Get the ids of the users that were interacted with
for every interaction

                if "tweets" not in follower:
                    print "Skipped follower " + str(follower_id) + "
```

```
because of missing tweets"
                    continue

                interactions = []

                for tweet in follower['tweets']:

                    date_str =
tweet['created_at'][:tweet['created_at'].rfind(':')-5]
                    date_str += tweet['created_at'][-4:]
                    tweet_created_at = datetime.strptime(date_str,
'%a %b %d %Y')

                    # Continue if tweet has been created more than 90
days ago
                    if str(date_limit -
tweet_created_at).find("day") != -1:
                        if int(str(date_limit -
tweet_created_at)[:str(date_limit - tweet_created_at).find(' ')]) >
90:
                            continue

                    for mention in
tweet['entities']['user_mentions']:
                        interactions.append(mention['id'])

                if len(interactions) == 0:
                    ESN = 0
                else:
                    ESN = interactions.count(user['id']) /
float(len(interactions))

                bonding += IAN * ESN
                bridging += (1-IAN) * ESN

                i += 1

                print "IAN: " +str(IAN) + " ESN: " + str(ESN)
                print "Processed " + str(i) + " out of " +
str(length) + " followers"

            print "Bridging: " + str(bridging) + " Bonding: " +
str(bonding)

            update_sc(bonding, bridging, "users", "degree_1",
user['_id'])

            print "Updated user " + str(user['id'])

        if n == 0: break

    # MongoDB cursor timeout
    except pymongo.errors.OperationFailure, e:
        msg = e.message
        if not ( msg.startswith( "cursor id" ) and msg.endswith( "not
valid at server" ) ):
            raise
```

2. Script running on web server to retrieve the IDs of users' friends and followers

```
from datetime import datetime
from sys import maxint
from functools import partial
from contextlib import contextmanager

from urllib2 import URLError
from httplib import BadStatusLine
from twython import TwythonError, TwythonRateLimitError,
TwythonAuthError

import twitter
import sys
import time
import os
import fcntl

# Scripts including authentication information
from t_auth import twython_auth
from t_auth import twitter_auth

def make_twython_request(twitter_api_func, *args, **kw):

    error_count = 0
    while True:
        try:
            return twitter_api_func(*args, **kw)
        except TwythonRateLimitError:
            print >> sys.stderr, 'Encountered 429 Error (Rate Limit
Exceeded) with application-only authentication'
            return 429
        except TwythonAuthError:
            print >> sys.stderr, 'Encountered 401 Error (Not
Authorized)'
            return None
        except TwythonError, e:
            if str(e).find("404") != -1:
                print >> sys.stderr, 'Encountered 404 Error (Not
Found)'
                return None
            elif error_count < 5:
                print >> sys.stderr, 'Encountered unexpected Error.
Retrying in 10 seconds'
                print >> sys.stderr, str(e)
                time.sleep(10)
                error_count += 1
            else:
                print >> sys.stderr, 'Too many retries. Skipping
request...'
                return None


def make_twitter_request(twitter_api_func, max_errors=10, *args,
**kw):
```

```
    def handle_twitter_http_error(e, wait_period=2,
sleep_when_rate_limited=False):

        if wait_period > 3600: # Seconds
            print >> sys.stderr, 'Too many retries. Quitting.'
            raise e

        if e.e.code == 401:
            print >> sys.stderr, 'Encountered 401 Error (Not
Authorized)'
            return None
        elif e.e.code == 404:
            print >> sys.stderr, 'Encountered 404 Error (Not Found)'
            return None
        elif e.e.code == 429:
            print >> sys.stderr, 'Encountered 429 Error (Rate Limit
Exceeded) with user authentication'
            if sleep_when_rate_limited:
                print >> sys.stderr, "Retrying in 15 min-
utes...ZzZ..."
                sys.stderr.flush()
                time.sleep(60*15 + 5)
                print >> sys.stderr, '...ZzZ...Awake now and trying
again.'
                return 2
            else:
                return 429 # Caller must handle the rate limiting
issue
        elif e.e.code in (500, 502, 503, 504):
            print >> sys.stderr, 'Encountered %i Error. Retrying in
%i seconds' % \
                (e.e.code, wait_period)
            time.sleep(wait_period)
            wait_period *= 1.5
            return wait_period
        else:
            raise e

    wait_period = 2
    error_count = 0

    while True:
        try:
            return twitter_api_func(*args, **kw)
        except twitter.api.TwitterHTTPError, e:
            error_count = 0
            wait_period = handle_twitter_http_error(e, wait_period)
            if wait_period is None:
                return
            if wait_period == 429:
                return 429
        except URLError, e:
            error_count += 1
            time.sleep(wait_period)
            wait_period *= 1.5
            print >> sys.stderr, "URLError encountered. Continuing."
            if error_count > max_errors:
                print >> sys.stderr, "Too many consecutive er-
```

```
rors...bailing out."
                raise
        except BadStatusLine, e:
            error_count += 1
            time.sleep(wait_period)
            wait_period *= 1.5
            print >> sys.stderr, "BadStatusLine encountered.
Continuing."
            if error_count > max_errors:
                print >> sys.stderr, "Too many consecutive er-
rors...bailing out."
                raise



def get_friends_followers_ids(twitter_api, screen_name=None,
user_id=None,
                              friends_limit=maxint, followers_lim-
it=maxint):


    assert (screen_name != None) != (user_id != None), \
    "Must have screen_name or user_id, but not both"


    get_friends_ids, get_followers_ids = [],[]

    if isinstance(twitter_api, twitter.api.Twitter):
        get_friends_ids = partial(make_twitter_request, twit-
ter_api.friends.ids,
                          count=5000)
        get_followers_ids = partial(make_twitter_request, twit-
ter_api.followers.ids,
                            count=5000)

    else:
        get_friends_ids = partial(make_twython_request, twit-
ter_api.get_friends_ids,
                                  count=5000)
        get_followers_ids = partial(make_twython_request, twit-
ter_api.get_followers_ids,
                                    count=5000)

    friends_ids, followers_ids = [], []

    for twitter_api_func, limit, ids, label in [
                    [get_friends_ids, friends_limit, friends_ids,
"friends"],
                    [get_followers_ids, followers_limit, follow-
ers_ids, "followers"]
                ]:

        if limit == 0: continue

        cursor = -1
        while cursor != 0:

            if screen_name:
                response = twitter_api_func(screen_name=screen_name,
```

```
cursor=cursor)
            else: # user_id
                response = twitter_api_func(user_id=user_id, cur-
sor=cursor)

            if response == 429:
                return 429, 429

            if response is not None:
                ids += response['ids']
                cursor = response['next_cursor']

            print >> sys.stderr, 'Fetched {0} total {1} ids for
{2}'.format(len(ids),
                                                    label, (user_id
or screen_name))


            if len(ids) >= limit or response is None:
                break


    return friends_ids[:friends_limit], followers_ids[:follow-
ers_limit]


def lockFile(lockfile):
    fp = open(lockfile, 'w')

    try:
        fcntl.lockf(fp, fcntl.LOCK_EX | fcntl.LOCK_NB)
    except IOError:
        return False

    return True


# Ensure that only one instance of the script is running
@contextmanager
def file_lock(lock_file):
    if not lockFile(lock_file):
        print 'Only one script can run at once. '\
              'Script is locked with %s' % lock_file
        sys.exit(-1)
    else:
        fp = open(lock_file, 'w')
        fcntl.lockf(fp, fcntl.LOCK_EX | fcntl.LOCK_NB)
        try:
            yield
        finally:
            os.remove(lock_file)


with file_lock('ts-lock'):

    twitter_api = twitter_auth()
    twython_api = twython_auth()
```

```
    read_user_id = False
    has_friends_ids = False
    has_followers_ids = False

    user_ids = []


    use_twitter_api = True
    window_started_at = datetime.now()

    nRequests=0
    id = ''

    friends_ids, followers_ids = [], []

    ids = []


    f = open(os.path.dirname(__file__)+"/ids4.txt", "r")


    DEGREE = f.readline()

    while True:
        line = f.readline()
        if line == '':
            break
        else:
            ids.append([line.rstrip(), f.readline().rstrip()])

    f.close()


    if not os.path.exists(os.path.dirname(__file__)+"/data/"):
        os.makedirs(os.path.dirname(__file__)+"/data/")

    started_at = datetime.now()

    f = open(os.path.dirname(__file__)+"/data/logfile", "w")
    f.write("Started at: " + str(started_at) +"\n")
    f.close()

    nUpdates = 0

    for object_id, user_id in ids:

        if not os.path.exists(os.path.dirname(__file__)+"/data/"+ob-
ject_id):

            while True:

                if use_twitter_api:
                    friends_ids, followers_ids = get_friends_follow-
ers_ids(twitter_api, user_id=user_id, friends_limit=5000, follow-
ers_limit=5000)

                    if friends_ids == 429 and followers_ids == 429:
                        use_twitter_api = False
                        print >> sys.stderr, "Switching to applica-
```

```
tion-only authentication in a minute..."
                        time.sleep(60)
                        continue


                if not use_twitter_api:
                    friends_ids, followers_ids = get_friends_follow-
ers_ids(twython_api, user_id=user_id, friends_limit=5000, follow-
ers_limit=5000)

                    if friends_ids == 429 and followers_ids == 429:
                        if (datetime.now()-window_started_at).sec-
onds > 905:
                            print >> sys.stderr, "Switching to user
authentication..."
                        else:
                            print >> sys.stderr, "Retrying in " +
str((900-((datetime.now()-window_started_at).seconds))/60) + "
minutes"
                            sys.stderr.flush()
                            time.sleep(60*15-(datetime.now()-win-
dow_started_at).seconds + 5)
                            print >> sys.stderr, '...ZzZ...Awake now
and trying again.'

                        use_twitter_api = True
                        window_started_at = datetime.now()
                        continue

                break

            f = open(os.path.dirname(__file__)+"/data/"+object_id,
"w")

            f.write("USER_ID\n" + user_id + "\n")

            f.write("FRIENDS_IDS"+"\n")
            for x in friends_ids:
                f.write(str(x)+"\n")

            f.write("FOLLOWERS_IDS"+"\n")
            for x in followers_ids:
                f.write(str(x)+"\n")

            f.close()

            if len(friends_ids) == 0 and len(followers_ids) == 0:
                f = open(os.path.dirname(__file__)+"/data/logfile",
"a")
                f.write("No friends or followers retrieved for user "
+ user_id + "\n")
                f.close()

            nUpdates += 1

            if (nUpdates + 1) % 100 == 0:
                f = open(os.path.dirname(__file__)+"/data/logfile",
"a")
                f.write("Updated " + str(nUpdates) + " users in " +
```

```
str((datetime.now()-started_at).seconds/60) + " minutes\n")
                f.close()


    f = open(os.path.dirname(__file__)+"/data/logfile", "a")
    f.write("Updated a total of " + str(nUpdates) + " users in " +
str((datetime.now()-started_at).seconds/60) + " minutes")
    f.close()
```

3. Script used to collect the tweets of the followers of the 214 users in the sample

```
from twython import Twython
from twython import TwythonError, TwythonRateLimitError,
TwythonAuthError
from datetime import datetime

import pymongo
import twitter
import sys
import time

# Scripts including authentication information
from t_auth import twython_auth
from t_auth import twitter_auth


def load_from_mongo(mongo_db, mongo_db_coll, return_cursor=False,
                    criteria=None, projection=None, **mongo_conn_kw):

    client = pymongo.MongoClient(**mongo_conn_kw)
    db = client[mongo_db]
    coll = db[mongo_db_coll]

    if criteria is None:
        criteria = {}

    if projection is None:
        cursor = coll.find(criteria)
    else:
        cursor = coll.find(criteria, projection)

    if return_cursor:
        return cursor
    else:
        return [ item for item in cursor ]


def update_tweets(tweets, mongo_db, mongo_db_coll, mongo_document_id,
**mongo_conn_kw):

    client = pymongo.MongoClient(**mongo_conn_kw)

    db = client[mongo_db]

    coll = db[mongo_db_coll]
```

```
    coll.update({ "_id": mongo_document_id },
                { "$set": {"tweets": tweets}})


def make_twython_request(twitter_api_func, *args, **kw):

    error_count = 0
    while True:
        try:
            return twitter_api_func(*args, **kw)
        except TwythonRateLimitError:
            print >> sys.stderr, 'Encountered 429 Error (Rate Limit
Exceeded) with application-only authentication'
            return 429
        except TwythonAuthError:
            print >> sys.stderr, 'Encountered 401 Error (Not
Authorized)'
            return None
        except TwythonError, e:
            if error_count < 10:
                print >> sys.stderr, 'Encountered unexpected Error.
Retrying in 10 seconds' #% (str(msg))
                #print str(msg)
                print str(e)
                time.sleep(10)
                error_count += 1
            else:
                print >> sys.stderr, 'Too many retries. Aborting...'
                raise


def make_twitter_request(twitter_api_func, max_errors=30, *args,
**kw):

    def handle_twitter_http_error(e, wait_period=2,
sleep_when_rate_limited=False):

        if wait_period > 3600: # Seconds
            print >> sys.stderr, 'Too many retries. Quitting.'
            raise e

        # See https://dev.twitter.com/docs/error-codes-responses for
common codes

        if e.e.code == 401:
            print >> sys.stderr, 'Encountered 401 Error (Not
Authorized)'
            return None
        elif e.e.code == 404:
            print >> sys.stderr, 'Encountered 404 Error (Not Found)'
            return None
        elif e.e.code == 429:
            print >> sys.stderr, 'Encountered 429 Error (Rate Limit
Exceeded) with user authentication'
            if sleep_when_rate_limited:
                print >> sys.stderr, "Retrying in 15
minutes...ZzZ..."
                sys.stderr.flush()
                time.sleep(60*15 + 5)
```

```
                print >> sys.stderr, '...ZzZ...Awake now and trying
again.'
                return 2
            else:
                return 429 # Caller must handle the rate limiting
issue
        elif e.e.code in (500, 502, 503, 504):
            print >> sys.stderr, 'Encountered %i Error. Retrying in
%i seconds' % \
                (e.e.code, wait_period)
            time.sleep(wait_period)
            wait_period *= 1.5
            return wait_period
        else:
            raise e

    # End of nested helper function

    wait_period = 2
    error_count = 0

    while True:
        try:
            return twitter_api_func(*args, **kw)
        except twitter.api.TwitterHTTPError, e:
            error_count = 0
            wait_period = handle_twitter_http_error(e, wait_period)
            if wait_period is None:
                return
            if wait_period == 429:
                return 429
        except URLError, e:
            error_count += 1
            time.sleep(wait_period)
            wait_period *= 1.5
            print >> sys.stderr, "URLError encountered. Continuing."
            if error_count > max_errors:
                print >> sys.stderr, "Too many consecutive
errors...bailing out."
                raise
        except BadStatusLine, e:
            error_count += 1
            time.sleep(wait_period)
            wait_period *= 1.5
            print >> sys.stderr, "BadStatusLine encountered.
Continuing."
            if error_count > max_errors:
                print >> sys.stderr, "Too many consecutive
errors...bailing out."
                raise


use_twitter_api = False
window_started_at = datetime.now()

def harvest_user_timeline(twitter_api, twython_api, screen_name=None,
user_id=None, max_results=1000):

    assert (screen_name != None) != (user_id != None), \
```

```
    "Must have screen_name or user_id, but not both"

    kw = {  # Keyword args for the Twitter API call
        'count': 200,
        'trim_user': 'true',
        'include_rts' : 'true',
        'since_id' : 1
        }

    if screen_name:
        kw['screen_name'] = screen_name
    else:
        kw['user_id'] = user_id

    max_pages = 16
    results = []


    global window_started_at
    global use_twitter_api

    while True:
        if use_twitter_api:
            tweets =
make_twitter_request(twitter_api.statuses.user_timeline, **kw)
            if tweets == 429:
                use_twitter_api = False
                print "Switching to application-only
authentication..."
            else:
                if tweets is None: # 401 (Not Authorized) - Need to
bail out on loop entry
                    tweets = []
                results += tweets

        else:
            tweets =
make_twython_request(twython_api.get_user_timeline, **kw)
            if tweets == 429:
                if (datetime.now()-window_started_at).seconds > 905:
                    print "Switching to user authentication..."
                else:
                    print >> sys.stderr, "Retrying in " + str((900-
((datetime.now()-window_started_at).seconds))/60) + " minutes"
                    sys.stderr.flush()
                    time.sleep(60*15-(datetime.now()-
window_started_at).seconds + 5)
                    print >> sys.stderr, '...ZzZ...Awake now and
trying again.'

                use_twitter_api = True
                window_started_at = datetime.now()
                continue

            else:
                if tweets is None: # 401 (Not Authorized) - Need to
bail out on loop entry
                    tweets = []
                results += tweets
```

```
        break


    if tweets is None: # 401 (Not Authorized) - Need to bail out on
loop entry
        tweets = []

    results += tweets

    print >> sys.stderr, 'Fetched %i tweets' % len(tweets)

    page_num = 1

    date_limit = datetime.strptime("Wed Apr 16 2014", '%a %b %d %Y')


    if tweets != []:
        date_str =
tweets[len(tweets)-1]['created_at'][:tweets[len(tweets)-1]['created_a
t'].rfind(':')-5]
        date_str += tweets[len(tweets)-1]['created_at'][-4:]
        tweet_created_at = datetime.strptime(date_str, '%a %b %d %Y')

        # Return if oldest tweet has been created more than 90 days
ago
        if str(date_limit - tweet_created_at).find("day") != -1:
            if int(str(date_limit - tweet_created_at)[:str(date_limit
- tweet_created_at).find(' ')]) > 90:
                return results[:max_results]

    if max_results == kw['count']:
        page_num = max_pages # Prevent loop entry

    while page_num < max_pages and len(tweets) > 0 and len(results)
< max_results:

        # Necessary for traversing the timeline
        kw['max_id'] = min([ tweet['id'] for tweet in tweets]) - 1

        while True:
            if use_twitter_api:
                tweets =
make_twitter_request(twitter_api.statuses.user_timeline, **kw)
                if tweets == 429:
                    use_twitter_api = False
                    print "Switching to application-only
authentication..."
                else:
                    results += tweets

            if not use_twitter_api:
                tweets =
make_twython_request(twython_api.get_user_timeline, **kw)
                if tweets == 429:
                    if (datetime.now()-window_started_at).seconds >
905:
                        print "Switching to user authentication..."
                    else:
                        print >> sys.stderr, "Retrying in " +
```

```
str((900-((datetime.now()-window_started_at).seconds))/60) + "
minutes"
                        sys.stderr.flush()
                        time.sleep(60*15-(datetime.now()-
window_started_at).seconds + 5)
                        print >> sys.stderr, '...ZzZ...Awake now and
trying again.'

                    use_twitter_api = True
                    window_started_at = datetime.now()
                    continue
                else:
                    results += tweets
            break

        print >> sys.stderr, 'Fetched %i tweets' % (len(tweets),)

        page_num += 1

        # Break if oldest tweet has been created more than 90 days
ago

        if tweets != []:
            date_str =
tweets[len(tweets)-1]['created_at'][:tweets[len(tweets)-1]['created_a
t'].rfind(':')-5]
            date_str += tweets[len(tweets)-1]['created_at'][-4:]
            tweet_created_at = datetime.strptime(date_str, '%a %b %d
%Y')

            if str(date_limit - tweet_created_at).find("day") != -1:
                if int(str(date_limit -
tweet_created_at)[:str(date_limit - tweet_created_at).find(' ')]) >
90:
                    break

    print >> sys.stderr, 'Done fetching tweets'

    return results[:max_results]



twitter_api = twitter_auth()
twython_api = twython_auth()

sample = load_from_mongo("users", "degree_1", return_cursor=True)

followers_ids = set()

# Get followers of users in the sample
for x in sample:
    if 'followers_ids' in x:
        followers_ids |= set(x['followers_ids'])

criteria = {"tweets": {'$exists': False}, 'followers_count': {'$mod':
[4,0]}}

while True:
```

```
    users = load_from_mongo("users", "degree_2", return_cursor=True,
criteria=criteria)

    print "Created new cursor"

    nUsersUpdated = 0
    try:
        for x in users:
            if 'tweets' not in x and x['id'] in followers_ids:
                tweets = harvest_user_timeline(twitter_api,
twython_api, user_id=x['id'], max_results=3200)
                update_tweets(tweets, 'users', 'degree_2', x['_id'])
                print "Updated user " + str(x['id'])
                nUsersUpdated += 1

    except pymongo.errors.OperationFailure, e:
        msg = e.message
        if not ( msg.startswith( "cursor id" ) and msg.endswith( "not
valid at server" ) ):
            raise


    if nUsersUpdated == 0: # Nothing left to update
        break
```