Author(s): Nelimarkka, Matti & Karikoski, Juuso

Title: Categorizing and measuring social ties

Year: 2012

Version: Final published version

Please cite the original version:

Nelimarkka, Matti & Karikoski, Juuso. 2012. Categorizing and measuring social ties. RC33 Eighth International Conference on Social Science Methodology.

Categorizing and measuring social ties

Matti Nelimarkka
Helsinki Institute for Information Technology
Aalto University
matti.nelimarkka@hiit.fi

Juuso Karikoski
Department of Communications and Networking
Aalto University
juuso.karikoski@aalto.fi

Introduction

Social network analysis (SNA) has increased significantly recently, especially thanks to online social networking services and the data that are being collected from them. These services and the overall increase of mediated communication are changing the way we communicate and keep in touch with other persons (Licoppe & Smoreda, 2005).

As said, the increased significance of online social networks has increased both the academic and the business interest to understand these new methods of communication. This may be due to easier access to social networking data, as the online environment and databases do not require significant surveying or observing people – as previous methods usually did. The easiness of systematic data gathering allows research to examine larger samples and frees resources to detailed examination. Secondly, with information collected via online services or other means of data logging and mining, it is easy to argue that the reliability has increased: human errors and other classical problems of data collection are not present in the digital environment. Naturally, data collection in digital environment may bring up its own reliability issues, for instance problems due to intended and unintended use of services.

However, social ties are subjective experiences (e.g., Adams, 2010), and the SNA scholars acknowledge this. For example, Eagle et al. (2010) emphasize the meaning to understand what a tie represents and what constitutes a tie. Ackland's (2009) observation discusses the difference of Facebook-friendship and a classical understanding of friendship in the 'real world'. This means that the interpretation of the Facebook friendship remains open. Therefore, the validity of a social tie is unclear: not all researchers focus on the interpretation of the tie.

This paper examines the concept of social ties in detail. Based on the existing literature, a two dimensional model of social ties is presented and discussed. The first dimension is the publicity of the social tie to others. The second dimension measures how active the tie is, or differently: if the persons communicate actively or if there only is a low level of interactivity. We accept the fact that the model is preliminary, however, we see it as a way to approach the questions on research validity. In this context good validity means that the research measures the correct phenomena. Therefore, the first contribution of this work is to discuss what phenomena are really studied in social network analysis.

The second contribution made in this paper focuses on the practice of using several sources of data to measure a tie. Again, when the number of indicators increases we assume that the validity of the results increases at the same time. The main argument for this is that studies have shown that different media channels are used for different purposes. Therefore different channels reveal different social structures and thus the social networks also differ. Using several indicators provides an additional insight into the social structure and captures more of the ‘reality.’ Using several indicators is not a novel approach, but still in minority in the field. In this work, we discuss the previous practices and also present some views and potential solutions to the problems experienced with extensive data logging.

A categorization of social ties

As observed above, our aim in this work is to present a model that helps to understand the nature of different social ties and allows researchers to discuss the domain and validity of their observations. It has been observed that data collected from different kinds of communication channels lead to different kind of detected social networks. Therefore, a researcher studying social networks must understand what kind of social network is studied and what kind of arguments can be presented soundly from that network.

We claim that social networks differ in functional level. Van Cleemput (2010) observed that among a sample of Belgian adolescence population, different communication channels and intensity of communication indicate different levels of friendship. Her major observation is that the face-to-face communication is a key element in building friendship, but mediated communication, such as text messaging (SMS) and instant messaging (IM), is used when two persons consider themselves to be very good friends. Not surprisingly, her results also indicate that the number of different channels used is closely related to whether people consider being friends or good friends.

Similarly, Kim et al. (2007) demonstrate the difference in the media choices in the Republic of Korea. They observed that for different tie categories (e.g., close family, work colleagues, etc.) one uses different communication channels. For instance, telephone communication is used with closest ties, like spouse and children and e-mail is used with weaker ties. In this study, face-to-face communication was seen more common than mediated communication among all social relationship categories. Kim et al. (2007) also observe that different social groups, such as students and office workers, choose different communication channels. So, whereas van Cleemput (2010) only focused on one social group, Kim et al.'s (2007) study indicates that different social groups also use different communication channels.

Finally, our study (Karikoski & Nelimarkka, 2011) combined social networks collected from the mobile phone and online service for a campus. We observed that different communication channels presented us different social networks. Or, said differently: by examining any of the channel and the network formed by that channel, we would have a different view. Only by combining these networks were we able to examine a broader view of the social structure, even while it still was limited. This will be discussed in detail in the next chapter.

In this chapter we have discussed differences on social networks. We have detected that studies show that people use different communication channels for their social networks. Based on this evidence, we argue that researchers should focus more on what is the nature of the network studied. Different communication channels may represent different kinds of social ties. To help researchers identify what kind of social tie their data represents, we overlay a model of social networks. This model is based on existing literature, but combines these observations in a novel way. The model currently presents two different attributes, or dimensions, of social ties. These dimensions allow us to see details of the social ties and they are the following:

1. Whether the display of the ties is *public* or *private*. In the first case, the tie between two persons is visible to others than just the two persons, for example, the others connected to these users (often called friends) or everyone accessing their profile pages. In the latter case, the tie is visible only to the two persons connected, if the nature of the tie is reciprocal, or to one of the persons connected, if the tie is non-

reciprocal. Note, that the publicity of a tie is not only an attribute of the technology-mediated ties: for example, face-to-face meetings may take place in a public venue or only in a private venue.

2. Whether the tie is *active* or *passiv*e. This dimension measures the level of communication between the two persons. Of course, determining this level is far from trivial, but an easy dichotomy is to detect if any kind of communication takes place between the two persons. When applied, a more detailed measurement tool should be used to detect different levels of activity and passivity.

To illustrate the model, these dimensions are displayed in Table 1. Based on these two dimensions, we can distinct four types of social relations: *social*, *personal*, *nominal* and *latent*.

Table 1. Two dimensions of social ties with examples

		The tie is	
		public	private
The communication is	active	**social relation** commenting on the Wall in Facebook, @responses in Twitter	**personal relation** instant messaging (IM), telephone calls, text messages (SMS)
	passive	**nominal relation** friend lists in an online service	**latent relation** address book

The *social relations* are visible to all people or a subset of the community, they also include active communication between the persons engaged in this kind of tie. Examples of this kind of communication include commenting on others' walls in Facebook, @-responses in Twitter or discussion in an IRC channel.

The *personal relations* are active, but only visible to those taking part in the communication. For instance, email messages between persons, telephone calls and text messages, and IM between the two persons are examples of this kind of communication.

The *nominal relations* are public and visible to other persons, but unlike in the social relations there is no active communication. For example, accepting one's friendship request in an online social network, if the connection is shown in the system, is a good example of a nominal relation.

Compared to the nominal relation the *latent relation* is characterized with the privacy of the tie. One's contact book on the mobile phone is an example of this: the contact book is considered private, but does not necessarily imply active communication between the two persons.

In this work we have used social tie or tie with the meaning that certain link exists between these persons, such as friendship in social networking site or communication using telephones. Now we use term relation to classify the tie: all of the relations discussed above are also social ties, but as suggested, they are different. Therefore, we could discuss of a social tie with the type of social relation, or ties based on social and nominal relations.

Also, as noted above, this model is developed based on both our prior and on-going work and review of existing literature on this topic. Both of the two dimensions have been documented in the existing literature (e.g., Donath & boyd, 2004; Donath, 2007; Huberman et al., 2009; Golder et al., 2007) and these dimensions therefore have theoretical soundness. We shall take a detailed look into this in the following two sections.

Public-private –dimension

The publicity of the social tie may affect how both participants perceive the tie. Several online networks allow public or semi-public display of one's connections, whereas other methods of communication, such as telephony and e-mails are not disclosed in public. We define private ties as the ties that can only be detected by the two persons forming the tie and public ties as those ties that can be detected by an external observer.

Scholars have been interested in the impact of publicity of the online friendships. Donath & boyd (2004) and Donath (2007) discuss that the public nature of the friendship may be used to, e.g., signal one's social status in the group. According to them, the friendship information provides additional cues or signals that allow others to estimate one's reliability. This means in practice that friendship information can be seen as part of self-presentation and in fact this may lead to a biased friendship acceptance and selection strategy. Also, publicity has been shown to impact one's decisions in consumption (e.g., Ratner & Kahn, 2002; Silfverberg et al., 2011) as it is again related to self-presentation.

Naturally, we accept that the venue of potential publicity matters a lot. Most likely the career orientated LinkedIn may present a different network than the more casual Facebook or the interest focused networks. Therefore, when focusing in the public dimension, understanding the venue of publicity is an important factor. Therefore, we argue that the potential differences in the public online networks highlight how the network is used for self-presentation, as explained above. Furthermore, we see that this supports our claim that the publicity of the social tie matters in the analysis, but at the same time accept that there may be other factors affecting this choice. As we will discuss in the next chapter, using several indicators for a tie is a key to understand the social phenomena creating a social tie.

As shown above, the previous research indicates that the publicity of a tie may serve a specific social function, such as serve as an apparatus for self-presentation. Also, we discussed how publicity changes behaviour. Based on these two notes, we argue that public social ties and private social ties are perceived in a different way and therefore need to be separated in the analysis. We do accept that there may be other factors also affecting the forming of social ties, such as the venue of publicity, i.e., not all public social networks are similar to each other.

Active-passive –dimension

In the previous section we discussed social ties and their publicity. This dimension focuses on the other aspect, the intensity of the social tie. Especially in online social network analysis, a variety of studies focus in the network structure and behaviour. One common observation in those studies is that only a part of the social ties are engaged in active communication.

Scholars like Huberman et al. (2009) with Twitter, Golder et al. (2007) with Facebook, Zinoview & Duong (2009) with Odnoklassniki.Ru, for instance, have reported these kinds of observations. Huberman et al. (2009) observed that in Twitter, reciprocal communication takes place with less than 20 % of one's published list of friends. They conclude that a social connection does not lead to social engagement. Golder et al. (2007) suggest based on their data, that being a friend is required, but not a sufficient indicator for friendship and later propose that using active communication may provide better insight into the 'real' network. However, in our view, these data should be interpreted in a way, that there exists different social networks within one service, and depending on the research questions, all of them may provide answers for the researcher.

To continue the study of friendship as a concept in online domain, Zinoview & Duong (2009) observe, that in the Russian social networking site Odnoklassniki.Ru, the users labelled only 25 % of the online friends as friends; others were good acquaintances or even just random people. All of these studies support the common observation, that the engaged social network is smaller than the known social network.

In classical social network analysis, terms such as strong tie and weak tie are used to reference similar phenomena (e.g., Granovetter, 1973). Strong ties are characterized with active communication, whereas in weak ties less communication takes place. However, the concepts of strong and weak ties also include elements that are not measurable via network data, such as emotional intensity, intimacy, closeness or reciprocity (Granovetter, 1973; Krackhardt, 1992; Marsden & Campbell, 1984; Gilbert & Karahalios, 2009). Therefore, even while this axis is similar to the strong and weak ties, there exists a nuance difference with it. We do claim that strong and weak ties are linked to this axis, but as Marsden & Campbell (1984) discuss, the frequency and duration of communication are bad estimators for the tie strength. Active communication might lead to strong tie, but it does not imply it.

In this section, we have reviewed empirical results that indicate that not all social ties in the online environments are similar; rather there exist different types of ties. Certain scholars have started to discuss with terms like ‘true friendship’ (Zinoview & Duong, 2009) focused in other aspects to classify the ties. One such an approach could be the classification to the strong and weak ties, but they do include several aspects, not all that can be collected from the social network tie data.

Therefore, we have suggested a division between active ties and passive ties that can be measured from the number and length of interactions two persons have, for example. This model also extends the view to a continuum, i.e., this classification might not be a dichotomy even while it is drawn as one. This activity can be considered as the weight of the social tie, or as some other attribute related to the ties.

Multiple and dynamic social network datasets

This chapter discusses the usage of multiple and dynamic datasets when measuring social ties. As argued in the previous chapter, researchers should first confirm that the data measure the phenomenon desired and that the right kind of social network is analysed, i.e., that the data should be valid. To get a more detailed view of the social structures behind the group of people under study, we argue that multiple social network datasets should be used in the analysis.

The main aim in our suggestion for using several channels is the fact that people use several communication channels (e.g., Boase, 2008) and different communication media for different purposes depending, for example, on the strength of the tie (Granovetter, 1973) and context of the person (Karikoski & Soikkeli, 2011). A lot of analysis has been done in the difference of communication channels. For instance, Szabo & Barabasi (2006) analyzed a large dataset of mobile phone service usage and concluded that email is used consistently across communities, while IM usage is separated in communities. Thus one could argue that IM is used more for maintaining strong ties with other IM users, while email can also be used to connect to new and weaker ties (Kim et al., 2007). Moreover, Haythornthwaite (2001) and van Cleemput (2010) have observed that the stronger the tie between people is, the more media they use to communicate to one another. Consequently, when measuring strong social ties, using multiple datasets in social network analysis becomes even more critical. Karikoski & Soikkeli (2011) have studied the effect of use context on mobile communication service usage. They observe that mobile communication services are used differently depending on the mobile phone use context and conclude that use context is one of the key factors affecting the selection of mobile communication media. In addition to the issues discussed above, also the user characteristics affect how relationships are managed - young students manage a more general configuration of relationships through multiple media, while workers seem to manage different sets of relationships through each medium (Kim et al., 2007).

Previous work with multiple and dynamic datasets

Naturally, we are not the first researchers arguing for the use of multiple social network datasets in social network analysis. Already more than a decade ago McPherson et al. (2001) concluded that the priority for future social network researchers should be to gather data on multiple social ties. Moreover, they claim that sociology offers a solid base of empirical knowledge and theory to support such a challenging objective. Banford et al. (2010) have studied how tie strength can be estimated in communication networks. They have studied different metrics of calls and proximity as proxies for tie strength, and although the proxies are variably successful in estimating the tie strengths by themselves, the accuracy of the estimation could be increased by combining different proxies. In our previous research (Karikoski & Nelimarkka, 2011) we have studied social networks of a group of people based on online social media services and mobile phone communication. We have analysed data collected from the usage logs of the social media services and directly from the handsets of the persons and shown that the social networks are different depending on the service used. This is why researchers should focus more on understanding the type of relation (social, personal, nominal and latent in this work) they study and then understand which communication channels reflect this kind of a relation.

In addition to using multiple datasets, one should also strive for dynamic, not static, social network data as McPherson et al. (2001) point out. Networks evolve over time and longitudinal (or panel) studies can be used to capture such phenomena. Large-scale datasets have become available in the recent years, which enable studying the structure and dynamics of real human communication networks.

This field of research has been named computational social science (Lazer et al., 2009). However, these datasets are usually limited to a single communication channel, like mobile voice calls (Onnela et al., 2007), Facebook (Backstrom et al., 2012) or email (Eckmann et al., 2004), for instance. Thus, to make these dynamic large-scale datasets even richer, the researchers in the area of computational social science should strive to collect and analyze data from multiple communication channels. Furthermore, social ties can be inferred from data that do not necessarily depict person-to-person communication as such, but rather the proximity of persons. For example, Nefedov (2011) has studied communities in multiple layers by analyzing phone call logs and Bluetooth connections. He has observed that users' profiles may significantly vary across layers. In addition to Bluetooth, proximity has also been inferred with WLAN (Wireless Local Area Network) radios (Krumm & Hinckley, 2004), RFID (Radio Frequency Identification) tags (Cattuto et al., 2010) and mobile network cell IDs (Calabrese et al, 2011), for example. Therefore, there are several techniques to collect data, each of them with its own characteristics and, as we suggest, a specific relation that is revealed by analysing them.

Based on Eagle et al. (2009), 95% of the self-reported friendships can be inferred from objective measurements of physical proximity and calling patterns, demonstrating the potential of technology to study the social structures, ties and relations. However, Adams (2010) points out a fundamental difference between subjective self-perceived data and objective behavioral data - he claims that if the goal is to study social ties (or friendship networks), there should be an aim to use subjective data, whereas if the goal is to study social or physical contagion then using the objective behavioral data should be the aim. However, Eagle et al. (2010) point out that significant social influence can travel also via unperceived ties and thus using an appropriate combination of self-report data and behavioral data can reveal information about the underlying dynamics of the group of people under study. Acquiring dynamic data on multiple layers is a formidable task, however, and next the challenges and opportunities related to the collection of such data are discussed in more detail.

Challenges related to bias, identity and research disciplines

By studying existing research and based on our own empirical research (Karikoski & Nelimarkka, 2011), we discuss the challenges and opportunities of collecting and analyzing multiple and dynamic social network datasets covering the same group of people. This discussion is important both from a sociological and practical perspective, especially now as large-scale human communication datasets are being collected and analyzed more and more.

First of all, the challenges in all longitudinal panel studies include challenges such as attrition bias, panel selection bias and conditioning effects as described by Lohse et al. (2000). Attrition bias refers to the loss of panelists (i.e., people participating in the data collection panel) over time, panel selection bias refers to the panel sample being different from the population to be studied and conditioning effects refer to the panel affecting the behaviour of participants. Wunsch et al. (2010) have also described sample selection and follow up biases, which mean that some individuals might be more reluctant to join a longitudinal study (sample selection bias) whereas some individuals might be more eager to drop out of a study (follow up bias). Moreover, the longer the panel study continues, the more changes in the individuals participating in the research and their context can be expected - same applies to the personnel conducting the study as well (Wunsch et al., 2010).

When combining longitudinal data, all these challenges need to be overcome. The challenge with attrition bias (or follow up bias) becomes even more salient when multiple datasets are collected from the same individuals - the individual needs only to drop out from one data collection effort to make the combined dataset broken. In our previous research (Karikoski & Nelimarkka, 2011) we have observed that this was indeed one of the biggest challenges which ultimately led the combined dataset to be fairly limited in terms of sample size. Conditioning effects is a challenge that we have discovered in our previous research as well - in our empirical studies with mobile phones, we have discovered that the participants in our handset-based measurements are biased towards early adopters of mobile phones and services (Karikoski, 2012). Thus, the external validity of the results can be questioned, unless the aim is to study early adopters specifically. There is also concern that installing a research application to the participants' mobile phones might affect how they use their phones.

In order to be able to combine multiple social network datasets, you need to have some sort of a common identifier for all individuals across all datasets. Unless you are collecting all the data yourself, there is a challenge in identifying individuals whose data are to be combined. Furthermore, combining large-scale datasets collected by a service provider such as Facebook is challenging because usually they do not disclose the data to third parties, and are even more unwilling to disclose any identifiable information which could enable combining multiple social network datasets. Some services such as Twitter do disclose some data via their APIs, but still no identifiable information can be acquired. As Lazer et al. (2009) point out, there need to be robust collaboration models and data sharing agreements between academia and corporations in the future in order for the large-scale analysis of human behavior to become more common and not restricted to privileged researchers whose research and results cannot be critiqued or replicated. The most natural source for large-scale social network data from multiple sources would be a mobile operator which offers also other services than just the basic services - voice calls, text messages and multimedia messages. The examples of this kind of research are few, but Szabo & Barabasi (2006), for instance, have studied several mobile communication services available to millions of customers of a mobile operator and concluded that the underlying social network indeed impacts the usage patterns of many services.

If we exclude the large-scale datasets and focus on data that can be collected in academia, we believe that handset-based measurements are one method that provides a good possibility to collect multiple social network datasets from a group of people via their mobile phones or smartphones. These measurements capture all the usage of the mobile phone, also the offline services and the communication services not offered by the operator, by installing a data collection application to the devices of the participants. Naturally the sample size is smaller in handset-based measurements than in large-scale datasets and there are also differences in mobile operating systems and devices that affect the type of data that can be collected. However, the advantage is that as the mobile phone can be seen as a hub of different communication channels for the user, it works as a good research platform for measuring multiple and dynamic social ties. It has to be noted, however, that although the mobile phone is the most used personal digital communication device in the world (with more than four billion individual subscribers), communication happens also outside the device. However, with the mobile phone you can also collect multiple types of proximity data in addition to the person-to-person communication data.

In order to combine objective behavioral data and subjective self-perceived data, for example, researchers from different disciplines (e.g., computer science and social science) need to collaborate. The differences between disciplines are multifaceted, however, which creates challenges for conducting cross-disciplinary research with multiple social network datasets. For instance, Ackland (2009) claims that differences between disciplines exist in three dimensions: theory, methods and availability of appropriate tools. Lazer et al. (2009) also state that the computational resources in social sciences are fewer and even the physical distance between departments of different disciplines tends to create institutional obstacles. To overcome these challenges, the paradigms in different disciplines need to become more receptive in the future, so that collaboration across disciplines can be developed.

Discussion

Social network analysis has increased significantly in the recent years. Researchers especially use social networking sites to analyse social ties, but also other methods, such as surveys and handset-based measurements (Raento, 2009; Karikoski, 2012) have been used. Less focus has been on the validity of the results: even while ties have been studied, the question on what a tie represents is still rather open. This work contributes to this question both in a theoretical level and in by discussing methods of studying this.

We have proposed a method to understand the function of the social ties. Our solution is a categorization of social ties that makes a distinction based on the tie's publicity and activity. The social tie can be visible to everyone or a significant subgroup - public - or only known by the person - private. Based on the existing literature (e.g., Donath & boyd, 2004; Donath, 2007) the public social ties are linked to self-presentation and signalling one's social status to other people. The activity of a social tie also varies: the communication between the two persons in a tie may take place daily or never. Again, the previous research (e.g., Huberman et al., 2009; Golder et al., 2008) has shown that in online domain tie's activity levels differ. Using these two dimensions we categorize four different tie types, or relations: social, personal, nominal and latent.

Above we have argued that ties vary based on the publicity and activeness. We see that this discussion contributes to the discussion on validity. At the same time, we also see that the reliability of social network analysis should be discussed. We argue that to understand social phenomena, several proxies should be used to estimate the social tie. One method to achieve this is to use multiple sources of data. For example, using data from several online services or from an online service and a mobile phone will enrich the data studied.

Using several sources of data is important, as the previous research (e.g., Haythornthwaite, 2001; van Cleemput, 2010) indicates that people choose different kind of services for different functions. Therefore, only by combining the data one can approximate the actual social structure and the social ties. Even while the empirical research still mostly focuses into the study of a single dataset, this idea is not novel (e.g., McPherson et al., 2001). We acknowledge the challenges that rise from this practice, such as the need to have a common identifier among the services to combine the data and especially in longitudinal studies, the decrease of the sample size due to attrition bias or failure to connect users across the services. However, we suggest that the increased validity of the results justifies the effort.

In this work we have contributed to the discussion on validity in social network analysis. We have above proposed two alternative approaches to increase the validity. Firstly, we have examined social ties and developed a categorization of social relations represented by different kinds of ties. We argue that understanding these relations is significant for interpreting the results of any social network analysis. Secondly, we have suggested using several proxies, such as online services, handsets, different sensors on the handset and surveys, to capture the social network. We suggest that this method will enable more detailed and fruitful analysis of the social phenomena experienced by people.

The core argument, however, is that at this time, research in social network analysis has not emphasized core questions, such as those of validity and reliability, enough. We assume that raising these questions will help the field to develop further and later provide better answers to those applying the knowledge in practice in, e.g. service design.

Acknowledgments

We thank Kalevi Kilkki for discussion related to the categorization terms, and Airi Lampinen for commenting the manuscript. We also thank both the Future Internet Graduate School (FIGS) and the OtaSizzle research project for funding this research. The OtaSizzle research project is co-funded by the Aalto University's MIDE program and Helsinki University of

Technology TKK's 'Technology for Life' campaign donations from private companies and communities.

References

Ackland, R., 2009. Social network services as data sources and platforms for e-researching social networks. *Social Science Computer Review,* 27 (4), pp.481-492.

Adams, J., 2010. Distant friends, close strangers? Inferring friendships from behavior. In: *Proceedings of the National Academy of Sciences (PNAS)*, 107 (9), E29–E30.

Backstrom, L., Boldi, P., Rosa, M., Ugander, J. & Vigna, S., 2012. Four Degrees of Separation. *Arxiv preprint physics*: arXiv:1111.4570v3.

Banford, J., McDiarmid, A. & Irvine, J., 2010. Estimating the Strength of Ties in Communication Networks with a Small Number of Users. In: *The 6th International Conference on Wireless and Mobile Communications (ICWMC2010)*. Valencia, Spain, 20-25 September 2010.

Boase, J., 2008. Personal Networks and the Personal Communication System - Using multiple media to connect. *Information, Communication & Society*, 11 (4), pp.490-508.

Calabrese, F., Smoreda, Z., Blondel, V.D. & Ratti, C., 2011. Interplay between telecommunications and face-to-face interactions: a study using mobile phone data. *PLoS ONE*, 6 (7), e20814.

Cattuto, C., Van den Broeck, W., Barrat, A., Colizza, V., Pinton, J.-F. & Vespignani, A., 2010. Dynamics of Person-to-Person Interactions from Distributed RFID Sensor Networks. *PLoS ONE*, 5 (7), e11596.

Donath, J., 2007. Signals in Social Supernets. *Journal of Computer-Mediated Communication*, 13 (1), pp.231-251.

Donath, J. & oyd, D., 2004. Public displays of connection. *BT Technology Journal*, 22 (4), pp.71-82.

Eagle, N., Pentland, A. & Lazer, D., 2009. Inferring social network structure using mobile phone data. In: *Proceedings of the National Academy of Sciences (PNAS)*, 106 (36), pp.15274-15278.

Eagle, N., Clauset, A., Pentland, A. & Lazer, D., 2010. Reply to Adams: Multi-dimensional edge inference. In: *Proceedings of the National Academy of Sciences (PNAS)*, 107 (9), E31.

Eckmann, J.P., Moses, E. & Sergi, D., 2004. Entropy of dialogues creates coherent structure in e-mail traffic. In: *Proceedings of the National Academy of Sciences (PNAS)*, 101 (40), pp.14333-14337.

Golder, S.A., Wilkinson, D. & Huberman, B.A., 2007. Rhythms of social interaction: messaging within a massive online network. In: *The 3rd International Conference on Communities and Technologies (C&T2007)*. East Lansing, Michigan, USA, 28-30 June 2007.

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

Gilbert, E. & Karahalios, K., 2009. Predicting Tie Strength With Social Media. In: *The 27th International Conference on Human Factors in Computing Systems (CHI '09)*. Boston, USA, 4-9 April 2009.

Haythornthwaite, C., 2001. Exploring multiplexity: social network structures in a computer supported distance learning class. *The Information Society*, 17 (3), pp.211-226.

Huberman, B., Romero, D.M. & Wu, F., 2009. Social networks that matter: Twitter under microscope. *First Monday*, 14 (1-5).

Karikoski, J. & Nelimarkka, M., 2011. Measuring social relations with multiple data sets. *International Journal of Social Computing and Cyber-Physical Systems (IJSCCPS)*, 1 (1), pp.98-113.

Karikoski, J. & Soikkeli, T., 2011. Contextual usage patterns in smartphone communication services. *Personal and Ubiquitous Computing (Online First Collection)*.

Karikoski, J., 2012. Handset-based data collection process and participant attitudes. Accepted for publication at *International Journal of Handheld Computing Research (IJHCR)*.

Kim, H., Kim, G.J., Park, H.W. & Rice, R.E., 2007. Configurations of relationships in different media: FtF, email, instant messenger, mobile phone, and SMS. *Journal of Computer-Mediated Communication*, 12 (4), pp.1183-1207.

Krackhart, D., 1992. The Strength of Strong Ties: The Importance of Philos in Organizations. In: N. Nohria & R. Eccles (eds.). *Networks and Organizations: Structure, Form, and Action*. Boston, MA: Harvard Business School Press, pp.216-239.

Krumm, J. & Hinckley, K., 2004. The NearMe Wireless Proximity Server. In: *The Sixth International Conference on Ubiquitous Computing (UbiComp 2004)*. Nottingham, England, 7-10 September 2004.

Lazer, D., Pentland, A., Adamic, L., Aral, S., Barabási, A-L., Brewer, D., Christakis, N., Contractor, N., Fowler, J., Gutmann, M., Jebara, T., King, G., Macy, M., Roy, D. & Van Alstyne, M., 2009. Computational Social Science. *Science*, 323 (5915), pp.721-723.

Licoppe, C. & Smoreda, Z., 2005. Are social networks technologically embedded? How networks are changing today with changes in communication technology. *Social Networks*, 27 (4), pp.317-335.

Lohse, G.L., Bellman, S. & Johnson, E.J., 2000. Consumer Buying Behavior on the Internet: Findings from Panel Data. *Journal of Interactive Marketing*, 14 (1), pp.15-29.

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

Marsden, P.V. & Campbell, K.E., 1984. Measuring Tie Strength. *Social Forces*, 63 (2), pp.482-501.

Nefedov, N., 2011. Multiple-Membership Communities Detection in Mobile Networks. In: *The 1st International Conference on Web Intelligence, Mining and Semantics (WIMS'11)*. Sogndal, Norway, 25-27 May 2011.

Onnela, J.P., Saramaki, J., Hyvonen, J., Szabo, G., Lazer, D., Kaski, K., Kertesz, J. & Barabasi, A.L., 2007. Structure and tie strengths in mobile communication networks. Proceedings of the National Academy of Sciences (PNAS), 104 (18), pp.7332-7336.

Raento, M., Oulasvirta, A., & Eagle, N. (2009). Smartphones: An emerging tool for social scientists. Sociological Methods and Research, 37(2), pp.426‐454.

Ratner, R.K. & Kahn, B.E., 2002. The Impact of Private versus Public Consumption on Variety Seeking Behavior. *Journal of Consumer Research*, 29 (2), pp.246-257.

Silfverberg, S., Liikkanen, L.A. & Lampinen, A., 2011. "I'll press Play, but I won't listen": Profile Work in a Music-focused Social Network Service. In: *The 2011 ACM Conference on Computer Supported Cooperative Work (CSCW 2011)*. Hangzhou, China, March 19-23 2011.

Szabo, G. & Barabasi, A.L., 2006. Network effects in service usage. *Arxiv preprint physics: arXiv:physics/0611177v1*.

van Cleemput, K., 2010. I'll see you on IM, text or call you: a social network approach of adolescents' use of communication media. *Bulletin of Science, Technology & Society*, 30 (2), pp.75–85.

Wunsch, G., Russo, F. & Mouchart, M., 2010. Do We Necessarily Need Longitudinal Data to Infer Causal Relations? *Bulletin de Methodologie Sociologique*, 106, pp.5-18.

Zinoviev, D. & Duong V., 2009. Toward Understanding Friendship in Online Social Networks. *The International Journal of Technology, Knowledge and Society*, 5 (2), pp.1-8.