

---

## Measuring social relations with multiple datasets

---

### Juuso Karikoski\*

Department of Communications and Networking,  
Aalto University,  
P.O. Box 13000, 00076 Aalto, Finland  
E-mail: juuso.karikoski@aalto.fi  
\*Corresponding author

### Matti Nelimarkka

Helsinki Institute for Information Technology HIIT,  
Aalto University,  
P.O. Box 15400, 00076 Aalto, Finland  
E-mail: matti.nelimarkka@hiit.fi

**Abstract:** Because people have different levels of engagement with each other, measuring social relations is difficult. In this work, we propose a method of measuring social relations with multiple datasets and demonstrate the differences with empirical evidence. Our empirical findings demonstrate that people use different communication media channels differently. Therefore, we suggest that in order to understand social structures, one should use several kinds of data sources and not just depend on a single dataset. Our datasets include mobile phone data gathered with handset-based measurements and data from OtaSizzle online social media services. By means of social network analysis, we show that the online social media services have a different friendship network than the networks based on mobile phone communication. The mobile phone communication networks, however, have a very similar structure. These results are encouraging as previous research also indicates differences in the communication networks.

**Keywords:** social network analysis; social computing; data collection methods; social media services; user behaviour; privacy; social relations.

**Reference** to this paper should be made as follows: Karikoski, J. and Nelimarkka, M. (2011) 'Measuring social relations with multiple datasets', *Int. J. Social Computing and Cyber-Physical Systems*, Vol. 1, No. 1, pp.98–113.

**Biographical notes:** Juuso Karikoski works as a Researcher at the Department of Communications and Networking, Aalto University, Finland. His research interests include handset-based measurements, mobile social media and communication service usage and social network analysis. He received his MSc (Technology) from Helsinki University of Technology.

Matti Nelimarkka works in Helsinki Institute for Information Technology HIIT, Aalto University, Finland. His research interests include human behaviour in mobile and online services, mobile and online services design and the influence of technology on society. He received his BSocSc from the University of Helsinki. This research was partly done while working in Nokia Corporation.

## **1 Introduction**

Researchers have for a longer time pondered what online friendship means in practice. Huberman et al. (2009) showed that users communicate only with a part of their online contacts, indicating that friendship may not be linked to whom we call our friends in the online environment. Similarly, Golder et al. (2007) examined a different service, but concluded (p.5): “Because so many people are listed as friends as compared to those who engage in active messaging, for research purposes, being a friend in Facebook can be considered a necessary but not a sufficient condition for being a friend in the conventional sense”. These two studies demonstrate that the online friendship network is different from the offline friendship network. Donath and Boyd (2004), Fono and Raynes-Goldie (2007) and Donath (2007) explain that this is due to the public nature of social networks in online environments. Thus, the social network of a user may be used to signal, e.g., social status or trustworthiness in the system, not to present the user’s actual friendships. We also hypothesise, that due to fact that the effort put in the online friendship is minimal, the online social networks are easier to maintain than conventional friendship.

For academic research, this raises questions, such as what can we explore using the online friendship network. Researchers have an easy access to online networks, but how well do they represent the conventional friendship? Are they totally separate from the conventional friendships, therefore creating a new social structure, or are conventional friendships a subset of online friendships? For those more practice-orientated, understanding and measuring social relations is also vital. This information can be used, e.g., to tell apart meaningful contacts from all contacts, thus allowing a smooth and seamless user experience.

Both Huberman et al. (2009) and Golder et al. (2007) used direct messaging between users as the disjunctive factor to separate more conventional friendships from the unconventional, i.e., online, friendships. However, in our view they did not observe that we communicate using multiple tools and platforms, using whichever is the most convenient for that time. In a similar way, this behaviour has been observed in mobile telephony: people only contact frequently a small subset of their contact book (Jung et al., 2008).

van Cleemput (2010) has investigated this communication pattern on multiple different kinds of services, such as instant messages, phone calls and face-to-face communication. Her major observation is that these kinds of networks are different in their nature. However, even while she has a large group (larger than this study), people do make mistakes in their estimates. This has been examined by Eagle et al. (2009) and Eagle and Pentland (2006), who, by using mobile phones as sensors, show that our estimates on with whom we spend our time are not exact enough. Based on these and other experiments, Raento et al. (2009) conclude, that mobile phones can be used by social scientists to examine and understand the human behaviour in more detail than what is possible with traditional observations or questionnaires.

In this study we examine the use of different datasets to measure social relations. By ‘social relation’ we mean different kinds of closeness and interaction: physical presence, online communication and presence, or direct communication. As we have argued above, we should examine something more than just the network from a social networking system to understand social relations in detail. We have also indicated that the use of more than one source of data could provide the needed insights, and if nothing else, help

observe the differences in communication patterns. In this study, we focus on mobile phone communications data (SMS and phone calls) gathered with handset-based measurements and a network from experimental social networking services from a project called OtaSizzle. OtaSizzle is a service platform to host different kinds of social media applications, both for mobile phones and desktop-based approaches (Mäntylä et al., 2009). The OtaSizzle services have different kinds of personal and application data stored in them, for example, the data on friendship connections.

In this work, we argue for the use of several data sources for understanding and, in the end, measuring social relations. To provide a valid presentation on this, we first present and discuss the research platform and privacy issues, after which we present the first set of empirical results using the three sources of data: phone calls, text messages (SMS) and the online service network (OtaSizzle). Naturally, our study has limiting factors, such as sample size and cultural differences, but based on the empirical data and existing research we can discuss the differences we have observed in these networks, and the implications of these for a wider content.

## **2 Research platform**

This chapter presents the research platform used for measuring social relations. The platform is implemented in the OtaSizzle project, which will be introduced shortly. The data sources will also be presented together with privacy and ethical considerations regarding the collection of behavioural data about users.

### *2.1 OtaSizzle*

The research platform is implemented in OtaSizzle, which is a project focusing on social media services and their creation and use among the students and staff of the Aalto University. Aalto University forms a tight knit community of three previously separate Universities, namely Helsinki University of Technology, Helsinki School of Economics and University of Art and Design Helsinki. The aim of OtaSizzle is firstly to provide services for the student community and engage in service design with them. Secondly the project aims at allowing experimentation and examination of different designs and providing data for research purposes.

There are several different OtaSizzle social media services developed and available and they all share the social network, i.e., users who are friends in one service are that also in the other ones. Up to date, the most successful service has been Kassi (Suhonen et al., 2009), which is a social web service for exchanging goods and services. In November 2010, the OtaSizzle services had more than 2,500 registered users in total.

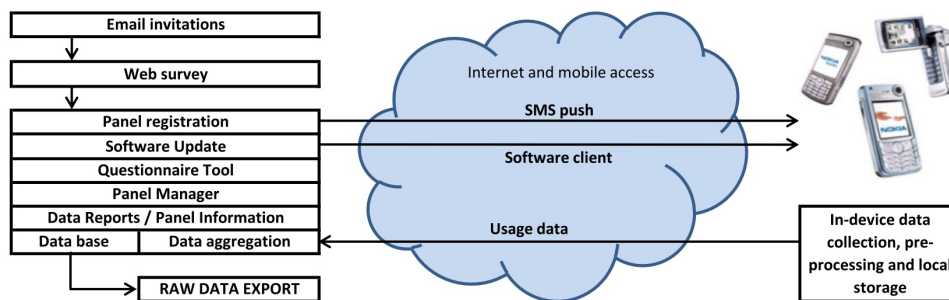
### *2.2 Handset-based data*

Because of technical limitations, the attempts of acquiring subscriber-level usage data using handsets has previously been challenging. Because of the smartphone, however, and its ability to install third party software on it, the acquisition of near-perfect mobile subscriber data is now possible. The smartphone is a personal device, always on and always carried by the subscribers. Raento et al. (2009) have also identified the smartphone as a useful tool for social sciences, because of its unique capabilities in

collecting behavioural data about the user. There are different tools that have been used in collecting mobile phone usage data from smartphone users. These tools include the ContextPhone (Raento et al., 2005), SocioXensor (ter Hofte, 2007), MyExperience (Froehlich et al., 2007) and the Smartphone360 platform (Nokia, 2007), for example. Eagle et al. (2009) have recently even used mobile phone data in inferring social network structure. Thus, the smartphone is used in this research as an essential part of analysing the social relations among the users.

The handset-based data collection in OtaSizzle is implemented with MobiTrack, which is a mobile audience measurement platform utilising smartphone-based software to measure real-life user behaviour, usage of devices and mobile services, and various technical parameters. The data types collected with the software include application usage, application installations, processes, battery levels and charging, Bluetooth and WiFi entries, phone calls, SMSs, MMSs, URL entries, network sessions and uploads. The platform also provides a possibility to collect user feedback through contextual surveys. The software is available for several operating systems (Symbian, Google Android, Windows Mobile and BlackBerry), but only the Symbian application of the MobiTrack software was used in this research.

**Figure 1** Data acquisition process (see online version for colours)



The data acquisition process is illustrated in Figure 1, and it is very similar to the one used by Verkasalo (2009). The users are recruited with e-mail invitations. In the OtaSizzle project, the prospective participants are all students and staff of the Aalto University, and thus all the e-mails were sent to members of these two groups. To increase the number of participants the invitation also includes some sort of compensation of participation for the users, such as vouchers and device lotteries. This kind of a direct compensation should, however, be avoided as it might have an effect on the usage patterns and produce biased results as discussed by Verkasalo (2009). In the invitations, the user is directed to fill in a short survey about contact details and basic demographics. The users have to opt-in, so after they have agreed to join the study, an SMS is sent to the users where a link to the download page of the software client is provided. If the user successfully installs the software, he or she becomes an active data producer. The data produced by a user in a week is less than half megabytes of size, so the users need to have some kind of a data plan in their subscriptions. The data are pre-processed, stored locally in the device and sent encrypted and compressed daily to the MobiTrack servers. In case of international roaming, the data will only be sent to the servers once in three days. The user can also disable the application during international roaming and send the data when returning to the home land. The data are aggregated

from all the participants and the raw data are then exported to the OtaSizzle researchers by request. If the researchers also wish to implement pop-up surveys with the software, there is a questionnaire tool for them as well.

The handset-based data used in this study were gathered during autumn 2009. In total 113 users produced data that was logged by the software. Inside this group of users we identified a social network of 20 users and this group was selected as the target for social network analysis. The users had to have at least one week of activity (i.e., data traces from seven days) in the handset-based measurements so that they would be included in the social network analysis. On average these users had data traces from 59 days. Only the phone call and SMS logs collected with MobiTrack were used in the analysis and the respective identified social networks are illustrated in the results chapter.

### *2.3 OtaSizzle service data*

Compared to the handset-based measurements described above, data collection from the OtaSizzle services is more straightforward, as the data are already gathered in the service database and logs. In the OtaSizzle-project, a main interface (ASI, Aalto Social Interface) to user data has been developed, and all OtaSizzle services use this interface to access identity data, information on groups and channels – and the friendship network of users. This will allow several services to use the same data, thus enriching the social experience and allowing seamless experience while chaining from one application to another, as the core data do not need to be re-inputted.

Currently, there exist three core applications using ASI: Kassi is a service for exchanging favours and items, Ossi is a mobile group-based social networking service and NordSecMob is a service for exchange students to get familiar with Finnish customs and practices. In all these services, the social network can be developed. Friends are not a core feature of use, however, but rather an additional function. As there exist different kinds of services, the meaning of a ‘friend’ in these online services is not easy to define. The friendship process requires mutual agreement, however, in a similar way that Facebook requires one to accept a friendship request. This means that the receiver of a friend request needs to acknowledge and approve a friendship request for the link to exist between the users.

These services have been marketed for the students of Aalto University, especially to those in the beginning of their studies via, e.g., festivals and IT education. The usage has been voluntary to all students, however, and there has not been active pushing of these services, e.g., via e-mail. During the phase of this study, March 2010, there were circa 1,500 users, and as stated above, in November 2010 the number of users was about 2,500. Naturally, only a part of these users use the services actively. One indication of the number of active users is, that during March 2010, the network inside OtaSizzle was 441, i.e., only less than a third of the users had one or more friends in the system. As we have indicated, we focus on 20 of them in more detail.

Of these 20 users, all have been in the service for at least three months, thus the differences in the network structure should not be caused by new users still adapting to the system. However, as we have indicated, the usage behaviour may be different between the users, i.e., some users may use these services actively whereas others may have logged in only once or twice. These differences are not part of our analysis.

#### 2.4 *Privacy and ethical considerations*

In general, the processing of private data is restricted by law. In case the person gives consent, however, then almost any processing is allowed. The consent has to be specific and informed and there has to be an acceptable purpose to process personal data. It is therefore essential what the end-user knows and understands about the processing of personal data (Kosta and Dumortier, 2008). In OtaSizzle, the consent is given when the users accept the research agreements, terms of use and privacy policies related to the data collection both from the services and the smartphones. Because the users explicitly give consent when participating in the handset-based measurements, the research ethics have a very important role in this research.

On international level there exists a broad consensus on the basic standards for fair information practice and the protection of citizen privacy. The basis for this is provided by the Organisation for Economic Co-operation and Development (OECD, 1980) guidelines on the protection of privacy and transborder flows of personal data. There are, however, substantial differences in how privacy policies are implemented in different countries. In the European Union, where this research is being conducted, the legislation essentially consists of Directive 95/46/EC on the protection of personal data and Directive 2002/58/EC on privacy and electronic communications. These directives together constitute a comprehensive legal rights approach to data protection and privacy. However, because of the rapid technological developments regarding social networking sites and cloud computing, for example, the European Commission (2010) has recently launched a review of the current legal framework to address the challenges incurred by these advances in technology. This will result in a proposed revised legislation in 2011, where the impacts of new technologies on personal data protection are addressed. In the USA there is no comparable overarching privacy law, however, and thus it is not always evident which law or regulation should be applied and which authority is concerned in a data collection effort (Kivi, 2009). Laws and regulation of the USA will not be discussed further in this study, however, since the research is conducted in the European Union.

Although the software used in the handset-based data collection resembles spyware or adware, the data collection procedures and the software are completely legal. The data are only collected for research purposes and SMS and MMS content, for example, are not recorded at all. Also mobile phone numbers are recorded as cryptographic hash values. The users have to opt-in to participate in the research and perform the application installation by themselves. Thus they are aware of the data collection and can decide themselves whether to take part in the data collection or not. The smartphone is a personal device and thus the data collected directly from the users' devices is highly sensitive to privacy issues. Only anonymised, aggregate level data are thus being analysed and unique identifiers are removed. Naturally only a very limited number of registered researchers working under non-disclosure agreements (NDA) can perform data analysis and any personal data relating to the participants will be destroyed at the end of the project (i.e., not later than December 31, 2012). The participants can quit the handset-based data collection whenever they want by deleting the application and can also request Aalto University to stop processing their personal data. Regarding the OtaSizzle online services, the data are collected for research purposes and the users need to agree on the research agreement to be able to use the services. The research team got access only to the friendship data from the server, so no

other personal data were accessed. Similarly to the handset-data, the data were handled only by a specific team of researchers, all working under NDAs. The research team did not use any personal data in the analysis but instead used numerical identifiers to identify the users.

Langheinrich (2001) has presented six principles for guiding ubiquitous computing system design and they are based on a set of fair information practices that are in use today, and can be used as guidelines for privacy design in the OtaSizzle project as well. The principles are notice, choice and consent, anonymity and pseudonymity, proximity and locality, adequate security, and access and recourse. *Notice* means that data collection should not happen unnoticed of the subject who is being monitored, *choice and consent* means that the data collectors need to receive an explicit consent from the data subjects to enable data collection and *anonymity and pseudonymity* means that the data collection should not pose a threat to the privacy of the individual. *Proximity and locality* means that data should only be collected whenever the user is present, *adequate security* means that the communications and storage methods used in the data collection need to be secure and *access and recourse* means that data should only be collected for a well-defined purpose and only relevant to that purpose, and the data should also only be kept as long as it is necessary for the purpose.

Regarding the handset-based data these principles are adequately fulfilled. Notice is fulfilled, because the data collection is not happening unnoticed of the subject that is being monitored as described above. Choice and consent is fulfilled with opt-in and the agreements presented above. Anonymity and pseudonymity is more difficult to define than the previous principles, because it is a subject for debate what types of data can be linked back to a person. As described earlier, all the unique identifiers are naturally removed from the data but since there are no universal standards regarding the anonymity of these kinds of behavioural data, the solutions tend to be ad hoc and there clearly is a need for standards related to the anonymisation of behavioural data as Eagle (2009) suggests. One possible answer to this problem is letting the users own their data and be able to access and remove their data from databases at any point in time as Pentland (2008) has suggested. It remains to be seen whether this model will become the standard in protecting behavioural data of users. The proximity and locality principle refers to a situation where data are only collected when the user is present. Because we are dealing mainly with usage data, this principle is more applicable to ubiquitous computing scenarios with sensor networks collecting data, for example. Adequate security refers mainly to secure communications and storage methods. With MobiTrack the data transfers are encrypted and the data are stored in local servers within the university to ensure adequate security. Access and recourse belongs to the realm of legal practice and purpose. As described above this principle is also fulfilled in the OtaSizzle handset-based data collection.

Regarding the OtaSizzle service data, the purpose of OtaSizzle as a research project is highlighted during registration, providing a specific research agreement that the user has to agree. Thus, users need to have agreed on the research agreement if they want to use the service. This is enforced by the service, so the users cannot opt-out from this. Naturally, this would be problematic if the services were a core part of studies or college life. The OtaSizzle services are not core services for the student community, however, and are not necessary for completing the studies, for instance. As with the handset-based measurements, the data are anonymised from the unique identifiers for research purposes, but it is subject to debate whether this is adequate when dealing

with behavioural data. Proximity and locality has a bit different meaning on online services when compared to the handset-based data collection or ubiquitous computing scenarios. We analyse only the actions users have done, especially as friendship in this service requires mutual agreement, and thus no data that the users would not be familiar with are collected. Regarding security, access and recourse, the data collection facilities have been secured, and only authorised requests are allowed. In this study, the online data collected, i.e., friendship data, are not collected only for research purposes as such, but are also part of the service data required to keep the service running. Therefore, data are collected for the specific purpose and are necessary for the functioning of the services.

### 3 Empirical results

SMS and phone call logs of the handset-based measurement data were used to identify the social network under analysis. Then the network of the identified users was studied from the OtaSizzle service data. As a result we have three different layers of social networks for analysis purposes – networks based on the SMS and phone call usage as well as the network based on OtaSizzle service usage. This chapter first describes the user group and illustrates the identified social network using the different datasets. The measures used in the analysis are described second and lastly the social network is analysed with these measures. The social network analysis was performed with the Ucinet software (Borgatti et al., 2002).

#### 3.1 Identified social network

As indicated in the data acquisition process in Figure 1, the basic demographics of the users were gathered using a short background questionnaire. Naturally the participants also had to have a smartphone where the research application could be installed to be able to participate in the handset-based measurements. OtaSizzle project distributed a hundred Nokia N97 smartphones to the users in the project and all of the users in this study also had such a device. All of the 20 users under analysis were Finnish and students at the Aalto University School of Science and Technology. Due to legal restrictions, the participants have to be over eighteen years of age and the average birth year of the participants is 1989. The group consists of eighteen male and two female subjects.

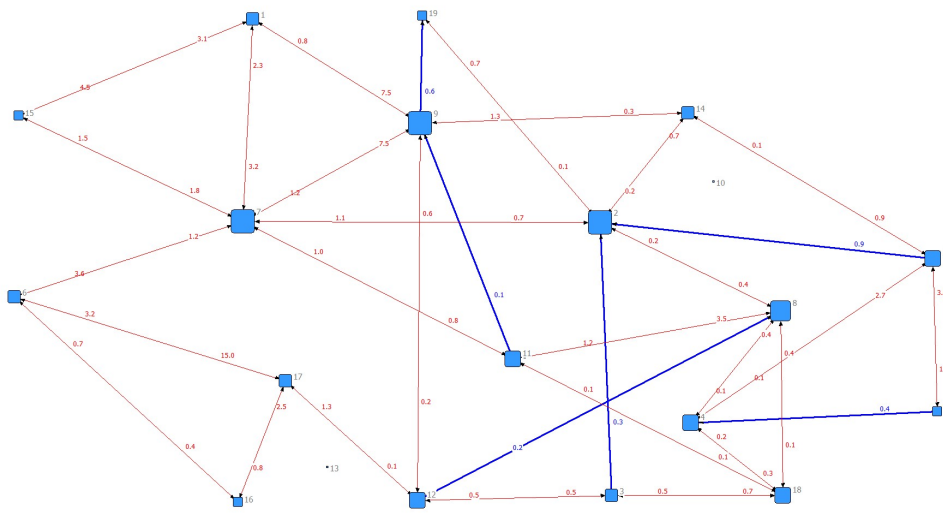
Figure 2 presents the identified social network based on the SMS logs of the handset-based measurement data. The degrees of the nodes set the node sizes. The networks have reciprocal and non-reciprocal ties, the former being illustrated with thin red lines and the latter with thick blue lines. The weights of the links indicate the number of messages sent per active smartphone usage day. For illustrational purposes the weights of the links are multiplied by ten (i.e., a weight of 1.0 indicates that an SMS has been sent once per ten usage days). Figure 3 depicts the social network based on the phone call logs of the handset-based measurements and the notations are the same as indicated above, except that the weight indicates the number of phone calls dialled.

Figure 4 depicts the social network of the users derived from the OtaSizzle service data. This network is different from the above presented networks, because it only

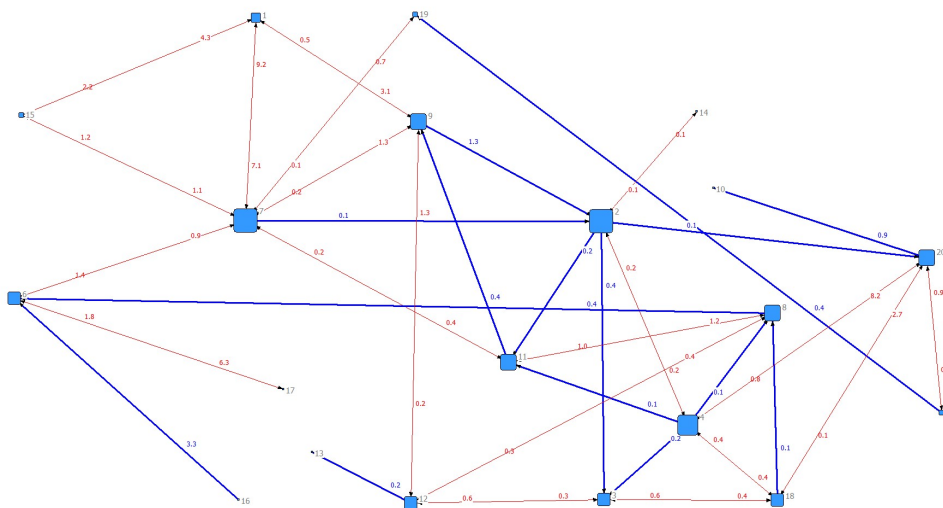


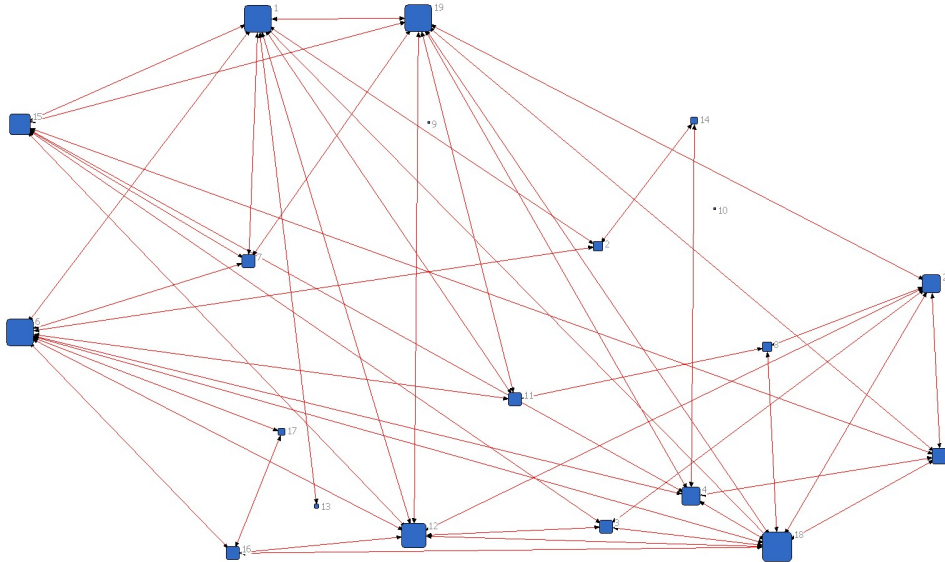
has reciprocal links. In the OtaSizzle services the receiver of a friend invitation has to accept the invitation for the link between two users to become active. Thus it is assumed that all the links are reciprocal. Although communication between users is possible in the OtaSizzle services, such data were not available for this research and thus there are no weights for the links available to indicate the frequency of communication between the users, as was done by Huberman et al. (2009) or Golder et al. (2007), for example. For illustrational purposes the nodes are positioned in a similar manner in all of the figures and the node sizes indicate the degree of a certain node as in Figures 2 and 3.

**Figure 2** Social network based on SMS data (see online version for colours)



**Figure 3** Social network based on phone call data (see online version for colours)



**Figure 4** Social network based on the OtaSizzle service data (see online version for colours)

### 3.2 Measures

This chapter shortly presents the measures used in the social network analysis from centrality and cohesion point of views. After this chapter the actual social network analysis results are presented and discussed.

#### 3.2.1 Centrality

The degree of a node indicates the number of links that the node has to other nodes in the networks. Degree is used as a simple measure of centrality in this study. Degree can be divided to indegree and outdegree, with the former indicating the number of links directed towards a certain node and the latter indicating the number of links that are originated from a certain node. With our data indegree refers to phone calls or SMSs received by a user and outdegree to phone calls or SMSs originating from a user. If the relations are non-directional, the degree centrality is easily calculated based on the node degrees. However, if the relations are directional the degree centrality is calculated based on the outdegrees of the nodes to better focus on the choices made by the nodes (Wasserman and Faust, 1994). The mean degree of the nodes is used as a measure of the network centrality and because the networks under analysis are of the same size we do not need to standardise the degree centrality measures by dividing the degrees with the network size.

Degree is an overly simplified measure of centrality, however, and thus we have also used betweenness as a measure of centrality. The nodes that occur on many shortest paths between other nodes of the network have a higher betweenness than the ones that do not. Contrary to the case with degree centrality, the algorithm to find betweenness indices for non-directional relations can be applied to directional relations as well. Thus the betweenness indices between the networks are compared in this study. Freeman's

approach (Hanneman and Riddle, 2005) of the Ucinet software is used to calculate the network betweenness as the mean of the node betweenness indices of the network. The network centralisation index is then calculated as a percentage of the maximum possible betweenness that the network can have.

### 3.2.2 Cohesion

Density is a measure of network cohesion. It is calculated by dividing the number of existing links (in + out) in the network with maximum number of links in the network. Links from a node to itself and the effects of link weights are excluded in the density calculations, which means that there can exist maximum 380 links for a network of 20 nodes. Diameter is another measure of network cohesion and it is calculated as the length of the longest shortest path between any pair of nodes in the network (Wasserman and Faust, 1994). Diameter measures how many steps are needed to get from one side of the network to the other and thus it is intuitively the ‘size’ of the network.

Density and diameter are simple measures of cohesion and thus we have also studied transitivity. Transitivity studies triplets of nodes and the links between them. A triad of nodes X, Y and Z is transitive if X directs a link to Y, and Y directs a link to Z, and then X also directs a link to Z (Wasserman and Faust, 1994). Triads can be vacuously transitive which means that either of the two conditions of the previous statement are not met (i.e., there is no link from X to Y and/or there is no link from Y to Z). By dividing the number of non-vacuous transitive triads with the number of triads of all kinds in the network the transitivity of a network can be calculated. In our calculations, the adjacency type of transitivity (Hanneman and Riddle, 2005) of the Ucinet software is used, meaning that the link weights have been excluded from analysis.

### 3.3 Social network analysis

The social network analysis was performed using the measures introduced in the previous subchapter. The differences of the networks were analysed both on aggregate and node level and the tables below summarise the differences. These differences will be discussed in the next chapter in more detail. The calculations were performed with the researchers themselves and with the Ucinet software.

**Table 1** Network structure of the different data sources

	<i>Phone calls</i>	<i>SMSs</i>	<i>OtaSizzle services</i>
Mean degree	3.5	3.3	4.8
Mean outdegree	2.8	3.0	4.8
Mean betweenness	26.1	21.9	6.7
Network centralisation index	30.1%	19.5%	12.2%
Density	0.145	0.158	0.253
Diameter	5	6	3
Transitivity	3.42%	0.57%	0.56%
Isolates	0	2	2
Average link weight	1.3	1.4	-

**Table 2** Node degree correlation between the networks

	<i>Phone calls</i>	<i>SMSs</i>	<i>OtaSizzle services</i>
Phone calls	1		
SMSs	0.8554	1	
OtaSizzle services	0.1330	0.0414	1

**Table 3** Correlation between nodes in phone call and SMS networks

Degree	0.8589
Outdegree	0.8123
Indegree	0.7243
Total degree (in + out)	0.8554
Average link weight out	0.7945
Average link weight in	0.7103

#### 4 Discussion

Table 1 demonstrates that, as we have suggested, networks are different. Characteristics of the OtaSizzle online service network show that it has a higher degree, density and transitivity while mean betweenness, network centralisation index and diameter are lower than in the SMS and phone call networks. This means that the OtaSizzle service network is denser and has a larger ego network for a single user on average than the other two networks. Furthermore, the phone call network and the SMS network are rather similar. This is also demonstrated in Table 2, where one can see strong correlation ( $> 0.7$ ) between the SMS network and the phone call network. Also, we see only weak positive correlation ( $< 0.3$ ) with the online service network. The strong correlation is significant in 1% level, where as the online network's correlation with phone calls and the SMS network is not significant, i.e., the null hypothesis cannot be rejected, and it may be that the networks are not correlated at all. This, however, further supports our claim that services are different, and thus to measure social relations, one should use more than one dataset. Naturally, our results are preliminary, and more investigation on this topic is needed. Limitations, such as the small sample size, may bias the result. Nevertheless, our research findings are supported by previous research in this field. Both Huberman et al. (2009) and Golder et al. (2007) discussed that online service users only communicate actively with part of their network, already indicating a difference in behaviour – and in the network structure. Moreover, van Cleemput (2010) has examined young adolescence communication patterns, and one of the many observations she had is that online service networks are not similar to those based on mobile telephony.

van Cleemput (2010) has also shown that mobile communications are similar in their nature, as we also have demonstrated in this study. In Table 3, we focus more closely to the commonalities between phone calls and SMS messaging, showing a strong positive correlation between them. In addition to the correlation of the degree measures in phone call and SMS networks, the link weights also correlate strongly. This issue further confirms that the SMS and phone call networks are very similar in terms of link weights and degree while the OtaSizzle service network differs substantially from these two

networks in terms of degree. Thus the claim that different kinds of communication channels create different kinds of networks is supported. Some of the networks are more alike, indicating similar usage pattern and strategies, whereas some are different, indicating that the usage – and the meaning of the link – is different.

Naturally, the small sample size might cause skewness in our results, especially considering the two isolates in SMS and OtaSizzle service networks. Also, the assumption that all online relations are two-way (i.e., reciprocal) is a potential source of skewness. We do, however, assume that these datasets show a valid difference between networks that requires more examination, especially since it has also been observed in other studies.

#### *4.1 Future research*

As we have indicated, this paper describes only the first empirical results that can be derived from the OtaSizzle platform and the research project. We assume that the OtaSizzle services will gain more momentum and thus we believe that more diverse and larger studies could be conducted in the future. Naturally, one of the core aims is to increase the number of users in our study.

Furthermore, additional data sources should be investigated. One of them is the addition of direct messaging capabilities in OtaSizzle, allowing a weighted graph to be examined. However, even more interesting would be to examine more global social media services, such as Facebook. Firstly, the use of Facebook is hugely more popular compared to OtaSizzle, thus allowing more realistic comparisons. Secondly, two different online social media services would allow us to examine the differences between these services, indicating whether the local service has different bonding behaviour compared to a global service.

We also believe more questionnaires should be used in our research. At this study, we used them to provide demographics of the users, but they could also be used to explain our findings, i.e., differences on communication between persons and differences we observe in the service. Especially the use of mobile surveys would be beneficial, as instead of passive logging, the mobile phone can be used to trigger questionnaires after specific actions and contexts. The MobiTrack platform would allow this kind of experience sampling which has been used in social sciences already by Froehlich et al. (2007), for example.

Also, modern smartphones have more sensor technology in them, which allows passive sensing of the user's surrounding. Technologies, such as Bluetooth (e.g., Eagle et al., 2009), have been used already to understand with whom people spend their time. This is an important social network which we are currently missing. Other context aware measures, such as location might also be fruitful, e.g., to examine if there are geographical differences on user behaviour.

The OtaSizzle research project allows also longitudinal data gathering. This way we can examine how the communication patterns and patterns of making friends change over time, as the current freshmen of Aalto University continue their studies. The dynamics of social networks are interesting in many aspects – how will the group identified here change over time, will there still be the same core persons with high degree or will the communication patterns divide the group into isolates, for example. Secondly, the OtaSizzle platform and services have been exported to other countries and cultures, and this will enable cross-cultural studies in the future as well.

## 5 Conclusions

This work aims to discuss how social relations should be measured. Based on previous research and our empirical findings, we demonstrated that people use different kinds of communication media channels differently, and thus their behaviour leads to different networks. Our solution for this is to suggest that, instead of using a single dataset, one should use several kinds of data sources to understand the social structures.

Our empirical evidence is based on a small sample ( $n = 20$ ) of college students in Aalto University. For these users, we were able to access mobile phone data, i.e., the SMS and phone call history, and OtaSizzle online social media service data. We show that the online social media services have a different kind of a friendship structure compared to those based on mobile phone communication. However, the SMS message and phone call networks are similar. The observation that instead of one network, the usage of different networks for different channels – and ultimately different purposes – is encouraging, as previous research in the online social networking validates this finding.

In our further research, the sample size will be increased and more data sources will be added. Possible data sources include other online services, questionnaires, and contextual data gathering. In this way, we will be able to investigate further how the communication channels, e.g., in global services or face to face, have different kinds of social structures. Furthermore, we aim to explain why this is the case, i.e., how and why people use different channels differently, and to provide suggestions for good measures for social relations.

## Acknowledgements

The work has been supported by the OtaSizzle research project that is funded by Aalto University's MIDE programme and Helsinki University of Technology TKK's 'Technology for Life' campaign donations from private companies and communities. The work was partly supported by the MoMI2 project and carried out in the Econ@Tel COST605 context. We also wish to thank MobiTrack Innovations Ltd. for providing the mobile audience measurement platform and acknowledge the sponsoring from Nokia and Elisa to our work.

## References

- Borgatti, S.P., Everett, M.G. and Freeman, L.C. (2002) *Ucinet for Windows: Software for Social Network Analysis*, Analytic Technologies, Harvard, MA.
- Donath, J. (2007) 'Signals in social supernets', *Journal of Computer-Mediated Communication*, Vol. 13, No. 1, Article 12.
- Donath, J. and Boyd, D. (2004) 'Public displays of connection', *BT Technology Journal*, Vol. 22, No. 4, pp.71–82.
- Eagle, N. (2009) 'Engineering a common good: fair use of aggregated, anonymized behavioral data', *IEEE Engaging Data*, Boston, MA.
- Eagle, N. and Pentland, A. (2006) 'Reality mining: sensing complex social systems', *Personal and Ubiquitous Computing*, Vol. 10, No. 4, pp.255–268.

- Eagle, N., Pentland, A. and Lazer, D. (2009) 'Inferring social network structure using mobile phone data', *Proceedings of the National Academy of Sciences (PNAS)*, Vol. 106, No. 36, pp.15274–15278.
- European Commission (2010) 'Consultation on the commission's comprehensive approach on personal data protection in the European Union', 4 November, Draft document COM (2010) 609/3.
- Fono, D. and Raynes-Goldie, K. (2007) 'Hyperfriends and beyond: friendship and social norms on LiveJournal', in Consalvo, M. and Haythornthwaite, C. (Eds.): *Internet Research Annual Volume 4: Selected Papers from the Association of Internet Researchers Conference*, New York, Peter Lang, pp.91–104.
- Froehlich, J., Chen, M.Y., Consolvo, S., Harrison, B. and Landay, J.A. (2007) 'MyExperience: a system for in situ tracing and capturing of user feedback on mobile phones', Paper presented at the *5th international Conference on Mobile Systems, Applications and Services*, June 2007, San Juan, Puerto Rico.
- Golder, S.A., Wilkinson, D. and Huberman, B.A. (2007) 'Rhythms of social interaction: messaging within a massive online network', Paper presented at the *3rd International Conference on Communities and Technologies (CT2007)*, June 2007.
- Hanneman, R.A. and Riddle, M. (2005) *Introduction to Social Network Methods*, University of California, Riverside, Riverside, CA.
- Huberman, B., Romero, D.M. and Wu, F. (2009) 'Social networks that matter: Twitter under microscope', *First Monday*, Vol. 14, Nos. 1–5.
- Jung, Y., Anttila, A. and Blom, J. (2008) 'Designing for the evolution of mobile contacts application', Paper presented at the *10th International Conference on Human Computer Interaction with Mobile Devices and Services (MobileHCI'08)*, September 2008.
- Kivi, A. (2009) 'Measuring mobile service usage: methods and measurement points', *International Journal of Mobile Communications*, Vol. 7, No. 4, pp.415–435.
- Kosta, E. and Dumortier, J. (2008) 'Searching the man behind the tag: privacy implications of RFID technology', *International Journal of Intellectual Property Management (IJIPM), Special Issue on: 'Identity, Privacy and New Technologies'*.
- Langheinrich, M. (2001) 'Privacy by design – principles of privacy-aware ubiquitous systems', in Abowd, G.D., Brumitt, B. and Shafer, S. (Eds.): *Ubicomp 2001 Proceedings, Lecture Notes in Computer Science*, Vol. 2201, pp.273–291, Springer, Berlin/Heidelberg, Germany.
- Mäntylä, M. et al. (2009) 'SizzleLab: building an experimentation platform for mobile social interaction', Paper presented at *Mobile Living Labs 09: Methods and Tools for Evaluation in the Wild*, September 2009, Bonn, Germany.
- Nokia (2007) 'Nokia Smartphone360 study: responding to what the user wants and needs', Nokia Corporation press release, October 2007, available at [http://www.nokia.com/NOKIA\\_COM\\_1/Press/twvln/press\\_kit/Nokia\\_Smartphone360\\_study\\_Press\\_Backgrounder\\_October\\_2007.pdf](http://www.nokia.com/NOKIA_COM_1/Press/twvln/press_kit/Nokia_Smartphone360_study_Press_Backgrounder_October_2007.pdf) [accessed 1/11/2010].
- OECD (1980) 'Recommendations of the council concerning guidelines governing the protection of privacy and transborder flows of personal data', 23 September, OECD Doc. C58.
- Pentland, A. (2008) 'Reality mining of mobile communications: toward a new deal on data', *The Global Information Technology Report 2008–2009*.
- Raento, M., Oulasvirta, A. and Eagle, N. (2009) 'Smartphones: an emerging tool for social scientists', *Sociological Methods & Research*, Vol. 37, No. 3, pp.426–454.
- Raento, M., Oulasvirta, A., Petit, R. and Toivonen, H. (2005) 'ContextPhone: a prototyping platform for context-aware mobile applications', *IEEE Pervasive Computing*, Vol. 4, No. 2, pp.51–59.
- Suhonen, E. et al. (2009) 'Kassi: everyday favors in social media – matching resources by means of campussourcing', Paper presented at the *European Conference on Computer Supported Cooperative Work (ECSCW 2009)*, September 2009, Vienna, Austria.

- ter Hofte, G.H. (2007) 'What's that hot thing in my pocket? SocioXensor, a smartphone data collector', Paper presented at the *Third International Conference on e-Social Science*, October 2007, Ann Arbor, Michigan, USA.
- 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*, Vol. 30, No. 2, pp.75–85.
- Verkasalo, H. (2009) 'Handset-based analysis of mobile service usage', PhD thesis, Helsinki University of Technology, Espoo, Finland.
- Wasserman, S. and Faust, K. (1994) *Social Network Analysis, Methods and Applications*, Cambridge University Press, Cambridge, UK.