Social Capital Companion: Capturing Personal Networks as They are Lived

Juergen Lerner*, Miranda Jessica Lubbers**, José Luis Molina** & Ulrik Brandes*

Abstract

Personal networks are complex variables characterizing the social resources and constraints of individuals and can be used to explain outcome such as behaviour, health, or job performance. Collecting qualitatively rich personal network data, however, is costly and difficulties increase if longitudinal data is needed. In this paper we identify requirements that collected personal network data should satisfy to be suitable for longitudinal analysis. Likewise we identify requirements on the tools for collecting personal networks and study which of these are satisfied by current data collection tools. We present design aspects and a prototypical implementation of a smartphone app (dubbed the Social Capital Companion) allowing users to keep a personal network diary on their own mobile devices over potentially long periods of time.

1 Introduction

A *personal network* consists of a focal person (called *ego*), all persons that are known to ego (called *alters*), and all relations between ego and any alter and between any pair of alters. Personal networks can be seen as complex, high-dimensional variables characterizing individuals. Typically a personal network study does not analyze just one network but a whole collection of them, sampled from a specified population of egos.

The study of personal networks is complementary to the study of *sociocentric* networks (also known as *whole networks*); compare Hennig et al. (2012). To illustrate the difference, consider a study that seeks to explain the smoking behaviour of teenagers by their social networks. A researcher following the sociocentric approach

---

* Department of Computer & Information Science, University of Konstanz (Germany).

** Department of Social and Cultural Anthropology – egolab-GRAFO (UAB).

could, for instance, sample from a population of school classes and ask every student in every sampled class about his or her behaviour (including smoking) as well as about his or her friends *within this school class*. While classmates can certainly be influential on a teenager's behaviour, the sociocentric approach might miss relevant social contacts. Indeed, some teenagers might be influenced by members of a sports-club who are not necessarily classmates, others might be very active in a religious association, some teenagers might be more influenced by family members than others, and so on. Increasing the network boundary by including members from every possibly relevant club, association, or family and collecting all the pairwise ties in this increased network is practically infeasible. A personal network approach to the same research question could sample from a set of teenagers (instead of schools) and survey for each of them his or her personal network. These personal networks would include for some teenagers large parts of the schoolmates, for others they would encompass parts of a sports club, or a religious association, or family members, etc. By definition, every relevant alter is included in ego's network. Since the alter-alter ties are only surveyed within a given network, the complexity of data collection scales linearly with the number of egos. For personal networks, as for whole networks, there is a growing literature addressed to identify the main factors that account for social network dynamics, including social influence through alters, and social selection based on homophily among ego and his/her alters; see Snijders et al. (2007); Steglich et al. (2010). A personal network study that seeks to separate *social influence* from *social selection* has to collect longitudinal network data, since these two hypothetical network effects are indistinguishable in cross-sectional data.

This paper deals with the issue of *collecting* personal network data and with the inherent difficulties in doing so. We argue that many studies require a type of personal network data which is hard to collect with existing means. Consequently we explore design aspects of a novel tool for collecting longitudinal personal network data and present a prototypical implementation.

While we believe that our tool can be beneficial for personal network re-searchers working in different areas, we introduce a particular application area — social entrepreneurship— to make the exposition more concrete. *Social*

entrepreneurship refers to entrepreneurial activities that are market-oriented and attempt to provide social value. Hoogendoorn lists examples of the objectives pursued by social entrepreneurs, including: "*reducing social, financial, and economic exclusion; providing access to health care, water, and sanitation in slum areas; reintegrating the long-term unemployed; revitalizing deprived communities; and generating solutions to environmental degradation*" (Hoogendoorn 2011, p.2). Many of the challenges faced by social entrepreneurs—among others, gaining access to funding and information or achieving visibility—can at least partially be tackled by effective networking. Therefore several researchers seek to explain the success[2] or failure of social entrepreneurs by the composition, structure, and dynamics of their personal networks: Greve and Salaf (2003); Schutjens and Volker (2010); Doh and Zolnik (2011). We stress that this paper does not attempt to shed light on the relationship between the personal networks and the success of social entrepreneurs; rather we take this general research question as an illustrative application area in which a new type of personal network data would be beneficial.

2 Requirements

We continue by identifying requirements on personal network data that might be used in our illustrative application: effective networking of social entrepreneurs. We believe that many of these requirements also apply to personal network data that is collected and analyzed in a different context.

2.1 Data Requirements

Ties in a personal network are associated with diferent types of information, among others, cognitive and behavioural relational dimensions. *Cognitive* data associated with ego-alter ties can comprise relational variables such as trust, esteem, like/dislike, emotional closeness, or being-at-ease-with. *Behavioural* data records social actions such as phone calls, emails, letters, or face-to-face meetings. Note that cognitive data can also contain information about the respondent's thinking of how often he/she is in contact with alters; ego's perception of the frequency of contact might be different from the actual behaviour (Killworth and Bernard 1980). As

---

[2] We leave the term *success* on an intuitive level in this paper; see, e. g., Kramer (2005) for an overview about measuring social innovation.

a matter of fact, different kinds of data can be collected with different means. Some behavioural communication events can be reliably and with low effort obtained from ego's calendar, email list, or phone call log. In contrast, cognitive data has to be provided, in one way or the other, by the respondent and might be hard to obtain at all after a certain elapsed time-span. Note that there are also mixed forms of data: for instance, a phone call comprises a behavioural event while the atmosphere or friendliness of the phone call is a cognitive dimension. Both types of data can be relevant in our illustrative application. Cognitive attitudes towards the social entrepreneur are certainly crucial determinants for whether a potential investor is willing to fund a start-up business. On the other hand, the frequency of contact or communication events shapes the opportunities to inform others about a new project and convince them to support it actively, financially, or otherwise. Thus, we note our first requirement on the collected data.

> *(D1) The collected personal network data should contain cognitive and behavioural dimensions.*

Many types of research questions require longitudinal network data, for instance, if the goal is to separate whether entrepreneurial success results from personal network characteristics (e. g., entrepreneurs with a larger support network are more successful) or, vice versa, whether entrepreneurial success explains parts of the network (e. g., successful entrepreneurs attract more supporters)[3].

Different relational dimensions in personal networks evolve in different time scales and exhibit two fundamentally different types of dynamics: state-type dynamics and event-type dynamics. *Relational states* (such as friendship, esteem, or trust) are assumed to have inertia and change gradually over time, albeit at a different speed. On the other hand, *relational events* (such as emails or phone calls) define point-processes where events occur in continuous time and many events can happen in a given time interval. These variations in the data dynamics pose challenges to both data collection and analysis but measuring and storing all

---

[3] Note that this issue is similar to the separation of selection and influence in longitudinal network analysis, see Steglich et al. (2010).

dimensions with the same sampling rate would constitute an inadequate data modelling. We summarize this in the second data requirement.

(D2) Different data dimensions should be measured and stored in potentially different time scales and the collected data should allow for state-type dynamics as well as for event-type dynamics.

Personal networks encode not only the alters and their ties to ego but also the alter-alter ties. Indeed, for explaining entrepreneurial success by personal networks it might matter not only the *network composition* (who is in the network) but also the *network structure* (how are they connected). For instance, if a social entrepreneur knows both friends having a positive attitude towards him or her, and potential investors who might fund the project, then it still makes a difference whether these two groups are interconnected or not (see Berrou and Combarnous (2008) for a discussion about structure and performance). We note this in our third requirement on personal network data.

(D3) The collected personal network data should encode network composition and network structure.

2.2 Tool Requirements

Besides the abovementioned requirements on the collected data, we also identify the following requirements on tools for collecting personal networks. Some of these tool requirements merely result from the requirements on the data while others constitute novel aspects.

As we mentioned above, different dimensions of personal network data might evolve at different time scales and therefore should also be measured with potentially different sampling rates. A possibility to achieve this is not to define the sampling rate a priory but rather let the respondent determine it in the way that data is updated whenever it changes or whenever some event happens. This requirement excludes the use of interviews for collecting the data; rather respondents should be able to have the tool for data collection continuously with them and to update whenever something happens. Such a tool would also minimize problems of irreliability caused by retrospection. Previous work has shown that personal network data collected with

a low sampling rate might underrepresent weak ties, might be biased towards more recent encounters, and depends on the momentary state of mind of the respondent (see Bondonio (1998) for a review).

(T1) The network collection tool should smoothly integrate into everyday life and continuously provide the possibility for data entry.

Besides providing an adequate and efficient means for data collection, the tool should also offer possibilities to view and explore the data collected so far. There are two potential benefits resulting from this requirement. First, the possibility to visually explore the own personal network might imply added value to the respondent and provide and intrinsic motivation to keep the data up to date. Secondly, appropriate visual network representations can lead to the recognition of errors or inconsistencies by the respondent and might improve data quality.

(T2) Besides collecting and storing the data, the tool should also offer the possibility to explore the own personal network data and thereby provide added value to the respondent and facilitate the recognition of inconsistencies.

We already indicated above that especially cognitive dimensions are hard to collect automatically but rather imply effort from the part of the respondent. On the other hand, it would be inefficient to make respondents also to enter computer-mediated communication and contact events that are already avail- able for automatic collection. An efficient tool should thus offer means for a hybrid approach for data collection. Communication and contact data that is available by automatic means should be imported with little effort; on the other hand the respondent should always have the possibility to correct such data or augment it by adding cognitive dimensions.

(T3) The network collection tool should automatically import com- munication and contact data whenever possible but should also allow to add cognitive dimensions.

Last but not least, a requirement on the tool for personal network collection states that the data should be controlled by the respondent. This means the data is

stored on the user's device, is not automatically sent to any server, and can be entirely or partially deleted by the respondent at any time. Furthermore, the tool should provide the possibility to remove or anonymize parts of the data before sending the network to others, e. g., to researchers who want to analyze the data. Many sensitive parts of the data, such as the alters' names, addresses, or phone numbers, are indeed not necessary for tackling common research questions and can be removed before exporting the data. It can be hoped that the possibility to control his/her own data increases the willingness for using the tool.

(T4) Ego should be able to control his or her data and to anonymize or remove parts before sending the network to others.

3 Current Tools for Measuring Personal Networks

In this section we give an overview about conceptually different existing means or tools for collecting personal networks. We are especially interested in assessing which of the abovementioned requirements are fulfilled and which not.

Automatically collected ICT data. Modern means for communication and social interaction produces vast amounts of data about whom we talk to or interact with. Examples include lists of phone calls or emails, electronic address books or calendars, online social networks, and many more. The advantages of this form of data collection are that it implies little effort from the part of the respondent, it produces longitudinal data with a very fine-grained temporal dimension and the data is reliable, objective, and does not suffer from limitations to recall the past. On the downside, such automatically collected data comprises only a very narrow view of the whole personal network. Cognitive data or information about face-to-face meetings is typically not included at all. Furthermore, many parts of automatically collected data are hard to interpret. For instance, is a facebook friend also a friend in real life or just a user account whose friendship request we did not decline? Is an email address with which we communicated frequently associated with an important alter or is it just a service email? Personal networks that contain only automatically collected data would certainly be insufficient to explain success of social entrepreneurs, as we would like to do in our example application. We conclude that

automatically collected data is an important component in a tool for personal network collection but that it must be supplemented by other data sources.

Interviews. Another approach to collect personal networks is to interview respondents at one or several points in time. Such interviews can be assisted by specialized software such as EgoNet[4] or VennMaker[5]. In these interviews respondents give information about themselves, list a number of alters, give information about each alter, and evaluate all alter-alter ties. While these tools can provide relatively reliable information about a person's core network, they are generally not well-suited for capturing information about the more ex- tended network including peripheral or weak ties (which nonetheless might be important for the analysis). In addition, data quality might suffer from retrospection, might be biased towards more recent encounters, and might depend on the momentary mood of the respondent. Also note that behavioural data (such as phone, email, or face-to-face contacts) can at most be measured in an aggregated way (e. g., evaluated by questions like "how often do you meet this alter?"). Due to the high respondent burden it is also not possible to collect the personal network dynamics with a high sampling rate. On the positive side, both EgoNet and VennMaker allow visual network exploration and can thus provide increased understanding of the own personal network and can be used for qualitative interviews in which the data can be interpreted or inconsistencies might be identified.

Contact diaries. The network collection method that comes closest to the tool that we envision is the use of a contact diary, see Fu (2007). In a contact diary respondents fill out regularly and frequently (for instance, every day) diary logs about the actual contacts that they had. For each contact on each day respondents provide characteristics about the alters (such as gender, age, or education), the circumstance of the specific contact (e. g., form, duration, place, audience, or feeling), and the characteristics of the ego-alter ties (such as familiarity, importance, or like/dislike). The contact diary method, thus, can collect data on cognitive and behavioral dimensions. The contact diary also limits the problems arising from retrospection since contact events are entered after a relatively short elapsed time.

---

[4] http://sourceforge.net/projects/egonet/
[5] http://www.vennmaker.com/

All data dimensions are measured in the same time-scale (e. g., once per day) implying too frequent updates on some of the alter and ego-alter characteristics. However, a slight modification would make it possible to update some dimensions only if there was a change to the last entry. On the downside, alter-alter ties are not collected and seem to be hard to collect with the diary log method. Besides this there are several issues with the usability of the tool. Keeping a contact diary entails a very high respondent burden and does not provide the possibility to view or explore the own network. Thus, respondents are unlikely to be intrinsically motivated to keep the diary but need other incentives, for instance financial recompensation. The diary log (in its basic form) is also inefficient in the sense that it requires the respondent to enter phone or email contacts which could be collected automatically. Finally, while the respondent controls his or her data and could in principle decide to delete it and not to give it to the researchers, there is no convenient possibility to anonymize or remove just parts of the data before export. The design and implementation of the Social Capital Companion could be seen as an attempt to overcome the limitations or drawbacks of the contact diary approach while keeping its advantages.

4 Social Capital Companion

The Social Capital Companion, or SCC for short, is a smartphone app allowing users to keep a personal network diary on their own mobile devices. In this section we discuss design aspects and describe a prototypical implementation which has been developed for the Android platform[6].

4.1 Dynamic Personal Networks

The underlying data structure of the SCC is a dynamic, labelled graph with a distinguished node representing ego. All parts of the data (i. e., the node and edge sets as well as the additional data labels) can change in continuous time. The components of a dynamic, labelled personal network are explained more formally in the following.

Ego and alters. The set of *actors* at time t is denoted by $V^{(t)}$. The only constraint is that a special actor (named *ego*) is always a member of $V^{(t)}$; all other actors are

---

[6] A preview version of the app is available at http://www.inf.uni-konstanz.de/algo/software/scc/

named *alters* and may enter or leave the set of actors at any time. The set of time points t at which a particular alter v is a member of $V^{(t)}$ is called the *lifetime* of v. The SCC allows for lifetimes that are finite unions of time intervals. That is the lifetime is defined by (and stored as) a list of k pairs of timestamps

$$[t_1, t´_1], \ldots , [t_k, t´_k], \text{ where } t_1 < t´_1 < t_2 < \ldots < t´_k,$$

having the meaning that v is a member of the network from time t_1 to $t´_1$, not in the network from t_1 to t_2, in the network from t_2 to $t´_2$ and so on. While we do not expect it to occur very often that an alter joins and leaves the network several times, we will apply the same concept of a lifetime to alter-alter ties and to the values of additional data labels associated with ego, alters, or ties, where it is more likely that values change several times.

Dyads and ties. The set of *dyads* $D^{(t)}$ -- $V^{(t)} \times V^{(t)} \setminus \{(v, v) : v\ V^{(t)} \}$ is the set of all pairs of different actors. Thus, the set of dyads at a point in time is already implied by the set of actors at the same point in time. The pairs involving *ego* are called *ego-alter dyads* and the pairs involving two different alters are called *alter-alter dyads*. The set of ties $E^{(t)}$ at time t is a subset of the set of dyads. A tie encodes that the incident actors are connected in a sense that depends on the particular application or that might be a function of additional data associated with actors or dyads. Note that there might be data associated with a dyad (u, v) even if u and v are not connected by a tie. In a typical point-line diagram of a network, ties are shown as lines while dyads are not shown, since they build a complete graph.

The lifetime of a particular dyad (u, v) is already defined by the lifetimes of the actors u and v (the dyad (u, v) is in the network at all points in time in which u and v are in the network). On the other hand, the lifetime of a tie might be individually constrained (so that two alters might be connected at some points in time and not at others—even when both are in the network). It is assumed that ego is connected by a tie to every alter throughout the lifetime of this alter. The lifetime of a particular tie can be of the same form as the lifetime of an alter, sketched above, i. e., it is a finite union of time intervals.

Data labels. The personal network can contain *labels* assigning values to elements from four different domains: ego, alters, ego-alter dyads, or alter-alter dyads. These data labels encode additional information such as characteristics of ego or alters, cognitive attitudes from ego to an alter or between two alters, or contact events among these. An arbitrary number of data labels can be declared for each domain. Labels for ego-alter dyads or alter-alter dyads can be symmetric or asymmetric. A symmetric label always assigns the same value to a dyad (u, v) as to the reverse dyad (*v, u*). On the other hand, the values of an asymmetric label associated with these two elements can be different or one can be defined and the other undefined.

Data labels have a *value type* which can be binary, categorical, numeric, or text. A *binary* label can only assume the values true or false, i. e., the element can have this label not without any additional information. For instance, family would be a binary ego-alter label with the interpretation that some alters are from ego's family and others not. A *categorical* label can assume one out of a finite number of predefined values. For instance, type of relation would be a categorical ego-alter label which can, e. g., assume the values family, friend, work, or other. Note that such a categorical label would restrict alters to be in just one type of relation to ego. On the other hand, coding each of the values as a binary label would allow for multiple roles (e. g., some alters are work colleagues and friends). A *numeric* label can assume only numbers as values and a label of type *text* assumes unconstrained text values.

All data labels can be dynamic in continuous time. This means that values associated with a particular element (e. g., a particular alter or a particular alter-alter dyad) can change at any point in time. Technically we make a similar restriction as for the lifetime of alters or ties: the *support* of a given label for a given element is a finite union of time intervals; each interval can map to a potentially different value; the label is considered as undefined or missing (for the particular element) at all points in time that do not belong to one of these intervals. Note that the same label can have different supports for different elements.

Labels declare their *dynamic type* as being one of *state-type dynamic* or *event-type dynamic*. These types of dynamics imply a different interpretation of data

updates. For instance, if a label encoding a relational state (e. g., esteem), changes its value (e. g., from *low* to *high*) then this is interpreted in the sense that the previous value is overwritten (the value was equal to low up to a given point in time, then it changed to *high*). On the other hand, subsequent relational events (e. g., emails) do not overwrite previous events but rather get appended to the list of previous events (if ego sends one email to an alter at time t and then another email later at time t then the list of previous events at time *t´* contains the first and the second email; the content of the first email is not overwritten by the second).

Secondary data labels. The data structure of the SCC allows to attach additional information to individual data entries. These secondary data labels encode meta information such as the point in time when the data has been updated (which might be different from the actual event time) or the origin of the data (e. g., does the data entry stem from some automated data collection such as importing from a phone call log or a facebook API query or was it entered by the user "by hand"). Such information might be important when analyzing the data, for assessing its validity, or for pointing the user to data entries that might need confirmation.

Data storage and access. Technically, the personal network sketched above is stored in an SQLite[7] database on the user's own device. Data access and update is efficiently implemented in the sense that query operations run in time proportional to the size of the result set (e. g., a query for all neighbours of a particular alter at a particular time need times proportional to the size of this neighbour set) and data update operations need time proportional to the number of data entries changed. The user never has to save the data explicitly. Thus, no data loss can occur when the SCC app is moved to the background (e. g., due to an incoming phone call) or when the device is switched of or runs out of battery. The data is visible only via the Social Capital Companion but can be exported to a GraphML[8] file if the user wants to send it to others; also see Section 4.3 below.

4.2 Graphical User Interface

---

[7] See http://www.sqlite.org/

[8] See http://graphml.graphdrawing.org

GRAFO *Working Papers*

The graphical user interface (GUI) of the Social Capital Companion determines how the user can view, explore, interact with, or update his/her own personal network. A good design is crucial for achieving efficient usability of the app and will be an important issue for future work. In this section we briefly describe the current implementation and indicate ideas for future development. A central paradigm of the SCC GUI is that the personal network can be explored from different view points and in different view spaces. More precisely, the network can be explored focussing on the alters or on the labels, i. e., on the resources that are in the network and the user can switch between viewing the data as lists of elements (which are convenient for searching, ordering, and selecting), as network visualizations (familiar point-line diagrams that reveal the structure of the network), or from the history point of view (allowing data exploration in time). Whatever the user is just viewing, it can be updated in place by opening edit dialogs.

Alters. Coming from the point of view of the alters (dubbed contacts in the GUI), the user can overview the list of all alters together with small hints showing the values of some of their labels (see Fig. 1). This list can be reordered by specific properties (e. g., alphabetically by name, by the values of specified labels, or by structural properties such as the degree in the network) and the user can search for alters matching specified properties.
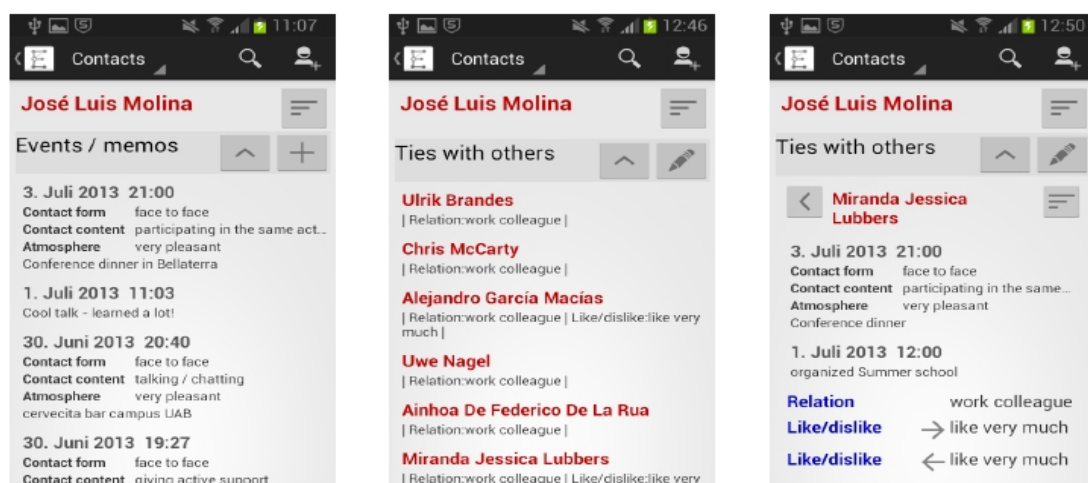

Figure 1: The contacts view shows characteristics of and events associated with alters, ego-alter ties, and alter-alter ties (see text for explanation).

Alter details are displayed in four categories: events involving the alter, his/her profile which encodes alter labels, the labels characterizing the tie be- tween ego and this alter ("*Ties with me*"), and the ties to other alters. Every category can be shown or hidden and the user can open edit dialogs to update the data or add new entries.

Labels. Viewing the labels offers a complementary way to explore the data (see Fig. 2). Instead of displaying all information associated with a particular alter, the label view offers the possibility to view the values of all alters or ties for one selected label. Depending on the particular label, these views could be used to explore social capital, e. g., to see the distribution of alters who are willing to support ego actively, financially, or otherwise.
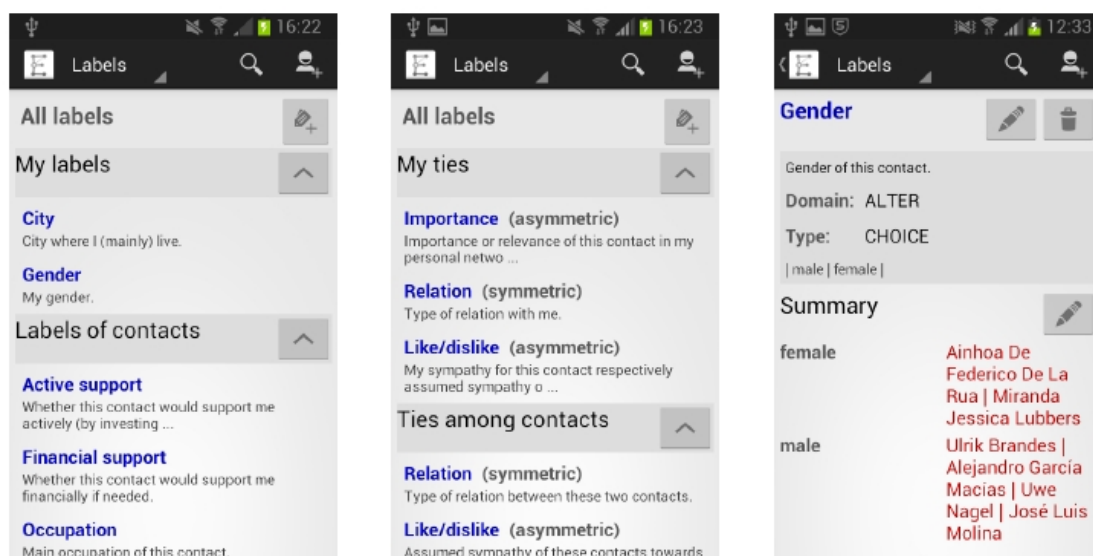

Figure 2: The *label view* shows overview and detail of data labels (see text for explanation).

Opening an edit dialog from the label view also offers a complementary approach for data update. While the edit dialogs started from the alters view allow updating information *by alter*, editing from the label view allows to assign the values of one particular label for all alters or for all ties. Furthermore, the user can define new labels for the personal network. The possibility to switch between overview (a list of all labels) and detail is analogous to the alter view.

View spaces. Besides browsing through lists of elements, the user can view his/her personal network also as a point-line diagram (left image in Fig. 3) or in its development over time (*history view*; see the right-hand image in Fig. 3). Thus, the user is offered different view spaces. The network view displays the structure of the network and allows to navigate over its ties. For instance, starting from an alter, see the other alters connected to him/her, etc. The history view displays the evolution of the network, indicates the points in time when the data changed, and allows to navigate over time. For instance, the user might spot that a significant change occurred at a certain point in time, can zoom in and see what happens just before or after.

Figure 3: *Network view* and *history view* provide the possibility to navigate over the network structure or to navigate in time (see text for explanation).

It is planned, but currently not implemented, to ofer yet another view space allowing to navigate in geographic space. Many alters or contact events can be associated with geographic locations and the user could be enabled to explore the surroundings, e. g., navigate to nearby alters or events.

4.3 Data Input and Sharing

As mentioned above, the user has always the possibility to update the part of the data he/she is just viewing. A typical usage scenario is that the user has met or is going to meet another person, locates this alter in the Social Capital Companion (by one of the many ways to search or navigate, sketched above) or adds him/her to the network if not yet in, explores the information associated with this actor, and potentially updates the data entry, for instance, reporting the new contact event. This way of data entry is called *event contingent* in Fu (2007).

However, the Social Capital Companion offers another way to enter the data which is implemented by the "*interview me*" functionality. In this scenario, the user has some spare time, for e. g., while sitting in a train or waiting for the bus, and could fill this time (or, stated differently, make use of this time) by updating his/her network. For this purpose, the SCC offers an "*interview me*" button, which presents a random or purposefully chosen data entry to update, see Fig. 4. In the current implementation, the SCC asks upon clicking the interview me button, with a certain probability to add more alters to the network; this probability decreases with the number of alters that are already in. With the complementary probability it asks to update an already existing element: either information associated with a randomly chosen alter or the values of a randomly chosen label. In future implementations we plan to modify the selection of the entry to update in the sense that elements are no longer chosen uniformly at random but biased towards entries that are in need for update. Such entries could be determined by considering the time since the last update (so that no elements remain forgotten), by identifying structural inconsistencies that need verification by the user (e. g., two alters with many common neighbours who are not directly connected themselves), or by presenting contact events that have been automatically imported and could be augmented by cognitive data by the user. It can be recorded in the secondary data labels (see above) that a particular entry has been updated after it was presented by the interview scheduler; indeed such information can be relevant when analyzing the data.
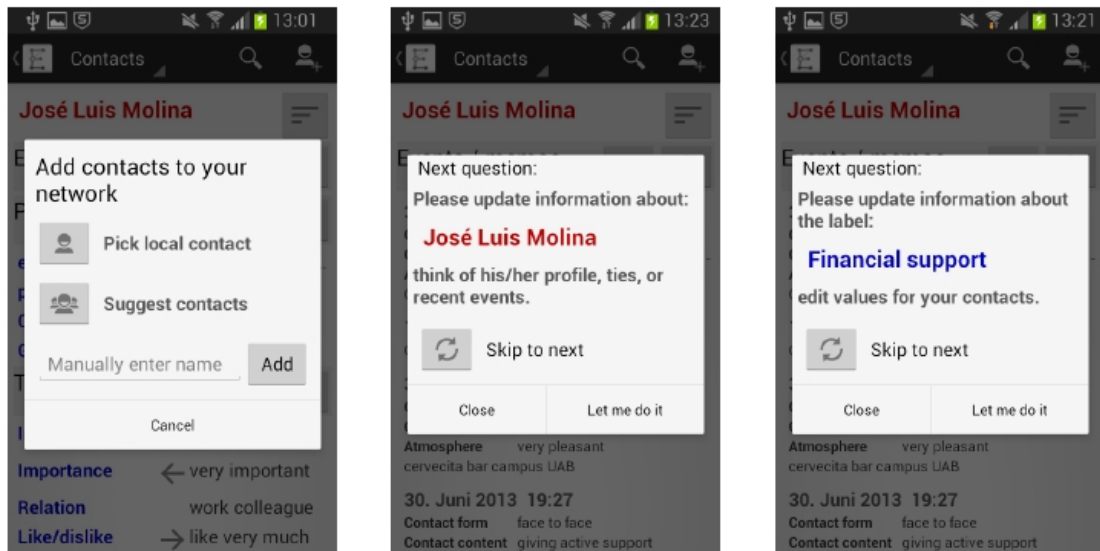
Figure 4: Possibilities to add new alters (*left*) and random entries presented by the *interview scheduler* for potential update (see text for explanation).

The two ways to update the personal network that have been sketched above could still be considered as manual data entry (they vary just in the way how elements in the network are selected). Besides this, the SCC offers the possibility to import computer-mediated contact and communication data automatically. In the current implementation the user can read the list of contacts on his/her mobile device, pick selected elements from it, get suggestions of contacts to add based on recent activity, and import phone calls and calendar events along with their attendees. In the future we plan to implement interfaces to more sources of data, including online social network and sent and received emails. Automatically imported data is recorded as such in the secondary data labels and the user has the possibility to add cognitive dimensions or further augment the data.

As mentioned before the personal network data is stored on the user's own device and not automatically sent to any server. We argue that it is a crucial feature of the SCC to let the user control his/her own data. Currently the only possibility to share the data is to export it in GraphML format and then share this file by any convenient means (e. g., sending it as an email attachment, copying it via a USB cable, etc). Users who participate in a study would have to go this way to give the data to researchers. It is planned but currently not yet implemented that the user can

decide which parts of the data are included in the exported GraphML file. For instance, it would be easy to leave out names or contact information of alters and in some cases aggregated statistics might be sufficient for research purposes.

5 Summary and Future Work

The Social Capital Companion presented in this paper is a smartphone app that allows users to keep a personal network diary on their own mobile device. The SCC is designed to collect longitudinal personal network data with a high sampling rate and allows to add both cognitive and behavioural data. It applies a hybrid approach for data collection by automatically importing data whenever possible but allowing the user to augment or correct such data manually. The resulting network is stored in a dynamic, labelled graph evolving in continuous time. Besides improving the graphical user interface (which arguably is a long- term task) we mention the following avenues for future work.

So far the SCC has been presented as a software that can collect and store arbitrary data associated with alters and ties. However, when it should be used as a data collection tool in an empirical study, researchers would need the possibility to constrain the collected data. For instance, a question like "*who of your alters could help you with household chores*? " would make sense in a survey of disabled persons but less so for studying social entrepreneurship. Thus, researchers designing a personal network study need a possibility to configure the SCC. In the current implementation we already offer to load an EgoNet study definition file whose questions about ego, alters, and alter-alter ties are then converted to labels in the SCC. This approach of adapting the app by making respondents load a particular configuration file seems to be feasible but needs to be further elaborated. For instance, researchers might not only need to ensure particular data labels but also constrain the means of data entry. In future work we plan to develop tools for designing particular data collection scenarios with the help of the SCC.

In this paper we focussed exclusively on the collection of personal network data and not on its analysis. The data collected by the SCC comprises a combination of networks of relational states and networks of relational events, both evolving in continuous time. We are not aware of any statistical model for such longitudinal

personal network data but want to point out that established models for the individual components exist. The stochastic actor-oriented models of Steglich et al. (2010) model networks of relational states and actor-level behaviour in continuous time. The model framework of Butts (2008) can model networks of relational events. We are confident that these two models can be combined in future work to make them applicable to the type of personal network data that we envision her.

References

Berrou, J.-P. and Combarnous, F. (2008). "Ties configuration in entrepreneurs' personal network and economic performances in African urban informal economy," Tech. rep., *GREThA*, no. 2008-25, http://ideas.repec.org/p/grt/wpegrt/2008-25.html [visited:22/07/2013].

Bondonio, D. (1998). "Predictors of accuracy in perceiving informal social networks". *Social Networks*, 20, pp. 301-330.

Butts, C. T. (2008). "A Relational Event Framework for Social Action". *Sociological Methodology*, 38, pp. 155-200.

Doh, S. and Zolnik, E. J. (2011). "Social capital and entrepreneurship: An exploratory analysis". *African Journal of Business Management*, 5, pp. 4961- 4975.

Fu, Y.C. (2007). "Contact Diaries: Building Archives of Actual and Comprehensive Personal Networks". *Field Methods*, 19, pp. 194-217.

Greve, A. and Salaf, J. (2003). "Social Networks and Entrepreneurship". *Entrepreneurship: Theory & Practice*, 28, pp. 1-22.

Hennig, M., Brandes, U., Pfefer, J., and Mergel, I. (2012). *Studying Social Networks,* Frankfurt: Campus-Verlag.

Hoogendoorn, B. (2011). "Social Entrepreneurship in the Modern Economy". Ph.D. thesis. Erasmus University Rotterdam.

Killworth, P. D. and Bernard, H. R. (1980). "Informant accuracy in social network data III: A comparison of triadic structure in behavioral and cognitive data". *Social Networks,* 2, pp.19-46.

Kramer, M. R. (2005). "Measuring Innovation: Evaluation in the Field of Social Entrepreneurship". *Tech. rep.*, The Skoll Foundation.

Schutjens, V. and Volker, B. (2010). "Space and social capital: The degree of locality in entrepreneurs' contacts and its consequences for firm success". *European Planning Studies*, 18, pp. 941-963.

Snijders, T. A., Steglich, C. E., and Schweinberger, M. (2007). "Modeling the co-evolution of networks and behavior" en *Longitudinal Models in the Behavioral and Related Sciences*, eds. van Montfort, K., Oud, H., and Satorra, A., Lawrence Erlbaum, pp. 41-71.

Steglich, C. E., Snijders, T. A., and Pearson, M. (2010). "Dynamic Networks and Behavior: Separating Selection from Influence," *Sociological Methodology*, 40, pp. 329-393.