# *TellUsWho*: Guided Social Network Data Collection

Stephen T. Ricken
New Jersey Institute of Technology
sr82@njit.edu

Richard P. Schuler
New Jersey Institute of Technology
rps22@njit.edu

Sukeshini A. Grandhi
New Jersey Institute of Technology
sg49@njit.edu

Quentin Jones
New Jersey Institute of Technology
qgjones@acm.org

## Abstract

*Significant gaps exist in our knowledge of real world social network structures, which in turn limit our understanding of how to design social software. One important reason for this has been that researchers have not been able to systematically probe individuals in sufficient detail about 'who' and 'how' they interact with in the social networks they wish to study. To address this shortcoming we designed TellUsWho, a web-based social network survey tool. We explored the tool's utility by studying the social ties of 141 students. TellUsWho supported the collection of rich social network data in a relatively short time period. Within an average of 34 minutes, respondents were able to describe their egocentric ties with people they regularly keep in contact. On average, respondents listed 42 alters, for each of which they answered 27 questions, resulting in 1134 responses. This compares favorably to traditional methods, which could require up to 15 hours per subject.*

## 1. Introduction

Social software such as MySpace, Facebook, Twitter, and Flickr, has captured the attention and imagination of researchers and designers. They offer individuals many new and exciting means of interaction and communication that was unimaginable only a decade ago. As a result of the popularity of social software there is a growing desire for greater understanding about the structure and effects of social networks. Understanding about how social networks form, change over time, and are structured in general would improve our ability to design and build systems. Unfortunately, existing data collection methods for social network analysis do not provide adequate means for systematically collecting large amounts of rich, detailed data [3]. As a result, researchers currently struggle to collect targeted data about specific communities and associated social network structures.

Researchers conduct social network analysis in order to explore the structure of the network, the relationships between entities, the flow of information, and its changes over time [26]. In order to conduct informative analysis, adequate data must be collected. However, there are challenges with the current methods, such as the inability to perform data collection both in breadth (number of social ties) and depth (detailed information about each tie and the nature of the tie itself) at a reasonable scale without overburdening respondents. Researchers have expressed interest in overcoming these and other common data collection challenges in order to facilitate the collection of richer and more detailed information regarding social network ties [3][4]. Current computer-aided social network survey tools attempt to address some of these challenges with limited success by computerizing earlier pen-and-paper based survey approaches [10].

We developed the *TellUsWho* tool in order to improve our ability to collect social network data directly from individuals. This tool allows the collection of in-depth information from a large number of individuals within short timeframes by allowing users to visually organize their social contacts [24][29]. This paper begins by briefly discussing the currently employed methods of social network data collection, as well as their strengths and weaknesses. This is followed by a presentation of the *TellUsWho* tool, and how it incorporates current methods as well as attempts to build upon them. Finally, we discuss how this tool was used to collect data about a university campus community from a sample of 141 respondents.

## 2. Background

Researchers employ various data collection techniques including structured and semi-structured interviews, surveys, observations, and data mining to collect social network data. Perhaps the most common data collection method is the questionnaire composed of questions that aim to solicit details about the respondent's social ties [26]. Questionnaires employ a variety of solicitation methods and attempt to balance between required effort of respondents to answer

IEEE computer society

978-0-7695-3869-3/10 $26.00 

questions and quality and/or quantity of the collected data. These solicitation methods are:

1) Roster vs. Free Recall - Roster presents a list of contacts for the subject to select from in answering questions, while free recall makes respondents recall alters by memory.
2) Fixed Choice vs. Free Choice - Fixed choice makes the subject select a specific number of alters, while free choice does not have such a constraint.
3) Ratings vs. Complete rankings - Complete rankings have the user rank order relationship ties over other alters, while ratings are applied to each individual tie.

When collecting social network data directly from respondents, either via questionnaires or interviews, two steps are performed: 1) 'name generation' and 2) 'name interpretation' [4]. During name generation, a respondent recalls, or selects, the names of people (alters) in their social network. Then, the respondent performs name interpretation, where details are provided on alters they named, such as demographics, communication characteristics, and relationship ties. Sometimes, as is the case with the study we present of our tool, the process of name generation is combined with name interpretation. In this case respondents generate alters which meet certain characteristics (e.g. listing work colleagues with whom they are in regular social contact).

Depending upon specific research goals, social network researchers may be interested in whole networks (a network of an entire group or organization), or egocentric networks (a network of an individual's alters). Due to the effort required of respondents to engage in name generation and name interpretation [17], various methods are employed to reduce or limit the name generation and/or name interpretation steps. This is typically done by presenting questions that have the respondent select only a subset of their social network, e.g. "list the five people you talk to the most," or limiting the attributes generated for each alter, e.g. "select the best description for each contact: friend, family, acquaintance, or colleague."

The subsections that follow will discuss examples of name generation, name interpretation, and data mining methods aimed at discovering information about social networks, how they are useful for data collection, and the shortfalls that exist with currently employed techniques.

## 2.1 Name Generation

Data collection methods that concentrate on name generation focus the participant's attention on recalling a number of names and the type of relationships rather than on answering detailed questions about social ties. Data typically consists of a list of alters and one or two dimensions of relationship to the participant.

The diary method of data collection asks participants to keep a record of other people with whom they interact for a period of time [26]. Kim et al. [18] directed participants to record each person they came in contact with in a diary over a two-week period in order to collect a list of daily contacts. While in theory this results in an accurate list of contacts, it requires diary keeping by participants and the management of this process by researchers, and may result in important contacts being missed if interaction does not occur during the study period.

Another technique for name generation is the employment of a hand-drawn network hierarchy [9] that focuses on respondents drawing graphs of contacts and connecting them to visualize social ties. Liao, et al. [21] used this method while investigating patterns of communication in work environments. Respondents drew a graph with his or herself in the center and then wrote names of groups and people whom they contacted in their organization. These names were then connected by similarity (e.g. same department) and degrees of closeness for each were indicated via a 5-point scale measure. This method works well for describing networks in small organizations, however, this approach is impractical for drawing a sizeable amount of alters in large organizations or describing a full egocentric network because the resulting graph becomes too large and cumbersome and its creation places a heavy time and recall burden upon the participant.

## 2.2 Name Interpretation

Researchers use name interpretation to obtain details from respondents such as relationship types, context, communication methods, frequency of contact, etc., in order to understand social ties, their strengths, and their social contexts. Usually, a small number of alters are generated or a small sample are selected from the name generation step for investigation.

McCarty et al. [22] conducted a phone interview of 1525 individuals with the goal of generating a representative sample of personal networks. They created a fixed list of 50 first names; excluding gender-neutral (e.g. Pat), and ethnically inclined (e.g. Letisha) names. The interviewer asked respondents to indicate if they knew someone who matched each name as they were read from the list and stopping when 14 names were known or the end of the list was reached. Respondents then answered five name interpretation

questions about each of the 14 alters. Each survey lasted approximately fifteen minutes.

Wellman and Wortley [28] interviewed a sample of twenty-nine Toronto residents in order to understand social support within a community. Each subject was interviewed for 10 to 15 hours and answered questions about their relationships with active members of their social network. Close ties were discussed, but less intimate alters were not examined. This study gathered a large amount of data, however, the number of respondents was small and the interviews were time-consuming.

Carrasco et al. [3] used an interview method where respondents were instructed to record contacts on post-it notes and position them within various concentric circles based on degree of closeness [12]. A sample of 15 contacts was selected for name interpretation, demographics were collected and four-to-six questions about communication and interaction patterns are queried. Finally, 6 of the 15 contacts were selected and the participants were asked about the last social encounter with each, what they did, when, where, who else was involved, etc. Although a large amount of social network information was collected using this approach, the interview was time-consuming, averaging 2.5 hours per participant. Also, the number of respondents was limited due to the length of each interview. It is noted that each part of the survey required using a smaller subset of alters to reduce complexity, thus a larger scale study would be increasingly complex.

## 2.3 Mining data

Various data can be mined for social network analysis. This includes: the interpersonal connections listed by users of social network sites (SNS), organizational communication, online mass interaction/communication [31][25], and personal communication histories. Analysis of such data can create rich visualizations. Here we review research using such data and outline how minable data sets do not typically contain information about key aspects of the social network under study. For example, information about relevant social ties is often missing because communication occurs via channels outside of the database such as face-to-face communication.

**2.3.1 Social network sites.** SNS are increasingly being used for social interaction over long periods of time [20]. The common ways in which SNS are used includes connecting with old friends and acquaintances and making new contacts [19]. When a user sets up their SNS friends list, they are engaged in name generation and mining. This, combined with the communication data, can provide insight into their social networks. However, SNS are not designed for the scientific probing of individuals about their social ties and as a result, do not always collect the type of user data desired. Heer, et al. [11] devised Prefuse and Vizster, social network visualization tools, which utilized Friendster data to create social network visualizations. The tool mines a user's network and creates a spring graph representation of connections. The visualization includes filters by demographic information (e.g., gender, interests, age). Although this tool creates scalable social network visualizations, relationship data beyond ties is not included. Further, it is not possible to probe users with more information such as relationship characteristics or degrees of closeness.

Gilbert et al. [14] mined MySpace, collecting data to examine the difference in usage between users from rural and urban areas. Data included relationships and demographics plus characteristics given by numbers of "wall posts" and private messages between users. The data mined from 4000 users supported their hypotheses on usage differences between rural and urban areas based on number of friends and distance from a user. Information such as ego-alter relationship ties (e.g., immediate family, relatives, friends, colleagues) could not be determined with the given data since this information is not a feature of MySpace. This relationship information could have been obtained if the researchers surveyed users after mining data.

Mislove et al. [23] mined four SNS to examine patterns within several networks rather than focusing on one single site. They found that online social networks were structurally different than other social networks and suggested that this work could benefit SNS application designers. This study outlined two important problems with mining SNS. First, the researchers could not obtain user data from the site owners; instead they were forced to "crawl" each SNS, a technique where the system follows links to every node (each node is a contact), beginning at a particular node. The system continues to record each tie from one contact to another. Second, data mining methods for social network sites themselves present challenges for researchers. Crawling an entire network is often unfeasible due to the number of connections. In this case the authors used a partial crawl, where they recorded a small number of degrees separated from the ego, then stopped. A partial crawl can be biased due to overestimation of degrees between nodes and underestimation of network symmetry [1].

While usage statistics of SNS may be useful to researchers, there are some challenges using such data for social network analysis. SNS Terms of Service (TOS) impact how researchers can use and gather data.

For example, a Facebook TOS change denied researchers from saving information such as users' profile and their related contacts [8]. This change rendered the SNS useless for data mining purposes. Mining SNS social tie data may help to reveal important and missing contacts that an interviewed participant may not mention or recall. However, the networks represented are influenced by the design and use of each SNS and may not be an accurate representation of a social network for various reasons such as: 1) Not all SNS users with offline connections are tied within an SNS [1]; 2) There are contacts that may not be part of SNS; and 3) Many have SNS contacts that would not be considered relevant contacts to social network analysts.

**2.3.2 Organizational communication.** Mining organizational communication such as email can result in data about egocentric networks [30], and can generate data corpuses, which can be utilized for SNA. Such was the case with the Enron email corpus. As part of their investigation into the collapse of Enron in 2002 the Federal Energy Regulatory Commission (FERC) released a corpus of email from 158 employees. This collection provides researchers a rare glimpse into the social network of a large business organization. Diesner et al. [6] analyzed the email data to examine the effects of a major organizational crisis on its communication network. In order to conduct the analysis they performed substantial data enhancement by including 525 missing employees, linking individuals to multiple email addresses, and incorporating career history information. Additional study of the Enron email corpus demonstrate that content of emails, vocabulary, names, and other information can also be used to distinguish social ties and details [15] within social groups and organizations. However, the challenge of mining a single form of electronic communication, such as email, is that within an organization there may be preference for many communication styles including synchronous communication with co-located employees [32]. In the same vein, an individual's social network mined through email may not be rich without mining the communication between groups of individuals. Further, such corpuses are currently not linked to tools that

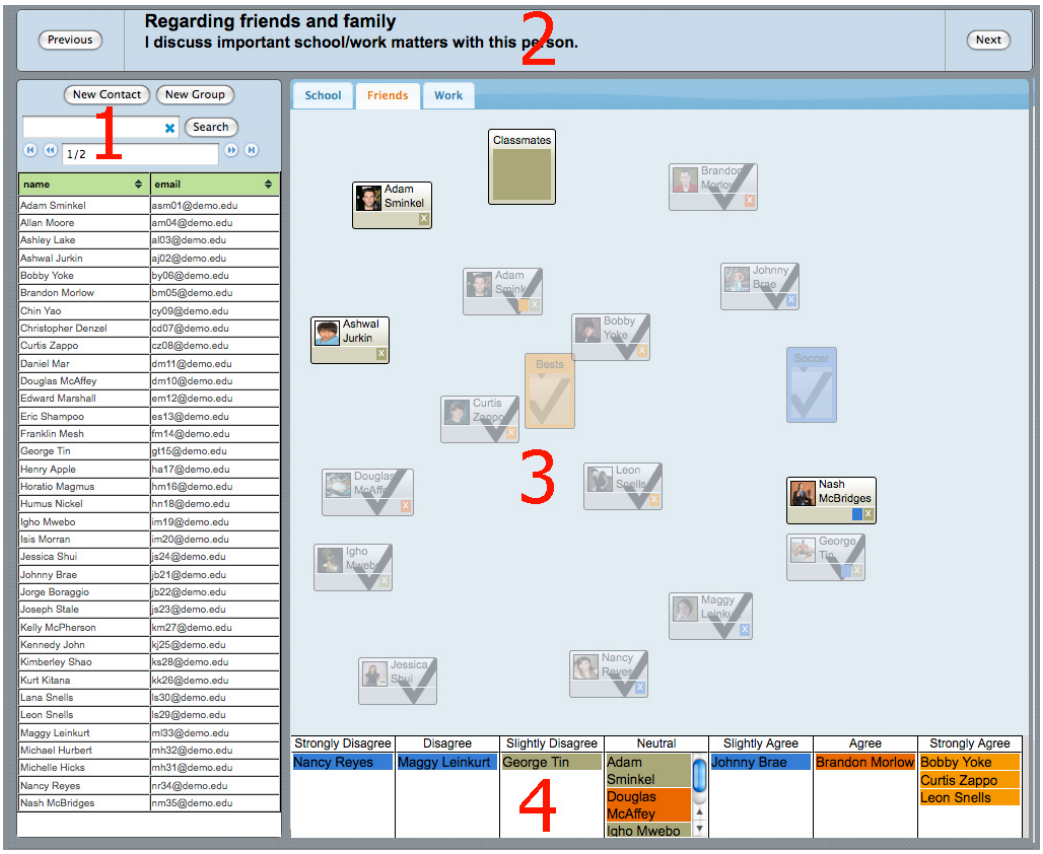



**Figure 1. TellUsWho user interface**

allow for individuals to be systematically probed about their historical and ongoing social ties.

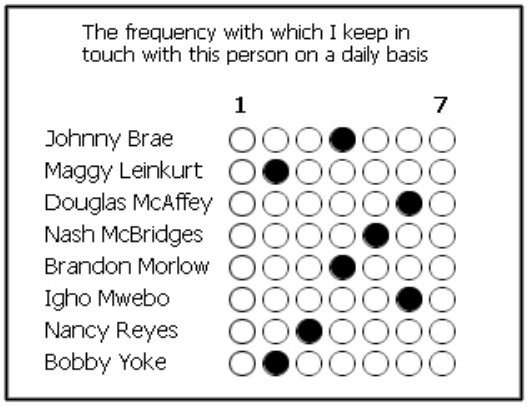
Another use of mining organizational communication is the employment of organizational charts (a chart of the structure of an organization with relationships) that map members to each other in a visualization. Sociologists who study social networks within organizations [30] sometimes use an organizational chart to display connections for participant recall purposes. By using this chart as a roster, they can probe respondents about individuals and groups, such as "How much time did working with employee X save you?" Using this method, researchers can discover workflows and the usefulness of connections in order to increase business efficiency [5]. While the social tie visualization may be revealed through a system, as seen from mining SNS, this data may not be available for individuals outside of an organization.

**2.3.3. Online mass interaction.** Usenet, a worldwide-distributed Internet discussion system, has been studied by social network analysts because of its international diversity of users and their massive numbers of crosspostings (taking a posted message from one topic and authoring it to another topic) between groups. Hierarchies categorize topics into groups, with eight at the core (e.g. alt, comp, misc, news, rec, sci, soc, and talk). Smith [25] mapped a sample of messages and found that crossposting groups formed "meta-clusters" because nearly every message within the sample was crossposted to another group. This discovery highlights the fact that group social network data can be mined from online mass communication, though from this data there may not be clear individual social networks due to the anonymity of some posts and the ability for anyone to browse posts without correspondence.

**2.3.4 Personal communication history.** Another approach to mining interpersonal communications is to find ties in order to visually organize contacts. While ContactMap [29][30] was not developed for social network research, mining contacts for users to place them on a desktop for easy organization is appealing for gaining insight into how an individual would visualize their own social network. Developing this approach into a social network survey tool could prove useful in allowing users to self-report connections and groups of contacts.

## 2.4 Motivation

Each of the previously mentioned social network data collection methods is unable to gather data consisting of both many alters and rich social tie details. Our goal was to create a computer-aided survey tool that allows for the rich collection of greater quantity and quality social tie data to that of current survey methods while reducing the time requirements for both the researcher and the respondents. This work was also motivated by a need to map the egocentric network structure of our students.




| Never | Yearly | A few times a year | Monthly | Weekly | Daily | Several times a day |
|---|---|---|---|---|---|---|
| | Maggy Leinkurt | Nancy Reyes | Johnny Brae | George Tin | Igho Mwebo | |
| | Bobby Yoke | | Brandon Morlow | Nash McBridges | Douglas McAffey | |

**Figure 2. Top: common survey method Bottom: Tool survey method.**

## 3. Guided social network survey

Most computer assisted social network survey tools have attempted to replicate previous pen and paper methods, and as a result, have only achieved modest gains in the amount and quality of data collected [10][13]. The survey tool presented in this paper advances the use of computer-aided surveys through a unique visual method. The authors implemented a user interface similar to ContactMap [29], where the respondent may visually group alters on the screen.

The *TellUsWho* survey tool provides support for many of the main question formats, roster, free recall, free and fixed choice, rankings and ratings, combining them into a computer supported and user-guiding interface. The reduction in respondent time and effort comes from the ability for a user to answer questions using drag-and-drop; rather than having to answer a question multiple times for several alters. A user is also able to apply an answer to multiple alters at once by using the in-depth response area.

### 3.1 User interface

The user interface of the tool consists of four main components (see Figure 1): 1) a computer generated list of the user's contacts mined from other applications such as email - shown on the left; 2) a question-space where questions are presented - shown above the network visualization space; 3) a desktop area where users engage in the guided visualization of their social network and in so doing answer numerous key questions; and 4) beneath the desktop area is an in-depth response area to answer detailed name interpretation questions.

**3.1.1 Contact list.** The contact list presents to the user their contacts that may be mined from various sources. Its purpose is to facilitate the recall of contacts and their addition to the desktop. For the field trial presented in this paper every contact from the user's university and Gmail email accounts were listed. The list allows ascending or descending alphabetical ordering by name or email address; and to facilitate a large number of contacts it is paginated and a search box is provided.

**3.1.2 Question space.** The question space specifies what the user should do: name generation questions specify that the user drag contacts onto the desktop, name interpretation questions specify the user to place alters into the in-depth response area. Other prompts may be to group certain alters, to organize the desktop, or to perform other tasks. Users navigate the survey via the previous and back buttons on the left and right respectively.

**3.1.3 Desktop.** Each contact placed on the desktop appears in a box with their name, like an avatar, and becomes an alter in their social network. These avatars can be dragged around the desktop and placed wherever the respondent wants. A respondent can add a contact to the desktop by drag-and-dropping the individual from the contact list to the desktop or they may click "new contact" and type in a name and/or email for a contact. Group avatars can also be created to organize alters visually through color.

Users have two ways of organizing alters on the desktop. Implicitly, they can physically group alters by piling avatars close together. Explicitly, users can create Group avatars. When the user clicks "New Group", the user types in a group name, and an avatar with the name appears on the desktop with a particular background color. When a user adds an alter to that group, the background of that alter will become the group color (signifying the primary group). When the user adds an alter to multiple groups, the color of the additional groups will appear in small boxes under the alter to display multiple group associations. By using groups, the subject reports information on the relationship of others and their impact on the participant.

**3.1.4 In-depth response area.** The in-depth response area, which appears below the desktop, is used to answer detailed questions of different types (e.g., a single box, yes/no designators, or scales) that are unanswerable by merely adding contacts to the desktop. Respondents answer each question by placing the alters on the desktop into the response area. Alters placed in the boxes appear as a list under their response and can be dragged between responses to change an answer. Respondents may place several alters into the response area at the same time in order to rate individuals and rank order between alters. For example, in Figure 2, nine alters are used to answer a single question, rather than repeating the same question nine times for each individual.

### 3.2. Data collection

The *TellUsWho* survey tool collects different types of information in order to gather a rich description of social ties. Each phase comprises of several data collection techniques.

During name generation, the subject's placement of alters on the desktop enables the researchers to use cluster analysis to determine interrelationship

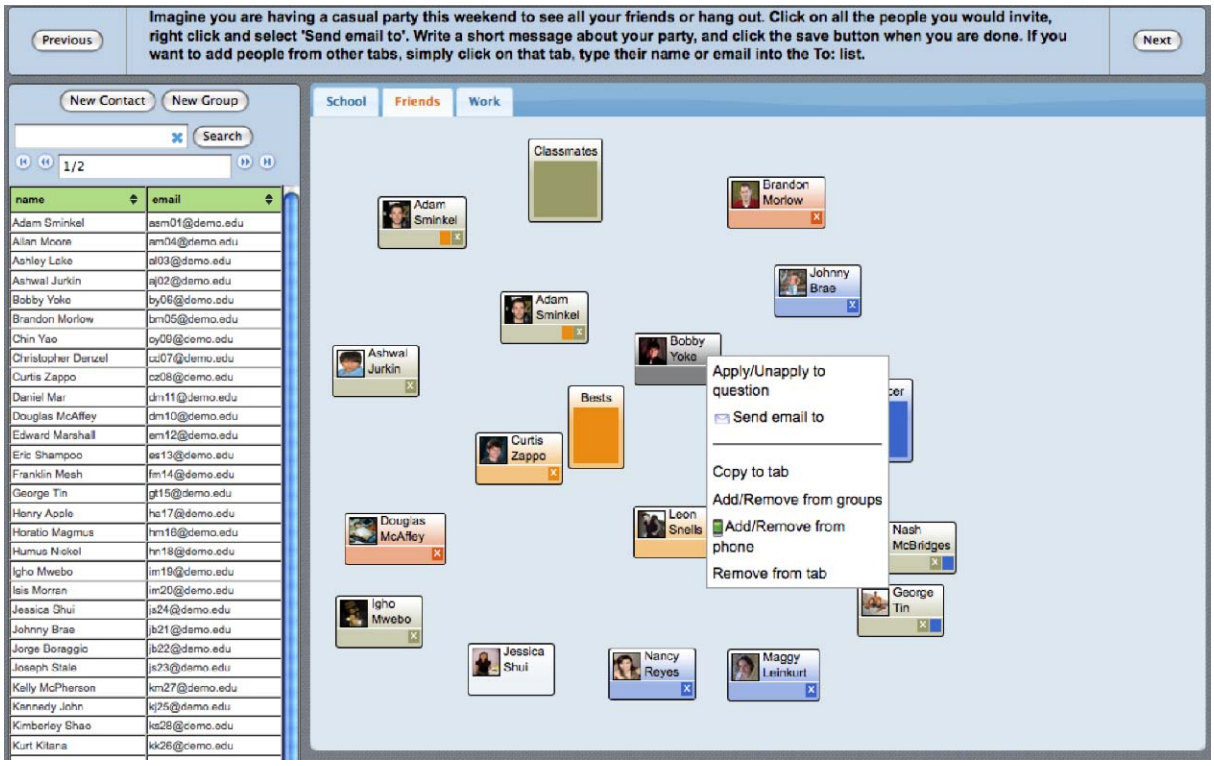



**Figure 3. Context menu**

groupings. Also, respondents are answering relationship questions to add alters to the desktop, so we explicitly collect general information on relationship ties. Respondents may also create "groups" to place alters in, adding an additional element of relationship detail.

During name interpretation, data collected on each question includes both ratings and complete rankings. Respondents drag any or all alters on the desktop into the response area. Each name they add is rated itself, but the subject also sees the other names they added so they may rank their relationships.

## 4. *TellUsWho* user trial

A user-trial of the tool was conducted by surveying 141 students (undergraduate and graduate) at a North Eastern US university. The aim of the trial was to see how well the tool performed in detailed social network data collection from students within a short time period of taking the survey. The survey was carried out in computer labs with groups averaging 25 respondents. Before entering the lab, respondents supplied their credentials to their university and Gmail email. The system pulled every name and email from headers found in their inbox to populate the contact list.

Respondents signed up for an hour and a half time block to complete the survey and were allowed to leave upon completion of the survey. Before starting the survey respondents viewed a 15-minute introduction and video explaining what they would be doing for the survey.

While respondents can complete this survey independently online, the authors used a supervised lab condition to address bugs and test the usability of the tool.

### 4.1. Name generation

Once a respondent began the survey their first task was to manually build their social network through name generation. To do this, the system prompted 21 specific questions (Figure 4) and the respondent added contacts via the roster technique (drag-and-drop from the contact list), or free recall (manually entering an alter's name through the "new contact" dialog). Using these questions and the contact list, respondents not only visualized their social network, but also related each alter to themselves, rather than the system automatically mining ties. The desktop was split into three tabbed desktops, School (people the participant related to University), Friends (people the participant related to friends and family outside of University), and Work (people the participant related to their employment, if applicable). These tabs kept each desktop separate such that the respondent could place

many contacts without cluttering space, and helped the respondent focus their recall efforts, plus it inherently supplied social context data.

**Figure 4. Questions asked during desktop placement**

| |
|---|
| **Questions about School** |
| 1. Place your roommates |
| 2. The person who introduces you to other people / the most popular friend you have |
| 3. Place the people you socialize/hang out with on campus |
| 4. Place the people with whom you are in sports, clubs, or organizations |
| 5. Place the people with whom you eat meals (breakfast, lunch, dinner, or snack) |
| 6. Place the classmates / studio-mates and people who you work with on projects |
| 7. Place the school friends you talk to/socialize with online (such as talk to on instant messenger or in chat rooms, play online games with, etc.) |
| 8. Place people who are acquaintances – people who you may greet upon passing, but do not normally converse or socialize with |
| 9. Place your current professors, advisors, and other NJIT faculty/staff you communicate with |
| 10. Place professors with whom you still keep in contact |
| **Questions about Friends & Family** |
| 11. Place your roommates |
| 12. Place the people you socialize/hang out with |
| 13. Place the people with whom you are in sports, clubs, or organizations |
| 14. Place the people with whom you go out to lunch or dinner with |
| 15. Place the people you talk to/socialize with online (such as talk to on instant messenger or chat rooms, play online games with, etc.) |
| 16. The most popular friend you have |
| 17. Family members you see or talk to on a daily/weekly basis |
| **Work** |
| 18. Place the people with whom you socialize or hang out with outside of work |
| 19. Place the people with whom you socialize with while at work |
| 20. Place your coworkers |
| 21. Place your boss or supervisor |

### 4.2. Email task

Once the respondent completed building their social network, they were given the task to produce two different emails from the contacts they added to their desktop. The first email assignment asked the respondent to invite friends to an imaginary party he or she was holding, through email. The respondent selected the alters, then clicked "send email to" and produced a fake email with subject and body (Figure 3). Respondents understood that the email was not actually being sent. The second email, reproduced in the same manner, prompted the respondent to email people in school they would ask for help in their toughest class. By asking respondents to reenact these scenarios, we remind them of important alters they may have missed while placing people on the desktop. The data collected from these situations also indicate the relevance of certain alters for social affairs.

### 4.3. Name interpretation

After a respondent had answered all the questions to build their social network, they were prompted to answer 6 questions about each alter they placed on the desktop (Figure 5) by dragging them into the survey block. The respondent could drag multiple alters onto the survey block per question, which lead to both ratings and complete rankings techniques. The respondent can easily 1) singularly relate the alter to the question and 2) relate the alter to all other alters used to answer the question. Drag-and-drop allows the respondent to answer each question for many alters at once. This timesaving method allows for detail questions to be asked without user fatigue.

**Figure 5. Questions forcused on name interpretation**

| | |
|---|---|
| 1. | The frequency with which I keep in touch with this person on a daily basis |
| 2. | I discuss important school/work matters with this person |
| 3. | I discuss important personal matters with this person |
| 4. | This person is there for me if I need help with work/school |
| 5. | This person is there for me if I need help with personal matters |
| 6. | The primary method of communication I have with this person is: Face-to-Face; Phone conversation; Text Message (SMS); Instant messenger (AIM, GTalk, Yahoo, IRC); or E-mail |

## 5. Observations and Discussion

The tool performed well in generating social network data, with a very minimal number of bugs and a small number of frequently asked questions by respondents when running the user trial. The video tutorial made it possible to outline the survey instructions in easy to follow steps.

For the 141 respondents in the study, an average of 42 alters was generated per respondent. The number of alters generated ranged from 5 to 150 per respondent and a total of 6213 alters were produced using this tool. Respondents took an average of 34 minutes to complete the survey, the minimum time being 7 minutes and the maximum time being 141 minutes with a standard deviation of 17 minutes. In this time respondents answered an average of 1134 probes about their social network stemming from 21 name-generation questions and 6 name-interpretation

questions. Respondents were able to complete the survey without running into severe technical difficulties or misunderstandings.

While this tool was found to be effective, there are limitations to the user trial. This study was run as a single-use case and could potentially benefit from being placed inside of a social network software application. This would add the advantage of collecting usage statistics while also gathering data through survey questions over time. Also, this work focuses only on the design of a tool to facilitate social network data collection and does not examine the reliability of the collected data. Follow-up studies should compare the amount of information collected using this tool to the other more traditional methods mentioned above.

The tool itself may also be refined to enhance the user experience. This tool currently supports respondents' university email and Gmail; however, it can be programmed to handle almost any email account in order to capture more contacts. Collecting IM buddy list information and mobile phone contacts would also be worthwhile. Running more user studies with these additions to the roster technique could test the advantage of having multiple sources available for contact recall. Further as a next step one could consider showing respondents an automatic visualization of his or her social network ties while they use the survey tool. By providing the visualization, respondents can see how they are connected to others while interpreting their social ties.

Finally, we ran the study in a laboratory where we could observe users and provide support in case of difficulty. While a small number of frequently asked questions were raised it was clear that monitoring in the lab was unnecessary. By addressing the frequently asked questions, a revised version that will streamline the instructions and allow for the survey to be taken independently online can be implemented as the tool runs in a web browser. An obvious next step is to allow respondents to use the survey tool in the comfort of their home computer rather than in a computer laboratory allowing for much larger user participation.

## 6. Conclusion

As social computing applications become more integrated within daily life, the importance of understanding social networks they support increases and so too will the ability to collect and leverage richer social network data. The *TellUsWho* tool is a means to collect large-scale and in-depth social network data, which is necessary to understand the growing complexity of social networks. In this paper we reviewed various social network data collection methods and highlighted the tension that exists between collecting detailed in-depth data about few alters and collecting restricted data of many alters through these methods. Mining social data produces large quantities of social tie information; however, useful qualitative data and missing ties can only be obtained by probing individuals on their connections. Surveying individuals can be costly for researchers and respondents due to the time required to gather both social network ties and detailed information on each social tie.

We sought to reduce this tension by creating a tool that allows for collecting in-depth details about a large number of alters in a short period of time. The *TellUsWho* social network survey tool incorporates a mined contact list facilitating easy name generation, an interface that allows respondents to quickly visualize their contacts and group them, and then answer in-depth questions about them. By combining data mining and social network visualization techniques into a survey tool we were able to produce a "best of both worlds" approach. The creation of this tool is a first step in collecting richer social network information, and we are moving forward to learn more by using this tool in conjunction with other collection methods.

## 7. Acknowledgements

This research is partially supported by the National Science Foundation Grant NSF CNS 0454081 and DST 0534520. The opinions expressed are those of the authors and may not reflect those of the NSF.

## 8. References

[1] Becchetti, L., Castillo, C., Donato, D., and Fazzone, A. A Comparison of Sampling Techniques for Web Graph Characterization. Proc. of the Workshop on Link Analysis (LinkKDD'06), Philadelphia, PA, (Aug 2006).

[2] Binder, J., Howes, A., and Sutcliffe, A. The problem of conflicting social spheres: effects of network structure on experienced tension in social network sites. *Proc. 27th International conference on Human factors in computing systems* ACM Press (2009) pp. 965-974.

[3] Carrasco, J.A., Hogan, B., Wellman, B., and Miller, E. J. Collecting Social Network Data to Study Social Activity-Travel Behaviour: an Egocentric Approach. *85th Transportation Research Board Meeting*, (2006).

[4] Carrington, P., Scott, J., and Wasserman, S. *Models and methods in social network analysis*. (2005) *ch. 1-3*.

[5] Cross, R. L., Martin, R. D., Weiss, L. M. *Mapping the value of employee collaboration.* The McKinsey Quarterly 2006, 3 pp. 28-41.

[6] Diesner, J., Frantz, T. L., and Carley, K. M. Communication Networks from the Enron Email Corpus "It's Always About the People. Enron is no Different". *Computational & Mathematical Organization Theory* 11(3), (2005) pp. 201-228.

[7] Ellison, N. B., Steinfield, C., and Lampe, C. The benefits of Facebook "friends:" Social capital and college students' use of online social network sites. *Journal of Computer-Mediated Communication*, 12(4), (2007) pp. 1143-1168.

[8] Facebook Terms of Service. http://wiki.developers.facebook.com/index.php/Platform_Policy_Overview.

[9] Freeman, L. C. Visualizing Social Networks. *Journal of Social Structure* 1, (2000).

[10] Hampton, K. N. Computer Assisted Interviewing: The Design and Application of Survey Software to the Wired Suburb Project. *Bulletin de Methode Sociologique (BMS)*, 62 pp. 49-68 (1999).

[11] Heer, J, Card, S. K., and Landay, J. A. prefuse: A Toolkit for Interactive Information Visualization. *Proc. CHI 2005*, ACM Press (2005).

[12] Hogan, B., Carrasco, J. A., and Wellman, B. Visualizing Personal Networks: Working with Participant-aided Sociograms. *Field Methods* 19, 2 (2007), pp. 116-144.

[13] Gerich, G., Lehner, R. Collection of Ego-Centered Network Data with Computer-Assisted Interviews. *Methodology* 2, 1 (2006) pp. 7-15.

[14] Gilbert, E., Karahalios, K., Sandvig, C. The Network in the Garden: An Empirical Analysis of Social Media in Rural Life. *Proc. CHI 2008*, ACM Press (2008).

[15] Keila, P.S. and D.B. Skillicorn (2005), "Structure in the Enron Email Dataset," *Proc. of Workshop on Link Analysis, Counterterrorism and Security, SIAM International Conference on Data Mining* (2005).

[16] Granovetter, M. The Strength of Weak Ties, *American Journal of Sociology* 78 (6): May 1973, pp. 1360-1380.

[17] Killworth, P.D., Johnsen, E.C., and Bernard, H. R. Estimating the Size of Personal Networks. *Social Networks*.

[18] Kim, H., Kim, G. J., Park, H. W., and Rice, R. E. Configurations of Relationships in Different Media: FtF, Email, Instant Messenger, Mobile Phone, and SMS. *Journal of Computer-Mediated Communication* 12, 4 (2007).

[19] Lampe, C., Ellison, N., and Steinfield, C. A Face(book) in the Crowd: Social Searching vs. Social Browsing. *Proc. CSCW 2006*, ACM Press (2006), pp. 167-170.

[20] Lampe, C., Ellison, N., and Steinfield, C. Changes in Use and Perception of Facebook. *Proc. CSCW 2008*, ACM Press (2008), pp. 721-730.

[21] Liao, Q. and Li, Q. NetWORK Patterns: Designing Effective User Interfaces for Connections Management at Work. *Proc. CSCW 2008*, ACM Press (2008), pp. 707-710.

[22] McCarty, C., Bernard, H.R., Killworth, P.D., Shelley, G.A., and Johnsen, E.C. Eliciting Representative Samples of Personal Networks. *Social Networks* 19, (1997), pp. 303-323.

[23] Mislove, A., Marcon, M., Gummadi, K. P., Druschel, P., Bhattacharjee, B. Measurement and Analysis of Online Social Networks. *Proc. 7th ACM SIGCOMM conference on Internet measurement*. ACM Press (2007), pp. 29-42.

[24] Nardi, B. A., Whittaker, S., Isaacs, E., Creech M., Johnson, J., and Hainsworth, J. Integrating Communication and information Through ContactMap. Communications of the ACM, 4, 45 ACM Press (2002), pp. 89-95.

[25] Smith, M. A., Kollock, P. *Communities in cyberspace*. Routledge NY 1999 pp. 195-218.

[26] Wasserman, S., Faust, K. Social Network Analysis: Methods and Applications. Cambridge University Press, UK (1994).

[27] Wellman, B. For a social network analysis of computer networks. *Proc. SIGCPR/SIGMIS 1996*, ACM Press (1996), pp. 1-11.

[28] Wellman, B., Wortley S. Different Strokes from Different Folks: Community Ties and Social Support. *The American Journal of Sociology*, 96, 3 (1990), pp. 558-588.

[29] Whittaker, S., Jones, Q, Nardi, B., et al. ContactMap: Organizing communication in a social desktop. *ACM Transactions on Computer-Human Interaction* 11, 4 (2004), pp. 445-471.

[30] Whittaker, S., Jones, Q., and Terveen, L. Contact Management: Identifying Contacts to Support Long-Term Communication. Proc. *CSCW 2002*, ACM Press (2002), pp. 216 - 225.

[31] Whittaker, S., Terveen, L., Hill, W., and Cherny, L. (1998). "The dynamics of mass interaction." *Proceedings of the ACM's Conference on. Computer Supported Cooperative Work.* (CSCW 1998). Seattle, WA, USA. ACM Press.

[32] Zilli, A., Laubacher, R., Gloor, P. A. E-mail May Not Reflect The Social Network. North American Association for Computational Social and Organizational Science (July 2006).