Evaluating the Paper-to-Screen Translation of Participant-Aided Sociograms with High-Risk Participants

Bernie Hogan
Oxford Internet Institute
University of Oxford
bernie.hogan@oii.ox.ac.uk

Joshua R. Melville
Oxford Internet Institute
University of Oxford

Gregory Lee Philips II,
Patrick Janulis
Feinberg School of Medicine
Northwestern University

Noshir Contractor
Kellogg School of
Management
Northwestern University

Brian S. Mustanski,
Michelle Birkett
Feinberg School of Medicine
Northwestern University

ABSTRACT

While much social network data exists online, key network metrics for high-risk populations must still be captured through self-report. This practice has suffered from numerous limitations in workflow and response burden. However, advances in technology, network drawing libraries and databases are making interactive network drawing increasingly feasible. We describe the translation of an analog-based technique for capturing personal networks into a digital framework termed *netCanvas* that addresses many existing shortcomings such as: 1) complex data entry; 2) extensive interviewer intervention and field setup; 3) difficulties in data reuse; and 4) a lack of dynamic visualizations.

We test this implementation within a health behavior study of a high-risk and difficult-to-reach population. We provide a within–subjects comparison between paper and touchscreens. We assert that touchscreen-based social network capture is now a viable alternative for highly sensitive data and social network data entry tasks.

Author Keywords

Participant-aided sociograms; Social networks

ACM Classification Keywords

D.2.1 Requirements/Specifications: Elicitation methods; E.1 Data Structures: Graphs and networks; H.5.3 Group and Organization Interfaces: Web-based interaction

INTRODUCTION

The rise of social media and large scale social data sources has led to significant interest in the collection and analysis of social network data [8]. Much of these data comes from behavioral actions as encoded on platforms, such as Facebook likes, emails sent and messages retweeted. Such 'trace' data can reveal insights about human behavior. However, there are numerous situations when no single passive source of data will be able to capture the specific networks of interest. In particular, the sensitive health behaviors of populations, such as sexual practices and the use of drugs, cannot be reliably captured through trace or non-reactive data. In these cases, self-report data are essential [7].

Since disease transmission often includes a social dimension, health researchers in particular have substantial interest in capturing social network data to develop epidemiological understanding of disease and to inform intervention development. Unfortunately, few data capture tools are designed for such uses. Those that do exist have not kept pace with advances in data management or user interface design. Most social network data capture still requires tedious analog surveying [15]. While some creative solutions have been developed [23, 43], few tools are capable of efficiently capturing complex network data. This is largely due to placing an overly high burden on participants, researchers, or both.

Several technological advancements — in particular the recent emergence of cost-effective capacitive touch screen displays, the increasing power of standardized web technologies, and the development of graph databases — make it plausible to translate the prior analog approach into a digital framework. In theory, this digital translation should significantly improve the usability and decrease the complexity of network data capture for both respondents and researchers.

This paper describes our effort to address the challenges presented by the state of the art in both analog and digital tools for social network capture. We describe the use of a novel open source framework, *netCanvas*[1], built using many of the technological advancements noted above. We apply *netCanvas* to a research study of young men who have sex with men (YMSM), a hard to reach and highly sensitive population, using a specific protocol called *netCanvas-R*. This protocol

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than ACM must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from Permissions@acm.org.
CHI '16, May 07–12, 2016, San Jose, CA, USA
ACM 978-1-4503-3362-7/16/05...$15.00
DOI: http://dx.doi.org/10.1145/2858036.2858368

[1] http://networkcanvas.io

was designed to reflect existing analog approaches to data collection and to handle complex multi-layered (i.e., *multiplex*) social, sexual, and drug use networks. We evaluate the efficacy of this protocol compared to a prior analog data capture technique using many of the same participants. When conducting clinical research studies, particularly longitudinal research with vulnerable populations, high respondent satisfaction and ease of use are essential to ensure participant retention and the capture of high quality data. We find that our technique not only captures equivalent data more efficiently than analog techniques, but provides high levels of user satisfaction.

RELATED WORK

Surveying personal networks

Surveying has been an essential cornerstone of social research for over a century. In that time, there have been innumerable studies on methodological advances as well as research on the transition towards computer-based surveying [16]. Social network studies, however, have proven to be especially challenging for existing survey techniques.

For many surveys, there are few dependencies between questions beyond straightforward skip-flow logic. One of the key distinctions between social network surveys and traditional approaches is the need to have alters and relationships persist throughout the survey. In social network visualization it is conventional for people mentioned by a respondent to be represented as nodes, and relationships between them as edges. By treating nodes and edges as discrete classes of objects, these visualization packages not only allow for the drawing of networks, but higher order calculations such as community detection and network layout. The *Prefuse* package, for example, contains methods to automatically arrange nodes using common social network layouts such as Fruchterman Reingold [22].

Most existing survey programs do not have facilities for representing ego (i.e. the respondent) and alters as discrete entities. Instead, alters are reduced simply to names that are piped between questions using a skip-flow logic. Responses to questions concerning network actors are therefore represented as simple fields in the survey data model, and not a special class of object with relational properties (as would be especially useful in managing the collection of social network data).

Here, we discuss the fundamental process of surveying social network data, and link this to advances in social network visualization. While these two practices (data collection and visualization) have often been considered separate, we believe that as research methodologies become more advanced, data collection will necessarily become increasingly intertwined with effective user-centered design.

Name generators as data capture devices

Name generation has been one of the cornerstones of social network analysis (SNA) since its early inception in Moreno's sociometry in the 1930s [19]. In the 1960s, the emphasis shifted from reporting on simply 'who knows who' towards the use of targeted name generators within a sample. In one of the earliest incarnations, the Detroit Area Study [29], name generators were used among a sampled population to assess the breadth and diversity of social support networks. Participants were asked to list N names before completing a 'dyadic census', wherein they systematically indicate any relationships between alters (does A know B, does A know C, does A know D, etc.).[2] This approach has been reproduced with some modifications in numerous sociological studies [45, 17, 10, 33].

The "name generator" approach to social network data collection is comprised of three stages: Name generating, name interpreting, and edge generating.

Name generating is the first stage in network data capture, in which respondents are asked to name members of their network. While the wording of prompts ranges between studies, the boundaries for who is a member of the social network will vary based on the study. There is no universal statement on who should be included or excluded from a social network [31].

Name interpreting is when respondents give additional data about their alters indicated in the prior stage. Name interpretation follows name generation rather than operates in parallel in order to minimize respondent bias. Examples of measures administered during this stage might include alter demographics, ratings of the strength of the relationship between the ego and alter, and the frequency of contact between ego and alter.

Edge generating refers to the respondent providing relational data linking the alters. It is the most complex stage, and the one that has led to substantial differences in research methods. The complexity is due to the fact that for n alters there are $n(n-1)$ possible directed relationships (such as who calls whom) and $\frac{n(n-1)}{2}$ undirected relationships (such as who spends time with whom). In the case of *multiplex* networks, there are even more links. In our study, numerous link types are assessed between alters, such as those who spend time together, use drugs together, have sex with each other, and participate in group sex together. Thus, the complete set of links would rise to $\frac{4n(n-1)}{2}$. As mentioned previously, the earliest method of capturing these data was through a 'dyadic census'. For even 20 alters, a dyadic census requires 190 questions. Responses to this challenge tend to cluster into two camps: visual analog approaches and computer-assisted interviews.

Visual analog approaches

Visual approaches are mentioned by Freeman [20] as a cornerstone of the SNA paradigm. Historically, visualization has provided a means for researchers to view and interact with complex network data, allowing the generation of intuitions and the visual representation of macro-level phenomena. However, when displaying a large number of nodes or relationships, graphs quickly become too complicated to easily interpret, or to be used to develop clear understandings of micro-level structural details. Visualization has been particularly useful, therefore, in studies of close personal networks,

[2] As is convention in social network analysis, we refer to relations (e.g., friendships) between ego and alter are as ties or *ego-alter* ties. The relations between any two alters are referred to as *alter-alter* ties.

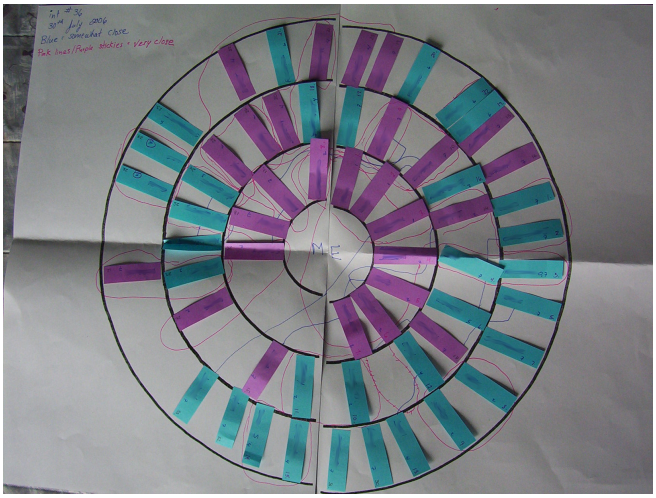


Figure 1. Particpant-aided sociogram from the Connected Lives project. (Previously unpublished image ©Wellman Associates 2005.)

where there are fewer nodes and linkages.[3] In the case of studies that involve the participant in interpreting the visualization, the effectiveness is often further enhanced by their inherent expertise regarding the data [24].

Visual analog approaches in personal network studies begin with visual representations that are built upon the 'scaffolding' of a series of concentric circles around a focal individual. Beginning in the 1980s, the concentric circle approach to network collection was utilized repeatedly by researchers such as Antonucci [3] and Spencer and Pahl [35]. While the arrangement of the nodes was considered important in this work, these studies did not collect alter-alter relational information.

In 2004, Wellman's NetLab extended data collection using a concentric rings approach by adding manipulated physical elements (in this case, post-it notes) that made the interview interactive, tactile, and most importantly, *participant focused* [23]. These physical and interactive elements not only improved the participant experience, but also made the capture of alter-alter relational information more feasible. They titled this approach a 'participant-aided sociogram' (PAS; Figure 1)[23]. Their methods have been subsequently applied in a number of studies, including those examining vulnerable populations [26, 5, 34].

Challenges with current visual analog approaches

Although the recent approaches to PAS networks are highly intuitive for respondents, it remains tedious to code the structural data produced in PAS interviews (e.g. lines drawn in pen between two analog elements). Sometimes lines bend, intersect, or merge in ways that make data imputation slow and prone to errors in dense networks. Further, errors in the imputation of alter names may cause serious issues, particularly in studies interested in matching and consolidating alters across networks [6]. Somewhat counter-intuitively, analog interfaces can quickly become rigid and inflexible once their initial 'openness' is collapsed by the permanence of ink-drawn lines. As an example, consider research examining multiplex networks. On a sheet of paper, one must either erase any existing ties (thus eliminating data) before moving on, or else use different colored pens, thereby multiplying the visual complexity.[4]

As a further illustration, consider a scenario where a question must be asked about an alter with certain qualities. For the interviewer to locate 'the alter with the highest degree' might require tedious manual counting. For a computer this counting is trivial. By extension, there is the matter of sampling within a network. That is, what is the best way to take a subset of alters? The complicated sampling of [12] suggests automation could be an improvement.

Computer-assisted approaches

While analog approaches focus on ease of use and user empowerment, they suffer from complex data collection and a lack of data flexibility *in situ*. Alternative computer-assisted approaches, of which there are now a number, largely bypass these issues through an emphasis on creating structured, manageable data. They range from adaptations of the most direct and generalist survey tools (which simply ask for N alters and then iterate through the dyadic census automatically), to complex real-time network creation as is done in EgoWeb [39].

One of the earliest noted digital capture techniques is McCarty's EgoNet [30]. This tool functions primarily as a survey tool for capturing network data. It is built in Java and requires extensive configuration, but can show real-time features: as one nominates new alters, they appear on a graph. These alters then become connected to the rest of the graph through alter-alter questions that are asked to the respondent. Other tools, such as VennMaker, also utilize a Java-based framework for drawing nodes and edges, and similarly allow the creation of real-time representations [21].

Each of these tools (see Figure 2 for screen-shots of existing digital approaches) were built with small screens and keyboard and mouse interaction in mind. While in-keeping with the technology available at the time, this is at odds with human-computer interaction (HCI) literature indicating that movement and interaction with certain elements is better accomplished with a finger [18, 11].

Compared to analog approaches, computer-assisted approaches front load technical demands. That is, paper-based approaches can be very flexible until before the interview starts. By contrast, computer-assisted approaches require immense specification, both in terms of the interview protocol and the back-end data structure, prior to the start of data collection. Once these are constructed and data captured digitally, it is possible to automate any further processes, such as the act of converting data from its existing format to one amenable to analysis. This is particularly advantageous for longitudinal and ongoing data capture. Removing the time necessary for data entry and validation, while being able to store, analyze,

[3] Nevertheless, work continues to assess how large scale networks can be arranged and visualized in useful ways [42].

[4] See [26] for an example of this process of redrawing multiple networks on a whiteboard.

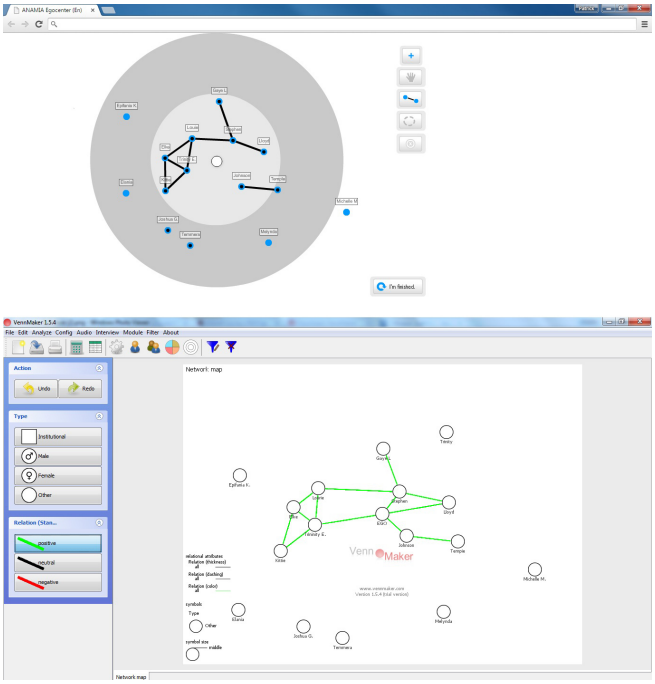


Figure 2. Example digital approaches to the collection of personal networks, Anamia (Top) and VennMaker (Bottom).

retrieve, and query data in near real time is a completely new paradigm in network analysis.

Challenges with current computer-assisted approaches

Despite these advantages, current computational approaches suffer from a series of issues. First they tend to be overly complex in their interface design, predominately as a result of focusing on the requirements of the expert user (often the researcher) rather than the novice (often the research subject). A single screen may be overloaded with controls in order to represent any number of different settings or options. This approach is not without forethought: such interfaces are designed to help an experienced user complete an action they are already familiar with more quickly. An analogy might be made with professional applications in other areas, such as photo-editing or music production, where interfaces often prioritize long-term productivity over short-term intelligibility. VennMaker, for example, functions much akin to a specialized drawing program like OmniGraffle or Visio, with an interface designed around expert requirements rather than co-production.

Though these design goals are reasonable when software is to be used by experienced users, we assert that when interfaces expose novice respondents to complex metaphors, options, and visual representations, they can intimidate and confuse. Tufte's notion of the 'data-ink ratio' [44] seems pertinent: interfaces should in general seek to minimise the redundant 'data ink' used merely to frame the relevant data. Re-enforcing this, past work in human factors has evaluated a number of means for simplifying software, e.g. McGrenere and colleagues demonstrating the benefits of showing fewer elements to novice users [32].

Most existing approaches present interfaces that assume a multi-tasking environment. A review of existing software shows that virtually none offer a full-screen (or single task) experience, or are intended to be used on a large format screen; this severely restricts the space available for interface elements, and makes drawing tasks potentially distraction prone and confusing. Some programs sidestep this issue by not having the sociogram drawn by the participant at all. For example, EgoWeb 2.0 only shows the sociogram after the dyadic census. In this case, the network is drawn algorithmically rather than being co-created. For the designers of EgoWeb 2.0, this was a deliberate decision, since the sociogram is meant primarily as a conversational object rather than one for tie generation. We consider this a design decision rather than a methodological necessity.

The Anamia software by Tubaro and colleagues promotes the sociogram to the status of an interface in its own right, and also operates in a quasi-full screen mode (with only the web-browser and operating system façades still present outside of the software itself) [43]. However, network generation, layout, and edge creation tasks all occur on the same interface, which is restricted by the use of Adobe Flash technology to a predefined screen-size. When combined with the need for the software to be used by an unassisted respondent remotely, all interface elements for all tasks must be visible at the same time, which reproduces many of the issues associated with congested interfaces noted above.

These approaches are also hindered by back-end inflexibility. Much software is intimately tied to a specific data model, often developed for the requirements of a specific study or research exercise. For example, many software requires ties to be uniplex (of only one type), does not allow complex skip-flow logic, or does not allow assignment of complex node attributes such as overlapping group membership. While some software provides a degree of configurability with regards to the data model, this happens primarily via bespoke GUI wizards, or arcane configuration files with proprietary syntax. This limits the type and complexity of data that are able to be captured, as well as the ability to use the software within varied research contexts.

One final challenge with current computer-assisted methods stems from the fact that once data have been captured, existing software is rarely designed for the straightforward aggregation or management of that data. By its nature, network data produces both structural data (ties between nodes) and attribute data (attributes of nodes; attributes of ties).

Storing this data is far from straightforward, yet existing software has been slow to adapt to more recent approaches, and has tended either to defer to notions inherited from relational databases, or to fall back on the even more outmoded system of flat files. If using relational databases, the researcher is often required to create complex queries across many tables to reconstitute the graph from its component parts during analysis. In the case of flat files, equally extensive file manipulation is required.

DESIGN GOALS: A HYBRID APPROACH

The approach taken in our current research has been to draw upon a new framework, *netCanvas*, which seeks to address prior shortcomings of both analog and computer-assisted approaches by using web technologies and a touch-optimized interface. The overall aim of this framework is to replicate as much of the front-end of analog techniques as possible while still preserving the data quality and data flexibility advantages of current digital approaches. Specific requirements which guided the design of this tool were:

1. Similar to PAS techniques, and to improve usability, generated alters should be treated as 'objects' which can be moved and shuffled by the respondent.

2. Similar to PAS techniques and counter to many computer-assisted approaches, data entry must be tactile.

3. Drawing from the HCI literature, interface elements should be designed for optimal respondent ease, engagement, and comprehension. This means simple and colorful displays that are free from excess radio buttons, check boxes, and any elements that are not of use to the interviewee, and that could potentially be distracting[41].

4. The screen should be able to accommodate upwards of 50 nodes clearly without occlusions. We opt for 50 nodes since the literature on personal networks indicates this characterize over 95 percent of personal networks according to past studies[23, 17].

5. Finally, the *netCanvas* framework must be flexible in order to work with multiple diverse protocols that can be specified by various research groups.

While *netCanvas* is the framework, the protocol file defines the specific items, screens, and logic. This paper therefore focuses on the evaluation of one such protocol, *netCanvas-R*.

RESEARCH HYPOTHESES

To evaluate the effectiveness of our implementations of the above requirements, we present four hypotheses:

Hypothesis 1. Users who have completed both an analog-based PAS and *netCanvas-R* will report either equal or more alters on *netCanvas-R*. Further, participants will disclose equal or more information about sexual behaviors with each alter within *netCanvas-R*.

Hypothesis 2. Users who have completed both an analog-based PAS and *netCanvas-R* will be faster on the digital platform.

Hypothesis 3. There will be no evidence that efficiency gains will be due to a learner effect about 'social network thinking'. That is, those who have done a name generator in a different medium will take as long as those who have never done a name generator before.

Hypothesis 4. Users will indicate satisfaction and high usability of *netCanvas-R*. This will include expressions of enjoyment with the data collection process.

These first two hypotheses suggest that *ceterus paribus*, *netCanvas-R* is a preferred approach as it will lead to either no difference or higher levels of disclosure about the social network. Thus, if data quality is either unaffected or improved using *netCanvas-R*, this validates our choice to translate this tool digitally. In addition to these hypotheses, a fourth hypothesis is related to the design and experience of completing the digitized PAS.

METHODS

The design and implementation of the *netCanvas-R* protocol

We use *netCanvas* as a framework that can accommodate the requirements for analog PAS. The framework itself provides a generic set of survey 'screens' (or interfaces) that can be fit to specific tasks, as well as a comprehensive set of APIs for interacting with a network data model and managing an interview session.

To employ *netCanvas*, one must create a 'protocol,' in much the same way as a video game console requires the creation of specific games to run on it. In this case, we implemented a protocol, *netCanvas-R*, that reflects much of the requirements of an analog PAS while also using utilizing novel features like complex skip-flow logic, pan and zoomable maps, and recalling alters between waves that would more complicated within an analog framework.

The interviewer-assisted interview protocol *netCanvas-R* was used to elicit social, drug, and sexual connections by the participant. Below we outline the specifics of how this was implemented, with reference both to the issues with prior approaches set out above, and our stated design goals.

Technological Foundations

The *netCanvas* framework is built entirely using open web technologies, primarily HTML5, CSS, and JavaScript. Historically, these technologies have not been considered sufficiently powerful to create full-fledged applications. As recently as 2011, fundamental interactions, such as scrolling using touch, were unsatisfactory when compared with the 'native' experience provided by applications directly built using operating system APIs [13].

In recent years there have been a number of dramatic improvements in this situation. The ratification of the HTML5 specification introduced APIs for full-screen display, local storage of data, touch events, and hardware-accelerated rasterized rendering using the `<canvas>` element. Of particular importance for an implementation of the analog PAS method are touch interactions and canvas rendering, since, when combined, they allow the arbitrary creation and manipulation of digital 'objects', via touch (or multi-touch) interfaces, wholly within the browser. This functionality gives developers a means to create programmable platform-agnostic applications without choosing between anachronistic UI frameworks (e.g., NetBeans) or proprietary and soon-to-be obsolete technologies such as Flash. Similarly the CSS3 specification has provided a raft of powerful new abilities. Among them is the ability to animate any property of an HTML element with arbitrary or user-defined "easing" functions. These functions allow designers

total control over the movement and feel of interface elements, allowing complex and meaningful interaction design. Finally, JavaScript is now robust and performant enough to be used for large-scale applications. The 'Node' server-side JavaScript framework, for example, handles high volume websites such as flickr.com, groupon.com, and yellowpages.com.

This choice of technologies puts *netCanvas*, and by extension *netCanvas-R*, in a unique position compared with contemporaries: it is capable of running across all modern web-browsers on all modern platforms (from Windows, OSX, and Linux, to Apple's iOS platform and Google's Android). While not necessarily advantageous in the context of a single-instance study such as is discussed here, this approach reflects a broader commitment by the authors to create reusable and general purpose academic software, which can be configured to suit specific needs.

Protocol specification

The *netCanvas-R* protocol itself is defined within a protocol file, which is loaded by *netCanvas*. This file consists of a simple and logical JSON schema (See Listing 1) familiar to anyone who has experience with JavaScript syntax. Since the protocol is defined in a clear-text format, it is directly editable by researchers, albeit with the caveat that they must understand the syntax. Still, this represents an improvement over black-box GUI configuration, or proprietary custom configuration files. Our use of open formats also allows the creation of additional tools for developing and managing protocol schemes in the future.

```
var protocol = {
  sessionParameters: {
    name: "RADAR"
  },
  stages: [
    {label:'Introduction', page:'intro.html'},
    {label:'Name Generator', page:'ng.html'},
    {label:'Alter Linking', page:'sociogram.html'},
    {label:'Name Interpreter', page:'ordbin.html'},
    // etc.
  ]
};
```

Listing 1. Example netCanvas protocol file

The protocol is defined as a series of discreet 'stages', with each stage manifest in a single self-contained HTML file. Conceptually, we aim to break the interview process up into logical single-step tasks. This dramatically simplifies the interface requirements for each screen, and establishes a predictable pattern that allows respondents to quickly understand the flow of the interview.

The use of the *netCanvas* framework also enables complex automatic skip logic (also defined in the protocol file) based on data collected from the current session, or even previous visits. This skip logic extends traditional survey programs where skip logic is based solely on responses to specific questions. Since we treat nodes and edges as special classes (creating an underlying network data model during the interview), *netCanvas* operates more like existing social network visualization software than either existing visual network data entry programs like VennMaker or traditional survey programs such as Qualtrics. For example in the case of the GUESS package, one can use

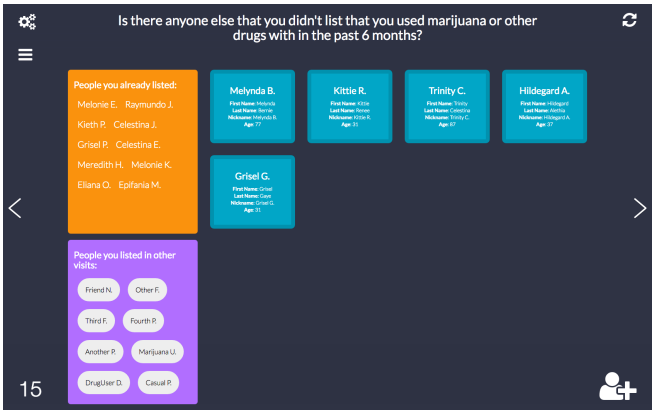


Figure 3. Name generator interface

the modified python (termed gython) to make boolean and set logic queries to nodes and edges[1]. In our case, a specific protocol can make structural queries to the network data model using the netCanvas API. This can be used to, for example, show a screen only if a fully connected clique is present (such as a fully connected set of sex partners).

The *netCanvas-R* protocol embeds automatic skip logic to provide several key benefits. First, by accounting for a participant's prior responses, later items that are deemed irrelevant are able to be skipped, and the length of the survey is greatly reduced. While *prima facie* this is conventional in survey programs, the difference is that we provide a skip logic based on the qualities of nodes. By contrast, fields in conventional surveys do not themselves have queryable attributes. For example, in our case, we can accommodate sensitive issues with gendered pronouns in a straightforward manner that would otherwise require extensive JavaScript hacks in programs such as Qualtrics or LimeSurvey (considering that our participants may be included for having sex with men but identify as either transwomen or genderqueer in addition to the conventional male gender). Although skip logic seems like an obvious inclusion, it is not as easy to permit in studies where all data collection aspects happen on a single screen.

Interfaces and affordances

In keeping with the steps required by the PAS method, *netCanvas* provides predefined interface modules for the three most important steps: name generation, name interpretation, and alter-alter linking on the sociogram. Each stage defined in the *netCanvas-R* protocol file implements one of these interfaces, with simple text fields changed to reflect different prompts, variable names, or other configuration options.

Name Generation (Figure 3) takes place on a simple card-based interface, showing only a prompt, a counter to keep track of alter numbers, and a button for triggering the 'add new alter' panel. Only those interface elements that are necessary for the particular task are shown, with context-specific panels appearing and disappearing as necessary. For example, triggering the new alter form dims all background interface elements, giving the form dramatically increased visual emphasis. Figure 3 demonstrates this minimalist and context-driven ap-

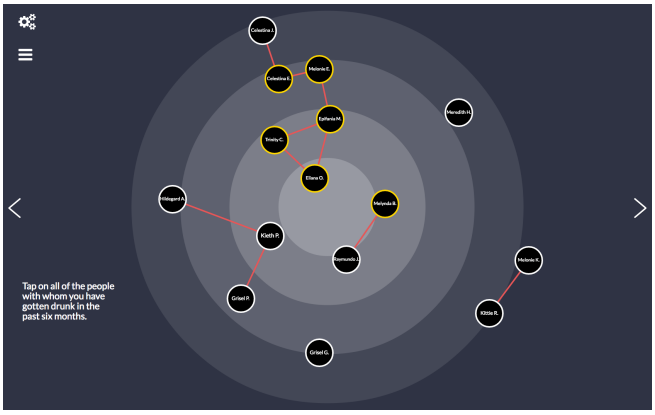


Figure 4. Alter-alter linking (red lines) and alter nomination (yellow highlights)

proach: UI panels are displayed showing previously elicited alters (orange), or alters from a previous interview visit (purple) entirely dynamically, based on the presence of the data they are designed to display. Thus, if a panel has no data to display, it disappears, thereby creating more room for other interface elements to expand into. This process occurs entirely automatically, with no interviewer or respondent involvement.

As with all *netCanvas-R* interfaces, interaction with the name generator interface occurs through touch, gestures, and keyboard for text entry. Since the keyboard is separate from the touchscreen, it can be used by the interviewer or the participant depending on the mobility needs of the participant. Tapping on buttons or cards (representing alters) triggers predictable actions such as opening, editing, or activating. More complex interactions include the ability to drag alters elicited in previous visits into the current interview session. This gesture is both symbolic and practical: it is designed to give participants the sensation of bringing an individual into the present context, while re-enforcing our stated design goal of making alter interactions as tactile as possible. Extending this, we do not use hidden or unconventional user actions, such as long-press for context menus.

The use of longitudinal data in these panels also demonstrates a technique that would have been impractical from a data-management perspective using an analog approach, and formidable from an interface perspective using an existing computational approach.

Within *netCanvas-R*, name generator steps successfully implemented in the earlier LYNC Study were adapted and implemented using the core *netCanvas* interfaces. As in LYNC, the alter's first, last, and nicknames were elicited for each step, along with ego's perception regarding their role(s).

Alter-alter linking (Figure 4) is an example of an interface that benefits particularly from digitization. To indicate a tie, the participant first taps an alter (which is highlighted), and then taps another alter. At this point, an edge is created between the alters, which is anchored to them should either be moved later.

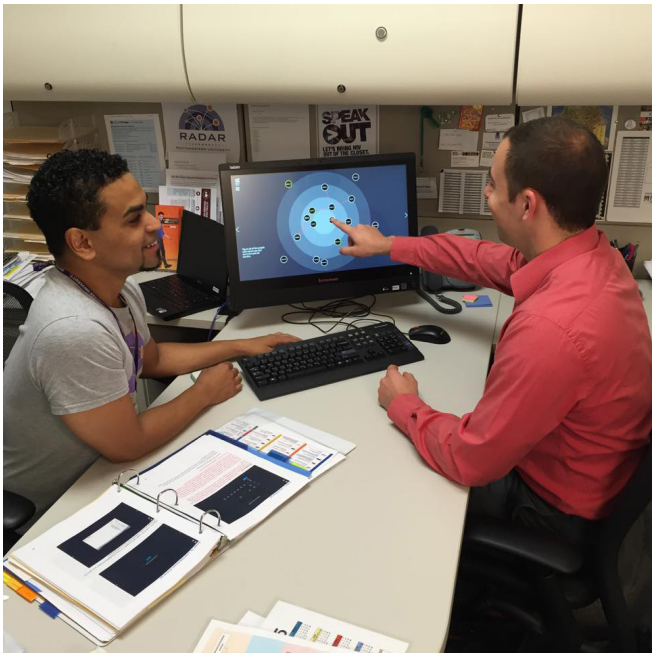


Figure 5. netCanvas-R in situ

In *netCanvas-R* we leverage the flexibility of showing and hiding ties in order to look at three distinct kinds of ties: sexual, social, and drug use.

Finally, **name interpreter** steps take place on one of three interfaces, depending on the type of variable being captured.

For ordinal and nominal variables, we devised a gesture-based binning interface, which allows respondents to quickly categorize alters based on a given prompt. The ordinal and nominal binning interfaces are used within the *netCanvas-R* protocol to collect data on contact frequency, relationship strength, sexuality, gender identity, racial identity, coarse (region level) location, and drug use frequency.

For binary attributes (primarily nomination variables such as 'got drunk with' or 'got advice from') we used the concentric circles interface, drawing again on the analog PAS approach (Figure 4). The participant selects an alter as having a given attribute through a simple tap, which results in it being visually highlighted.

Several other key variables related to the overall aims of the project were also collected, including important attributes of ties (e.g., estimated dates of first and most recent sex), and an interactive map in which alter residence was coded by US Census tract.

Research context

netCanvas-R was designed to understand the social, sexual, and drug use networks of YMSM. YMSM are an important target for HIV research due to their alarming HIV prevalence, as they accounted for 67% of infections in youth in 2008 and increased to 80% in 2013 [2]. They are one of the only risk groups showing an increasing rate of infections [38]. Despite

this high prevalence, and despite HIV's high transmission dependence on drug and sexual network dynamics [25, 27], few studies have been able to efficiently capture social, sexual, and drug network data with this population [25].

Data come from two projects focused on understanding the social, sexual, and drug use networks of YMSM. The first is LYNC, an existing study on sexual health, drug use, and networks that collected data from 2011 to 2012. *netCanvas-R* was introduced in the second study, RADAR (hence the use of the '-R' suffix), which began collecting data in 2015. Participants enrolled in LYNC were subsequently invited to join the RADAR study, which allowed a natural experiment where the reliability of the touchscreen interface could be tested.

LYNC

Within the LYNC study, analog SNA interviews were conducted with 175 YMSM who were a part of a larger longitudinal cohort study aimed at characterizing the prevalence, course, and predictors of co-occurring health problems among YMSM in Chicago. To enroll in the parent study, participants must have been assigned male sex at birth, been between the ages of 16 and 20 years at enrollment, resided in the Chicago metropolitan area, and either identified as gay/bisexual/queer or reported sex with another man. SNA interviews described the sexual, support, and drug use networks of YMSM via a method modified from [23]. Network visits occurred between June 2011 and October 2012.

The analog interviews relied on a combination of two interviewer administered procedures to collect these data: 1) completion of a pre-numbered list form (i.e., in paper-and-pencil form) to enumerate alters and to capture alter attributes; and 2) PAS to capture respondent report of interactions between alters (sexual, substance-using, and social networks). Name generators as well as alter characteristic elicitation questions and procedures were based on prior studies of populations at risk of HIV infection [4, 28, 36, 37, 40]. Egos were asked to elicit up to 40 alters on a list that could be moved around on a whiteboard. After the initial list of supportive individuals was generated, participants were then asked which individuals on that list they had "used drugs or alcohol with" or "had sex with." Then they were asked to name anyone else that they had not yet listed that they had "used drugs or alcohol with" or "had sex with." Finally they were asked if there was anyone that they had not listed yet that has "used drugs or alcohol" or "had sex" with at least two the people already mentioned.

RADAR

The second study, RADAR, is an ongoing longitudinal study of a cohort of YMSM in Chicago. Participants came from a combination of recruitment from existing research cohorts of YMSM, venue-based recruitment, and peer referral. To enroll, participants must have been assigned a male sex at birth, been between the ages of 16 and 29 years, and either identified as gay/bisexual/queer or reported sex with another man. Study visits began in 2015 and were comprised of three sections — a network interview (netCanvas-R; Figure 5), a self-report psychosocial interview (H-RASP), and biological testing. The research activities conducted under both RADAR and LYNC were approved by the Northwestern University Institutional Review Board.

The proposed analysis

To evaluate the quality of data captured within *netCanvas-R*, research hypotheses 1-3 will be assessed via two comparisons. First, to allow the most direct comparisons of an analog PAS to the *netCanvas-R* protocol, the subset of the participants recruited into RADAR who had also participated in the LYNC study ($N = 67$) will be analyzed. The number of alters who were indicated to be sex partners and drug partners within LYNC will be compared to data captured later in RADAR. Additionally, to further evaluate research hypothesis 1, the sexual partner information collected by *netCanvas-R* of all current RADAR participants ($N = 280$), will be compared to data captured under self-report via the computerized self-administered HIV-Risk Assessment for Sexual Partnerships (H-RASP) [34]. This instrument evaluates up to 6 sexual partnerships during the 6 months prior to each interview.

To evaluate the efficiency of data capture with *netCanvas-R* versus PAS, interview times will be compared. Comparisons will occur for all participants ($N = 280$) and for all participants minus the subset involved in both LYNC and RADAR ($N = 213$). Research hypothesis 4 will be evaluated by an analysis of two subsets of participants who at the end of the network interview completed surveys assessing *netCanvas-R*. The first subsample ($N = 65$) provided feedback on their experience with using *netCanvas-R* by responding to three Likert scale questions corresponding to three interfaces (network layout, the calendar widget, and the mapping screen) as well as providing qualitative feedback. The second subsample ($N = 60$) completed a modified version of the System Usability Measure [9] indicating their level of support (1 = strongly disagree to 5 = strongly agree) for the measure's usability. To note, our only modification was to change the wording to clarify that the respondent was evaluating the netCanvas application.

RESULTS

Comparing network disclosure across approaches

In order to compare the quality of the data captured in *netCanvas-R*, network size and reported high-risk sexual behaviors were examined across the three approaches: LYNC, which used the analog PAS name generator; *netCanvas-R*, which used the digital name generator; and H-RASP, which asked respondents to provide a single number of sex or drug partners.

Comparing netCanvas-R to LYNC PAS

A total of 67 participants completed both the LYNC network survey and the *netCanvas-R* interview as a participant in RADAR. For these participants, the median number of drug partners in the 6 months prior to the interview was 4 (interquartile range $[IQR] = 7.0$) for LYNC and 4 ($IQR = 7.5$) for RADAR (Wilcoxon signed-rank $V = 644, p = 0.86$). Similarly, the median number of sex partners in the six months prior to the interview was 1 ($IQR = 2.0$) for LYNC and 2 ($IQR = 3.0$) for RADAR (Wilcoxon signed-rank $V = 565, p = 0.35$). These results indicated that equivalent numbers of sex and drug

partners were obtained from *netCanvas-R* compared to analog PAS. Additionally, the number of partners respondents indicated under analog PAS was highly correlated with the number indicated under *netCanvas-R*. Results showed significant associations between both the number of drug partners (Spearman $r = 0.56, p < 0.001$) and the number of sex partners (Spearman $r = 0.63, p < 0.001$) named by each participant across interviews, representing strong effect sizes [14].

Comparing netCanvas-R to H-RASP

Of the 280 participants that completed both the *netCanvas-R* interview and the H-RASP as part of the RADAR study, 179 (63.9%) named the same number of sex partners in both data collection tools, while $57(20.4\%)$ named more sex partners in the H-RASP and $41(14.6\%)$ named more sex partners in *netCanvas-R*. The median number of sex alters named during the *netCanvas-R* interview and the H-RASP was 2 ($IQR = 2$ and $IQR = 3$, respectively), indicating there was no significant difference between the two reports (Wilcoxon signed-rank $V = 2269, p = 0.08$).

Accurate estimates of specific sex behaviors are important for understanding HIV transmission. Therefore reports of sexual behavior with up to 6 of the most recent partners were examined. A comparison of the distribution of sexual behavior reported in *netCanvas-R* and H-RASP can be found in Figure 6. Participants reported similar numbers of anal sex acts (ASA) in H-RASP $(Median = 6, IQR = 23)$ and in *netCanvas-R* $(Median = 5, IQR = 20)$, although this did obtain statistical significance using the Wilcoxon signed-rank test $(V = 6851.5, p = 0.003)$. Conversely, there was not a significant difference (Wilcoxon signed-rank $V = 4212.5, p = 0.24$) in the number of condomless ASA (CASA) reported in the H-RASP $(Median = 1, IQR = 11)$ as compared to *netCanvas-R* $(Median = 1, IQR = 10)$. These results indicated equivalent reports of CASA, but not ASA. However, the number of high risk sexual acts indicated under H-RASP was significantly correlated with the number indicated under *netCanvas-R*. Both the number of ASA (Spearman $r = 0.71, p < 0.001$) and number of CASA (Spearman $r = 0.75, p < 0.001$) were strongly correlated across H-RASP and *netCanvas-R*.

Efficiency during and post-data collection

Despite our results suggesting similarities in data quality, there were stark differences in the time taken for data collection. In a comparison of median times, *netCanvas-R* took participants significantly less time to complete (Wilcoxon signed-rank $V = 35515, p < 0.001$) with the LYNC interview taking approximately 50% longer than *netCanvas-R* (*netCanvas-R*: $Median = 32$ minutes, $IQR = 18.4$; LYNC: $Median = 47$ minutes, $IQR = 36.0$). Furthermore, this difference was unaffected by removing the 67 LYNC participants who also completed *netCanvas-R* (Wilcoxon signed-rank $V = 27105.5, p < 0.001$).

User satisfaction and usability

To assess user satisfaction and usability, we asked a subsample of RADAR participants to provide feedback on their experience with using *netCanvas-R*. The first sample ($N = 65$) were asked to provide a response to three Likert scale questions

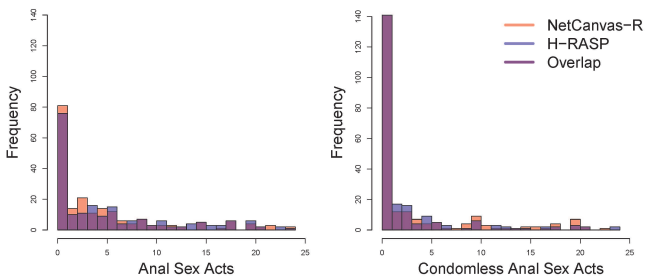


Figure 6. Comparison of distribution of anal sex acts (ASA) and condomless ASA (CASA) in *netCanvas-R* and H-RASP

corresponding to the ease of use (1 = 'very easy' to 5 = 'very hard') of three controls (network arrangement screen, calendar widget, and mapping interface). Across all three controls, scale ratings indicated that respondents found the *netCanvas* system to be very easy to use (network arrangement screen: $Mean = 1.23, SD = 0.52$; calendar widget: $Mean = 1.35, SD = 0.80$; mapping interface: $Mean = 1.72, SD = 1.17$). The respondents were also asked to provide one word descriptions of each of these screens and any suggestions they might have to make the interview more comfortable. Nearly all feedback was strongly positive in nature, with the five most frequently mentioned words being 'Interesting', 'Interactive', 'Fun', 'Simple', and 'Easy.' Further, direct quotes from our query for participant suggestions include: "No, I really enjoyed my time."; "It was pretty fun using the technology!"; "Everything was comfortable"; "Completely comfortable"; and "It all was pretty cool to see visually".

A second sample ($N = 60$) were asked to complete a modified version of the System Usability Measure [9] indicating their level of support (1 = 'strongly disagree' to 5 = 'strongly agree') for the measure's usability. The scale showed good internal consistency ($\alpha = 0.79$), and participants reported strong support of *netCanvas-R*'s usability, with a mean score of 4.45 ($SD = 0.48$).

DISCUSSION

Hypothesis 1. Users who have completed both an analog-based PAS and *netCanvas* will report either equal or more alters on *netCanvas*. Further, participants will disclose equal or more information about sexual behaviors with each alter within *netCanvas*.

Our results support that the data captured from *netCanvas-R* was of comparable quality to the data captured in both the analog PAS and in the traditional individual self report of specific sexual behaviors (H-RASP). Both the number of partners and the specific behaviors reported with these partners was shown to be equivalent. Given these findings, we can assert that the touchscreen-based approach is comparable to existing research techniques with regards to both data quality and the disclosure of sensitive data.

Hypothesis 2. Users who have completed both an analog-based PAS and *netCanvas* will be faster on the digital platform.

While the data are comparable across platforms, the digital platform demonstrated high efficiency. Interviews were substantially shorter within *netCanvas-R* than in LYNC. Using *netCanvas-R* took a little over half the time taken in LYNC. Also, it is important to note that efficiency extends to the data analytic side as well.The digital platform merges seamlessly with the database and analytic technologies. Therefore, once the data workflow had been constructed, analysis was able to occur immediately after interviews. By contrast, the analog PAS within LYNC required 2-3 hours of additional coding per respondent to enter this data and was unable to accommodate longitudinal observations. These efficiency gains mean lower respondent burden, as well as a better user experience, since alters from time $t-1$ can be pulled into the interview at time t with no additional challenges.

Hypothesis 3. There will be no evidence that efficiency gains will be due to a learner effect about 'social network thinking'. That is, those who have done a name generator in a different medium will take as long as those who have never done a name generator before.

Our analyses controlled for prior experience with the analog PAS, and therefore demonstrated that the efficiency gains observed within *netCanvas-R* were not solely due to a learner effect.

Hypothesis 4. Users will indicate satisfaction and high usability of *netCanvas-R*. This will include expressions of enjoyment with the data collection process.

Results consistently demonstrate that users responded positively to the *netCanvas* platform. Across both open-ended responses and multiple scales of usability, the results show that users were happy with *netCanvas-R*, and especially with the act of laying out and interacting with networks. This suggests that a digital approach does not sacrifice user experience.

CONCLUSIONS

In this paper we describe both a general framework for collecting data on touchscreens termed *netCanvas*, which takes inspiration from paper-based techniques as well as from social network visualization frameworks, and an instantiation of this framework using a protocol tested in the field (*netCanvas-R*). This testing allows us to assess the general claim that touchscreen data collection can now be as reliable and far more efficient than paper based methods.

Because of the need for specific and highly sensitive social data unable to be provided via trace information, tools such as *netCanvas-R* are necessary to facilitate individual data capture. These tools hold a special interest for health and infectious disease researchers, who until now have been limited due to the high participant and researcher burden involved in network capture.

Although in [23] the authors indicated reluctance to use digital tools because of their potential inflexibility, *netCanvas-R* has demonstrated strong success in a large study of the health of a high-risk community. By making use of recent technological advancements and building upon valid participant-aided analog techniques, *netCanvas-R* has been able to strongly improve the capture and management of individual network data. The tool we have designed is able to capture complex multilayered social, sexual, and drug use networks. We find that our technique not only captures equivalent data more efficiently than analog techniques, but provides remarkably high levels of user satisfaction. When conducting clinical research studies, particularly longitudinal research with vulnerable populations, high respondent satisfaction and ease of use are essential to ensure participant retention and the capture of high quality data. Therefore, the prior aphorism of "keep high technology in the lab and low technology in the field"[23] be updated to "keep complexity in the lab and simplicity in the field".

Given the sensitivity of the data collection and the complexity of designing the instrument, it was not possible to directly compare the *netCanvas-R* protocol to other digital approaches. That said, none of the computer-assisted approaches reviewed have incorporated HCI research as thoroughly, nor are they designed to meet the complex requirements we noted such as the importing of longitudinal data, multiplex edges, skip-flow logic with multiple dependencies, and storage of the data within a graph database.

While there are many benefits to the current tool, some limitations still exist. For example, despite the successes demonstrated in this analysis, it must be emphasized that *netCanvas-R* is only a tool and does not substitute for strong interviewer training. For example, netCanvas does not limit the number of nodes one can create. Thus, without interviewer intervention adding more than fifty leads to a very cramped interface. We cannot claim *netCanvas-R* will be as successful in a self-administered setting. Additionally, as of yet there is no easy method for researchers to construct surveys using the *netCanvas* framework without being well versed in HTML5 and JavaScript. However, this team is actively pursuing the development of a survey-building wizard for the development of future protocols. Finally, it must be noted that the *netCanvas-R* protocol exports a notably complex nested JSON file that still must be parsed to be inserted into a Neo4j database or other similar graph structure. Yet, the complexity of data management is hidden from the user. By keeping the design of the interface straightforward and linear, we have been able to capture complex network data with a simplified interface and high fidelity.

The *netCanvas* framework and the *netCanvas-R* interview protocol represent a collaboration between a trans-disciplinary team that spans multiple academic disciplines and institutions. By merging technological advances in network visualization with prior participant-aided sociogram techniques, informed by the needs of a high-risk population, we demonstrate the validity, utility, and feasibility of complex data capture for personal social networks.

ACKNOWLEDGEMENTS

This work was supported by the National Institute on Drug Abuse and the National Institute of Allergy and Infectious Diseases of the National Institute of Health (U01DA036939, PI: Mustanski; R01DA025548-S1, PI: Mustanski; R03DA033906, PI: Birkett; K08DA037825, PI: Birkett; P30AI117943, PI: D'Aquila; P30AI117943-S1, PI: Phillips).

REFERENCES

1. Eytan Adar. 2006. GUESS: A Language and Interface for Graph Exploration. *Proceedings of the ACM Conference on Human Factors in Computing Systems (CHI 06)* (2006), 791800.
2. Centre for Disease Control and Prevention. 2015. Diagnoses of HIV Infection in the United States and Dependent Areas, 2011.HIV Surveillance Report, Volume 23. *HIV Surveillance Report* 25 (2015), 1–82. http://www.cdc.gov/hiv/library/reports/surveillance/2011/surveillance
3. Toni Antonucci. 1986. Measuring Social Support Networks: Hierarchical Mapping Technique. *Generations* 10 (1986), 10–12.
4. Colette L Auerswald, Eiko Sugano, Jonathan M Ellen, and Jeffrey D Klausner. 2006. Street-based STD testing and treatment of homeless youth are feasible, acceptable and effective. *Journal of Adolescent Health* 38, 3 (2006), 208–212.
5. Michelle Birkett, L. Kuhns, C. Latkin, S. Muth, and Brian S Mustanski. 2015a. The sexual networks of racially diverse young men who have sex with men. *Archives of Sexual Behavior* Forthcoming (2015).
6. Michelle Birkett, Lisa M Kuhns, Brian S Mustanski, and Stephen Q Muth. 2015b. The creation and analysis of a macro-network of young men who have sex with men. In *35th International Sunbelt Social Networks Conference [SUNBELT '16]*. Brighton, UK.
7. Stephen P Borgatti, Martin G Everett, and Jeffrey C Johnson. 2013. *Analyzing social networks*. SAGE Publications Limited, Thousand Oaks, CA.
8. Stephen P Borgatti, Ajay Mehra, Daniel J Brass, and Giuseppe Labianca. 2009. Network analysis in the social sciences. *Science (New York, N.Y.)* 323, 5916 (2009), 892–895. DOI: http://dx.doi.org/10.1126/science.1165821
9. John Brooke. 1996. SUS-A quick and dirty usability scale. *Usability evaluation in industry* 189, 194 (1996), 4–7.
10. Ronald Burt. 1984. Network Items and the General Social Survey. *Social Networks* 6, 4 (1984), 293–339.
11. William Buxton, Ralph Hill, and Peter Rowley. 1985. Issues and techniques in touch-sensitive tablet input. *ACM SIGGRAPH Computer Graphics* 19, 3 (1985), 215–224.
12. Juan Carrasco, Bernie Hogan, Barry Wellman, and Eric Miller. 2008. Collecting social network data to study social activity-travel behavior: An egocentered approach. *Environment and Planning B* 35, 6 (2008), 961–980. DOI: http://dx.doi.org/10.1068/b3317t
13. Andre Charland and Brian Leroux. 2011. Mobile application development: web vs. native. *Commun. ACM* 54, 5 (2011), 49–53.
14. Jacob Cohen. 1992. Statistical power analysis. *Current directions in psychological science* 1, 3 (1992), 98–101.
15. Nick Crossley, Elisa Bellotti, Gemma Edwards, Martin G Everett, Johan Koskinen, and Mark Tranmer. 2015. *Social Network Analysis for Ego-Nets: Social Network Analysis for Actor-Centred Networks*. Sage Publications, London, UK.
16. Don A Dillman. 2011. *Mail and Internet surveys: The tailored design method–2007 Update with new Internet, visual, and mixed-mode guide*. John Wiley & Sons.
17. Claude S Fischer. 1982. *To Dwell among Friends: Personal Networks in Town and City* (1 ed.). University Of Chicago Press. 459 pages.
18. Clifton Forlines, Daniel Wigdor, Chia Shen, and Ravin Balakrishnan. 2007. Direct-touch vs. mouse input for tabletop displays. *Proceedings of the SIGCHI conference on Human factors in computing systems - CHI '07* (2007), 647–656. DOI: http://dx.doi.org/10.1145/1240624.1240726
19. Linton C Freeman. 2000. Visualizing Social Networks. *Journal of Social Structure* 1, 1 (2000), 1–20.
20. Linton C Freeman. 2004. *The Development of Social Network Analysis: A Study in the Sociology of Science*. Empirical Press, Vancouver, BC.
21. Markus Gamper, Michael Schönhuth, and Michael Kronenwett. 2012. Social Networking and Community Behavior Modeling : Qualitative and Quantitative Measures. In *Social Networking and Community Behavior Modeling : Qualitative and Quantitative Measures*, Maytham Safar and Khaled A. Madhi (Eds.). IGI Global, Hershey, PA, 193–213.
22. Jeffrey Heer, Stuart K Card, and James A Landay. 2005. Prefuse: a toolkit for interactive information visualization. In *Proceedings of the SIGCHI conference on Human factors in computing systems*. ACM, 421–430.
23. Bernie Hogan, Juan Carrasco, and Barry Wellman. 2007. Visualizing Personal Networks: Working with Participant Aided Sociograms. *Field Methods* 19, 2 (2007), 116–144.
24. Grace YoungJoo Jeon, Nicole B. Ellison, Bernie Hogan, and Christine Greenhow. 2016. First-Generation Students and College: The Role of Facebook Networks as Information Sources. In *Proceedings of the 2016 ACM Conference on Computer-Supported Cooperative Work and Social Computing [CSCW '16]*. San Francisco, CA.
25. Blair T Johnson, Colleen A Redding, Ralph J DiClemente, Brian S Mustanski, Brian Dodge, Paschal Sheeran, Michelle R Warren, Rick S Zimmerman, William A Fisher, Mark T Conner, and Others. 2010. A network-individual-resource model for HIV prevention. *AIDS and Behavior* 14, 2 (2010), 204–221.
26. Lisa M Kuhns, Michelle Birkett, Brian Mustanski, S. Q. Muth, C. Latkin, I. Ortiz-Estes, and R. Garofalo. 2015. Methods for Collection of Participant-aided Sociograms for the Study of Social , Sexual and Substance-using Networks Among Young Men Who Have Sex with Men. *Connections* 35, 1 (2015). DOI: http://dx.doi.org/10.17266/35.1.1

27. Carl A Latkin, Danielle German, David Vlahov, and Sandro Galea. 2013. Neighborhoods and HIV: a social ecological approach to prevention and care. *American Psychologist* 68, 4 (2013), 210.

28. C A Latkin and A R Knowlton. 2005. Micro-social structural approaches to HIV prevention: a social ecological perspective. *AIDS Care* 17, S1 (2005), 102–113.

29. E.O. Laumann, J. Galaskiewicz, and P.V. Marsden. 1978. Community structure as interorganizational linkages. *Annual Review of Sociology* 4 (1978), 455–484. http://www.jstor.org/stable/2945978

30. Christopher McCarty and Sama Govindaramanujam. 2005. A Modified Elicitation of Personal Networks Using Dynamic Visualization. *Connections* 26, 2 (2005), 61–69.

31. Tyler H. McCormick, Matthew J. Salganik, and Tian Zheng. 2010. How Many People Do You Know?: Efficiently Estimating Personal Network Size. *J. Amer. Statist. Assoc.* 105, 489 (mar 2010), 59–70. DOI: http://dx.doi.org/10.1198/jasa.2009.ap08518

32. Joanna Mcgrenere, Ronald M Baecker, and Kellogg S Booth. 2002. An evaluation of a multiple interface design solution for bloated software. *Chi '02* (2002), 164–170. DOI:http://dx.doi.org/{http://doi.acm.org/10.1145/503376.503406{\protect\T1\textbraceright},acmid={\protect\T1\textbraceleft}503406}

33. J Miller McPherson, Lynn Smith-Lovin, and Matthew Brashears. 2006. Changes in Core Discussion Networks over Two Decades. *American Sociological Review* 71, 3 (2006), 353–375.

34. Brian Mustanski, Tyrel Starks, and Michael E Newcomb. 2014. Methods for the design and analysis of relationship and partner effects on sexual health. *Archives of sexual behavior* 43, 1 (2014), 21–33.

35. Ray Pahl and Liz Spencer. 2004. Capturing Personal Communities. In *Social Networks and Social Exclusion*, Chris Phillipson, Graham Allan, and David Morgan (Eds.). Ashgate, Aldershot, UK, 72–96.

36. John J Potterat, Richard B Rothenberg, and Steven Q Muth. 1999. Network structural dynamics and infectious disease propagation. *International journal of STD & AIDS* 10, 3 (1999), 182–185.

37. J J Potterat, D E Woodhouse, S Q Muth, R Rothenberg, W W Darrow, A S Klovdahl, and J B Muth. 2004. Network dynamism: history and lessons of the Colorado Springs study. *Network epidemiology: a handbook for survey design and data collection* (2004), 87–114.

38. Joseph Prejean, Ruiguang Song, Angela Hernandez, Rebecca Ziebell, Timothy Green, Frances Walker, Lillian S Lin, Qian An, Jonathan Mermin, Amy Lansky, H Irene Hall, and for the HIV Incidence Surveillance Group. 2011. Estimated HIV Incidence in the United States, 2006–2009. *PLoS ONE* 6, 8 (2011), e17502. DOI: http://dx.doi.org/10.1371/journal.pone.0017502

39. Harmony Rhoades, Suzanne L. Wenzel, Daniela Golinelli, Joan S. Tucker, David P. Kennedy, Harold D. Green, and Annie Zhou. 2011. The social context of homeless men's substance use. *Drug and Alcohol Dependence* 118, 2-3 (2011), 320–325. DOI: http://dx.doi.org/10.1016/j.drugalcdep.2011.04.011

40. Richard Rothenberg, Julie Baldwin, Robert Trotter, and Stephen Q Muth. 2001. The risk environment for HIV transmission: Results from the Atlanta and Flagstaff network studies. *Journal of Urban Health* 78, 3 (2001), 419–432.

41. Ben Shneiderman and Catherine Plaisant. 2003. *Designing the user interface* (fifth ed.). Pearson Education, Boston.

42. Charles D Stolper, Florian Foerster, Minsuk Kahng, Zhiyuan Lin, Aakash Goel, John Stasko, and Duen Horng Chau. 2014. GLOs: graph-level operations for exploratory network visualization. In *CHI'14 Extended Abstracts on Human Factors in Computing Systems*. ACM, 1375–1380.

43. Paola Tubaro, Antonio. A. Casilli, and Lise Mounier. 2014. Eliciting Personal Network Data in Web Surveys through Participant-generated Sociograms. *Field Methods* 26, 2 (2014), 107–125. DOI: http://dx.doi.org/10.1177/1525822X13491861

44. Edward R Tufte and P R Graves-Morris. 1983. *The visual display of quantitative information*. Vol. 2. Graphics press Cheshire, CT.

45. Barry Wellman. 1979. The Community Question: The Intimate Networks of East Yorkers. *Amer. J. Sociology* 84, 5 (1979), 1201–1233.