ORGANISING ORGANISATIONS

Collecting Social Network Information in Organizations: Successfully Navigating the Five Main Phases

Filip Agneessens, University of Surrey
Giuseppe (Joe) Labianca

Collecting social network data among organization members is challenging. It requires a well-considered strategy that goes beyond merely developing and administering a survey or accessing electronic trace data. This manuscript identifies and discusses five major phases in the data collection process, including: 1) defining your research question; 2) negotiating access to the organisation; 3) identifying the network's boundary and the relevant formal organizational structures that affect social networks; 4) deciding how to approach research subjects and how to collect data; and, 5) providing useful and ethically-sensitive feedback to the organization and its members. Decisions in each of these phases and the alignment among the phases, particularly with respect to the chosen research question, is crucial. We will offer guidelines and provide examples for each of these phases based on a wide variety of past experiences in both private and public organisations.

Phase 1. Defining your research question. Having a good, well-defined research question prior to deciding what and how you wish to collect data is central to any good research design. We will discuss design issues related to generalizability and scope of the research, and then discuss whether to use ego-network, snowball sampling, or complete network approaches; we will also consider (multistage) sampling choices for individuals, teams and/or organizations. **Phase 2. Negotiating access to organisation(s).** The first step of a data collection strategy, whether qualitative or quantitative, is the need for a good strategy to approach organization(s). We will deal with issues such as how to motivate and convince organizations to participate, gaining the appropriate level of commitment from top management prior to data collection, what is offered to organizations and their members employees in return for participation (see step 5), and up-front discussions about confidentiality and ethical considerations. We also discuss who needs to be involved in the decision, and ethical and legal considerations to network data collection in different countries. **Phase 3. Identifying network boundaries and formal structures.** A clear *a priori* network boundary identifying who will be included or excluded from the study is critical, as is identifying the formal structures that create structured foci for interaction within the network in order to distinguish voluntary from mandated interaction prior to collecting network data.

Phase 4. Approaching organizational members and collecting network data. Retrieving SNA data is dependent on the research question and the practical situation. We will focus primarily on quantitative network approaches, including surveys and secondary data such as email data, but will also discuss mixed-method approaches. We will discuss strategies to minimize non-response, electronic and paper-and-pencil surveys, and how to build appropriate network questions for surveys. We will also discuss the use of recall vs. recognition questions, and how to collect data on sensitive network relations, such as conflict or undermining.

Phase 5. Providing feedback to the organization and its members. Debriefing the study is important to the organization, the study's main proponents, and to respondents. Central to this last stage is offering advice to the organization on how to improve shortcomings that the study uncovered without violating organizational members' confidentiality and ethical concerns. We will provide examples of benefits for both organizations and individuals.

Remember GIGO?

Dean Lusher, Swinburne University of Technology
Bopha Roden, Swinburne University of Technology
Julien Brailly, Swinburne University of Technology
Tom Spurling, Swinburne University of Technology
Till Klein, Swinburne University of Technology

George Fuechsel –an early IBM programmer– was given credit for coining the term GIGO (garbage in, garbage out) as a concise way of reminding users that a computer can only process what it is given. In order to get good data, one must pay attention to the way the data is collected. However, even when intentions are good and researchers do their best to obtain the best data possible, there are still many issues that may derail the collection of good data. This paper examines the difficulties experienced in data collection for an ambitious, international study of the commercialisation of inventions using 'controlled radical polymerisation' (CRP), a new way for making better polymers. CRP is a process platform technology that allow the 'leveraging of existing technological competencies and the application of technological know-how in different business lines' (Jolly and Nasiriyar, 2007). In short, CRP can be applied in multiple ways, in multiple contexts, making it more adaptable than a single use innovation, and providing an ideal opportunity to compare and contrast the different attempts to commercialise a product developed using CRP – or so we thought.

The paper follows a research project specifically focused on how public research organisations collaborated with private firms to develop products or services that utilised this CRP technology. The research aimed to investigate and understand the specific nature of barriers to increased collaboration between the public and the private sectors and how these barriers can be overcome. This was a multilevel network study investigating the social interactions between individuals within and between organisations, and ties (e.g., research contracts, licences) between organisations.

However, the data gathered did not enable the research team to address the key research aims. In this paper we critically examine the problems encountered and provide a guide to researchers regarding issues to consider in advance during the project design phase, and how these might be ameliorated or avoided. For example, the key research partner underwent major re-structure during the study and was in a state of constant flux, with multiple changes of staff occurring during the set-up and data collection period. This was extremely important as the partner was going to act as a broker and provide the research team with access to other public research organisations as well as private firms. Further, the research team totally under-estimated how secretive industry organisations would be about their research, even with an introduction from their public research collaborators. Innovation is a secretive business, and as a result only a handful of industry firms agreed to participate in the study even though a number of strategies were employed to try and maximise engagement.

This paper details these issues, oversights and unexpected turns. It critically appraises what could have been avoided, and provides guidelines for future research projects.

"Fuck off" and Other Responses: Measuring and Manipulating Valued Networks to Improve a Difficult Workplace

Petr Matous, University of Sydney
Julien Pollack, University of Sydney
Jane Helm, University of Sydney

We describe challenges during a network intervention in a difficult workplace. A large organization contracted us to help one of their units work better as a team. The combination of consulting with research posed an ethical challenge. Formally, this was resolved by contractually distinguishing consulting aimed at helping the client and academic research based on data produced and released to us separately. In practice, however, it was more complicated.Because the project was driven by the management, collaboration from employees was expected. We made extra effort to ensure that no one feels pressured into participation and everyone understands exactly what we will do with the data.

In an initial survey, we asked the 21 team members to rate one another in terms of comfort and frequency of personal and work-related communication. Analysis of the data revealed two factions and their existence was confirmed with the team. Correct identification of the factions resulted in both appreciation and apprehension of network methods by the participants.

The factions emerged clearly on a sub-network of the strongest links (scored 10 on 1-10 scale). Possibly because of social desirability bias, the respondents apparently used the top score for everyone with a positive personal relationship. While it would have been impossible to ask about negative ties, the sub-median scores appear to capture this information for analytical purposes. We organized a dyadic intervention on pairs of co-workers from the two sides of the divide to increase the connectivity of the network by strengthening crucial dyads. We selected the pairs based on analytical measures and consultations.

We asked the pairs to spend some time together, unrecorded and unobserved away from others, to discuss about pre-selected topics on private matters. Some participants perceived this as a confronting top-down initiative and provided us with hostile feedback immediately after the exercise.

One of the authors spent long hours in the organization talking with people on both sides of the divide, learning about their concerns and regaining their trust. Worries of some participants were alleviated during long phone calls. This provided invaluable contextual understanding and made us increase the level of aggregation of the results to be shared with the team.

Four months later, we surveyed the networks again. Despite the challenges, most individuals reported more comfort communicating with their intervention partners. No one reported less comfort. The frequency of friendly chats increased within the intervention dyads but other relationships did not change. The positive shift was not biased by expectations of what others will see because this information was not to be disclosed at the relationship-level.

Overall, we learnt that voluntary research participation in professional organizations is problematic, especially in networks surveys and interventions when respondents' answers and actions may impact others. The participants need a guarantee that they will not get penalized for their honesty but they also want to see that volunteering their time will be meaningful. Constructing networks that can be used for organizational improvement requires certain levels of private information disclosure. Finding the right level is an iterative act between the researchers and the participants.

BIG & BEAUTIFUL

The Dynamics of Multiplex Social Network[s] – Insights from the Swiss StudentLife Study

Christoph Stadtfeld, ETH Zürich
András Vörös, ETH Zürich
Zsófia Boda, ETH Zürich
Timon Elmer, ETH Zürich
Marion Hoffman, ETH Zürich
Kieran Mepham, ETH Zürich
Isa Raabe, Nuffield College, University of Oxford

We present an overview of the data collection strategies employed in the Swiss StudentLife Study, conducted by the chair of Social Networks at ETH Zürich in 2016-2018.

The central goal of the study was to closely follow the emergence of an undergraduate student community. Over the course of the first year, we aimed at assessing the main factors that explain how students – who rarely know each other on their first day at university – develop densely knit social networks through time and how these networks matter for their well-being, motivation, and academic success. In order to get a detailed picture of the underlying social processes we combined a number of data collection strategies that have rarely been used together. First, we developed a longitudinal network survey design. Participants filled out the questionnaires in computer labs or from home (six questionnaires within a year that required a time investment of about 45 minutes each; >75% participation rate). These "long questionnaires" assessed a range of network perceptions (e.g., friendship, co-studying, group membership, dislike) and individual variables (e.g., emotional well-being, motivation, academic self-assessment). Second, to collect information about social processes that we assumed to operate on faster time-scales, we employed short, phone-based surveys (a dozen "short questionnaires" that took about 5 minutes each). Thereby, we could, for example, assess whether students interacted with each other regularly and how much they would study on a weekly basis. The intervals between subsequent long and short survey waves were increased through time. Third, we combined these observational data with experimental elements by randomly assigning students to specific settings (e.g., seating arrangements, information groups) in an attempt to understand in how far early random meeting opportunities affect the formation of social network ties. Fourth, we closely followed a two-day social event at the end of the first study week to investigate how interpersonal communication (collected through RFID sensors) relates to the formation of stable interpersonal perceptions.

A specific focus of the study was to consider the multiplex, multi-dimensional nature of social networks. We therefore assessed more than 20 types of interpersonal relations, perceptions, and activities. Specific considerations went into designing a strategy for collecting group perceptions and group identities.

We will present the data collection strategies and preliminary findings that relate to the multiplex dynamics of the social networks in the Swiss StudentLife community. We will additionally discuss ethical concerns we faced when designing the study and our experience in achieving high participation rates.

Collecting Large Personal Networks in a Representative Sample of Dutch Women
Gert Stulp, University of Groningen

Introduction
The extent to which social influence can affect people depends on several characteristics of their personal network. Such characteristics of personal networks can only be reliably obtained when collecting data on large personal networks. Collecting such data can be a significant burden on respondents, both in time and repetitiveness. For instance, asking information about 25 alters requires the respondent to come up with 25 names, answering repeated questions about each of these alters, and—if one is interested in relations between alters and structural characteristics of the network—requires 300 separate assessments. This probably underlies why personal network data collection typically involves only a limited number of alters. Occasionally large personal networks are collected, although often in small, selective samples.

This study
Here we report on our experiences with collecting large personal network data from a representative sample of women. In an attempt to reduce the burden to the respondents, we made use of a recently developed and freely available software tool GENSI (Stark & Krosnick, 2017; Social Network) that focuses more on the use of visualisations than standard survey software and that increases respondent enjoyment. We discuss how respondents experienced the survey, focusing on the duration of different elements, missing values, and how different characteristics of the respondents come into play.

Method
A sample of 758 women were recruited through the LISS-panel; a longitudinal survey of Dutch men and women, in which respondents fill in surveys online. All women in the LISS-panel between 18 and 40 were invited to participate in a study on "social networks and fertility behaviour". They were instructed that the survey would take about 30 minutes, and they were paid 12.50 euro if they would complete the survey. In addition to some background questions, respondents were asked to name exactly 25 alters. Sixteen questions were asked about the alters, and respondents were also asked to all 300 alter-alter relations.

Results
Nearly all (98%) respondents reported on 25 alters (13 out of 758 did not). respondents took on average 23±6 (mean±SD) minutes, and 87% finished the survey within 30 minutes. Listing 25 alter took 3±2 minutes, where answering all the alter question took longest with a time of 14±4 minutes. Reporting on the associations between all alters took 3±1 minutes. We'll further discuss: demographic differences between the respondents in the way that they filled in the survey will be discussed; how respondents rated their experiences while filling in the survey; and the quality of the data.

Conclusion
Collecting large personal network data means a significant time investment for respondents, but it is clearly feasible. Less standard and more graphic ways of collecting network data, such as GENSI, may significantly reduce respondent burden. The value of data on large personal networks in representative samples will be discussed for sociological questions.

Predicting Data Quality of Proxy Reports in Ego-Centered Network Studies

Tobias H. Stark, Utrecht University/ICS
Volker Stocké, University of Kassel

Proxy reports play a central role in ego-centered network studies in which respondents (egos) are first asked to name social contacts (alters) and, subsequently, are asked to provide information about these contacts in so-called name interpreter questions (Marsden 2011). A fundamental concern of network researchers is the extent to which proxy reports are an accurate representation of the answers that alters would give themselves (Marsden 1990). Previous research has produced inconsistent findings with regard to a long list of factors that may be associated with the congruence between proxy report and alters' self-reports (Roydhouse and Wilson, 2017). For example, some studies found proxy reports to be more in line with alters' self-reports, the more frequently ego and alter had contact or the closer their relationship was (Bickart et al., 2006; Jäger, 2005; Pearcy et al., 2008). Other investigations did, however, not find evidence in line with these ideas (Dawe and Knight, 1997; White and Watkins, 2000).

Based on the theory of survey satisficing (Krosnick, 1991), the present research proposes a theoretical framework that allows predicting when proxy reports are more likely to match self-reports. Congruence should be higher if respondents possess the motivation and ability to answer a proxy question effortfully, and if the task is not too difficult. Moreover, the theory of survey satisficing states that motivation, abilities, and task difficulty are not independent of each other. The latter offers an explanation for the inconsistent findings in the past literature. For instance, some studies could not confirm the prediction that egos have higher abilities to answer a proxy question accurately the longer they know the alters (Harling et al., 2015; White and Watkins, 2000). However, these studies did not take ego's motivation to provide an accurate answer or the difficulty of the proxy question into account. Higher abilities may, for instance, not compensate for the increased difficulty when one is required to answer a proxy question about non-observable attitudes of alter.

Results from two ego-centered network studies study among adults (*N* = 756) and among middle school students (*N* = 679), in which nominated alters were also interviewed, were in line with these hypotheses. Different indicators of respondents' motivation, their abilities, and the difficulty of the proxy questions predicted congruence between proxy reports and alters' self-reports. In addition, the effects of these predictors depended on each other. For instance, higher motivation in terms of interest in the topic of the survey (Sudman et al. 1994) was only related to more congruent proxy answers in easy proxy questions about socio-demographic characteristics but it did not compensate for the increased difficulty of answering a proxy question about alters' attitudes. Results of this study can guide the design of future ego-centered network studies. Moreover, the findings clearly indicate that predictors of congruence of proxy reports should not be studied in isolation.

CANVASSING CRIME

Recruitment and Territoriality in Organized Crime in the UK: A Network Analysis
Paolo Campana, University of Cambridge
Federico Varese, University of Oxford

A number of prominent thinkers have argued that contemporary society is a 'global network' (Castells, 2000; Sennett, 1999, 2006; Bauman 2000; 2002). Globalization and 'liquid modernity' have also affected the underworld. Organized crime is no more a phenomenon rooted in a particular territory, and a "Global Crime Inc." transcending territorial control has emerged (Shelley, Picarelli and Corpora 2003; see also Barak 2001; Castells 2000). Several scholars have put forward the opposite view, namely that various forms of organized crime continue to be local in scope, are hard to export and the liquidity metaphor is misplaced (Reuter 1993; Gambetta 1993: Varese 2010). Several factors conspire to keep gangs local. The first refers to problems of asymmetric information and principal-agent problems. A related mechanism to ensure control, and compliancy, of members is shared ethnicity. A third key factor that keeps gangs local is the level of policing. The prediction of such theoretical arguments is that gang membership tends to be homogeneous in relation to nationality and ethnicity, leading to a high level of homophily within the gang.

Data Challenges in the study of OC

The study of organized crime is notoriously difficult. First, it is hard to pin down what the phenomenon actually is. Definitions range from minimum membership, to the presence of a hierarchical structure, continuity over time, the ability to influence politics, to what actually groups do (for reviews, see Varese 2010). What is defined as 'gang' in the US scholarship is often labelled as 'OCG' across the Atlantic, making it hard for scholars to enter into a meaningful dialogue. Second, people who engage in organized crime are a hidden population that actively tries to stay hidden and is reluctant to open up to outsiders.

Most network studies of gangs rely on anonymized police arrest records (McGloin and Piquero 2010; Papachristos et al. 2012; Papachristos and Widelman 2014; Shaefer 2012; Grund and Densley 2012). Such data set allow scholars to unpack the internal structure of the OCG using individual-level network data, and map sub-groups and cliques. However, such data are limited by the fact that not every offence leads to arrest (Hughes 2005) and police enforcement is selective: some individuals might be targeted more than others depending on race or place of residence (Black 1970). Finally, co-arrest data do not in themselves distinguish between gang members and non-gang members (Pyrooz et al 2010).

Other studies have gone beyond arrest data by relying on police field intelligence observation cards recording non-criminal encounters with the police (Papachristos, Braga and Hureau 2012) and expert interviews with law enforcement personnel (Braga et al. 2001; Kennedy, Braga and Piehl 1997). This study goes in the same direction of moving beyond the use of co-arrest data. We were given access to a rich data set collected by a police force in the UK. The data include demographic data, intelligence reports, co-arrest data as well as non-criminal encounters with the police, and local profile crime data. We also conducted interviews with key law enforcement personnel. Relying on these data, in this paper we explore a number of substantive issues related to gangs' territoriality and recruitment, the role of ethnicity, and the relationship between type of crime and gang membership.

SNA and Criminal Justice Data: Benefits, Challenges, and the Way Forward
David Bright, Flinders University
Carlo Morselli, Université de Montréal

Achieving access to good quality data is one of the main challenges faced by researchers who conduct social network analysis (SNA) on criminal groups. Law enforcement and surveillance agencies are understandably unlikely to share data on open investigations, or cases that are or will be before the courts. Researchers have utilised a variety of criminal justice data sources in order to conduct SNA on a range of criminal networks including co-offending networks, organised criminal groups, and terrorist groups. Data has been sourced from all points in the criminal justice process, from crime and arrest data, electric surveillance records, and interrogation transcripts to prosecution files, court transcripts, and judges' sentencing comments. Research over the last two decades has demonstrated that such data can facilitate SNA, including analysis of actor roles and resources, relations or ties, types of ties, and flows of resources. Such data has been used to reach conclusions regarding network structure, cliques and clusters, and the positional importance of network actors. Nonetheless, such data sources also suffer limitations, mainly with respect to accuracy, validity and reliability. Data from criminal justice source are likely to include errors, both intentional (e.g., aliases, false information) and unintentional (e.g., transcription errors). Such data is also likely to suffer from missing information, including missing actors and/or characteristics of actors (e.g., roles and resources) and missing links and/or characteristics of links (e.g., strength of ties, resource flows). Criminal justice data present particular problems with defining the network boundary as the boundary as determined by law enforcement or prosecution agencies may not correspond to the boundary as defined by network members. Finally, such data may produce biased centrality scores, with such scores reflecting the focus of the investigation or prosecutorial strategy. Given the limitations of such data, one solution is the use of multiple criminal justice data sources in order that each additional data source compensates for the limitations of other sources. In this chapter, we outline the types of criminal justice data sources, articulate the benefits and challenges of such sources, and provide recommendations about the considered use of criminal justice data to undertake social network analysis.

Collecting Social Network Data in Prison and During Community Re-Entry

Corey Whichard, Pennsylvania State University
David Schaefer, University of California
Derek Kreager, Pennsylvania State University

In the era of mass incarceration, the prison has reached an ever-widening portion of the population, yet our understanding of the inmate experience and prisoner social organization have paradoxically declined. Social network data and methods provide unique opportunities to illuminate this hard-to-reach context. However, prison conditions simultaneously produce unique challenges for conducting primary research on inmate populations, as interviewers must negotiate a system of total control and find ways to establish inmate trust and successfully motivate them to participate in research. These challenges intensify when collecting prison social network data, where high response rates are crucial and core aspects of the data collection process are in tension with prison norms against "snitching" on fellow inmates. Despite these challenges, the Prison Inmate Network Study (PINS) and its associated projects have successfully collected global and egocentric network data from hundreds of incarcerated respondents across multiple prison settings. To date, the PINS research team has collected two waves of global network data from a good behavior unit of a rural men's prison, ten waves of data from a drug and alcohol treatment unit of an urban men's prison, and one wave of data from a good behavior unit of a rural women's prison. In addition, PINS researchers have collected longitudinal egocentric network data from a subsample of parole-eligible respondents from each research site as they transition from prison into the community.

In this presentation, the authors will discuss key issues associated with designing and administering network surveys for inmate and re-entrant populations. For the global prison surveys, we discuss various measurement strategies, including boundary selection, the availability and utility of inmate rosters, the wording of network questions, and the inclusion of open-ended responses to promote rapport. We draw on insights from an original pilot study and subsequent field experience collecting network data from prisoners to highlight survey items that were effective for studying this population, as well as questions that provoked strong resistance and were dropped from later surveys. We also discuss effective tactics for recruiting respondents, managing the data collection process within the institution while minimizing burden on correctional staff, and the pros and cons of computerized network surveys in the high-security and low-technology prison setting.

For the egocentric re-entry surveys, we discuss the benefits of data visualization for promoting respondent engagement, strategies for managing cognitive demand and respondent resistance to name-generator designs, and the broader challenges associated with collecting such data from members of a hard-to-reach population. We also discuss the innovative collection of "expected" network data from soon-to-be-released inmates—a variant of cognitive social structures—that we compare against the "actual" networks that emerged during community re-entry. Qualitative interview data was collected in conjunction with egocentric networks to investigate how inmates form expectations for future social relationships while preparing for prison release, the extent to which such relationships manifest during re-entry, and the interactive processes that account for asymmetries between expected and actualized networks.

Towards a Unified Approach to Data Collection on Covert Networks

Tomáš Diviák, University of Groningen

The interest in the study of covert networks among practitioners and academics alike has increased tremendously since the events of 9/11. Numerous types of covert networks have been studied, for instance terrorist networks, drug trafficking networks, mafias, or gangs (Bichler, Malm, & Cooper, 2017; Morselli, 2014a; van der Hulst, 2011).

The defining feature of these networks is that actors involved in them aim to avoid detection (Morselli, 2009). This makes any sort of data collection a daunting task as it directly collides with the main interest of actors whom researcher would like to observe (Morselli, 2014b). Yet, both practice and rigorous research require good data.

This paper seeks to propose a unified approach to data collection on covert networks in order to facilitate further research in the area, particularly with regard to systematic reviews of multiple studies and meta-analysis of results (Bichler et al., 2017; Gutfraind & Genkin, 2017).

The problem central to this paper is a variation on the classical problem of boundary specification in network analysis (Laumann, Marsden, & Prensky, 1983). However, in covert networks, it is essential to account for specific situation surrounding the data collection, which stems from two main problems related to the nature of covert networks – reliance on secondary data and proneness to missing data. As direct observation or primary data collection by the means of questionnaires is generally not available for covert networks, researchers rely on secondary data from various sources (police, courts or media; Bright, Hughes, & Chalmers, 2012), which are all susceptible to systematic errors and biases. Related to these issues, missing information is a severe yet frequently untreated issue in covert networks (Campana, 2016).

Secondary and missing data are intertwined with five key questions of boundary specification in covert networks. 1) Criteria for inclusion of nodes are often unclear, which substantially affects comparability across different studies, because such comparison relies on some common basis. Morselli (2009) introduced the idea of criminal justice rings based on how law enforcement defines the involvement of actors in particular case. 2) The criteria for inclusion of ties need to acknowledge the multiplexity of ties and that simply aggregating all ties may lead to erroneous conclusions (Gerdes, 2015). 3) Systematic analysis of individual attributes is also necessary in order to advance our understanding of covert networks (Robins, 2009), however there is still no consensus on a generalizable coding scheme which would allow for comparison of actors with similar attributes. 4) The multilevel aspect has been largely neglected in covert networks. Criminological theory emphasizes the importance of convergence settings (Felson, 2006), but there is no systematic account on the role spatial or social settings in the structure of crime. 5) Covert networks have been labelled as dynamic and adaptable, but in order to asses this claim empirically, we need to be able to clearly distinguish time periods of the existence of particular criminal networks.

The paper will further elaborate on these issues.

MAKING MEANING

Socio-Semantic-Material Network Data Collection

Nikita Basov, St. Petersburg State University
Irina Kretser, St. Petersburg State University

Mixed data, including both in-depth ethnographic data and social network data, is usually a challenge to collect, put together, and analyze. Especially, when it includes the data on cultural structures, that later are to be formalized and subjected to statistical network analysis. In order to find out how social structure interplays with cultural (material and semantic) structures (see Basov, 2018; Basov and Brennecke, 2017), in 2014-2015 our international team conducted a longitudinal four-wave study of five European collectives of artists based in Barcelona, Hamburg, London, Madrid, and St. Petersburg. The dataset capturing relations between words, individuals, and physical objects (see Figure 1) includes verified sociometric surveys on multiple social network ties in each collective, over 200 semi-structured interviews with the artists, 108 visual ethnographies in the artistic spaces, over 200 photo elicitations, and a vast collection of texts written by the artists and amounting to over 1,000,000 words. Out of this dataset, we constructed a three-level structure including six types of relations between individuals, physical objects, and words, both within and across the three orders. To secure the possibility to formalize subsequently the network data on the three interrelated orders, before, during, and after the fieldwork, we adapted, developed, tested, integrated, and transformed our set of tools and techniques. The paper will discuss the toolkit for longitudinal socio-semantic-material network data collection, as well as the peculiarities of its application. It allows addressing a number of issues in mixed network data collection, such as securing continuous access to the field throughout the study, involving the problem of repetitive data collection in a longitudinal ethnographic study, triangulation techniques for different types of relations, tension between observed and reported relational data, multiple social network data collection, issues of multilevel network data collection that imply simultaneous collection of data on relations of different nature (such as social, material, and semantic) between different node types (such as actors, objects, and words), formalization of ethnographic data, and many others.

Constructing the Network in situ: Notes from the Design of a Participatory Egocentric Research Instrument

Bernie Hogan, University of Oxford
Joshua Melville, Northwestern University
Patrick Janulis, Northwestern University
Gregory L. Phillips II, Northwestern University
Noshir Contractor, Northwestern University
Michelle Birkett, Northwestern University

Beyond the tedium of collecting personal network data is the question of how reliable is this data to begin with? The immediate answer to this question would make an appeal to 'ground truth' data. In many cases, however, it is not merely that such a ground truth is not accessible to the researcher. Rather, this ground truth does not exist outside the interview to begin with. From choice blindness on discussing important matters to the imposition of closeness as a unifying metric for name recall, the personal network interview is not merely an exercise in recollection; it is a constructivist exercise. This is not to dismiss the exercise as contaminated by subjective whims and the opinions of the researcher. Rather, it is to acknowledge how the many little steps involved in the development and deployment of a network interview that serve to create specific kind of social network. The resulting network ideally both addresses the research questions of the researcher as well as the lived experience of the respondent.

This paper discusses the design decisions involved in the development of Network Canvas, a software suite for the capture of personal social network data that extends existing name generator methods using touchscreen interfaces. Past work on Network Canvas has demonstrated its reliability vis-à-vis paper based methods. In this paper we go a step further to describe the philosophy of 'network co-creation' as an epistemological backdrop to the design and collection of network data. Using a mixed-methods qualitative approach we detail how we navigated the tension between a general-purpose research tool and one that makes too many assumptions about what constitutes a network study. Investigating, and to a limited extent resolving, this tension has led us to reconsider the epistemic bases of reliability in network instruments. Through reports from field sites, interviewers, test subjects and peers in personal network collection, we have discovered that materiality, question ordering, and even aesthetics play a role in reliability.

To this end, we emphasize the crucial role of the interviewer, the interview designer, and the constraints of current technological interfaces as being reflexive foci. These matters are not mere biases to be overcome, but actors in the process of constructing a network. Further, in confronting the limits of personal network research and designing for the interview-as-theatre, we believe we are being particularly faithful to the use of network analysis as a means to describe and analyse social structure as a lived experience. We advance this claim with qualitative evidence from early design work, pre-tests, ethnographic work with field sites, reports from interviewers as well as a small amount of descriptive quantitative data from previous studies using Network Canvas.

Giving Artefacts a Voice: Collecting SNA Data on Non-participatory Populations

Daniel Tischer, University of Bristol
Adam Leaver, University of Sheffield
with support from Martin Everett, University of Manchester
& Johan Koskinen, University of Manchester, University of Melbourne, University of Linkoping,
& Swinburne University of Technology

Finance is ubiquitous and whilst we have a notional understanding of the careers of senior figures on Wall Street or in the City of London, there has been less opportunity to explore patterns of career progression within the financial industry more generally. This is particularly true for less transparent, non-public-facing functions where it is difficult to define populations and locate suitable data. Moreover, getting access to setting where finance is performed and people proves difficult, and whilst this has been done for specific cases, it may prove impossible in situations where the data covers a larger number of individuals working at numerous firms.

However, the financial crisis has created an opportunity to rectify this because the widespread failure of Collateralised Debt Obligations (CDOs) has moved obscure documents – offering circulars and pitchbooks – into the public sphere. Traditionally of concern to investors, these documents became subject of public investigations (FCIC 2011) and of an object of concern to the wider social sciences (see Hardie & MacKenzie 2014, Tischer & Leaver 2017, Chernenko 2017). For the purpose of this study we are interested in the biographical information contained in these documents for a particular actor: Collateral Management (CM). CM was a central risk-mitigating function and aimed at protecting buyers of CDOs by managing underlying assets. In doing so, these financial actors claim independence; the documents are keen to point out that collateral management firms (CMF) do not work for the investment bank, even though it is the investment bank that hires them.

As private agreements between two parties, CDOs are not regulated and thus locating an exhaustive repository of CDOs proves difficult[1]. Still, we can locate these documents on specialist exchanges. Our dataset includes 367 CDOs for which documentation is available from the Irish Stock Exchange. The level of detail provided differs from CDO to CDO but at its most basic level each disclosure provides us with a biographical overview of CMFs' employees: name, educational achievements, past employment, and current position, etc. But importantly, by attaining the names of individuals we can begin to empirically define our population – names of people employed by CMFs prior to 2008 – and generate attributes files. And we can triangulate and complete this dataset through other sources – LinkedIn, Bloomberg Executive, company websites etc.

By collecting this data for a large number of CMF employees (approximately 100 CMFs and 2000 employees) we get insights into career developments up until the financial crisis and, for example, begin to understand if and how the claim of independence is undermined by latent ties between individuals and previous co-workers. But we can also follow up on their post-crisis careers. What do these actors do once the market implodes, their particular function becomes obsolete and their activity becomes subject of high-profile lawsuits?

---

[1] *The "CDO Library" provided by the FCIC only holds records of 63 CDOs. See*
https://fcic.law.stanford.edu/resource/staff-data-projects/cdo-Library

Defining and Eliciting Relationships in a Traditional Society: Social Networks in Rural Cambodia

Emiel de Lange, University of Edinburgh
E.J. Milner-Gulland, University of Oxford
Aidan Keane, University of Edinburgh

Social network analysis was developed, and has largely been applied, in western modern societies. These societies are highly individualistic, and relations between individuals tend to be highly specialised. In collectivist cultures, individual identity is closely linked to the group, and the boundaries between relationships, such as personal and business, are often blurred. This is likely to have implications for how network methods can be applied in traditional collectivist societies. For example, how can the ties relevant to the research question or intervention be identified and measured?

I will present experience and data from a social network study in rural Cambodia as a case study for exploring these issues. We set out to measure agricultural advice networks, as is commonly done in the natural resource management literature. However, we found that these ties could not be reported on by villagers, nor could influential farmers (i.e. opinion leaders) be identified. We therefore conducted a scoping study, using qualitative (focus group discussions and key informant interviews) and ethnographic methods in ten villages, to understand the interactions through which relevant information might flow. We also drew on anthropological literature on Khmer society. Focus groups were structured to probe knowledge about specific agriculture practices, and then to ask how participants learned about these practices. Key informants were asked to recall the entry of now widespread practices into their community and to describe how these spread. Data indicated that agricultural information is likely to spread through more general everyday interactions such as home visits or incidental contact along the road. In addition, external sources of information were found to be important, such as visiting salesmen.

Subsequently, I conducted a social network analysis in one village, interviewing all 366 adults (aged 18+) present at the time. Initially, a household questionnaire was used to measure household level variables and interactions. We also used this to record the identities of all adults present in the household, which allowed us to revisit until each had been surveyed individually. Name-generator questionnaires were administered individually to measure a wide range of interactions, including kinship, home visits, and material exchange, as well as interactions outside the village, such as with agricultural middlemen. We encountered several further issues. In Cambodia, names are not used as frequently as in western society, so there were significant issues with the accuracy of responses. I used reported kinship and ages to identify and correct aliases, however a significant number of mistakes remain. A simple solution to this problem remains elusive. Finally, social desirability biases may occur for hierarchical interactions, such as conflict resolution, with respondents rarely deviating from official hierarchies in their responses.

Collecting Social Network Data to Assess a Poverty Alleviation Intervention in the Philippines

Lincoln L. Lau, Univeristy of Toronto, International Care Ministries
Petr Matous, University of Sydney
Krisha Lim, International Care Ministries
Peng Wang, Swinburne University of Technology
Donald C. Cole, Univeristy of Toronto

We are piloting longitudinal socio-centric data collection instruments as an evaluation tool of the social impacts of large-scale community-based poverty-alleviation interventions in the Philippines. The intervention targets the poorest individuals in low-income neighbourhoods with a 16-week collective training to strengthen their networks, livelihoods, and health. We report on pilot data gathering in ten communities and subsequent analysis.

The first challenge was establishing network boundaries of each "community". Initially included in the rosters for sociocentric network data gathering were people screened as eligible for the program using a poverty assessment test. However, some preselected participants did not join the program and others were welcomed into the program as replacements. Alters' unique identification was a challenge, as the organization conducting the intervention used official full names but people are often known only by their nicknames. (We may use photographs next time.) For all these reasons, initial and final rosters and nodesets were slightly different.

Using the constructed rosters, a fixed-form pen and paper survey was implemented by hired and trained third-party enumerators. We applied four different name generators to elicit expressive and instrumental networks starting with (1) Do you know this person? If positive, then: (2) Has this person visited your house or have you visited this person's house in the past month? ; (3) Do you know this person well enough to discuss private matter such as personal health issues together?; (4) Do you know this person well enough that you could ask to borrow 50 Pesos (without interest) if needed? We also asked if respondents had links outside of the participant group but because identification or follow up would be hard, we asked only for overall number. None of the questions were judged too sensitive by the enumerators or respondents.

We fitted longitudinal exponential random graph models to the networks of discussion of private matters (3). The changes in network boundaries, household-level (intervention focus) versus individual-level data (unit of social network data gathering), and inclusion of multiple independent communities required developing new extensions of the models.

We expected incoming links to be a reliable indicator of social support because these are counted from answers of all community members and are unaffected by the degree of the cooperation of the respondent with the survey but incoming-links based measures were unrelated to the intervention outcomes. Surprisingly, network effects based on the number of alters named by the respondent were most strongly associated with improvement in health over time. Individuals' subjective perceptions of how many people they could talk to about private matters seemed to be more relevant for their health than the actual number of peers who considered them to be their discussion partner. Possibly, egocentric (personal network) data might have been sufficient.

While the identified implementation challenges need to be addressed, the participating organizations see the potential of regularizing network data collection for the assessment of social impact of poverty-alleviation programs.

Networks of Peace in Conflict: Practical and Ethical Challenges of Collecting Social Network Data in Rural Eastern DR Congo

Patrycja Stys, LSE, Swinburne University of Technology, & University of Edinburgh
Johan Koskinen, University of Manchester, University of Melbourne, University of Linkoping, & Swinburne University of Technology, Batumike Papy Muzuri, Goma, DR Congo

As Social Network Analysis (SNA) increases in popularity across many disciplines, researchers are collecting social network data across a diverse range of contexts: from the dark web to cross-cultural contexts in developing countries of the global South. This paper focuses on research design and implementation in rural, conflict-affected eastern Democratic Republic of Congo (DRC), from the viewpoint of one of the authors, supporting the team on the ground from 'the North', and those of the other two, translating data collection instruments and implementing the design on the ground in 'the South'. Research was conducted over seven months in 2016, yielding 396 partially overlapping ego-networks (ego-nets) which include members of all active armed groups in the area, the national army, and civilians. The final paper will include experiences from a similar project undertaken in the summer of 2018.

This paper first presents the tailoring of participant-aided sociogrammes and the cultural translation of personal support networks, designed in the North and employed in the South. We consider working with under-educated (and sometimes armed) respondents in areas of heightened insecurity, where communities are suspicious of such projects and rumours are rife, impacting the data collected and the safety of enumerators. Likewise, we consider the effects of budgeting and logistics on the ground on the research – the lack of electricity, running water, mobile service, and reliable infrastructure. As much as these are practical concerns, they are also ethical ones, impacting both the researchers and the researched communities in terms of data collection and future dissemination of findings.

We conclude with key takeaways concerning link-tracing designs adapted to and developed from such research, what data is collected and what biases it inevitably harbours due to the selection of the seed set and subsequent respondents. Thus, we aim to shed light on the tensions between research design and the realities of the field, arguing for flexibility in their resolution. Despite such challenges, we posit that analytical techniques for SNA are advanced enough to cope with real-world data, and it is such data that advances our understanding of social systems in areas where the state is either absent or predatory. Findings can inform future conflict resolution interventions as well as development programmes, as much as research methodology in terms of design and subsequent statistical analysis. Likewise, we posit that such large-scale projects offer opportunities for less (if not non-) extractive and exploitative research in the global South – but only if deeply inscribed institutional conventions are successfully challenged.

DIGITAL DIVIDES

The Observation of Valued, Time-stamped and Signed Social Interaction Processes
Jürgen Lerner, University of Konstanz
Alessandro Lomi, University of Italian Switzerland

Network data come from social relations that are typically assumed to be binary, and directly observed either at a single point in time, or at discrete time moments. Yet, relational social processes rarely produce data that conform to these restrictive research design assumptions. A recent trend in social network research is beginning to reflect on the problems posed by differences between the richness of the data produced by relational social systems, and the simplification imposed by prevalent conceptualizations of social networks. Inspired by this trend, our talk is organized in two main parts. In the first, we discuss problems posed by the collection of directed relational data produced by event recurrences, rather than state transitions, as it is standard practice in network research. More specifically, we focus on problems created by the collection of network data that are: (i) valued (i.e., contain information about relational intensity), and (ii) time-stamped (i.e., contain information on the exact timing at which events connecting 'senders' and 'receivers' happen). In this first part, we discuss research design strategies that may help to alleviate some of the constraints inherent in the graph- theoretic representation of social network data inherited from sociometry. In the second part of our talk, we discuss how to collect data on negative relations, i.e., data on relations that are typically not directly observable. More specifically, we focus on the collection of data produced by signed relational processes – social processes defined in terms of non-mutually exclusive positive and negative relational among participants. Problems of relational data that are (i) valued (non-binary); (ii) time-stamped (continually observed), and (iii) signed (positive and negative) pose statistical and substantive issues that are typically addressed by simplifications implying considerable loss of information. For example, valued data are binarized, time-stamped data are aggregated in conventional time frames, and positive and negative relations are defined as mutually exclusive thus precluding their co-existence. These various data transformations involve a considerable impoverishment of the information content of network data, and impose drastic limitations on the range of conclusions that we can draw form an analysis of the evolutionary dynamics of systems of social relations. Throughout our talk, we exemplify our argument by building on our extensive experience with data produced by the Wikipedia editing network. Focusing on editing activities connecting participants to individual words, we reconstruct negative and positive relations between any pair of Wikipedia editors in terms of continually observed individual acts of text addition and deletion. A weighted, directed negative relation exists between i to j if i deletes text that j has contributed. The weight of this relation is equal to the number of words contributed by j that are deleted by i. A positive relation from i to j is recorded whenever i undeletes, reinstates or somehow protects the text contributed by j from deletion attempted by a third party (k). We rely on our experience with Wikipedia also to reflect on the large number of specific coding choices we had to face – choices that may seem idiosyncratic and context-specific, but in fact depend delicately on general assumptions about the ontology of connected social systems.

Collecting Data in an IT-Security Team: Considerations Regarding the Appropriate Collection of Network Data in a High-risk, Professional Environment

Laurin B. Weissinger, Nuffield College, University of Oxford

As part of a larger project on trust and cooperation in IT-Security, the author collected network data from a high-level, international security team specialised in cyber-incident response, also called Security Operations Centres (SOC). This specific SOC had between 18-25 members during observation. Over the course of 18 months, network data were collected at five points in time, leading to a dataset containing five waves of 15 network types, e.g. advice giving and friendship. In addition to these network data, the author also interviewed all members of the team to gather node-data, e.g. specialisation, prior assignments, career path, etc. Generally, data collection in security fields is risky, as information about one's defences can become a powerful tool should an adversary attain access. Thus, collecting these data securely but also in a worthwhile manner academically involved a variety of challenges:

First, the team in question is associated with a very large international firm, and therefore, various legal and procedural issues had to be considered. Second, actors in IT-Security are acutely aware of how data can be misused, requiring the author to build trust and demonstrate appropriate data handling. Third, precautions were necessary due to the actual security impact that would result from an adversary getting access to these data. Thus, various ethical and security concerns had to be addressed and mitigated: for example, how could the author decrease the risk for the host organisation, the team members but also to himself and the university? To do so, it was necessary to negotiate various precautions, some technical, some procedural, with the security team and higher management to avoid data becoming accessible, and sufficient anonymization and other data protection measures to ensure that even the full dataset would not be useful to a potential adversary; for example, some data were never recorded digitally. Yet, not all precautions initially designed and agreed were actually feasible in practice.

As alluded to, it was necessary to explain the project and the author's precautions and to build interpersonal trust, including by demonstrating technical understanding, and to reassure the team that the project was useful and neither a security risk, nor a way for the employer to track their work and relationships. Thus, sufficient security precautions and anonymization were essential for collecting these data. While attaining management support was a hurdle, it was helpful later on to build trust and work with the team members. Furthermore, the team and environment kept changing, which the author had to keep track of in a qualitative manner to manage data collection and ensure sufficient continuity, in some cases changing procedures.

Due to wide-ranging technical and procedural measures regarding data handling, the project was approved and also accepted by the team members. In order to support future network collection projects, this presentation will focus on pertinent issues of data collection in a security environment and give insights into what precautions worked and which failed. These practical insights are applicable to organisational and high-security settings more generally and thus helpful for other empirical researchers.

Using Think-tank Policy Reports to Construct Knowledge Production Networks: A Case Study in the Sociology of Intellectual Interventions

Jordan Soukias Tchilingirian, University of Bath

Background

Think-tanks and their research staff are important actors within national and transnational policy-knowledge regimes. The name think-tank is an umbrella term describing a variety of actors who differ in levels of access to economic resources and expertise, and who cross the ideological and organisational spectrum. Though they differ considerably, these actors can be seen to be unified by their close association with (but never fully belonging to) the worlds politics, media, academia, and business (Medvetz, 2012). This liminal status means think-tanks are objects of inquiry for scholars with interests across the social and political sciences. However think-tank scholarship tends to bracket out the interstitial nature of these organisations by binding think-tanks to one of the many established communities they are located between. This approach has stunted the study of knowledge production within, and the intellectual life of, think-tanks.

Approach

To resist privileging a set of think-tank relations with a particular field, I focused on the emergent networks of think-tanks and their intellectuals. Emergent networks tend to be less formally structured, visible and interstitial, as they draw disparate actors together (Kadushin, 1976, p. 770). Collecting emergent networks required bypassing the data sources traditionally used in think-tank scholarship, namely interviews with senior staff (Pautz, 2012) or published board memberships (Burris, 2008). Inspired by the 'sociology of interventions' within the sociology of knowledge (Eyal & Buchholz, 2010), I focused on the actual products think-tank intellectuals create and deploy in political debates, namely the policy report.

In order to apply this 'sociology of interventions' approach, and to identify the actors enrolled in the production of policy research, I constructed a novel mixed research design. First, policy reports were mined for names and professional affiliations of report authors, project advisors, and funders. Names, affiliations, and biographies actors were cleaned and verified using several sources, including LinkedIn, and organisational websites. This formed the basis of a co-funding, co-authorship and, co-advice network. One and multimode quantitative analysis was conducted which investigated homophily within the British think-tank community and uncovered the various the ideational and material resources that are active and valued by these liminal intellectuals.

I also conducted personal network interviews with junior and senior think-tank intellectuals about the production of a single policy report. This narrative approach to networks offered three key insights: First, an account of the manner by which information from a variety of sites is found and translated into policy-focussed social science. Secondly, it explored the tensions underpinning creative networks. Thirdly, interviews identified actors hidden from more public networks.

Assessment

Whilst the extraction and verification of relational data was time consuming; using policy reports as the data source for network analysis produced novel findings which challenged established conceptions of the intellectual life of think-tanks. Starting with intellectual interventions also captured the dynamics and politics of policy research. This is important, as creative networks are not static structures, they are social relations; they are based on positions, communication, power relations, and ultimately the acceptance of some actors and the exclusions of others.