Egocentric Network Analysis of the National Health Interview Survey Data

Azadeh *Hemmati* & Kon Shing Kenneth *Chung*

Complex Systems Research Group
Project Management Program
Faculty of Engineering & Information Technologies
The University of Sydney, NSW 2006, Australia
Telephone: +61 2 9036 7517
Email: ahem5793@uni.sydney.edu.au & ken.chung@sydney.edu.au

Abstract

In this study, we develop a methodology for extracting social network data from a social survey – namely, the U.S. National Health Interview Survey (NHIS) 2010. We (i) demonstrate how relational data is extracted for (ii) investigating the association between egocentric network properties (structure, position and relations) and Quality of Life (QOL) in the context of cancer care. Here, we highlight the data collection procedure, its benefits and limitations and standard measures of social network data. This novel approach is beneficial for network science academics, industry professionals and organizational network analysts to analyse social surveys from a social network's perspective. Results show that there are significant differences in the network properties (density, degree, tie strength, efficiency and constraint) of those experiencing high and low QOL. These findings are critical to influencing interventions and policy development for enhanced QOL in cancer care.

Keywords—Social network; measurement; Egocentric Network; Social Survey; General Social Survey

Introduction

The network approach seeks to describe social structure in terms of networks and interpret behavior of individuals (actors) considering their change in position within social structure (Marsden, 1990). Social network analysis (SNA) offers a unique methodology for visualizing and investigating social structures and relations based on the theoretical constructs of sociology, mathematical foundations of graph theory and recent developments in computer hardware and software (Galaskiewicz and Wasserman, 1994). In social network research, organizations are viewed as clusters of people joined by a variety of links. Such research focuses on patterns of relationships between people rather than on people in isolation from one another (Brass, 1995). SNA allows analysing structure of relationships among social entities, plus the impact of structure on other social phenomena. However a general social survey usually allows for studying individuals' properties as the prime context for explaining outcome. The main difference between traditional social science data and network data is that traditional behavioral studies focus on actors and attributes while network data concentrates on actors and relations (Hanneman and Riddle, 2005).

Based on earlier network surveys (Fischer, 1982; Verbrugge, 1977; Wellman, 1979), the General Social Survey (GSS) measured the national U.S. social system of ego-networks for the first time in 1985 (Burt, 1984; Marsden, 1987). Burt (1984) mentioned that the data obtained through General Social Survey needs to be selected carefully. Cost of developing survey design and network analysis, proposed by Burt (1984), is inconsiderable in compare to its benefits. The costs are 5 to 11 minutes of interview and associated data processing time. On the other hand, benefits are doubled including increased accuracy of GSS measures in social context and expanded research opportunities. "Conceptual developments in network analysis offer a variety of indicators describing theoretically significant aspects of an individual's interpersonal environment; social integration, social participation, and exposure to normative pressures. Secondly, network data offer (in interaction with existing GSS items) insights into the ways in which a respondent's interpersonal environment distorts and enriches the respondent's abilities, aspirations, attitudes, and behaviors" (Burt, 1984, p.294).

Social media services such as Facebook or Twitter enabled new source of collaboration and online interaction. Despite providing opportunity of accessing broad range of social network data through social media, most methodological research on network measurement has focused on data obtained through general social surveys and questionnaire.

This research study develops an analytical framework that shifts from the traditional focus on the individual to a relational analysis. Our aim is to demonstrate that it's possible to extract relational data from traditional social surveys and census data to provide further insights into the association between social network properties of the individual and their individual attributes such as quality of life (QOL).

Traditional Social Network Data Collection

Social network data includes two types of information: individual's data and relational data. This means there are two different sampling units: individual respondents, and the partnerships. Types of relational data depend on research questions. They might include social or interpersonal evaluation or else behavioral interactions such as friendship or teacher-student. While individuals are still the source of all information, the information they provide is not limited to their own attributes and includes data that will assist establishing the pattern of relationships between them and everyone else. Therefore the sampling frames for attribute and relational data are nested. To collect data, we first need to choose how to sample individual respondents, and then select how to sample partnerships and relational data from these respondents. The main difference in network study designs are derived from the ways in which these two types of data are sampled (Morris, 2004).

Network data have been obtained through surveys and questionnaires, archives, observation, electronic track downs, and experiments (Marsden, 1990; Hemmati and Chung, 2014a; Hemmati and Chung, 2014b; Rebehy and Chung, 2013; Chung et al., 2014). The method of measurement is driven by fundamental questions related to individual respondents and their social relations such as "Whom do you go to for advice?" or "Who provides you with emotional support?" Measurements of network relations can be binary or valued. Binary relations indicate existence or absence of connection between respondents (actors) and other contacts, while valued relations measure frequency of the interactions between respondents (network actors) and other contacts. Examples of binary and valued relations are the marital relations between network members and amount of friendship or closeness between friends. In terms of representing the data, network data is demonstrated in Sociomatrix or graph format. While Sociomatrix is simpler to manage for analysis, graph format allows for visualization of the social network data. There are two main approaches to social network data collection (a) Sociocentric or complete networks and (b) Egocentric networks. In egocentric (or local) networks, an individual is located at the hub of a wheel, with the edge outlining social contacts and the roads indicating the ties which connects the individual to contacts. In the social network context, the individual at the center is named "ego". The people or contacts referred to or by the "ego" as advisor, friend, or relative are called "alters". In egocentric approach only information about individual's immediate contacts and associated interconnections are studied (Chung, 2012; Chung et al., 2005). On the other hand, Sociocentric or complete networks are developed by gathering information related to alters, both those who influence and those being influenced. It focuses on the whole network analysis and measuring structural pattern of interactions and impact of patterns on outcomes such as concentration of power (Chung et al., 2005; Smith and Christakis, 2008; Fisher, 2005; Marsden, 2002; Mizruchi and Marquis, 2006; Scott, 2000). Thus Sociocentric network studies are infrequent as scale of data required for the analysis is extensive. On the other hand, they provide an opportunity for novel insights and are more suitable for indicating emergent quality of networks (Smith and Christakis, 2008).

The first instrument for collecting egocentric network data from social surveys was administered in the 1985 General Social Survey (GSS) by Ron Burt. That is the most famous and generally used instrument for collecting egocentric network data from social surveys, among other instruments. The instrument introduced by Burt (1984) was concentrated on a ‘‘name generator’’ approach asking respondents the following question to extract the contacts who were considered by participants in a survey respondent's network:

"From time to time, most people discuss important matters with other people. Looking back over the last 6 months, who are the people with whom you discussed matters important to you? Just tell me their first names or initials."(Bailey and Marsden, 1999, p.288).

Name generator which was introduced firstly in the 1960s (Laumann, 1966) and administered through surveys or interviews, is a standard method to identify networks and describe network characteristics and structure. Participants are given one or a series of questions that produce a list of network relations (alters), for instance "those people with whom they discuss important matters (Burt, 1984)", or "the people with whom they chat or visit (Campbell and Lee, 1991)". When list of names have been generated, participants are given series of "name interpreters" which are follow up questions that collect information on demographic characteristics of each relation (alter), the relationship between respondent (ego) and connections (alter), and the relationships between alters themselves. "Data collected via name generators and interpreters provide individual profiles of respondents' personal network members that can be aggregated into measures of network composition, such as average tie strength, mean alter characteristics, communication activity, network range and density" (Marin and Hampton, 2007, p.164).

Social network studies are concerned with studying patterns of social structure. The collected network data might not represent the correct structure due to being collected from participants who are willing to contribute. Thus it may not include individuals with vital roles in the network. Additionally, collecting the complete network data is time consuming and sometimes complete network information is unavailable.

SNA incorporates the social context to explain individual or group outcomes. For instance, relational concepts are used to measure amount of cohesion within a group of individuals (e.g. network density); the degree to which an individual initiates interactions within a social network (e.g. actor centrality); the extent to which people in a network initiate interactions with a particular individual (e.g. actor prestige); and whether some individuals remain unconnected or isolated from network interactions. Network researchers who use SNA have the opportunity to choose the level of analysis at which to study the network collaborations such as the dyadic, triadic, or the whole network and view actions as socially "embedded" or "nested" in one, or several networks (Galaskiewicz and Wasserman, 1994).

Burt (1984) made a strong argument on importance of collecting relational data in the administration of a general social survey (GSS). He stated that the GSS is sociology's premier source of national survey data which can be utilized for theoretically empirical research. Furthermore, the cost of collecting relational data in the administration of GSS is 5 to 11 minutes of interview time in addition to data processing time. More importantly, the inclusion of the relational data allows for describing and understanding important aspects of an individual's interpersonal and social environment besides offering richer insights to explain social outcomes through analyzing attribute and relational data.

In this study, we develop a methodology for extracting social network data from a social survey and demonstrate how relational data is extracted for investigating the association between egocentric network properties (structure, position and relations) and Quality of Life (QOL) in the context of cancer care. This research builds upon the egocentric network methodology, to study the effect of social networks on QOL while considering the information about individual's immediate contacts and associated interconnections. This is a novel approach and beneficial for network science academics, industry professionals and organizational network analysts to analyse social surveys from social networks perspective. In the following section, we describe the methodology and techniques that were used to extract egocentric network data of 2010 National Health Interview Survey (NHIS) including assumptions which made through the data analysis.

Methodology : Egocentric Analysis of National Health Interview Survey (NHIS) Data

The context for this study concerns adult cancer patients (18 years of age or over) with different types of cancer including blood, breast, brain, bone and so on, who live in one of the 50 states of America and Columbia. The dataset was obtained from the 2010 National Health Interview Survey (NHIS) which is a multiuse health survey managed by the National Center for Health Statistics (NCHS), Centers for Disease Control and Prevention (CDC), and is the primary source of information on the health of the resident, non-institutionalized, household population of the United States. In the 2010 NHIS survey, a randomly selected household was interviewed personally in their residence using computer-assisted interviewing technology or through telephone (National Center of Health Statistics, 2010).

We analyzed the data released by NHIS in 2010, which contained eight questionnaires and nine data files including Household, Family, Adult, Adult Cancer, Child, Cover, Disability Test, and QOL. The 'QOL' section was done as part of the 2010 NHIS Sample Adult module. Almost one quarter of the sample adults were randomly chosen to receive the QOL questionnaires. As a result of the selection process, different response measures (weights) were generated for these respondents and NHIS staff decided to create a separate, stand-alone file for the QOL variables, rather than adding the variables to the 2010 Sample Adult file (National Center of Health Statistics, 2010).

As the subject of this study is evaluating QOL of adult cancer patients and examine impact of support network or interaction network on quality of life or life satisfaction of these patients, data of 2010

NHIS ‘Sample Adult Cancer’ and ‘Quality of Life’ was utilized and merged for the analysis. We extracted social network properties of the respondent (or the ‘ego’) such as personal network size (the number of ego’s alters), density (the degree to which alters know each other) and boundedness (the degree to which alters come from different categories of familiarity, such as relatives or neighbor). Tie characteristics analysed including level of closeness, frequency of contact, or diversity of resources flowing through tie connections.

Survey Design

Non-institutional civilian Americans (both citizens and non-citizens) were sampled by household, for analysis of medical behavior at the family-level. These households were located in one of the 50 states of America and Columbia. One adult and one child (if any were present) selected from a household to complete the “Sample Adult” and “Sample Child” components of the survey. Approximately 35,000 households containing 87,500 individuals were interviewed every year (National Center of Health Statistics, 2013).

The NHIS covers "core questions" or “Basic Module” including three components called “Family core”, “Sample Adult Core”, and “Sample Child core”. The 2010 NHIS covers supplementary questions/modules called “Sample Adult Cancer Control” and “Sample Adult Quality of Life”. The data released by NHIS in 2010, contained eight questionnaires and nine data files including Household, Family, Adult, Adult Cancer, Child, Cover, Disability Test, and QOL. Almost one quarter of the sample adults were randomly chosen to receive the QOL questionnaires. As a result of the selection process, different response measures (weights) were generated for these respondents and NHIS staff decided to create a separate, stand-alone file for the QOL variables, rather than adding them to the 2010 Sample Adult file (National Center of Health Statistics, 2010).

All members of the household 18 years of age and over, who were at home at the time of the interview, invited to participate and respond for themselves. For children and those adults who were not at home during the interview, information was provided by a knowledgeable adult family member (18 years of age or over) lived in the household. Information for the “Sample Child” questionnaire gained from a knowledgeable adult who lived in the household. For the “Sample Adult” questionnaire, one adult per family randomly selected to respond to questions individually. A knowledgeable proxy assisted individuals with physical or mental disabilities (reported around 350 cases per year). “Sample Child” and “sample Adult” questionnaires were slightly different however they both gathered fundamental information on health status, health care services, and behaviour. Flowchart of the 2010 NHIS questionnaire components can be found in Figure 1.

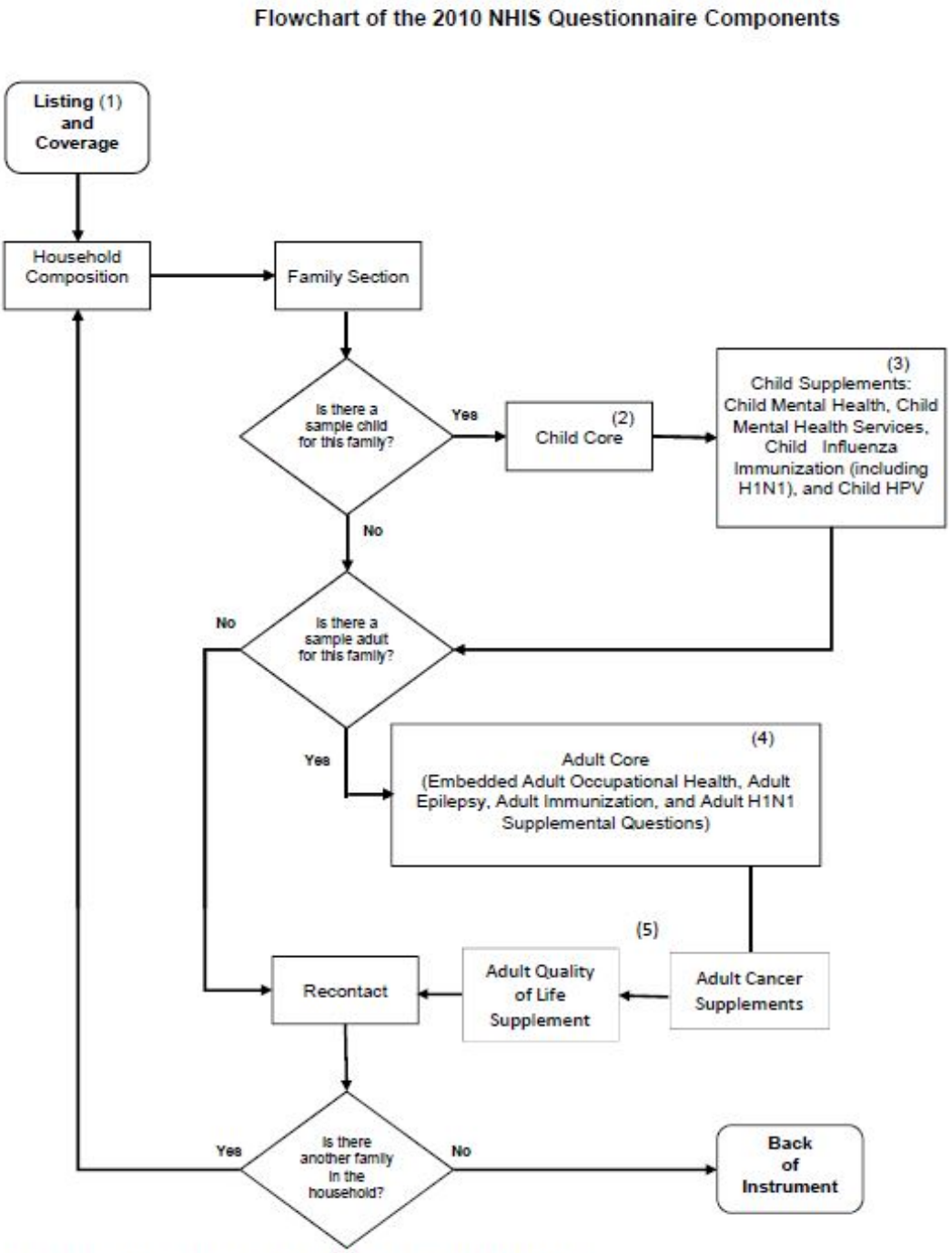

Figure 1: Flowchart of the 2010 NHIS Questionnaire Components (National Center of Health Statistics, 2010)

List of output data files from the 2010 NHIS components are as below (National Center of Health Statistics, 2010):

- Cancer (CANCERXX)
- Family (FAMILYXX)
- Household (HOUSEHLD)
- Injury and Poison Episode (INJPOIEP)
- Injury and Poison Verbatim (INJVERBT)
- Person (PERSONSX)
- Sample Adult (SAMADULT)
- Sample Child (SAMCHILD)
- Paradata (PARADATA)
- Disability Questions Tests 2010 (DISBTEST)
- Quality of Life (QUALLIFE)

Each released file contains the following seven components and additional documents such as survey description, questionnaires, interview manual and data access instruction.

- Data - Compressed ASCII format (.ZIP-->.DAT)
- File Variable Summary - PDF format (_SUMMARY.PDF)
- File Layout (Codebook)- PDF format (_LAYOUT.PDF)
- File Frequency Report - PDF format (_FREQ.PDF)
- SAS Program Sample - ASCII format (.SAS)
- SPSS Program Sample - ASCII format (.SPS)
- STATA Program Sample - ASCII format (.DO)

As noted before, only the data of 2010 NHIS 'Sample Adult Cancer' and 'Quality of Life' was utilized and merged for the analysis. The remaining data files, which are listed above, were not used in this study's analysis. This process is shown in the following figure.

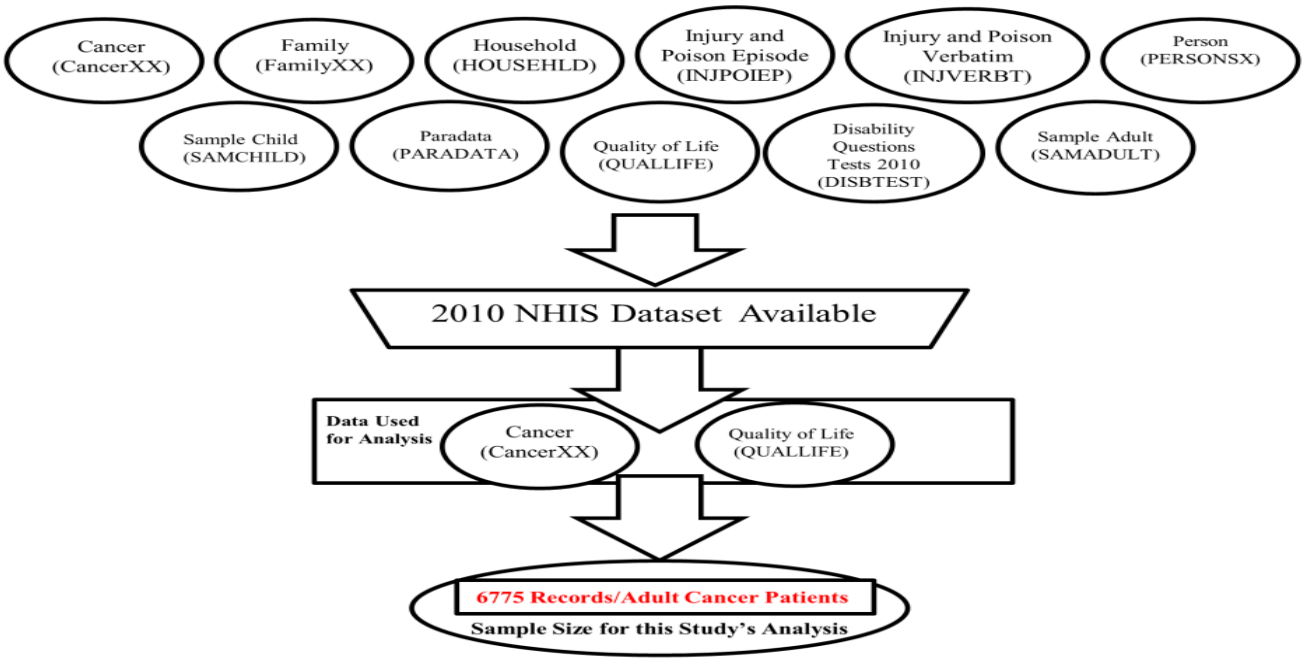

Figure 2: Process of Data Extraction from 2010 NHIS Dataset

Sample Size and Response Rate

The number of households interviewed in 2010 NHIS sample was 34,329 including 89,976 persons in 35,177 families. The number of interviewed sample for the "Sample Adult" component was 27,157 persons 18 years of age and older, who didn't require any assistance to complete the survey. Educated proxies assisted 378 cases to respond the sample adult survey. The total household response rate was 79.5%. The 12.9 percentage of non-interviewed households rejected to participate. The remaining 7.6 percentage points were the result of failure to locate an eligible respondent at home after repeated contact efforts. The final response rate for the family component was 78.7% while the final response rate for the Adult Sample component was 60.8%. In this study's analysis, data of 2010 NHIS 'Sample Adult Cancer' and 'Quality of Life' was merged. Number of respondents who completed both 'Sample Adult Cancer' and 'Quality of Life' was 6775 adult cancer individuals.

Data Extraction

Each record in the NHIS data files has a unique Household Serial Number (HHX), Family Serial Number (FMX) and Person Sequence Number (FPX), which were utilized to merge data of 'Sample Adult Cancer' and 'Quality of Life' files. The resulting data file that contained 6775 records was evaluated in order to locate variables, which captured social interactions (including contact with friends, relatives or attend religious service or social activities) and QOL data. For instance, households were asked if they have ever had a colonoscopy or breast MRI test. If they answered YES to these questions, a social relation between the respondent and gastroenterologist or MRI Radiologist developed indicating that person had a social interaction with those types of health professionals. Thus, the egocentric network data consisting of interactions between patients and their contacts were developed by analysing the responses and identifying the type of contacts (e.g. general practitioner, nurse) and frequency with whom the patients contacted. In total 49 variables (including attribute data) were selected from 'Sample Adult Cancer' and 'Quality of Life' codebooks. Interaction network in this study refers to social or professional network relations with people whom adult cancer patients associate or interact with during their illness and cancer treatment journey. Thus, it is possible that in the ego network representation, that there are interactions between patient & health care professional, patient & health care organization, health care professional & health care professional, health care professional & health care organization, health care organization & health care organization. List of variables which capture support network and attribute data can be found in the following table. The unique codes assigned to each support network variable will be used later, in data analysis to generate egocentric network data.

Variable Type	Description of Variable	Variable Name	Key Phrase indicating interaction	Unique code
Attribute	Family No	FMX	Not Assigned	Not Assigned
	Household No	HHX	Not Assigned	Not Assigned
	Person No	FPX	Not Assigned	Not Assigned
	REGION	REGION	Not Assigned	Not Assigned
	Gender	SEX	Not Assigned	Not Assigned
	Marital Status	R_MARITL	Spouse	3
	Language used most often	SPSPEAK	Not Assigned	Not Assigned
Support Network	Use someone's assistance	MOB_3F	Someone's Assistance	1
	Use other type of equipment or help	MOB_3G	Other Type of Equipment or Help	2
	MD recommends exercise/physical activity	MDEXER	Doctor or Other Health Professional	4
	Use telephone help/quit line to stop smoking	FQUITB_1	Telephone Helpline or Quitline	5
	Use one-on-one counseling to stop smoking	FQUITB_2	One-on-one Counseling	6
	Use a stop smoking clinic, class, or support group to stop smoking	FQUITB_3	Stop Smoking Clinic/ Class, or Support Group	7
	Use telephone help/quit line to try to quit smoking	CQUITB_1	Telephone Helpline or Quitline	5

Variable Type	Description of Variable	Variable Name	Key Phrase indicating interaction	Unique code
Support Network	Use one-on-one counseling to try to quit smoking	CQUITB_2	One-on-one Counseling	6
	Use a stop smoking clinic, class, or support group to try to quit smoking	CQUITB_3	Stop Smoking Clinic/ Class, or Support Group	7
	Health professional advised quit smoking/using other kinds of tobacco	MDTOB2	Doctor or Other Health Professional	4
	MD advised quit smoking/using other kinds of tobacco	HPTOB21	Medical Doctor	8
	Dentist advised quit smoking/using other kinds of tobacco	HPTOB22	Dentist	9
	Nurse advised quit smoking/using other kinds of tobacco	HPTOB23	Nurse	10
	Dental Hygienist advised quit smoking/using other kinds of tobacco	HPTOB24	Dental Hygienist	11
	Other health prof advised quit smoking/using other kinds of tobacco	HPTOB25	Other Health Professional	12
	Ever had Pap smear/Pap test	PAPHAD1	Medical Doctor	8
	Doctor recommended Pap test	MDRECP1	Doctor or Other Health Professional	4
	Told by doctor you had HPV	HPVHAD	Doctor or Other Health Professional	4
	Ever had a mammogram	MAMHAD	Mammogram Radiographer	13
	How breast cancer was found	MFOUND	Either Spouse, Doctor or Health Professional, Mammogram Radiographer or MRI Radiologist	3,4,13 or 14
	Doctor recommended mammogram	MDRECMAM	Doctor or Other Health Professional	4
	Ever had a breast exam	CBEHAD	Doctor or Other Health Professional	4
	Ever had a breast MRI	MRIHAD	MRI Radiologist	14
	Had a chest x-ray in last 12 months	CHESTX	Radiographer	15
	Had CAT scan or CT scan in last 12 months	CAT12	Radiographer	15
	Doctor recommended a PSA test	PSAREC	Medical Doctor	8
	Ever had PSA test	PSAHAD	Medical Doctor	8
	Doctor talked to you about advantages of PSA test	PSAADV	Medical Doctor	8
	Doctor talked to you about disadvantages of PSA test	PSADISAV	Medical Doctor	8
	Doctor told you some experts disagree about whether men should have PSA tests	PSAEXP	Medical Doctor	8
	Ever had a colonoscopy	COLHAD	Gastroenterologist	16
	Doctor recommended you be tested for colon/rectum problems, past 12 months	COLPROB	Doctor or Other Health Professional	4
	Doctor recommended particular tests	COLTEST	Doctor or Other Health Professional	4
	Ever had blood stool test collected during doctor office visit	FOBHAD	Doctor or Other Health Professional	4
	Discussed genetic test with MD	GTPOSS	Doctor or Other Health Professional	4

Variable Type	Description of Variable	Variable Name	Key Phrase indicating interaction	Unique code
Support Network	Advised to have genetic test for cancer	GTADVISE	Doctor or Other Health Professional	4
	Had genetic test for cancer risk	GTGRISK	Medical Geneticists	17
	Told cancer had come back	RECUR	Doctor or Other Health Professional	4
	Received advice from a health care professional about routine cancer check-ups after completing cancer treatment	FUADVEV	Doctor, Nurse or Other Health Professional	18
	Health professional spoke about how cancer may affect emotions/relationships	HPCAEROP	Doctor, Nurse or Other Health Professional	18
	Received professional counseling/joined support group after cancer diagnosis	ACDPCSG	Professional Counseling or Support Group	19
	Participated in research study/clinical trial as part of cancer treatment	CTRSCT	Research Study or Clinical Trial	20

Table 1: List of variables capturing support network and attribute data

In terms of data collection, this study adopted the egocentric approach, focusing on social relations and interactions that surround cancer patients rather than the whole network analysis and measuring structural pattern of interactions among all adult cancer patients in the dataset and their interaction networks. In the egocentric approach, the focus is to measure patterns of interactions between individuals and how these patterns affect outcomes (Burt, 1992). There are challenges for collecting social network data using whole network approach in the context of this study. To conduct the whole network data, we need to have name of all respondents and their contacts including name of health professionals, friends, family members and community workers in each patient's social network which were unavailable in the NHIS dataset.

The resulting data from merging 'Sample Adult Cancer' and 'Quality of Life' and selection of 49 variables, was imported into a MySQL database. A PHP application was developed to parse those variables, which capture support network and QOL data to generate egocentric network data including egos, alters and alter to alter connections for each cancer patient in the dataset. In this study, support network variables were grouped and categorized into "Family", "Health Professional", "Community" and "Other" for analysing functional diversity. The health professional category includes nurse, medical doctor or dentist. Equipment or services which help patients to cope with their situation or fulfil their physical needs such as wheelchair or telephone helpline are categorized as "Other". List of support network variables categorized into "family", "community" and "other" is shown in the following table. The remaining support network variables, which are mentioned in table 1, were classified as "health professional".

Group Name	Variable Name
Family	R_MARITL
	MFOUND
Community	FQUITB_3
	CQUITB_3
	ACDPCSG
	CTRSCT
Other	MOB_3F
	MOB_3G
	FQUITB_1
	CQUITB_1

Table 2: List of Interaction Network Data Categorized into Family, Community and Other

The PHP application analyzed value of those capturing support network and QOL variables for each patient. In case patient's response was YES (or numeric '1'), tie data was generated between that ego (patient) and the support network variable, which was defined by a unique code in the application (unique codes can be found in table 1). Tie data represented existence, frequency and strength of interaction between cancer patients and their interaction network such as medical doctors, health professional and so on.

In this case, an ego is a particular patient (or a case in the NHIS dataset) while medical doctors, nurses, telephone helpline, friend, family members, people from religious or cancer support communities whom an adult cancer patient associates with are the alters. The connections between the patients (egos) and their contacts (alters) were derived from their responses in the dataset and by identifying variables which capture interactions between them. To develop the 'alter to alter' ties (or connections between contacts) in each ego network four assumptions were made. For instance a patient who responded YES to the questions "Received professional counseling/joined support group after cancer diagnosis" and "Participated in research study/clinical trial as part of cancer treatment" have both of these contact types (counselling/ support group, study/clinical trial group) as an alter in the ego network. Information about research studies is usually distributed by health professionals and groups which are working in the same field. Generally, supporting groups and professionals who directly deal with cancer patients, share information about research studies with patients or encourage them to participate in clinical trials or research studies. Thus alter-to-alter connections were developed on the basis of discussion between these two types of supporting contacts in the patient's egocentric network. Other assumptions are listed as below:

- Telephone helplines are commonly a toll free number, which provide basic information to callers. In most cases they refer individuals to contact community or medical centers, which provide more professional and one-on-one counseling services. Thus patients, who have both telephone helpline and one-on-one counseling service in their interaction network, have alter to alter connection between these two types of support network variables.

- Patients who answered YES to both questions asking if they "Ever had Pap Smear/Pap test" and "Doctor recommended Pap test" are assumed to have alter to alter connection between their support network contacts as Pap test is commonly done by the doctor who recommended doing the test or doctor's assistant nurse.

- Health professional, doctor or nurse who provides information on how cancer may impact emotions and relationship of patients with other people, normally refer the patient to join supporting groups or supply information on how the patient can find supporting groups or professional counseling to cope with the sickness. Thus those patients who answered YES to both questions "Health professional, doctor or nurse spoke about how cancer may affect emotions and relationships with other people" (support network variable called HPCAEROP) and "Received professional counseling/joined support group to help patient to cope with the illness after cancer diagnosis" (support network variable called ACDPCSG) have alter to alter connection between these two types of support network variables. This process is shown in the following figure.

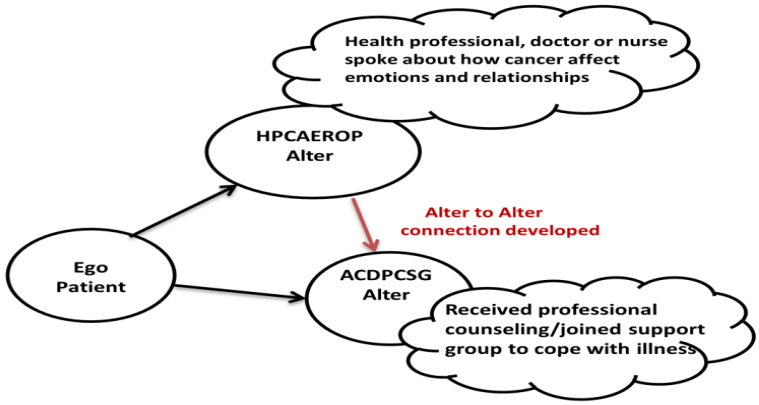

Figure 3: Developing 'alter to alter' Connection on the basis of Forth Assumption

Repeating this process 6775 times (for each case) generated the egocentric network data containing node and tie data in UCINet-Netdraw's VNA file format for all cases in the dataset. Netdraw was used to construct the Sociograms and extract the degree, efficiency and constraint values based on their respective equations. The last stage of data analysis process involved querying the MySQL database for attribute and QOL data, then consolidating the resulting data with social network data and developing a comma separated values (csv) file for analysis in SPSS. The processes discussed above are summarized in Figure 5.

The relationship between a patient and interaction network members is assumed to be directional. This assumption is made on the basis that patients contact health professionals to seek help or advice. It is very rare for health professionals to contact patients unless there is another type of relation apart from doctor and patient between them, such as friendship, colleague or family relationship. As there is no information about interaction network members to locate if there is a friendship relation between a patient and health professional members thus, an assumption is made that contacts are initiated by patients and the relationship direction is from patient's side. Process of extracting relational data is demonstrated in the following figure.

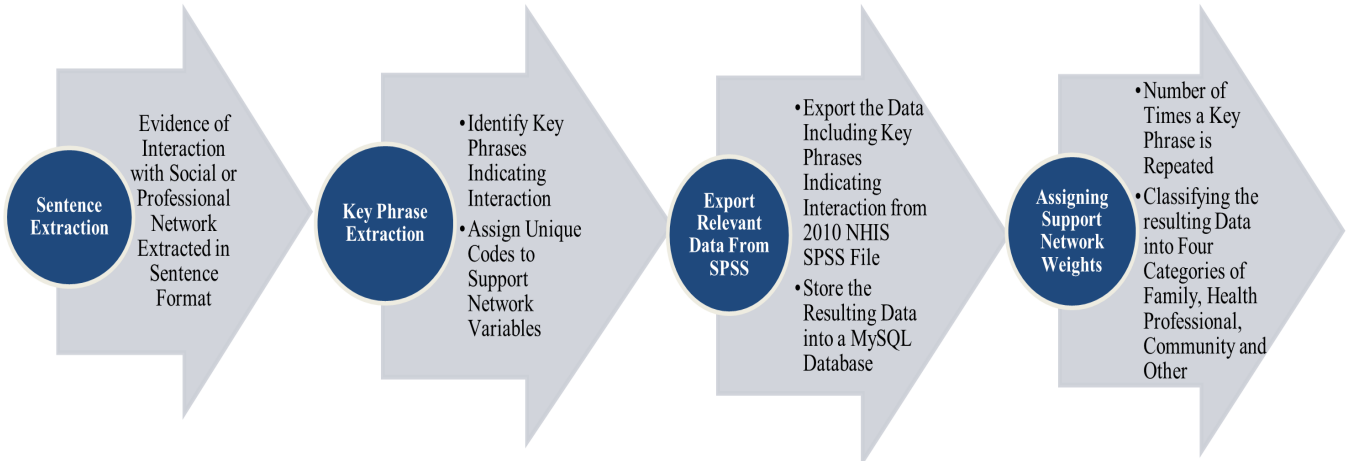

Figure 4: Process of Extracting Relational Data

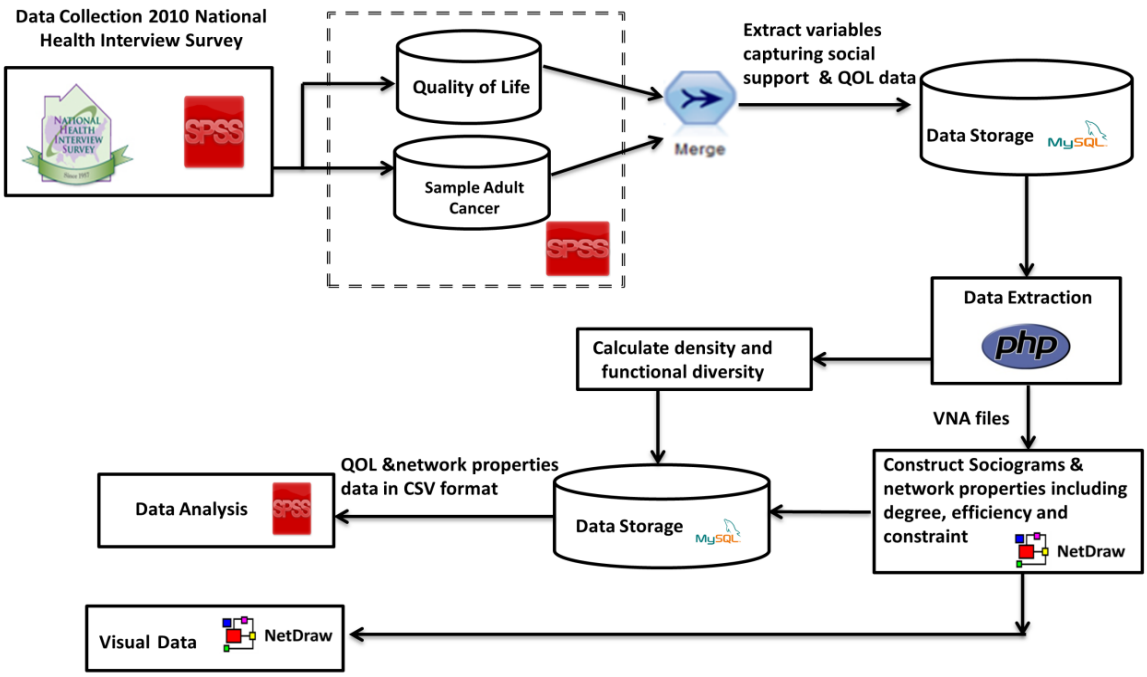

Figure 5: Research Procedure Outlook

Data Analysis

The first phase of analysis engaged eliciting ego network data of each cancer patient including tie data, which represent existence, frequency and strength of interaction between cancer patients and their interaction network such as medical doctors, health professional. Additionally, producing 'alter to alter' data and converting these data into VNA format for analysis in NetDraw. As explained earlier, the resulting data from merging 'Sample Adult Cancer' and 'Quality of Life' and selection of 49 variables, was imported into a MySQL database. A PHP application was developed to parse those interaction network variables and generate egocentric network data for each cancer patient in the dataset. MYSQL database was used for data storage which enabled efficient retrieval of data, based on MYSQL queries and flexibility to extract significant information. The PHP application generated egocentric network data, containing node and tie data in UCINET-NetDraw's VNA file format (Borgatti et al., 2002), for all 6775 cases in the dataset. Network properties values including degree, efficiency and constraint were extracted through NetDraw and stored in the MySQL database. We also calculated density and functional diversity (Hemmati and Chung, 2014a; Hemmati and Chung, 2014b). The last stage of data analysis process involved querying the MySQL database for attribute and QOL data then consolidating the resulting data with social network data and developing a comma separated values (csv) file for analysis in SPSS.

In the second phase of data analysis process, which was done in SPSS, preliminary analysis of the data distribution such as descriptive statistics, histograms, tests of normality and scatterplots were performed to select the best suited statistical tests. In the event that the data is normally distributed,

statistical tests which investigate relationships between variables such as Pearson's correlations and Independent-samples t-test can be used. However, if the distribution is not normal, then non-parametric tests such as Spearman's rank order correlation and Mann-Whitney U tests need to be considered.

Figure 6 illustrates interaction network of respondents who had more than 10 alters (support network members). The central nodes are demonstrating the egos or cancer patients. They are named by a unique number in the database (such as C3116). Other nodes connected to the central node are indicating alters, such as family members, medical doctors or friends and recognized by unique codes defined in Table 1. Node's shape and color are defined based on their attribute data of marital status. For instance a "black rectangle" indicates that actor's marital status was "divorced". Strength of tie is shown by thickness of lines, which are drawn between actors and alters (1 = distant, 7 = very close) and their color. Thus those ties/lines which drawn thicker, denote stronger social relationship between cancer patients and interaction network members. Those ties/lines colored in 'light orange or numeric 1' are indicating the lowest tie strength while others colored in 'green or numeric 7' are representing the highest tie strength.

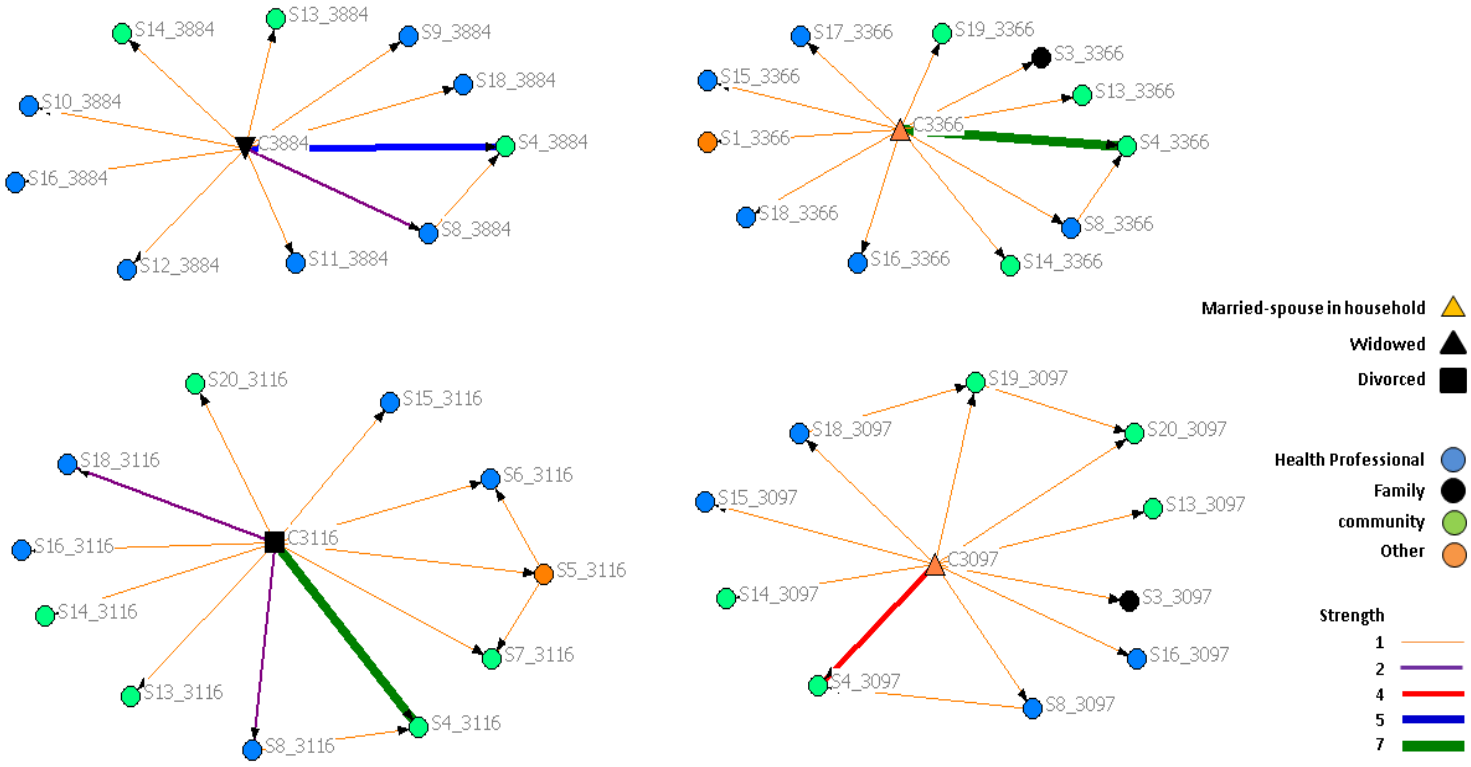

Figure 6: Ego Network Diagram of Respondents (Actors) with More than 10 Support Network Members (Alters)

Discussion

This research study develops an analytical framework that shifts from the traditional focus on individual attributes to a relational analysis of an interactive network. This is a fundamental move from standard behavioural research approach to a networks perspective method for understanding social outcomes. As described earlier, two main approaches to social network data collection are Sociocentric (or complete networks) and Egocentric networks. This research builds upon the egocentric network methodology, to study the effect of social networks on QOL while considering the information about individual's immediate contacts and associated interconnections. Although the network is partial, as the entire network information is unavailable, the patterns and attitudes of ego within the sample are apparent and generalisable to a reasonable degree.

We described the process of extracting key phrases which capture social support data from 2010 NHIS dataset and collecting network and attribute data for exploring the relationship between social network properties and quality in life of cancer patients. While number of assumptions was made in driving the network data from 2010 NHIS dataset, the processes of data collection, classification, storage and extraction are operationalized. Thus this methodology can be replicated in future research studies. This research study demonstrated that it is possible to extract relational data from traditional social surveys and census data to provide further insights into the association between social network properties of the individual and their individual attributes such as QOL. This, we believe, is a significant contribution, which network science academics, industry professionals and organizational network analysts can benefit from. With abundance of census data collected from nations world-wide, this research provides ways forward for eliciting and analysing social network data from social surveys.

In terms of limitations of the study, we made a number of assumptions in deriving our network data listed in the methodology section and developing alter to alter connections between patients and their interaction network. This was due to the fact that this study relied on secondary data (2010 NHIS dataset) and that primary data could not be collected. While the reliability of this aspect may be challenged, we are confident that replication of our process is not an issue. Considering the limitation presented above, it is hoped that the results from this study are able to derive new conversations, discussions and generate new questions to develop and utilize methodologies to extract relational data from social surveys.

Conclusion and Future Directions

This paper documents a novel methodology for a relational analysis of pre-existing survey data such as National Health Interview Survey (NHIS). We demonstrated the relational data extraction process, and in particular utilised the egocentric network method for personal network analysis. The National Health Interview Survey (NHIS), which was initiated in July 1957, is the key source of information on the health of the civilian non-institutionalized population of the United States (NCHS). NHIS is

one of the major data collection programs of the National Center for Health Statistics. The NHIS provides persistent survey and special studies to secure accurate and current statistical information on the amount, distribution, and effects of illness in the United States. As stated earlier, Burt (1984) made a strong argument on importance of collecting relational data in the administration of a general social survey (GSS). He stated that national general social survey is sociology's premier source of national survey data which can be utilized for theoretically empirical research. Furthermore, the cost of collecting relational data in the administration of GSS is 5 to 11 minutes of interview time in addition to data processing time. More importantly, the inclusion of the relational data allows for describing and understanding important aspects of an individual's interpersonal and social environment besides offering richer insights to explain social outcomes through analyzing attribute and relational data. For researchers interested in studying social network and extracting relational data from social surveys, the following key factors need to be considered when approaching social network analysis of social surveys:

- Analyse the dataset to locate evidence of implicit or explicit interaction with social or professional network that respondents associate or interact with.
- Extract those questions, which indicate any type of collaboration including referral (direct or indirect), consultation, awareness, emotional support and knowledge sharing.
- Perform key phrase extraction to identify type of social or professional network contacts and classify the relational data types.
- Measure strength of social or professional relations by number of times a key phrase is repeated.

We believe that this study may potentially generate considerable interest and discussion, and may motivate organisations such as the Australian Bureau of Statistics to collect network or relational data. It would also be valuable to apply the proposed methodology in the context of another domain, analyzing social network data from a social survey undertaken by an Australian organization such as Australian Bureau of Statistics or Cancer Council NSW. This would definitely serve as a valuable complement to this research.

References

BAILEY, S. & MARSDEN, P. V. 1999. Interpretation and interview context: examining the General SocialSurvey name generator using cognitive methods. *Social Networks,* 21**,** 287 - 309.

BORGATTI, S. P., EVERETT, M. G. & FREEMAN, L. C. 2002. Ucinet for Windows: Software for Social Network Analysis. *Harvard, MA: Analytic Technologies.* .

BRASS, D. J. 1995. A social network perspective on human resources management. *In G. R. Ferris (Ed.),Research in personnel and human resources management. Greenwich, CT:J AI Press,* 13**,** 39-79.

BURT, R. 1984. "Network Items and the General Social Survey". *Social Networks,* 6**,** 293-339.

BURT, R. S. 1992. Structural holes: The social structure of competition.

CAMPBELL, K. E. & LEE, B. A. 1991. "Name Generators in Surveys of Personal Networks.". *Social Networks,* 13**,** 203-221.

CHUNG, K. K. S. 2012. Towards a Social Network Model for Understanding Information and Communication Technology Use for Knowledge-Intensive Workers. *Taipei: IGI Global Publishers*.

CHUNG, K. K. S., HOSSAIN, L. & DAVIS, J. 2005. Exploring sociocentric and egocentric approaches for social network analysis. *International Conference on Knowledge Management Asia Pacific***,** 1-8.

CHUNG, K. S. K., YOUNG, J. & WHITE, K. 2014. Using Social Network-based Complexity Profiles to Understand Attitudes to Medical Care. *Health Systems,* 3**,** 199-213.

FISCHER, C. S. 1982. To Dwell among Friends: Personal Networks in Town and City.

FISHER, D. 2005. Using Egocentric Networks to Understand Communication. *Internet Computing,IEEE*.

GALASKIEWICZ, J. & WASSERMAN, S. 1994. Advances in social network analysis : research in the social and behavioral sciences.

HANNEMAN, R. A. & RIDDLE, M. 2005. Introduction to social network methods.

HEMMATI, A. & CHUNG, K. S. K. 2014a. Personal Social Network Properties and Mental Health in Cancer Care. *Paper presented at the The 4th Workshop on Social Network Analysis in Applications, Beijing, China*.

HEMMATI, A. & CHUNG, K. S. K. 2014b. Social Networks and Quality of Life: the National Health Interview Survey. *Paper presented at the Advances in Social Network and Mining (ASONAM), Beijing, China.*

LAUMANN, E. 1966. Prestige and Association in an Urban Community. . *New York: Bobbs-Merrill.* .

MARIN, A. & HAMPTON, K. N. 2007. Simplifying the Personal Network Name Generator: Alternatives to Traditional Multiple and Single Name Generators. *Field Methods,* 19**,** 163 – 193.

MARSDEN, P. V. 1987. "Core Discussion Networks of Americans.". *American Sociological Review,* 52.

MARSDEN, P. V. 1990. Network data and measurement. *Annual Review of Sociology,* 16**,** 435 - 463.

MARSDEN, P. V. 2002. Egocentric and sociocentric measures of network centrality. 24**,** 407-422.

MIZRUCHI, M. S. & MARQUIS, C. 2006. Egocentric, sociocentric, or dyadic?: Identifying the appropriate level of analysis in the study of organizational networks. 28**,** 187-208.

MORRIS, M. 2004. Network epidemiology: a handbook for survey design and data collection. *International studies in demography***,** 252.

REBEHY, S. & CHUNG, K. S. K. 2013. Measuring Performance & Interrelatedness in Complex Social Networks of Knowledge Workers. *Paper presented at the Advances in Social Network and Mining (ASONAM), Niagara Falls, Canada.*

SCOTT, J. 2000. Social Network Analysis.

SMITH, K. P. & CHRISTAKIS, N. A. 2008. Social Networks and Health. *Annual Review of Sociology,* 34: 405-429 (Volume publication date August 2008).

STATISTICS, N. C. O. H. 2010. National Centre of Health Statistics, http://www.cdc.gov/nchs/nhis/quest_data_related_1997_forward.htm.

STATISTICS, N. C. O. H. 2013. National Center of Health Statistics, http://www.cdc.gov/nchs/nhis/about_nhis.htm.

VERBRUGGE, L. M. 1977. "The Structure of Adult Friendship Choices.". *Social Forces* 56**,** 576-97.

WELLMAN, B. 1979. The Community Question: The Intimate Networks of East Yorkers. *American Journal of Sociology,* 84.