7 Summary and conclusions

In social network studies there is a growing demand for (practical) sampling designs. This demand stems from actual network research which is more and more concerned with studies of large network structures. Unlike studies of small and well-bounded entire networks where the tools of social network analysis (SNA) can be applied, studies of large networks give several problems. A major theoretical problem is that of the significance of social structure. In a small network the significance of social structure is not the same as that of a large social structure. Consequently to describe large network structures with estimators derived from the small entire network concepts is debatable. A related problem with this, is that the network sampling literature mainly focuses on estimators for network properties that describe aspects of total structure; furthermore this literature does not reckon with the practical problems of a network study. For instance to ask some respondent selected in a sample to give relational information about, say, the 200 other persons that constitute the entire network of study, will decrease the quality of the data and is not practical and economic.

To cope with these problems and to elaborate some relevant approach for studying large network structures the following steps were considered in this study:

1. What do we want to know about a large network structure?
2. What kind of network parameters are relevant?
3. What are practical sampling designs to estimate relevant parameters?
4. How to estimate the relevant parameters?

In this final chapter, we will briefly summarize these steps and discuss some future issues.
What do we want to know about a large network structure?

To study a small and well-bounded entire network such as the employees of some small organization, the relational concepts and tools of mainstream SNA will usually be sufficient. The purpose of SNA is to measure the influence of relations on behaviour and, in turn, to measure the influence of behaviour on the relations in a group or network; patterns of relations are referred to as structure, specific patterns are measured by network or structural parameters. Population network parameters intend to summarize aspects of the total structure; they measure the restrictions (positive and negative) of total structure on the behaviour of the network members. For instance, the density is a frequently applied population parameter in SNA. This parameter can be viewed as the average proportion of relations of the actors in a network. In a small network the social distance is usually small between the actors. Therefore one may assume that the density of a small network indicates the restrictions of relational structure for the group at large. In a large network such a kind of influence of total structure on the behaviour of the actors and vice versa is debatable.

Applying the theoretical concepts of SNA in a large network is problematic because the social distances between the actors will tend to be large in large networks. To elaborate an approach to large network structures we propose to describe this structure from the viewpoint of an individual (ego). From ego's perspective the significance of structure depends mainly of the persons with whom he has a direct relation (alters). The alters are important for ego because they can directly influence ego or be influenced by ego. Furthermore also direct contacts of ego's alters who are not alters themselves are supposed to be relevant for ego in certain situations. Such persons are said to be located at a sociometric distance 2 of ego in the large network and called second-order alters. Next, direct contacts of second-order alters who are not first-order or second-order alters themselves could also be considered as relevant for ego. This way of describing a large network structure can be continued until all actors are considered. However, in this study we propose that the structural influence of and for ego in a large network is mainly restricted to his first- and second-order alters. Therefore the influence of a large network structure is mainly locally determined; consequently relevant network population parameters must reflect the overall local structural properties of the network.

What I
What kind of network parameters are relevant?

A natural way to describe a large network by its local structural properties is by focusing on the properties of the population of personal networks. In this study we have considered three types of personal networks:

- $U_0(i)$, the zero-order network of ego $i$ defined as the collection of relations between ego and his first-order alters
- $U_1(i)$, the first-order network of ego $i$ defined by $U_0(i)$ and the collection of relations between the first-order alters of ego $i$
- $U_2(i)$, the second-order network of ego $i$ defined by $U_1(i)$ and the collection of relations between the first-order and second-order alters of ego $i$.

The first- and zero-order network are embedded in the second-order and the zero-order in the first-order network. The decision to distinguish these three types of personal networks stems from empirical and theoretical restrictions; each type of personal network is supposed to reflect empirical and theoretical restrictions imposed by the data collection or the researchers. For instance, limited financial or time resources may result in collecting only first-order networks. Or the substantive research questions simply refer only to zero-order networks. Relations within the three types of personal networks may be directed or undirected.

To elaborate network parameters that describe the structure in the different types of personal networks we adopted the general meaning of a relation in SNA: a relation between two persons is a viewed as a channel for transfer or “flow” of resources (either material or nonmaterial) (Wasserman & Faust, 1994). Focusing attention on a population of personal networks with only one relevant binary relation of interest (a relation exist or does not exist between two persons), we have defined several network parameters. A parameter summarizing the relational structure of a personal network is called a local parameter and is understood as measuring the structural embeddedness of an ego. The corresponding global parameter is defined as the population average of this local parameter.

Several structural embeddedness parameters are introduced for the different types of personal networks (directed and undirected). The number of
relations of ego with his first-order alters is understood as an important parameter. It gives the total number of structural opportunities for ego to channel resources directly, but the number of first-order alters can also be understood as ego’s structural opportunities to channel indirectly resources with second-order alters. In an undirected zero-order network the number of first-order alters is called the degree of ego; in the directed zero-order network this number is called ego’s outdegree. The number of first-order alters who have an arc to ego (mutual relation) is called ego’s reciprocated indegree (arcs from actors to ego with whom ego himself has no direct relation are not considered).

Structure in first-order networks (directed or undirected) is summarized by two graph-theoretical concepts: the ego-triad and the sociometric distance. An ego-triad is defined as a triple of actors that consists of the (possible) relations between ego and two first-order alters. In the undirected first-order network only two types of ego-triads can be distinguished; the transitive (there is a relation between ego’s first-order alters) and intransitive (there is no relation between ego’s first-order alters) ego-triad (see Figure 3.1). Recall that a relation is understood as a structural opportunity to channel resources. Consequently a relation between two first-order alters of ego is understood as an additional indirect opportunity for ego to channel resources. The proportion of transitive ego-triads is the density of relations in ego’s first-order network. This measure expresses for ego his degree of additional indirect opportunities of the structure between his first-order alters. For the directed first-order network 10 different types of ego-triad types are defined (see Figure 3.6). Based on the relations from first-order alters to ego a further distinction is made:

- directed ego-triads where both first-order alters do not have a direct relation to ego
- directed ego-triads where one of the first-order alters has a direct relation to ego and the other has not
- directed ego-triads where both first-order alters have a direct relation to ego.

Structural opportunities of different types are specific concepts that each the indirect structural opportunities of ego’s second-order alters.
Structural embeddedness parameters are defined for relevant combinations of the different types of directed ego-triads (see formula 3.15). The relevance of a specific combination is dependent on the appropriate theoretical problem. Note that each type of directed ego-triad can be understood in terms of additional indirect structural opportunities for ego.

The sociometric distance between two first-order alters \(j\) and \(h\) of ego is defined as the length of the shortest path between \(j\) and \(h\) in the undirected case, and from \(j\) to \(h\) in the directed case. This concept is mainly used to summarize extra additional opportunities to channel resources for ego. An extra additional opportunity for ego to benefit from the channeling of resources between his first-order alters is in this study restricted to sociometric distance 2, i.e. first-order alters \(j\) and \(h\) do not have a direct relation but they both have a direct relation with a mutual first-order alter \(m\). The number of sociometric distances smaller or equal than 2 is used to define a personal segmentation index (to what extent is a first-order network segmented into hardly connected groups of first-order alters). Related proportional measures such as the proportion of first-order alters connected at a mutual or directed distance 1 or 2 are also defined.

The structure of the second-order network (directed and undirected) is summarized in terms of the influence on ego’s first-order alters. First the number of second-order alters of ego, and related measures such as the number of second-order alters ego may reach via his mutual connected first-order alters, are viewed as important structural embeddedness parameters. Second, the concept of sociometric distance is applied to ego’s second-order network in order to summarize the influence second-order alters may have on the structure of ego’s first-order network. This influence is measured by the number of sociometric distances 2 between ego’s first-order alters by taking into account also his second-order alters. A sociometric distance larger than 2 between a pair of first-order alters (considering all other first-order alters) is reduced to distance 2 if they have a mutual second-order alter. Consequently the personal segmentation index and related proportions can be extended by taking into account the structural influence of ego’s second-order alters.
All proposed structural embeddedness parameters are illustrated with personal network data from a sample of heroin users of the drug assistance in the city of Utrecht.

**What are practical sampling designs for estimating structural embeddedness parameters?**

In the network sampling literature several sampling designs have been introduced to estimate population parameters such as the number of arcs, dyads, and triads (Frank, 1971 and further). These designs, however, do not reckon with the local significance of structure for actors in a large network. To collect data for the proposed structural embeddedness parameters a two-stage sample design is used: first a simple random sample of egos is drawn from the entire network, and ego is asked to mention his first-order alters, second (if practical) a simple random sample from the first-order alters of ego is drawn. An important distinction between a conventional and a network survey is the distinction between sampling procedure and method of measurement. In a network survey the structural part of the network to be observed is dependent on the method of measurement; in a survey often only individual characteristics are observed.

To estimate the introduced structural embeddedness parameters we use in this study as common sampling procedure the two-stage sample design; the methods of measurement differ. The different types of observational procedures reflect the various empirical restrictions such as the limited time and financial resources one may meet in a network survey. Basically two methods of measurement are used: all relations between sampled and between sampled and non-sampled first-order alters are observed or only the relations between sampled first-order alters are observed. The first is called measurement method I, the second measurement method II. In our study a further distinction is made between self-reported and perceptual information. Self-reported information is defined as the relational information given by a sampled person about his own relations. Thus ego as well as a sampled first-order alter may give self-reported information. Perceptual information is relational information given by ego about relations between his first-order alters, i.e. in the first-stage of the sample ego may give his perception about relations between (all or a sample of) his first-order alters. To collect second-order alters each selected first-order alter is asked to mention persons that are not sampled (I or II), we call those measurement errors are made.

Based on structural embeddedness parameters, we use in this study the Horvitz-Thompson estimator to estimate graph totals. A set of ordered valued dyads or valued triads is counted for such parameters and if the values are not observed, and if the value is not known.

To estimate the introduced structural embeddedness parameters we use in this study as common sampling procedure the two-stage sample design; the methods of measurement differ. The different types of observational procedures reflect the various empirical restrictions such as the limited time and financial resources one may meet in a network survey. Basically two methods of measurement are used: all relations between sampled and between sampled and non-sampled first-order alters are observed or only the relations between sampled first-order alters are observed. The first is called measurement method I, the second measurement method II. In our study a further distinction is made between self-reported and perceptual information. Self-reported information is defined as the relational information given by a sampled person about his own relations. Thus ego as well as a sampled first-order alter may give self-reported information. Perceptual information is relational information given by ego about relations between his first-order alters, i.e. in the first-stage of the sample ego may give his perception about relations between (all or a sample of) his first-order alters. To collect second-order alters each selected first-order alter is asked to mention persons that are not sampled (I or II), we call those measurement errors are made. Based on structural embeddedness parameters, we use in this study the Horvitz-Thompson estimator. A set of ordered valued dyads or valued triads is counted for such parameters and if the values are not observed, and if the value is not known.

Some results are given in summary and conclusions.
parameters that are absent on the submitted list (generated by measurement method I or II), we call this the second-order measurement method. When no measurement errors are made, and the second-order measurement method is applied, measurement methods I and II result in the same relational data.

Based on the work of Frank (1971, and further), possibilities are investigated to estimate structural embeddedness parameters in terms of Horvitz-Thompson estimation theory. Frank (1971, and further) showed that the classical Horvitz-Thompson theory could be extended to entire networks for estimating graph totals. A graph total is defined as the sum of a real-valued function on the set of ordered pairs of vertices in a graph. Examples are the number of certain valued dyads or simply the number of lines in a graph. HT estimation is possible for such parameters only if the inclusion probability of a pair of vertices is known and if the value for the sampled pair can be uniquely determined from the sample data. Although Frank elaborated his survey sampling techniques for entire networks, the results can also be applied to personal networks by considering each personal network as an entire network. Model-based estimators are discussed for those structural embeddedness parameters that cannot be estimated by HT-estimation theory.

**Some results and conclusions**

In general HT-estimators can be defined for structural embeddedness parameters based on the number of pairs at distance 1; to estimate structural embeddedness parameters based on the number of pairs at distance 2 mainly model-based estimators must be defined. To define an unbiased estimator of the variance of a structural embeddedness HT-estimator the first-order alter sample must be larger or equal to 4. In general the variance of a structural embeddedness estimator will be large. This is because the size of a first-order alter sample will be rather small. Consequently, method of measurement I will usually lead to lower variance estimates than method II, simply because the first method collects more structural information than the second. The various estimators are illustrated by a (directed and undirected) personal network of an employee in some civil organization.
For the number of undirected relations, i.e. the number of transitive ego-triads, between ego’s first-order alters an unbiased HT estimator can be defined, as well for method of measurement I as for II. Because relational data collected according to measurement method I will be larger than for measurement method II, the variance of the size estimator for the first method will be smaller than the second. To estimate the number of arcs from first-order alters to ego (reciprocated indegree) also an unbiased HT-estimator is defined. However, to estimate the total counts of (some combination of) the 10 directed ego-triads, only method of measurement II can be applied. Only for those first-order alters selected in the first-order alter sample their arc to ego is known. This restriction has its influence on the corresponding variance estimators, i.e. variances tend to be high because only relations between sampled first-order alters can be used. A HT-estimator can also be defined for the number of arcs between first- and second-order alters. In this case the number of arcs from the selected first-order alters is viewed as an individual attribute.

To estimate the number of pairs at sociometric distance 2 HT-estimators cannot be defined except in the undirected case when applying measurement method I. Here for each pair observed in the sample, observing a distance 2 means that the distance in the population is 2. For the other distance 2-based parameters a sampled distance larger than 2 may be actually 2 in the population. To cope with such problems we have investigated some model-based estimators for estimating the number of pairs at (directed and undirected) sociometric distance 2. The model is a simple Bernoulli graph model. A key assumption for these model-based estimators is that the probability of observing an arc within the sample approximates the probability of an arc outside the sample. This may be an unrealistic assumption, and other model-based estimators must be considered in the future.

Two model-based estimators are proposed to estimate ego’s number of second-order alters. The first is again an estimator based on the model of a Bernoulli graph, the second a multiple recapture estimator. Both have disadvantages; for the Bernoulli estimator the probability of an arc must be estimated from the sample data, for the multiple recapture estimator each sampled first-order alter must have at least one unique second-order alter. This is a strong restriction that can only be met in formal\(^2\).
Summary and conclusions

As a summary: from a statistical point of view it is possible to estimate the proposed structural embeddedness parameters (directed and undirected) in a rather formal way. For undirected distance-1 based parameters unbiased HT-estimators are always possible irrespective which method of measurement is applied. For directed ego-triads parameters unbiased HT-estimators are only possible for measurement method II. For distance-2 based parameters a HT estimator is possible for only one special case, the other cases must be estimated using model-based estimators. Although statistical inference is possible, we observe two empirical weaknesses of a network survey based on self-reported information only. First, the estimates will be rather unprecise because of the large variances due to small sample sizes. Second in order to estimate the variance of the HT estimators at least 4 first-order alters must be selected per ego which may be a too strong restriction in empirical research.

There are several arguments to use perceptional reported relational data besides the conventional collected self-reported relational data in a network survey. A practical argument is the relatively cheap and easily way of collecting such data. Relational information given by ego about (possible) relations between his first-order alters is not very time expensive (ego is interviewed anyway). A statistical argument is the possibility to define several types of difference estimators, which make use of perceptional information in combination with the later collected self-reported data of first-order alters. Main purpose of this combination is to find more precise estimators for the structural embeddedness parameters. By using difference estimators for different methods of measurements, it was investigated whether perceptional data could be used to lower the variances of the structural embeddedness parameters. Some examples were given for cases in which the variances of corresponding self-reported estimators were decreased by using difference estimators based on perceptional data. To relax the "four first-order alters per ego" condition for estimating unbiased variances of local network...
parameters the use of empirical Bayes estimators was investigated. Especially in combination with perceptual information such estimators may relax this condition. However, this part of the study must be viewed as a beginning for further research about the role of perceptual data in network surveys.

Samenvatting

Dit proefschrift is gewijd aan het onderzoek naar structurele kenmerken van sociale netwerken. Deze methoden zijn in het bijzonder gericht op analysemethoden die in staat zijn om perceptuele informatie te gebruiken om een betere beschrijving van een populatie te verkrijgen. Bijvoorbeeld, de politieke voorkeuren van een populatie kunnen worden bepaald door te onderzoeken welke persoon de meeste invloed heeft op de populatie. Dit kan worden gedaan door te onderzoeken welke persoons eigenschappen het meeste invloed hebben op de populatie.

Veronderstel dat sociale contacten invloed hebben op de invloed van netwerkonderzoek. Deze invloed is van sociale contacten die iedere persoon in de populatie kan hebben, maar ook naar de invloed van een persoon wordt gedefinieerd als functionele eigenschappen. Deze invloed wordt opgevat als een maat voor de invloed van zijn relationele eigenschappen. Technieken van de sociale netwerken kunnen echter een probleem zijn bij het beoordelen van invloed van netwerkonderzoek.