TOM A. B. SNIJDERS

MARINUS SPREEN

Segmentation in personal networks

SEGMENTATION IN PERSONAL NETWORKS

Tom A.B. SNIJDERS[1] , Marinus SPREEN[1]

RÉSUMÉ — La segmentation des réseaux personnels

On propose un concept et plusieurs mesures de la segmentation des réseaux personnels. Les auteurs défendent la thèse que les implications de la segmentation des réseaux personnels sont, d'un certain point de vue, opposées de celles des réseaux totaux. Les mesures sont illustrées par l'exemple d'un réseau de relations de confiance dans un service de fonctionnaires. Des estimateurs du degré de segmentation sont proposés pour le cas où les relations dans le réseau personnel sont observées sur la base d'un échantillonnage plutôt que dans leur totalité.

SUMMARY — *A concept and several measures for segmentation of personal networks are proposed. It is argued that the implications of segmentation of personal networks are, in a sense, the opposite of those of segmentation of entire networks. The measures are illustrated by the example of the trust network in a civil service department. For the case where relations in the personal network are observed by a sample rather than completely, estimators for the segmentation measures are given.*

INTRODUCTION

The concept of segmentation was proposed in Baerveldt & Snijders (1994) as a structural concept for complete networks. The defined segmentation index intends to measure the degree of division of a social network into subgroups with high within-group and low between-group densities. The theoretical definition of segmentation is given as follows (Baerveldt & Snijders, p.214):

> *Segmentation is the degree to which there is, for actors in the network, a contrast, or distance, between their personal network and the rest; one might say, between in-group and out-group.*

In this paper a local, or ego-oriented, concept of segmentation is introduced by using personal networks as unit of analysis. Differences in interpretation of segmentation of personal and of complete networks will be briefly discussed and some hypotheses will be formulated. The personal segmentation index will be illustrated by data from a network study in a civil organisation conducted and reported by Bulder, Flap & Leeuw (1993). For those situations in which it is impossible to interview all members of the personal network (too expensive, too much time, etc.), the personal network index must be estimated; some sampling possibilities will be discussed.

---

[1] Department of Statistics & Measurement Theory/ICS, University of Groningen, The Netherlands.

SEGMENTATION IN COMPLETE NETWORKS

Consider an undirected graph with N vertices and adjacency matrix X, representing a set of social actors and some relationship between them. The adjacency matrix is defined by $X_{ij} = 1$ if there is an edge between vertices i and j, and $X_{ij} = 0$ otherwise ($X_{ii} = 0$ for all i). Since the graph is undirected $X_{ij} = X_{ji}$ for all i, j. The degree of vertex i is $X_{i+} = X_{+i} = \Sigma_j X_{ij}$. We say that i and j are acquainted if $X_{ij} = 1$.

The segmentation index is based on the concept of sociometric distance in a graph, defined as the length of the shortest path between two vertices in a graph: $d(i,i) = 0$, $d(i,j) = 1$ if there is a relation between i and j, $d(i,j) = 2$ if there is no relation between i and j but both are acquainted with a common vertex k, etc.. If no path is observed between i and j, i.e., they are vertices in different components of the graph, the sociometric distance is infinite, denoted $d(i,j) = \infty$. In an undirected graph of N persons, there are $N(N-1)/2$ distances, some of which may be infinite. The segmentation index reflects a particular kind of dispersion of these $N(N-1)/2$ distances. The mathematical definition for segmentation in a complete network defined by an undirected graph on N vertices is the following:

- Define by D_r the number of pairs of vertices at mutual distance r;
- Define the fraction of pairs of vertices at mutual distance r by

$$F_r = \frac{2D_r}{\{N(N-1)\}} \quad ;$$

- Define the fractions of pairs at distance r or greater by

$$P_r = F_r + F_{r+1} + \ldots + F_{N-1} + F_\infty = 1 - \Sigma_{q \geq r} F_q \quad .$$

Then F_1 is the density of the graph while $P_2 = 1 - F_1$ is the fraction of pairs of vertices that are not acquainted. The segmentation measure is then defined as:

$$S_r = \frac{P_r}{P_2} ,$$

i.e., the fractions of vertex pairs at distance r or greater among those pairs that are not directly connected. It is briefly argued in Baerveldt & Snijders (1994) that in a complete network distance 2 may be interpreted as a close distance, 3 as intermediate, and 4 or more as long. Therefore Baerveldt & Snijders propose to work with S_3 or S_4. Note that:

$$0 \leq S_\infty \leq S_{N-1} \leq \ldots \leq S_4 \leq S_3 \leq S_2 = 1 ,$$

and that $S_r = 1$ for $r \geq 3$ if and only if the graph consists of 2 or more disconnected cliques.

To illustrate the segmentation index consider Figure 1. Figure 1 is the network of relationships based on mutual trust among 29 employees in a civil organization as reported at page 46 in Bulder et al. (1993). The research about this organisation is also discussed in Bulder, Leeuw, and Flap (1996). Mutual trust relations are defined as those relations between employees in which the subjects of discussion are not solely "neutral" work-related topics, such as the future of the organization. The authors consider trust relations as an indication for a certain degree of personal involvement with the organization and for a certain degree of trust in each other. Therefore the relations are symmetrical by nature.
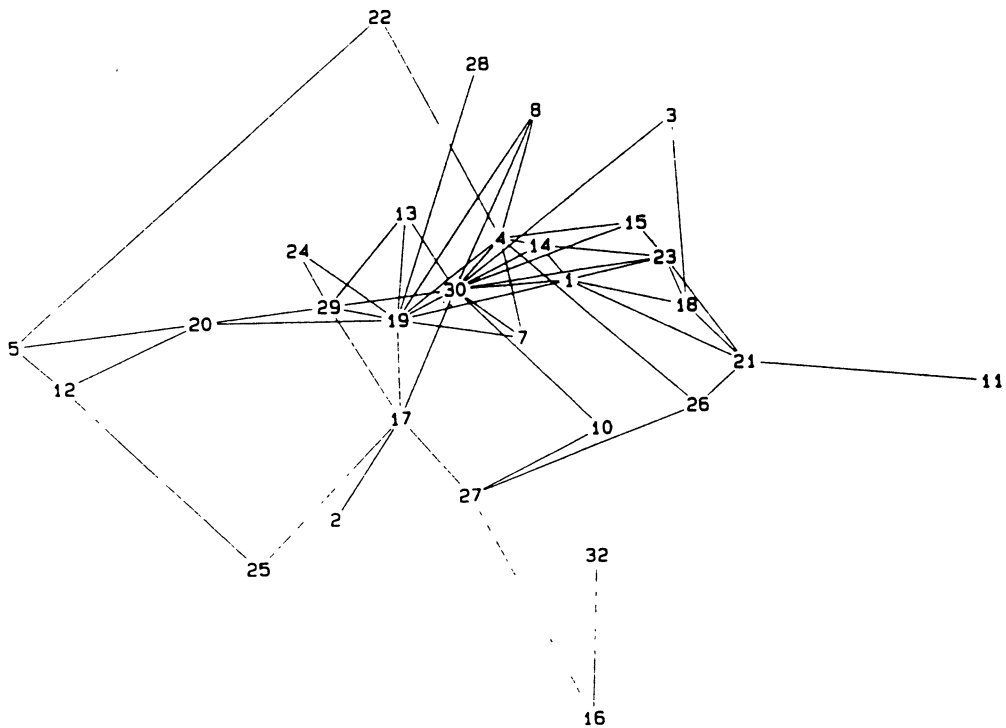
Figure 1. Mutual trust relations in civil organization

To determine the frequency distribution of sociometric distances between the 406 relations the network program GRADAP (Sprenger & Stokman, 1989) was used (see table 1).

Distance	Frequency	Fractions F_r
1	57	.14
2	146	.36
3	129	.32
4	53	.13
5	20	.05
6	2	.00
TOTAL	406	1.00

Table 1. Frequency distribution of distances within the network

Note that the density of the network is $F_1 = .14$, and that the network consists of 1 connected component, so that infinite distances do not occur. The frequency distribution of distances leads to the following segmentation measures:

$$S_3 = .58 \quad \text{and} \quad S_4 = .21 .$$

SEGMENTATION IN PERSONAL NETWORKS

The concept of segmentation make sense also in personal networks. The personal network of ego may be said to be highly segmented when the set of his direct acquaintances (alters) is composed of two or more subgroups with high within-group and low between-group densities. The theoretical definition of segmentation of the personal network is analogous to that of complete networks:

Segmentation is the degree to which, within the set of actors directly connected to ego (the "alters"), there is for each alter a gap between the other alters to whom he is himself directly connected, and the other alters to whom he is not directly connected.

The segmentation of ego's personal network is defined as the segmentation of the network formed by his alters. The segmentation index thus is based on distances between pairs of alters. The formula will be given below.

The interpretation of segmentation of personal networks is quite distinct from the interpretation of segmentation of complete networks. If ego has a highly segmented network, then:

(a) - ego has more opportunities to "manage" what others know about him: he can offer different views of himself to alters who are not, or only distantly, related to each other; this will give him more freedom of action;

(b) - it is likely, although not necessary, that there will also be differences in the characteristics of the alters (segmentation is often associated with segregation as described by Freeman, 1978); if this is the case, then ego has access to more varied social resources, and his social capital will be larger.

This implies that, in a certain sense, the consequences of this local version of segmentation are just the reverse of those of the global segmentation:

• persons/actors are less influenced by others when they have a more strongly segmented personal network;

• behaviour of persons/actors is less predictable when they have more strongly segmented personal networks.

This reversion can be easily understood by noticing that a person with a highly segmented personal network is a "bridge" between parts of the network that, without him, would have been not, or only distantly, connected. Therefore, individuals with highly segmented personal networks lead to a smaller segmentation of the complete network. Note that this way of reasoning is similar to the "structural holes" theory as proposed by Burt (1992, 1995). The concept of personal network segmentation, however, is different from the concept of structural holes. The latter is based on the strength of the strategic position of ego with respect to his alters. E.g., when the personal network is composed of two disconnected cliques then the segmentation is maximal whereas Burt's measure for structural holes is moderate.

Just as in the case of segmentation of a complete network, the *density* is an essential concept that can be regarded as more basic than segmentation. The segmentation measure must control in some sense for the density. For the mathematical notation we define:

$$U_1(i) = \{j{\neq}i \mid X_{ij} = 1\} = \{j \mid d(i,j) = 1\},$$

the *neighbourhood* at distance 1 of vertex i. This neighbourhood will be identified with the graph generated by vertex set $U_1(i)$. Both the vertex set and the graph commonly are referred to as the (first-order) personal network of i. The number of elements of $U_1(i)$ is the degree of vertex i, denoted X_{i+}.

The sociometric distance within the graph $U_1(i)$ is denoted d_{1i}; thus, for vertices j and h that both are elements of $U_1(i)$, $d_{1i}(j,h)$ is the length of the shortest path contained in $U_1(i)$ between j and h. This distance cannot be shorter than the distance in the entire network, so that $d(j,h) \leq d_{1i}(j,h)$.

The density of the personal network is the density of the graph $U_1(i)$,

$$\eta(i) = \frac{\displaystyle\sum_{j,h \,\in\, U_{1(i)}} X_{jh}}{X_{i+}^{(2)}} \,,$$

where $X_{i+}^{(2)} = X_{i+} (X_{i+} - 1)$.

The segmentation of the personal network is defined as:

$$S_r(i) = \frac{\#\,\{ (j,h) \mid j,h \in U_1(i), \, d_{1i}(j,h) \geq r \, \}}{\#\,\{ (j,h) \mid j,h \in U_1(i), \, d_{1i}(j,h) \geq 2\}}$$

$$= \frac{\#\,\{ (j,h) \mid j,h \in U_1(i), \, d_{1i}(j,h) \geq r \, \}}{(1 - \eta(i) \, X_{i+}^{(2)}} \,.$$

Note that:

$$(1 - \eta(i) \, X_{i+}^{(2)} = \#\,\{ (j,h) \mid j,h \in U_1(i), X_{jh} = 0\} \,,$$

i.e., the number of pairs in the personal network that are not mutually directly related.

If $\eta(i)$ is low, then there are few direct relations between the acquaintances of i. The parameter $\eta(i)$ can be interpreted as a local transitivity parameter: the degree to which i's acquaintances are mutually acquainted among themselves. If $S_r(i)$ is high, then there is, among the acquaintances of i, the tendency that either they are directly related, or they are far apart (as measured within the set of i's acquaintances).

For an illustration of $S_r(i)$ consider the first-order network of vertex *17* in Figure 1. The degree of vertex *17* is *6* and its neighbourhood is $U_1(17) = \{2,19,25,27,29,30\}$. Figure 2 gives the corresponding graph.
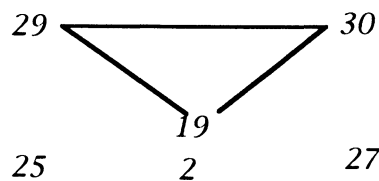

Figure 2. Graph $U_1(17)$

Of the *15* pairs of alters in graph $U_1(17)$, *3* pairs are at distance *1* and *12* pairs at distance ∞. The density is low; $\eta(17) = .10$. The personal segmentation index for node *17* is $S_\infty(17) = 1$, i.e., the personal network is maximally segmented because the graph consists of a 1-clique and *3* isolated nodes.

Which distances may be considered to be long, when measured within first order personal networks? There are two reasons why it is sensible to use a lower threshold for "long" distances in personal than in complete networks. In the first place, personal networks will often be situated within smaller social circles than complete networks, so the same absolute value for a distance may be, relative to an overall distribution of distances, considered larger in a personal than in a complete network. In the second place, when two persons are distantly related within the personal network of individual *i*, e.g., $d_{1i}(h,j) = 4$, it is quite well possible that in the surrounding complete network they are more closely related, e.g., $d(h,j) = 2$ or *3*. For the meaning of the indirect social relationship between two acquaintances of *i*, it is of minor importance whether the shortest social path between them consists entirely of other individuals who are also acquaintances of *i*, or whether this shortest path is not contained within the network of *i*'s direct acquaintances. It was remarked above that $d_{1i}(h,j) \geq d(h,j)$ for all *i, h, j*. However, $d_{1i}(h,j) = 1$ if and only if $d(j,h) = 1$: direct relations between *i*'s acquaintances can be traced already within his first-order personal network.

For the two reasons mentioned, we shall interpret a sociometric distance of *3* within a personal network already as a large distance. The proposed measure of local segmentation is:

$$S_3(i).$$

The preceding argument implies that it also makes sense to consider distances between acquaintances *j* and *h* of person *i* as shortest lengths of paths not necessarily contained within *i*'s personal network $U_1(i)$. Since we are thinking of personal networks, it is of doubtful applicability to try to study large distances between individuals as defined in the whole network: large distances will be exceedingly hard to measure. We restrict attention to distances in the second-order personal network, defined a:

$$U_2(i) = \{j \mid d(i,j) = 1 \text{ or } 2\} .$$

All definitions given above for segmentation measures can be repeated, substituting distances within $U_2(i)$ for distances within $U_1(i)$. The distance within the graph $U_2(i)$ is denoted d_{2i}.

An alternative segmentation index of the personal network is defined by:

$$S_r'(i) = \frac{\# \{(j,h) \mid j,h \in U_1(i), d_{2i}(j,h) \geq r \}}{\# \{(j,h) \mid j,h \in U_1(i), d_{2i}(j,h) \geq 2\}}$$

$$= \frac{\# \{(j,h) \mid j,h \in U_1(i), d_{2i}(j,h) \geq r \}}{(1-\eta(i)) X_{i+}^{(2)}} .$$

The second equality holds because $d_{1i}(j,h) = 1$ if and only if $d_{2i}(j,h) = 1$ if and only if $d(j,h) = 1$ (see above). For segmentation measures of personal networks that refer to the direct as well as the indirect social surroundings of the focal individual, we propose the measure:

$$S_3'(i) .$$

Since for direct acquaintances j and h of i, the distance $d_{2i}(j,h) = 2$ if and only if $d(j,h) = 2$, the definition of $S_3'(i)$ can also be given as

$$S_3'(i) = \frac{\#\ \{(j,h)\ |\ j,h \in U_1(i),\ d(j,h) \geq 3\}}{\#\ \{(j,h)\ |\ j,h \in U_1(i),\ d(j,h) \geq 2\}}.$$

This means that for the definition of $S_3'(i)$, it is not necessary to make reference specifically to $U_2(i)$.

To compare the relevance of the two proposed measures $S_3(i)$ and $S_3'(i)$, the following considerations can be put forward:

• Measure $S_3'(i)$ is conceptually more meaningful, because for the indirect relations between individuals in i's social surroundings, it is a rather irrelevant restriction to measure them by paths that must be contained within i's direct social surroundings.

• For the determination of $S_3'(i)$, more information is required than for the determination of $S_3(i)$; therefore, $S_3(i)$ can be used in cases when the value of $S_3'(i)$ cannot be determined.

• The influence of the second-order network on the segmentation of ego's direct social environment can be expressed as $\Delta(i) = (S_3(i) - S_3'(i))$. This is non-negative.

To calculate these local segmentation measures, it is not necessary to observe the total network: it is sufficient to observe the relations in the ego-centered network ($U_1(i)$ or $U_2(i)$, respectively).

For an illustration of this measure, consider figure 3, which is the second order personal environment of vertex 17 in the total network. The dashed lines are the relations of the alters of vertex 17 to vertices in $U_2(i)$ that are not relevant for computing $S_3'(i)$. The circle illustrates the boundary between $U_1(17)$ and $U_2(17)$.
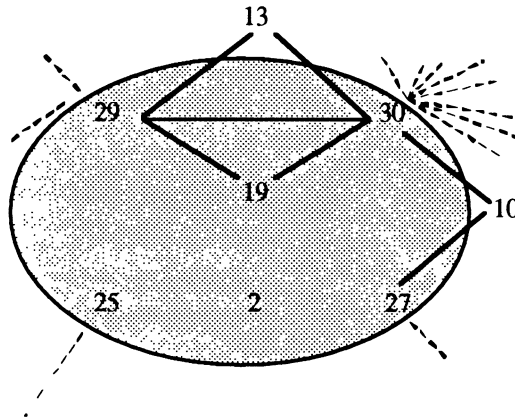

Figure 3. Graph $U_2(17)$ (the part that is relevant for computing $S_3'(i)$).

Only those persons in ego's second-order environment who reduce the distance between pairs of alters in the first-order environment, are relevant for computing the personal segmentation index.

In figure *3*, vertex *10* reduces the sociometric distance between pair (*27, 30*) from infinite to 2. Vertex *13* also connects two alters of ego, but the sociometric distance between these two alters remains *1*. This means that by taking into account the second-order network of vertex *17* the frequency distribution of distances changes into *3* pairs at distance *1*, one pair at distance 2, and *11* pairs at distance ∞, which results in $S_3(i)= .92$. This implies that the segmentation of the environment of vertex *17* remains high, but the direct environment of vertex *17* consists now of *1* component of *4* vertices and 2 isolated vertices.

Table *2* gives a summary of all personal (local) segmentation indices of the employees in the confidence network.

Vertex	degree	$\eta(i)$	$S_3(i)$	$S'_3(i)$	$\Delta(i)$
1	6	.23	.25	.25	0.00
2	1	–	–	–	–
3	2	0.00	1.00	0.00	1.00
4	8	.13	.62	.62	0.00
5	3	.17	1.00	0.00	1.00
7	3	1.00	0.00	0.00	0.00
8	3	1.00	0.00	0.00	0.00
10	2	0.00	1.00	0.00	1.00
11	1	–	–	–	–
12	3	.17	1.00	0.00	1.00
13	3	1.00	0.00	0.00	0.00
14	4	.33	0.00	0.00	0.00
15	3	.33	0.00	0.00	0.00
16	2	0.00	1.00	1.00	1.00
17	6	.10	1.00	.92	.08
18	4	.25	1.00	0.00	1.00
19	11	.12	.43	.43	0.00
20	4	.17	1.00	1.00	0.00
21	5	.15	1.00	1.00	0.00
22	2	0.00	1.00	1.00	0.00
23	6	.23	.25	.25	0.00
24	2	1.00	0.00	0.00	0.00
25	2	0.00	1.00	1.00	0.00
26	3	0.00	1.00	.67	.33
27	4	0.00	1.00	.83	.17
28	1	–	–	–	–
29	6	.23	0.00	0.00	0.00
30	13	.11	.56	.52	.04
32	1	–	–	–	–

Table 2. Personal segmenttion indices of confidence network

It is obvious that for vertices of degree 1, the personal network segmentation indices are not defined. For vertices of degree 2 and 3, $S_3(i)$ and $S_3'(i)$ are defined, but for such low degrees the interpretation of these indices is somewhat troublesome. The segmentation indices of a vertex of degree 2 can have two values: $S_3(i) = 1$ if no line is observed between the two alters or 0 otherwise; $S_3'(i) = 1$ if there is no person connecting both alters or 0 otherwise. A vertex of degree 3 can have two values for $S_3(i)$: 1 if no lines or only 1 line is observed in $U_1(i)$ or 0 otherwise; $S_3'(i)$ can already have 5 values (see figure 4, values 0 and 1 omitted).

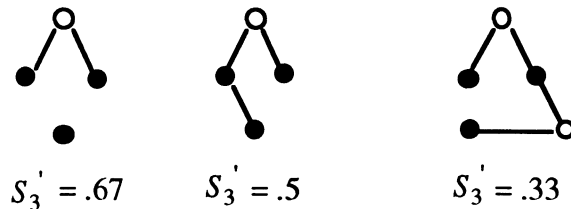$$S_3' = .67 \qquad S_3' = .5 \qquad S_3' = .33$$

Figure 4. Possible values $S_3'(i)$ strictly between 0 and 1 for degree 3 if $S_3(i) = 1$
(vertices in $U_1(i)$ are black, in $U_2(i)$ are white)

The influence $\Delta(i)$ of the second-order network on the segmentation of ego's direct environment for vertices of degree 2 is 0 or 1. For vertices of degree 3, $\Delta(i)$ is $0, .33, .5, .67$, or 1. For these cases, $\Delta(i)$ can be used as an indication for examining the degree of connectedness of vertex i in the total network by assuming that vertices of degree 2 or 3 are more isolated or peripheral. In table 2 there are 13 vertices of degree 2 or 3, of which 8 have a maximal first-order segmentation. For 5 of them (vertices $3, 5, 10, 12$, and 26) their second-order network does count for a reduction of the segmentation. This implies that they are more influenced by the total network structure than the other vertices whose first-order network is not influenced by total network structure. In other words, the other 8 vertices can be viewed as more isolated or peripheral.

For the 12 vertices of degree 4 or higher the following observations in the first-order network can be made:

- The personal networks of 2 vertices are not segmented $(S_3(i) = 0)$.
- The personal networks of 5 vertices are segmented in the range $.25 - .62$.
- The personal networks of 5 vertices are maximally segmented $(S_3(i) = 1)$.

Examining the influence $\Delta(i)$ of the second-order environment for vertices i with degree 4 or higher, note that only 3 vertices ($17, 27, 30$) show positive influence of this environment. However, these influences are rather low. Note that vertices 4 and 19 with both a relatively high degree and high first-order segmentation have the same second-order segmentation. Also vertex 30 with the highest degree has more or less the same first-order and second-order segmentation index. Recalling that the density of the total network is low $(.14)$, as a conclusion one might say that the segmentation of mutual trust relations in the direct personal environments of the employees is quite locally determined.

ESTIMATION OF $S_3(i)$ AND $S_3'(i)$

The mutual trust network is a relatively small network which can be "easily" observed. For larger networks in which it is impractical to interview all network members, a sample is needed to estimate the proposed segmentation indices. For instance, interviewing only 10 ego's each mentioning 15 alters already requires 150 interviews to compute the indices.

To compute segmentation indices $S_3(i)$ and $S_3'(i)$ in personal network data, several steps are needed in the data collection procedure.

1. From individual i, in his/her role of "ego", the set of his acquaintances is elicited through a name-generating procedure.
2. All the acquaintances j of ego i are presented with a list of other acquaintances (known from step 1), and each j is asked to indicate for each of the other acquaintances whether or not j has the particular type, or types, of relation with him/her.
3. All the acquaintances j of ego i are interviewed and for each alter j the set of his/her acquaintances is elicited through a name-generating procedure. Because of step 2, it is sufficient to ask information on the set of j's acquaintances that are not already contained in the set of i's acquaintances.

To compute $S_3(i)$ step 1 and 2 are enough, to compute $S_3'(i)$ step 3 is needed. For situations in which these three steps can be applied, both segmentation indices can be computed. For those situations in which it is impractical to interview each individual or alter of ego i, sampling and observation designs are needed to estimate the segmentation indices. Several sampling and observation designs are possible of which some will be discussed.

SAMPLING AND OBSERVATION DESIGN FOR $S_3(i)$

Step 1: Each ego i mentions his set of acquaintances through some name-generating procedure.

Step 2: A random sample $S(i)$ without replacement of size n is drawn from the acquaintances j of i, and each j is presented the list of all other $(X_{i+} - 1)$ acquaintances to indicate whether or not j has the particular type of relation with them.

Which characteristics are known in this observation situation, and which have to be estimated? From step 1, we know X_{i+}. What is not known and has to be estimated is the density $\eta(i)$ and the number of pairs of i's acquaintances at distance 3 and larger within .

For estimating $\eta(i)$, the number of relations $R_1(i)$ in $U_1(i)$ must be computed. This number is called the size of graph $U_1(i)$. Several statistical graph-size estimators are discussed by Capobianco & Frank (1981). First, step 2 is considered.

A random sample $S(i)$ of size n is drawn from $U_1(i)$. The probability that acquaintance j is drawn is denoted $\pi_j = \frac{n}{X_{i+}}$. For each $j \in S(i)$ his relations with the $(X_{i+} - 1)$ acquaintances are observed. Capobianco & Frank (1981) define various estimators for $R_1(i)$. One of these, which we also propose to use here, was denoted in their paper by $\widehat{R}$ ' and is given by

$$\widehat{R}_1(i) = \frac{1}{1 - q_2}\left[\frac{1}{2}\sum_{j,h \in S(i)} X_{jh} + \sum_{j \in S(i), h \in U_1(i)\backslash S(i)} X_{jh}\right]$$

where

$$q_2 = \left[\frac{(X_{i+} - n)(X_{i+} - n - 1)}{X_{i+}^{(2)}} \right],$$

the probability that neither of 2 specified vertices are in the sample.

Note that we can write:

$$S_3(i) = \frac{X_{i+}^{(2)} - R_1(i) - R_2(i)}{X_{i+}^{(2)} - R_1(i)},$$

where $R_2(i) = \# \{ (j,h) \mid j,h \in U_1(i), d_{1i}(j,h) = 2 \}$.

We now have estimated $R_1(i)$, and we still must find an estimator for $R_2(i)$.

For vertices j and h in $U_1(i)$, a distance $d_{1i}(j,h)$ is observed in the sampling design if they are not directly related (i.e., $X_{jh} = 0$) while at least one vertex k is observed acquainted to them both (i.e., $max_k X_{jk} X_{hk} = 1$). The Horvitz-Thompson estimator (cf. Cochran (1977) or another text on sampling theory) for $R_2(i)$ is the number of such pairs (j,h) divided by the probability that j and h are in the sample,

$$\widehat{R}_2(i) = \frac{1}{p_2} \sum_{j,h \in S(i)} (1 - X_{jh}) \, max_{k \in U_1(i)} X_{jk} X_{hk},$$

where $p_2 = \dfrac{n(n-1)}{X_{i+}^{(2)}}$. This leads to the estimator

$$\widehat{S}_3(i) = \frac{X_{i+}^{(2)} - \widehat{R}_1(i) - \widehat{R}_2(i)}{X_{i+}^{(2)} - \widehat{R}_1(i)}.$$

SAMPLING AND OBSERVATION DESIGN FOR $S_3'(i)$

To estimate the second-order personal network segmentation index $S_3'(i)$, step 2 must be replaced by step 3.

Step 3: each selected acquaintance j of ego i is presented with a list of all the other $(X_{i+} - 1)$ acquaintances, and each j is asked to mention his acquaintances on this list and those not on the list of ego i.

For the estimation of $S_3'(i)$, it is convenient to write:

$$S_3'(i) = \frac{X_{i+}^{(2)} - R_1(i) - R_2(i) - H(i)}{X_{i+}^{(2)} - R_1(i)},$$

where $H(i)$ is the number of i's acquaintances who are indirectly related because both are acquainted with the same k, but not with a common k who is himself acquainted with i. Step 3 allows to estimate $H(i)$ in the same way as $R_2(i)$. The estimator can be written as:

$$\widehat{R}_2(i) + \widehat{H}(i) = \frac{1}{p2} \sum_{j,h \in S(i)} (1 - X_{jh}) \max_{\substack{k=1 \\ k \neq i}}^{N} X_{jk} X_{hk}.$$

The estimator for $S_3'(i)$ now follows straightforwardly.

BIBLIOGRAPHY

BAERVELDT, C. & SNIJDERS, T.A.B., 1994, "Influences on and from the segmentation of networks: hypotheses and tests", *Social Networks*, 16, 213-232.

BULDER, A, FLAP, H. D. & LEEUW F. L., 1993, "Netwerken, Overheidsorganisaties en effectiviteit" (Networks, Government Organizations, and Efficiency), Den Haag, Ministerie van Binnenlandse Zaken. (*In Dutch*).

BULDER, A, LEEUW F. L. & FLAP, H. D., 1996, "Networks and evaluating public-sector reforms", *Evaluation*, 2, 261-276.

BURT, R. S., 1992, *Structural Holes: The social structure of competition*, Harvard University Press, Cambridge, Massachusetts, and London, England.

BURT, R. S., 1995, "Le capital social, les trous structuraux et l'entrepreneur", *Revue Francaise de Sociologie*, 36, 599-628.

CAPOBIANCO, M. & FRANK, O., 1982, "Comparison of statistical graph-size estimators", *Journal of Statistical Planning and Inference*, 6, 87-97.

COCHRAN, W.G., 1977, *Sampling techniques,* New York, Wiley and Sons.

FREEMAN, L. C., 1978, "Segregation in social networks", *Sociological Methods and Research*, 6, 411-429.

SPRENGER, C.J.A. & StOKMAN, F.N., 1989, GRADAP: *Graph Definition and Analysis Package*. Groningen, The Netherlands, iec ProGAMMA.
99599746999695494