MATHÉMATIQUES ET SCIENCES HUMAINES

HANS J. HUMMELL
WOLFGANG SODEUR

Structural analysis of social networks with respect to different levels of aggregation

Mathématiques et sciences humaines, tome 137 (1997), p. 37-60
<http://www.numdam.org/item?id=MSH_1997__137__37_0>

Math. Inf. Sci. hum. (35e année, n°137, *1997*, pp.37-60)

STRUCTURAL ANALYSIS OF SOCIAL NETWORKS WITH RESPECT TO DIFFERENT LEVELS OF AGGREGATION

Hans J. HUMMELL and Wolfgang SODEUR[1]

RÉSUMÉ — Analyse structurale des réseaux sociaux selon différents niveaux d'agrégation
Cet article vise à rapprocher deux traditions de recherche, l'analyse multi-niveaux et les réseaux sociaux. On présente dans ce but une stratégie qui trouve son origine dans la typologie des unités d'analyse et de leurs propriétés proposée par Paul F. Lazarsfeld et Herbert Menzel. Leur classification est tout d'abord élargie de manière à prendre en considération un plus grand nombre de concepts relationnels qu'ils n'en avaient retenus à l'époque ; elle est ensuite utilisée pour traduire une question de recherche qui se présente au premier abord comme relevant purement de l'analyse de réseaux en un problème classique d'analyse des relations entre variables concernant différentes unités à différents niveaux d'agrégation. Bien que les données analysées soient choisies dans le domaine de la sociométrie et que les concepts d'équilibre et de transitivité viennent de la psychologie sociale, il devrait apparaître clair que la stratégie de recherche proposée ne voit pas sa pertinence limitée aux conditions théoriques et empiriques propres à l'exemple choisi comme illustration.

SUMMARY — *The article aims at the integration of the two research traditions of multi-level and of network analysis. To this effect, a strategy is presented which can be traced back to P.F. Lazarsfeld and H. Menzel's typology of units and of their properties. After having extended their classification to take account of more network concepts than was needed at their time, the Lazarsfeld-Menzel-Classification is used as a conceptual instrument to translate a research question, which first looks like a specialty of network analysis, into a standard problem of treating relationships between variables of different units at different levels of aggregation. Though the data set to be analyzed is taken from sociometric research and the selected theoretical framework from "balance" and "transitivity" theory of social psychology, it should be clear that the proposed research strategy is not restricted to the empirical or theoretical specificities of the example chosen for its illustration.*

INTRODUCTION

The concepts of multi-level analysis constitute a coherent analytical perspective, which is well established within sociology (cf. e.g. H.J. Hummell 1972). Though its main notions have been primarily developed within non-network research, they can also fruitfully be applied to the structural analysis of social networks. Thus, our aim is to contribute to the integration of the two research traditions of multi-level and of network analysis. To this effect, we will go back to the typology of units and their properties as it was formulated by Paul F. Lazarsfeld and Herbert Menzel in their classical paper, which was published in 1961. We will slightly modify and extend their classification in order to include more of network concepts than was needed at their time. Then this analytical perspective, based on the Lazarsfeld-Menzel-

[1] University of Duisburg, Germany/University of Essen, Germany.

classification, will be used in the substantive analysis of a specific empirical data set within a specific theoretical framework. Though the data set is taken from sociometric research and the selected theoretical framework refers to what in socio-psychological research is called "balance" or "transitivity" theory (cf. e.g. D. Cartright and F. Harary 1956; J.A. Davis 1967; J.A. Davis and S. Leinhardt 1972; P.W. Holland and S. Leinhardt 1971, 1975), we will try to make clear that the research strategy to be presented is not necessarily restricted to the empirical or theoretical specificities of the example chosen for its illustration.

1. INDIVIDUAL AND COLLECTIVE PROPERTIES

In their often-cited article "On the Relation between Individual and Collective Properties" Paul F. Lazarsfeld and Herbert Menzel (1961) suggest a useful classification of "properties" on different levels of analysis which are used as variables in theoretically oriented empirical research. On the level of specific elements or units ("properties of members", "individual properties"; p.431f.) they make the following distinctions :

(m1) absolute properties (concerning the element or unit directly)

(m2) relational properties (concerning the relations of the element to other elements)

(m3) comparative properties (concerning an absolute or a relational property of the element in comparison to the distribution of this property in the surrounding population ("collective", s.below, c1-c2)) and

(m4) contextual properties (concerning a property of the surrounding population or any other environment of the element which is then attributed to the element).

On the level of collectives ("properties of collectives", "collective properties"; p.426f.) they distinguish

(c1) analytical properties (properties of collectives which are generated from absolute properties (m1) of each individual element as a member of the collective, e.g. by calculating arithmetical means or other distributional characteristics)

(c2) structural properties (properties of collectives which are generated from the relations (m2) of each individual member of the collective to some or all other members of the collective)

(c3) global properties (properties concerning collectives without being generated from properties of their members).

An important function of the Lazarsfeld-Menzel-classification (LMC) is that it provides order in a very suggestive way

— for the many different possibilities of how units (elements, members) and their properties - in the way they are treated by data-gathering techniques - relate to those units and their properties to which the theoretical propositions refer; and

— for the procedures which are necessary to describe the intended facts by the available data.

Lazarsfeld and Menzel make it clear that the classification they propose is of no value in itself but that its value should be judged from its usefulness in solving research problems. As a matter of fact, their classification has turned out to be useful in a wide area of research. Therefore, we will look whether this is also the case for network analysis. Though Lazarsfeld and Menzel give some specific sociometric examples (pp.428,431f.) for relational properties of individuals and other individual and collective properties generated from them, their examples could not be exhaustive or representative of the field as we know it today.

Before showing the analytic power of the LMC for network analysis by specific examples we have to point to some vagueness in the central concept of "relational property". Lazarsfeld and Menzel relate this notion ordinarily to individual persons or other individual units ("relational properties of members...", "relational properties of individuals"; p.431f.) instead e.g. to the pair consisting of the two persons for which the relation is ascertained. As a different formulation of theirs shows this convention might be problematic ("...if the elements of a proposition are pairs of individuals", p.440).

Therefore, in the following, we will try to be as clear as possible,

— whether an element or unit is an "individual unit" in an"ordinary sense" and the property is directly related to this individual unit;

— whether - in a different perspective - a pair (or triple or quadruple etc. of individual units in the ordinary sense) might be treated as the unit itself and a relationship among the "constituents" of this pair as its (absolute) property;

— whether - in still another perspective - a pair (or triple etc.) might be considered as a collective "consisting" of individual units in the ordinary sense which are related in a specific way (Lazarsfeld and Menzel in relation to such "pairs" argued: "It would be artificial to call such notions 'propositions about collectives' "; p.440).

A precise delimination of the "collectives", within which relations among their "individual members" are to be studied, is always necessary, if the propositions refer both to "individuals" and to "collectives" or to different levels of "collectives" at the same time. Therefore, we will use the term "relational properties" (m2) always for units in well determined collectives, even if the propositions might refer only to units of one level. We will also use the term "structural" (and not: "relational") properties, if, when "combining" the relational properties, the boundaries of the collectives involved will be crossed. What that implies, will be made clear in the next chapter.

2. TREATING NETWORK-DATA WITH THE LMC

The empirical basis of almost all studies about the structure of networks are data about the relations between units (as members of a collective) considered as pairs, that is relational properties in their simplest form. If e.g. we take persons A,B,C etc. and the business firms X and Y as units and different types of relations among them as relational properties we might have the following data:

a) A informs B;
b) X is owned by Y;
c) A has high esteem for B; or: A despises B;
d) A and B communicate with each other;

e) B supports A financially by giving him 500 Dollars;
f) A works in X; and: B works in X.

These data refer to

— pairs of units of the same (cases a-e) or different type (case f);
— directed (a-c,e,f) or undirected (d) relations;
— relational properties with two (a-d,f) or more (e) values;
— relations which describe the presence or absence of a (positive) characteristic (a,b,d-f) or pertain to one of two contrary values (c).

All these relational properties describe relations between two "individual units". Though in a different context the business firms X and Y might be regarded as collectives (of people working there), here they are treated as (indivisible) units.

In a different perspective "configurations" of two "individual units" could sometimes not only be considered as "collectives of (two) individuals" but even as a "collective of collectives"; i.e. the (unordered) pair (A,B) of persons A and B not only as one collective, but also as a collective of two ordered pairs (A->B) and (B->A), and each of them itself as a collective. Informational flow within a pair of two persons A and B might then be regarded as a structural (!) property of the collective (A->B,B ->A) with one of the following four values: (1) none of A and B informs the other; (2) A informs B but B does not inform A; (3) B informs A but A does not inform B; (4) both, A and B inform each other.

At the time of writing their article, Lazarsfeld and Menzel might have considered this elaboration as "artificial", too. And the foregoing differentiation does not seem necessary because of the simple situation involving only collectives of two ordered pairs. But when we consider directed relations between three and more (instead of only two) persons, the possible configurations will rapidly increase in complexity. Explicitly formulated rules concerning different configurations of relations within clearly determined collectives will then be of great analytical value.

2.1. Structural Properties (c2)

When working with structural properties, the (theoretical) propositions refer to a different level than the level where the (relational) data have been gathered. For instance, when we got data about ordered pairs of persons like "A informs B", and when a proposition about the unordered pair (A,B) is to be stated, all information about ordered pairs of these units is to be "aggregated". As we said already, this specific case is rather simple because we have only to deal with ordered pairs of the kind (A->B) and (B->A). But there might be larger collectives about which structural propositions are to be formulated, e.g. triples (three-person-configurations) or quadruples (four-person-configurations) etc.. For a structural characterization of a triple (A,B,C) we might need to "aggregate" information about 6 ordered pairs ((A->B);(B->A);(A->C);(C->A); (B->C);(C->B)) or information about the three (unordered) pairs ((A,B);(A,C);(B,C)), each of which itself "consists" of two ordered pairs.

When generating structural properties of units at a higher level from properties of the units at a lower level, the procedure of combining the relevant information (the "aggregation procedure") has to be defined carefully. Lazarsfeld and Menzel specify only that structural properties "are obtained by performing some operation on data about the relations..."(p.428). A school class e.g. could be characterized by its density of friendship relations by counting the number of pairwise actual friendship relations and standardizing this number by dividing it through the maximum number of possible friendship relations. This procedure parallels that which would be used outside network analysis when an analytical property of a collective (c1) (e.g. the proportion of Catholics among community residents) is constructed from the "corresponding" absolute individual property (i.e. religious affiliation of community residents).

With other combinations of relational properties the "relation" of the relations, i.e. the structure of their arrangement into complex configurations, has to be taken care of. For instance when we speak about the transitivity of a sociometric structure we need a more complicated "aggregation procedure" than just counting relations or averaging their frequencies etc.. Instead we search for a special aspect of how the relations involving three persons are arranged into complex configurations: Looking at a triple from A's point of view, we ask: Whenever there are three persons A,B,C - where A is related to B and B is related to C -, is A also related to C (is there a "transitively closed two-path")?

Or, looking at the collective "consisting" of all possible triples of persons as "units" we ask: What is the relative frequency of "transitively closed two-paths" among all "two-paths"?

The same holds for studying different degrees of hierarchization or different degrees of connectednes in informal groups by looking at chains of pairwise relations.

2.2. Contextual Properties (m4)

Following Lazarsfeld and Menzel (cp. p.433) contextual properties characterize members of collectives according to properties of their "surroundings" ("environment", "neighbourhood"). These surroundings could be characterized by analytical (c1), structural (c2) and/or global (c3) properties. Contextual descriptions of members on the basis of analytical as well as global properties are identical for all members of the same collective by definition. In so far a contextual description is a simplification. For those situations, where contextual properties are to be attributed to the members not in the same, but in a differentiating manner, Lazarsfeld and Menzel introduced the additional category of "comparative properties" (m3) to which we return below.

If structural properties are generated by "simple aggregation procedures" (i.e. disregarding the structure of the arrangement of the relations), the identical contextual property has to be attributed to all members of the collective indiscriminately. In this respect there is no difference between analytical, global and structural contextual properties. But when we take the arrangement of the relations into consideration then the situation changes: specific units might be placed at different "locations" within the "same" surrounding collective and therefore might be differentiated from some or all other units of the collective. We could say that their "local" environment is different.

For instance, if we are describing a school class by counting for each student the number of friendship choices which are received by him, the frequency distribution of the number of choices received is a characteristic of the structure of the friendship relation ("egalitarian vs. non-egalitarian" structure) within the class as the collective. This property of the class could be used for an identical contextual description of all students. But, from the point of view of individual students, the same context might be perceived very differently. Thus, for the (sociometric) "star", who gets choices from almost every other student, the context might look very different from how it is seen by the "isolate" who gets almost no choice.

Therefore, instead of describing all students by the same value of the contextual variable, every student could be characterized by a different value of the contextual variable according to where he is located in this structure.

Also, "connectedness" among members of a collective could be used as a property describing the context of all members in the same manner or, alternatively, each member could be characterized by his specific individual way of being "connected" to some or all remaining members of his surrounding context.

2.3. Comparative Properties (m3)

"Comparative properties characterize a member by a comparison between his value on some (absolute or relational) property and the distribution of this property over the entire collective of which he is a member." (P.F. Lazarsfeld and H. Menzel, p.432).

While on the level of collectives Lazarsfeld and Menzel treat analytical and structural properties separately, in the context of comparative properties they do not distinguish between collective variables based on absolute or on relational properties. Obviously, this hides intermediate steps of data generation and/or attributing.

Using the terminology of Lazarsfeld and Menzel comparative properties involve either:

— the comparison of absolute properties of individual units with those analytical properties of the collective which are generated from them (e.g. comparing the age of a student with the average age of all students of his class) or

— the comparison of "relational properties" with those structural properties of the collective which are generated from them (e.g. number of friends of a specific student with average number of friends of students in his class, a parameter, i.e. property, of the frequency distribution, which is a structural property itself).

It is obvious, however, that the second case does not parallel the first, because it is not the original relational property (*k* receives a friendship choice from *i*) which is to be compared with a collective property. Instead, the original data about directed links first are aggregated to the more complex relational property "number of friends of a student within his class"; then from this a frequency distribution and its parameters as structural properties of the class are generated, which then are treated as contextual variables of students and attributed to them identically.

It is even more complicated, however, when we consider structural properties which take the arrangement of the relations into account; direct comparability of the original relational property and its structural "derivative" a fortiori is not possible here. To these structural properties (in the sense of the above mentioned examples of connectedness or of hierarchization) there do not exist obvious "corresponding" relational properties on the level of individual units so that a comparison can not be made straightforwardly.

3. TRANSFORMATION OF NETWORK DATA

Following the conceptual ideas of Lazarsfeld and Menzel, we now propose a general strategy to handle network data: We will try to keep specialities of network analysis as small as possible or positively speaking - we will use network analysis exclusively for concept definition and measurement ("scaling") of certain properties based on relational data. Afterwards, network analysis will have done its work, and we shall follow the paths of "ordinary" multi-level data analysis.

3.1.Theory

We will outline this strategy by means of an example which is well-known from literature, i.e. the development of network structure over time, possibly dependent on individual tendencies towards "balance" or "transitivity" (apart from the literature cited in the Introduction cf. H.J. Hummell and W. Sodeur 1990,1992).

We will first discuss which properties of which units at which levels may be involved, and then reformulate the problem as one about "ordinary" relationships between variables. Then we will show with data from "Newcomb's fraternity" (Th. Newcomb 1956, 1961) in a step by step procedure which data have been gathered originally, which intentions are guiding the data transformation, especially the generation of "derived" properties, and which steps of data transformation consequently are to be taken. Eventually, we will have a new ("rectangular") data set where the structural and contextual properties, which were generated from network data, play exactly the same role as properties obtained differently. This data set will then be analysed with standard statistical procedures, here with comparisons between percentages and with some generalized linear models (GLIM).

The units of our analysis are:

— persons *i* as actors who will decide about their;

— directed links *i-->k* to other persons, i.e. whether to create a link to a person *k*, to whom there was no link so far, or to delete a link to a person, to whom there did exist a link in the past.

The creation or deletion of links are the events which are to be explained.

These "decision processes" of creating or deleting links are conceptualized to be influenced by different properties concerning different units on different levels.

In the analysis to be presented here, we will not deal with properties of the actor himself, for instance his involvement in social relationships or his relative isolation in the past.

However, we will deal explicitly

— with some properties of the "target" person *k* to whom the "decision link" *i-->k* is directed, for instance the popularity of this person, measured by the number of links which are directed to him;

— with properties of the behavior or sentiment of the "target" person *k* towards the actor *i*, measured by the existence of a directed link from *k* to *i* (*k-->i*);

Both of these properties may alternatively be interpreted as aspects of the actor *i*'s social environment or as properties of the (ordered or unordered) pair comprising the decision link.

— with properties of the triadic environment of the actor i or of his decision link $i\text{-->}k$, especially of the transitivity status of triplets ijk and ikj, into which the link $i\text{-->}k$ is embedded:

On the ground of balance or transitivity theory different utilities will be related to different outcomes (of the actor's decision about his link $i\text{-->}k$ to the target person k), depending on the transitivity status of the triplets ijk and ikj, into which the link is embedded. Furthermore, these utilities may depend on weights which are related to the significance of both persons (j,k) who belong to the actor's triplet environment. We may assume, for instance, that the transitivity status of surrounding triplets will influence the actor's decisions about a link especially if "significant" persons are involved. Significant persons (from the actor's point of view) will either be persons who are very popular (getting many "choices", having many "incoming links") or persons who are related in a specific way to the actor (i.e. "choosing" him, "directing links to him").

— with properties of an even more complex triadic environment of actor i or his decision link $i\text{-->}k$, defined by the configuration of all possible triplets ijk and ikj in the environment of the decision link $i\text{-->}$k, where the position of the third person j is taken by each of the remaining n-2 persons different from i and k.

Substantially, these conditions may influence the actor's behavior by connecting decisions about the creation of new or the deletion of old links with the "utility" of different outcomes:

— Avoiding intransitivity may reduce conflicts in the social environment. This holds especially if the other persons in the triadic relationship are important for the actor.

— Producing transitivity may save time while enabling a person to take care of more than one social relationship at the same time. Moreover, taking care of multiple relationships is necessary especially with "significant" others.

— Creating a link to a person k may be more valuable if this person is resourceful or influential.

— The same holds, if one can expect that this person will reciprocate, i.e. react positively to a link.

Specifically, we will expect, that - ceteris paribus - an actor i will create a new link with higher (or will delete an old link with lower) probability,

(1) the more the corresponding "target" persons k are liked in the population (the larger their number of choices received);

(2) the more the corresponding "target" persons k positively react to the actor i (the more i receives choices by k);

(3) the more the links to the "target" persons k are compatible with the structure of existing relationships in the actor's social environment (the more often $i\text{-->}k$ is embedded in transitive and/or not embedded in intransitive triplets);

(4) the more important the ("third") persons j are, who are involved in transitive or intransitive triplets which enclose the actor i and the link $i\text{-->}k$ under decision.

There is some ambiguity in the choice of the proper level on which all these variables should be combined and analysed (H.J.Hummell and W.Sodeur 1990).

(1) On the level of specific actors it is asked:

How do actors choose a special decision link (out of n-1 possible links in a population of n people) to decide about?

Assuming an actor's complete information about all possible links and the conditions related to them (i.e. their social environment), a "rational" actor will decide first about the most important link, which is the link with the relative highest utility gain (cf. E. Zeggelink 1993).

(2) On the level of a given specific decision link (however it may have been selected) it is asked:

How will the actor i decide whether to keep or to change the actual status of this relationship, i.e. whether to delete or maintain the link, if there exists one, or to create a link or not, if there exists none?

Assuming now the actor's complete information about all the conditions of the social environment specifically related to this link, a "rational" actor will decide about this link without considering other decision links and their possibly higher outcomes at the same time.

(3) On the level of a given specific decision link in the context of a specific aspect of its social environment, it is asked:

How will the actor decide whether to maintain or to change its actual status?

Assuming - now even more restricted - the actor's information about this specific aspect of his social environment, - for instance the location of the decision link i-->k within a specific triad ijk (taking into account only one specific third person j out of n-2 possible ones) -, a "rational" actor will decide about the "optimal" status of the link in this specific triad without considering its "optimal" status in other triads or even considering other decision links.

Decisional complexity decreases from version (1) to (3), but so does the number of necessary assumptions about the actor's information and his information processing capacities.

In the extreme cases (1) and (3), we must be willing to accept either the assumption of the actor's complete information about the (relevant aspects of the) whole network (cf. E. Zeggelink 1993) or - depending on the actual situation - a very narrowly bounded horizon where his decision about a single link i-->k is handled as if it could be made separately and independently within each of the n-2 triplet contexts (cf. H.J. Hummell and W. Sodeur 1990).

In this paper we prefer version (2), the level of directed links. Its main advantage is that the units at this level are naturally related to the empirically observed "events": i.e. the existence (or non-existence) of links at specific moments in time or the stability or change in their presence or absence (change in their "status") over time.

3.2. Data

We choose the data about Newcomb's fraternity, 2nd group (fall semester 1955/56), as it was distributed in the UCINET (Version 2)-package (see Th. Newcomb 1961; P. Nordlie 1958;

L.C. Freeman without year): 17 male students rank ordered each of the remaining 16 with respect to their preferences for each of 15 weeks (weeks no. 0 to 8 and no. 10 to 15; week no. 9 missing). Thus, for 13 out of 14 pairs of consecutive weeks there is a time gap of one week between two panel waves; for one pair (weeks no. 8 and 10) the gap amounts to two weeks. Later in this chapter we ignore this difference, renumber the weeks from 1 to 15 (t=1,2,3...15), and look at the pairs of consecutive weeks (t,t+1), taking no longer into account whether a pair of weeks belongs to the beginning, the middle or the end of the 15-weeks-period (see below).

Table 3.2.1 shows the rank orders as reported for week no.1. To transform this matrix of rank orders into a binary matrix (which simply reports who is connected to whom or who "chooses" whom), one has to decide how many of the first rankings of each student express a "positive relationship" or which of the last rankings express a neutral or even a negative relationship towards the other person.

For reasons of simplicity we just take the first 3 preference choices of every student (and for every week), so that there is by definition a fixed outdegree of three for every person and every week. In the resulting binary matrix (see Table 3.2.2) there will be "ones" in all cells, where the corresponding cells of the rank order matrix contain the ranks 1, 2, or 3, and "zeros" in all other cells[2].

Stud.	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17
1	00	07	12	11	10	04	13	14	15	16	03	09	01	05	08	06	02
2	08	00	16	01	11	12	02	14	10	13	15	06	07	09	05	03	04
3	13	10	00	07	08	11	09	15	06	05	02	01	16	12	04	14	03
4	13	01	15	00	14	04	03	16	12	07	06	09	08	11	10	05	02
5	14	10	11	07	00	16	12	04	05	06	02	03	13	15	08	09	01
6	07	13	11	03	15	00	10	02	04	16	14	05	01	12	09	08	06
7	15	04	11	03	16	08	00	06	09	10	05	02	14	12	13	07	01
8	09	08	16	07	10	01	14	00	11	03	02	05	04	15	12	13	06
9	06	16	08	14	13	11	04	15	00	07	01	02	09	05	12	10	03
10	02	16	09	14	11	04	03	10	07	00	15	08	12	13	01	06	05
11	12	07	04	08	06	14	09	16	03	13	00	02	10	15	11	05	01
12	15	11	02	06	05	14	07	13	10	04	03	00	16	08	09	12	01
13	01	15	16	07	04	02	12	14	13	08	06	11	00	10	03	09	05
14	14	05	08	06	13	09	02	16	01	03	12	07	15	00	04	11	10
15	16	09	04	08	01	13	11	12	06	02	03	05	10	15	00	14	07
16	08	11	15	03	13	16	14	12	01	09	02	06	10	07	05	00	04
17	09	15	10	02	04	11	05	12	03	07	08	01	06	16	14	13	00
etc. for weeks 2 to 15																	

Table 3.2.1: Newcomb's Fraternity (2nd group, week 1).
Matrix of preference rank orders

2 In spite of being a standard procedure in network analysis, this obviously produces some problems. First it seems unrealistic to believe that all students have the „same capacity“ of creating and maintaining social relationships. Second and more severely: Arbitrarily fixing the outdegrees of all students at the same level (here 3 outgoing links) implies that a „popular“ student has no possibility of choosing all those students who did choose him. Consequently, the amount of asymmetry in the relationships among students, as it is described by the binary matrix, will be artificially increased. The degree of symmetry (or asymmetry) in any network, however, is for purely technical („formal“) reasons related to the degree of transitivity (for details see: K. Echterhagen et al. 1981). In other words: This procedure of fixing outdegrees will as an unintended consequence manipulate transitivity, a key variable of our analysis.

Stud.	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17
1	0	0	0	0	0	0	0	0	0	0	1	0	1	0	0	0	1
2	0	0	0	1	0	0	1	0	0	0	0	0	0	0	0	1	0
3	0	0	0	0	0	0	0	0	0	0	1	1	0	0	0	0	1
4	0	1	0	0	0	0	1	0	0	0	0	0	0	0	0	0	1
5	0	0	0	0	0	0	0	0	0	0	1	1	0	0	0	0	1
6	0	0	0	1	0	0	0	1	0	0	0	0	1	0	0	0	0
7	0	0	0	1	0	0	0	0	0	0	0	1	0	0	0	0	1
8	0	0	0	0	0	1	0	0	0	1	1	0	0	0	0	0	0
9	0	0	0	0	0	0	0	0	0	0	1	1	0	0	0	0	1
10	1	0	0	0	0	0	1	0	0	0	0	0	0	0	1	0	0
11	0	0	0	0	0	0	0	0	1	0	0	1	0	0	0	0	1
12	0	0	1	0	0	0	0	0	0	0	1	0	0	0	0	0	1
13	1	0	0	0	0	1	0	0	0	0	0	0	0	0	1	0	0
14	0	0	0	0	0	0	1	0	1	1	0	0	0	0	0	0	0
15	0	0	0	0	1	0	0	0	0	1	1	0	0	0	0	0	0
16	0	0	0	1	0	0	0	0	1	0	1	0	0	0	0	0	0
17	0	0	0	1	0	0	0	0	1	0	0	1	0	0	0	0	0
etc. for weeks 2 to 15																	

Table 3.2.2: Newcomb's Fraternity (2nd group, week 1)
Binary Matrix: First three preferences

The binary matrices (see table 3.2-2 for week no.1 as an example) about directed links among students for weeks 1 to 15 are the data base which is used exclusively for the following data transformation and analysis.

a) In a first step properties about students, directed and undirected pairs of students, triplets and triplet configurations are derived from this common data base.

With data about Newcomb's fraternity we will get a couple of properties about units at different levels (Figure 3.2.1). To show how these properties are generated in line with Lazarsfeld and Menzel's classification, we give two examples:

(a1) target person *k*'s popularity:
The concept of popularity of students refers to "target" persons *k* as well as to the class as their surrounding population.

The database consists in preference choices (directed links *j->k* for *j#k*). From this, by counting for *k* the number of choices he gets in his class, we obtain *k*'s popularity; this could also be called a relational property, but it is a more complex one than the original preference choices.

(a2) configuration of triplets, which surround *ik*:

The concept "configuration of surrounding triplets" refers to the student actor's "decision links" *ik* as well as to his surrounding population; the later is here (somewhat artificially) defined as consisting of all triplets with all students *j* (*j#i,k*). The database again consists of preference choices.

The configuration of surrounding triplets is obtained by:

— first taking directed relations as (member) units of "triplet collectives" and generating structural properties (c2) of each "triplet collective" based on the arrangement of the directed relations within it;

— then aggregating all triplets, into which a specific link *ik* is embedded, thus obtaining a structural property of a second-order collective (c2) which surrounds *ik* and

— finally attributing the last mentioned property as a contextual property to the directed link *ik* (m4).

Unit	Name of unit	Number of cases	V: variables D: used to describe
i	actor	$n = 17$	
k	"target" persons	$n\text{-}1 = 16$	V: *k*'s popularity D: (1) pair *i,k*: pairs significance D: (2) triplet *ikj, ijk* triplet's significance
j	third persons	$n\text{-}2 = 15$	V: *j*'s popularity D: triplets *(ikj, ijk)* significance
i-->k	"decision link"	$n*(n\text{-}1) = 272$	V; dependent variable (change over time) D: actor's behavior
k-->i	"incoming" link	$n*(n\text{-}1) = 272$	V: pair relation D: (1) pair *i,k*: pair's significance D: (2) triplet *ikj, ijk*: triplet's significance
j-->i	"incoming" link	$n*(n\text{-}1) = 272$	V: pair relation D: triplets' *(ikj, ijk)* significance
i,k	"surrounding" pair	$n*(n\text{-}1)/2 = 136$	V (see above: *k, k-->i)* D: social context for decision link
ijk) *ikj)*	"surrounding" triplets	$n*(n\text{-}1)*(n\text{-}2) = 4080$	V: transitivity status, (see also, *j, k,* *k-->i, j-->i*) D: social context for decision link
i(.)k	configuration of *ik* 'surrounding" triplets	$n*(n\text{-}1) = 272$	V: distribution of $(n\text{-}2)$ transitivity statuses D: social context for decision link

Figure 3.2.1: Units, properties and levels of analysis

b) In a second step these generated variables have to be combined into a single rectangular matrix. In this case, we have to combine

— properties of single actors i and single target persons k (n=17) with

— properties of directed links ik, ki between them (n*(n-1)=272) with

— properties of triplets which surround the decision links (n*(n-1)*(n-2)=4080).

The easiest way of combining units of different levels is to do it on the "lowest level" involved, that is the level with the largest number of units. The resulting rectangular data matrix would then consist of 4080 rows where of course the data about "higher levels" have been multiplied, i.e. data of each of the 17 target persons k now appear (n-1)*(n-2)=240 times. Substantially, however, this data matrix would stress the correct level.

c) As we explained in chapter (3.1) with respect to three different ways of choosing an adequate level of analysis, we will take the second; i.e. the dependent variable shall be the decision link i-->k with n*(n-1) related units. Now, in a third step the corresponding data set has to be generated. That means, for the rectangular data matrix we obtained so far, each of n-2 rows belonging to the same decision link i-->k have to be aggregated. Information about the n-2=15 triplets as the context, into which each decision link is embedded, is preserved in form of frequency distributions which summarize the states of all these triplets.

Elsewhere we proposed a typology of triplet contexts (cf. H.J. Hummell and W. Sodeur 1990). The main idea is to classify all those triplets into the same type which would change their status (with respect to transitivity) in the same way, if a new decision link would be created. There are 5 possibilities, how triplets could change their transitivity status.

triplet type	time t (ijk	ikj)	--->	time t+1 (ijk	ikj)
TRAB = 1	n	n		n	i
TRAB = 2	n	n		n	n
TRAB = 3	n	n		n	t
TRAB = 4	i	n		t	n
TRAB = 5	i	n		t	t

Figure 3.2.2: transitivity status before (t) and after (t+1) creating a new link ik (i: intransitive; t: transitive; n:neutral) for the two triplets (ijk) and (ikj)

Here we use only triplet contexts of type TRAB=1 as well as of type TRAB=4 and TRAB=5 which will later be combined; i.e. for each decision link, frequencies of these types will describe the actor's utility (in the light of transitivity) which is related to the action of creating a new link ik (see below). A similar definition of 5 types holds true for the case of deleting an existing decision link ik (cf. H.J. Hummell and W. Sodeur 1990).

Finally some remarks about the handling of dynamic aspects of change : Our aim is to explain the stability or change in the presence or absence of decision links over time, i.e. change of their state between specific points in time.

To simplify the analysis, we

— look for changes of directed links among students for pairs of consecutive weeks only;

— regard relations ik, for which preference links do not exist at time t, separately from those relations ik, where preference links do exist at time t.

In the set of events which will be treated here, a link will either remain absent (--) or it will be created (-+) between time t and t+1.

(The corresponding set of events, where a link will either remain present (++) or will be deleted (+-) between time t and t+1, will not be treated).

— explain the development of decision links between time t and t+1 by properties of persons, properties of directed links and of triplets at time t only, i.e. we will not consider earlier points of time (t-1, 2..);

— assume a "stationary process", where decisions about directed links follow the same rules regardless of time, i.e. by and large are identical within all 14 pairs of consecutive weeks.

This assumption is not completely correct, because empirically, it turns out that there is an overall tendency that effects are fading out over the period of 15 weeks. But beyond this general tendency there is some zig-zag in the change of relationships between single pairs of weeks. May be, because of too few empirical observations, we could not find any regular development.

— As a consequence of the stationarity assumption we handle changes of the state of decision links ("present" vs. "absent") between different pairs of consecutive weeks as events which are independent from each other. Thus, we now have to explain $n*(n-1)*\mathrm{NPW}=3808$ events (i.e. $n*(n-1)=272$ decision links times the number NPW=14 of pairs of weeks).

Tables 3.2.3 and 3.2.4 show both data sets in a condensed form, where all events with the same pattern of the values of the independent variables are aggregated and where for each pattern the different types of outcomes of the decision processes (i.e. possible changes of the state of decision links over time) are counted[3].

To sum up: As a minimal specific tool set for network analysis we used procedures to generate network data on the levels of persons, of directed links between persons and on the level of triplets containing the directed links. Successive steps of the data transformation process do not need special solutions. General data base systems attach data from different levels to each other in the same way, as it is done with non-network data. They may also help to condense the data set if necessary. Once having generated an "ordinary" rectangular data matrix, general statistical procedures will be used for analysing the interrelationships between network variables.

---

[3] In this case, due to the relative small number of units and properties, the data set (as it is described) can be use with limited resources even without any condensation, for instance with GLIM and less than 600 KB memory available for the program. However, if there are networks with a larger number of knots or if units on the level of triads or even triplets are chosen, very large data sets may be generated. Under such circumstances one has to condense the data set by combining all units with the same pattern of values of the independent variables and to use the frequency of the cases as weight, as is done here.
There are programs which
(a) generate different network properties on monadic, dyadic and triadic levels, given data about directed links (NETZDIAL),
(b) produce aggregate data files with up to 10 integer variables to (weighted) patterns (AGGREGAT). These programs are available from Wolfgang.Sodeur@Uni-Koeln.de (email).

vv <-- LINK: Link INcoming from K (i<--k)
! vv <-- NINK: Number of links INcoming at K
! ! 1 4 5 <-- Fx, Frequencies of triplet environments of type x
! ! ! ! ! vvv <----- } number of units with same pattern of LINK, NINK, F1, F4, F5
! ! ! ! ! +/+ +/- <- } where link ik remains present (+/+) or disappears (+/-).

LINK	NINK	F1	F4	F5	+/+	+/-
1	1	1	0	0	2	3
1	1	2	0	0	7	10
1	1	3	0	0	4	2
1	2	1	0	0	0	2
1	2	1	0	1	2	1
1	2	2	0	0	6	4
1	2	2	0	1	6	0
1	2	2	1	0	3	0
1	2	3	0	0	12	7
1	3	1	0	0	0	1
1	3	1	0	1	0	3
1	3	2	0	0	2	3
1	3	2	0	1	10	4
1	3	3	0	0	8	5
1	3	3	1	0	0	2
1	4	1	0	0	1	0
1	4	1	0	1	1	0
1	4	1	0	2	2	0
1	4	2	0	0	3	2
1	4	2	0	1	13	5
1	4	2	1	1	1	0
1	4	3	0	0	14	3
1	4	3	1	0	2	1
1	5	1	0	1	1	0
1	5	1	0	2	1	0
1	5	2	0	0	8	1
1	5	2	0	1	9	8
1	5	2	1	0	0	1
1	5	2	1	1	3	2
1	5	3	0	0	15	7
1	5	3	1	0	3	1
1	6	1	0	2	0	2
1	6	2	0	0	3	0
1	6	2	0	1	5	3
1	6	2	1	0	2	0
1	6	2	1	1	1	2

continued:

LINK	NINK	F1	F4	F5	+/+	+/-
1	6	3	0	0	6	2
1	6	3	1	0	2	0
1	7	2	0	0	2	2
1	7	2	0	1	5	0
1	7	2	1	0	1	1
1	7	2	1	1	2	0
1	7	3	0	0	4	4
1	7	3	1	0	3	2
1	7	3	2	0	2	0
1	8	1	0	1	0	2
1	8	1	0	2	5	0
1	8	2	0	0	2	3
1	8	2	0	1	17	5
1	8	2	1	0	2	0
1	8	2	1	1	10	3
1	8	3	0	0	13	1
1	8	3	1	0	13	4
1	8	3	2	0	1	0
1	9	2	0	1	10	2
1	9	2	1	0	0	1
1	9	2	1	1	8	2
1	9	3	0	0	9	2
1	9	3	1	0	8	1
1	9	3	2	0	1	1
1	10	2	0	1	1	1
1	10	2	1	1	1	1
1	10	3	0	0	2	0
1	10	3	1	0	1	0
2	1	0	0	0	1	0
2	1	1	0	0	4	2
2	1	2	0	0	6	2

continued:

LINK	NINK	F1	F4	F5	+/+	+/-
2	2	0	0	0	1	0
2	2	0	0	1	1	0
2	2	1	0	0	15	2
2	2	1	0	1	5	1
2	2	1	1	0	1	0
2	2	2	0	0	26	7
2	2	2	1	0	1	1
2	3	0	0	1	4	0
2	3	1	0	0	7	1
2	3	1	0	1	9	2
2	3	1	1	0	1	0
2	3	1	1	1	1	0
2	3	2	0	0	9	6
2	3	2	1	0	5	0
2	3	2	2	0	1	0
2	4	0	0	0	2	0
2	4	0	0	1	1	0
2	4	0	0	2	1	1
2	4	1	0	0	5	1
2	4	1	0	1	9	2
2	4	1	1	0	1	2
2	4	1	1	1	3	0
2	4	2	0	0	11	3
2	4	2	1	0	3	0
2	4	2	2	0	3	0
2	5	0	0	1	2	0
2	5	1	0	0	7	0
2	5	1	0	1	10	4
2	5	1	1	0	2	0
2	5	1	1	1	4	0
2	5	2	0	0	14	3

continued:

LINK	NINK	F1	F4	F5	+/+	+/-
2	5	2	1	0	3	0
2	5	2	2	0	1	0
2	6	0	0	1	1	0
2	6	0	0	2	1	0
2	6	1	0	0	1	0
2	6	1	0	1	0	1
2	6	1	1	0	1	0
2	6	1	1	1	1	1
2	6	2	0	0	2	1
2	6	2	1	0	3	1
2	7	0	0	1	0	1
2	7	1	0	0	1	0
2	7	1	0	1	4	1
2	7	2	0	0	4	0
2	7	2	1	0	2	0
2	7	2	2	0	1	0
2	8	0	0	1	1	0
2	8	1	0	0	3	0
2	8	1	0	1	5	1
2	8	1	1	0	1	1
2	8	1	1	1	5	0
2	8	2	0	0	12	1
2	8	2	1	0	7	1
2	8	2	2	0	0	1
2	9	1	0	1	2	2
2	9	1	1	1	3	1
2	9	2	0	0	6	1
2	9	2	1	0	3	0
2	10	1	0	1	1	0
2	10	1	1	1	1	0
2	10	2	1	0	1	0

sum(all: +/+ plus +/-): 714

sum: 534 180

Table 3.2.3.: Complete data in condensed form about the development of 714 decision links existing at time *t* (definition of variables see above)

vv <-- LINK: Link INcoming from K (i<--k)
! vv <-- NINK: Number of links INcoming at K
! ! 1 4 5 <-- Fx, Frequencies of triplet environments of type x
! ! ! ! ! vv <----- } number of units with same pattern of LINK, NINK, F1, F4, F5
! ! ! ! ! -/- -/+<- } where link ik remains absent (-/-) or is created (-/+).

1	0	0	0	0	5	0
1	0	1	0	0	79	5
1	0	2	0	0	212	5
1	0	3	0	0	120	3
1	1	0	0	0	9	0
1	1	0	0	1	2	0
1	1	1	0	0	82	2
1	1	1	0	1	16	3
1	1	1	1	0	6	0
1	1	2	0	0	211	6
1	1	2	0	1	16	2
1	1	2	1	0	9	0
1	1	3	0	0	150	2
1	1	3	1	0	13	2
1	2	0	0	0	1	0
1	2	0	0	2	1	0
1	2	1	0	0	25	1
1	2	1	0	1	41	4
1	2	1	0	2	11	2
1	2	1	1	0	3	0
1	2	1	1	1	1	0
1	2	2	0	0	146	6
1	2	2	0	1	104	14
1	2	2	1	0	26	2
1	2	2	1	1	9	0
1	2	2	2	0	1	0
1	2	3	0	0	184	5
1	2	3	1	0	45	0
1	2	3	2	0	0	1
1	3	0	0	1	1	0
1	3	1	0	0	6	1
1	3	1	0	1	31	2
1	3	1	0	2	11	0
1	3	2	0	0	46	2
1	3	2	0	1	54	2
1	3	2	1	0	10	0
1	3	2	1	1	12	0
1	3	2	2	0	1	0
1	3	2	2	1	1	0
1	3	3	0	0	112	5
1	3	3	1	0	28	0
1	3	3	2	0	1	0
1	4	0	0	2	2	0
1	4	1	0	0	5	0

continued:

1	4	1	0	1	6	1
1	4	1	0	2	6	1
1	4	1	1	1	3	0
1	4	1	1	2	1	0
1	4	2	0	0	30	5
1	4	2	0	1	42	6
1	4	2	1	0	11	1
1	4	2	1	1	12	1
1	4	2	2	0	2	0
1	4	2	2	1	1	0
1	4	3	0	0	92	4
1	4	3	1	0	26	1
1	4	3	2	0	5	0
1	5	1	0	1	14	3
1	5	1	0	2	5	0
1	5	1	1	1	3	0
1	5	1	1	2	2	1
1	5	2	0	0	23	1
1	5	2	0	1	47	7
1	5	2	1	0	17	2
1	5	2	1	1	9	1
1	5	3	0	0	61	4
1	5	3	1	0	23	1
1	5	3	2	0	2	0
1	6	0	0	2	1	0
1	6	1	0	1	7	0
1	6	1	0	2	1	0
1	6	1	1	1	1	1
1	6	2	0	0	4	0
1	6	2	0	1	7	1
1	6	2	1	0	5	0
1	6	2	1	1	12	2
1	6	2	2	1	3	1
1	6	3	0	0	5	0
1	6	3	1	0	8	1
1	6	3	2	0	2	1
1	7	1	0	1	2	0
1	7	1	0	2	1	1
1	7	1	1	2	2	0
1	7	2	0	0	6	0
1	7	2	0	1	9	1

continued:

1	7	2	1	1	6	0	2
1	7	2	2	1	1	0	2
1	7	3	0	0	8	0	2
1	7	3	1	0	7	1	2
1	7	3	2	0	4	1	2
1	8	1	0	1	4	0	2
1	8	1	0	2	7	0	2
1	8	1	1	1	2	1	2
1	8	1	1	2	4	1	2
1	8	2	0	0	5	1	2
1	8	2	0	1	20	0	2
1	8	2	1	1	18	4	2
1	8	2	2	0	2	0	2
1	8	2	2	1	10	1	2
1	8	3	0	0	9	2	2
1	8	3	1	0	14	1	2
1	8	3	2	0	8	0	2
1	9	1	0	2	2	0	2
1	9	1	1	1	2	0	2
1	9	1	1	2	4	0	2
1	9	2	0	0	1	0	2
1	9	2	0	1	8	0	2
1	9	2	1	0	2	0	2
1	9	2	1	1	12	4	2
1	9	2	2	1	3	0	2
1	9	3	1	0	6	0	2
1	9	3	2	0	2	0	2
1	10	1	0	2	2	0	2
1	10	2	0	1	1	0	2
1	10	2	1	1	1	0	2
1	10	3	1	0	2	0	2
2	0	0	0	0	5	0	2
2	0	1	0	0	51	0	2
2	0	2	0	0	41	2	2
2	1	0	0	0	5	1	2
2	1	0	0	1	0	2	2
2	1	1	0	0	56	1	2
2	1	1	0	1	8	1	2
2	1	1	1	0	2	0	2
2	1	2	0	0	35	0	2
2	1	2	1	0	3	0	2

continued:

2	0	0	0	3	2
2	0	0	1	1	2
2	0	0	2	1	0
2	1	0	0	2	0
2	1	0	1	19	0
2	1	1	0	5	0
2	1	1	1	2	0
2	2	0	0	24	4
2	2	1	0	9	1
3	0	0	1	1	1
3	0	1	1	1	0
3	1	0	0	4	0
3	1	0	1	7	0
3	1	1	0	2	0
3	1	1	1	0	1
3	2	0	0	14	1
3	2	1	0	4	2
4	0	0	1	1	0
4	0	0	2	0	1
4	1	0	0	3	1
4	1	0	1	4	2
4	1	1	0	3	0
4	1	1	1	1	1
4	2	0	0	5	0
4	2	1	0	1	1
5	1	0	0	1	0
5	1	0	1	3	2
5	1	1	1	1	1
5	2	0	0	4	0
5	2	1	0	3	1
6	1	0	1	1	1
6	2	0	0	3	1
6	2	1	0	1	0
7	0	1	2	1	0
7	2	0	0	2	0
7	2	1	0	1	0
8	0	0	2	1	0
8	1	1	1	0	2
8	2	0	0	3	0
9	1	1	1	0	1
9	2	0	0	1	1

sum(all: -/- plus -/+): 3094 — sum: 2914 180

Table 3.2.4.: Complete data in condensed form about the development of 3094 decision links not existing at time *t* (definition of variables see above)

3.3. Data Analysis

Given the data set which was described above, we will explain the proportion of changing links between two consecutive weeks (out of all links "under the same conditions") according to the hypothesis which was proposed above by the following effects (names of variables in parentheses):

Ceteris paribus - an actor *i* will create new links *ik* to "target persons" *k* the more frequently,

(1) the more person *k* is liked in the population (number of popularity choices received by *k*) (NINK);
(2) the more person *k* positively reacts to the actor *i* (*i* receives a popularity choice by *k*) (LINK);
(3) the more a link to person *k* corresponds with the structure of existing relationships in the actor's social environment, considered under the aspect of its transitivity (TRAB);

If decision links *ik* are embedded in triplets of type TRAB=1 (s.above) at time *t*, this will reduce the probability of creating a new link between *t* and *t*+1, because under such conditions a triplet *ikj* with a formerly neutral transitivity status would become intransitive.

On the other side, environments of triplets of type TRAB=4 or of type TRAB=5 at time *t* will increase the probability of creating a new link between *t* and *t*+1, because under such conditions a formerly intransitive triplet *ijk* would become transitive.

In environments of triplets of type TRAB=5, compared to those of type TRAB=4, we would expect even more creations of new links because in these cases new links would not only change intransitive triplets *ijk* to transitive ones but, in addition, would also change triplets *ikj* (i.e. *ijk* from an "other point of view", cf. Figure 3.2.2; see also H.J. Hummell and W. Sodeur 1990) from neutral to transitive.

Figure 3.3.1 summarizes the properties which will be used here to explain stability or change of decision links between consecutive weeks:

Name of Variable	Description	Number of possible categories	Number of observed categories	Values of observed categories
LINK :	Link INcoming at *i* from *k* (*i* <--- *k*:	2	2	0,1
NINK :	Number of link INcoming at *k*:	16	11	0 to 10
F1 :	Frequencies of triplets with TRAB = 1 :	15	4	0 to 3
F4	Frequencies of triplets with TRAB = 4 :	15	3	0 to 2
F5	Frequencies of triplets with TRAB = 5 :	15	3	0 to 2

Restrictions: sums of frequencies (TRAB = *i*; with *i* = 1...5) = 15

Figure 3.3.1. Properties with possible influence on the development of decision links

We skip the discussion of how to recode the independent variables appropriately (see Appendix A). As a result of these modifications the following set of independent variables will be used to explain the development of decision links for pairs of weeks:

LINK: with two categories: existence of a Link INcoming from K: (*i*<--*k*) or not;

NINK3: Number of links INcoming at K with three categories: 0-1, 2-3, 4-10 received preference choices;

F1: number of triplets of type TRAB=1 in the environment of the decision link *ik* with three categories: 0-1, 2, 3 triplets of type TRAB=1;

F45: number of triplets of type TRAB=4 or of type TRAB=5 in the environment of the decision link *ik* with two categories: 0, 1-3 triplets of type TRAB=4 or of type TRAB=5.

We look at the process of "creating new links", i.e. at those 17*(16-3)=221 decision links which did not exist in the respective first of two consecutive weeks. Because transitions of these 221 links have been observed over 14 pairs of consecutive weeks in total, there are 221*14=3094 (number of links times number of pairs of weeks) events under study.

(As we mentioned already, an analogical analysis could be done with the 17*3*14=714 events, where existing links might possibly be "deleted". For reasons of simplicity this will be left out here).

Working with the recoded variables LINK, NINK3, F1 and F45, it turned out that only 30 of the possible 36 (=2*3*3*2) combinations of their categories occured empirically (see Table 3.3-1). That is, as a matter of fact the 3094 events happened under only 30 different patterns of the values of the (recoded) independent variables.

Variables LINK, NINK3, F1 and F45 are recoded (see above)

```
vv <-- LINK : Link INcoming from K (i<--k)
!vv <-- NINK3 : Number of links INcoming at K, 0 to 1, 2 to 3, 4 to 10 choices received
! ! 1  4+5 <-- Fx, Frequencies of triplet environments of type x: F1 : 0 to 1, 2, 3 triplets;
! ! ! !                                                            F45: 0, 1 to 3 triplets
! ! ! ! vvv <----- } number of units with same pattern of LINK, NINK3, F1, F45
! ! ! !  -/-   -/+ <- }         where link ik remains absent (-/-) or is created (-/+).
```

1 1 1 0	175	7	continued :			continued :			continued :		
1 1 1 1	24	3									
1 1 2 0	423	11	1 2 2 1	218	18	1 3 3 0	175	10	2 2 1 1	39	4
1 1 2 1	25	2	1 2 3 0	296	10	1 3 3 1	109	7	2 2 2 0	38	5
1 1 3 0	270	5	1 2 3 1	74	1	2 1 1 0	117	2	2 2 2 1	13	3
1 1 3 1	13	2	1 3 1 0	5	0	2 1 1 1	10	3	2 3 1 0	4	1
1 2 1 0	32	2	1 3 1 1	84	10	2 1 2 0	76	2	2 3 1 1	16	11
1 2 1 1	100	8	1 3 2 0	69	7	2 1 2 1	3	0	2 3 2 0	18	2
1 2 2 0	192	8	1 3 2 1	261	32	2 2 1 0	29	2	2 3 2 1	6	6
sum (all: -/- plus -/+: 3094)									sum:	2914	180

Table 3.3.1. Complete data in condensed form about the development of 3094 decision links not existing at time *t*

Depending on different contextual conditions actors *i* may decide either to maintain the actual situation (no directed link to *k*) or to create the respective decision link. Therefore, the dependent variable is the relative frequency of links created between time *t* and *t*+1 out of all 3094 decision links which did not exist at time *t*. The variability of these frequencies is to be explained by the patterns of contextual conditions.

Looking for an adequate model we used a generalized linear system (GLIM)[4] assuming a binomial error function. Table 3.3.2 shows the results of a model where all 4 variables are introduced simultaneously. Some variations of the basic model with different interaction terms will give some ideas about how to modify the model.

Model	deviance	d.f. / change
1	103.64	29
------------------------- basic model -------------------------		
1 + LINK + NINK3 + F1 + F45	27.255	23
------------------------- compared with basic model : change in deviance and d.f. -------------------------		
1 + LINK + NINK3 + F1 + F45 + LINK.F1	-0.018	-1
1 + LINK + NINK3 + F1 + F45 + NINK3.F1	-2.270	-4
1 + LINK + NINK3 + F1 + F45 + LINK.F45	-5.186	-1
1 + LINK + NINK3 + F1 + F45 + NINK3.F45	-7.261	-2
1 + LINK + NINK3 + F1 + F45 + LINK.NINK3	-6.611	-2
1 + LINK + NINK3 + F1 + F45 + F1.F45	-2.659	-2

Table 3.3.2. Basic Model and some variations

Balancing economy and explanatory power, one has to decide about how many interaction effects should be taken into account. Here we only choose the most influential ones, i.e. NINK3.F45 and LINK.F45. The resulting model will be described with more details in Table 3.3.3:

scaled deviance = 1 5.731
d.f = 20
terms = 1 + LINK + NINK3 + F1 + F45 + LINK.F45 + NINK3.F45

Line	Estimate	s.e.	Parameter	Comments
1	-3.551	0.2550	1	
2	0.1883	0.3205	LINK(2)	link incoming (*i* <-- *k*)
3	0.7433	0.2818	NINK3(2)	2 to 3 links incoming at *k*
4	1.311	0.3146	NINK3(3)	4 to 10 links incoming at *k*
5	-0.06568	0.1998	F1(2)	link embedded in 2 triplets of type TRAB=1
6	-0.6306	0.2657	F1(3)	link embedded in 3 triplets of type TRAB=1
7	1.372	0.4125	F45(2)	link embedded in 1 to 3 triplets of type TRAB=4 or 5
8	0.8430	0.4125	LINK(2).F45(2)	interaction lines 2 and 7
9	-1.217	0.4789	NINK3(2).F45(2)	interaction lines 3 and 7
10	-1.122	0.4863	NINK3(3).F45(2)	interaction lines 4 and 7

Table 3.3.3. "Explanation" of relative frequencies of newly created links *i*-->*k* under different conditions

[4] Generalized Linear Interactive System, rel.3.77; Manual. C.D.Payne (ed.) 1987; Royal Statistical Society.

All effects relate to the first category of the respective variable, i.e.:

LINK(1) :	no incoming link;
NINK(1) :	0 to 1 links incoming at *k*;
F1(1) :	0 to 1 triplets of type TRAB=1 (transitivity status of triplet *ikj* will change from neutral to intransitive and *ijk* from neutral to neutral);
F45(1) :	0 triplets of type TRAB=4 or 5 (transitivity status of triplet *ikj* will change from neutral to neutral (4) or transitive (5) and *ijk* from intransitive to transitive).

With one "exception" (see below) all main effects are in line with theoretical expectations.

Decision links *ik* have been created between two consecutive weeks more often, if:

— the "target person" *k* was popular among students (lines 3,4),
— the creation of the new link changed at least one of the surrounding *n*-2 triplets from intransitive to transitive (line 7).

Decision links *ik* have been created less often, if they

— transformed many of the surrounding triplets from neutral to intransitive (lines 6).

The exception is the main effect of LINK (preference choice from *k* to *i*) which is superseded by a strong interaction effect. The main effect of LINK disappears when the interaction term LINK.F45 is introduced into the model. In its place there is a strong interaction effect which is in line with our assumptions.

If the target person *k* gave a preference choice to actor *i* (LINK) at time *t*, the effect of triplet contexts of type TRAB=4 or type TRAB=5 (new links reduce intransitivity and/or increase transitivity) becomes larger, i.e. triplet environments with target persons who reciprocate are more influential (see line 8).

The size of the other interaction term (NINK3.F45) tells us, however, that the influence of the two most important single factors will cumulate only to a moderate extent. Contrary to the first result this does not support the expectation that triplet environments with popular persons in the position of the target person *k* will have an extraordinary influence.

Finally we translate the model into simple comparisons between percentages of changing links (Table 3.3.4). The table tells the same story as the GLIM model before. However, quantitative comparisons between effects are more obvious. Comparisons of relative frequencies of new links under different contextual conditions show effects of single factors and of the interaction terms as well as the damped cumulation of effects in situations where the target person is very popular and the creation of a link to him would change the structure of social relationships in the direction of more transitivity.

Legend of the Table 3.3.4		: SW (SWitch : creating a link) was coded with
	0	: no change,
	100	: change.

The "mean" in this case mirrors the percentage of events where a link was newly created.

(a) main effect

Summaries of	SW				
By levels of	LIN	Link incoming from *k*			
Variable	Value	Label	Mean	Std Dev	Cases
For Entire Population			5.8177	23.4116	3094
LINK	1.00	no link incoming	5.3199	22.4473	2688
LINK	2.00	link incoming	9.1133	28.8153	406
Summaries of	SW				
By levels of	NINK	Number of links incoming at *k*			
Variable	Value	Label	Mean	Std Dev	Cases
For Entire Population			5.8177	23.4146	3094
NINK3	1.00	0-1 received	3.1543	17.4855	1173
NINK3	2.00	2-3 received	5.5861	22.9758	1092
NINK3	3.00	4-10 received	9.8914	29.8727	829
Summaries of	SW				
By levels of	F1	Triplet environment, TRAB=1			
Variable	Value	Label	Mean	Std Dev	Cases
For Entire Population			5.8177	23.4116	3094
F1	1.00	0-1 triplet	7.7035	26.6841	688
F1	2.00	2 triplets	6.4156	24.5116	1434
F1	3.00	3 triplets	3.6008	18.6407	972
Summaries of	SW				
By levels of	F45	Triplet environment, TRAB=4 or 5			
Variable	Value	Label	Mean	Std Dev	Cases
For Entire Population			5.8177	23.4116	3094
F45	.00	0 triplet	3.7130	18.9128	1993
F45	1.00	1 + triplets	9.6276	29.5104	1101

(b) interaction effects

Summaries of	SW				
By levels of	LINK	link incoming from *k*			
	F45	Triplet environment, TRAB=4 or 5			
Variable	Value	Label	Mean	Std Dev	Cases
For Entire Population			5.8177	23.4116	3094
LINK	1	no link incoming	5.3199	22.4473	2688
F45	0	0 triplet	3.5357	18.4734	1697
F45	1	1 + triplets	8.3754	27.7158	991
LINK	2	link incoming	9.1133	28.8153	406
F45B	0	0 triplet	4.7297	212634	296
F45B	1	1 + triplets	20.9091	40.8521	110
Summaries of	SW				
By levels of	NINK3	Number of links incoming at *k*			
	F45	Triplet environment, TRAB=4 or 5			
Variable	Value	Label	Mean	Std Dev	Cases
For Entire Population			5.8177	23.4116	3094
NINK3	1.00	0-1 received	3.1543	17.4855	1173
F45	.00	0 triplet	2.4816	15.5636	1088
F45	1.00	1 + triplet	11.7647	32.4102	85
NINK3	2.00	2-3 received	5.5861	22.9758	1092
F45	.00	0 triplet	4.3974	20.5204	614
F45	1.00	1 + triplets	7.1130	25.7311	478
NINK3	3.00	4-10 received	9.8914	29.8727	829
F45	.00	0 triplet	6.8729	25.3428	291
F45	1.00	1 + triplets	11.5242	31.9611	538

Table 3.3.4. Illustration of effects with ordinary percent comparisons

3.4. Conclusion

In this article we followed a strategy which can be traced back to Lazarsfeld and Menzel (1961). Using as an example the development of a sociometric network over time we showed, where the analysis needs special network tools and solutions and where, however, it may follow standard procedures developed outside network analysis.

Lazarsfeld and Menzel's well known classification of units and their properties was the conceptional instrument to translate a research question, which first looks like a speciality of network analysis, into a standard problem of treating relationships between variables of different units at different levels.

Network analysis was used to define units at different levels in such a way that they closely mirror the theoretical concepts under study, and to derive statistics which describe the respective properties. However, and in line with the main idea of this paper, specialities of network analysis have been restricticted to these two topics. The subsequent analysis - meant only as an illustration of the general strategy - exclusively used standard procedures of statistics.

BIBLIOGRAPHY

CARTWRIGHT, D., and F. HARARY, 1956, "Structural Balance : A Generalization of Heider's Theory", *in The Psychological Review* 63, 277-293.
DAVIS, J. A., 1967, "Clustering and Structural Balance in Graphs", *in Human Relations* 20, 181-187.
DAVIS, J. A., and S. LEINHARDT, 1972, "The Structure of Positive Interpersonal Relations in Small Groups", *in* J. Berger et.al. (eds.), *Sociological Theories in Progress*, vol.2, Boston, Houghton & Mifflin, 218-251.
ECHTERHAGEN, K., H.J. HUMMELL, L. KREMPEL and W. SODEUR, 1981, "Base Line Models for Evaluating Interpersonal Tendencies for Balance", *in Theories of the Davis-Holland-Leinhardt Type*, Duisburg/Wuppertal, mimeo.
FREEMAN, L.C., "UCINET", Version 2, School of Social Sciences, University of California, Irvine, CA.
HOLLAND, P. W., and S. LEINHARDT, 1971, "Transitivity in Structural Models of Small Groups", *in Comparative Group Studies* 2, 107-124.
HOLLAND, P.W., and S. LEINHARDT, 1975, "Local Structure in Social Networks", *in*: D.Heise (ed.), 1975, *Sociological Methodology* 1976, San Francisco, Jossey Bass, 1-45.
HUMMELL, H.J., 1972, *Probleme der Mehrebenenanalyse*, Stuttgart, Teubner.
HUMMELL, H.J. and W. SODEUR, 1987, "Triaden- und Triplettzensus als Mittel der Strukturbeschreibung", *in* F.U. Pappi (ed.), 1987a, *Methoden der Netzwerkanalyse*. Muenchen, Oldenbourg, 129-161.
HUMMELL, H.J., and W. SODEUR, 1990, Evaluating Models of Change, *in Triadic Sociometric Structures*, *in* J. Weesie and H.Flap (eds.), *Social Networks Through Time*. Utrecht (ISOR), 281305.
HUMMELL, H.J., and W. SODEUR, 1992, Multivariate Analyse von Struktureigenschaften auf mehreren Ebenen, Netzwerkanalyse als "messtheoretisches" Konzept, *in* H.J. Andress et al. (eds.), *Theorie - Daten - Methoden. Neue Modelle und Verfahrensweisen in den Sozialwissenschaften*, Muenchen, Oldenbourg, 269-294.
LAZARSFELD, P. F., and H. MENZEL, 1961, "On the Relation between Individual and Collective Properties", *in* A. Etzioni (ed.), *Complex Organizations*, New York, Holt, Rinehart & Winston, 422-440.
NEWCOMB, Th., 1953, "An Approach to the Study of Communicative Acts", *in: Psychological Review* 60, 393-404.
NEWCOMB, Th., 1956, "The Prediction of Interpersonal Attraction", *in American Psychologist* 11, 575-586.

NEWCOMB, Th., 1961, *The Acquaintance Process*, New York, Holt, Rinehart & Winston.
NORDLIE, P.G., 1958, *A Longitudinal Study of Interpersonal Attraction in a Natural Group Setting*, Dissertation, University of Michigan.
PAPPI, F. U.(ed.), 1987a, *Methoden der Netzwerkanalyse*, Muenchen, Oldenbourg.
PAPPI, F. U., 1987b, "Die Netzwerkanalyse aus soziologischer Perspektive", *in* F. U. Pappi (ed.), 1987a, 11-37.
ZEGGELINK, E., 1993, *Strangers into Friends. The Evolution of Friendship Networks Using an Individual Oriented Modeling Approach*, Amsterdam, Thesis Publishers.

APPENDIX A

Social environment of (non existent) decision links at time *t*.
Frequency distribution of the original and recoded variables.

LINK	Link Incoming from K (i <--k)	
Value	Frequency	Percent
1.00	2688	86.9
2.00	406	13.1
TOTAL	3094	100.0

NINK			NINK3	Number of links incoming at k		
Value	Frequency	Percent	Value Label	Value	Frequency	Percent
.00	528	17.1	0-1 received	1.00	1173	37.9
1.00	645	20.8	2-3 received	2.00	1092	35.3
2.00	728	23.5	4-10 received	3.00	829	26.8
3.00	364	11.8				
4.00	288	9.3				
5.00	242	7.8				
6.00	70	2.3				
7.00	54	1.7				
8.00	120	3.9				
9.00	49	1.6				
10.00	6	.2				
TOTAL	3094	100.0				

F1 Triplet environment, TRAB = 1			F1 (recoded)	Triplet environment, TRAB = 1		
Value	Frequency	Percent	Value Label	Value	Frequency	Percent
.00	51	1.6	0-1 triplet	1.00	688	22.2
1.00	637	20.6	2 triplets	2.00	1434	46.3
2.00	1434	46.3	3 triplets	3.00	972	31.4
3.00	972	31.4				
TOTAL	3094	100.0	TOTAL		3094	100.0

F4 Triplet environment, TRAB = 4		
Value	Frequency	Percent
.00	2584	83.5
1.00	456	14.7
2.00	54	1.7
TOTAL	3094	100.0

F5 Triplet environment, TRAB = 5			F45	Triplet environment, TRAB = 4 or 5		
Value	Frequency	Percent	Value Label	Value	Frequency	Percent
.00	2338	75.6	0 triplet	.00	1993	64.4
1.00	683	22.1	1 + triplets	1.00	1101	35.6
2.00	73	2.4				
TOTAL	3094	100.0		TOTAL	3094	100.0