Social Networks, 3 (1981) 215–223

The NEGOPY Network Analysis Program

William D. Richards
*Simon Fraser University**

Ronald E. Rice
*Stanford University***

A method and its associated computer program for (specifically communication) network analysis are described. The program described here, NEGOPY, is relational, or linkage-based. The conceptual orientation, computational algorithm, operating characteristics, format and availability of NEGOPY are described. Finally, a partial bibliography of works describing other aspects of NEGOPY and research studies using NEGOPY is included.

Key words: network analysis, computer program, relational method, network algorithm, NEGOPY.

Introduction

NEGOPY is an implementation of a discrete linkage-based clique-detection method for the structural analysis of large networks. It is *linkage-based* because it uses for data a description of the system in terms of pairwise relationships (links) between nodes. This is the form of the data throughout the analysis; the program does not utilize a statistical 'distance' or 'similarities' form of the data at any time. Thus, NEGOPY provides *relational* and not *positional* analysis (see Burt 1980). The method is *discrete* because it results in a classification of all the nodes in the network into a number of distinct categories, based on patterns of interconnection. The program is able to analyze data from large networks—ones in which there are up to 2^{12} nodes and 2^{15} (2^{16} in some implementations) links.

The conceptual orientation of this analytic approach and an efficient computational algorithm that has been developed to implement the approach are described below. Finally, the operating characteristics of the NEGOPY program (data format, user-controllable parameters, output options, etc.) are briefly summarized.

*Department of Communication, Simon Fraser University, Burnaby, British Columbia, Canada.
**Institute for Communication Research, Stanford University, Stanford, CA, U.S.A.

Conceptual orientation

NEGOPY assumes the following definitions:

(1) *System*
A set of elements, components, or units (in the case of a communication network, these may be persons, groups of persons, or even whole organizations) that are related to one another in such a way that their behavior is interdependent.

(2) *Network*
A set of 'nodes' or 'points' connected by 'links' or 'edges' that correspond to the individuals in the system and th their pairwise relationships.

(3) *Relationship*
A condition of interdependence between two or more individuals. A relationship may be symmetrical (non-directed) or asymmetrical (directed).

(4) *Link*
A link between a pair of nodes in a network indicates that there is relationship between the corresponding pair of individuals in the system. A link may be directed or non-directed (as the corresponding relationship is asymmetrical or symmetrical). A link may also be reciprocated or unreciprocated, but this is another matter entirely, since it is tightly bound to the process of measurement.

(5) *Measurement*
The process by which it is determined which individuals in the system are connected by the type of relationship under consideration.

(6) *Reciprocation*
Agreement on the nature and strength of the relationship between a pair of individuals. (Note that this definition makes no reference to direction or symmetricality.)

(7) *Strength*
The extent to which a pair of individuals in the system are interrelated by the relationship under consideration. In the network representation, strength may be simplified by dichotomizing, so that a link is either present or absent; or it may be represented as a scalar quantity which may vary more or less continuously over a range of values.

(8) *Role*
A category into which network members are placed, based on their patterns of interconnection with other members. The roles include:

(A) *Isolate*

These nodes are minimally connected to the rest of the network, having a most one link with others (isolate Type 2) or having no links with others (Type 1).

(B) *Participants*

These individuals all have at least two links with others.

(i) *Group member.* These nodes have most of their interaction with other members of the same group. Note that the criterion involves *amount* of interaction, rather than simply the *number*, or even the *presence* or *absence*, of links. If, for example, amount of interaction is operationalized as number of hours spent in conversation, an individual having a total of 12 hours of interaction with members of the network, eight of which are with members of a single group, will be categorized as a member of the group, even if those eight hours are accounted for by only two of that individual's six links. A *group* is more precisely defined as a set of individuals who satisfy the following criteria:

(a) There are at least three individuals (nodes).

(b) The nodes pass the membership test (i.e., have more than a set threshold percentage of interaction with other members of the same group; NEGOPY's default percentage is 50.01%).

(c) The nodes are connected. That is, there must be some path, lying entirely within the group, from each member to every other member.

(d) There are no critical nodes. A node is 'critical' if its removal from the group causes the group to fail to meet any of the other criteria.

(ii) *Liaison.* A node which connects two or more groups within a system without belonging to any group. A liaison Type 1 is only one link from any group; a liaison Type 2 is at least two links from at least one group.

The computational algorithm

The NEGOPY program utilizes a highly efficient two-stage algorithm to accomplish the categorization of members according to the criteria outlined above. The first stage is a heuristic procedure designed to give a good (but inexact) first approximation to a structural description. It is supplemented with a second stage, which refines the description so that it meets the formal criteria.

The first stage may be characterized as using a heuristic pattern-recognition technique. It has two parts. The first is an iterative process that performs an operation that is in some ways analogous to simultaneously permuting the rows and columns of an adjacency matrix, and which gives results that resemble those of the old matrix manipulation techniques. The second part involves a scanning process that draws boundaries around clusters of tightly packed subsets in the (virtual) adjacency matrix, giving a first approximation to a description of the group structure in the network.

The iterative process involves the repeated application of the following equation to each individual in the network:

$$M(i, t+1) = \frac{\sum_j M(j, t) \cdot S(i, j) \cdot WF(i, j)}{\sum_j S(i, j) \cdot WF(i, j)}$$

$M(i, t)$ is the 'mean' for individual i at iteration t. $S(i, j)$ is the strength of the link from individual i to individual j. $WF(i, j)$ is the number of individuals with which both i and j have links (i.e., the size of the overlap of the first-order zones of i and j). The identification number of each node is used as the initial value for that node's mean.

In order to make the process converge faster and to prevent it from converging to a useless solution, a normalizing operation is performed at the end of each iteration. This operation closes up very large gaps between means and enhances the pattern that is emerging by increasing or decreasing the size of the distance between successive means in such a way as to prevent ties, to reduce the distance between means that are relatively close to one another and to increase the distance between means that are relatively far apart.

The end result of four or five iterations is a continuum with a scattering of points along its length. The points corresponding to sets of tightly interconnected individuals will be found in very tight clusters on the continuum. These clusters are identified by a scanning procedure which counts the number of points visible through a window which is moved along the length of the continuum. The location of boundaries is determined by the results of a test that measures the amount of change in the distribution of points visible through the window at successive points along the scan. The final product of this part of the algorithm is a tentative specification of the group structure in the network.

The tentative solution is refined by the application of a series of logical tests which verify the correctness of the original assignments of individuals to role categories. Incorrect assignments are rectified by relocating the individual(s) concerned into the proper classification. The final result of this process is a solution in which the members of the network are correctly described in terms of the logical criteria outlined above.

Operating characteristics of the NEGOPY program

(1) *Parameters and program control*

The operation of NEGOPY is controlled by a number of parameters which are used for a variety of purposes. Some describe the format and other relevant characteristics of the data that are to be input; some control various aspects of the program's execution; and others control the output of the program. All parameters take on default values unless the user wishes to specify other values for any particular run. While there are about forty parameters altogether, it is rarely necessary to set more than six.

Parameters are set by means of 'parameter cards' which are read in before the actual data. Values are specified by naming each parameter and stating the value the user wishes it to take.

(2) *Input specification and options*

The actual network data take the form of a list (or series of lists) rather than a matrix. This is because, first, the internal representation of the data is a list structure; and, second, a list of the links that exist is the most economical form of input data, especially in large networks which tend to have very sparse matrix representations. The links for an individual are specified by giving that individual's identification number (ID), and, for each input record, sets of link data. Each set includes the ID of an individual named by the subject individual (or obtained by a variety of data-gathering procedures such as observation, monitoring, questionnaire, diary, etc.) and the strength of the link. The links may be read in in any order—input sequence within a dataset is irrelevant. The strength may be specified by zero, one, or two numbers, called 'weights'. The simplest case would involve no weights, with a value of '1' provided by NEGOPY to indicate the presence of a link. Or, one weight could be a scalar measure of frequency or evaluation. For one or two weights, there are provisions for a wide variety of transformations to combine or rescale the weights in order to give, for each link, a single value that is an approximation to a ratio-level indicator of the strength of the relationship between the corresponding individuals in the system. These transformations may include scale reversals, additive or multiplicative combinations of multiple weights, and exponentiation. The user has full control over these operations and may elect to have the program do anything from very complex transformations to no transformations at all. It is thus possible to do a number of runs on the same basic dataset, each time testing the effects of different assumptions on the data and the way they are scaled.

Besides the network data, it is possible (but not necessary) to read up to 20 columns of alphanumeric information for each individual or node in the network. This information is referred to as the individual's 'name' (which indicates one possible use), although it may include any kind of alphanumeric information. The individual's 'name' is printed every time the individual is referred to in the analysis.

(3) *Execution options*

NEGOPY will accept binary or scalar-valued links for relationships that may be assumed to be either directed or non-directed. It can be instructed to perform a variety of 'cleaning' operations on the data. For example, cutoff values may be set, so that all links with a strength (after transformation) value lower than the cutoff are not included in the analysis. In addition, there are four ways in which the program can be instructed to deal with unreciprocated links. Under the assumption of directed links, the program may be instructed to use all the links that appear in the data, whether or not they are reciprocated. Under both directed and non-directed options, the program may be instructed to remove all unreciprocated links from the analysis, or to keep unreciprocated links if their strengths are above a specified cutoff value and to reject them otherwise. Finally, under the non-directed option, the program may be instructed to force reciprocation by adding the 'missing halves' of all unreciprocated links.

Other parameters may be used to control various aspects of the iterative process that is used in group identification or to direct the operation of the logical testing processes that follow the 'tentative' phase. It is generally not necessary to use these parameters. Since the operation of the program may be altered in subtle or profound ways with the aid of these parameters, the user is advised to proceed with knowledge of the data, of the research assumptions, and the effects of parameter changes.

(4) *Output options*

The user may ask for a number of types of information to be printed during or after a network run. The options include:

(A) A listing of all the individuals in the network and the links of each. The description of each individual's links may include information about link strengths, discrepancies between strength as reported by the two individuals involved (for reciprocated links), etc.

(B) A distance matrix ***D*** for each group. There is a row and a column for each member of the group; the value in ***D***(*i*, *j*) is the number of steps in the shortest path from individual *i* to individual *j*. Several types of summary statistics are also presented with each matrix.

(C) A final description and breakdown for each group. The links of each member in the group are described in terms of the identity of the other end of the links, the strengths, and the type of link (within-group, between-group, liaison, etc.)

(D) Descriptions of the links of individuals not included in any of the groups. These will include all liaisons and isolates that have links. The analysis here is similar to that provided for the members of each group.

(E) A brief summary file specifying the role category of each individual, the group to which the individual is assigned (for group members), and the density of the first-order zone (the integrativeness) of each individual in the network.

(F) A file containing the whole data structure, which includes information about the classification of each individual, the membership of the groups, all links in the network, and so on. This file is simply a listing of the internal data structure. It is used as input for additional routines that carry out further analysis, such as NETPLOT (Lesniak *et al.* 1977) which produces a graphical representation of the network.

Program format and availability

NEGOPY was written in CDC FORTRAN extended. It uses that language's ability to perform logical shifting and masking (bit-by-bit Boolean algebra) operations in order to create a pseudo list processing language that makes it possible to set up the linked-list structures needed to handle large networks efficiently. This approach to the data problem gives the program its large capacity. The price that is paid for the extra capacity is that the program is presently available only for CDC computers having 60-bit words. Versions of the program are now available to run on both IBM and UNIVAC machines. The new package will be a series of modular units that give the user greater flexibility and power.

The current version of NEGOPY is running at about a dozen universities in North America and Europe. Some users, unfortunately, have modified their copies of NEGOPY without authorization, and some of these versions produce erroneous results, such as the errors reported by Rice (1978, 1979a, b). Users should contact the authors for further information on this issue, for user manuals, or for copies of the program. The bibliography lists works describing other aspects of NEGOPY and research studies using NEGOPY.

Bibliography

Burt, R.
1980 "Models of network structure". *Annual Review of Sociology* 6:79–141.

Cavalcanti, C.
1980 *Diffusion of Innovations in Rural Brazil.* Unpublished doctoral dissertation. Stanford, CA: Stanford University.

Chavers, D.
1976 *Social Structure and the Diffusion of Innovations: A Network Analysis.* Unpublished doctoral dissertation. Stanford, CA: Stanford University.

Danowski, J. A.
1976 "Environmental uncertainty and communication network complexity: a cross-system, cross-cultural test". Paper presented at the annual meeting of the International Communication Association, Portland, Oregon.

Danowski, J. A. and R. V. Farace
1974 "Communication network integration and group uniformity in a complex organization". Paper presented at the annual meeting of the International Communication Association, New Orleans.

Dennis, H. S., G. M. Goldhaber and M. P. Yates
1978 "Organizational communication theory and research: an overview of research methods". In D. R. Brent (ed.), *Communication Yearbook 2.* New Brunswick, NJ: Transactional Books.

Edwards, J. A. and P. R. Monge
1977 "The validation of mathematical indices of communication structure". In D. R. Brent (ed.), *Communication Yearbook 1.* New Brunswick, NJ: Transactional Books.

Farace, R. and T. Mabee
1981 "Communication network analysis methods". In P. Monge and J. Cappella (eds.), *Multivariate Analysis Techniques in Communication Research.* New York: Academic Press.

Farace, R. V., P. R. Monge and H. M. Russell
1977 *Communicating and Organizing.* Reading, MA: Addison-Wesley.

Farace, R. V., J. A. Taylor and J. P. Stewart
1978 "Criteria for evaluation of organizational communication effectiveness: review and synthesis". In D. R. Brent (ed.), *Communication Yearbook 2.* New Brunswick, NJ: Transactional Books.

Fisher, F. A.
1978 "Information systems theory and research: an overview". In D. R. Brent (ed.), *Communication Yearbook 2.* New Brunswick, NJ: Transactional Books.

Goldhaber, G.
1976 "The ICA communication audit: rationale and development". Paper presented at the annual meeting of the Academy of Management, Kansas City.
1978 *Organizational Communication.* New York: William C. Brown. 2nd edn.

Goldhaber, G., M. Yates, D. Porter and R. Lesniak
1978 "Organizational communication 1978". *Human Communication Research* 5:76–96.

Krippendorff, K.
1977 "Information systems theory and research: an overview". In D. R. Brent (ed.), *Communication Yearbook 1.* New Brunswick, NJ: Transactional Books.

Lesniak, R., N. Yates, G. Goldhaber and R. Richards
1977 "NETPLOT: An original computer program for interpreting NEGOPY". Paper presented at the International Communication Association, Berlin.

MacDonald, D.
1976 "Communication roles and communication networks in a formal organization". *Human Communication Research* 2: 365–375.

Moch, M., J. Feather and D. Fitzgibbons
1980 "Conceptualizing and measuring the relational structure in organizations". In S. Seashore *et al.* (eds.), *Assessing Organizations: A Guide to Practice.* New York: Wiley-Interscience.

Monge, P. R., J. A. Edwards and K. K. Kirste
1978 "The determinants of communication and communication structure in large organizations: a review of research". In D. R. Brent (ed.), *Communication Yearbook 2.* New Brunswick, NJ: Transactional Books.

O'Reilly, C. A. and K. H. Roberts
1974 "Information filtration in organizations: three experiments". *Organizational Behavior and Human Performance 11*:253–265.

Porter, L. and K. H. Roberts
1972 *Communication in Organizations.* Technical Report #12. Contract #N00014-69-A-0200-9001, NR151-315, Washington, DC: Office of Naval Research.

Rice, R.
1978 *NEGOPY User's Manual.* Stanford, CA: Institute for Communication Research.
1979a "Investigations into the validity and reliability of NEGOPY: a computer program for network analysis". Paper presented at the annual meeting of the International Communication Association, Philadelphia.
1979b "NEGOPY: parameter effects". Paper presented at the Seminar on Communication Network Analysis, East–West Communication Center, Honolulu.

Rice, R. and W. Richards, Jr.
1983 "An overview of network analysis methods". In Dervin, B. & M. Voigt (eds.) Progress in Communication Sciences, Vol. 6. Norwood, NJ. Ablex, 1983, in press.

Richards, W. D.
1975 *A Manual for Network Analysis Using the NEGOPY Program.* Stanford, CA: Institute for Communication Research.
1978 "Network analysis methods: a conceptual and operational comparison". Paper presented at the Seminar on Communication Network Analysis, East–West Communication Center, Honolulu.
1979 "Measurement problems in network analysis I: reciprocity and directed links". Paper presented at the annual meeting of the International Communication Association, Philadelphia.

1980 "A cognitive-constructivist approach to measurement in communication network research". — Paper presented at the founding meeting of the Canadian Communication Association, Montreal.

Richards, W. D. and G. Lindsey
1979 "Social network analysis: a review of recent developments". In K. Krippendorff (ed.), *Communication and Control in Society*. New York: Gordon and Breach.

Richetto, G. M.
1977 "Organizational communication theory and research: an overview". In D. R. Brent (ed.), *Communication Yearbook 1*. New Brunswick, NJ: Transactional Books.

Roberts, K. H. and C. O'Reilly
1974 Measuring organizational communication. *Journal of Applied Psychology* 59:321–326.
1978 "Organizations as communication structures: an empirical approach". *Human Communication Research* 4:283–293.

Roberts, K. H., C. O'Reilly, G. Bretton and L. Porter
1973 *Organizational Theory and Organizational Communication: A Communication Failure?* Technical Report #2. Contract #N000-314-69-A-0222. Washington, DC: Office of Naval Research.

Rogers, E. M. and R. Agarwala-Rogers
1976 *Communication in Organizations*. New York: Free Press.

Rogers, E. M. and L. Kincaid
1981 *Communication Networks: Toward a New Paradigm for Research*. New York: Free Press.

Schwartz, D. F.
1968 *Liaison-Communication Roles in a Formal Organization*. Communimetrics Reports, No. 1. Fargo: North Dakota State University.

Wigand, R.
1976 "Communication and inter-organizational relationships among complex organizations in social service settings". Paper presented at the annual meeting of the International Communication Association, Portland, Oregon.
1977 "Some recent developments in organizational communication: network analysis—a systematic representation of communication relationships". *International Journal of Communication Research* 2:181–200.
1979 "Communication network analysis: a computerized tool in the analysis of large social systems with applications for organizations". In G. Klir *et al.* (eds.), *Basic and Applied General Systems Research*. New York: Gordon and Breach.