COMPUTER PROGRAMS AND SOCIAL NETWORK ANALYSIS

Linton C. Freeman,
University of California, Irvine

This is a review of the history and current state of the art in computer application for social network analysis. It documents a development from small task-specific programs to large general purpose program packages. It also shows a trend away from mainframe computers and toward micro-computers for network analysis. Seven currently available programs are examined in an attempt to show the kinds of analyses for which each is designed.

I. Introduction

The aim of the present paper is to review the role played by computers in the development of social network analysis. The first section will describe the approach of the network analyst and attempt to characterize the computational demands generated by that approach. The second section will review some of the highlights in the history of the development of programs for network computing. And the last section will examine the current generation of programs in order to uncover their strengths and weaknesses for various kinds of applications and, finally, to make some guesses about future trends.

II.The Computational Demands of Social Networks Analysis

Social network analysis and the use of computers in research have developed side by side over the past thirty years. As a matter of fact, as Wolfe (1978) has suggested, the development of network analysis could not have occurred in a context without computers. The sheer magnitude of the computational task required by network analysts eliminates the possibility of hand calculation.

Traditional procedures for data analysis in most social sciences are relatively simple by comparison. We have always focused on the attributes of individuals or other social actors. Thus, data come to us in the form of a rectangular, individual by attribute, matrix. Such have permitted us to use the same standard linear or log-linear analyses that are used in biology and psychology. For these kinds of analyses, the basic unit of study is the individual, and, for any one attribute, the magnitude of the computing task depends only on the number of individuals studied.

In contrast, social network analysis is not focused on the individual. It is explicitely concerned with the social relations -- the ties -- that link people to each other. Thus, the fundamental data for the network analyst are records of who is socially linked to whom. Many different kinds of social relations are studied, and data are collected on whether a given pair of individuals is, or is not, linked, or on the strength of the relation linking that pair of individuals.

Thus, because network analysts study relations among individuals, their data are recorded in square, person by person, matrices. Each relation recorded generates a matrix where each row and each column represents one of the actors studied. A given cell in the matrix contains a record of whether the row actor has the relation with the column actor, or the amount of the row actor's relation with the column actor.

For the network data, then, the unit of analysis is the pair, and, for any given relation, the size of the computing task is a function of the square of the number of individuals. This means that unless they are content to study very small networks, network analysts require considerably more computing power than that needed by the traditional analysts who study individuals.

A single network study might generate several of these matrices, each recording a different relation over the same set of individuals. But whether a study is focused on one or many relations, it is the job of the network analyst to uncover structural "patterns" in, and across, the data matrices. Thus since network analysis is explicitely focused on social structure, it requires not only big computing facilities , but entirely new kinds of programs that embody novel computational procedures.

Since the needed computational power was not generally available until sometime in the mid late 1960s, network analysis did not begin to develop as a specialty until the 70s (Alba, 1982). Some important, but small scale, network research was conducted in the mid-1950s (Barnes, 1954; Bott, 1955). But it was not until the end of the 1960s that we begin to see the emergence of large scale network data collection and analysis (Mitchell, 1969).

Typically, programs for analyzing network data involve the search for "socially important" structural patterns in data matrices. But what do network analysts see as socially important? Two main kinds of patternings have emerged as foci of interest in this context.

First, motivated by traditional concerns with "social groups," "social circles," "cliques" and the like, analysts have been concerned with uncovering collections of individuals that are both tightly linked together and more or less clearly separated from others. This is subgroup analysis and it is based on finding sets of people who are socially close to each other or proximate.

The second approach is motivated by traditional interests in "statuses," "roles," "social positions," "stations" and the like. Here analysts look for sets of individuals who are equivalent in the sense that they are linked in the same ways to the same or to equivalent others. This kind of analysis uncovers positions in the structure and it requires finding sets of people who are similar in the ways in which they are linked into the total network.

Most analytic procedures developed for the study of networks embody one or the other of these approaches. Thus, although network analysts have developed new procedures for collecting and analyzing data, their intuitive foundations are grounded in traditional social science concerns. The contribution of social network analysis lies in specifying these traditional coccerns explicitly and in developing rigorous methods for studying them empirically.

III. The development of Programs for Social Network Analysis

The earliest efforts at programming in social network analysis were all relatively simple and task specific. Each specifies a structural property of interest, and each embodied a single algorithm to uncover that property. But, as time has passed, the trend has been toward creating larger and more complex programs. These newer programs are also more apt to be general purpose in orientation. In this section the older task specific programs will be reviewed, and in the next section we will examine the newer programs.

The first network analysis programs were directed toward subgroup analysis. In 1949 Luce and Perry (1949) had specified a graph-theoretic definition of a clique. But developing an algorithm to actually find such structures in data turned out to be non-trivial. Harary and Ross (1957) finally specified an algorithm (but not a program) to find Luce-Perry cliques.

Programs were written, but the clique notion itself demanded such extensive computations that it was (and still is) painfully slow. So investigators began to seek other methods. And, lacking alternative theoretical foundations, various *ad hoc* methods for finding clique-like clusters were introduced. Many proposals were made, and finally, in 1960, Coleman and MacRae (1960) introduced a program for finding clusters that was the first really large scale computer program for network analysis. Their program would find all of the more or less distinct clusters in a group of up to 1000 individuals.

Attempts to develop new theoretical foundations for subgroup programs continued through the 1970s. Alba (1973) generalized the earlier theoretical conception of Luce and Perry, and, along with Gutman (1972), he produced and distributed a general set of computer programs, called SOCK, to calculate and display various kinds of graph-theory-based cliques. Seidman and Foster (1978) derived an alternative graph-theoretic generalization of cliques and included it in a general-purpose network analysis program, SONET, that began distribution that same year. Since SONET is a general-purpose program, it will be reviewed in the next section. Finally, Mokken (1979) proposed still another generalization based on graph theory that has, so far, not been fully implemented in a program.

Work on subgroups that was not explicitly grounded in theory continued also. In 1973 Bernard and Killworth (1973) introduced CATIJ,a computer program designed to locate subgroups through a kind of factor analysis. And in 1975 Richards (1975) released an early version of an elaborate subgroup finder, NEGOPY. This program is essentially task specific; it finds subgroups. But NEGOPY is a large and complex program. It includes, not one, but a collection of routines. Therefore, like SONET, NEGOPY will be described in detail in the discussion of contemporary programs below.

Like the earlier work on subgroups, that focused on social positions, these efforts began with a foundational theoretical essay. Lorrain and H. White (1971) soon introduced a program called BLOCKER and Breiger, Boorman and Arabie (1975) introduced another called CONCOR. Soon, Burt (1976) produced a third algorithm that he incorporated into a program called STRUCTURE. All three of these programs partition the individuals in a network into subsets, or "blocks" that each contain individuals that are approximately structurally equivalent. The Burt program, STRUCTURE, has been greatly expanded and has become quite elaborate; it too will be examined in the next section.

More recently, Sailer (1978), Everett (1982), Wu (1983), Mandel (1983), Winship and Mandel (1984), D. White and Reitz (1983) and Breiger and Pattison (1986) have introduced various modifications and extensions of the original algebraic formalisms introduced by Lorrain and H. White. In effect, these provide alternative theories of social position. And all of these alternative theories have led to the development of new programs. Each of these new programs reveals something different about the way people are positioned in a social structure.

Finally, from the very beginning, programs have been developed that are not directly motivated either by the notion of subgroups or that of positions. Most of these programs have been designed to calculate various structural parameters of social networks. And most of them have been grounded in graph theory.

The largest effort of this sort was initiated by Bavelas (1948). Bavelas drew from the graph-theoretic notion of "center" and defined the centrality of individuals in a network. The stimulus of Bavelas' work led others to try their hands, and various other definitions of centrality were proposed over the years. Some were *ad hoc* and some were grounded in theory. I reviewed them all (Freeman, 1979) and rederived those that could be given graph-theory foundations. They were embodied in a program called CENTER.

This, then, was the state of the art in network computing at the end of the 70s. There were literally dozens of individual programs that were more or less available to network data analysts. Almost all of them were task-specific. Each was designed to permit the calculation of one, or at most a few, properties of a social network. They were written in everything from APL to PL/I, and, typically, each had been implemented on only one mainframe computer. A few, like Alba and Gutmann's SOCK, were widely distributed, but most were available only to a small ingroup of potential users. Network computing was still very much in its infancy.

IV. The Eighties: the Era of General Purpose Network Computing

By the late 1970s it was widely accepted that progress in social network analysis awaited some further developments in computer programming. What seemed to be needed was a general-purpose program that could integrate some or all of the existing task-specific programs into a single package. In addition, transportability was a problem. A desirable package should be easily adapted to a wide range of local computing environments all over the world.

Four efforts to produce such a transportable general-purpose package were initiated. Heil undertook the task of producing an integrated set of network analysis routines at Toronto. At the same time, Mokken and Stockman led a Dutch group centered at Amsterdam in a similar task. And D. White and Sailer at the University of California, and Payne, Deans and Mitchell at Oxford did likewise.

Unfortunately, three of these four projects were ill fated. Payne and company lost their funding in 1981. And both Heil and the D. White-Sailer team made the mistake of trying to write their program in an IBM language (no longer supported by IBM) called APL. APL is easy to write and it is useful for turning out quick and dirty "use-one-time-then-throw-away" computer codes. But APL has the unfortunate characteristic of being completely unstructured. This means that it is almost impossible to update and maintain, and it is therefore inappropriate for producing programs for general distribution. For that reason, both Heil's and D.White and Sailer's attempts faltered; only the Dutch program, called GRADAP, succeeded.

Two new efforts along the same lines were started in 1983. One of these, led by Pappi and Kappelhoff at Christian-Albrechts University in Kiel, resulted in a package called SONIS. The other, developed by a team at the University of California at Irvine, was named UCINET.

By 1984, then, six network-oriented computer programs were being distributed. They were:

- SONIS by Pappi and Kappelhoff,
- UCINET by Freeman,
- GRADAP by Mokken and Sockman,
- SONET by Seidman and Foster,
- NEGOPY by Richards, and
- STRUCTURE by Burt.

These six programs were, and still are, (along with one newer program called AL that will be described below) "the state of the art" for network computing. Each is unique in form, generality and transportability. But, in terms of their basic design philosophies, they can be divided into two distinct sets.

One set contains SONIS, UCINET and GRADAP. These programs are all very similar in conception. They are designed to be general-purpose. They each provide broad and wide ranging collections of routines that permit the use of network methods in almost any area of application. Thus each of them includes many of the task-specific subgroup, position and centrality programs outlined above. All of these programs, moreover, also include collections of network data that can be used for instruction or simply to demonstrate the routines.

SONIS was originally implemented on Siemens and CDC mainframes as well as on a DEC mini-computer. For the Siemens and the DEC, however, the portability of the program is potentially limited since parts were written in an operating-system-specific assembly language. The CDC version, on the other hand, is written entirely in PASCAL and FORTRAN-5, so it has a greater potential for being adapted for use elsewhere. For English speaking users it should be noted that the command language of SONIS is German!

UCINET was originally written in a composite composed of Basic, Fortran and Turbo-Pascal for the IBM-PC micro-computer operating under DOS. Because of the universality of DOS, the transportability of UCINET was markedly easier than that of SONIS. On the other hand, this easy transportability involved a cost. UCINET, so far, is relatively slow and may be used only on fairly small data matrices.

In addition, early versions of UCINET were inadequate in handling data transformations. Because of that limitation, Borgatti developed a program that was originally no more than a set of data transformation routines called AL. AL was written at Irvine in Turbo-Pascal for the IBM-PC.

Since later versions of UCINET have eliminated the original problem by incorporating AL-like procedures to transform data, AL has moved off in a new direction. The latest release of AL embodies high speed versions of several of the UCINET routines as well as a virtual memory option that permits (*very*) slow analysis of matrices up to 1000 by 1000 in size. Moreover, the the new version of AL includes a whole set of the newest experimental routines that may reflect future trends in the analysis of network data. AL should interest those who are in the business of developing new analytic methods for network data. GRADAP was originally written in 1981 in the ALGOL 60 language for CDC mainframes. From the beginning, therefore, it has been able to handle large data sets. Like SONIS and UCINET, it has several network data sets integrated into the program itself, but unlike those other programs, the data contained in GRADAP are large. Moreover, its routines are consistent with and can be integrated into the standard SPSS statistical package.

GRADAP emphasizes graph-theoretic procedures in measuring structural characteristics. It stresses particularly the calculation of various kinds of centralities. This means that it is somewhat less general and includes fewer procedures than SONIS or UCINET. But the use of graph theory means that the user always knows exactly what is being calculated and why.

The other set of programs, NEGOPY, STRUCTURE and SONET are somwhat narrower in orientation than the three discussed so far. Rather than attempting to provide tools for all kinds of applications, these programs are designed to meet the specific needs of particular users. NEGOPY was written to meet the needs of people working in communications; STRUCTURE was designed specifically for applications in sociology; and SONET was produced as an aid for research in anthropology. Thus, while SONIS, UCINET and GRADAP all aim towards increasing the breadth of applications, NEGOPY, STRUCTURE and SONET aim to provide depth for users with particular disciplinary concerns.

NEGOPY was originally developed by Richards at Michigan State University. It was programmed in CDC Fortran Extended for the CDC 6500 mainframe. Because Fortran Extended was machine specific, the program could not originally be run on any other machine. Subsequent IBM mainframe versions were developed, and finally, to make it transportable, NEGOPY was rewritten in standard Fortran-77.

NEGOPY is essentially a practical routine to locate subgroups. It embodies a number of traditional intuitive ideas about groups and it strings those ideas together into a sequence of procedures. As a whole, this sequence of procedures seems to uncover those subgroups that are consistent with ethnographic intuition.

STRUCTURE was originally developed by Burt at the University of California, Berkeley. It was written in a general version of Fortran, so it has always been reasonably transportable. Originally, STRUCTURE was simply just another special purpose program. Its purpose was to find positions in a network through a new kind of approximation to structural equivalence developed by Burt. But the program has been extended to include a whole range of procedures of particular interest in sociological research. Although its main emphasis is still on positions, it now includes routines for subgroup detection as well as procedures to examine contagion, structural autonomy and equilibria (Burt, 1982, 1986; Burt and Minor, 1983).

SONET was written in PL/I for IBM mainframes. Like GRADAP, it utilizes procedures based on graph theory. But, while GRADAP stresses general procedures for calculating standard structural parameters, SONET includes some special routines based on ideas from traditional kinship analysis. It is designed primarily for applications in anthropology. Moreover, its portability is somewhat limited by the fact that it was written in a language that is only available on IBM mainframes, but for those who have access to such machines, SONET is a powerful program.

The most recent developments include new versions of all of these programs. With the collaboration of Bruce MacEvoy, UCINET was, in 1987, released in an entirely new version written completely in BASIC. This permits greater flexibility in the sizes of matrices that may be analyzed. STRUCTURE, was recently adapted to the IBM-

PC micro-computer, and in 1987 a new expanded version was released. Both NEGOPY and GRADAP have also been tested in IBM-PC versions and both are now ready for release. And SONIS has undergone similar testing and may be available by this time. Thus, like UCINET, all of these programs are now available to almost any user. Only SONET has, so far, not been adapted for micros. But a new expanded version of SONET has been produced and is available for mainframe applications where PL/I compilers are available.

The future of network computing can be read in the current trends. All the programs are moving in the direction of micro-computer implementation. At the moment this means that they will be slow. But the incoming generation of micros is introducing much faster processors and that problem should disappear soon. More important perhaps, current micro-computer operating systems impose serious limitations on the sizes of the data matrices they can handle. New operating systems that will eliminate this problem have been promised, but so far they have not appeared. When they are available, micro-computers will be able to process data set of a magnitude equal to that of those currently processed on mainframes.

Bibliography

Alba, Richard d. 1982. Taking stock of network analysis: A decade's results. *Research in the Sociology of Organisations, 1,* 39-74.

Alba, Richard d. and Myron P. Gutmann. 1972. SOCK: a sociometric analysis system. *Behavioral Science, 17,* 326.

Barnes, John A. 1954. Class and committees in a Norwegian island parish. *Human Relations, 7,* 39-58.

Bavelas, Alex. 1948. A mathematical model for group structures. *Human Organisations, 7,* 16-30.

Bernard, H. Russel and Peter D. Killworth. 1973. On the social structure of an ocean-going research vessel and other important things. *Social Science Research, 2,* 145-184.

Bott, Elisabeth. 1957. *Family and Social Network.* London: Tavistock.

Breiger, Ronald L., Scott A. Boorman and Phipps Arabie. 1975. An algorithm for clustering relational data, with applications to social network analysis and comparison with multidimensional scaling. *Journal of Mathematical Psychology, 12,* 328-383.

Breiger, Ronald L. and Philippa Pattison. 1986. Cumulated social roles: the duality of persons and their algebras. *Social Networks, 8,* 215-256.

Burt, Ronald S. 1976. Positions in networks. *Social Forces, 55,* 93-122.

Burt, Ronald S. 1982. *Toward a Structural Theory of Action.* New york: Academic Press.

Burt, Ronald S. 1987. Social contagion and innovation: cohesion versus structural equivalence. *American Journal of Sociology, 92,* 1287-1335.

Burt, Ronald S. and Michael Minor. 1983. *Applied Network Analysis.* Beverly Hills: Sage.

Coleman, James S. and Duncan MacRae. 1960. Electronic processing of sociometric data for groups up to 1,000 in size. *American Sociological Review, 25,* 722-727.

Everett, Martin. 1982. A graph-theoretic blocking procedure for social networks. *Social Networks, 4,* 147-167.

Freeman, Linton C. 1979. Centrality in social networks I: conceptual clarification. *Social Networks, 1,* 215-239.

Harary, Frank and I. A. Ross. 1957. A procedure for clique detection using the group matrix. *Sociometry, 20,* 205-215.

Heil, Gregory H. and Harrison C. White. 1976. An algorithm for finding simultaneous homomorphic correspondences between graphs and their image graphs. *Behavioral Science, 21,* 26-35.

Lorrain, Francois and Harrisson C. White. 1971. Structural equivalence of individuals in social networks. *Journal of Mathematical Sociology, 1,* 49-80.

Luce, R. Duncan and A. Perry. 1949. A method of matrix analysis of group structure. *Psychometrika, 14,* 95-116.

Mandel, Michael. 1983. Local roles and social networks. *American Sociological Review, 48,* 376-386.

Mitchell, J. Clyde. 1969. *Social Networks in Urban Situations.* London: Tavistock.

Mokken, Robert J. 1979. Cliques, clubs and clans. *Quantity and Quality, 13,* 61-75.

Richards, William D., Jr. 1975. A manual for network analysis (using the NEGOPY network analysis program). Unpublished ms., Institute for Communication Research, Standford University . Sailer, Lee.

Richards, William D., Jr. 1978. Structural equivalence: meaning and definition, computation and application. *Social Networks, 1,* 73-90.

Seidman, Steven and Brian L. Foster. 1978. A graph-theoretic generalization of the clique concept. *Journal of Mathematical Sociology, 6,* 139-154.

White, Douglas R. and Karl Reitz. 1983. Graph and semigroup homomorphisms on semigroups of relations. *Social Networks, 5,* 193-234.

Winship, Christopher and Michael Mandel. 1983. Roles and positions: a critique and extention of the blockmodeling approach. In S. Leinhardt, ed., *Sociological Methodology 1983-84.* San Francisco: Jossey-Bass.

Wolfe, Alvin W. 1978. The rise of network thinking in Anthropology. *Social Networks, 1,* 53-64.

Wu, L. 1983. Local blockmodel algebras for analyzing social networks. In S. Leinhardt, ed., *Sociological Methodology 1983-84.* San Francisco, Jossey-Bass.

Addresses at which available network analysis programs may be obtained:

- SONET -- Brian L. Foster, Dean, Graduate School, Arizona State University,Tempe, AZ 85287.
- GRADAP -- Inter University Project Group GRADAP, Technical Center FSW, University of Amsterdam, Roetersstraat 15, 1018 WB Amsterdam, The Netherlands.
- SONIS -- Institut fur Soziologie, Christian-Albrechts Universitat,Olshausenstrasse 40, D-2300 Kiel 1, Federal Republic of Germany.
- UCINET -- Linton C. Freeman, Program in Mathematical Social Science, School of Social Science, University of California, Irvine CA 92717.
- NEGOPY -- William d. Richards, Jr., Department of Communication, Simon Fraser University, Vancouver, BC, Canada.
- STRUCTURE -- Ronald S. Burt, Reseach Program in Structural Analysis, Center for the Social Sciences, Columbia University, New York, NY 10027.
- AL -- Steve Borgatti, Program in Mathematical Social Science, School of Social Sciences, University of California, Irvine, CA 92717.