

## PROGRAM ABSTRACTS/ALGORITHMS

### Program package GRADAP: Graph definition and analysis package

C. VAN DE WIJGAART

*Technisch Centrum FSW, University of Amsterdam  
Amsterdam 1000 HE, The Netherlands*

Graph analysis is applicable in every situation in which "objects" (points) have relations with (lines toward) other objects and in which the structure of these relations is to be studied. Some examples are: relations between persons (family, friends, commercial relations), relations between functions (e.g., functions fulfilled by the same person), and functions between political parties and states, where a common border can be considered as the relation defining the graph. The aim of graph analysis is the detection of patterns in such networks, the interdependence of patterns formed by different relations, the implications of patterns for the behavior of objects, and the impact of the qualities of the objects on the patterns. Graph analysis, therefore, consists of methods that, on given graphs, try to find: subgraphs (subsets of points in the graph) that meet certain requirements of relations (like strong relations), partial graphs (all points, with a subset of lines) in selecting relations, relations between such subgraphs, detection of trees in the graph (e.g., the shortest spanning tree), and detection of paths (e.g., the shortest path). Of course, the analysis becomes more complicated when: a directed graph is used (lines run in only one direction), the graph is "signed" (each directed line has a positive or negative sign), the points and/or lines are valued (e.g., the "distance" between two points as a value for the connecting line), or more than one type of graph is used at the same time (a multigraph). Also, more than one relation between points may be considered in the same analysis.

Secondary analysis may consist of the analysis of properties of points within detected clusters, the comparison of clusters, and so on.

A simple example of a graph is the sociogram in Figure 1. In Figure 1, A to I represent children in a class. An arrow from A toward B represents A choosing B as a playmate. Other relations, defined for the same children, could be "previously studied together," "are members of the same club," and so on. Additional information includes age, sex, school results, and so on. An example of additional information is a preference order

In order to develop powerful and user-friendly software in this field, an interuniversity project group was formed, consisting of universities of Amsterdam, Groningen, and Nijmegen (The Netherlands). The project has been supported by a Grant 40-9 of the Netherlands Organization for the Advancement of Pure Research (Z.W.O.).

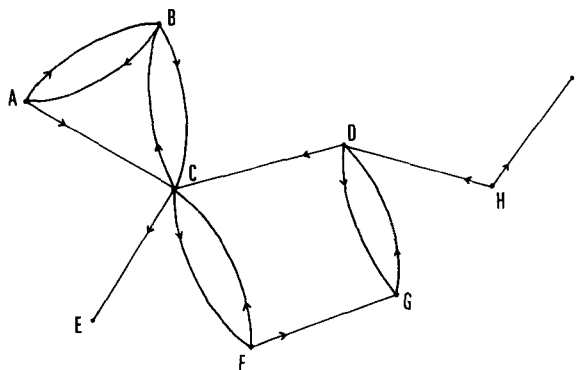


Figure 1. A directed graph.

(A's first choice of a playmate is B; A's second choice is C). A general introduction to graph analysis is given in Harary, Norman, and Cartwright (1965).

**GRADAP.** The GRADAP package is an organized set of programs with a standard data file, providing flexible graph entry possibilities and large opportunities for different graph representations, graph management, and analysis. Results of graph management and analysis can be transferred directly to SPSS and STAP (the statistical appendix to SPSS, developed at the University of Amsterdam). SPSS system files (as produced by SPSS-6000 Version 8 and SIR: Scientific Information Retrieval System) can be submitted directly to GRADAP. The package is comparatively fast and can handle up to 6,000 points and 60,000 lines. The current version is suited for use on CDC Cyber installations (operating systems NOS and NOS/BE) and is written in standard FORTRAN IV, augmented with CDC assembler (e.g., in memory management and input/output) and comment: FORTRAN, 56,400 lines; assembler, 16,400 lines; comment, 29,400 lines. Conversions to FORTRAN V and to IBM/370 are being studied. The GRADAP control language is patterned according to the syntax of the SPSS control language.

In GRADAP, a graph is defined by four types of data: points, lines (between points), sets of points, and sets of lines (including the sets of all points and of all lines, further user defined), and information on the points and lines. The user-defined sets of points and lines provide subgraphs and partial graphs. Graph transformation may be performed through selection, induction, aggregation, and definition of new constructs (e.g., now point sets).

GRADAP offers a number of analytic procedures. **ADJACENCY** computes the adjacency matrix and optionally lists the isolated points and, for each non-isolated point, its neighbors. **CENTRALITY** computes a number of point and graph centrality measures based on both the adjacency and the distance matrix.

DISTANCE computes the distance matrix and corresponding statistics based on neighborhood. REDUCE is a procedure that reduces the graph in an iterative process to the null graph, according to criteria based on adjacency and incidence. RUSH computes the centrality measure Rush (Anthonisse, Note 1) for points and, optionally, for lines. This measure is based on the structural property "betweenness." SUBGRAPHS detects subgraphs, such as strong/weak components and cliques (Alba, 1973; Mokken, 1979).

Nonanalytical, informative procedures include the following. LIST GRAPHINFO prints descriptive information for all the objects of the graph (cf. LIST FILEINFO in SPSS). LIST POINTS and LIST LINES print a list of points/lines with the specified point/line information (cf. LIST CASES in SPSS). WRITE POINT/LINES writes the specified point/line information to the alternate output file (cf. WRITE CASES). WRITE GRAPHINFO writes all graph definition and documentation information to the alternate output file (cf. WRITE FILEINFO in SPSS).

An example of a GRADAP set-up, using an analytical procedure on a predefined graph, which is stored on a system file "CLASS," is as follows:

```
GET GRAPH CLASS
ANALYSE SETS INTRA=PLAY
SUBGRAPHS DETECT=STRONG/MINSIZE=2
OPTIONS 2
FINISH
```

Applied on the school data, shown in Figure 2, this set-up has the following effect: ANALYSE SETS selects the subgraph, defined by the relation "likes to play with" (i.e., a subset of lines, maintaining all points). On this subset, the analysis SUBGRAPHS is performed, according to the specification DETECT=STRONG and MINSIZE=2. This means that all "strong" subgraphs are selected (i.e., all subgraphs in which there is a path between all pairs of points) and that the minimal number of points of a subgraph will be two. The analysis results in the subgraphs shown in Figure 2. OPTIONS 2 causes output on the alternate output file.

**Documentation.** The first volume of the *GRADAP User's Manual* (Stokman & van Veen, 1981) consists of an introduction, a description of graph management and informative procedures, a summary of all non-

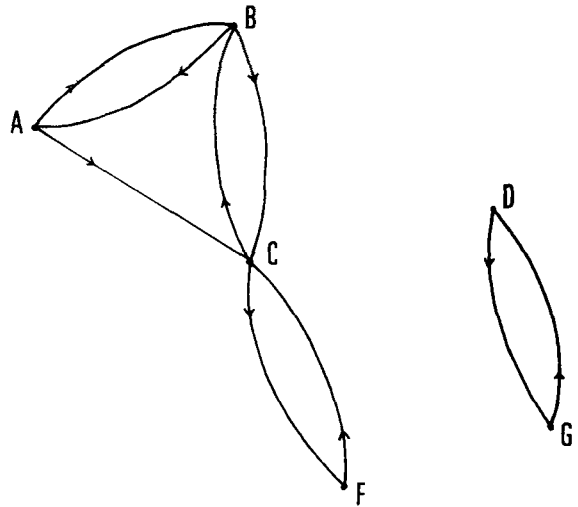


Figure 2. Strong components of the graph of Figure 1.

procedure and procedure statements, and a recommended sequence of statements in a GRADAP set-up. The second volume consists of six chapters, each describing an analytic procedure. The description of the analytic procedures consists of a theoretical introduction to the method used, a description of the program, a control statement description, including options and statistics of the procedure, and sample output.

The package GRADAP (CDC version) and the *GRADAP User's Manual* can be obtained from the author. For academic institutions, the initial license fee is \$500; the annual renewal is \$200. The price of the *GRADAP User's Manual* is \$13.

#### REFERENCE NOTE

1. Anthonisse, J. M. *The rush in a directed graph*. Amsterdam: Mathematical Center, 1971.

#### REFERENCES

- ALBA, R. D. A graph theoretic definition of a sociometric clique. *Mathematical Sociology*, 1973, 3, 113-126.
- HARARY, F., NORMAN, R. Z., & CARTWRIGHT, D. *Structural models. An introduction to the theory of directed graphs*. New York: Wiley, 1965.
- MOKKEN, R. J. Cliques, clubs and clans. *Quality and Quantity*, 1979, 13, 161-173.
- STOKMAN, F. N., & VAN VEEN, F. J. A. M. (Eds.). *GRADAP user's manual*. Amsterdam: University of Amsterdam, 1981.

(Accepted for publication December 30, 1981.)