Anthropology

Analyzing Large Kinship and Marriage Networks With Pgraph and Pajek

DOUGLAS R. WHITE
University of California, Irvine

VLADIMIR BATAGELJ
ANDREJ MRVAR
University of Ljubljana, Slovenia

Five key problems of kinship networks are boundedness, cohesion, size and cohesive relinking, types of relations and relinking, and groups or roles. Approaches to solving these problems include formats available for electronic storage of genealogical data and representations of genealogies using graphs. P-graphs represent couples and uncoupled children as vertices, whereas parent-child links are the arcs connecting nodes both within and between different nuclear families. Using results from graph theory, P-graphs are shown to lend themselves to solutions of the problems discussed. Relinking of families through marriage, for example, can be formally defined as sets of bounded groups that are the cohesive cores of kinship networks, with nodes at various distances from such cores. The structure of such cores yields an analytic decomposition of kinship networks and constituent group and role relationships. The Pgraph and Pajek programs for large network analysis help both to represent kinship networks and their patterns and to solve problems of analysis.

Keywords: kinship, marriage networks, large network analysis, graph theory, cohesion, social groups and roles, genealogy, graphic representation, Pajek, Pgraph

INTRODUCTION

In the study of contemporary and historical societies, ethnography, social demography, political and economic elites, social class, ethnicity, and numerous other fields, the analysis of those social networks that are partially ascriptive, such as those constructed through relationships of kinship, might present formidable problems. Computerized genealogical data are now available for tens of millions of people as well as for large genetic, demographic, ethnographic, and historical databases.

Five Key Problems of Kinship Networks

Even small-scale kinship studies require that we consider the first and foremost problem of limits—how to bound the field of study. Social networks constructed through inter-

Social Science Computer Review, Vol. 17 No. 3, Fall 1999 245-274

marriage, unlike egocentric genealogical studies of ancestries, pose problems of how to anchor and bound the network.

A second problem is conceptual. If we look for socially cohesive networks of kinship and marriage as one of the possible ways in which to bound such a study, then what do we mean by *cohesion*? Traditionally, cohesion in social networks studies has begun from the small-groups approach and defined the paradigm of social cohesion as the clique in which there is a complete set of connections (e.g., friendships) between pairs of individuals. Cohesion in kinship and marriage networks is more diffuse and might require an altogether different set of conceptual models.

A third and practical problem, related to that of limits and concepts of diffuse cohesion, is that of size. Is it possible to study the structure of large social networks, and up to what point is it feasible and useful to visualize network structures in terms of graphical representations?

A fourth problem is how to define and code the basic social relations of kinship and marriage. Genealogists typically do so in terms of the primary relations of parent and child and the derived relationships of consanguinity and affinity (i.e., relations between two parents or connecting through a marriage or couple). Geneticists are concerned only with consanguinity. Anthropologists problematize the cultural construction not only of marriage in all of its myriad forms, including questions as to whether marriage is universal, but also of kinship relations in general. They question the universality of the assumption that biological parentage is the basic kinship relationship. Besides the question of how to treat the multiple types of parentage (e.g., biological, sociological, adoptive, ritual), cultural definitions of who are kin and who are not often depend on behavior, not simply ascription, so that both levels are significant.

A fifth set of problems is related to both cohesion and the difficulty of characterizing the social relations of kinship—those of the kin groups and of the networks of social roles that are either achieved or ascribed on the basis of kinship links. Given the fluidity and frequent overlap among different kinship groups, anthropologists worry a great deal about problems of validity in characterizing cohesive groups or interlocked sets of roles.

In this article, some approaches to solving these five problems are described. In the first part, different available formats for storing genealogical data in electronic form are presented. Then, possible presentations of genealogies using networks or graphs are described. It is shown that parentage graphs (P-graphs) are the most suitable presentation for our purposes. Some results are explained from graph theory that can be used to solve the problems posed. The final part describes the use of the Pajek program for representing kinship patterns and solving kinship problems.

AVAILABLE FORMATS FOR CODING KINSHIP DATA

Against this background of the problematics of kinship, a number of advances have been made that facilitate a network approach to kinship and marriage and that address the more general question as to why social scientists would want to study kinship and marriage networks in the first place.

White and Jorion (1992) address the problem of what are kinship links or how to study kinship networks with a minimal set of assumptions. They take as their starting point that kinship links are culturally constructed but that what makes them *kinship* as opposed to economic, political, affective, or religious ties (functions that clearly can be taken on by kinship links) is the underlying model of the relation of parentage in which parents precede their children as point of origin (for birth or nurturance) in childhood. Whereas the relations of parenting are variable, the potential existence of a parental couple is at least a recognized norm

TABLE 1
Coding Individuals and Their Relationships

5 Demo Data: Family Schmidt
(I4,1X,A21,A1,3I4,3x,I2)

1 John Schmidt	M	2	4	5	1920
1 John Schmidt	M	3	4	5	1935
2 Mary Finley	F	1	8	9	1920
3 Sara Gonzalez	F	1	0	0	1935
4 Adam Schmidt	M	5	0	0	1895
5 Elizabeth Markowitz	F	4	0	0	1895
6 Maude Schmidt	F	0	1	2	
7 John Schmidt, Jr.	M	0	1	2	
8 ?	M	9	0	0	
9 ?	F	8	0	0	

END OF DATA

Variables and variable labels:

1	2	3	4	5	6	7
Ego	Name	Sex	Spouse	Father	Mother	Marriage date

modeled on sexual reproduction. One can push this flexible definition in a biological direction to find the “unique” pair of genetic parents or open it up to allow a range of different types of parents and parental couples and relations to be defined in different cultural contexts.

Database Construction

A kinship network is constructed as a database by taking individuals as a starting point and noting their gender, their different conjoints (sexual or marriage partners in whatever form), and their children or their parents (e.g., biological, sociological, adoptive). Because parent-child is a reciprocal relation, it is sufficient in a database simply to list every individual and his or her parents and conjoints. A system of coding for individuals, conjoints, and parents is exemplified in Table 1. Here there are seven variables: a 4-digit *ego number*, a 20-character *name*, a character to indicate *gender*, three 4-digit numbers to identify a *spouse* and the *father* and *mother* of ego, and a number for the *year or decade* of the marriage (many other variables can, of course, be added to this format). In this system, *when a person has two spouses, as in the case of John Schmidt, they are given multiple lines in the database, each indicating a different spouse*.

Part of the construction of genealogical or kinship databases, as in the GEDCOM standard[1] format for exchanging computerized genealogical data, is the assignment of a unique set of numbers or identifiers not only to individuals, as in Table 1, but also to nuclear families, defined as conjoints and their offspring (if any). Every individual whose parent or parents are known can be assigned a corresponding FAMC (FAMily of Child) number, and every individual can be assigned a FAMS (FAMily of Spouse) number for each nuclear family in which he or she has a spouse or child. Sociologists call these the family of orientation versus the family of procreation (Murdock, 1949). White and Jorion (1992) take family numbering a step further and assign a unique nuclear family number even to unmarried children or to those individuals (e.g., spinsters, bachelors) who are not recorded as having any conjoint relationships. When two people marry, their conjoint unit is coded under a common FAMS number.[2] This double system for identifying the fundamental units of kinship, individuals, and families leads to a double system for coding kinship relations. In the ego-centered system, the kinship relations are between identified individuals. In the family-centered system or P-graph, the kinship relations are between families. Among the many commercial programs for drawing genealogies, there is only one that uses the family-centered system with couples at the vertices of the graph, and it is not well suited for network analysis of marriages.[3]

Data Transformation From Individuals to Couples: Ego2Cpl

The Ego2Cpl program (Ego and Cpl stand for individual vs. couple or family numbers) is the entry-level database engine[4] of the Pgraph system (White & Skyhorse, in press) originally created by White and Jorion (1992); henceforth, we use Pgraph for a software package as opposed to the P-graph genealogy drawing format. It converts from the numbering of individuals such as in Table 1 to family (FAMS/FAMC) numbering for Pgraph and GEDCOM files.

Table 1 is typical of formats used in computer files in which kinship data are collected. As noted, here Variables 1 to 3 are the ego identifier number, name, and gender; Variables 4 to 6 are the numbers of the spouse, father, and mother; and the final variable is the date of the marriage. Many other variables can be added to such records, but the first six variables (in whatever order)[5] are essential. The input file for Ego2Cpl conversion requires two header cards: one to identify the order and type of input variables (here with a Type 3 format that is discussed later) and another to identify the format for reading the variables. The format `(I4,1X,A21,A1,3I4,3x,I2)`, as shown in Table 1, specifies a 4-digit integer, a space, a 21- followed by a 1-character field, three 4-digit integers, 3 spaces, and a 2-digit integer that truncates the year of marriage into a decade (e.g., `1920` into `92`).

The program creates FAMS numbers for couples and remaining individuals and uses these numbers to assign FAMC numbers to children. It produces five sets of output. First, it flags errors in data entry and stores them in an output file. An individual, for example, might be listed once as a husband and again as a mother (gender error) or might be listed as his or her own parent (axiomatic error).[6] Second, it checks to see whether any individuals who are multiply married lack references to parents and creates a file of links to “virtual” parents needed in Pgraph and GEDCOM formats to keep track of multiply married persons.[7] Third, it creates a standard genealogical GEDCOM file (with extension *.GED), of which Tables 2 to 4 give an example. Fourth, it creates a file with the extension *.NET for use with the Pajek program and allows the user to choose different types of graphic labels. Fifth, it creates a series of files for specialized use with the Pgraph programs. The Pgraph and Pajek packages are discussed later.

GEDCOM Standards

The GEDCOM (*.GED) file header information produced by Ego2Cpl is shown in Table 2. The input file listed is one containing the data in Table 1. Records in *.GED files have a line starting with `0` and continue on lines starting with `1`.

The header data in the GEDCOM file produced as output from the data in Table 1 are followed by a series of records, one for each individual, as shown in Table 3. Each individual is assigned two family numbers, where the FAMC given to a child corresponds to the FAMS number of his or her parents. A list of the families follows, as in Table 4, giving the number of each family and the individual numbers for its members—husband, wife, and children.

TABLE 2
GEDCOM Coding of the Data in Table 1: Headers

```
0   HEAD
1   SOUR    PAF 2.2
1   DEST    PAF
1   DATE    03 MAR 95
1   FILE    p-dem.ged          (the file shown in Table 1)
1   CHAR    ANSEL
0   @S1@    SUBM
1   NAME    Doug White
1   ADDR    School Social Science
2   CONT    UC Irvine 92697
2   CONT    Internet Email address: drwhite@uci.edu
1   PHON    949-824-5893
```

The directed graphs drawn inside Tables 3 and 4, where child-to-parent links from lower to upper vertices (or, if transposed to parent-to-child, from upper to lower vertices) are stipulated to be asymmetric, represent the two different types of networks that can be coded in the ordinary manner of social networks: the first, with individuals as vertices, and the second, in Pgraph format, where conjoints (couples or single individuals) are the vertices. The double horizontal lines in the first graph indicate a symmetric marriage relation. This graph, in Table 3, has two types of relations—those with parents (solid lines) and those between spouses (double lines)—and the labeling of vertices must indicate gender of individuals. If we remove the marriage relation, then we have the genetic graph of Ore (1963), used by geneticists, in which no graphic relation is given between couples who have no children (e.g., Individuals 1 and 3). Where there are two or more children, the skeleton or symmetric graph of parental relations within a nuclear family forms one or more cycles.

Figure 1 shows a P-graph of the basic kinship data in Table 1 in terms of relations among the six nuclear family or conjoint vertices, F1 to F6. The labels on the lines connecting the families indicate linking individuals. Individual 2, for example, is a wife and mother in Family F1, but her parents are in F4. A coding of two types of lines (dotted and solid) shows whether the linking individual is female (solid) or male (dotted)[8]—in the case of Individual 2, female. Note that the line for Female 3 is a demiarc (Harary, 1971) because it does not link to a set of parents (the identities of her parents are unknown). Pairs of demiarcs also have been drawn for Families F4 and F2—one for the husband and one for the wife in each case. An individual with plural marriages, such as Individual 1, will have separate arcs for each marriage, as is the case for Individual 1 in the figure.

SOLVING FIVE KEY PROBLEMS OF KINSHIP NETWORKS USING P-GRAPHS

For the study of our five theoretical problems—bounded subgraphs, cohesion, size, social relations, and groups—the P-graph conventions used in Figure 1 have considerable advantages over conventional network diagrams. To see these advantages, we have only to break the expected Western paradigm of "kinship as a tree" and examine the relevance of closed social circles of marriage among different families for opening up a host of questions of interest to kinship analysis. Thus, we add a link to Figure 1 to exemplify what happens when families "relink" through marriage.

TABLE 3
GEDCOM Coding of the Data in Table 1: Individuals

```
0 @I1@ INDI
1 NAME          John Schmidt
1 SEX M
? ?
1 FAMS @F1@
1 FAMS @F3@
1 FAMC @F2@

0 @I2@ INDI
1 NAME          Mary Finley
1 SEX F
1 FAMS @F1@
1 FAMC @F4@

0 @I4@ INDI
]1 NAME         Adam Schmidt
1 SEX M
1 FAMS @F2@

0 @I5@ INDI
1 NAME          Elizabeth Markowitz
1 SEX F
1 FAMS @F2@

0 @I3@ INDI
1 NAME          Sara Gonzalez
1 SEX F
1 FAMS @F3@

0 @I8@ INDI
1 NAME
1 SEX M
1 FAMS @F4@

0 @I9@ INDI
1 NAME
1 SEX F
1 FAMS @F4@

0 @I6@ INDI
1 NAME          Maude Schmidt
1 SEX F
1 FAMS @F5@
1 FAMC @F1@

0 @I7@ INDI
1 NAME          John Schmidt, Jr.
1 SEX M
1 FAMS @F6@
1 FAMC @F1@
```

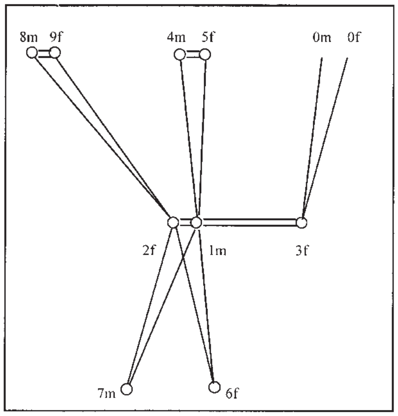
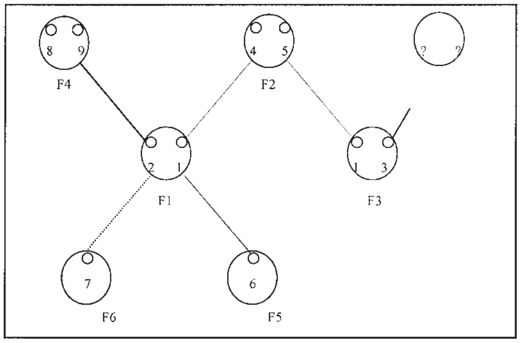


TABLE 4
GEDCOM Coding of the Data in Table 1: Families

```
0 @F1@ FAM
1 REFN F1
1 HUSB @I1@
1 WIFE @I2@
1 CHIL @I6@
1 CHIL @I7@

0 @F2@ FAM
1 REFN F2
1 HUSB @I4@
1 WIFE @I5@
1 CHIL @I1@

0 @F3@ FAM
1 REFN F3
1 HUSB @I1@
1 WIFE @I3@

0 @F4@ FAM
1 REFN F4
1 HUSB @I8@
1 WIFE @I9@
1 CHIL @I2@

0 @F5@ FAM
1 REFN F5
1 WIFE @I6@

0 @F6@ FAM
1 REFN F6
1 HUSB @I7@

0 TRLR
```



Numbers on the vertices are for families.

Figure 2 illustrates a marriage that relinks families already linked, what anthropologists call a relinking marriage. To show such a marriage, we added to Figure 1 a new link between Families F5 and F3: Ahmed Schmidt, son of F3, marries a paternal cousin, Maude Schmidt. Any marriage that completes a circle of ties in the skeleton of a P-graph (in this example, the circle is among Families F1, F2, F3, and F5) constitutes a relinking. Not all relinkings need be blood marriages; there also are affinal relinkings through one or more prior marriages. To illustrate an affinal relinking, we arrange for Ahmed's brother, Habib, to marry a daughter of F4, his stepmother's sister. New data would be entered in Table 1 for these two illustrations as follows:

10 Ahmed Schmidt	M	6	1	3	(Ahmed marries 6; his parents are 1 and 3)
11 Habib Schmidt	M	12	1	3	(Habib marries 12; his parents are 1 and 3)
12 Mathilde Finley	F	11	8	9	(Mathilde marries 11; her parents are 8 and 9)

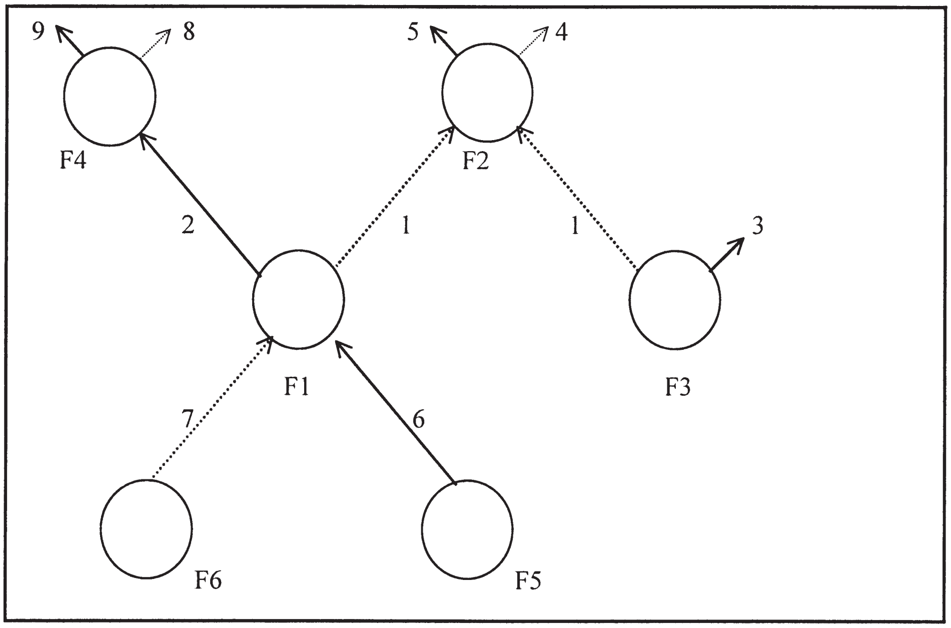


Figure 1: P-graph for the Data in Table 1
NOTE: Numbers on the lines are for individuals. Numbers on the vertices are for families.

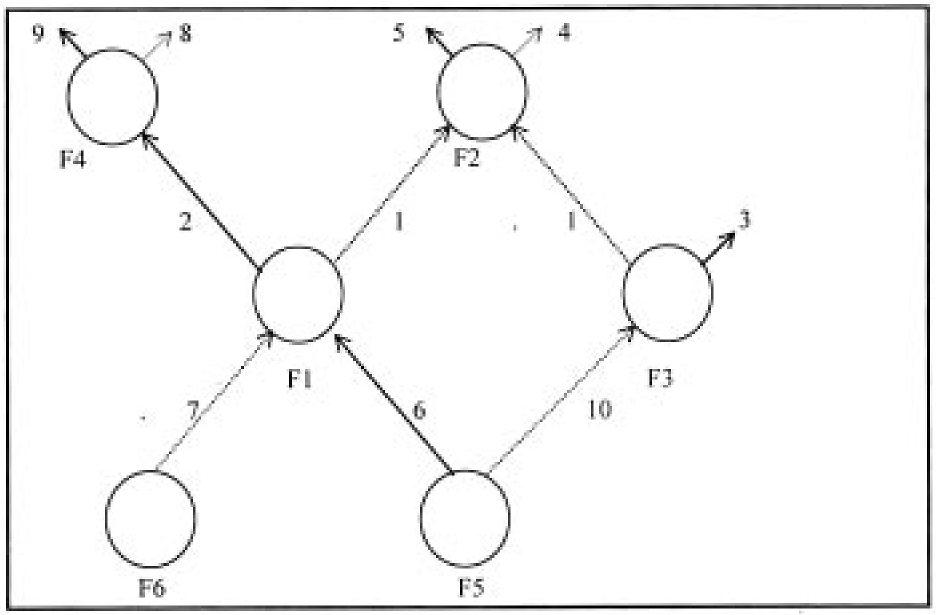


Figure 2: P-graph of Figure 1 Plus a New "Relinking" Marriage
NOTE: Numbers on the lines are for individuals.

Network Boundedness: Problem 1

Relinkings are relevant to our first and second theoretical problems, of the limits or boundaries of social groups, in terms of the network of kinship relations. Every relinking defines a bounded set of families not only connected but also doubly connected by a marriage circle. In any given body of genealogical data, there may be many such circles of varying lengths. The one in Figure 2 is of length 4. If Ahmed's brother married a daughter of F4, then the length of that circle would be 5. Ahmed's and Habib's marriage circles would overlap in sharing three of the same families, or two of the edges in the graph, in their marriage circles.

Those marriage circles in the undirected version (skeleton) of a P-graph that overlap in sharing one or more edges define a bounded social unit or subgraph, a set of families of which any pair is connected by one or more marriage circles. In graph theory, maximal subgraphs of this sort are called the 2-connected components of a graph, which we refer to here as *bicomponents*. A bicomponent of a Graph G is a subgraph of G with three or more vertices in which any pair of vertices is connected by two independent paths, hence, by a set of edges that form a cycle. Two distinct bicomponents of a graph can have at most one vertex in common between them. Hence, any given edge in the skeleton of a P-graph will belong to at most one bicomponent. By definition and formal proof (Harary, 1969, pp. 27-28), the following statements are equivalent for a Graph G:

G is a bicomponent.
G contains no cutpoint.
Every two vertices of G lie on a common cycle.
Every two edges of G lie on a common cycle.
There is a path joining any two vertices of G that uses any given edge of G.
There is a path joining any two vertices of G that uses every other vertex of G.
There is a path joining any two vertices of G that avoids any other vertex of G.

As self-bounding subgraphs of mutually exclusive sets of edges in a P-graph, bicomponents are easy and quick to determine by computer because the time required for computation is a linear function of the size of the network, unlike those many network problems that are polynomial of higher order or even exponential (nonpolynomial) in the time required for computation. White (1997) and Brudner and White (1997) developed both the vocabulary and our sociological understanding of the substantive relevance of identifying bicomponents as bounded subgraphs within kinship networks. They defined the populations in those marriages contained within bicomponents as bounded subgraphs or social units with structural endogamy.

Cohesion: Problem 2

Structural endogamy has substantive implications for the second of our theoretical problems: What are the socially cohesive subgraphs of a kinship and marriage network? The social implication of structural endogamy through network links within a set of living persons is that information (or any other resource) that might flow through these links is able to travel from anyone to anyone else in the set through multiple independent links. This enhances the conditions for cultural sharing, for the cross-checking of information and enforcement of sanctions, and for the maintenance of social boundaries, even where cycles are of considerable length.

White and Harary (1998) carry these ideas further to characterize the concept of social cohesion for social networks with the following insight: Because no cohesive group should contain a cutpoint whose removal leads to the disconnection of the cohesive group into two parts, cohesive groups must, at a minimum, have the graph theoretic structure of bicomponents. Bicomponents do not necessarily contain cliques; hence, the concept of overlapping cliques is not a necessary condition for bicomponents. White and Harary (1997, 1998) propose two measures of the social cohesiveness of the network graph, G, for a social group: (a) the *connectivity* of G or minimum number of independent paths between any two vertices in the group and (b) the *cycle density* of G. In a regular P-graph, vertices can have at most outdegree 2, however, the connectivity never will exceed 2.[9] Hence, the cycle density is the appropriate measure of cohesion. Mrvar and Batagelj (1998) also independently propose the cycle density of the undirected version of a P-graph as a *relinking index* for a genealogy:

$$P = \frac{k + m - n}{k + n - 2M},$$

where k = number of weakly connected components of the P-graph, n = number of vertices, m = number of child-parent links, and M = number of apical ancestral vertices having outdegree 0. In general, $k + m - n$ is the number of independent (basic) cycles in a graph. In a regular P-graph, where all but M ancestral vertices can have at most outdegree 2, the maximum number of edges is $2(n - M)$, so that the maximal number of cycles is $k + 2(n - M) - n = k + n - 2M$. Hence, the relinking index, p, is the cycle density of a P-graph. Mrvar and Batagelj (1998) prove that $0 < \mathrm{P} < 1$, where $P = 0$ if the genealogy is a forest, and that there exist arbitrarily large genealogies with $P = 1$ (e.g., two brothers marry two sisters). White and Harary (1997, 1998) show the same results independently and propose that the relinking index is most relevant when it is computed on a bicomponent of a P-graph, that is, for a structurally endogamous group. Otherwise, arbitrary data that happened to be collected outside of the structurally endogamous group will affect the index of relinking.

Network Size, Radial Cohesion, and Linked Vertices: Problem 3

Rethinking the concept of cohesion led White, Schnegg, and Brudner (in press) to define forms of cohesion that might be suitable for the study of the types of diffuse cohesiveness that often are found in the kinship networks of complex societies, where the length of marriage circles often is very large. They define the radial cohesiveness, R_k, of a bicomponent as the ratio D/Z_k, where D is the density of the bicomponent and Z_k is the density of the sum of ties in the kth order zone of ties at distance k from each ego. Thus, Z_1 is the density of the ties of those who are directly linked to a common ego, Z_2 is the density of the ties of those who are distance 2 from a common ego, and so forth. For Z_1 to be greater than zero in a P-graph, some marriages must be as close as uncles marrying nieces; hence, the ratio D/Z_1 often will be infinite. Z_2 will be greater than zero only if some marriages are as close as between cousins. For bicomponents in some populations, $D/Z_k > 1$, and as a statistically significant departure from chance, for up to quite large values of k. Such bicomponents are cohesive in the sense of structural endogamy as defined by Brudner and White (1997).

Having established that structurally endogamous clusters of kinship and marriage connections may be cohesive, another problem of boundaries and network size remains: What about vertices that are outside bicomponents? If they are connected to structurally endogamous clusters, then these vertices are potentially as important as those within the clusters because they may be the children, parents, spouses, or other relatives of those within the bicomponent to which they are attached. These linked vertices are ordered, however, by distance from the bicomponent, and they may be connected to more than one bicomponent at different distances.

Structurally endogamous subgraphs, as indicated by the frequency of other vertices linked to them, virtually never constitute populations that are endogamous to the point of closure or lack of marriage with those outside the structurally endogamous subgraphs. It often is the case in such an empirical subgraph that the majority of the kinship links or marriages are with those outside the subgraph. In no way do structurally endogamous subgraphs resemble caste-like marriage structures. They are merely a way of characterizing *that relinking occurs* within a population so as to create and reinforce social cohesion. In Brudner and White's (1997) Austrian village study, they show that some children of farm families marry or are linked into a single structurally endogamous bicomponent that runs through a whole ethnic segment of the farming valley of the region. The majority of children of these families, however, do not marry or have children who marry so as to become members of the bicomponent. This is a social system of farmsteads, however, that tend to pass intact to single heirs and in which the heirs and their spouses form a distinct social class of landed families as opposed to laborers, craftspersons, salaried workers, and professionals. The relevance of the study of relinking is that they show a strong correlation between those who relink and those who inherit or marry heirs. Brudner and White argue that structural endogamy is the basis for the formation of a landed family social class in this region, a class division that makes sharp divisions even within sibling groups.

Social Relations and Types of Relinking: Problem 4

The *specific ways in which relinking occurs* within a population to reinforce social cohesion poses a fourth theoretical problem—how to define basic and derived social relations of kinship and marriage networks. White (1997) discusses the relevance of different types of relinkings—the number of distinct families involved, the length of the cycles, the number of generations. The various types of relinking are easily characterized using P-graphs.[10]

The significance of the P-graph representation for the problems of bounded subgraphs and social cohesion in a given population is that *every circle of relinking marriages is reflected in a cycle of the undirected version (skeleton) of the P-graph, and every cycle of the P-graph corresponds to a circle of relinking marriages*. This is not true for the genetic graph such as in the vertical links of the graph within Table 3, where there are cycles inside each nuclear family having two or more children. It is even less true in the typical egocentric graph that contains both descent and marriage relations, so that even the father-mother-child triad is a complete graph that contains a cycle. In egocentric kinship graphs, there are many cycles that do not reflect relinkings and that clutter our ability to visualize and to analyze kinship and marriage structures.

Groups and Roles: Problem 5

As a coding of social relations, the P-graph is merely the most reduced possible graph of kinship and marriage networks. Many of the derived relations, starting with various types of siblingship (full siblings, half siblings), can be computed from different concatenations of the primary relation of parentage (e.g., full siblings = share common parents) or different paths in the P-graph. That different societies may emphasize different types of dyads or paths is part of the general observation that each case study may require understanding of different social processes and social emphases and that kinship relations at the behavioral or affective level may depend on patterns of behavior and sentiment that must be studied in their own right. In this double sense, the P-graph is simply a minimal network armature for the study of kinship networks and a sparse representation to which other social networks may be added. (As one wag remarked, “Kinship is to anthropology as the nude is to drawing.”)

If compound or distant kinship relations have a redoubling effect that heightens the importance of certain kinship linkages, then these effects are even more intensified if they are embedded within marriage circles or structurally endogamous bicomponents of the minimalist P-graph kinship network. Examining the distribution of those kinship relations that are activated in important behavioral transactions bears on our fifth theoretical problem, that of identifying social roles or groups. We cannot solve this problem of analysis in the abstract, but we would argue that for the network study of social organization—whether an analysis of social class, elites, or kinship or a succession to religious (or symbolic), economic (e.g., material), or political offices (or property)—it is a very useful adjunct to have a complete kinship scaffolding.

STUDIES USING P-GRAPHS

P-graph studies of kinship at present have the character of a large and open cooperative research project involving French, German, American, and Mexican social scientists as well as scholars of Asia, Africa, and the Pacific.[11] Exploration of new means of scientific visualization and computability have been two of the principal objectives in developing P-graph paradigms for a series of empirical research projects on marriage and kinship networks in a wide variety of theoretical and regional contexts. The programs and methods of visualization used in these projects between 1991 and 1998 were mostly those developed within the Pgraph program package (White & Skyhorse, in press) written by White. In 1998, the authors of Pajek, which is discussed later, adapted the P-graph format as the default for representing and displaying GEDCOM kinship network data. Whereas much had been learned about graphic displays of kinship networks up to that time (cf. Schweizer & White, 1998), Pajek opened up a whole new set of possibilities for actively manipulating images and exporting them to PostScript, VRML, and chemistry visualization formats, which allow color printing and three-dimensional (3D) rendering through Web browser plug-ins for virtual reality or molecular images.

One set of hypotheses currently being used in a wide number of P-graph studies are those exploring the links between structural endogamy (bicomponents) and their linked vertices as a means of identifying social units or subgraphs and cohesive phenomena such as social class, political or economic elites, ethnicity, succession to office or religious leadership, and the formation of social, economic, or symbolic capital. The transmission of power among Central American elites (White, 1996), the transmission of estates among Austrian farmers (Brudner & White, 1997), the ethnogenesis of Turkish nomads (White & Johansen, 1998), radial political cohesion of Tlaxcalan villagers (White et al., in press), and the biblical kinship charter for transmission of religious leadership (White & Jorion, 1992) provide examples using the P-graph as a framework for societal-level studies.

A second set of hypotheses seeks to explore specific genres of social organization (e.g., dual organization in marriage structures, generalized exchange, radial cohesion, semicomplex kinship systems) through network analysis. Houseman and White (1998a, 1998b) argue that the bicomponent structure of a kinship network contains much of the structural information for the study of marriage rules and emergent alliance structures. Their study of Amazonian societies (Houseman & White, 1998a) provides elements of a new vocabulary for the identification of structural regularities emergent out of marriage choices and, therefore, open to change as a product of social interaction. Their restudy (Houseman & White, 1998b) of Leach's monograph on Pul Eliya (Sri Lanka) provides a detailed case study in which the existence of dual organization is self-organizing at the level of behavior and egocentric kinship terminology in a context of bilateral kinship without a societal-level coding into descent-based moieties. As opposed to descent-based moieties, which have been the sole means to recognize dual organization in anthropology to date, they provide a novel network approach to the study of multiple forms of dual organization in marriage systems.

There are many other aspects of the theory of kinship developed through the vehicle of P-graph representations (White & Jorion, 1992, 1996; Héran, 1995) and remaining to be developed. We will return to the advances in computability using P-graph representations and data formats for the statistical study of bicomponents and various structural types of dual organization, relinking, and blood marriages. Some of the major advances in this area are

- new ways of controlling for confounding demographic variables when assessing the relative frequencies of various types of relinking and blood marriages;
- new ways of simulating comparative baseline frequencies for different constraints on random marriage regimes while controlling for confounding demographic variables; and
- statistics for evaluating network-emergent dual organization.

Because here we are addressing the widest possible audience in the social sciences, we next discuss the Pajek network analytic and visualization package used since 1998 (displacing earlier graphics programs in the Pgraph package) as the vehicle for graphics and network analysis in P-graph representation format. The easiest way in which to enter kinship data into Pajek is to begin with a format such as that of Table 1 (see Note 4) and transform the data to a *.NET or *.GED file using the Ego2Cpl program. There is an advantage of having Ego2Cpl produce network or *.NET files indigenous to Pajek in that names or numbers of individuals, and the colors and types of lines used in the graphs, are more easily managed.[12] Pajek network files also can handle many other types of social relations and attribute data.[13]

PAJEK: A PACKAGE FOR LARGE NETWORK ANALYSIS

Pajek (the Slovenian word for "spider") is a Windows (32-bit) program for network visualization and analysis, written by Batagelj and Mrvar (1997a, 1997b).[14] Because it is oriented to analysis and graphic representation of large networks, meaning thousands of vertices, it is especially well suited to the analysis of kinship and marriage networks along with a host of other types of data that may be available on either small or large populations of individuals.

Genealogical Data Entry

GEDCOM files are one of Pajek's standard input formats. The file in Tables 2 to 4, produced by the Ego2Cpl program in the Pgraph package or as *.GED output files of numerous standard genealogical programs, can be read directly into Pajek and visualized as a graph using the P-graph format. The P-graph representation format became Pajek's default for kinship data as of Version 19 in December 1997. The network or *.NET file, however, is the native Pajek file. Table 5 shows the data of Table 1 (plus Individuals 10, 11, and 12), having been processed through Ego2Cpl to produce a *.NET file where nuclear family numbers identify the nodes to which individuals are attached. Only one of the Pajek data formats is shown, with `*Vertices` and `*Arcs` as headers for two of Pajek's data types. The first data

TABLE 5
Schmidt Family Data in Native Pajek Format as a *.NET File

```
*Vertices    7
    1 “1 2”                                      0.3500    0.5000    0.5000
    2 “4 5”                                      0.5341    0.0589    0.5000
    3 “1 3”                                      0.7310    0.4982    0.5000
    4 “8 9”                                      0.1706    0.0589    0.5000
    5 “10 6”                                     0.5210    0.9214    0.5000
    6 “7 0”                                      0.1864    0.9214    0.5000
    7 “11 12”                                    0.3950    0.8625    0.5000
*Arcs
    1   2        1 c Black p Solid l “John Schmidt”
    3   2        1 c Black p Solid l “John Schmidt”
    5   3        1 c Black p Solid l “Ahmed Schmidt”
    6   1        1 c Black p Solid l “John Schmidt, Jr.”
    7   3        1 c Black p Solid l “Habib Schmidt”
    1   4        2 c Red p Dots l “Mary Finley”
    5   1        2 c Red p Dots l “Maude Schmidt”
    7   4        2 c Red p Dots l “Mathilde Finley”
```

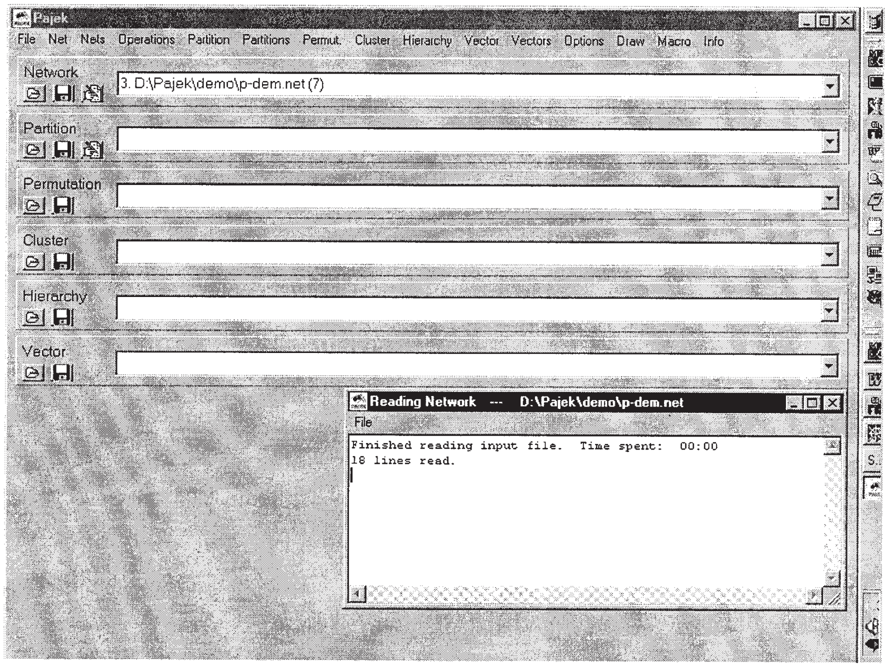
type lists the seven vertices or families (with labels in quotes giving husband’s and wife’s names). The second (`*Arcs`) lists two types of directed links between children and their parents in P-graph format. Here, the first entry indicates that Couple 1 is linked by John Schmidt (the husband: Relation Type 1) to his parents in Couple 2. The graphic instruction `c Black p Solid l “John Schmidt”` assigns black solid lines and a label to his arc. The second entry shows the same information for John Schmidt, this time from the perspective of Couple 3 where he has a different wife. The first five entries are for males (husband’s parents), and the last three are for females (Relation Type 2: wife’s parents), to be shown as red lines composed of dots with labels. This file has been read, and its data have been graphed, by Pajek, which has added the *x*, *y*, and *z* coordinates for each of the couples or individuals.

Algorithms for the Analysis of Large Networks

The philosophy underlying Pajek is to analyze networks in terms of a number of different types of objects (e.g., networks, partitions, hierarchies, clusters, permutations, vectors) and to implement algorithms that are subquadratic (i.e., with run time less than $O[n^2]$), where n is the number of vertices, that is, increasing less rapidly than the square of the number of vertices. The algorithm for finding the bicomponents of a network, for example, runs in $O(n + m)$, where m is the number of edges. Large networks often are sparse, and many sparse network algorithms run in $O(n)$, $O(n \log n)$, or $O(n\sqrt{n})$ time. Pajek provides very fast algorithms for selected large network problems. Finding all the possible paths (or shortest paths) between two vertices in a genealogical database of 5,000 people, for example, is implemented to run very quickly.

Pajek Menus

Pajek’s main menu contains pull-down options at the top of a screen that is largely composed of six small windows for showing which of the basic types of objects currently are in use in the analysis. Because these screens initially are blank, the place to begin is by choosing to read a network file, in either *.NET or *.GED format, in the uppermost window. To the left


Pajek
File Net Nets Operations Partition Partitions Permut. Cluster Hierarchy Vector Vectors Options Draw Macro Info
Network
Partition
Permutation
Cluster
Hierarchy
Vector
File

3. D:\Pajek\demo\p-dem.net (7)

Reading Network --- D:\Pajek\demo\p-dem.net

```
Finished reading input file.  Time spent:  00:00
18 lines read.
```

Figure 3: Screen Snap of Data Objects in Pajek's Main Menu

of each window are speed buttons for reading, saving, and editing of files. Figure 3 illustrates a screen snap after the user has clicked open the file P-DEM.NET of our illustrative input data. When a speed button is clicked for opening a file, a directory of files in the current directory with *.NET extensions (i.e., network files) appears. To read a *.GED file, one must click the option to change the data type and then click *.GED. The GEDCOM files created by Ego2Cpl will appear if they are in the current directory, and one may be clicked to open it. Once a network file is read, the options in the upper menu can be used:

```
File  Net  Nets  Operations  Partition  Partitions  Permut.
  Cluster  Hierarchy  Vector  Vectors  Options  Draw
  Macro  Info.
```

Options Versus Macros

Each of Pajek's menu choices activates submenu trees that are three to four levels deep, as in the manual listing menu choices available at Pajek's Internet address.[15] We refer here to menu item command sequences (including sub- and sub-sub-menu choices) by command labels (which are clicked) separated by slashes. For example, Info/Network shows the number of vertices in the network. Net/Transform/Transpose is an important option to consider at the start of analysis because it reverses the direction of arcs from child-to-parent (the P-graph convention) to the more intuitive parent-to-child direction.

The options Macro/Record and Macro/Play allow the user to save as in a command file and to play back a series of operations as a macro command. Standard macros for genealogical analysis, included with the program, are shown in Table 6. When macros are accessed by Macro/Play, a directory opens from which existing macro (*.MCR) files can be clicked.

TABLE 6
Pajek Macros

Name	*Operation*
	Compute genealogical layers and view initial picture (macros of the same name in the Draw subdirectory also will draw and show the graph)
Layers1	2D graphic with genealogy depth partition; use x and y keys to rotate (the picture still can be improved using averaging *x* coordinate, minimizing total length)
Layersz	3D graphic with spring embedding; use x, y, and z keys to rotate
	Find connections between people (enter two persons/couples)
Path	Shortest undirected path
AllPaths	All paths from *i* to *j* (to an ancestor)
	After bicomponents, enter bicomponent number to extract
HieCompL (Hierarchy Component Layers)	Extracts bicomponent from genealogy, computes genealogical layers, and draws genealogy
HieClNet (Hierarchy Cluster Network)	Extracts bicomponent and selects its cluster from genealogy
	Predecessors and successors of a person or a couple (the couple will be partitioned by distance)
3Bef3Lat	Three generations before and three generations after person or couple (enter the same person/couple twice)
AllBef	All predecessors before selected person or couple
AllLat	All successors after selected person or couple

AllBef and AllLat, for selecting predecessors or successors (all descendants, all ancestors, or all relatives, in the case of genealogical P-graph files), are important macros (equivalent to Net/K-Neighbors/Input or /Output) because these commands often are used to select a particular genealogical line within a potentially huge database (e.g., selecting all of Queen Victoria's descendants from the European royalty genealogy). The result of such commands is a partition in terms of distance from ego (Partition 0) to the farthest connected relative (Partition *k* at this distance [those unconnected are in the partition labeled "unknown" and are colored dark brown]). Info/Partition will show the distribution of distances. Operations/Extract from Network/Partition/Select classes from . . . to . . . will result in a new object shown in the Network window containing only the network selected. Draw/Draw Partition now will show this reduced network with different colors assigned according to the shortest distance to or from other connected nodes.

Organizing Visual Network Displays

The Layers1 macro used to create generational levels and draw a genealogy is equivalent to two sets of options. The first set, Net/Partition/Depth/Genealogy, causes a descriptor to appear in the Partition object window for a partition file containing the generation level of each vertex. Any such object appearing in a window may be saved to a file. At this point,

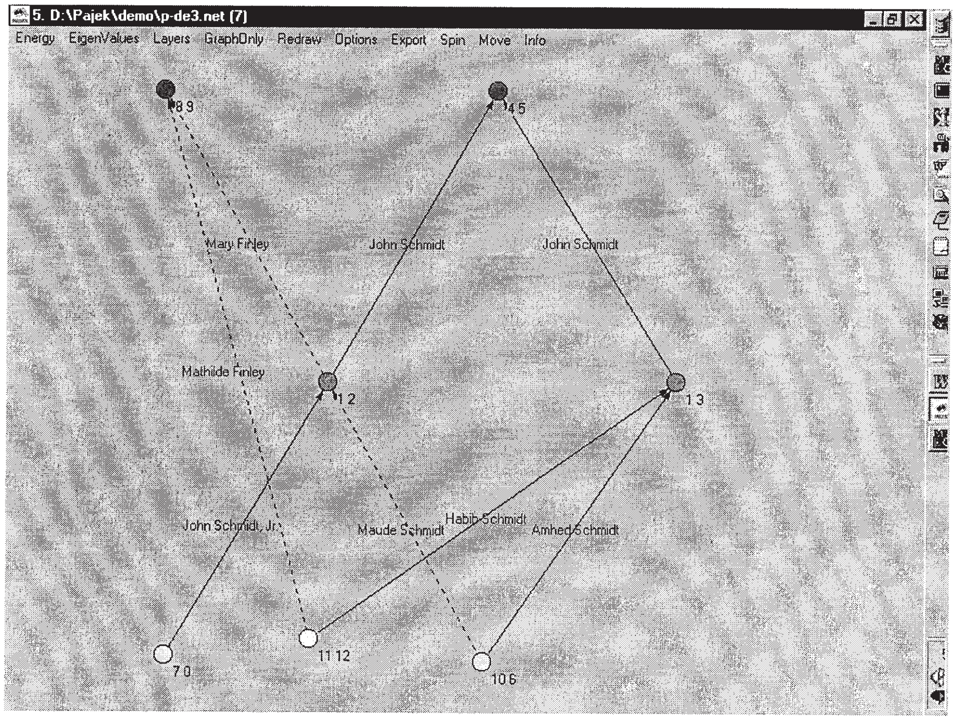


Figure 4: A Pajek Screen for the Schmidt Family

Info/Partition will show the number of vertices in each generation. The second set of options, Draw/Draw Partition, initiates the actual drawing of the current network according to the generational levels computed.

A two-dimensional (2D) graphic such as in Figure 4 is drawn by a slight variant from the Layers1 macro. The options used are Net/Partitions/Depth/Acyclic (acyclic is the variant) and Draw/Draw partition, followed by Layers/in y direction in the Draw screen's Graphics menu, and then holding down the *z* key to spin the graph 180°[16] so that descendants (colored yellow) are on the bottom. Then, Options/Lines/Mark labels/ will label the arcs, and Options/Mark vertices using/Numbers labels the vertices.

The input file for Figure 4 is from Table 1, plus the added persons (10-12), using either the *.NET or *.GED output of Ego2Cpl. This is a P-graph, where the arcs are labeled with names (note that Sara Gonzalez is missing because we did not have a link to her parents; for this reason, and because of multiple spouses, it sometimes is useful to create dummy parents) and vertices are labeled with both individual numbers. Other choices (e.g., names for couples at vertices, numbers for arcs) would have been possible with Ego2Cpl's *.NET output file as Pajek's input.

The Pajek Draw screen in which the drawing appears, as in Figure 4, has its own menu:

```
Energy  EigenValues  Layers  GraphOnly  Redraw  Options  Export
  Spin  Move  Info.
```

When the Draw screen is opened, it might appear at first to be empty, depending on whether coordinates saved from a previous session are included with the *.NET file (they are not with the *.GED file). Clicking Layers/in y direction will fix generations into layers and make a 2D picture appear. Once the graph is visible, the mouse can be used to click and move points (the lines move automatically), and the picture can be enhanced in a variety of ways. The first option under Layers/, /Type of Layout, provides 2D or 3D choices and affects the other options under Layers/. In 2D, the options Layers/in y direction, Layers/Averaging x coordinate, and Layers/Optimize layers in x direction will alter the picture in systematic ways. In 3D, the Layers/ options that change the parameters for the picture are Layers/in z direction, Layers/Averaging x and y coordinates, and Layers/Optimize layers in xy plane.

Options/ in the Draw window menu provides for different types of labeling, coloring, and size of the vertices and arcs. Options/Mark vertices using/Labels will show the direction and type of arrows as well as the names for husband and wife from the original data set. If no vertex labels are wanted, then Options/Mark vertices using/No labels will suppress the labels but keep the direction and type of arrows; arrows here (not transposed) are upward to parents. If the *.NET file was made by Ego2Cpl, then Options/Lines/Mark will label the lines with each person’s name or number, depending on the Ego2Cpl option chosen previously. This also is true if a *.GED file is read using Options/ReadWrite/Pgraph and labels. The settings chosen, using Options/ReadWrite in the Main window and Options/Lines/MarkLines in the Draw screen, are saved when work with Pajek is finished and is restored when Pajek is run again.

Move/ in the Draw window provides parameters for screen editing with the mouse. Move/Fix y (or /Fix x), for example, lets you move vertices only in the horizontal (or vertical) direction, whereas Move/Grid allows you to change format for the grid spacing between vertices.

For a 3D graphic, Macro/Layersz (clicking the LAYERSZ.MCR file in a directory) is equivalent to clicking, in the Graph window, Layers/Type of layout/3D-layers in z direction /Energy/Starting positions/Given z coordinates/2D-Fruchterman Reingold. This begins a minimum-energy configuration algorithm commonly called spring embedding. The algorithm pushes unconnected pairs of vertices (with no child-parent arcs) toward opposite perimeters of the graph and pulls connected vertices toward one another (and, therefore, toward the center) until both the length of links is minimized and the distance between unconnected points is maximized. The 3D graph can be rotated around the *x*, *y*, or *z* axis by the use of the corresponding key (i.e., holding down the shift key along with the x, y or z key reverses each of these rotations). Starting configuration options may make for different outcomes on each run, so outcomes should be saved as files.

Viewing and Printing Exported Graphics

Because Pajek is a Windows program, pressing the PrintScreen key copies the screen image into the Windows clipboard. For printing or further enhancement, the clipboard image is then inserted into a word processor or graphics editor.

Export/ from the Draw window includes options not only for printer and graphic format files but also for saving animated displays to files that can be viewed by Web browser plug-ins. The following are the current export formats.

PostScript *.PS and *.EPS (Encapsulated PS) files are standard formats that can be printed with PostScript (or GhostScript)[17] and GhostView programs. Pajek’s Draw window provides Export/Options for borders, shapes, and backgrounds; the graph corresponds to what is seen on the screen.

VRML *.WRL files are produced from the Draw screen with Export/Options for size of lines and vertices and for background colors and are read with Virtual Reality Internet browser plug-ins (e.g., Cosmo Player) that allow movement within the 3D structure and viewing labels by clicking vertices.[18]

MDL *.MOL files are molecular structure output that make beautiful 3D images that can be animated for rotation, rotated with the mouse, viewed in stereo, and endowed with other features (Freeman, Webster, & Kirke, 1998). It requires the Chemscape Chime viewer.[19] Vertex labels are lost in the current implementation, however, so these images are mostly useful for viewing general structures.

KineMage *.KIN files offer a spectacular molecular structure output that provides most of the features of the *.MOL files (Freeman, 1998a; Freeman et al., 1998) but also include options for labeling vertices. A series of slides may be animated, and vertices may be clicked to view or suppress their labels. The advantage for presentations is a series of different slides or views of a structure. The KineMage option that we jointly conceived as a result of writing this article (as with a number options discussed here), and that Mrvar implemented in Pajek, is an animation of birth, death, and marriage processes unfolding as a moving window image over generations. The Draw screen option Export/KineMage asks “How many neighbor classes to show” and “Resize option”; the responses “–4” and “2” will show moving windows of four successive generations as a dynamic evolutionary view of the kinship structure. This moving image is especially effective if the vertices are partitioned not just by generation but also by birthdate or birth intervals such as decades.

KineMage is the only one of the present VRML and chemistry viewers that distinguishes different types of arcs while also providing labeling for vertices.[20] Labels in Chime, for example, are chemical (not social) network labels, and Chime does not distinguish different types of arcs. None of these viewers at present, however, provides labeling for arcs. Many of the chemistry graphic engines are available for modification, and rapid evolution is occurring among 3D viewers.

Further Strategies for Organizing Visual Information and Analyzing Pajek Variables

The philosophy underlying graphic display in Pajek is to create and manipulate sets of objects on the graphic screen, either with the mouse or by embedded commands within the *.NET file, as exemplified in Table 5. Embedded commands to control the appearance of nodes, arcs, and labels are explained more fully in documentation attached[21] to the Web site for the manual listing menu choices for Pajek.

The strategy for on-screen editing emphasizes creating partitions that show as sets of colored nodes on the Draw screen and then moving single nodes by clicking on the node and moving the cursor (here it is useful to fix the *x* or *y* coordinate with the Move/Fix x or /Fix y option so as not to get nodes out of alignment with others in their generation) or moving clumps of colored nodes by clicking near a node of the selected color and then moving the cursor. To create partitions, we already have seen a number of Pajek menu operations. A direct way in which to make a new partition, however, is to create a cluster (*.CLS) file in ASCII format with a text editor (putting the node numbers for the new partition in successive rows of a *.CLS file) and reading the file into Pajek by selecting it within the Cluster object window. A partition containing these nodes is then created by the option Cluster/Make Partition. With Draw/Partition, the new partition will become a set of colored nodes within the original network. Clusters and Partitions can also be extracted from Hierarchies by the command Hierarchy/Extract Cluster or /Make Partition.

Defining and moving matrilineages in a P-graph is obtained by changing the Options/ReadWrite Threshold to 1 (be careful to set it back to the default = 0 when done because it is otherwise reset for the next run) and saving a separate file with the female network after moving elements. Reconstituting the new configuration involves the multiple network option Nets/First Network and Nets/Second Network, entering the file containing the female network in the first case and the entire network in the second case.

Defining patrilineages also involves the multiple network option Nets/First Network and Nets/Second Network, again entering the saved file containing the female network in the second case and the entire network in the first case. Nets/Difference will then subtract the arcs in the second case from the first case, and the result will be the male lines, which can be saved to a file after points are moved about.

In using separate files for patrilines versus matrilines, the option Partition/Components/ Weak will identify separate lineages that can be moved as blocks (clicking near, but not on, a point of chosen color) on the graphic screen and can then be saved. To recombine the two images, the multiple network option Nets/First Network and Nets/Second Network is used again, entering the separate saved files in the order wanted so that the first file will define the graphic configuration of the new image.

In the process of graphic analysis, a number of partitions may be defined and saved as *.CLU files—patrilineage and matrilineage memberships, connected components and relinked bicomponents, other data entered manually (e.g., a partition of places of residence). Because these are simply column variables in ASCII files, they are easily read or pasted into standard statistical packages such as SAS, SPSS, and SYSTAT for cross-tabulation and statistical analysis.

Graphic Aesthetics

Aesthetic criteria are used to compute some properties of large but well-organized pictures and to draw graphics accordingly:

- closest vertices (vertices should not be too close);
- smallest angle between lines with a common vertex (should not be too small);
- shortest/longest line (lines should not be too short or too long);
- number of crossings (the fewer, the better); and
- vertex closeness to line (vertices should not be too close to lines).

Using measures such as these, provided by Pajek, the user can check for the quality of a picture before and after reorganization. Making a partition (by which vertices are colored in the graph using the Draw/Draw Partition option) may help to organize the graph. To see a lineage as a partition of vertices, for example, Net/K-Neighbors/Input/[vertex number:*input ancestral couple number*]/0 will color all the descendants of the selected vertex, and Partition/Binarize/[range of distances from ancestor:*input maximum distance*] will give the same color to all the descendants up to the selected distance. Then, clicking not on but rather near a vertex of the color of the lineage, the mouse will move all the vertices of this color, in tandem, to a new location on the screen. Hence, the simpler options such as Draw/Layers/Optimize layers in x direction do not limit the types of visual reorganization of the data on the screen that are possible.

Figure 5 represents a picture of a 2D graph made by partitioning vertices of the genealogy according to their neighborhood (relative to vertices with very high degree, a type of leader partition). This gives a general layout in clusters of “similar” vertices that can be further

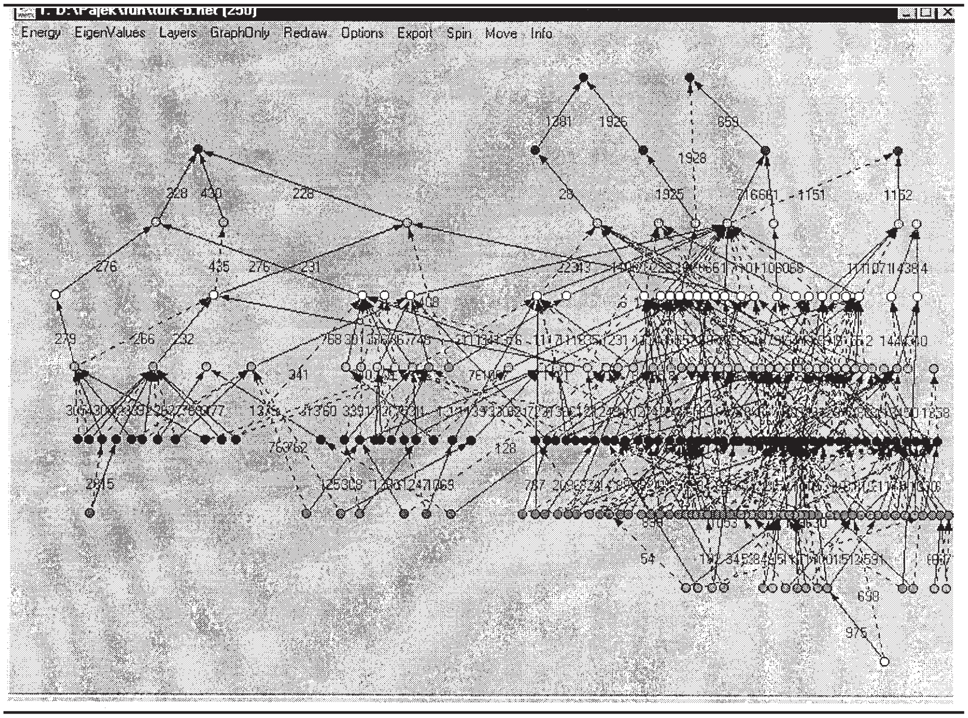


Figure 5: Optimized Drawing of the Bicomponent of Intermarrying Patrilineages in a Turkish Nomad Clan/Society

improved by manual editing. The image contains 416 arcs and 250 vertices with 14 apical ancestors, for an extremely high relinking index of 75%, but the network nonetheless has a marriage porosity (i.e., openness to the outside) of 25%. In spite of the density of the graph, many of the numeric labels on the uppermost arcs are readable, as are those in the most recent (lowermost) generations where marriages have only just begun. The image conveys many of the structural features of the Turkish nomad social networks given by White and Johansen (1998). There are few early ancestors (identified by numbers for brother pairs 1926-1381 and 228-430 and by brother-sister pairs 1152-1151 and 659-1928), and their descendants often relink through marriage not in one generation but rather within two or three. Marriages tend to be very close and within the same generation; they crosscut the patrilines, although many are within the patrilineages themselves, knitting them together from within.

Network Analysis

Figure 5 tells only part of the story of the 250 biconnected Turkish nomad couples. Figure 6 shows the entire genealogy for this society, with 1,114 arcs connecting more than 1,600 persons in 946 couples and 23 apical ancestors. The difference in the much lower relinking density of 19% is structural; those who do not relink are those who do not survive to marry or who emigrate from the clan. Pajek provides further information, using options for Net/Components/Weak and Net/Components/Bicomponents, that tell that there is a single connected component (everyone is a relative) and that all are connected (as children, affines, or ancestors) to the 25% of the couples who form the single large bicomponent.

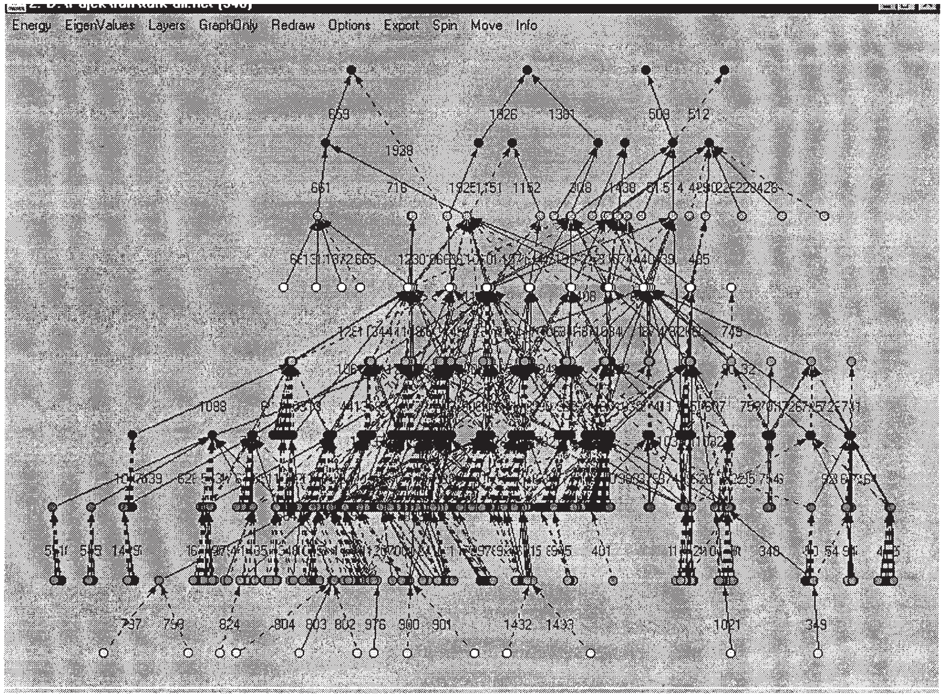


Figure 6: Optimized Drawing of the Entire Turkish Nomad Clan/Society

Pajek's main menu contains algorithms that provide solutions to some of the analytic problems of kinship analysis. Recall that a bicomponent (with three or more vertices), as noted previously, is a maximal set of vertices of which every pair is connected by two independent paths. Hence, every pair of vertices in a bicomponent is connected by marriage circles formed by relinking marriages. The bicomponents of a P-graph of kinship and marriage are mutually exclusive sets of arcs or bounded subgraphs or social units with structural endogamy (White, 1997; Brudner & White, 1997). To extract cohesive sets of marriages, after using Net/Components/Bicomponents, we double click the Hierarchy object window to view a list of components and their sizes[22] and then use Hierarchy/Extract/Cluster to extract by its number one (e.g., largest) clusters. Operations/Extract from Network/Clusters (the number of which already is selected) restricts the network (as in Figure 5) to only the bicomponent. Along the way, the menu Info/ (e.g., /Hierarchy) lets us see the distributions by size pertinent to the various objects we have computed. Returning to Draw/Draw Partition, the graph now shows only the chosen bicomponent.

Attributes of vertices can be constructed within Pajek or be imported as partition (*.CLU) files and similarly for permutations (*.PER), clusters (*.CLS), hierarchies (*.HIE), and vectors (*.VEC). One way in which to select a subgraph is to use Operations/Extract from Network/Partition (or /Cluster), which asks for criteria for selection. If the same or another partition is wanted to label or analyze the subgraph, then the options to use are Partitions/First partition (choosing some partition to be reduced for use with the subgraph) and Partitions/Second partition (criterion partition or the partition used to select the subgraph), followed by Partitions/Extract Second from First. The result is that the reduced graph and the reduced first partition are of the same dimension and can be used together for further analysis.

Pajek, as partly seen from the macros in Table 6, offers many other useful options such as

- searching for the shortest kinship paths among persons;
- determining all predecessors and successors of selected persons; and
- extracting the neighborhood of a selected person.

Options we have not discussed that are of somewhat greater complexity, as in our discussion of problems for kinship analysis (e.g., groups, roles, relations), include

- searching for interesting patterns in a genealogy;
- marriages among relatives, where and how often they occur (if a patterned type of subgraph does not occur very often in the genealogy, then this option is executed quite quickly even for large genealogies);
- parents having many children;
- persons married several times; and
- statistics—average number of children, maximum number of children.

The ability to output UCInet (Borgatti, Everett, & Freeman, 1998) *.DL files for further network analysis[23] lets us analyze features such as

- centralities of individuals and couples;
- block modeling of relational patterns in the network; and
- scalings of clusters or relatives or families.

Finally, Pajek is in no way limited simply to kinship and marriage networks or to P-graphs or family network representations; rather, it accommodates all sorts of social network and attribute data. There also is a P-graph format for a bipartite graph including both couples and individuals so that we can study relationships between individuals, between couples or families, or both at the same time. Figure 7 shows a bipartite graph of the data in Figure 4, with two types of vertices: squares to represent individuals and circles to represent their marriages (the graph is drawn to show the two slanting patrilines crosscut by three opposite-slanting matrilines). This type of graph may be useful where many individuals have multiple spouses (to separate such individuals from their siblings) and to depict distinct network relations between individuals as opposed to those between couples.

SPECIALIZED PGRAPH PROGRAMS (PARENTÉ SUITE)

The Parenté Suite of programs (White & Skyhorse, in press) includes the Pgraph program for drawing P-graphs, and the Ego2Cpl program makes the files needed by Pgraph or Pajek from a raw *.TXT file such as the one shown in Table 1. As noted previously, Ego2Cpl creates a *.GED file containing the ID numbers of individuals and assigns a new set of nuclear family (FAMS and FAMC) numbers to which individuals belong as either parents (FAMS) or children (FAMC). It also produces a series of compact files for Pgraph data analysis read by various Parenté Suite programs—the VEC or vector file (extensions *.VE*), DOC or documentation file (*.DO*), the NAM or names file (*.NA*), and (in some cases) additional coordinates (*.CO*) files. The last letter of the extension of Pgraph files is used by convention to distinguish raw data (D) files from subset (S) and bicomponent (C) subset files as well as from reduced (B) files of blood marriages or kinship model (M) files.

The VEC file lists the two vectors needed for Pgraph analysis. These arrays are indexed to the new series of unique family numbers, as explained by White and Jorion (1992). The first

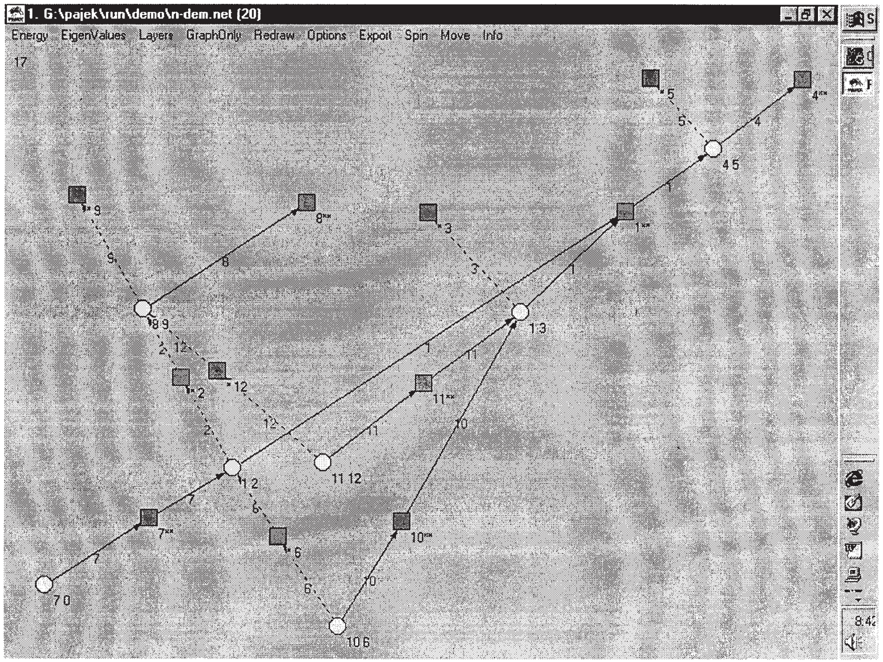

1. G:\pajek\run\demo\n-dem.net (20)
Energy EigenValues Layers GraphOnly Redraw Options Export Spin Move Info
17
5
4
4 5
1
9
8
3
8 9
13
12
2
11
10
1 2
11 12
7
6
7 0
10 6
8:42

Figure 7: P-graph Combining Individual and Family Vertices for Sample Data

array gives the family of orientation (FAMC) of a son who may be or may become a husband (FAMS), and the second array gives the family of orientation (FAMC) of a daughter who may be or may become a wife (FAMS). Additional arrays are just for purposes of keeping track of alternative family number and individual labels.

Par-Calc is a Pgraph program that reads the VEC file and does statistical analysis of the different types of blood marriages, compared to numbers of existing relatives of each type and ranging from siblings to fifth cousins in ego's generation as well as other types of relatives. The advantage of this procedure is that by computing as a baseline the number of times relatives of a given type exist (e.g., the number of men who have a mother's brother's daughter available to marry compared to those who actually marry such a relative), the rates of blood marriage are normed to control for demographic variables such as sibship size. The program also gives the percentage consistency of the observed blood marriages with dual matrimonial organization, statistics on generation differences between husbands and wives in blood marriages, and relative numbers of male and female links in such marriages.

Par-Link performs a similar statistical analysis on pairs of marriages that relink two families such as sister exchange, cousin exchange, and two brothers marrying two sisters. Unlike Par-Calc, however, it currently requires considerable sophistication in marriage alliance theory and involves a complex set of notational conventions to read the output.

Prior to the advent of Pajek, Pgraph was the principal program for network visualization within the Parenté Suite. Many of the Pgraph algorithms can now be found in Pajek and need not be separately described (White & Skyhorse, in press, provide a manual for program options). One of the continuing advantages of this program, however, is its options for creating screen images and saving coordinate files that can be converted into high-quality graphics using the HPGL language for printer control. Extremely large genealogies can be split into separate genealogies for each of the lineages following a chosen line of descent and the entire set of genealogical connections represented by lines of intermarriage that occur on the same page, together with identification numbers for the parents-in-law of spouses whose families of origin can be found on a separate page. White and Johansen (1998) use this method to print the entire labeled genealogy of the 250 relinked marriages shown in Figure 5 for the clan of nomadic Turkish pastoralists. Another feature of Pgraph is the capability to write to different types of output files—DOT graphics and UCI link lists.

Pgraph has two additional features that have not, as of this writing, been incorporated into the repertoire of Pajek algorithms. One is the option, noted earlier, to organize genealogical data by lineages. A second feature is an algorithm for reorder lineages to maximize any existing tendency for marriages to occur between opposing sides of a marriage network, as in systems of dual matrimonial organization (Houseman & White, 1998, in press). Par-Side is the accompanying auxiliary program for evaluating sidedness by calculating the probability of observed tendencies toward dual organization (bipartiteness of the graph) given the null hypothesis. Doreian and Mrvar (1996a,1996b) also provide an algorithm for finding the best approximation to bipartiteness in a graph.

The most important and original of Pgraph's algorithms is a random simulation option that keeps sibling sets intact but, within each generation, reassigns each sibling to a randomly chosen marriage partner of the same generation. By saving several sets of randomized data, any observed marriage regime can be compared to random baseline data, a task that is done by the Par-Bloc (White & Skyhorse, in press) auxiliary program.

CONCLUSION

Possibilities for the study of kinship and marriage networks are radically transformed by the evolution of software for large network analysis both in contemporary sociological contexts (Brudner & White, 1997) and in traditional anthropological contexts (Schweizer & White, 1998). Ours is not the first team of anthropologists and graph theorists (see also Foster & Seidman, 1979, 1989; Hage & Harary, 1983, 1991, 1996; Seidman & Foster, 1978) to struggle with large-scale anthropological network analysis, but our particular trio, including computer scientists, has worked to create program packages both to analyze and to visualize networked relationships. On the computer science side, we address the problems of

- fast algorithms for large networks;
- combining different types of objects of study;
- incorporating the fundaments of graph theoretic analysis; and
- aesthetics of 2D and 3D graph drawing and organizing visual displays.

"When principles of design replicate principles of thought, the act of arranging information becomes an act of insight" (Tufte, 1997, p. 9; see also Tufte, 1987, 1990).

Pgraph as a program package and P-graphs as a system of representation interface with Pajek, the program for large network analysis, to produce capabilities for solving theoretical problems in the analysis of kinship and marriage networks such as the five discussed here. These involve

- boundedness;
- cohesion;
- large-scale social systems;
- network characteristics of social relations; and
- identification of groups and roles.

Our approaches to these problems are strongly linked to graph theoretic findings:

- biconnected components of graphs as large-scale bounded subgraphs;
- the relinking index as an appropriate measure of cohesive density; and
- problems of studying cohesion on a large scale.

Granovetter's (1973) work on the strength of weak ties in very large networks for the flow of information, under certain social conditions, pointed the way to new understandings. Drawing on earlier work in graph theory, critical thresholds in large networks (Bollobás, 1985; Palmer, 1985; Watts & Strogatz, 1998) are now understood to be central to the emergence of self-organizing complexity (Kauffman, 1995).

Kinship and marriage have long been looked at "close up" in terms of the solidarity of close ties and the formation of corporate-like kinship groups. Alliance theories of marriage opened up new theoretical paradigms, from which sprang the current representation of P-graphs. Studies of kinship and marriage network on a large scale, as a necessary background and foundation for the study of other social processes, may prove to have large-scale properties that will change our conception of how societies, market systems, and major social institutions are organized. The emergence of social class, elite formation, succession to office and political domination, ownership of corporations, market exchange, and processes of inheritance and bequest need to be restudied and reconceptualized in a large-network context in which kinship and marriage network play a critical role as part of critical threshold emergent phenomena. The research tools discussed here can play a role in this development.

APPENDIX
Additional Information on Sources

Following is a list of URLs for Web sites at which software and documentation reviewed or mentioned in this article are available in their most recent versions (in alphabetical order):

Chime (Web browser plug-in by MDL Information Systems, featuring chemical structure displays): http://www.mdli.com/wel.html

Cosmo Player (Web browser plug-in from Cosmo Software, Platinum Technology, featuring VRML displays): http://cosmosoftware.com

Ego2Cpl (part of the Pgraph package for parentage networks analysis by Douglas R. White at the University of California, Irvine): http://eclectic.ss.uci.edu/~drwhite/ P-graph/ego2cpl.html. Documentation is available at links that begin at http://eclectic. ss.uci.edu/pgraph.... An introduction to P-graph analysis is found at http://eclectic. ss.uci.edu/knhe/str-endo.htm.

GIM (Genealogical Information Manager by D. Blaine Wasden and Brian C. Madsen, using GEDCOM format): http://www.mindspring.com/~dblaine/gimhome.html. For documentation on the GEDCOM standard by the LDS Church, see http://www.gendex.com/ gedcom55/55gcint.htm.

KineMage (protein display and viewing language developed by David Richardson). For introductions, see http://www.elsevier.com/locate/son and articles by Linton C. Freeman and others at http://tarski.ss.uci.edu/new.html and http://eclectic.ss. uci.edu/~lin/chem.html as well as a dynamic sample page for kinship images at http://eclectic.ss.uci.edu/~drwhite/pgraph/mage.html.

Kith and Kin (commercial software by SpanSoft for P-graph-style genealogies, available on a trial basis): http://www.rocketdownload.com/details/home/kithkin.htm

Pajek (program for large networks analysis by Vladimir Batagelj and Andrej Mrvar at the University of Ljubljana): http://vlado.fmf.uni-lj.si/pub/networks/pajek. See http://vlado.

fmf.uni-lj.si/pub/networks/pajek/pajekman.htm for the manual. An introduction to drawing genealogies with Pajek is found at http://vlado.fmf.uni-lj.si/pub/networks/doc/sitges.pdf. For a description of the program, see http://vlado.fmf.uni-lj.si/pub/networks/pajek/pajek.pdf or http:// vlado.fmf.uni-lj.si/pub/networks/pajek/sunbelt.97/pajek.htm.

UCINet (networks analysis package for sale from Analytic Technologies): http://eclectic.ss.uci.edu/~lin/order.html

NOTES

[1] The GEDCOM 5.5 standard is accessible at http://www.gendex.com/gedcom55/55gcint.htm. It is, of course, possible to enter data directly into a GEDCOM database such as Genealogical Information Manager, available at http://www.mindspring.com/~dblaine/gimhome.html, that assigns both individual and family numbers automatically. Pajek can read GEDCOM files directly.

[2] Hence, one of their two prior FAMS numbers, if any, can be reassigned to their conjoint unit, and the other can be reserved in case the corresponding individual marries again. Note that the FAMC number of a child equals the FAMS of the child’s parents.

[3] Kith and Kin (currently in Version 3.02), which runs in Windows 95 and which prints descendant and ancestral trees in P-graph format, is authored by SpanSoft and is available on a trial basis at http://www.rocketdownload.com/details/home/kithkin.htm. Pictures, maps, diagrams, and sounds may be embedded in or linked to a person or family. A printed book can be produced with family group details and a cross-referenced index.

[4] The Ego2Cpl program, as part of the Pgraph package, is available at http://eclectic.ss.uci.edu/~drwhite/p-graph/ego2cpl.html. It is a DOS program that is compatible with Windows, but a straight DOS version also is available at the same location.

[5] Ego2Cpl allows a choice of eight formats, depending on the order and type of variables: (1) ego number, name, sex, father number, mother number, spouse number; (2) ego number, sex, name, father number, mother number, spouse; (3) ego number, name, sex, spouse number, father number, mother number, decade of birth; (4) ego number, sex, spouse number, father number, mother number, name, other data; (5) ego number, sex, spouse number, father number, mother number, name, other data; (6) ego number, sex, spouse number, father number, mother number, premarital residence, decade of birth, postmarital residence, name, other data; (7) ego number, sex, spouse number, father number, mother number premarital residence, decade of birth, postmarital residence, purchase, name, other data; (8) ego number, sex, spouse number, father number, mother number, name, number, decade of birth, premarital residence, postmarital residence. Formats 4 to 7 are flexible in having the name and other data in last place in the data format so that new variables for data entry can be extended indefinitely. Formats 6 to 8 exemplify special cases in which certain types of data are coded (e.g., place of residence, pre- and postmarriage).

[6] It does not check beyond parents, however, for errors where egos are their own ancestors. This is done by Pajek and by the Par-Calc program in the Pgraph suite of programs.

[7] In a P-graph, multiply married persons need a common parental root to link their multiple marriage vertices.

[8] Which sex is assigned solid versus dotted (or other types of) arcs, and the coloring of arcs is a matter of convention and may vary from study to study depending on what the researcher wishes to emphasize. The default for both Pgraph and Pajek is to assign solid arcs to male links and dotted arcs to female links, which is the reverse of Figure 1.

[9] The connectivity of a regular genealogy (with a maximum of two parents for each individual) can only be 0 (several weak components), 1 (weakly connected), or 2 (2-connected).

[10] In egocentric representations, there are several types of structural ambiguities such as how to represent marriage, whether the existence of children necessarily implies the presence of a marriage or pseudo-marriage link, and whether comparable definitions of *marriage* in different societies are being used, definitions that can have artifactual effects on the analysis. In the P-graph, these problems do not occur because they do not affect the structure of the graph; rather, they merely affect what types of cultural labels are placed on the vertices where marriages, concubinage, informal sexual liaisons, and the like may occur.

[11] Development of the Pgraph program package was supported by a 1993-1995 National Science Foundation grant (“Network Analysis of Kinship, Social Transmission, and Exchange: Cooperative Research at UCI, UNI Cologne, and CNRS Paris”), the Alexander von Humboldt Foundation TransCoop program, and the Maison des Science de l’Homme-Paris (including the Maison Suger).

[12] Ego2Cpl allows as labels either individual names or numbers.

13. There also are ways in which to do this in the GEDCOM format, but it is difficult to find software to handle the extra data graphically.

14. Pajek is available, always in its most recent version, as a download from the Internet address `http://vlado.fmf.uni-lj.si/pub/networks/pajek` at the University of Ljubljana.

15. Available at `http://vlado.fmf.uni-lj.si/pub/networks/pajek/pajekman.htm`.

16. Draw/Spin/Normal and type 001 and Draw/Spin/Spin around and type 180 also will spin 180° in the *z* direction. A 180° spin is not required if the arcs are transposed to parent-to-child.

17. Available at `ftp://ftp.cs.wisc.edu/pub/ghost/`, `ftp://ftp.cs.wisc.edu/pub/ghost/rjl/`, and `ftp://ftp.cs.wisc.edu/pub/ghost/aladdin/`.

18. Available at `http://cosmosoftware.com`.

19. Available at `http://www.mdli.com/wel.html`.

20. Pajek provides an option to include labels in Mage images, but this might produce a crowded image. The other way in which to get fixed but selected labels is within the Mage program, using option Edit/Draw new/. On the right-hand side where options appear, Labels must be checked so that vertices subsequently may be clicked for a label to appear on the layout.

21. Embedded commands are listed at `http://vlado.fmf.uni-lj.si/pub/networks/draweps.htm`.

22. From this list of the hierarchy of bicomponents, clicking on the largest bicomponent and choosing Edit/Change type will toggle it to "(Close)". Then on the main menu, Operations/Shrink network/Hierarchy, selecting "Yes" for checking whether the network is simple and "1" for the minimum number of connections will result in the "Closed" group (the largest bicomponent) being shrunk to a single node. Net/Transform/Remove loops will remove the loops produced by the arcs inside the bicomponent. After that, the shrunken network can be treated as an ordinary genealogy, with the shrunken vertex having vertex number 1. Net/K-Neighbors/All with vertex label 1 and distance 0 (no limit) will sort the vertices by their distances from the largest bicomponent.

23. UCINet is for sale at `http://eclectic.ss.uci.edu/~lin/order.html`.

REFERENCES

Batagelj, V., & Mrvar, A. (1997, February). *Pajek: Program package for large networks analysis*. Paper presented at Sunbelt XVII conference, San Diego. [Online]. Available: `http://vlado.fmf.uni-lj.si/pub/networks/pajek/sunbelt.97/pajek.htm` and `http://vlado.fmf.uni-lj.si/pub/networks/pajek/pajek.pdf`

Batagelj, V., & Mrvar, A. (1998). *Pajek—A program for large network analysis*. [Online]. Available: `http://vlado.fmf.unk-lj.si/pub/networks/pajek/`

Bollobás, B. (1985). *Random graphs*. New York: Academic Press.

Borgatti, S. P., Everett, M. G., & Freeman, L. C. (1998). *UCINET 5.0 for Windows 95/NT*. Nantick, MA: Analytic Technologies.

Brudner, L. A., & White, D. R. (1997). Class, property, and structural endogamy: Visualizing networked histories. *Theory and Society*, *25*, 161-208.

Doreian, P., & Mrvar, A. (1996a). A partitioning approach to structural balance. *Social Networks*, *18*, 149-168.

Doreian, P., & Mrvar, A. (1996b). Structural balance and partitioning signed graphs. In A. Ferligoj & A. Kramberger (Eds.), *Developments in data analysis* (Metodoloski zvezki 12, pp. 195-208). Ljubljana, Slovenia: FDV.

Foster, B. L., & Seidman, S. B. (1979). Network structure and the kinship perspective. *American Ethnologist*, *8*, 329-355.

Foster, B. L., & Seidman, S. B. (1989). A formal unification of anthropological kinship and social network methods. In L. C. Freeman, D. R. White, & A. K. Romney (Eds.), *Research methods in social network analysis* (pp. 41-59). Fairfax, VA: George Mason University Press.

Freeman, L. C. (1998a). Using available graph theoretic or molecular programs in social network analysis. [Online]. Available: `http://tarski.ss.uci.edu/new.html`

Freeman, L. C. (1998b). *Using molecular modeling software in social networks: A practicum*. [Online]. Available: `http://eclectic.ss.uci.edu/~lin/chem.html`

Freeman, L. C., Webster, C. M., & Kirke, D. M. (1998). Exploring social structure using dynamic three-dimensional color images. *Social Networks*, *20*, 109-118. [Online]. Available: `http://www.elsevier.com/locate/son`

Granovetter, M. (1973). The strength of weak ties. *American Journal of Sociology*, *78*, 1360-1380.

Hage, P., & Harary, F. (1983). *Structural models in anthropology.* Cambridge, UK: Cambridge University Press.

Hage, P., & Harary, F. (1991). *Exchange in Oceania: A graph theoretic analysis*. Oxford, UK: Oxford University Press.

Hage, P., & Harary, F. (1996). *Island networks*. Cambridge, UK: Cambridge University Press.

Harary, F. (1969). *Graph theory*. Reading, MA: Addison-Wesley.

Harary, F. (1971). Demiarcs: An atomistic appoach to relational systems and group dynamics. *Journal of Mathematical Sociology*, *1*, 195-205.

Héran, F. (1995). *Figures et Légendes de la Parenté*. Paris: Institut national d'etudes démographiques.

Houseman, M., & White, D. R. (1998a). Network mediation of exchange structures: Ambilateral sidedness and property flows in Pul Eliya. In T. Schweizer & D. R. White (Eds.), *Kinship, networks, and exchange* (pp. 59-85). Cambridge, UK: Cambridge University Press.

Houseman, M., & White, D. R. (1998b). Taking sides: Marriage networks and Dravidian kinship in lowland South America. In M. Godelier, T. Trautmann, & F. E. Tjon Sie Fat (Eds.), *Transformations of kinship* (pp. 214-243). Washington, DC: Smithsonian Institution Press.

Kauffman, S. (1995). *At home in the universe: The search for laws of self-organization and complexity*. Oxford, UK: Oxford University Press.

Mrvar, A., & Batagelj, V. (1998, May). *Drawing genealogies*. Paper presented at Sunbelt XVIII conference, Sitges, Spain. [Online]. Available: `http://vlado.fmf.uni-lj.si/pub/networks/doc/sitges.pdf`

Murdock, G. P. (1949). *Social structure*. New York: MacMillan.

Ore, O. (1963). *Graphs and their uses*. New York: Random House.

Palmer, E. M. (1985). *Graphical evolution*. New York: John Wiley.

Schweizer, T., & White, D. R. (Eds.). (1998). *Kinship, networks, and exchange*. Cambridge, UK: Cambridge University Press.

Seidman, S. B., & Foster, B. L. (1978). A note on the potential for genuine cross-fertilization between anthropology and mathematics. *Social Networks*, *1*, 65-72.

Tufte, E. (1987). *The visual display of quantitative information*. Cheshire, CT: Graphics Press.

Tufte, E. (1990). *Envisioning information*. Cheshire, CT: Graphics Press.

Tufte, E. (1997). *Visual explanations*. Cheshire, CT: Graphics Press.

Watts, D. J., & Strogatz, S. H. (1998). Collective dynamics of "small-world" networks. *Nature*, *393*, 440-442.

White, D. R. (1996). Enfoque de redes al estudio de comunidades urbanas. *Estudios Demográficas y Urbanas*, *26*, 303-326.

White, D. R. (1997). Structural endogamy and the graphe de parenté. *Informatique, Mathématique et Sciences Humaines*, *137*, 107-125. [Online]. Available: `http://eclectic.ss.uci.edu/knhe/str-endo.htm`

White, D. R., & Harary, F. (1997). *Structural models and conditional densities for kinship and neighborhood*. Unpublished manuscript, University of California, Irvine.

White, D. R., & Harary, F. (1998). *Cohesiveness measurements for social groups*. Unpublished manuscript, University of California, Irvine.

White, D. R., & Johansen, U. (1998). *Social anatomy of a nomadic clan: An anthropological introduction to networked histories*. Unpublished manuscript, Institute of Ethnology, University of Cologne, Germany.

White, D. R., & Jorion, P. (1992). Representing and analyzing kinship: A network approach. *Current Anthropology*, *33*, 454-462.

White, D. R., & Jorion, P. (1996). Kinship networks and discrete structure theory: Applications and implications. *Social Networks*, *18*, 267-314.

White, D. R., Schnegg, M., & Brudner, L. A. (in press). Multiple connectivity, bounded cohesion and cross-cutting integration: Kinship and compadrazgo in rural Tlaxcala. In J. Gil & S. Schmidt (Eds.), *Social networks: Theory and applications*. Mexico, D. F.: UNAM Press.

White, D. R., and Skyhorse, P. (in press). Parenté Suite: User's manual for analysis of kinship and marriage networks. In V. Burton, T. Finnigan, & D. Herr (Eds.), *Multimedia Renaissance in social science computing* [CD-ROM]. Urbana: University of Illinois. [Online]. Available: `http://eclectic.ss.uci.edu/~drwhite/pgraph/p-graphs.html`

Douglas R. White is a member of the Social Networks graduate program, the Institute of Mathematical Behavioral Science, and the Department of Anthropology at the University of California, Irvine, CA 92697; e-mail: `drwhite@uci.edu`*. His current work is on complexity theory and the effects of emergent structure in social networks on a large scale including social class, large-scale cohesion, solidarity and exchange, elites, markets, and global processes. He does extensive collaboration on longitudinal social research projects in Europe, Latin America, Africa, and Asia.*

Vladimir Batagelj is a professor of discrete and computational mathematics at the University of Ljubljana, Slovenia. His main research interests are in mathematics and computer science, and combinatorics with emphases on graph theory, algorithms on graphs and networks, combinatorial optimization, algorithms and data structures, cluster analysis, and applications of information technology in education. He may be contacted at `vladimir.batagelj@uni-lj.si`.

Andrej Mrvar is a teaching assistant of computer science and statistics with the Faculty of Social Sciences at the University of Ljubljana, Slovenia, where he is completing his doctoral thesis on the analysis and visualization of large networks with the Faculty of Computer and Information Science. He may be contacted at the Faculty of Social Sciences, Kardeljeva pl. 5, 1000 Ljubljana, Slovenia; e-mail: `andrej.mrvar@uni-lj.si`; *Web:* `http://www.uni-lj.si/~fdmrvar/andrej.html`.