

# Kinship Network Analysis

Klaus Hamberger, Michael Houseman, and  
Douglas R. White

## KINSHIP IN A NETWORK PERSPECTIVE

Kinship, like language, is a structure, not a substance.<sup>1</sup> The distinctive features of kinship networks reside less in how their constitutive ties – be they biological, jural, ritual, symbolic, or whatever – are defined and established than in the way these ties are organized. Kinship network theory is thus not just another “application” of general network theoretic methods to a particular social domain but a specific branch of social network theory in itself, defined by its own axioms and described by its own theorems.

Kinship networks are characterized by the interplay of three fundamental principles: filiation, marriage, and gender. We ordinarily represent *filiation* by a set of arcs (*descent arcs*) that are directed from parents to children, and *marriage* by a set of undirected edges (*marriage edges*) between spouses (for alternative representations of kinship networks without edges, see 2.2 below). Kinship networks thus are mixed graphs, containing both arcs and edges. *Gender* is usually taken into account by a partitioning of the vertex set (the *gender partition*), usually into two or three disjoint classes (male, female, and possibly unknown sex).

The characteristic features of kinship networks can be described in terms of cyclicity. While kinship networks do not contain *oriented cycles* (nobody can be his or her own descendant<sup>2</sup>), they may contain *cycles* (where arc direction does not matter): people may marry persons with whom they are already linked by kinship or affinity. Now, such cyclic configurations occur not just randomly but in ways that are informative about the self-organizing behavior of the network. Some kinds of relatives hardly ever marry, while other

kinds of kinship ties between spouses may be overrepresented. Marriage rules and prohibitions, but also residential organization, social morphology, and so forth, affect the relative frequencies of different types of cycles in a kinship network. Analyzing the distribution of cycles therefore is the key to kinship network classification and interpretation.

## *Paths and cycles in kinship networks*

Kinship network theory thus rests on a theory of cyclic configurations. We call a *path* an alternating sequence of vertices and lines (edges or arcs of whatever direction), where every vertex is incident with the lines that precede and follow it in the sequence, and all vertices are distinct. If, by contrast, the first and the last vertex are identical (all others being distinct), we obtain a *cycle*. A path is said to be *closed* by a line connecting its first and its last vertex, so that adding that line turns the path into a cycle. A path or cycle is called *oriented* if all lines are arcs oriented in the same direction.<sup>3</sup> A *weakly acyclic* network is one that contains no oriented cycles and an *acyclic* network contains no cycles whatsoever.

Alternatively to their definition as (open or closed) *sequences* of vertices and lines, paths and cycles are also often defined as the *graphs* made up of these vertices and lines (in this perspective, a cycle is a connected graph where every vertex has degree 2, a path a connected graph where two vertices have degree 1, and all others degree 2). There is, however, a crucial difference between the two concepts. If we define a path as a sequence, the starting point matters. A path ABC

is distinguished from its inverse CBA, or to take a kinship example, the path “father’s wife’s daughter” (FWD) is distinguished from its inverse: “mother’s husband’s son” (MHS). If, by contrast, we define a path as a graph, ABC and CBA are only two different notations for one and the same mathematical object – for the kinship case, this means that FWD and MHS are two different ways to express the same kinship chain.

The ambiguity is less severe in the case of cycles, where, by convention, graph theorists treat starting and ending points as irrelevant: the sequences ABCA and BACB are considered to be identical. Nevertheless, in kinship network analysis we often need to distinguish them. In an ego-centric perspective, marrying one’s “father’s wife’s daughter” (FWD) is different from marrying one’s “son’s wife’s mother” (SWM). By contrast, in a socio-centric perspective – as it is required if, for example, we want to count all cycles of a given type in a kinship network – we have to treat FWD and SWM marriages as two different aspects of one and the same configuration. In other words, we should define paths and cycles as *sequences* when adopting an *ego-centric* view and as *graphs* when adopting a *socio-centric* view. In order to avoid any ambiguity, we shall reserve the terms “path” and “cycle” to sequences and use the terms “chain” and “circuit” when we speak of the corresponding graphs.<sup>4</sup> A *chain* is thus a graph made up by the vertices and lines of a single path, and a *circuit* is a graph made up by the vertices and lines of a single cycle (alternative definitions in terms of degrees and connectedness are given below).

A *path* is an alternating sequence of vertices and lines (arcs or edges), where each vertex is incident to the preceding and the succeeding line, the vertices preceding and succeeding a line are its endpoints, and all vertices are distinct.

A *cycle* is a sequence of vertices and lines sharing the properties of a path, except that the first and the last vertex are identical (all other vertices being distinct).

A path or cycle is called *oriented* if all its lines are arcs oriented in the same direction.

A *weakly acyclic* graph is one that contains no oriented cycles and an *acyclic* graph contains no cycles whatsoever.

A *chain* is a graph whose vertices and arcs form a single path (alternatively, it may be defined as a connected graph where two vertices have degree 1 and all others have degree 2).

A *circuit* is a graph whose vertices and arcs form a single cycle (alternatively, it may be defined as a connected graph where every vertex has degree 2).

The first and the last vertices of a path are called *endpoints*. Any line connecting the endpoints of a path is said to *close* it; the line’s addition to the path transforms the latter into a cycle.

Armed with these basic concepts, we can now describe kinship networks entirely in structural terms, without making statements regarding the nature of the relations involved. Characterizing kinship networks by weakly acyclic filiation, acyclic gendered descent, nonoccurrence of certain types of circuits, and so on, is not equivalent to stating that no one can be engendered by his or her own offspring, that it takes a man and a woman to produce children, or that self-organization is brought about by incest prohibitions that prevent people in certain kinship relations from having sex with each other. Filiation does not necessarily involve a biological tie, procreation is not the only basis for parenthood, and marriage prohibitions are not defined everywhere in terms of sexual relations. Kinship network structures may have biological explanations, but these are culturally determined, variable from one society to another, and far from universal.

Biology is the privileged model for kinship inasmuch as it affords a universally intelligible code for expressing the latter’s fundamental relations: procreation as a model for filiation, sex as a model for marriage, and so forth. However, being a model for a relationship does not mean being the essence of this relationship. There are many social networks that are clearly kinship networks, both according to structural criteria and in the minds of those concerned, but where filiation, marriage, and gender cannot be defined exclusively in reference

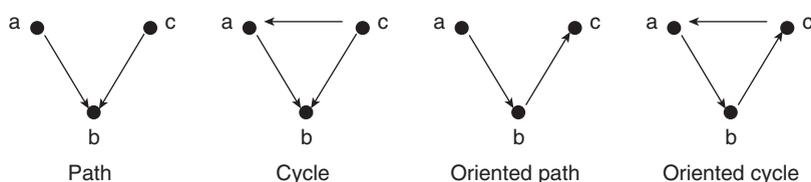


Figure 35.1

to biological givens. Kinship is a type of *social* structure, and if people all over the world have chosen to express its basic relations in a biological idiom, there are numerous cases where these relations are defined independently of, and sometimes even in opposition to, biological relationships. Defining kinship in network terms recognizes this fact.

## NETWORK REPRESENTATIONS OF KINSHIP

### *Ore-graph representation*

What we have described in our introductory statement is the most conventional representation of a kinship network, where vertices represent individuals, arcs represent filial ties, and edges represent marriages. Such networks are called *Ore-graphs*.<sup>5</sup> If not specified otherwise, kinship networks are treated in this article as Ore-graphs. Their characteristic feature is that they are weakly acyclic.<sup>6</sup> As a consequence, kinship networks contain a directed generational hierarchy.

Gender in Ore-graphs is represented by a partition of the vertex set. This gender partition gives rise to an analogous partition of arcs, according to the gender of the vertex from which they originate. The subgraph produced by arcs originating from parental vertices of the same gender can be termed a *descent graph*. Cycles in descent graphs are ruled out as soon as we impose the condition of *unique descent*, that is, filial arcs for only one parent of each gender: one father and one mother. Under this condition, descent graphs are acyclic, and their connected components are *trees* (the well-known unilinear descent trees of “lineages”). This condition is closely related to the condition of heterosexual marriage, according to which a marriage edge can only link vertices from a different gender. We speak of a *standard kinship network* if these two conditions – unique descent and heterosexual marriage – are satisfied. This allows for the possibility of nonstandard kinship networks, which, because they admit multiple descent<sup>7</sup> and homosexual marriage, require a more complex analysis. In this chapter, we will restrict ourselves to discussing standard kinship networks.

In many cases, marriage is the correlate, sometimes even the condition or equivalent of having children in common. It is therefore useful to distinguish, as a particular category of standard kinship networks, those networks that meet the condition of married co-parents: a standard kinship network will be called *canonical* if the presence of descent arcs from two parent vertices

to the same child vertex necessarily implies the existence of a marriage edge between the parent vertices.

A *kinship network* is a mixed graph  $G(V, E, A, \sim)$ , where  $V$  is a set of vertices, called *individual* vertices,  $E$  is a set of edges, called *marriage* edges,  $A$  is a set of arcs (directed from parents to children), called *descent* arcs, and  $\sim$  is an equivalence relation on  $V$  partitioning it into  $n$  disjoint classes  $V_i$  ( $i = 1, \dots, n$ ), called *genders* (usually  $n = 2$  for {male,female}), with the following property:

1 the network is weakly acyclic

The *descent graph* for the  $i$ th gender is the subgraph of a kinship network produced by all descent arcs springing from individuals of the gender  $i$ . They are called *agnatic* for male gender and *uterine* for female gender.

A kinship network is *regular* if

2 no vertex in a subgraph  $G(V, A)$  (the descent graph for the  $i$ th gender) has an indegree higher than 1 (which rules out cycles in descent graphs), where  $A_i$  is the subset of arcs with origin in  $V_i$  [unique descent]. Descent graphs of regular kinship networks are acyclic.

A regular kinship network is *standard* if

3 the subgraphs  $G(V_i, E)$  are empty [heterosexual marriage].

A standard kinship network is *canonical* if

4 any two vertices that are arc-adjacent to the same vertex<sup>8</sup> are edge-adjacent to each other [married co-parents].

Every kinship network  $G(V, E, A, \sim)$  is *partially ordered* on  $V$ , as is any weakly acyclic graph. This is the *generational* partial order relation. In addition, a kinship network may be *ordered* on  $V$  by assigning an arbitrary unique identity number to each individual (this is important for computation issues, but also for data storage in general, see below).

Two individuals linked by a descent arc are called *parent* and *child* with respect to each other. Two individuals linked by an oriented path of descent arcs are called *ascendant* and *descendant* with respect to each other. Two individuals linked by a marriage edge are called the *spouses* of each other. Two individuals are called *co-parents* if they are parents of the same child (arc-adjacent to the same vertex), *siblings* if they are children of the same parent (arc-adjacent from the same vertex), and *co-spouses* if they are spouses

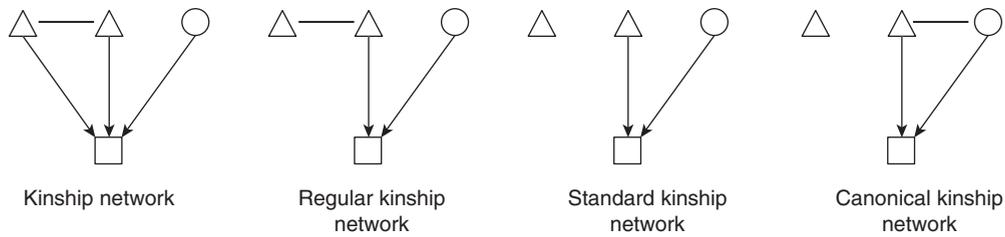


Figure 35.2

of the same spouse (edge-adjacent to the same vertex).

We can reformulate conditions (1)–(4) as follows:

- 1 In a kinship network, no individual can be his or her own ascendant or descendant.
- 2 In a regular kinship network, no individual has more than one parent of each gender.
- 3 In a standard kinship network, parents are regular and spouses are always of different gender.
- 4 In a canonical kinship network, spouses are standard and all co-parents are spouses.

### Simple digraph representations

There are a variety of ways of representing standard kinship networks as digraphs. In addition to an Ore-graph, two of these, P-graphs (White and Jorion, 1992, Harary and White 2001, also see White, this volume) and bipartite P-graphs (proposed by White, implemented by Batagelj and Mrvar, 2004), are represented in Figure 35.3 where letters indicate individuals (a married couple, wife **A** and husband **B**, who have a daughter **C** and a son **D**).<sup>9</sup> These are all isomorphic once arc-labels for P-graphs are included.

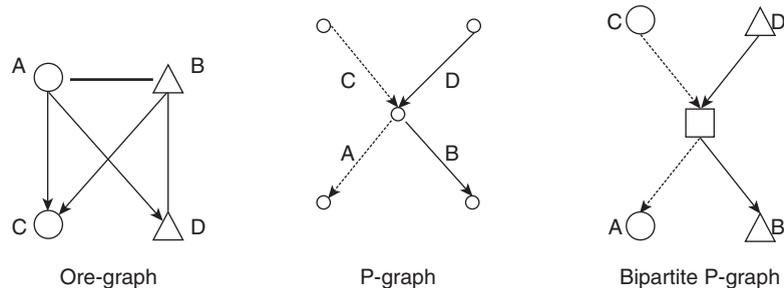


Figure 35.3

### P-graphs

*P-graphs*<sup>10</sup> represent couples as vertices and individuals as individually and gender-labeled lines. As these individuals are at once born of one couple and may become partners in another, the arcs that represent them run from the couple formed by an individual and his or her spouse to the couple formed by his or her parents. Unmarried individuals are treated like couples.

P-graphs have the advantages of incorporating fewer lines and vertices, allowing marriage cycles to be more easily detected. An individual who marries several times is represented by several lines that are numbered with individual IDs or can be given names. The way to distinguish two lines representing the same individual from those representing two same-gender full siblings is by either the line IDs or vectors for individual male or female IDs for each vertex. Apart from these two differences, P-graphs share the structural properties of Ore graphs, such as weak acyclicity and generational partial ordering.

### Bipartite P-graphs

*Bipartite P-graphs* are two-mode networks where individuals and couples are represented by

vertices. There are therefore no marriage edges, but there are arcs that run from individuals to couples and from couples to individuals.

Bipartite P-graphs are weakly acyclic and, because marriages are represented by vertices, it is possible to represent sibling relations even if the siblings' parents are unknown. In addition, marriages can be easily partitioned, for example, according to marriage dates.

## KINSHIP PATHS, KINSHIP RELATIONS, AND MATRIMONIAL CIRCUITS

### *Kinship paths and relations*

The most general definition of a kinship relation relies on the existence of a path linking one individual to another in a kinship network. If we abstract from the individual identity of vertices, kinship paths can be characterized by generic properties such as the *gender* of vertices and the *direction* of lines. This abstract form of a kinship path is an *elementary kinship relation*. Elementary kinship relations thus can be considered as abstract kinship paths (where vertex identity does not matter as long as vertex gender, line direction and vertex-line incidence are preserved). Whatever property we may state of a path without reference to individual vertices, we may consider as a property of the corresponding elementary kinship relation.

We shall call a *simple* kinship relation one that connects Ego and Alter by a single line. By contrast, we shall call a relation *compound* if it connects Ego and Alter by several consecutive lines. Compound relations can be defined by the composition of simple relations. For instance, "father" is a simple kinship relation, while "mother's father" is a compound one.

*Elementary* kinship relations are defined by a complete specification of the gender and the direction pattern of a single kinship path connecting Ego and Alter. *Complex* kinship relations are obtained by combining elementary relations by means of one or more logical operations ("and," "or," "not," etc.). If the logical connective is "or," we obtain a *disjunctive* (or "classificatory") relation (e.g., "uterine sibling" is defined as "mother's son *or* mother's daughter"). If the connective is "and," we obtain a *conjunctive* (or "multiple") relation (e.g., "full brother" is defined as "mother's son *and* father's son"). If the connective is "not," we obtain a *residual* relation (e.g., "nonagnatic kin").

Finally, we call a kinship relation *mixed* if properties of vertices and lines other than gender

and direction enter into its definition; for example, the definition of "widow" requires a supplementary partition of vertices (alive vs. dead), the definition of "elder brother" makes reference of a partial order relation defined on filial arcs, and so forth.

A *kinship path* is a path in a kinship network. The first vertex of a kinship path is called *Ego*; the last one is *Alter*.

The *direction* of a line in a kinship path is 0 ("horizontal") if it is a marriage edge, -1 ("descending") if it is an arc directed to the successor, and +1 ("ascending") if it is inversely directed.

Two kinship paths are *isomorphic* if there is a bijection between them that preserves the gender of vertices and the sequence and direction of lines.

An *elementary kinship relation* corresponds to a maximal set of isomorphic paths in a kinship network. Any of these paths *represents* the kinship relation. Any invariant property of these paths – beginning with the gender and direction sequence – can be considered as a property of the elementary relation.

A *complex kinship relation* is any relation that can be obtained by logical junction of several elementary kinship relations.

### *The notation of kinship paths and relations*

The conventional notation of kinship relations uses capital letters for indicating direction and the gender of Alter (the gender of Ego must be indicated by additional signs such as ♂ [male Ego] or ♀ [female Ego] placed before the initial letter): ascending arcs are F(ather) and M(other), descending arcs are S(on) and D(aughter), marriage edges are H(usband) and W(ife), plus supplementary letters for B(rother) and Z(Sister) relations. Examples of this are MBD (mother's brother's daughter, a matrilineal cross-cousin), ZH (sister's husband, a brother in-law), and FWS (father's wife's son, a stepbrother).

This conventional notation, a simple abbreviation of English kinship terminology, has the advantage of being easy to learn and to apply, but it is hardly the best tool for analysis. It may even obscure the structural similarities and the distinctive properties of kinship relations.

In this respect it can be contrasted with the alternative, positional notation developed by L. Barry (2004). Here, a kinship relation is represented by a sequence of letters specifying vertex labels (gender) and diacritical signs, which indicate the presence of a marriage edge (the point or full stop ".") and a change of arc direction (the parentheses "()"). All letters not separated by

a point represent vertices connected by arcs, vertices in “apical” position (that is, vertices that are not a neighbor’s children) are put into parentheses and, by convention, all arcs to the left of an apical vertex have ascending direction, all arcs to its right have descending direction, and the marriage point implies change of direction.

Consider, for example, the kinship path linking a male Ego (1) to a female Alter (5) who is Ego’s paternal sister’s husband’s daughter (Figure 35.4). In positional notation, this relation is written as 1 (2) 3 . (4) 5 where 2 is Ego’s and 3’s parents, 4 is Ego’s sister’s husband and 5 is the latter’s daughter.

By abstracting the concrete identity of the individuals concerned (represented by their number) and retaining their gender only (H for male and F for female<sup>11</sup>), one obtains the kinship relation in question: H(H)F.(H)F.<sup>12</sup> See Table 35.1 for an example of how standard and positional notations (for paths and relations) are used.

The principle of positional notation also applies to P-graphs. In P-graph notation, with labels for arcs rather than for vertices, the ZHD relation of Figure 35.4 is written HfH.hf, where H and F give the parents of a male and a female, respectively,<sup>13</sup> relation inverses  $h=H^{-1}$ ,  $f=F^{-1}$  give sons and daughters, and marriages are thus written fH and hF, respectively. The full stop “.” in H.h identifies the same individual in a different couple – a distinction that is necessary to clarify that 5 is the child of 4 but not of 3.

Positional notation has a number of important advantages. It incorporates the gender of Ego as well as that of Alter, and it provides a clear representation of the structural properties of kinship relations, which is not radically changed by symmetry transformations. Thus, for example, a man who marries his HF(H)F is his wife’s FH(F)H (in p-graph version: HFhf and FHfh), whereas in conventional notation, a man who marries his MBD is his wife’s FZS. Allowing for a

homogeneous representation of kinship paths (with individual numbers), of kinship relations (with gender letters), and of kinship relation classes (with gender variables), positional notation may be used not only as a means of notation but also as a classificatory or programming tool.

### The classification of kinship relations

In order to compare kinship relations, and to analyze the ways they combine so as to give rise to particular network structures, these relations may be classified according to different criteria. We restrict ourselves here to some basic definitions (for a more extensive treatment, see Hamberger and Daillant 2008 and Hamberger forthcoming).

A kinship relation is *linear* if it is oriented (we speak of “oriented” paths but of “linear” relations).

Any linear kinship relation can be represented by a *characteristic number*  $\lambda$

$$\lambda = \sum_{i=0}^{\kappa} (1 + \sigma_i) \cdot 2^i$$

where  $\kappa$  is the degree of the relation (see note 15) and  $\sigma_i$  is the gender number (0 = male, 1 = female) of the  $i$ th individual in ascending direction (starting with Ego  $i=0$ ), for example,  $\lambda = 1$  for male Ego,  $\lambda = 3$  for F/S,  $\lambda = 5$  for M/S,  $\lambda = 7$  for FF/SS,  $\lambda = 13$  for MM/DS, and so on.<sup>14</sup> Characteristic numbers impose an order on all linear kinship relations.

A kinship relation is *canonical* if it contains no linking children positions (that is, if the kinship path does not pass through parental triads as defined below).

A canonical kinship relation is *consanguineous* if it contains no marriage edge.

A *consanguineous component* of a kinship network is a maximal set of individuals linked to each other by consanguineous paths (an individual

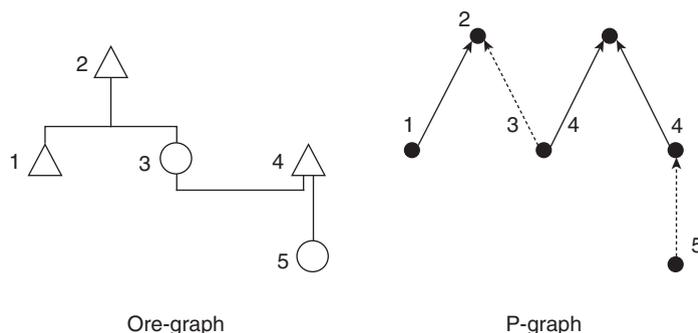


Figure 35.4

**Table 35.1 Matrimonial census**

111 marriages (2.06%) involving 211 (1.41%) individuals (105 men, 106 women) in 1,114 circuits of 37 different types (average frequency 3.08) of width 1 and height 3

ID	Standard	Positional	P-Graph	Marriages	Circuits	% Circuits
1	FBD	HH()HF	HHhf	3	3	2.63
2	FZD	HH()FF	HHff	5	5	4.39
3	FFSD	HH(H)HF	HHH.hhf	4	4	3.51
4	FFDD	HH(H)FF	HHH.hff	10	10	8.77
5	MBD	HF()HF	HFhf	9	9	7.89
6	MZD	HF()FF	HFff	2	2	1.75
7	MFSD	HF(H)HF	HFH.hhf	13	13	11.4
8	MFDD	HF(H)FF	HFH.hff	2	2	1.75

372 marriages (6.89%) involving 667 (4.47%) individuals (339 men, 328 women) in 267 circuits of 152 different types (average frequency 1.76) of width 2 and height 2

46	FWFSD	H(H).F(H)HF	HH.hfH.hhf	6	3	1.12
47	FWFDD	H(H).F(H)FF	HH.hFH.hff	2	1	0.37
48	FWFFSD	H(H).FH(H)HF	HH.hFHH.hhf	2	1	0.37
49	MHBD	H(F).H()HF	HF.fHhf	2	1	0.37
50	BW	H()H.(F)	HhF	20	10	3.75
51	BWMD	H()H.F(F)F	HhFF.ff	2	1	0.37

62 marriages (1.15%) involving 112 (0.75%) individuals (53 men, 59 women) in 23 circuits of 16 different types (average frequency 1.44) of width 3 and height 1

190	FWBWZ	H(H).F()H.F()F	HH.hFhFf	6	2	8.7
191	FWZHD	H(H).F()F.(H)F	HH.hFfH.hf	6	2	8.7
192	FWFSD	H(H).F(H)H.F(H)F	HH.hFH.hhFH.hf	3	1	4.35
193	FWFDHD	H(H).F(H)F.(H)F	HH.hFH.hfH.hf	6	2	8.7
194	BWBWFD	H()H.F()H.F(H)F	HhFhFH.hf	3	1	4.35

without consanguineous kin constitutes a consanguineous component in itself).

The *length* of a kinship relation is the number of arcs and edges it contains. The *height* of a kinship relation is the length of the longest linear kinship relation it contains. The *width* of a kinship relation is the number of consanguineous components it contains.<sup>15</sup> The kinship relation represented by the path in Figure 35.4, for example, has a length of 4, a height of 2, and a width of 2.

### Matrimonial circuits

#### The concept of a matrimonial circuit

Real-world kinship networks grow in two ways. On the one hand, new individuals are born, generally being assigned a father and a mother from birth. In other words, a new vertex emerges

together with two arcs linking it to its parents, who are, at the moment of birth, the individual's only direct neighbors in the network. If we assume that the parents are already linked by a marriage edge (i.e., we are dealing with a canonical kinship network), the emergence of a new child vertex does not create a chain between individuals who are not already linked by a shorter chain. Its impact on global structure may thus be said to be marginal: it enlarges the network but does not alter its connectivity.

On the other hand, kinship networks also grow by marriage, that is, by the creation of new edges between two vertices, each of which can already be linked to a neighborhood of other vertices by several different lines going in all directions. The new marriage edge creates new connections between all these neighbors. In this way, marriage changes social structure. But at the same time, the way in which social structure would be changed by a potential marriage influences the marriage

choice itself, be it explicitly (through marriage rules and incest prohibitions), implicitly (by virtue of preferences or strategies), directly (by taking into account kinship ties between potential spouses), or indirectly (by taking into account other factors that are in turn correlated with kinship). The probability of two potential partners becoming a couple thus depends upon the nature of the kinship chains between them. Now, the most direct way to study this dependency is to look for those kinship chains that actually connect marriage partners – in other words, to look for circuits.

It should be clear from the preceding remarks that we are not interested in just any type of circuit, but in those circuits that contain at least one marriage edge. However, this restriction is not enough. Every triangle formed by a couple and their child – a *parental triangle* – is a circuit containing a marriage edge, and yet the couple's relation to their common child is hardly a condition of marriage that we wish to consider. More generally, we want to exclude all circuits containing parents together with their children – *parental triads*. We can therefore define the types of circuits of interest to us as those that contain at least one marriage edge and no parental triads. But in fact, the second condition implies the first: as descent is weakly acyclic, the only way to form a circuit in a kinship network without passing through a parental triad is to pass through a marriage edge. We thus arrive at a simple definition of a *matrimonial circuit*:

A *parental triad* is a graph formed by three vertices and arcs pointing from two of them to the third (that is, by parents and their child). If the parents are joined by a marriage edge, the resulting circuit constitutes a *parental triangle*.

A *matrimonial circuit* is a circuit that does not contain a parental triad. Alternatively, it can be defined as a connected subgraph where every vertex has degree 2 but no vertex has arc-degree 2. Because of the acyclicity of descent (condition 1 in the definition of kinship networks), this definition implies that a matrimonial circuit necessarily contains at least one marriage edge. In P-graph representation (where marriages are vertices and not edges), every circuit is a matrimonial circuit.

Any vertex of a matrimonial circuit that is adjacent to a marriage edge (that is, an individual whose marriage forms part of the circuit) is called a *pivot* of the circuit.

Any consanguineous chain within a circuit connecting two pivots is called an *arch* of the circuit.

A *matrimonial path* is a kinship path that passes through all vertices of a matrimonial circuit as

well as through all of its lines except the *closing* marriage edge, which links the first and the last vertex of the path. For a matrimonial circuit containing  $n$  marriage edges, there are  $2n$  different matrimonial paths.

If the kinship network is ordered (for instance, by arbitrary identity numbers), there is a unique rule of selecting, for any matrimonial circuit, a *characteristic path*: it is the matrimonial path that has the lowest possible Ego and (if there are two such paths) the lowest possible Alter.

A *matrimonial circuit type* is a class of isomorphic matrimonial circuits. Any matrimonial circuit type can be represented as a complex kinship relation formed by a marriage relation and another elementary kinship relation. We can define a unique rule for selecting, among these relations, the *characteristic relation* of a matrimonial circuit type: for example, the relation that begins with the longest sequence of ascending arcs and (if there are several sequences of equal length) with the lowest characteristic number.

### *Circuit inclusion and rings*

Matrimonial circuits have been defined in a most general manner as circuits that do not pass through parental triads. However, there may be situations where we might want to reduce our analysis to only some of these circuits – namely, those that have no “shortcut” linking two of its vertices. Consider, for example, a marriage with the daughter of the maternal uncle's wife (MBWD). Now, if this woman is at the same time the maternal uncle's daughter (MBD), many anthropologists would consider it improper to count her as an MBWD and would want to distinguish such marriages from marriages with “true” MBWDs (stepdaughters rather than daughters of maternal uncles). There is no a priori answer to the question of whether or not to count circuits that contain “shortcuts” of this kind. The choice depends both on the ethnographical context<sup>16</sup> and on the type of circuit in question.<sup>17</sup> In either case, it is useful to distinguish circuits that contain shortcuts from circuits without shortcuts that link their vertices. The latter are called *rings* (White, 2004).<sup>18</sup>

We say that a circuit *A* includes another circuit *B* if all vertices of the circuit *B* form part of the circuit *A*. This may also be stated by saying that *B* forms part of the subgraph *induced* by the vertices of *A*, that is, the graph constituted by these vertices and *all* of the lines that connect them in the global network (if a circuit *A* contains all the vertices of circuit *B*, the subgraph induced by these vertices also includes the lines of *B*). An *induced circuit* or *ring* can thus be defined as a circuit that is its own induced subgraph.<sup>19</sup>

The subgraph of a graph  $G$  induced by a vertex set  $V$  is the maximal subgraph of  $G$  having  $V$  as its vertex set (see Harary, 1969: 11). The subgraph induced by a circuit in the kinship network  $G$  is the subgraph of  $G$  induced by the vertices of the circuit. We also call it briefly the induced subgraph of the circuit.

A circuit  $A$  includes a circuit  $B$  if every vertex of  $B$  is also a vertex of  $A$  (that is, if  $B$  lies in the induced subgraph of  $A$ ).

An induced circuit or ring is a circuit that does not include any other circuit (that is, a circuit that is its own induced subgraph). Alternatively, it can be defined as a circuit such that no two vertices of the circuit are connected by a line that is not itself part of the circuit.

**Circuit intersection and composition**

The definition of an induced circuit (or ring) rules out circuits whose vertices are connected by extra lines. It does not, however, exclude circuits whose vertices are connected by extra chains (consisting of more than one line). Consider, for example, Figure 35.5, in which a man marries his MMBDD, while his own mother is his father's FFZD.<sup>20</sup>

The chains 1-8-9-5 and 2-8-9-5 "shorten" the outer circuit 1-2-3-4-5-6-7-1 (of type MMBDD), forming two inner rings of type FFZD: 8-9-5-4-3-2-8 and 1-8-9-5-6-7-1. But note that the outer circuit also constitutes a ring: as the vertices 8 and 9 do not belong to it, neither of the inner rings is included in it.<sup>21</sup> However, it intersects with the two inner rings in the sense that it has one or more lines in common with them (for a further discussion of circuit intersection, see section 4.2). Moreover, the entire outer ring can be composed from the two inner rings and the parental triangle 1-2-8-1 by taking their union and deleting all lines that form part of more than one circuit (an operation called *circuit union*). A circuit that in this

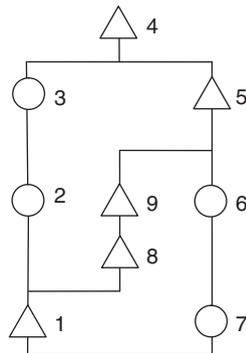


Figure 35.5

manner can be entirely decomposed into circuits shorter than itself is called *reducible*; if it cannot, it is called *irreducible*. By definition, every irreducible circuit is also a ring.

It should be stressed that reducible circuits are not necessarily sociologically less relevant than irreducible circuits. If a Fulani man, for instance, marries a FFBSD who is at the same time an MBD (due to an FBD marriage between the husband's parents), the apparent cross-cousin marriage MBD may simply be a by-product of successive parallel-cousin marriages (FBD and FFBSD): it is then the longer and not the shorter ring that matters for marriage decisions. Nor does the formation of a reducible circuit presuppose the previous formation of some irreducible circuit (see note 17).

The study of irreducible matrimonial circuits is closely related to the idea of a *cycle basis* in general graph theory. A cycle basis for a graph is a minimal set of circuits from which all circuits of a graph can be composed by a circuit union. Different sets of circuits can constitute a cycle basis. However, the number of circuits in the basis (also called the *circuit rank* or *cyclomatic number* of the graph) is invariant: we can compute it from the numbers of its arcs, edges, and components by means of a simple formula (see below). In order to find a cycle basis for a kinship network, it is reasonable to concentrate on irreducible circuits. Note, however, that the cycle basis may well be smaller than the set of irreducible circuits: if a man marries three sisters, we have three irreducible circuits, but the circuit rank is only two (as two circuits are sufficient to compose the third).

Two circuits intersect if they have lines in common.

They intersect matrimonially if they have marriage edges in common.

The union of two circuits is the graph that has the union of their line (vertex) set as its line (vertex) set. The circuit union of two intersecting circuits consists in taking their union and deleting the lines they have in common.

A circuit is irreducible if it cannot be composed by a circuit union from a set of circuits that are all shorter than itself.

A cycle basis of a graph is a minimal set of circuits whose union contains all the circuits of the graph.

The cyclomatic number or circuit rank of a graph is the number of circuits constituting its cycle basis (which is equivalent to the number of lines one has to remove from a graph to make it acyclic). For a graph with  $e$  lines,  $v$  vertices, and  $c$  components, it is calculated as

$$\gamma = e - v + c$$

### **Counting circuits: The matrimonial census**

A matrimonial census provides an exhaustive list of all matrimonial circuits of specified properties in a given kinship network, and it counts the occurrences of every distinct circuit type (which may or may not be aggregated into broader classes).

A circuit search usually has to be restricted to a limited set of circuit types. Even if the total number of circuits in a finite network is not infinite, it is usually so high that an unbounded search exceeds the capacities of the most advanced personal computers. For exploratory analysis, it is recommended to restrict matrimonial circuit search by criteria that are as neutral as possible, such as maximal width and height of circuits. Do not just count first-cousin marriages in order to decide whether you are dealing with “generalized exchange,” “arab marriage,” and so on – a look at higher degree consanguineous marriages, and at marriages between affines, may change the entire picture! Kinship structures constitute a whole and can only be understood if one considers them as such; they cannot be characterized by the frequency of this or that circuit type but only by the proportions and the interdependency of these frequencies. A matrimonial census must therefore be comprehensive in order to provide the basis for further analysis and interpretation, even if subsequent circuit searches may be more restricted and refined. It is clear that, even for small networks, such a task cannot be accomplished without computer support.

Circuit searches can be undertaken on the entire network or restricted to certain subsets of vertices. This restriction does not imply that *all* vertices of the matrimonial circuit have to belong to the subset – but their *pivots* have to lie within it (if, for example, we are interested in consanguineous marriages concluded between people born after 1800, we, of course, allow consanguineous chains to pass through ancestors born before 1800). Subsets may be defined according to “exogenous” criteria recorded for the individuals in the network (dates of birth, death or marriage, residence, occupation, etc.) but also according to “endogenous” criteria deriving from the kinship network itself (e.g., sibling group size, number of known ascendants, number of spouses, etc.). Such restrictions are not only convenient for comparative analysis and tests of representativity, but they may also be helpful in determining the optimal network to work with (see 6.2.1).

Table 35.1 presents excerpts from a matrimonial circuit census produced by the software Puck (*infra*) of a kinship network collected among the Watchi of Togo,<sup>22</sup> restricted to circuits of width

1 and height 3, circuits of width 2 and height 2, and circuits of width 3 and height 1.

### **NETWORK REPRESENTATIONS OF CIRCUIT STRUCTURES**

The set of matrimonial circuits thus obtained can in turn be studied with genuine network-analytic tools.

One of these tools is to construct the network that the circuits compose; this gives us a *subnetwork* of the original kinship network. A second tool consists of constructing the network of the structural relations between the circuits themselves; this gives us a *second order network* in which the circuits represent the vertices and their structural interrelations are represented by lines.

In the following section, we shall discuss one network of the first type, the matrimonial network, and one of the second type, the circuit intersection network.

### **Networks derived from circuit sets: The matrimonial network**

A *matrimonial network* is a subgraph of a kinship network resulting from the union of a set of matrimonial circuits, as, for instance, the circuits found by the matrimonial census in Table 35.1. Note that this is not equivalent to the subgraph *induced* by this set: for an arc or edge to be in the subgraph, it is not enough that each of its endpoints is in some circuit – the arc or edge must itself be part of a circuit. The matrimonial network derived from a set of matrimonial circuits found in a kinship network is thus simply the network composed of these circuits. It consists, in other words, of the matrimonially “interesting” regions of the original kinship network. The components of the matrimonial network (which we call *matrimonial components*) are connected subnetworks of matrimonial circuits, which may be studied from various perspectives. On the one hand, we may suppose that the frequent occurrence of particular matrimonial patterns is correlated with other properties of the network region concerned (for instance, social class, geographical region, or historical period); we may then apply several partitions to the network in order to evaluate the degree to which partition clusters correspond to matrimonial components. On the other hand, we may interpret the density of circuits as an effect of self-reinforcing social mechanisms (behavior transmission, imitation, or the presence of rules) or as

a simple network effect (circuits combining to compose other circuits) that we did not consider when defining the criteria for our initial circuit search.

The concept of a matrimonial network is also meaningful in and of itself, independent of any particular circuit set. Even in cases where it is not possible to precisely identify all matrimonial circuits (without limits of size) that may exist in a kinship network, it is possible to determine which part of the network is composed of matrimonial circuits (of whatever length). The result – the largest possible matrimonial network – is the *nucleus* of the kinship network, the union of all existing circuits in the network (see Grange and Houseman, 2008).

The concept of the nucleus can be more strictly delineated by introducing the concept of a *matrimonial bicomponent*, that is, a maximal subgraph in which every two vertices form part of a matrimonial circuit (note that this is a stricter condition than that of simply forming part of a circuit, as is required for the more general notion of a bicomponent). The nucleus is simply the union of all matrimonial bicomponents.<sup>23</sup>

As a result, matrimonial *components* (consisting only of circuits) are closely related to matrimonial *bicomponents*: both are line-biconnected (two distinct line series link each vertex to every other), but matrimonial bicomponents have the additional feature of being vertex-biconnected as well (the two interconnecting line series never run through the same vertex).

In the kinship network represented in Figure 35.6, for example, the shaded individuals and their interconnections within the bold boundaries constitute the nucleus, which is composed of two matrimonial components (A and B) and three matrimonial bicomponents (1, 2, and 3), two of which overlap (one individual is included in both 1 and 2).

Given a set of circuits  $R$  in a kinship network  $K$ , the *matrimonial network* derived from  $R$  is the subgraph of  $K$  resulting from the union of the circuits of  $R$ . In other words, it is a subgraph in which every line belongs to some matrimonial circuit of  $R$ . The components of a matrimonial network are called the *matrimonial components* of  $K$  with respect to  $R$ . Matrimonial components

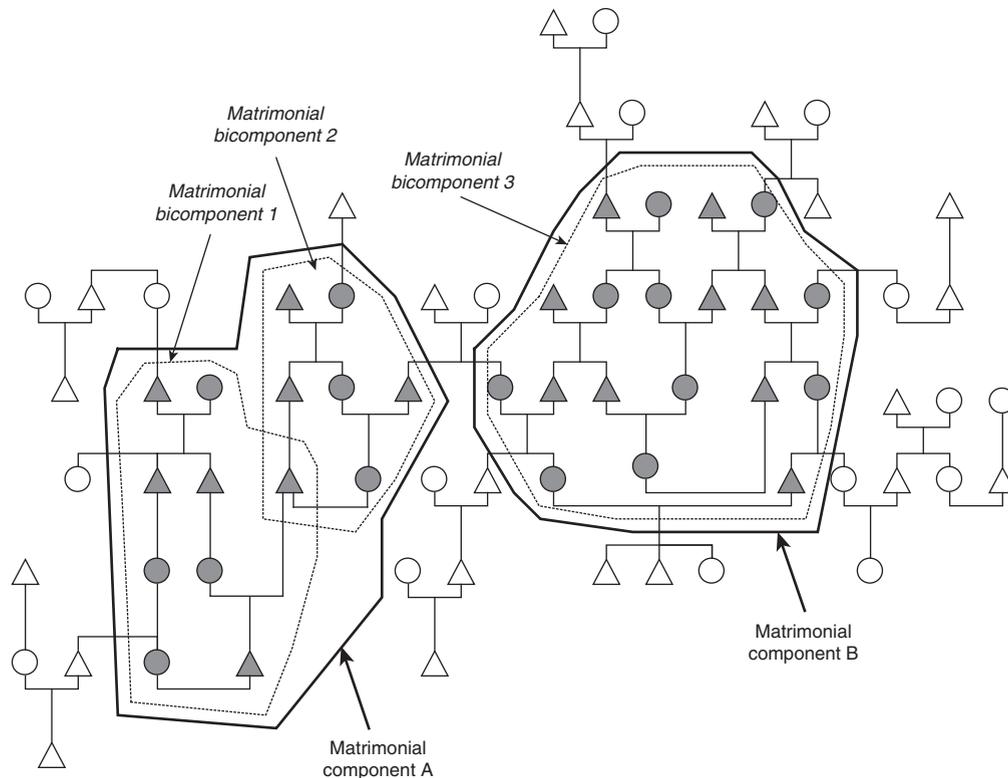


Figure 35.6

are line-biconnected but not necessarily vertex-biconnected.

A *matrimonial bicomponent* is a subgraph of  $K$  in which every pair of vertices (no matter how distant) belongs to a matrimonial circuit (hence no pair of vertices can be separated by removal of a single intermediary vertex). In  $P$ -graph representation, every bicomponent is a matrimonial bicomponent.

The *nucleus* of a kinship network is the network resulting from the union of all matrimonial circuits in  $K$ . The nucleus is equivalent to the union of all matrimonial bicomponents of the kinship network. The largest of these matrimonial bicomponents constitutes the *kernel* of the kinship network.

### **Networks of circuits: The circuit intersection network**

Circuit graphs are tools for analyzing the interdependence of matrimonial circuits. A simple example is the *circuit intersection network*, where circuit interdependence is measured by the frequency of shared marriage edges.

It is often the case that a given marriage forms part of two or several matrimonial circuits. Such overdetermination poses problems of sociological interpretation: if a man has married a woman who is at the same time his father's sister's daughter, his maternal aunt, and his sister-in-law, should we say that he married his cousin, his aunt, or his sister-in-law? And how shall we interpret such a situation if, for instance, it is considered to be very good to marry one's cousin but bad to marry one's aunt? Under such circumstances, it is clearly insufficient to simply count the frequency of maternal aunts that are spouses and affirm the presence of a high number of "bad" marriages, without considering the fraction of such aunts who are at the same time cousins and thus represent "good" spouses. But there is a further reason for wanting to determine the frequencies of circuit intersection: if a father's sister's daughter is at the same time a maternal aunt, such a configuration mathematically implies that the husband's father has married his sister's daughter. We shall therefore necessarily find a high number of niece marriages in our network, and if we did not yet search for them, such a finding will prompt us to do so.

Circuit intersection networks are an easy and intuitive way to approach these questions. In these networks, vertices represent circuit types, the size (or vector value) of vertices represent circuit frequencies, and the values of the lines connecting two vertices represent the number of marriages

that are simultaneously part of a circuit of the two corresponding types.

The *matrimonial intersection* of two circuit types  $A$  and  $B$  is the set of all marriage edges belonging simultaneously to a circuit of type  $A$  and to a circuit of type  $B$ .

The *circuit intersection network* corresponding to a set of circuit types  $T$  is a valued graph  $G$ , such that

- 1 each vertex of  $G$  corresponds to a circuit type of  $T$ , and the value (size) of the vertex is proportional to the frequency of the corresponding circuit type
- 2 each edge between two vertices corresponds to a nonempty matrimonial intersection of the corresponding circuit types and the weights of the edges correspond to the size of the intersections

### **ALLIANCE NETWORKS**

All matrimonial structures hitherto discussed have been defined in terms of filial and marriage relations between individuals. But kinship also has to do with relations between *groups* derived from relations between individuals. One convenient tool of analyzing these relations is the *alliance network*.

The *alliance network* corresponding to a given kinship network (real or simulated) is a network composed of vertices representing *groups* of individuals and arcs representing *marriage frequencies* between the groups, where the value of an arc from  $A$  to  $B$  indicates the number of marriages of a woman of  $A$  with a man of  $B$ . One can think of alliance networks as resulting from "shrinking" the kinship network with respect to some partition, after having transformed marriage edges into arcs that point in the husband's direction, that is, from the group of "wife-givers" to the group of "wife-takers." The clusters of the partition are those collectivities related by the marriages between their respective members. The partition used may be exogenous to kinship (for instance, if we are dealing with residential units, professional categories, or social classes), but it may also be derived from the kinship network itself. Thus, for example, clusters may represent the components of the agnatic or uterine subnetwork, that is, "lineages" within the kinship network.

Alliance networks no longer have the particularities of kinship networks. They are simple valued digraphs. Circuits in alliance networks may be called *connubial circuits* in order to distinguish them from matrimonial circuits in kinship

networks, and the connubial circuit structure of an alliance network can be represented by its matrix (the *alliance matrix*). In exogamous systems, the diagonal will have only zero values; in a perfect system of balanced bilateral marriage alliances (designated by Lévi-Strauss [1949 (1967)] as “restricted exchange”) it will be symmetric; in a system having a single directed Hamiltonian cycle (“generalized exchange”), it will be asymmetric and contain only one nonzero value in each line and column. Note, however, that real-world alliance structures are not as clear-cut. On the one hand, several circuit patterns may be superimposed, such that, for example, the resulting cumulative pattern may appear symmetrical even if partial patterns are asymmetric. On the other hand, the structure of alliance networks is highly sensitive to the definition of wife-giving and wife-taking groups, connubial circuit structures being in general not continuous with respect to changes in the level of group aggregation. Thus, what may appear to be endogamous unions on one level may be exogamous on another, with combinations of endogamy and balanced exchange on one level giving rise to asymmetrical patterns on another (for an example, see Gabail and Kyburz, 2008).

The *alliance network* corresponding to a partitioned kinship network  $K$  is a digraph  $G$  where

- 1 each vertex of  $G$  corresponds to a partition cluster of  $K$
- 2 each arc between two vertices of  $G$  corresponding to two clusters  $A$  and  $B$  corresponds to the existence of a marriage of a woman of  $A$  with a man of  $B$ , and the weight of the arcs correspond to the number of such marriages.

A *connubial circuit* is a circuit in an alliance network.

## APPLICATION ISSUES

### *Data collection and saving*

#### *Orientations of data collection*

Of course, the way kinship data collection is oriented depends largely on what one wants to find. However, in order to minimize biases and to allow the data to be used by others, there are some criteria that every corpus of kinship data should fulfill, regardless of the specific purpose of the collection:

- Document the dates and informant(s) for every part of the data set; this is not only useful for

checking possible errors and biases but may be an interesting datum in itself, for instance, to appraise genealogical memory.

- Document missing data; indicate if the spouses and children you have noted for a given individual are, according to your knowledge, complete.
- In case of contradictory data, keep records regarding alternative items of information (and their origin) and on the reasons for your choices.
- Try to avoid biases from the start by always following both male *and* female lines; do not record only those kinship ties that are easily given but make an effort to search those that are missing, even if this may be costly in the case of an extensive matrimonial area.
- If you are interested in the kinship relations between two individuals, do not be content to ask how they are related but try to establish their entire pedigree within given bounds – you will surely find quite a number of additional ties your informants did not mention spontaneously!
- Try to determine the relative order of births and marriages (in particular in the case of polygamy), even if absolute dates are not available. In interpreting marriages with affines or overlapping matrimonial circuits, this may turn out to be very important.
- Keep an account of the research method used and specify the purpose for which the corpus (or part of the corpus) is established.

#### *Storing data without a computer*

Data are not only a result but also a means of data collection. They should be easily accessible in order to guide your research and to cross-check informant answers. When dealing with archives, this is often fairly simple: you can take a computer with you. But in many fieldwork situations this is not possible. However, noting kinship “by hand” can be extremely fast and efficient, if some basic principles are observed.

- Always use a compact medium, such as a notebook. Do not use filesheets or loose papers. You cannot easily use them during interviews, and there is a high risk of losing some of them.
- Separate graphics and text. A good method is to use a notebook with the left page for drawing genealogies, the right page for listing the individuals and their properties, and numbers for identifying these individuals (if numbers get large, it is recommended to use, in addition, initial letters to prevent identification problems in case of numbering errors).
- Attribute an identity number to each individual and never attribute that number to another individual. If you have “duplicates,” make a link to the original number but do not re-assign it.

Gaps in the series of numbers do not cause any damage, but ambiguities in identity numbers cause much damage and are extremely difficult to detect.

- Do not use identity numbers as codes. Identity numbers serve to identify individuals and nothing else (except, perhaps, to recall the order in which you have entered them and to document the history of your corpus). If you want to convey information on individuals' gender, clan affiliation, residence, and so on, do not use identity numbers for that purpose.
- Never forget to make copies and store them in different places. This holds for all data, but especially for kinship data, due to the network properties of kinship: one lost notebook may render all other notebooks useless.

### **Data interpretation**

It is necessary to know the biases, gaps, and limits of a genealogical corpus. In some cases, this may lead one to conclude that certain analyses simply cannot reasonably be made. For instance, it makes little sense to search for consanguineous marriage circuits in a corpus with very shallow genealogies. Here, as in all cases, knowledge of the basic qualities of the corpus is essential for interpreting the findings.

In particular it is important to keep in mind that kinship networks are virtually infinite. Every corpus is necessarily only a part of a larger whole, and incomplete both with regard to individuals and with regard to the links that connect them. Choices pertaining to the delimitation and composition of genealogical corpuses arise not only during fieldwork but also prior to analysis as a means of reducing the collected data to a meaningful core.

### **Choosing bases and boundaries**

The usefulness of exogenous reductions bringing sociological, geographical, and demographical criteria into play is fairly self-evident. However, the value of endogenous reductions based on structural features of the network itself is perhaps less so, and depends on what one is trying to show. For example, one cannot first eliminate all unmarried individuals if one later wants to compare matrimonial circuit frequencies with those of kinship relations (e.g., White 1999). This applies not only to endogenous reductions but also to endogenous augmentations of the network, such as the creation of fictive individuals in order to preserve information on full siblingship in those cases where one or both parents are unknown.

A related problem concerns the delimitation of the network to be used for analytic purposes. To begin with, it is important that results relating to unconnected components of the matrimonial network not be combined indiscriminately, as each component represents an autonomous matrimonial universe whose patterning may not obey the same rules. This also applies, to a lesser degree, to matrimonial bicomponents. However, even when analysis is limited to the largest matrimonial bicomponent (kernel) of the kinship network, additional restrictions are often necessary. Because bicomponents contain matrimonial circuits of any length, height, and width, including those that incorporate very distant ties whose sociological relevance is questionable, it is helpful to confine analysis to subnetworks formed by matrimonial circuits of a certain maximal height or width. It is not always easy to determine the optimal criteria for such restricted matrimonial networks, choices being largely guided by two contrary principles. On the one hand, the resulting subnetwork should be as large as possible so as to be sufficiently representative of the wider kinship network from which it is drawn. On the other hand, in order to avoid redundancies and to keep the number of circuits and circuit types at manageable proportions, it is essential that the latter be kept to a minimum.

### **Determining representativity and significance**

Once a delimited data set has been chosen and first results obtained, another difficulty arises: to what extent do the regularities observed in the corpus provide information on the organization of the real social network? This, of course, is a general problem in network analysis; however, because issues of incompleteness are so omnipresent in genealogical research, questions regarding the representativity of kinship networks are particularly pressing.

One central question concerns the extent to which a corpus's boundaries correspond to endogenously definable subnetworks of the real kinship network. Another, related question concerns the over- and underrepresentation of persons belonging to certain kinship categories. These issues apply not only to the individuals (vertices) that make up the corpus, but to the relations (lines) between them. Kinship networks are often biased in that they favor some types of relations over others (agnatic over uterine relations, for example, for a population with patrilocal residence). It might be possible to adjust one's findings so as to eliminate such biases. However, these biases may themselves be a function of matrimonial behavior: kinship ties that form part of matrimonial circuits may be more easily remembered than others.

The stronger this correlation between network structure and collective memory is, the more difficult it is to do away with the biases concerned.

The problem of network biases is directly related to issues of significance: are the observed regularities indicative of a behavioral pattern, or are they simply due to chance? In order to treat this question by means of a comparison with random kinship networks, it is necessary to simulate a network that not only reproduces the demographic features of the concerned population but also the biases of the corpus itself. Because of this, random permutation (see White, 1999; White, this volume) or explicit simulation of the data collection process (virtual fieldwork) may prove more useful than conventional demographic simulation.

### *Interpretation of results*

If “interesting” regularities appear in analysis for which the known norms and institutions of the society do not provide a straightforward account, one is immediately tempted to take them as indicators of some hidden norms or institution. The most important thing in kinship data analysis is to resist this temptation. Before formulating a sociological hypothesis, one must always check the following:

- First, does the “interesting” structural feature simply reflect a bias of the corpus? (for instance, high incidences of patrilineal marriages in a corpus where uterine genealogies are shallow).
- Second, is the structural feature, if it does not correspond to any known rule, the result of a *combination* of known rules? (for example, in a society where uterine nieces are preferred spouses and maternal aunts are avoided, patrilineal cross-cousins have a greater chance of being at the same time maternal aunts and therefore avoided, even if no rule states that they should be avoided as such).
- Finally, remember that a sociological hypothesis is not validated by the simple fact that *no other* interpretation can account for the structural feature observed. Sociological hypotheses often can be validated only in the field.

### **Resources**

A wide variety of commercial software (Brother’s Keeper, Family Tree Maker, Legacy, Kith and Kin, etc.) and noncommercial programs (Personal Ancestral File, Gramps, etc.) exist for entering, storing, and outputting genealogical data. Some of them are particularly flexible and are used extensively by social scientists: G n atique is

used by many French historians; Alliance Project, developed by S. Sugito and S. Kubota in Japan (Sugito, 2004), is widely used by Australian anthropologists; Kinship Editor, written by M. Fischer, is a key element of the European Kinship and Social Security (KASS) project, and it allows both for systematic data collection and the modeling of kinship phenomena such as terminological usage. There are also additional software tools developed by demographers or historians for managing records and calculating demographic and other variables on the basis of empirical genealogical data (e.g., CASOAR [Hainsworth and Bardet, 1981]). None of these programs, however, allow for an in-depth analysis of kinship networks as such.

Such analyses may be undertaken using general purpose social network analysis tools. Perhaps the software most used in this way is the network analysis and visualization program Pajek, developed by V. Batagelj and A. Mrvar (de Nooy et al., 2005; Batagelj and Mrvar, 2004, 2008; Mrvar and Batagelj 2004). A number of kinship-centered macros have been developed for Pajek by the authors and by others (e.g., Tip4Pajek pack by K. Hamberger).

Finally, certain programs have been specifically developed for the analysis of kinship networks (including matrimonial circuit censuses). First attempts, such as Gen-Par by M. Selz (1987; see also H ritier, 1974) and Pgraph by D. White (1997, White et al. 1999), have since given way to more flexible and easier-to-use software such as Genos by L. Barry (2004) and, more recently, Puck by K. Hamberger (Hamberger et al., 2009).

An open depository of kinship networks from historical and anthropological sources, controlled by a scientific board, is hosted at the site of the *kinsources* project (<http://kinsource.net>).

### **NOTES**

1 We are grateful to Isabelle Daillant and Vladimir Batagelj as well as to the editors Peter Carrington and John Scott for helpful comments and discussions.

2 We are talking of individuals and not of classes. An individual may well belong to one’s parent’s parent’s marriage class, as in Australian alternating generation models.

3 In digraphs, the notions “path” and “cycle” are often restricted to oriented paths and cycles, whereas the terms “semipath” and “semicycle” are used if arcs are not consistently oriented. Because we are talking of mixed graphs (containing edges as well as arcs), we use “path” and “cycle” as the general terms (as in the case of undirected graphs) and

specify them as “oriented” if all of their lines are arcs pointing in the same direction.

4 The terms “chain” and “circuit” are equivalent to “path graph” and “cycle graph” in graph-theoretic literature. Note, however, that these terms are also sometimes used in a different meaning, namely, as synonyms for “walk” (open and closed, respectively).

5 Named after the Scandinavian mathematician Oystein Ore (1960). The representation of marriages by edges has been introduced by the computer program Pajek (see Batagelj and Mrvar, 2008).

6 Note that the weakly acyclic property of kinship networks holds for remarriage cycles because a cycle consisting entirely of edges is not an oriented cycle.

7 Kinship networks that incorporate adoptive, godparent, or co-genitor relations often have to allow for the possibility of more than one parent of a given gender (multiple descent).

8 A vertex  $x$  is arc-adjacent to another vertex  $y$  if the arc goes from  $x$  to  $y$ . It is arc-adjacent *from*  $y$  if the arc goes from  $y$  to  $x$ .

9 The upper and lower vertices of the P-graph and bipartite P-graph are reversed in Figure 35.3 compared to the Ore-graph. This is because their arrows run in reverse, from children to parents, usually as a unique function mapping a child to a unique parental couple. The inverse of this function gives the relation of parents to children.

10 The “p” stands for *parenté* (French for “kinship”).

11 From the French *homme* (man) and *femme* (woman).

12 This positional notation may also be used, within limits, to represent complex kinship relations, by leaving the parentheses empty (or by putting two letters in it) if the relation entails kinship paths that pass through apical ancestors of both genders. For instance, the MBD (mother’s brother’s daughter) relation is represented as HF(HF), the ZH (sister’s husband) relation as H(F.H), and the FWS (father’s wife’s son) relation as H(H).(F)H.

13 Original P-graph notation (White and Jorion, 1992) used letters G and F (from the French *garçon* (boy) and *filles* (girl)). In its present version, P-graph notation uses the same letters as positional notation (but applies them to lines while positional notation applies them to vertices).

14 This is a variant of the ahnentafel genealogical numbering system.

15 In the case of consanguineous kinship relations, length is also called roman (or civil) degree, whereas height is also called german (or canonic) degree (these terms derive from the history of European kinship). In the case of linear kinship relations, roman and german degree are identical, and we can simply speak of its degree.

16 For example, in the case of a Dravidian kinship network from southern India, where cross-cousin

marriages may be considered as redoublings of alliance relations rather than as consanguineous marriages (Dumont, [1953] 1975), counting all MBWDs, whether they are or are not also MBDs, would make good sense.

17 For instance, if two brothers marry two sisters, one brother dies, and the remaining brother marries the widow, the fact that the longer circuit (of type BWZ) includes two shorter ones (of types BW and WZ) does not in the least make it sociologically less relevant: the BWZ marriage clearly precedes the BW (and WZ) marriage (we are grateful to Isabelle Daillant for this example).

18 The distinction between circuits and rings (induced circuits) is a rather recent one. Much of what has been said regarding matrimonial rings in Hamberger et al. (2004) refers to matrimonial circuits in general, while *rings* in White (2004) refers to induced circuits.

19 Note, however, that this restriction may not mean the same thing in P-graph and in Ore-graph representation. The subgraph induced by the vertices of a circuit changes meaning according to whether vertices are defined as individuals (Ore-graph) or as marriages (P-graph). For instance, a marriage with an MMBDD who is at the same time an MBD constitutes a ring in Ore-graph but not in P-graph representation. If the context is not clear, one should therefore speak of Ore-rings and P-rings to avoid ambiguities.

20 This example holds in a P-graph as well as in Ore-graph representation: in both cases, the outer circuit is a ring. In general, Ore-rings are not always P-rings (see note 19).

21 This is equally true in P-graph representation, where the two individuals 8 and 9 are represented by two lines forming an extra chain between the vertices (representing marriages) of the outer circuit.

22 See <http://kinsource.net/kinsrc/bin/view/KinSources/Watchi>.

23 Every line of a matrimonial bicomponent is by definition in some matrimonial circuit. On the other hand, every matrimonial circuit is either a matrimonial bicomponent in itself or forms part of some larger matrimonial bicomponent. The concept of the nucleus is a restriction of the definition of the *core* by White and Jorion (1996; cf. Houseman and White, 1996) as the union of all matrimonial bicomponents and their single-link connections (equating with 2-core as defined by Seidman, 1983). Neither is necessarily connected.

## REFERENCES

- Barry, L. (2004) ‘Historique et spécificités techniques du programme Genos’, *École Collecte et traitement des données de parenté*, <http://llacan.vjf.cnrs.fr/SousSites/EcoleDonnees/extras/Genos.pdf>.

- Batagelj, V. and Mrvar, A. (2004) 'Analysis large genealogical networks with Pajek', paper presented at Sources et Ressources pour les Sciences Sociales, Paris, EHESS, December 9–11, 2004, <http://vlado.fmf.uni-lj.si/pub/networks/doc/seminar/Geneo04.pdf>.
- Batagelj, V. and Mrvar, A. (2008) 'Analysis of kinship relations with Pajek', *Social Science Computer Review*, 26(2): 224–46, <http://ssc.sagepub.com/cgi/content/abstract/26/2/224>.
- Dumont, L. ([1953] 1975) 'Le vocabulaire de parenté dravidien comme expression du mariage', in *Dravidien et Kariera, l'alliance de mariage dans l'Inde du Sud et en Australie*. Paris/La Haye: Mouton. pp. 85–100 and 145–46.
- Gabail, L. and Kyburz, O. (2008) 'Hurons chez les Toureg', *Annales de Démographie Historique*, 116:197–232.
- Grange, C. and Houseman, M. (2008) 'Objets d'analyse pour l'étude des réseaux de parenté', *Annales de Démographie Historique*, 116: 105–144.
- Hainsworth, M. and Bardet, J.-P. (1981) 'Logiciel C.A.S.O.A.R.: Calculs et analyses sur ordinateur appliqués aux reconstitutions', *Cahiers des Annales de Démographie Historique*, 1: 1–175.
- Hamberger, K. (forthcoming) 'Matrimonial circuits in kinship networks: calculation, enumeration and census', *Social Networks*
- Hamberger, K. and Daillant, I. (2008) 'L'analyse de réseaux de parenté: concepts et outils', *Annales de Démographie Historique* 116: 13–52.
- Hamberger, K., Houseman, M., Daillant, I., White, D.R. and Barry, L. (2004) 'Matrimonial ring structures', *Mathématiques et Sciences humaines/Mathematics and Social Sciences* 168 (*Les réseaux sociaux*, ed. Alain Degenne), pp. 83–119.
- Hamberger, K., Houseman, M. and Grange, C. (2009) 'La parenté radiographiée: Un nouveau logiciel pour l'analyse des réseaux matrimoniaux', *L'Homme*, 189: 107–137.
- Harary, F. (1969) *Graph Theory*. Reading, MA: Addison-Wesley.
- Harary, F. and White, D.R. (2001) 'P-systems: A structural model for kinship studies', *Connections*, 24(2): 35–46.
- Héritier, F. (1974) 'Systèmes Omaha de parenté et alliance. Etude sur ordinateur du fonctionnement réel d'une société africaine', in P.A. Ballonoff (ed.), *Genealogical Mathematics*, Paris-La Haye: Mouton.
- Houseman, M. and White, D.R. (1996) 'Structures réticulaires de la pratique matrimoniale', *L'Homme*, 36(139): 59–85.
- Lévi-Strauss, C. ([1949] 1967) *Les structures élémentaires de la parenté*. 2nd ed. Paris/La Haye: Mouton.
- Mrvar, A. and Batagelj, V. (2004) 'Relinking marriages in genealogies', *Metodolo kizvezki Advances in Methodology and Statistics* (Ljubljana), 1: 407–18, <http://mrvar.fdv.uni-lj.si/pub/mz/mz1.1/mrvar.pdf>.
- de Nooy, W., Mrvar, A. and Batagelj, V. (2005) *Exploratory Social Network Analysis with Pajek*. New York: Cambridge University Press.
- Ore, O. (1960) 'Sex in graphs', *Proceedings of the American Mathematical Society*, 11: 533–39.
- Seidman, S.B. (1983) 'Network structure and minimum degree', *Social Networks*, 5: 269–87.
- Selz, M. (1987) 'Parenté et Informatique', *Mathématique et Sciences Humaines*, 97: 57–66.
- Sugito, S. (2004) 'Possibility of genealogical study in population study', *Journal of Population Studies*, 34: 23–29 (in Japanese).
- White, D.R. (1997) 'Structural endogamy and the Graphe de Parenté', *Mathématiques, informatique, et sciences humaines*, 137: 107–25.
- White, D.R. (1999) 'Controlled simulation of marriage systems', *Journal of Artificial Societies and Social Simulation*, 2(3), <http://www.soc.surrey.ac.uk/JASSS/2/3/5.html>.
- White, D.R. (2004) 'Ring cohesion in marriage and social networks', *Mathématiques, informatique, et sciences humaines*, 168(4): 59–82.
- White, D. R., Batagelj, V. and Mrvar, A. (1999) 'Analyzing large kinship and marriage networks with Pgraph and Pajek', *Social Science Computer Review*, 17(3): 245–74.
- White, D.R. and Jorion, P. (1992) 'Representing and computing kinship: A new approach', *Current Anthropology*, 33(4): 454–62.
- White, D.R. and Jorion, P. (1996) 'Kinship networks and discrete structure theory: Applications and implications', *Social Networks*, 18: 267–314.

### Software

- Alliance Project*: <http://study.hs.sugiyama-u.ac.jp/e/>
- Brother's Keeper*: <http://www.bkwin.org/>
- Family Tree Maker*: <http://www.familytreemaker.com/>
- Généatique*: <http://www.cdip.com/>
- Gramps*: [http://gramps-project.org/wiki/index.php?title=Main\\_Page](http://gramps-project.org/wiki/index.php?title=Main_Page)
- Kinship Editor*: <http://era.anthropology.ac.uk/Kinship/>
- Kith and Kin*: <http://www.spansoft.org/>
- Legacy*: <http://www.legacyfamilytree.com/>
- Pajek*: <http://pajek.imfm.si/doku.php>
- Personal Ancestral File*: <http://www.familysearch.org/eng/paf/>
- Pgraph*: <http://eclctic.ss.uci.edu/~drwhite/pgraph/pgraph.html>
- Puck*: <http://kintip.net/>
- Tip4Pajek*: <http://kintip.net/>, <http://intersci.ss.uci.edu/wiki/>