P-Systems:
A Structural Model for Kinship Studies

Frank Harary
Department of Computer Science, New Mexico State University

Douglas R. White
Department of Anthropology, University of California, Irvine

November 15, 2000
We dedicate this paper to the memory of Oystein Ore

Several mathematical models have been proposed for kinship studies. We propose an alternate structural model designed to be so simple logically and intuitively that it can be understood and used by anyone, with a minimum of complication. It is called a P-system, which is short for parental system. The P-system incorporates the best features of each of the previous models of kinship: a single relation of parentage, graphs embedded within the nodes of other graphs, and segregation of higher level descent and marriage structure from nuclear family structure. The latter is also the key conceptual distinction used by Lévi-Strauss (1969) in the theory of marriage alliance. While a P-system is used to represent a concrete network of kinship and marriage relationships, this network also constitutes a system in the sense that it contains multiple levels where each level is a graph in which each node contains another graph structure. In sum, the connections between the nodes at the outer level in a P-system are especially useful in the analysis of marriage and descent, while at inner level we can describe how individuals are embedded in the kinship structure.

Introduction

Several mathematical models have been proposed for kinship studies. Those that are sufficiently general to allow a network analysis of kinship and marriage or the recording of genealogical data include the genetic graph proposed by the great Norwegian mathematician Oystein Ore (1960), the

---

FH Fax 505 646 6218; DW Fax 949 824 4717. This research was supported by NSF Award BCS-9978282 (1999-2001) to Douglas White, "Longitudinal Network Studies and Predictive Social Cohesion Theory," with Frank Harary as consultant.
multilevel graph of Harary and Batell (1981), and the p-graph of White and Jorion (1992, 1996), which ultimately derives from the algebraist André Weil (1969). Our present purpose is to propose an alternate structural model, called a P-system, designed to be so simple both logically and intuitively that it can be understood and used by anyone, with a minimum of complication.

**Genetic Graphs and Systems with Multiple Levels**

The seminal paper of Ore (1960), in which he proposes a model for biological descent, Weil (1969), who proposed an algebra of marriage systems, and Harary and Batell (1981), who embed graphs within graphs within graphs as a formulation for systems, are foundational to the mathematical formulation of our model. Some basic definitions are required to introduce their concepts. A **digraph** may have a symmetric pair of arcs, as in $D_1$ of Figure 1, or none as in $D_2$. The **underlying graph** (Harary, 1969) of a digraph $D$ has the same nodes as $D$, but each arc and each symmetric pair is replaced by an undirected edge, as in $G$.

![Figure 1](image1.png)

**Figure 1.** $D_1$: a digraph; $D_2$: an asymmetric digraph; $G$: the underlying graph of $D_1$ and $D_2$

Ore modelled two parents and one child by a digraph in which each of two nodes, $h$ and $w$, have an arc to a third node, $c$. Here, $h$ and $w$ denote husband and wife, and $c$ stands for child. When there are two children, $c_1$ and $c_2$, there are arcs from $h$ and $w$ to each of them. Figure 2 shows three digraphs, with five nodes, $h$, $w$, $c_1$, $c_2$, $c_3$, a husband and wife with three children; $D_1$ with nodes for a series of marriages, each couple having one child ($c_1$, $c_2$, $c_3$), where a wife $w_1$ has two successive husbands $h_1$ and $h_2$, and then $h_2$ has a new wife $w_2$; after which, in $D_3$ following an appropriate divorce, $h_1$ and $w_2$ marry and have a child, $c_4$. Being a professional mathematician, Ore had a compulsion to include at least one original theorem in each of his papers. An oriented graph is obtained from a graph $G$ by assigning a direction to each edge of $G$. His one theorem in the article (1960) is that when one considers the graph $G$ of his genetic oriented digraph, every cycle of $G$ has length divisible by four. Three cycles of length 4 are evident in the graph of $D_1$, none in $D_2$, and a single cycle of length 8 in $D_3$. The only way that a cycle of length 6 could be produced would be if $h_1$ and $h_2$ or $w_1$ and $w_2$ were to produce a child, which is a biological impossibility.

![Figure 2](image2.png)

**Figure 2.** Three digraphs illustrating Ore’s theorem about cycles

We must note that Ore’s theorem applies only to idealized assumptions such as when (1) only two generations are involved or (2) we do not consider marriages among persons previously related by descent. When Oedipus marries his mother, for example, the theorem does not apply (Harary 1982).
A limitation of Ore’s genetic digraphs, although they can represent empirical networks of descent, is that they do not take marriage into account as distinct from parentage, as does the marriage system formalization of Weil (1969). Weil’s algebraic discussions, however, are also based on idealized marriage rules, and do not handle the complexity of empirical kinship networks. What is lacking is some combination of elements of these two approaches that can model both marriage and descent in ways that lead to better intuitive understanding of kinship systems. We also note that Weil’s analysis treats idealized types of marriages, while Ore’s digraph uses nodes to represent individuals. Two different levels of analysis are thus involved in these two approaches, one using algebra and the other graphs.

If we are interested in combining models for marriage and descent into a more general systems model, Harary and Batell (1981) define a system as sets of relations among elements at different levels where each level is a graph in which each node may contain another graph structure. The embedded graphs approach to systems provides a way to integrate mathematical models of marriage and descent into what we call a P-system.

P-system 2

A P-system is neither a graph nor a digraph, as it may have three types of nodes representing a single female, 0, a single male, 1, or a reproducing couple, 01. It has, however, only one type of arc, as in Ore’s genetic digraph. Further, a P-system has two levels of nodes. Each node at level-1 in a P-system contains a graph at level-2. An arc from \( u \) to \( v \), where \( u,v \) are nodes of either type, represents parentage. At level-1, such an arc entails that \( u \) is either a parent or a parental couple (e.g., a married couple). If node \( u \) at level-1 is a married couple, then that node contains a pair of nodes at level-2, the husband and the wife. It will often be the case in a kinship network that node \( v \) at level-1 is a single child, hence node \( v \) will contain, at level-2, a node representing a single individual.

In a P-system there are three types of nodes: females (coded 0), males (coded 1), and couples (coded 01 for a female-male pair or 10 for a male-female pair, as convenient to simplify the diagram, or coded 2 = [0,1]=[1,0] if the order of the pair makes no difference). Conventionally, couples will be married. A relation of parentage may exist between any pair of nodes regardless of type, giving sixteen possible combinations of nodes joined by arcs, as shown in Figure 3. We stipulate that each of the edges is oriented here from left to right, giving 16 arcs that go between nodes of the four different types. The coding of nodes and the sixteen possible ordered pairs of nodes connected by arcs are shown for level-1 of any given P-system. Also shown are the level-2 interpretations of the graphs within the nodes at level-1. For example, if an arc goes from a node \( u \) of type 0 to a node \( v \), then node \( u \) is in a mother relation to node \( v \). If \( v \) is of type 0, then \( v \) is \( u \)'s daughter. If node \( v \) is coded 01, then \( v \) represents a daughter and her husband, and if \( v \) is 10, then \( v \) represents a son and his wife.

<table>
<thead>
<tr>
<th>Level-2 Parental Legend for Nodes</th>
<th>Arcs at Level-1</th>
<th>Level-1</th>
<th>Nodes at Level-2</th>
</tr>
</thead>
<tbody>
<tr>
<td>From Node</td>
<td>To Node</td>
<td>Child Legend</td>
<td>Level-2 Offspring</td>
</tr>
<tr>
<td>mother</td>
<td>0</td>
<td>0</td>
<td>daughter (single)</td>
</tr>
<tr>
<td>father</td>
<td>1</td>
<td>1</td>
<td>son (single)</td>
</tr>
<tr>
<td>mother and father</td>
<td>01</td>
<td>01</td>
<td>daughter</td>
</tr>
<tr>
<td>father and mother</td>
<td>10</td>
<td>10</td>
<td>son</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>daughter and her husband</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>son and his wife</td>
</tr>
</tbody>
</table>

Figure 3. The 16 possible parental connections at level-1

---

2 The P- in P-system is short for “parental.”
The 16 parental relationships (arcs) in Figure 3 are listed as follows:

- 0 - 0  mother to daughter
- 0 - 1  mother to son
- 0 - 01 mother to married daughter with husband
- 0 - 10 mother to married son with wife
- 1 - 1  father to daughter
- 1 - 1  father to son
- 1 - 01 father to married daughter with husband
- 1 - 10 father to married son with wife
- 01 - 0  mother-and-father to daughter
- 01 - 1  mother-and-father to son
- 01 - 01 mother-and-father to married daughter with husband
- 01 - 10 mother-and-father to married son with wife

The last four are equivalent (transposing mother-and-father) to:

- 10 - 0  father-and-mother to daughter
- 10 - 1  father-and-mother to son
- 10 - 01 father-and-mother to married daughter with husband
- 10 - 10 father-and-mother to married son with wife

Figure 4 shows a hypothetical example of a network of kinship relations drawn as a P-system. Arcs indicate parentage, and by convention, as indicated by the large downward arrow, all arcs are oriented downward from parents to children. The gender identifications of arcs are shown in the third column of Figure 3. When arcs go to nodes of type 01 or 10, to indicate a child who is married, we use the following conventions in drawing a level-1 graph such as Figure 4.

1) An arc to a couple from the upper left indicates that the left-most individual in that couple is the child, and hence the arc represents an individual of a particular gender. Hence we can see that arcs a and b in Figure 3 are male, while d and g are female.

2) Conversely, an arc to a couple from the upper right indicates that the rightmost individual in that couple is the child, and hence the arc again represents an individual of a particular gender. Hence we can see that c in Figure 4 is female and h, whose two marriages are indicated by two arcs labelled h, is male (see below).

3) Nodes of type 01 or 10 are equivalent in that both indicate a couple, but the significance of the order itself serves to indicate the gender of individuals represented by incoming arcs. If the order is not significant, we may use the equivalent symbol 2 for an unordered pair, 2 = |0,1| = |1,0|.

Each arc can be identified with the individual who is the son or daughter of the given parent or parents. Individual identities in Figure 4 are labeled by the letters a to h attached to arcs. Arcs that go to nodes of type 0 or 1 indicate single children: daughters and sons, respectively. There is a single arc for such individuals. An arc from mother to daughter, for example, can be identified with the daughter, and there may be multiple daughters descended from the same mother. If an individual is married twice, however, there will be two arcs descended from the same parents bearing labels for that individual and oriented towards the two different marriages, as shown by two arcs bearing the same label h in Figure 4. In general, several arcs for married individuals may bear the same labels.
The P-system incorporates the best features of each of the previous models: a single relation of parentage (Ore 1960), graphs embedded within the nodes of other graphs (Harary and Batell 1981), and segregation of higher level descent and marriage structure from nuclear family structure (White and Jorion 1992), which is also the key conceptual distinction used by Lévi-Strauss (1969) in his theory of marriage alliance. Hence, while a P-system is used to represent a concrete network of kinship and marriage relationships, this network also constitutes a system in the nested sense of Harary and Batell (1981). In sum, the connections between the nodes at level-1 in a P-system are especially useful in the analysis of marriage and descent, while at level-2 we can describe how individuals are embedded in the kinship structure.

If we replace the 0 and 1 labels within the nodes of Figure 4 with the conventional graphical symbols used in anthropological genealogies, viz., the circle and the triangle for females and males, respectively, we can create a conventional genealogical diagram, such as the one shown in Figure 5, which conveys the same information as Figure 4. The system of genealogical notation used in Figure 5 is due to Rivers (1910) and is still in use today by anthropologists. It does not define a proper graph or digraph, however, even allowing (as in a multigraph) for the two kinds of relations, one undirected (=) between spouses and the other directed downward from parents to child. Having two relations would not be a problem except that the parental second (asymmetric) relation goes from the first (symmetric) relation to one of the nodes. In graphs, edges can only go from nodes to nodes, not from edges to nodes, and similarly for digraphs.

By defining a P-system with two levels we are able to specify a structural model of kinship and marriage networks that contains digraphs at each level. At level-1 the relation of parentage is between nodes containing one or two persons, and at level-2 there is either a single node or a generalized “coupling” or marriage relation between individuals of opposite sex. Further, if we include the children descended from any of the nodes at level-1 in the level-2 relationships, as in a genetic digraph, then all nuclear family relationships are segregated at level-2 and all between-family relationships are at level-1. This is extremely useful for visualizing and analyzing the structure of kinship networks in general.

Figure 6 shows the underlying graph of the P-system in Figure 4. Each arc has been replaced by an edge (the direction of individual arcs has been removed). The graph contains a single cycle.

Compare Figure 6 with the Rivers-type diagram in Figure 7, which is the basis for Ore's (1960) genetic digraph in Figure 8. Figure 7 converts the relationships shown in Figures 4 and 5 into a pair of relations, parental and marital, among two types of nodes, black nodes for males and white nodes for females (hence Figure 7 is not strictly a digraph, although in terms of connections it is a "mixed graph" (Harary 1966), having both arcs and edges). The parental relation is defined by arcs between individuals (oriented downward as usual) rather than between a couple and a child. This multiplies
the number of parental arcs. The marital relation between individuals is shown by horizontal (darker) edges. To create a digraph with a single type of arc, Ore suppresses the distinction between male and female nodes and erases marriage bonds. The result is shown in Figure 8.

Figure 7: Rivers-type diagram

Figure 8: Genetic digraph

Figure 8 loses some of the information in the P-system of Figure 4: information is lost both about gender and marriage. Taking nodes g and h in Figure 8, for example, we know neither their gender nor the fact that they are married. We can restore information about gender through labeling, but analysis of the graph itself, without any gender distinctions, can tell us little about kinship structure. Figure 7, which adds gender and marriage to the genetic digraph, does so at the expense of different types of nodes and relations, a greater number of nodes and arcs and greater complexity of the representation.

The Meaning of Cycles

A path in a graph $G$ is a sequence of distinct nodes in which each sequential pair is adjacent in the graph. A directed path in a digraph $D$ is a sequence of distinct nodes in which each sequential $u,v$ pair is adjacent, i.e., by arc $(u,v)$. A cycle in $G$ is the union of a $u$-$v$ path of three or more nodes from $u$ to $v$ and a $uv$ edge. A directed cycle in $D$ is the union of a directed path of three or more nodes from $u$ to $v$ and a $(v,u)$ arc. A digraph is acyclic if it contains no directed cycles. P-systems derived from biological ancestries are acyclic digraphs (no one is her own ancestor).

If we ignore the direction of their arcs, the cycles in P-systems encode information about marriages within and between families. As an example of marriages within an ancestrally related family, the fact that cousins g and h have married is encoded in Figure 4 in the cycle with edges g-h-d-c-g. In Figure 7, this marriage between cousins is encoded in the cycle defined by nodes g-h-d-i-c-g (and also by cycle g-h-d-j-c-g; other cycles such as g-h-d-j-i-c-g do not imply the specifically consanguineal relation between g and h). In the genetic digraph of Figure 8, this marriage is not encoded at all. Shortly, we will examine and illustrate marriage cycles between families. In the underlying graph of a genetic digraph, cycles occur simply because of the existence of families with two parents and two or more children, such as the cycle e-c-f-b-e in Figure 8. These types of cycles do not occur at level-1, in P-systems.

A P-system is thus a more efficient coding of marriage patterns (reflected in cycles) than the genetic digraph. One might say that we have lost the information in a genetic digraph about cycles in the nuclear family. If we want to recover the relations among individuals as defined in the genetic digraph, however, we can draw at level-2 the appropriate genetic digraph for an individual, couple or nuclear family, as in Figure 9.
Figure 9. Coding the inner level in a P-system with an appropriate genetic digraph

With this combined representation, we have the following social science observations:

Observation 1: In the graph of a P-system, all cycles at level-1 are due to cycles created by marriages within or between families.

Observation 2: In the graph of a P-system, all cycles at level-2 are due to parental relations between two parents and two or more children within nuclear families, and are always of length 4. This observation instantiates Ore’s Theorem in an appropriate context.

Observation 3: If we extract the genetic digraphs of level-2 in a P-system and identify each set of nodes that represent a single individual (in several marriages), then we have a genetic digraph for the entire system. Ore’s theorem holds if we do this only for nodes representing two consecutive generations, disallowing Oedipal marriages (Harary 1982).

From P-Systems to P-Digraphs

Two further transformations move us from the P-system as a well defined mathematical structure to the P-graph of recent anthropological literature (White and Jorion 1992, 1996, Jorion 2000). First, we introduce a binary coding of the parental relation to indicate the gender of the offspring, and we define an appropriate type of graph to accommodate the binary coding. A signed graph (Harary 1953) is obtained from a graph by designating each edge as either positive (+) or negative (-). Figure 10 shows the four signed triangles. The negative edges are drawn dashed. A signed digraph (Cartwright and Harary 1956) is defined and drawn similarly.

Figure 10. The signed triangles

When the binary transformation to a signed graph is applied to the P-system in Figure 4, for example, we obtain a labeled P-graph as in Figure 11, where the solid arcs (+) are identified with females, and the dashed arcs (-) are identified with males. The assignment of signs by gender is arbitrary and can be reversed without loss of meaning to accommodate different examples. The large downward arrow indicates the direction of the arcs.

---

3 P- is a mnemonic for parental-graph or a graphe de parenté.
Second, since the level-2 coding of nodes into sets \( \{0\} \), \( \{1\} \) and \( \{01|10\} \) can be recovered from the structure of the digraph, we use a single type of node in the P-graph. Thus, we get the unlabeled P-graph of Figure 12, with the usual downward orientation of arcs. With proper instruction one may read types of marriage from patterns of the edges and their signs in cycles of the graph. A feature of Figure 12 is the marriage between cousins, which can be read from the four-node cycle as a marriage of a man with a mother’s sister’s daughter. All of the structural information in a P-system, such as that in Figure 4, is deducible from a corresponding P-graph, as in Figure 12.

> Figure 11: Labeled P-graph of a P-system (downward orientation of arcs)
> Figure 12: Corresponding unlabeled P-graph (solid lines for females, dotted for males)

**Relinking Marriages**

Cycles that occur in the level-1 graphs of a P-system may be created by marriage between two persons who are related by common descent. Anthropologists call these *consanguineal* marriages or marriages between blood relatives. The only other way that cycles occur at this level is described by anthropologists as *relinking* among a set of families who "marry in circles." They designate as *relinking marriages* those that create relinking among sets of families.

Different varieties of relinking can be given a series of graph theoretic definitions. A *subdigraph* of a digraph \( D \) is a subset of nodes in \( D \) together with the arcs between them, as illustrated in Figure 13 for digraphs and their graphs at P-system level-1. A *subgraph* of a digraph \( D \) is a subset of nodes in \( D \) where edges are substituted for arcs. An *ancestral* node of a subgraph of a digraph \( D \) is one that has no indegree in \( D \). A proper or two-family relinking in the level-2 graph (or P-graph) of a P-system is a

> Figure 13. Types of relinking as exemplified in subgraphs of P-systems
cycle that contains two ancestral nodes. One such relinking is exemplified in Figure 13. By extension, a *three-family relinking* is a cycle in the level-1 graph of a P-system that contains three ancestral nodes, and similarly for more than three families. By further extension, a consanguineal marriage is a relinking within a single family, although this is not a properly anthropological term. Figure 13 contains an example of marriage between cousins. The subgraphs in Figure 13 are far from the only examples of each type, since ancestral nodes relate to descendants in many ways and at different generational depths. The illustrations for consanguineal and two-family relinking are at depth 2 (relinkings among cousins), but the illustration for three-family relinking contains a combination of common ancestors at depth 2 (cousin) and depth 1 (sibling).

The P-graph in Figure 14 exemplifies relinking marriages among couples with common ancestors at shallow generational depth. Every node with two incoming arcs represents either a (consanguineal) marriage with a blood relative or one that relinks families in a cycle of marriages. There are many relinkings here between different sets of siblings. Darker lines represent men while lighter lines represent women, opposite to Figures 11 and 12: such adjustments of P-graph representations accommodate the needs of different studies. In this case, taken from the kinship network of a nomadic clan in Turkey studied by Johansen and White (forthcoming, White and Johansen ms.), only descent groups in the male line are recognized (labels on the lines might include codes for the historical generations, migration histories, lineage numbers and first initials of individuals). Sets of nodes with different shadings within the two large circles are two sets of relinked couples. These two sets share one parental node in common that contributes children to each of the relinked subsets, but there is no relinking between the two sets.

**Figure 14.** Relinking marriages in a P-graph

### Representing Complex Structures

A P-system and its corresponding P-graph, loosely designated, is a simple but sufficiently rich structure to provide a condensed structural representation of a kinship and marriage network, and when its nodes and arcs are fully labelled, it contains all of the information for an underlying P-system. We can identify patterns of intermarriage by studying the kinds of cycles in which marriages are contained. It can be see from Figure 14, for example, that marriages with parallel cousins in the same patriline are very common. We can also study patterns of marriages between families or lineages. In the study from which this example is taken (White and Johansen ms.), P-graphs are constructed for the entire society and analyzed for changing structural properties in successive time periods.
Mathematical Properties of P-graphs

The underlying P-digraph of a P-graph is the digraph produced by treating all arcs as belonging to a single generic type. Figure 15 shows the underlying P-digraph for the P-graph in Figure 12.

The underlying P-digraph of a P-system representing a biological kinship network plus marriage ties has these properties:

1. It is asymmetric and acyclic.
2. The maximum indegree of nodes is 2.

The signed P-digraph (P-graph) of a P-system representing a biological kinship network plus marriage ties has the additional property:

3. For arcs of each sign the maximum indegree of nodes is 1.

Mathematically speaking, then, a P-graph is a signed digraph for which the three axioms above are true. These properties derive from biological kinship. Parentage is not a symmetric relation, and does not permit a directed cycle of ancestry; any given couple or marriage has no more than two sets of parents; biologically, one parent is male and one female. P-graphs having these axioms have been extensively used to represent and analyze the structure of genealogical data collected in field studies (Brudner and White 1997, Houseman and White 1996, 1998a, 1998b, White 1997, White and Jorion 1996, White and Schweizer, 1998). These properties may also provide axioms for culturally defined parentage relations under appropriate circumstances, or they may be modified to take culturally defined parentage relations into account. What has been lacking to date and presented here is a precise mathematical formulation of the underlying type of graph-theoretical structure for kinship: the P-system. The P-graph used in anthropology, more specifically, is level-1 of a P-system in which arcs are given signs appropriate to the gender of individuals at level-2.

The P-system captures the details of marriage structure, and can represent the systems of marriage rules studied by Weil (1969), with the benefit of generations situated in time. It captures the parental relationships of the genetic graph for empirical networks, but in a much more parsimonious form that is suitable for the analysis of kinship and marriage structure, or to study changes in structure through time.
References


