Convergence Analysis of Communication Networks*

William D. Richards, Jr. and **Andrew J. Seary**
School of Communication
Simon Fraser University
Burnaby, B.C.
Canada

We begin with some of the central ideas from General Systems Theory — notions about the relation between parts and wholes, the source of emergent properties, and the necessity of approaching the entire set of relationships between the parts as a unitary thing. We take a few fundamental theories about how people process and make sense of information and, by putting them into the context of communication networks, show how the consequences of taking a systems approach leads to some most interesting and, we think, useful tools that can be used in the study of social networks in a wide range of situations.

0. SYSTEMS THEORY

In the late 1960's articles with the phrase "A Systems Analysis of …" or "A systems Approach to …" in their title began to appear with increasing frequency. A decade later, while there was a large body of literature dealing with uses for or applications of systems thinking, as well as a body of writing about the systems approach in general, theoretical terms, there was much less in the way of systems methodologies (Richards, 1976, p. 3). The systems approach was an alternative to the classical analytic method, a particular epistemological approach to understanding the universe. Rappaport and Horvath describe analysis as

> … an attempt to understand a complexity by examining its constituent parts. The parts being simpler, they are supposedly more amenable to understanding. The idea of analysis, then, is to understand the working of the parts . . . . The implied hope is that it is possible to 'build up' the understanding of a complexity by 'superimposing' the workings of the various parts" (1968. P. 87)

> [The relations between the parts] . . . . are determined one by one. These separate determinations are made possible by the method of controlled experiment. In order to bring out some causal relations free of disturbance by other factors, we deliberately try to hold constant all those factors suspected of having some influence. Thus the basic assumption underlying the empirical study of physical phenomena is that we can eliminate all disturbing phenomena and study the relation of interest alone. Next, by establishing several pairs of such relations, we can (we assume) combine them into a more general causality law, that is, an equation in which all the contributing factors appear as variables. This is called the *analytic* method. It has been phenomenally successful in the physical sciences (Rappaport, 1968, p. xiv).

The process as described proceeds in two fairly separate stages. First, the whole "complexity" is *decomposed*. That is, it is broken down into parts and the relations between the parts are examined one at a time, in isolation from one another and from any "disturbing phenomena". Second, an understanding of the whole is *synthesized* by "combining" or "superimposing" the partial descriptions. The idea here, of course, is that the whole is nothing more than the sum of its parts, an assumption that may be correct in simple mechanical systems, but not in more complex ones. Buckley complained that sociological theory had for decades been dependent on two conceptual models — the mechanism (the mechanical device which obeys the laws of physics) and the organism (the living, growing, evolving plant or animal). He said that "... current dominant theory is built on mechanical and organic (more exactly, organismic) systems models laid down during previous centuries" and which are quite inappropriate in dealing with the kinds of systems we are interested in (1967, p. 1).

Deutsch (1968) argued that the use of these models in attempting to understand social situations was consistent with advances in the physical and natural sciences, where these models were very successful. According to Monge, "the breakthrough by the natural sciences in developing viable conceptualizations inspired social scientists to adopt models based on the natural science conceptualizations of the world, all of which was done in hopes of achieving similar success in explaining human behavior; this, physical and biological science success bred social science imitation. Unfortunately, the hoped-for success has never been realized; social and communication science are still without a viable model of human behavior" (1973, p. 7).

The classical analytic method had the most difficulty when the situation under investigation involved closed causal loops, as, for example, did all systems with operating feedback loops. Two explanations for this failure were given. According to the vitalistic view, the classical analytic method failed because "living processes are

not governed by the same laws as nonliving processes." The other explanation, supported by those who took a more mechanistic or reductionistic approach, was that the problem was "the tremendous complexity of living processes" (Rapoport, 1968, p. xv).

The mechanistic approach is opposed by the organismic one in a dialectic where fundamental views of reality are contested. Where the organismic approach resulted from the opposition of the mechanistic method to situations not sutied to mechanical analysis, the systems approach comes from the dialectic. The solution to the animate vs. inanimate (i.e. mechanism vs. organism) dilemma was proposed by De la Mettrie in 1747 when he suggested that

> . . . matter was in itself neither organic nor inorganic, neither living nor dead, neither sensible nor insensible. The difference between these states or properties of material things sprang, not from the intrinsic natures of their raw materials, but from the different ways in which these materials were *organized* (in Toulmin and Goodfield, 1962, p. 318).

Although he had no suggestions on how to study systems, De la Mettrie got very close to what has turned out to be a crucial concept in modern systems theory — organization — the set of relationships among the parts of the system. In the decomposition stage of analysis, things are very much taken apart and taken out of the context in which they ordinarily exist. The problem with this is that the isolated, decontextualized variables and parts are not the same as they were before decomposition. The characteristics of complex systems owe part of their nature to the fact that they coexist and interact with one another, in the context of the other characteristics and relationships. This is in part a consequence of the fact that systems are not aggregates of independent individuals, but rather systems of interdependent interacting (communicating) individuals, where the behavior of individuals is altered both because it takes place and is interpreted in the context provided by the others in the group.

Furthermore, we have the behavior of the higher levels in the system to contend with. Gerard argues that ". . . the class [of individuals] is, of course, a kind of individual; and the more the members of the class interact — even to the extent of developing into differentiated subclasses — rather than coexist, the more does the superordinate group become a true individual rather than a collection of ordinate individuals "(1968, p. 53).

At higher levels of analysis, then, we have entirely new variables — the ones associated with what have been called the "emergent properties" of the system. The observer who shifts attention from one level of organization to the next . . . "expects to find obedience to all of the laws of the levels below. But upper levels of organization require specification of the *arrangement* of the lower units, which in turn generates richness and the basis of new and unexpected principles" (Wilson, 1977, p. 137, italics added). These emergent properties have their origins in the interactions among the individuals making up the system. A system of independent individuals is simply the sum of the individuals. Since the members of such a set are not altered by being members, there are no emergent properties.

In contrast, a system of interacting individuals includes not only the individuals but also *the relationships among them*. Since the patterns of interaction are included, it is possible to speak of the behavior of the system in terms of the interactions of the members. In a communication network, for example, it is possible to speak of the extent to which the network as a whole exhibits structural differentiation. Organizations with greater decentralization and less bureaucratization have been found to be more innovative in adopting new ideas and hence more effective (Zaltman et al., 1973). Other studies have shown that the communication structure of a system is a determinant of its performance and of its innovativeness (Lin, 1966, 1968; Allen, 1970; Shoemaker, 1971). Network-based proximity mechanisms in organizations have been studied by Dow (1988), Erickson (1988), Johnson and Miller(1986), Tushman & Romanelli (1983), and Dean & Brass (1985). Most of this research has focussed on various aspects of the structure of networks; there has been relatively little work done on the dynamic behavior of networks (e.g. Barnett, 1988; Tutzauer, 1985). These are examples of emergent properties that don't even make sense for either isolated individuals or sets of independent individuals.

I. PART 1: CONVERGENCE THEORY

Convergence theory (Kincaid, 1987, 1988; Rogers and Kincaid, 1981; Kincaid and others, 1983; Barnett, G.A. & D.L. Kincaid 1983; Becker, C. 1993), simply stated, says that the more two people (or a larger group of people) communicate with each other, the more similar their views of the world become. More precisely, "unrestricted information flow within the boundaries of a relatively closed communication system will lead to a convergence of beliefs, values, and behavior toward a state of greater uniformity or negentropy" (Kincaid, 1993, p. 132).

Erickson has argued that people are most likely to have attitudes similar to those with whom they have stronger ties (Erickson, 1988, p. 115). Laumann & Marsden (1979) showed the opinions and attitudes of members of dense networks to be more similar than those of less dense networks. Danowski (1980) confirmed this and showed the effect to be stronger in networks where the organizational activity is more closely related to the attitudes and beliefs of their members. Since, according to Festinger's (1954) social comparison theory, a high level of direct interaction with others similar to oneself would lead to convergence, one would expect to see convergence in highly cohesive cliques or subgroups (Kincaid, 1993, p. 132). If the system is closed to interaction with the environment, the network's members would be expected to eventually converge to a single position. But few, if any, social networks are closed systems. The more the members of an open system interact, the more likely the system is to develop into differentiated subsets — "clusters", "cliques", or "groups" — which become individuals in their own right, rather than simple aggregates (Barnett, 1973).

The convergence model turns out to be of particular importance to communication researchers for a number of reasons, some of which are clear and easily understood, while others are more complex and subtle. This chapter explores the convergence model, develops its consequences, and examines its value as an analytic approach and a theoretical heuristic.

A. Two approaches to convergence

There are two ways in which a system of interacting individuals can be seen as converging. In the first, the focus is primarily on concepts. The entire set of concepts can be seen as a set of points in a multidimensional space, where their locations describe both their relationship to the underlying cognitive dimensions and their distances from one another. The meaning of a concept is found in its relation to other concepts: if a pair of concepts have similar meanings, they will be close together in the space. For example, spaghetti and macaroni would probably be quite close to one another, as would Haydn and Mozart, but politics and automatic transmissions wouldn't. When two concepts move towards one another, their meanings become more alike. If they are located at the same place, their meanings are identical — they have completely converged. With this approach, the focus is on the entire set of concepts and the distances between them — in other words, on the configuration of the set of concepts. Since the location of a set of points in multidimensional space can be described by its matrix of coordinates, convergence for a pair of people would mean that differences between their matrices (their conceptual configurations) become small or disappear. In a converging system, the between-persons variance declines.

The mechanism by which this happens is communication between the people. The more communication there is between a set of people, the faster and more completely the differences between their ideational configurations decline. This is the approach taken by Kincaid (1987); Kincaid, Yum, Woelfel & Barnett (1983); etc. Their model (Woelfel & Fink, 1980; Barnett & Woelfel, 1988) is based on Linear Forced Aggregate Theory, which uses Newtonian laws of motion to describes the effect communication has on people's conceptual configurations (Woelfel, J. & Saltiel, J., 1978). Their approach focusses on concepts and aggregates people; it treats the set of individuals in the network as a more-or-less uniform statistical mass, while it describes the coordinates of concepts as they move in the space. They (Kincaid, Yum, Woelfel, Barnett, p.63) do this because ". . . it would be impossible to account for relative positions and changes of all the potential (n(n-1)/2) dyadic pairs of individuals in a given, naturally occurring social system of any substantial size." The starting point for analysis is data obtained by asking respondents to judge the differences between pairs of concepts. Interaction

patterns of the individuals are not examined. Since this model ignores the specific patterns of interactions among individuals, it is more a statistical theory than a systemic model.

A second approach to convergence has its primary focus on individual people and their social interactions. Where the first approach looked at the meaning of a concept as its location in a multidimensional space, the second locates people in a similar space, where a person's location describes the individual's position along the underlying cognitive dimensions. When the locations of two people converge, their ideational configurations become more alike. When two individuals are located in the same place, their configurations are identical — they have completely converged. When a pair of individuals interact, their ideational configurations become more similar. As in the concept-centered approach, the more communication there is between a set of people, the faster and more completely the differences between their configurations will decline. This model is based, ultimately, on Heider's theory of cognitive balance. Heider argued that ". . . a good deal of interpersonal behavior and social perception is determined — or at least co-determined — by simple cognitive configurations" (1946, p. 111). "Unbalanced" configurations involve implicit contradictions and are unstable. They tend to be converted to stable, "balanced" configurations by the introduction of changes that resolve the contradictions. These changes may require either a shift in position or the formation or elimination of a relationship with another person. If the interaction patterns with other people remain constant, it is the conceptual positions that will be adjusted to bring things into balance. This, we believe, is where convergence "happens."

Unlike to the researchers who take the concept-centered approach, we believe we *can* "account for relative positions and changes of all the potential (n(n-1)/2) dyadic pairs of individuals ..." Later in this chapter, we describe in considerable detail various formulations of the second approach and show how it forms the basis for both a theoretical model of significant utility and an analytic procedure with unusual descriptive power.

The two approaches are complimentary duals in the sense that the first aggregates and more-or-less ignores the people and their interaction patterns while it describes the movement of concepts as their meanings change; in contrast, the second approach aggregates and more-or-less ignores the individual concepts, while it describes the movement of people as their world-views change. The first approach uses Newtonian and thermodynamic laws to describe the movement of concepts in terms of elastic solids, while the second uses psychological and cognitive theories of how people evaluate and organize attitudes to explain how they change their attitudinal configurations. The first describes the trajectories of moving concepts, while the second describes the interactions between people and the effect these interactions have on the ideational configurations of the people. The first is a statistical description, while the second is a systemic model.

B. A mathematical formulation of convergence theory

One of the earliest formulations of what is known today as the convergence model was described in a paper by French (1956) entitled "A formal theory of social power". He extended the theories of Heider and Newcomb, which described two-person groups, and presented a theory of interpersonal influence in N-person groups. French begins his model by describing a two-person situation in which person *A* expresses an opinion or argues for it in a way that influences person *B*. Assume *A*'s attitudinal position is *a* and *B*'s position is *b*. *A*'s advocacy will exert a force pushing *B* towards *A*. At the same time, *B* will resist the change and exert a force in the opposite direction. The forces will balance at the point ½ $(a + b)$. As a consequence of these forces, *B*'s position will move toward the point of equilibrium.

> "Similar calculations of the resultant force and consequent changes can be made for *A* and for groups with any number of members by placing all members on the same scale and by assuming that the gradient of forces around each member represents both forces he can induce on others and forces he can set up as resistance against others" (p. 37).

The forces individuals exert in this process are a function of the amount of "interpersonal power" the individuals have over one another. The basis of this power, French argues, is the "more or less enduring relationship" between the individuals. A number of mechanisms have been suggested to explain how being in a relationship

with another gives rise to interpersonal power. Back (1951), French and Snyder (1957), and Lippitt, Polanski, and Rosen (1952) demonstrated that interpersonal *attraction* — *A*'s liking for *B* — is a basis for interpersonal influence. If *A* is seen by *B* to have superior knowledge and information, *A* will have *expert* power, as was demonstrated by Hovland and Weiss (1952) and others. Heider (1946) and Newcomb (1953) developed Balance Theory, "a more general conception which combines several types of power" (French, p. 38), and one of the most influential and enduring models of the dynamics of cognitive change.

French's model can be stated as an equation that describes what happens to the conceptual positions of the members of a communication network in each increment of time:

$$P_i' = \frac{1}{2}\left(P_i + \frac{\sum_{j \neq i} P_j A_{i,j}}{\sum_j A_{i,j}}\right) \tag{1}$$

The equation says that the new position of person *i* is the mean of her present position (P_i) and the weighted mean of the positions of the people in the system with whom she interacts (P_j). The weights ($A_{i,j}$) indicate the strengths of the relationships between pairs of individuals. The matrix of weights, *A* is simply the scalar adjacency matrix. After each round of calculations, each person's position is adjusted to the new value, and another set of means is calculated. In the equation shown above, individuals move to the equilibrium point, halfway between their current positions and the weighted means of the positions of the people with whom they interact.

```
      0 0 0 0 1 0 0 0 1 1 0 0 1 0 0 0 0 0
      0 0 0 0 0 0 0 1 0 0 0 0 1 1 0 1 0 0
      0 0 0 0 0 1 1 0 0 0 0 1 0 0 0 0 1 1
      0 0 0 0 1 0 0 0 0 0 1 0 0 0 1 0 1 0
      1 0 0 1 0 1 0 0 0 1 0 0 0 0 0 0 1 0
      0 0 1 0 1 0 1 0 1 0 0 1 0 0 0 0 0 1
      0 0 1 0 0 1 0 0 0 0 0 0 0 0 0 0 0 1
      0 1 0 0 0 0 0 0 1 0 0 0 0 1 0 1 0 0
A =   1 0 0 0 0 1 0 1 0 1 0 1 0 0 0 0 0 0 0
      1 0 0 0 1 0 0 0 1 0 0 0 1 0 0 0 0 0
      0 0 0 1 0 0 0 0 0 0 0 0 0 0 1 0 1 0
      0 0 1 0 0 1 0 0 0 0 0 0 0 0 0 1 0 1
      1 1 0 0 0 0 0 0 1 0 0 1 0 1 0 0 0
      0 1 0 0 0 0 0 1 0 0 0 1 0 0 1 0 0
      0 0 0 1 0 0 0 0 0 1 0 0 0 0 0 0 1 0
      0 1 0 0 0 0 0 1 0 0 0 1 1 1 0 0 0 0
      0 0 1 1 1 0 0 0 0 0 1 0 0 0 1 0 0 0
      0 0 1 0 0 1 1 0 0 0 0 1 0 0 0 0 0 0
```
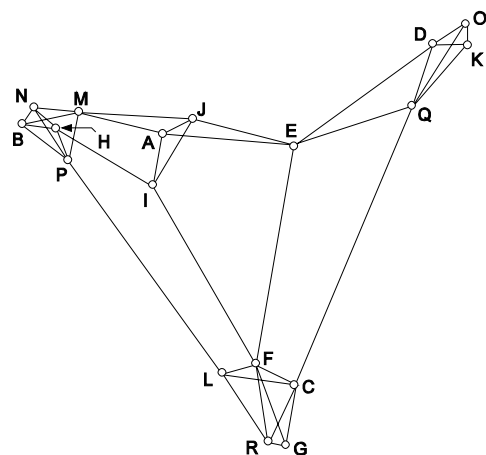

Figure 1.   A network and its binary adjacency matrix. This network is used throughout the chapter to illustrate the model.

In the first time period, individuals will experience only the forces from those with whom they interact directly. In subsequent times, however, they will be influenced indirectly by those with whom they do not interact. This happens because each individual's new position is a function of *all* the individuals with whom she interacts. Each of those individuals are also influenced by those with whom they interact, and their positions will be changed by the end of the first time period, which will be reflected by the influences they have in the second time period. The model thus describes an *iterative* process in which the effects of each round of interactions affect all subsequent interactions.

To observe the behavior of the model, an adjacency matrix, *A*, for a connected network[1], and a set of initial positions, *P*, are fed into Equation 1 and iterated. If the relationship connecting individuals in the network is symmetrical or bidirectional, the positions of individuals will become closer and closer to one another until they appear to converge to a single point, *regardless of the initial positions*. The results produced by three different starting configurations are shown in Figure 2. In every case, the point to which they converge is *the mean of the*
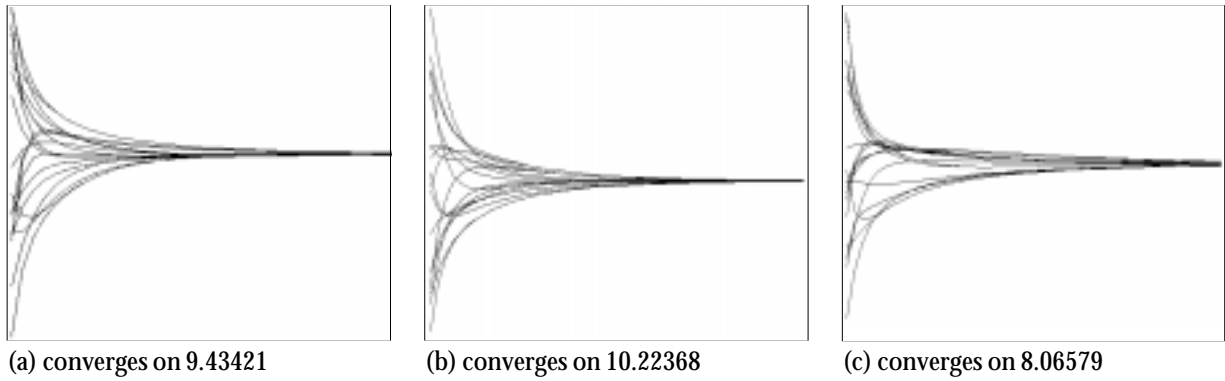
(a) converges on 9.43421   (b) converges on 10.22368   (c) converges on 8.06579

Figure 2. Results produced with three different starting configurations for the sample network *A*

original positions of the members of the network, weighted according to the total amount of interaction each member has (i.e., according to the row sums of *A*).

Note that only the amount of interaction (in a binary network, only the *number* of connections) each individual has matters; the specific pattern of connections among the members of the network is irrelevant. Although these results are similar to the ones French obtained with a closed system in which there was no interaction between members of the system and the environment, they don't adequately describe what happens in real social networks, where there is almost always some ideational variety, even among individuals who have been in very close contact for decades.

The apparent convergence to a single point is due to two things. First, the convergence is only *apparent*. If the calculations are done with sufficient precision, it will be clear that the convergence is not complete, but the positions of the individuals relative to one another usually cease to change after a few dozen iterations. Second, the behavior of the model is dominated by the overall contributions of individuals to the total amount of interaction (i.e., the row sums of *A*), rather than by the specific patterns of interactions among individuals. In this sense, it is a gross statistical model that describes a property of an aggregated set of individuals, rather than a systemic model that describes the specific pattern of interactions among the individuals in the network.

C. Convergence is a Markov process

Near the end of his paper, French developed a matrix formulation of two aspects of his model. He discussed the implications of raising the adjacency matrix to higher powers. He also presented a matrix formulation that shows the amount of influence each individual has on all individuals at any point in the process. Neither of these formulations, however, actually expressed the essence of the model. A number of properties of the model become apparent if the process described in Equation 1 is expressed in a matrix form. This can be done by making two changes to *A*. First, calculate the sum of the elements in each row and put them on the diagonal of the matrix. Second, divide the contents of each cell by twice its row sum. The result is the matrix *M*:

$$
M = \begin{bmatrix}
\dfrac{Rsum_1}{2\,Rsum_1} & \dfrac{A_{1,2}}{2\,Rsum_1} & \dfrac{A_{1,3}}{2\,Rsum_1} & \cdots \\[2ex]
\dfrac{A_{2,1}}{2\,Rsum_2} & \dfrac{Rsum_2}{2\,Rsum_2} & \dfrac{A_{2,3}}{2\,Rsum_2} & \cdots \\[2ex]
\dfrac{A_{3,1}}{2\,Rsum_3} & \dfrac{A_{3,2}}{2\,Rsum_3} & \dfrac{Rsum_3}{2\,Rsum_3} & \cdots \\[2ex]
\vdots & \vdots & \vdots
\end{bmatrix}
=
\begin{bmatrix}
\dfrac{\sum A_{1,j}}{2\sum A_{1,j}} & \dfrac{A_{1,2}}{2\sum A_{1,j}} & \dfrac{A_{1,2}}{2\sum A_{1,j}} & \cdots \\[2ex]
\dfrac{A_{2,1}}{2\sum A_{2,j}} & \dfrac{\sum A_{2,j}}{2\sum A_{2,j}} & \dfrac{A_{2,3}}{2\sum A_{2,j}} & \cdots \\[2ex]
\dfrac{A_{3,1}}{2\sum A_{3,j}} & \dfrac{A_{3,2}}{2\sum A_{3,j}} & \dfrac{\sum A_{3,j}}{2\sum A_{3,j}} & \cdots \\[2ex]
\vdots & \vdots & \vdots
\end{bmatrix}
$$

If the vector *P* is multiplied by the matrix *M*, the resulting value of $P_i{'}$ will be given by:

$$P_i^{'} = P_1\left(\frac{A_{i,1}}{2\sum A_{i,j}}\right) + P_2\left(\frac{A_{i,2}}{2\sum A_{i,j}}\right) + \ldots + P_i\left(\frac{\sum_j A_{i,j}}{2\sum A_{i,j}}\right) + \ldots + P_n\left(\frac{A_{i,n}}{2\sum A_{i,j}}\right)$$

$$= \frac{P_i}{2} + \frac{\sum_{j\neq i} P_j A_{i,j}}{2\sum A_{i,j}} = \frac{1}{2}\left(P_i + \frac{\sum_{j\neq i} P_j A_{i,j}}{\sum A_{i,j}}\right),$$

which is Equation 1. The results obtained by applying Equation 1 to the whole network can thus be obtained by multiplying the vector of original positions by the matrix *M*:

$$P^{'} = P \times M$$

The effect of two rounds of interactions is given by:

$$P^{''} = P \times M \times M$$

There are two important things to note about this matrix: first, it is *row-normalized* — the entries in each row of *M* sum to 1.0; second, because Equation 1 gives a weight to each node's position equal to the sum of the weights of all the other nodes' positions, the entries on the diagonal are all 0.5. The matrix *M* for the sample adjacency matrix *A* is shown on the top of the next page.

The operation that takes place in a round of interactions is described by multiplying the vector of positions by the matrix. The result of ten rounds of interactions will be given by multiplying the starting vector of positions by the matrix ten times, or by the matrix raised to the tenth power. The interaction process, as formulated in this operationalization of convergence theory, is thus a Markov process. Figure 3 shows what happens to the positions of the 18 members of the network on the previous page over 50 rounds of interactions.

```
         .5     .0    .0    .0    .125 .0    .0    .0    .125 .125 .0    .0    .125 .0    .0    .0    .0    .0
         .0     .5    .0    .0    .0   .0    .0    .125 .0   .0    .0    .0    .125 .125 .0    .125 .0    .0
         .0     .0    .5    .0    .0   .1    .1    .0    .0   .0    .0    .1    .0    .0    .0    .0    .1    .1
         .0     .0    .0    .5    .125 .0    .0    .0    .0   .0    .125 .0    .0    .0    .125 .0    .125 .0
         .1     .0    .0    .1    .5   .1    .0    .0    .0   .1    .0    .0    .0    .0    .0    .0    .1    .0
         .0     .0    .083  .0    .083 .5    .083 .0    .083 .0    .0    .083 .0    .0    .0    .0    .0    .083
         .0     .0    .167  .0    .0   .167  .5    .0    .0   .0    .0    .0    .0    .0    .0    .0    .0    .167
         .0     .125  .0    .0    .0   .0    .0    .5    .125 .0    .0    .0    .0    .125 .0    .125 .0    .0
M =    .125    .0    .0    .0    .0   .125  .0    .125 .5   .125 .0    .0    .0    .0    .0    .0    .0    .0
       .125    .0    .0    .0    .125 .0    .0    .0    .125 .5    .0    .125 .0    .0    .0    .0    .0    .0
         .0     .0    .0    .167  .0   .0    .0    .0    .0   .0    .5    .0    .0    .167 .0    .167 .0
         .0     .0    .125  .0    .0   .125  .0    .0    .0   .0    .0    .5    .0    .0    .0    .125 .0    .125
         .1     .1    .0    .0    .0   .0    .0    .0    .0   .1    .0    .0    .5    .1    .0    .1    .0    .0
         .0     .125  .0    .0    .0   .0    .0    .125 .0   .0    .0    .0    .125 .5    .0    .125 .0    .0
         .0     .0    .0    .167  .0   .0    .0    .0    .0   .0    .167 .0    .0    .0    .5    .0    .167 .0
         .0     .1    .0    .0    .0   .0    .0    .1    .0   .0    .0    .1    .1    .1    .0    .5    .0    .0
         .0     .0    .1    .1    .1   .0    .0    .0    .0   .0    .1    .0    .0    .0    .1    .0    .5    .0
         .0     .0    .125  .0    .0   .125  .125 .0    .0   .0    .0    .125 .0    .0    .0    .0    .0    .5
```

The ith row of matrix *M* describes the relative impact each node in the network has on the ith node's position after a round of interactions. For example, in the first row, (.5, .0, .0, .0, .125, .0, .0, .0, .125, .125, .0, .0, .125, .0, .0, .0, .0, .0), the .5 in the first column indicates that 50% of the first person's new position is determined by her current position. The .125 in the fifth, ninth, tenth, and thirteenth columns indicate that 12.5% of the first person's new position is determined by the current positions of the fifth, ninth, tenth, and thirteenth persons. *M* can thus be seen as a *Markov* or *stochastic* matrix — a matrix of transition probabilities from one set of positions to another, where the positions represent the cognitive configurations of the members of the network.

This appears to be the outcome French predicted and obtained when he studied a number of small social systems. It seems to be the common-sense result that one would expect to see if convergence theory is valid. However, this result shows only the mass behaviour of the system — the behavior that is due to the proportions of interaction associated with the individual members of the network (the row sums of A). Since this behavior is completely specified by the marginals of the adjacency matrix, it is equivalent to the "expecteds" in the calculation of chi-square, which are calculated on the basis of the assumption that the row variable is independent of the column variable. In the current situation, to assume this kind of independence would be the same as assuming that the particular pattern of connections in the network is irrelevant, something that we are reluctant to do, since it is precisely that pattern which is the object of our attention.
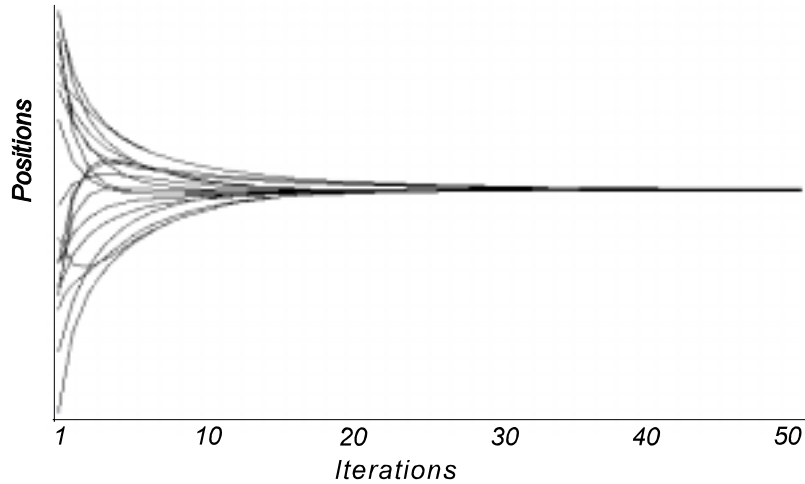

Figure 3. The positions of 18 members of the network in Figure 1, over 50 rounds of interactions.

As is the case in the interpretation of crosstabulation tables, we focus on the difference between the observed and the expected. That is, we subtract the expecteds from the matrix and examine what is left over. This is analogous to subtracting the "background" and studying the "signal." A different line of reasoning suggests how we might achieve this goal: Since we are interested in the positions of individuals with respect to the positions of others in the network, it would be reasonable to choose a coordinate system that describes only their *relative* positions. This can be accomplished by placing the mean position — the "center of mass" — at the origin by subtracting the mean from all positions before each iteration. Because individuals with many links will have a larger impact on the positions of others than individuals with few links, individuals are weighted by the row sums — the number of links or amount of interaction they have. This relocation of the origin will preserve information about differences between the positions of network members.

The revised model thus begins with:

$$P_i \ = \ P_i \ - \ \frac{\sum_i P_i\, Rsum_i}{\sum_i Rsum_i} \qquad \text{where} \qquad Rsum_i \ = \ \sum_j A_{i,j} \qquad\qquad 2.0$$

This change does not affect the fact that the range of positions grows smaller with each successive iteration, so the results are still somewhat difficult to work with because of the shrinkage of the range of values. This problem can be remedied if the range of values is expanded so it has a constant length (e.g., 2.0) before each iteration. To do this, each P_i is multiplied by a ratio calculated by dividing the desired length by the difference between the highest and lowest values obtained on the previous iteration (Equation 3). This adjustment, like

the previous one, is purely for convenience; it does not affect the relative positions of the P_i.

$$P_i = P_i \frac{2.0}{range} \tag{3}$$

The net effect of the transformations in Equations 2 and 3 is analogous to following the changing positions with a video camera with a zoom lens; Equation 2 moves the camera to keep the image in the center of the screen, while Equation 3 zooms in to keep the image a constant size. When the shrinkage of the range is controlled by Equation 3, the appearance of the results (see Fig. 4) differs considerably from the earlier ones. Nevertheless, if the calculation of Equation 1 is carried out to a sufficient level of precision, the results correlate highly (r = .99999999) with the results produced with Equations 2 and 3.
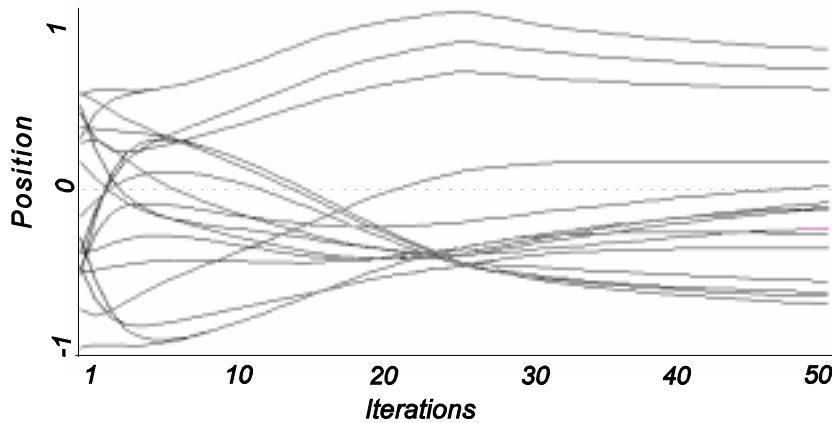
Figure 4. The positions of the nodes plotted against iteration number.

The adjustment produced by Equation 2 can be incorporated into the matrix M by subtracting the adjustment specified for P_i from each element in column i of M, which is equivalent to the $O - E$ step of calculating chi-square:

$$M_1 = \begin{array}{ccc} \dfrac{Rsum_1}{2\,Rsum_1} - \dfrac{Rsum_1}{\sum\limits_j Rsum_j} & \dfrac{A_{1,2}}{2\,Rsum_1} - \dfrac{Rsum_2}{\sum\limits_j Rsum_j} & \cdots \\[4ex] \dfrac{A_{2,1}}{2\,Rsum_2} - \dfrac{Rsum_1}{\sum\limits_j Rsum_j} & \dfrac{Rsum_2}{2\,Rsum_2} - \dfrac{Rsum_2}{\sum\limits_j Rsum_j} & \cdots \\[4ex] \dfrac{A_{3,1}}{2\,Rsum_3} - \dfrac{Rsum_1}{\sum\limits_j Rsum_j} & \dfrac{A_{3,2}}{2\,Rsum_3} - \dfrac{Rsum_2}{\sum\limits_j Rsum_j} & \cdots \\[2ex] \vdots & \vdots & \end{array}$$

The adjusted matrix for the example is shown on the top of the next page for comparison. This matrix is similar to the original Markov matrix, M, in that the effect of one round of interactions can be obtained by:

$$P' = P \times M_1$$

and the effect of two rounds by:

$$P'' = P \times M_1 \times M_1 = P \times M_1^2$$

and so on, but it differs in that it is no longer a Markov matrix — the entries in the rows no longer sum to 1.0. If the correction described in Equation 3 were added to M_1, the adjusted version of M, the results shown in Fig. 4 would be identical to the ones that would be obtained by multiplying the vector of positions by the result of

multiplying the matrix by itself 50 times.

$$M_1 = \begin{matrix}
\mathbf{.447} & -.053 & -.066 & -.053 & .059 & -.079 & -.039 & -.053 & .072 & .072 & -.039 & -.053 & .059 & -.053 & -.039 & -.066 & -.066 & -.053 \\
-.053 & \mathbf{.447} & -.066 & -.053 & -.066 & -.079 & -.039 & .072 & -.053 & -.053 & -.039 & -.053 & .059 & .072 & -.039 & .059 & -.066 & -.053 \\
-.053 & -.053 & \mathbf{.434} & -.053 & -.066 & .021 & .061 & -.053 & -.053 & -.053 & -.039 & .047 & -.066 & -.053 & -.039 & -.066 & .034 & .047 \\
-.053 & -.053 & -.066 & \mathbf{.447} & .059 & -.079 & -.039 & -.053 & -.053 & -.053 & .086 & -.053 & -.066 & -.053 & .086 & -.066 & .059 & -.053 \\
.047 & -.053 & -.066 & .047 & \mathbf{.434} & .021 & -.039 & -.053 & -.053 & .047 & -.039 & -.053 & -.066 & -.053 & -.039 & -.066 & .034 & -.053 \\
-.053 & -.053 & .018 & -.053 & .018 & \mathbf{.421} & .044 & -.053 & .031 & -.053 & -.039 & .031 & -.066 & -.053 & -.039 & -.066 & -.066 & .031 \\
-.053 & -.053 & .101 & -.053 & -.066 & .088 & \mathbf{.461} & -.053 & -.053 & -.053 & -.039 & -.053 & -.066 & -.053 & -.039 & -.066 & -.066 & .114 \\
-.053 & .072 & -.066 & -.053 & -.066 & -.079 & -.039 & \mathbf{.447} & .072 & -.053 & -.039 & -.053 & -.066 & .072 & -.039 & .059 & -.066 & -.053 \\
.072 & -.053 & -.066 & -.053 & -.066 & .046 & -.039 & .072 & \mathbf{.447} & .072 & -.039 & -.053 & -.066 & -.053 & -.039 & -.066 & -.066 & -.053 \\
.072 & -.053 & -.066 & -.053 & .059 & -.079 & -.039 & -.053 & .072 & \mathbf{.447} & -.039 & -.053 & .059 & -.053 & -.039 & -.066 & -.066 & -.053 \\
-.053 & -.053 & -.066 & .114 & -.066 & -.079 & -.039 & -.053 & -.053 & -.053 & \mathbf{.461} & -.053 & -.066 & -.053 & .127 & -.066 & .101 & -.053 \\
-.053 & -.053 & .059 & -.053 & -.066 & .046 & -.039 & -.053 & -.053 & -.053 & -.039 & \mathbf{.447} & -.066 & -.053 & -.039 & .059 & -.066 & .072 \\
.047 & .047 & -.066 & -.053 & -.066 & -.079 & -.039 & -.053 & -.053 & .047 & -.039 & -.053 & \mathbf{.434} & .047 & -.039 & .034 & -.066 & -.053 \\
-.053 & .072 & -.066 & -.053 & -.066 & -.079 & -.039 & .072 & -.053 & -.053 & -.039 & -.053 & .059 & \mathbf{.447} & -.039 & .059 & -.066 & -.053 \\
-.053 & -.053 & -.066 & .114 & -.066 & -.079 & -.039 & -.053 & -.053 & -.053 & .127 & -.053 & -.066 & -.053 & \mathbf{.461} & -.066 & .101 & -.053 \\
-.053 & .047 & -.066 & -.053 & -.066 & -.079 & -.039 & .047 & -.053 & -.053 & -.039 & .047 & .034 & .047 & -.039 & \mathbf{.434} & -.066 & -.053 \\
-.053 & -.053 & .034 & .047 & .034 & -.079 & -.039 & -.053 & -.053 & -.053 & .061 & -.053 & -.066 & -.053 & .061 & -.066 & \mathbf{.434} & -.053 \\
-.053 & -.053 & .059 & -.053 & -.066 & .046 & .086 & -.053 & -.053 & -.053 & -.039 & .072 & -.066 & -.053 & -.039 & -.066 & -.066 & \mathbf{.447}
\end{matrix}$$

A consequence of the fact that the convergence model is a Markov process is that the simple process described in equation 1 turns out to generate some rather complex results. While it may seem that this kind of calculation should move people who interact with one another close to one another, and that the members of clusters of interacting individuals will be ideationally homogeneous, this is the case only after a small number of iterations. If the calculations are repeated until the positions of individuals stop changing, the locations of individuals will be partitioned into two sets, as the dotted line on Figure 4 indicates.

D. Starting configuration doesn't matter

It was mentioned above that the positions of the network members converge to a common point regardless of the configuration of the initial positions. It is only the pattern of connections between individuals that influences the final positions; the initial positions have no effect on the final locations. To demonstrate this, several thousand randomly assigned starting configurations of positions were generated and fed into the model described above. For each configuration, we calculated the correlation between the positions after each iteration and the final set produced by using Equations 2 and 3 on the original data. *Without fail*, the correlation converged on either 1.0000 or -1.0000.
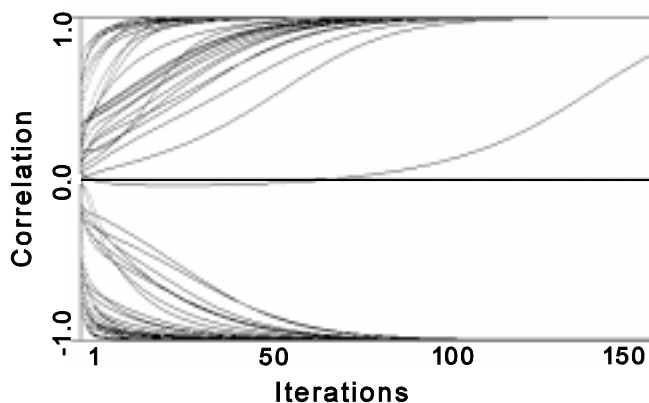

Figure 5. Correlations of position vector with final results obtained by iterating Equations 2 and 3 on the original data, plotted against iteration number. Results shown for 50 randomly-chosen starting configurations.

This result shows that the final state of the system is determined by the particular pattern of connections in the network, and that the initial configuration of positions is irrelevant. This demonstrates a systems-theoretic property called *equifinality*, an emergent property of the system, which in this case is a result of the set of connections between individuals. The reason this happens is the topic of Part 2 of this chapter, where we explore the mathematical properties of the model demonstrated above and develop an analytic procedure suggested by these results.

II. PART 2 — ANALYSIS

In Part 1 we introduced the Convergence Model as a systems theoretic approach to the dynamics of communication networks. The model is inherently iterative, in the sense that the results of the system's behaviour at one point in time are fed back into the system for processing at the next point in time. We explored the behaviour of the model and constructed a matrix formulation that made it possible to develop a mathematical framework that we could use to explore the model in greater depth.

In Part 2 we introduce some notation and use ideas from various branches of mathematics to explore the structure revealed by our approach to convergence, and to build a coherent set of explanations for the behaviour we saw in Part 1. We approach the problem from four perspectives: the algebraic view, where we consider how the eigenvectors partition the network; the geometric view, where we develop a geometric interpretation of the eigenvectors and consider an especially useful non-Euclidian coordinate system we call "display space"; the statistical view, where we give quantitative meanings to the eigenvalues and the components of the eigenvectors; and the pragmatic view, where we address a number of practical matters of interest to the researcher who wishes to apply the methods we are developing. In the course of our presentation we construct an analytic model that can be used to investigate the structure of communication networks. Because of its roots in Convergence theory, we call it "Convergence Analysis". We describe this model and show how it is related to existing approaches.

A. Background and mathematical foundation

In this section we develop some useful notation, describe some useful aspects of matrix algebra, and lay the groundwork for a deeper analysis of the eigen structure of networks. We begin the discussion with a brief description of five matrices that we use throughout this part of the chapter. The first one, A, is the familiar adjacency matrix that is used throughout most applications of network analysis. The other four are variations of the adjacency matrix, most of them normalized so that the sum of certain elements equals 1.0.

The adjacency matrix: A

We will always use the symbol A to refer to the *Adjacency Matrix* of a graph (Biggs, 1993). We assume that A is symmetric (and therefore square), non-negative, and without entries on the diagonal (so the graph has no self-loops, i.e. nodes are not connected to themselves). The simplest version of A is a matrix of 0's and 1's, where a_{ij}, the item in row i, column j of A, is 1 if node i is connected to node j, and 0 otherwise. An example of A and the graph it represents appears on the top of the next page.

$$A = \begin{matrix} 0 & 1 & 0 & 0 & 0 & 0 & 0 & 1 \\ 1 & 0 & 1 & 0 & 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 & 0 & 1 & 0 & 1 \\ 0 & 0 & 1 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 1 & 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 & 0 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \end{matrix}$$
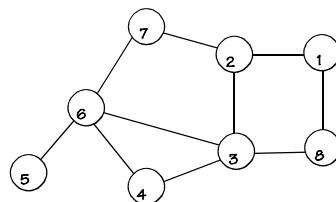
None of the results of that we have seen or will derive requires that **A** be binary (though it must be

non-negative). If the non-zero items of *A* are integers greater than 1, then *A* may be thought of as representing a *multigraph*, for which multiple connections exist between nodes. *A* may even contain non- integers (e.g. 18.2456 or 0.914), which we interpret as *weighted* connections between nodes. We will assume *A* represents a *connected* graph (where there is always a path from every node to every other node), though the analysis could be extended to cover disconnected graphs. Since *A* is connected, there will be at least one non-zero entry in each row (no isolated nodes).

The normalized adjacency matrix: *C*

It is often convenient to deal with the matrix *C*, which is derived from *A* by dividing each item of *A* by $\sum a_{ij}$, the sum of all the items in the matrix, which makes $\sum c_{ij} = 1$. In other words, *C* is a version of *A* that has been normalized so that the sum of all the elements equals **1.0**.

$$ C = \frac{A}{\sum_{i,j} a_{ij}} $$

For the example *A* given above, there are ten links, each of which appears twice in the adjacency matrix, so $C_{ij} = A_{ij}/20$.

$$ C = \begin{matrix}
\mathbf{.0} & .05 & .0 & .0 & .0 & .0 & .0 & .05 \\
.05 & \mathbf{.0} & .05 & .0 & .0 & .0 & .05 & .0 \\
.0 & .05 & \mathbf{.0} & .05 & .0 & .05 & .0 & .05 \\
.0 & .0 & .05 & \mathbf{.0} & .0 & .05 & .0 & .0 \\
.0 & .0 & .0 & .0 & \mathbf{.0} & .50 & .0 & .0 \\
.0 & .0 & .05 & .05 & .05 & \mathbf{.0} & .05 & .0 \\
.0 & .05 & .0 & .0 & .0 & .05 & \mathbf{.0} & .0 \\
.05 & .0 & .05 & .0 & .0 & .0 & .0 & \mathbf{.0}
\end{matrix} $$

Diagonal matrix of row proportions: D_r

Let r_i be the sum of the entries in row *i* of *C*:

$$ r_i = \sum_j c_{ij} $$

Then r_i will be the proportion of links contributed by node *i* or the marginal percentage of row *i* if the adjacency matrix *A* were treated as a crosstabulation table. We then form the matrix D_r, a diagonal matrix, by placing the *r*'s (the row sums of *C*) along the diagonal. We refer to the diagonal as *r*. Both *r* and D_r play central roles in the following presentation.

$$ r = \begin{matrix}
.10 \\ .15 \\ .20 \\ .10 \\ .05 \\ .20 \\ .10 \\ .10
\end{matrix} \qquad D_r = \begin{matrix}
\mathbf{.10} & .0 & .0 & .0 & .0 & .0 & .0 & .0 \\
.0 & \mathbf{.15} & .0 & .0 & .0 & .0 & .0 & .0 \\
.0 & .0 & \mathbf{.20} & .0 & .0 & .0 & .0 & .0 \\
.0 & .0 & .0 & \mathbf{.10} & .0 & .0 & .0 & .0 \\
.0 & .0 & .0 & .0 & \mathbf{.05} & .0 & .0 & .0 \\
.0 & .0 & .0 & .0 & .0 & \mathbf{.20} & .0 & .0 \\
.0 & .0 & .0 & .0 & .0 & .0 & \mathbf{.10} & .0 \\
.0 & .0 & .0 & .0 & .0 & .0 & .0 & \mathbf{.10}
\end{matrix} $$

A row-normalized adjacency matrix: M^*

If we normalize *A* so that the elements in each row sum to **1.00**, we get M^*. This matrix is defined as $D_r^{-1}C$, thus:

$$
M^* = \begin{bmatrix}
\mathbf{.0} & .5 & .0 & .0 & .0 & .0 & .0 & .5 \\
.333 & \mathbf{.0} & .333 & .0 & .0 & .0 & .333 & .0 \\
.0 & .25 & \mathbf{.0} & .25 & .0 & .25 & .0 & .25 \\
.0 & .0 & .5 & \mathbf{.0} & .0 & .5 & .0 & .0 \\
.0 & .0 & .0 & .0 & \mathbf{.0} & 1.0 & .0 & .0 \\
.0 & .0 & .25 & .25 & .25 & \mathbf{.0} & .25 & .0 \\
.0 & .5 & .0 & .0 & .0 & .5 & \mathbf{.0} & .0 \\
.5 & .0 & .5 & .0 & .0 & .0 & .0 & \mathbf{.0}
\end{bmatrix}
$$

It is derived from C by dividing each row of C by its row sum (or by dividing each row of A by its row sum), so:

$$
m^*_{ij} = \frac{c_{ij}}{\sum_j c_{ij}} = \frac{c_{ij}}{r_i} = \frac{a_{ij}}{\sum_j a_{ij}}
$$

An augmented row-normalized adjacency matrix: M

Finally, the matrix M is derived from A by adding its row sums to its diagonal, and then dividing each new row by the new row sums. It is easy to show that $M = .5M^* + .5I$, where I is an identity matrix.

$$
M = \begin{bmatrix}
\mathbf{.5} & .25 & .0 & .0 & .0 & .0 & .0 & .25 \\
.167 & \mathbf{.5} & .167 & .0 & .0 & .0 & .167 & .0 \\
.0 & .125 & \mathbf{.5} & .125 & .0 & .125 & .0 & .0 \\
.0 & .0 & .25 & \mathbf{.5} & .0 & .25 & .0 & .0 \\
.0 & .0 & .0 & .0 & \mathbf{.5} & .5 & .0 & .0 \\
.0 & .0 & .125 & .125 & .125 & \mathbf{.5} & .125 & .0 \\
.0 & .25 & .0 & .0 & .0 & .25 & \mathbf{.5} & .0 \\
.25 & .0 & .25 & .0 & .0 & .0 & .0 & \mathbf{.5}
\end{bmatrix}
$$

Recall from Part 1 that M is a matrix expression of Equation 1, the original statement of the convergence model. The entries of M are all non-negative, and the diagonal elements are equal to the sum of the off-diagonal elements in each row, so M is positive semi-definite.[2] This is not true for M^*, which is also non-negative, but which has zeros on the diagonal. We will see the importance of this later. *Note also that neither M^* nor M are likely to be symmetric.*

1. Stochastic matrices

A matrix with non-negative entries whose rows all sum to **1** is called a *row-stochastic* (or more generally *stochastic*) matrix (Feller, 1950). Stochastic matrices arise in the study of Markov chains (they are also called Markov matrices). A stochastic matrix may be thought of as a matrix of transition probabilities between states represented by the rows of the matrix. Since the rows of both M and M^* sum to **1**, they are stochastic matrices.

As a stochastic matrix (to be precise, it is row-stochastic) M has the *averaging property* (Senata, 1981): The minimum value of p is less than all values of Mp, and the maximum value of p is greater than all values of Mp. Multiplying the vector p by the stochastic matrix M reduces the range of values of p:

$$
\min_j p_j \leq \sum_j m_{ij} p_j \leq \max_j p_j
$$

We refer to this phenomenon as "shrinking of range". This is convergence.

2. The Eigenvalue problem

The eigenvalues of the *symmetric* matrix S are the numbers λ_j which, together with e_j, the eigenvectors of S, satisfy the Eigenvalue equation:

$$
Se_j = \lambda_j e_j
$$

The eigenvalue problem arises naturally by considering the following problem: maximize the projection $x^t S x$ for any symmetric matrix S and vector x (Joliffe, 1986). That is, choose a set of orthogonal axes in the multidimensional space occupied by the data and *rotate* them so that the first axis points in the direction of the greatest variability in the data; the second one, perpendicular to the first one, points in the direction of greatest remaining variability; and so on. This set of axes is a coordinate system that can be used to describe the relative positions of the set of data points. Most of the variability in the locations of points will be accounted for by the first few dimensions of this coordinate system. The coordinates of the points along each axis will be an eigenvector, and the length of the projection will be an eigenvalue. This projection is a sum of $x_i x_j$ weighted by s_{ij}, the elements of S.

We can thus write the eigenvalue problem as:

$$\text{minimize: } \sum_{i,j} x_i x_j s_{ij}$$

To solve this we add the constraint that $x^t x = 1$, so that x has unit length. Using the method of Lagrange multipliers[3], we write:

$$x^t Sx - \lambda x^t x - 1 = 0$$

where λ is a Lagrange multiplier (Courant and Hilbert, 1966). Since we are looking for a maximum, we take derivatives with respect to x and set the result to $\mathbf{0}$ to get:

$$Sx - \lambda x = 0 \text{ or } Sx = \lambda x,$$

which is the eigenvalue equation. Solving this for a series x_j (the eigenvectors) and λ_j (the eigenvalues) gives us a *set* of projections of S on x_j such that:

$$X^t SX = D_\lambda$$

where X is a matrix with columns x_j, and D_λ is a diagonal matrix with the λ_j along the diagonal. This allows us to express the information in S in a very simple form, and for this reason we say that the eigenvectors form a *natural coordinate system* for S. We may also think of the eigenvectors as a set of *axes*, so that the data in S projected along these axes has lengths D_λ. For simplicity, we will assume that the λ_j are distinct, that is, if $\lambda_i = \lambda_j$, then $i = j$.

Some useful facts about eigenvectors and eigenvalues will assist in the following discussion:

- The eigenvectors of $X + cI$ (where c is a constant, and I is an identity matrix) are *the same* as the eigenvectors of X. In other words, adding a constant, c, to the diagonal elements of X does not affect the eigenvectors. The eigenvalues of $X + c I$ are shifted by an amount equal to c. If the value 2.0 is added to all the diagonal elements of the matrix X, the eigenvalues will each be increased by 2.0.

- The eigenvalues of cX are c times the eigenvalues of X. In other words, if all elements of the matrix X are multiplied by a constant, the eigenvalues are also multiplied by the same constant. For example, the eigenvalues of $-X$ are $-$(the eigenvalues of X).

For reasons that will become clear in a few paragraphs, a variation of the eigenvalue problem called the *Generalized Eigenvalue Problem* is especially well-suited for working with network data. Instead of $Se_j = \lambda_j e_j$, we are interested in $Ce_j = \lambda_j D_r e_j$, where the matrix D_r enters the product on the right of the equals sign[4]. We will examine $Ce_j = \lambda_j D_r e_j$, which we can express in terms of matrix M:

$$D_r^{-1} C = \lambda_j e_j = Me_j$$

Since M is a positive semi-definite matrix, it has eigenvalues $\lambda \geq 0$. As a stochastic matrix, its greatest eigenvalue is $\mathbf{1}$, and all others are less than $\mathbf{1}$. We may write:

$$1 = \lambda_1 > \lambda_2 \geq \dots \lambda_n \geq 0$$

An important implication of the fact that the greatest eigenvalue is **1** and all the others are smaller[5] is that when $p \leftarrow Mp$ is iterated[6], p will converge monotonically, without oscillations, to the first eigenvector of M, which is what we saw in Part 1 (see Fig. 4 in Part 1 for a demonstration of this process).

In the following discussion we occasionally use the matrix M^*, another representation of A introduced earlier, which has the same eigenvectors as M, and its eigenvalues λ^* are related to those of M by a simple formula. Because $M^* = 2M - I$,

$$\lambda_j^* = 2\lambda_j - 1$$

Note that some of the λ^* *may* be negative, (in particular, $\lambda^* = $ -**1** when $\lambda = $**0**). Since the rows of M^* also sum to **1**, M^* is an example of a stochastic matrix , but since its diagonal elements are 0, it is *not* positive semi-definite. We will see that M is a better matrix to work with than M^*, but we are also interested in the negative eigenvectors[7] of M^*.

Finally, we repeat that M and M^* are *not likely to be symmetric*, since they are both only normalized on the rows — they are row-stochastic.

Some notation: The vector e_i is a column vector, while its transpose, e_i^t is a row vector. The *outer product* of a column vector and a row vector, such as $e_j e_j^t$, will be a square matrix. In contrast, the *inner product* of a row vector and a column vector, such as $e_j^t e_j$, will be a scalar number.

---

3. Spectral Decomposition

For a *symmetric* matrix S with distinct (for $i \neq j$, $\lambda_i \neq \lambda_j$) eigenvalues λ_j, we may write (Curtis, 1974):

$$S = \sum_j \lambda_j e_j e_j^t = \sum_j \lambda_j E_j$$

This is the spectral decomposition of S. It allows us to expresses a matrix S as a sum of E_j weighted by eigenvalues λ_j, where E_j is $e_j e_j^t$, the outer product of eigenvector e_j. E_j is an idempotent (i.e. $E_j^2 = E_j$) rank 1 (its columns are all multiples of the first column) matrix. E_j defines a *subspace* — a single structural dimension — of S. λ_i, the eigenvalue associated with E_i, indicates the relative contribution of E_i to the matrix S. Since the first eigenvalues are larger than the others, the first subspaces will account for larger proportions of the matrix.

Because S is symmetric, we can always select the e_j to be orthonormal, so the subspaces will be independent:

$$e_i e_i = 1 \text{ and } e_i e_j = 0 \quad (i \neq j)$$

so that:

$$e_i e_i^t \ e_i e_i^t \ = e_i(e_i^t e_i) \ e_i^t \ = \ e_i e_i^t \text{ and, when } i \neq j, \ e_i e_j^t \ e_i e_j^t = e_i(e_j^t e_i) \ e_j^t = 0$$

Alternatively,

$$E_i E_i = E_i \quad \text{and} \quad E_i E_j = 0$$

The square of S may therefore be written:

$$S^2 \ = \ \sum_j \lambda_j^2 e_j e_j^t e_j e_j^t \ = \ \sum_j \lambda_j^2 e_j e_j^t$$

and similarly for higher powers of S. Thus if a single eigenvalue λ_j is larger than all the others, higher powers of S will be dominated by the rank **1** matrix $e_j \ e_j^t$, and if that eigenvalue happens to be exactly **1** (as it is for stochastic matrices), all the other subspaces will vanish since, for $\lambda_i < 1$, $\lambda_i^n \rightarrow 0$ as $n \rightarrow \infty$. This is the basis of the *Power Method* for finding the largest eigenvalue and corresponding eigenvector for a matrix. The Power Method consists of iterating:

$$p \leftarrow S p$$

until p does not change direction (and we have seen this in Part 1). Also, after a sufficient number of iterations, the ratio $S^{k+1}p \big/ S^k p$ will approach the largest eigenvalue.

4. Spectral Decomposition of M

Since M is *not* symmetric, its eigenvectors are *not* orthonormal (Press, *et. al.*, 335-380,1986), so we need to make some changes in the spectral decomposition representation of M. D_r, the diagonal matrix of row proportions defined above, is used here. We introduce matrices F_j, constructed by forming:

$$F_j = e_j D_r e_j^{\,t}$$

Then $F_i F_i = F_i$ and $F_i F_j = 0$ for $i \neq j$ (as in the symmetric case) where, again, the F_i are idempotent rank 1 matrices. We may write:

$$M = \sum_j \lambda_j F_j$$

for the spectral decomposition of M. Now we have a representation of M which shows that higher powers of M will be dominated by the first eigenvector of M, *whose eigenvalue we know to be 1*. Furthermore, by the averaging property of M, we know that for any starting vector p, the result of many iterations will move to the mean value of p weighted according to row sums. This is exactly the behaviour we saw in Part 1.

The Spectral Decomposition of M is not useful only for calculating high powers of M. It also has an important Least Squares property. Define the partial sum M_k:

$$M_k = \sum_{j=1}^{k} \lambda_j F_j$$

The sum using only the first k eigenvectors is the best Least Squares Approximation to M in the sense that it minimizes the sum of the element-by-element distances between M and M_k (Curtis, 1974):

$$\sum_{i,j} (m_{i,j} - m(k)_{i,j})^2$$

where $m(k)_{i,j}$ is the i,j element of M_k. (This is a goodness of fit measure.) Furthermore, this spectral notation suggests that we may use the Power Method to get the second eigenvector of M by iterating

$$p \leftarrow M_1 p$$

where $M_1 = M - \lambda_1 F_1$ and $F_1 = e_1 e_1^{\,t}$. The first eigenvalue, λ_1, is **1**, so M_1 is simply $M - F_1$. (This is the "correction" applied by Equation 2 in Part 1.) The leading eigenvalue of M_1 will be less than **1** and the range of values in P will decrease with each iteration (shrinkage of range), so it will be necessary to expand the position vector P after each iteration, which, again, is what we saw in part 1.

We may continue this to find the next eigenvector by iterating $M_2 = M - F_1 - \lambda_2 F_2$, although this method, called Deflation (Acton, 1990), is not recommended because rounding error builds up and the system becomes unstable. There are much more efficient and stable methods (such as Generalized Singular Value Decomposition) for finding the Spectral Decomposition of M. However, these methods are $O(n^3)$ in time and $O(n^2)$ in space[8], and so are expensive and time-consuming to apply to the large matrices often found in social network analysis. Also, we are not always interested in a complete eigen decomposition, but may wish only to find a low (2 or 3) dimensional representation of the network so that we may examine it for partitions, clusters, etc. The Least Squares property of the Spectral Decomposition indicates that low dimensional approximation may be good, as long as the eigenvalues left out are "small", in a sense that we will quantify later.

We have now developed sufficient notation to examine our problem from a number of different viewpoints. Our goal is to provide insight into the quantitative meanings of both eigenvalues and eigenvectors of M.

B. The Algebraic View

In this section we will compare our results with those of Algebraic Graph Theory, which is largely concerned with the eigen decomposition of the original adjacency matrix *A*. We will refer to this as the *Standard Method*.

It is easy to show that *A* is completely determined by its Spectral Decomposition (both eigenvalues and eigenvectors) (Cvetkovic, *et al.* 1988). However, the Standard method usually concentrates on the eigen*values*, and pays small attention to the eigenvectors. The set of eigenvalues is called the *Spectrum* of *A*(G) where *A* is the adjacency matrix of the graph **G**. The Standard method is concerned with distributions of the spectra of graphs (such as the occurrence and meaning of multiple eigenvalues), and proving results about the connections between eigenvalues and structure (such as number of cycles, longest and shortest cycles, etc.) Also, the Standard method is largely concerned with various types of regular graphs — those in which all nodes have the same number of connections. These graphs are necessarily symmetric and their eigenvectors are always orthonormal.

Rather than *A*, which is neither stochastic nor positive semi-definite, we focus our attention on *M*. By doing this, we normalize the spectrum ($1 \geq \lambda \geq 0$). Accordingly we refer to this as the *Normal method*. The
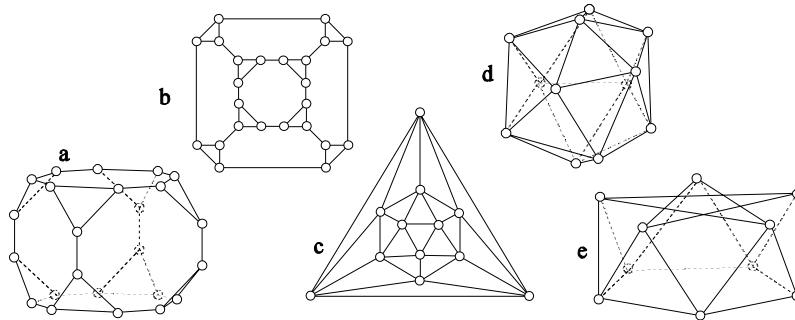

Figure 6.   Three regular graphs. *b* and *c* are planar representations of *a* and *d*. *e* is not planar; it cannot be shown in two dimensions without crossed lines.

normalization removes some of the information found in the standard spectrum (though, as we shall see, much important information remains). We are much more concerned with the eigen*vectors* as a source of structural information, and we focus on a different type of structural information (partitions and clusterings) than that provided by the Standard method. As stated above, our eigenvectors are not, in general, orthonormal (since *M* is not symmetric), and our graphs are unlikely to have multiple eigenvalues, so we have simplified our presentation here by assuming they are distinct.

Also, we have little interest in regular graphs, though it is useful to note that in the case of a regular graph, the Normal method gives exactly the same results as the Standard method (except for a constant multiple of the spectrum). It is easy to show this by considering the matrix *M* for a regular graph. Since both the rows and the columns all have the same sums in this case, *M* will be symmetric. We use this as a starting point to contrast the results of the two methods.

1. Partitions

The Standard eigenvector belonging to the largest eigenvalue of a *regular* graph is a vector of constant values (just as is the corresponding eigenvector of the Normal method for *any* graph). For non-regular graphs, however, the first eigenvector will not be a vector of constant values, but it will have constant sign (all positive or all negative), and the actual sign is arbitrary (Biggs, 1993). The next "largest" eigenvector, being orthonormal to the first, must therefore have both positive and negative values. Similarly, the third "largest", being orthonormal to the first two, must have a different pattern of signs, and so on. These patterns of signs *partition* the nodes of the graph into biants, quadrants, octants, etc. Up to a point, nodes will tend to go into the same

partition if they are connected (we will show why in the next two sections) (Barnes, 1982).

We can make exactly the same argument about Normal eigenvectors, though we must be careful about the fact that the vectors are not orthogonal. We can, in effect, force orthogonality by considering $e_j\,(r_j)^{\frac{1}{2}}$ instead of the e_j, that is, by weighting the eigenvectors by the square root of the row proportions. (This is the reason for the D_r term used in the definition of F_j for the Spectral Decomposition of M.) Since all elements of r_j are positive, this weighting will not affect the distributions of signs for the eigenvectors (though they will not, in general have the *same* distribution as in the Standard method), and we can follow exactly the same reasoning we used for signs and partitions in the previous paragraph. This partitioning process was seen in Part 1 as a natural consequence of the iteration procedure. For reasons we discuss later, the partitions induced by the Normal method are very useful.

Figure 8 on the next page shows the signs of the first three "non-trivial" Normal eigenvectors for the network used throughout Part One — first in order according to the original rows and columns of the adjacency matrix A, and then in order according to the signs of the vectors. The rows and columns of the adjacency matrix shown in the figure have been permuted (the order has been changed) according to the second ordering of the signs of the eigenvectors. Even a quick glance at the matrix will show that there are four clusters of nodes. Eigenvector 2 partitions the network into two subsets: (A, I, J, M, B, H, N, P) and (L, F, C, G, R, E, D, K, O, Q). Vector 3 pulls out (L, F, C, G, R), and vector 4 separates (A, I, J, M) from (B, H, N, P). The ordering in the matrix was produced by sorting the nodes with positive and negative signs on vector 2, then by sorting the nodes in each of those partitions according to the signs on vector 3, and then on vector 4. The use of more vectors to produce further partitions would eventually put each node in its own partition. It is likely that this will happen before all n eigenvectors have been used. If we use the signs of *all* the eigenvectors, we will have put every node into its own 2^n-ant (Aspvall & Gilbert, 1984). If we are interested in forming clusters, we want to stop *before* this starts to happen, but . . . how do we know where to stop?

```
     2   3   4        2   3   4         A I J M B H N P L F C G R E D K O Q
A   -1   1  -1    A  -1   1  -1     A   0 1 1 1 0 0 0 0 0 0 0 0 0 1 0 0 0 0
B   -1   1   1    I  -1   1  -1     I   1 0 1 0 0 1 0 0 0 1 0 0 0 0 0 0 0 0
C    1  -1   1    J  -1   1  -1     J   1 1 0 1 0 0 0 0 0 0 0 0 0 1 0 0 0 0
D    1   1   1    M  -1   1  -1     M   1 0 1 0 1 0 1 1 0 0 0 0 0 0 0 0 0 0
E    1   1  -1    B  -1   1   1     B   0 0 0 1 0 1 1 1 0 0 0 0 0 0 0 0 0 0
F    1  -1  -1    H  -1   1   1     H   0 1 0 0 1 0 1 1 0 0 0 0 0 0 0 0 0 0
G    1  -1   1    N  -1   1   1     N   0 0 0 1 1 1 0 1 0 0 0 0 0 0 0 0 0 0
H   -1   1   1    P  -1   1   1     P   0 0 0 1 1 1 1 0 1 0 0 0 0 0 0 0 0 0
I   -1   1  -1    L   0  -1   1     L   0 0 0 0 0 0 0 1 0 1 1 0 1 0 0 0 0 0
J   -1   1  -1    F   1  -1  -1     F   0 1 0 0 0 0 0 0 1 0 1 1 1 1 0 0 0 0
K    1   1   1    C   1  -1   1     C   0 0 0 0 0 0 0 0 1 1 0 1 1 0 0 0 0 1
L    0  -1   1    G   1  -1   1     G   0 0 0 0 0 0 0 0 0 1 1 0 1 0 0 0 0 0
M   -1   1  -1    R   1  -1   1     R   0 0 0 0 0 0 0 0 1 1 1 1 0 0 0 0 0 0
N   -1   1   1    E   1   1  -1     E   1 0 1 0 0 0 0 0 0 1 0 0 0 0 1 0 0 1
O    1   1   1    D   1   1   1     D   0 0 0 0 0 0 0 0 0 0 0 0 0 1 0 1 1 1
P   -1   1   1    K   1   1   1     K   0 0 0 0 0 0 0 0 0 0 0 0 0 0 1 0 1 1
Q    1   1   1    O   1   1   1     O   0 0 0 0 0 0 0 0 0 0 0 0 0 0 1 1 0 1
R    1  -1   1    Q   1   1   1     Q   0 0 0 0 0 0 0 0 0 0 1 0 0 1 1 1 1 0

  signs of         ordered by       adjacency matrix ordered by sign patterns
  evecs 2 3 4      sign  patterns
```

Fig. 7.   The network used to illustrate Part 1, with the rows and columns of the adjacency matrix permuted according to the signs of three eigenvectors.

2. Colourings

We define a *correct colouring* of a graph as a labeling (the colour) of each node such that no adjacent nodes have the same colour. The *minimum* correct colouring uses the smallest possible number of colours. This problem is known to be NP-hard in general (i.e., finding a solution is hard, but checking to see if it is correct is easy) (Garey, 1979).

The *trace* of a matrix is the sum of the values along its diagonal. The sum of the eigenvalues of a matrix is

equal to its trace. In Normal analysis, we use the matrix M, where the sum of the diagonal is **n/2**. We already know all the λ_j will be $\geq$ **0.** In Standard analysis, however, we start with the adjacency matrix A, which has zeroes along its diagonal. Since the sum of its eigenvalues is zero, its spectrum must have both positive and negative values. In the discussion above, we have been implicitly using only the positive ("largest") eigenvalues. Now we will concentrate on the negative eigenvalues.

For the negative eigenvalues of A, we can make an argument very similar to the one we used above: that the sign pattern of the two "smallest" (most negative) eigenvectors must differ, and therefore partition the nodes. In this case the nodes go into *different* partitions if they are *connected* (again, we will show why below) (Aspvall & Gilbert, 1984). See Figure 9 for an illustration of a network with negative eigenvalues.

This can continue until each node has its own colour (based on its own unique pattern of signs) if we include enough "smallest" eigenvectors. This trivially solves the coloring problem, but not the minimum coloring problem. It has been conjectured that the trivial coloring is produced by considering *only* the negative eigenvalues (Aspvall & Gilbert, 1984). For a certain class of regular graphs, (the "block regular," e.g., regular multi-partite graphs) we can always get a minimum correct colouring from a small number of eigenvectors.

Since we are not concerned with the Standard eigenvalues, why worry about colourings at all, since we know all our eigenvalues are $\geq$**0**? In fact, since the diagonal of M is n copies of **.5**, we know the trace is *.5n*, so the eigenvalues are distributed about **.5**, not **0**. The answer is that we do *not* worry precisely *because* of the diagonal of M — our largest eigenvalues are *always* positive. But consider M^*, where the trace is zero (see above), and which has the same eigenvectors as M, and eigenvalues $1 \geq \lambda_n{}^* \geq$ *-1* (and $\lambda_n{}^* = 2\lambda_n$ - *1*). In iterating M^*, we are in the same position as in iterating A. If the eigenvalue *largest in absolute value* happens to be negative, then we will get the corresponding eigenvector first.

```
                              D E O Q S C F J N P A B H I L G K M R T
A   1 -1      D -1 -1    D  0 0 0 0 0 0 1 0 0 0 0 0 1 0 0 1 1 1 1 0
B   1 -1      E -1 -1    E  0 0 0 0 0 0 0 0 0 0 0 0 0 0 1 0 1 1 1 1
C  -1  1      O -1 -1    O  0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 1 0 1 1 1
D  -1 -1      Q -1 -1    Q  0 0 0 0 0 0 0 0 0 0 0 0 1 0 0 1 1 1 0 1
E  -1 -1      S -1 -1    S  0 0 0 0 0 0 1 0 0 0 0 0 1 0 0 1 1 0 1 1
F  -1  1      C -1  1    C  0 0 0 0 0 0 0 0 0 0 1 1 0 1 0 1 0 0 0 0
G   1  1      F -1  1    F  1 0 0 0 1 0 0 0 0 0 0 1 1 1 1 0 0 0 0 0
H   1 -1      J -1  1    J  0 0 0 0 0 0 0 0 0 0 1 1 0 1 0 1 0 1 1 1
I   1 -1      N -1  1    N  0 0 0 0 0 0 0 0 0 0 1 1 1 0 1 0 0 0 0 0
J  -1  1      P -1  1    P  0 0 0 0 0 0 0 0 0 0 1 0 0 1 1 0 0 0 0 0
K   1  1      A  1 -1    A  0 0 0 0 0 1 0 1 1 1 0 0 0 0 0 0 0 1 0 1
L   1 -1      B  1 -1    B  0 0 0 0 0 1 1 1 1 0 0 0 0 0 0 0 0 0 0 0
M   1  1      H  1 -1    H  1 0 0 1 1 0 1 0 1 0 0 0 0 0 0 0 0 0 0 0
N  -1  1      I  1 -1    I  0 0 0 0 0 1 1 1 0 1 0 0 0 0 0 0 0 0 0 0
O  -1 -1      L  1 -1    L  0 1 0 0 0 0 1 0 1 1 0 0 0 0 0 0 0 0 0 0
P  -1  1      G  1  1    G  1 0 1 1 1 1 0 1 0 0 0 0 0 0 0 0 0 0 0 0
Q  -1 -1      K  1  1    K  1 1 0 1 1 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
R   1  1      M  1  1    M  1 1 1 1 0 0 0 1 0 0 1 0 0 0 0 0 0 0 0 0
S  -1 -1      R  1  1    R  1 1 1 0 1 0 0 1 0 0 0 0 0 0 0 0 0 0 0 0
T   1  1      T  1  1    T  0 1 1 1 1 0 0 1 0 0 1 0 0 0 0 0 0 0 0 0

Signs of      Ordered       Adjacency matrix with rows and columns
evecs 2 3     by signs      ordered according to signs of evecs
```

Fig. 8. A network with negative eigenvalues. The adjacency matrix has been ordered according to the signs of the components of the first two eigenvectors.

In Part 1 we introduced the Convergence model:

$$P_i^{\,/} = \frac{1}{2}\left(P_i + \frac{\sum_{j \neq i} P_j A_{i,j}}{\sum_j A_{i,j}}\right) \tag{1}$$

This equation translates into iteration using the matrix M. If we iterate M^* instead, two things happen: first,

we get both positive *and negative* eigenvalues. Second, we change the model in such a way that people who are connected do not necessarily move closer to one another. In fact, when the first eigenvector is negative, they move *further apart*. Clearly this is not convergence. It is at this point that we shift from convergence to analysis.

Since we are interested in the positive eigenvalues, we iterate $M = .5M^* + .5I$ rather than M^*.[9] Once we have found the eigenvalues for M, it is very easy to get the eigenvalues of M^*, and so to determine whether the eigenvectors should be used for clustering or colouring. (Our experience is that it is not unusual for a social network to have a large negative second or third eigenvalue, which is why we emphasize this point. For certain types of analysis, the coloring eigenvectors may be of most interest).

In some important ways, the Normal method *generalizes* the Standard method to non-regular graphs. Some important features of the Standard eigen decomposition of a regular graph are retained by the Normal decomposition of a non-regular graph. For example, for a non-regular graph, the first Normal eigenvector has constant components, whereas the first Standard eigenvector only has constant sign. We can always express the Normal eigenvectors as linear combinations of the Standard, and vice versa; the same information is there, but it is organized differently. We lose orthogonality, but the partitioning information is not so "spread out" (into the first eigenvector, for instance) (Barnes, 1982).

C. Geometric View

The algebraic discussion was mainly concerned with the signs of the eigenvectors, and we stated that the Normal eigenvectors had good partition properties. Here we will begin to describe what these properties are by giving a geometric interpretation for the eigenvectors. This section begins with a variation of a common method for introducing eigenvectors, which we extend by considering the case for a non-Euclidean coordinate system (Hall, 1970). In this section we find it useful to work with the matrix C defined by:

$$c_{ij} = \frac{a_{ij}}{\sum_{i,j} a_{ij}}$$

We first introduce the concept of a *metric*: a method for measuring distance between two points. The familiar example is the Euclidean metric expressed by the following expressions between axes:

$$x_i{}^t x_j = 1 \text{ if } i = j , \mathbf{0} \text{ if } i \neq j$$

This is also a definition of orthonormality for eigenvectors, so we see that orthonormal eigenvectors give us the Euclidean metric. This makes it easy (and familiar) to calculate the distance between two points x and y (in n-dimensional space) as:

$$(z_1{}^2 + \ldots + z_n{}^2)^{\frac{1}{2}}, \quad \text{where } z_i = x_i - y_i$$

Since the eigenvectors of M are *not* orthogonal, we introduce a new metric (called, for now, the D_r metric) to define:

$$x_i{}^t D_r x_j = 1 \text{ if } i = j, \mathbf{0} \text{ if } i \neq j$$

where D_r is the diagonal matrix with r along the diagonal (we have already used this matrix in the spectral decomposition of M).

In this metric, we can now give coordinates to each node in the normalized adjacency matrix C. If we form a matrix V in which each column is an eigenvector of M, then row j of V contains the coordinates of node j of C. We refer to the rows of V as $v_{i\bullet}$ and the columns as $v_{\bullet i}$ so that the *squared distance* between two nodes i and j is simply

$$d_{ij}^2 = \sum_k (v_{ik} - v_{jk})^2 = (v_{i\bullet} - v_{j\bullet})^t (v_{i\bullet} - v_{j\bullet})$$

We will not attempt to express these D_r metric distances in terms of the original matrices M or C. In the next section we show how to describe distances in these matrices in a closely-related metric.

1. The D_r metric

In this section, we are more concerned with distances in the D_r metric, and for this reason we introduce the notion of a *display*. A display is simply what we would see if we looked at the actual components of the eigenvectors, e.g., a table of numbers, or a graphic representation on the page or on a computer monitor. We will show that for a display of a network we can use the eigenvectors of M to *minimize the total length of lines connecting nodes*, where length is measured in display (i.e., D_r metric) coordinates.

We consider the following minimization problem:

$$\text{minimize} \quad z = \sum_{i,j=1}^{n} \sum_{j=1}^{n} (x_i - x_j)^2 a_{ij} \tag{4}$$

where a_{ij} are the entries of adjacency matrix A. This is a weighted sum of squares, where the weights are the connections between the nodes. We wish to minimize z, the total distance between all connected nodes. We will show that the eigenvectors of M provide the solution to this problem.

Following Hall, (1970), we now derive a convenient matrix formula for the total distance between connected nodes, using C:

$$
\begin{aligned}
z &= \sum_{i,j=1}^{n} (x_i - x_j)^2 c_{ij} \\
&= \sum_{i,j} (x_i^2 - x_j^2) c_{ij} - \sum_{i}^{j \neq i} 2 x_i x_j c_{ij} \\
&= 2 \sum_{i,j} x_i^2 \sum_j c_{ij} - 2 \sum_{i}^{j \neq i} x_i x_j c_{ij} \\
&= 2 x^t D_r x - 2 x^t C x
\end{aligned}
$$

If we set $L = D_r - C$, we can re-write equation 4 as:

$$\text{minimize} \quad z = x^t L x \tag{4a}$$

$$
L = D_r - C =
\begin{array}{cccccccc}
\mathbf{.10} & -.05 & .00 & .00 & .00 & .00 & .00 & -.05 \\
-.05 & \mathbf{.15} & -.05 & .00 & .00 & .00 & -.05 & .00 \\
.00 & -.05 & \mathbf{.20} & -.05 & .00 & -.05 & .00 & -.05 \\
.00 & .00 & -.05 & \mathbf{.10} & .00 & -.05 & .00 & .00 \\
.00 & .00 & .00 & .00 & \mathbf{.05} & -.50 & .00 & .00 \\
.00 & .00 & -.05 & -.05 & -.05 & \mathbf{.20} & -.05 & .00 \\
.00 & -.05 & .00 & .00 & .00 & -.05 & \mathbf{.10} & .00 \\
-.05 & .00 & -.05 & .00 & .00 & .00 & .00 & \mathbf{.10}
\end{array}
$$

This has a trivial solution $z = 0$ for $x_i = x_j$ for all i,j, which we call the solution in *zero* dimensions (a single point where every node has the same coordinates). To get a more useful *one* dimensional solution which will spread the nodes along a line, we add the constraint that $x^t D_r x = 1$ (the D_r metric) to get a pair of equations:

$$\text{minimize} \quad z = x^t L x \quad \text{such that } x^t D_r x = 1 \tag{5}$$

This constraint forces the points to fall along a line. Once again we use a Lagrange multiplier to combine the parts of Equation 5 into a single equation:

$$z = x^t L x - \zeta (x^t D_r x - 1)$$

Following the same method used earlier, we take the derivative of this with respect to x and set it to **0** (we want to minimize) to get:

$$L\,x - \zeta\,D_r\,x = 0 \quad \text{or} \quad L\,x = \zeta\,D_r\,x$$

which looks very like the definition of an eigenvalue problem[10]. This form defines the Generalized Eigenvalue Problem, which we can solve quite easily since D_r^{-1} is just $D_{1/r}$, and so we can write:

$$D_r^{-1}\,L\,x = \zeta\,x$$

But $D_r^{-1}\,L = D_r^{-1}\,D_r - D_r^{-1}\,C = I - M^*$, where I is an identity matrix, so we can write:

$$(I - M^*)x = \zeta\,x$$

which is close to the problem we have been solving all along. We expect to find a *set* of ζ_j which are related to λ_j^* (the eigenvalues of M^*) and to λ_j (the eigenvalues of M) by the equation:

$$\zeta_j = (1 - \lambda_j^*) = (1 - 2(\lambda_j - 1)) = 2(1 - \lambda_j),$$

and which have exactly the same eigenvectors as M and M^*.

Since $1 = \lambda_1 > \lambda_2 \geq \ldots \geq \lambda_n \geq 0$, then $0 = \zeta_1 > \zeta_2 \geq \ldots \geq \zeta_n \geq 2$.

If we use the largest positive λ's, we will get a picture in which clusters of tightly connected nodes will be visible, which is consistent with what the convergence model would predict. On the other hand, if we use the largest negative λ^*'s, we will get a picture in which nodes are not close to those with which they are linked, but rather with those with which they *are not* linked. Perhaps more importantly, they will be close to nodes which have similar patterns of connections with other sets of nodes — nodes to which they are structurally equivalent.

We now have three sets of eigenvalues associated with the network:

$$\text{for } M\text{:} \qquad \lambda_1 = 1 > \; \ldots \; \geq \lambda_n \geq 0$$

$$\text{for } M^*\text{:} \qquad \lambda_1^* = 1 > \; \ldots \; \geq \lambda_n^* \geq -1$$

$$\text{for } I - M^*\text{:} \qquad \zeta_1 = 0 > \; \ldots \; \geq \zeta_n \geq 2$$

We have already seen the zero-dimensional solution for $\zeta_1 = 0$, so we choose $\zeta_2 = 2(1 - \lambda_2)$ to get the minimal one-dimensional solution, meaning we choose x to be the second eigenvector of M. We can extend Equation 5 to *two* dimensions, by adding another constraint which forces the points into a two-dimensional display:

$$z = x^t\,L\,x - \zeta_2\,(x^t\,D_r\,x - 1) - \zeta_3\,(y^t\,D_r\,y - 1) \tag{6}$$

This leads us to choose the second and third ($\zeta_3 = 2(1 - \lambda_3)$) eigenvectors of M to give the minimal two-dimensional solution (probably the most useful one, since it can be displayed on a page). Since we are minimizing the distances between *connected* nodes, we would expect to see approximate clusterings in this display: nodes connected to one another are put into the *same* partitions.

We can continue this to give *three-* and *four-*dimensional displays (if display software supports it), but we would expect diminishing returns at some point. As we keep adding dimensions, the total lengths of lines connecting nodes increases. Higher dimensions add information on how to partition, for example, two connected nodes whose other connections are all the same (so-called "symmetries") that we are not really interested in. When ζ exceeds **1** (i.e. $\lambda < $ **0.5** and $\lambda^* < $ **0**), we have moved from partitions that cluster to those that color.

We can also view Equation 5 as a maximization problem. In fact, if we choose

$$\zeta_n = 2(1 - \lambda_n)$$

then we have the *maximum* 1-dimensional display, where connected nodes are *as far apart as possible*. This explains why the smallest values of the λ_j (and the most negative values of the λ_j*) produce approximate colorings: to maximize distance, connected nodes are put into *different* partitions.

Note that the metric D_r plays a crucial role in the above, since it reduces $D_r^{-1}L$ to a very simple form which allows us to relate the solutions ζ_j to λ_j, the eigenvalues of M, in a simple one-to-one fashion.

We conclude this section with a *very* simple example: a cycle of 5 nodes (this is regular and so the Standard method will give the same results as the Normal method). In this case we see that the positions of the nodes defined by the second and third eigenvectors are indeed as close as they can get in two dimensions: they form a regular polygon with each node on the circumference of a circle.   If we instead use the solution from maximization (the largest ζ), we see that no two connecting nodes are adjacent, since the distances between connecting nodes is maximized.
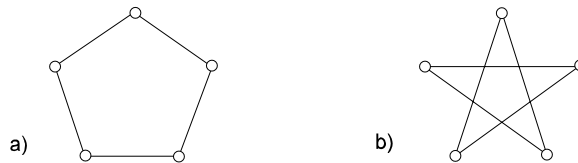

Figure 10.   Minimizing versus maximizing the distances. a) using the largest positive λ*'s to minimize the distances between linked nodes; b) using the largest negative λ*'s to maximize the distance between linked nodes.

D. The Statistical View

In this section, we give quantitative meanings to the eigenvalues and the components of the eigenvectors. We start with the usual expression:

$$Mx = \lambda x,$$

which we have seen can also be written as:

$$Ax = D_r \lambda x$$

One way of solving such a problem (Parlett, 1980) is to re-write it as:

$$D_r^{-1/2} A D_r^{-1/2} = \lambda x$$

which is exactly the form of equation used by the statistical procedure known as Correspondence Analysis (CA) — sometimes called Dual Scaling (Weller & Romney, 1990).

In CA, matrix A is generally not symmetric, or even square, (and can even have "apple" rows and "orange" columns) so the usual expression involves both $D_r^{-1/2}$ and $D_c^{-1/2}$ for row and column sums, and the solution consists of a pair of eigenspaces for rows and columns, and a common set of eigenvalues. Since we are considering only symmetric A, we only need a single eigenspace, but we can still use all of the machinery of CA to show results about the eigenspaces and interpret the eigenvalues, and to suggest methods for extending Convergence Analysis to the more general asymmetric case. We will not attempt to prove any of these results, since the topic is well-treated by (Greenacre, 1984; Jolliffe, 1986; Barnett, 1993).

The most important result is that CA is essentially eigen decomposition of the square root of the χ^2 matrix $(C - E)^2/E$ where E is the matrix of expected values of C, in our case simply rr^t. Any non-zero entry in $\chi(C) = (C - E)/E^{1/2}$ is:

$$x_{ij} = c_{ij} / (r_i^{1/2})(r_j^{1/2})$$

The squares of the eigenvalues are simply a partition of χ^2 *(C)* except for the first, which corresponds to the *expected value*, which we have not yet subtracted off! Thus: χ^2 of *C* is:

$$\chi^2 (C) \quad = \sum_{i=2}^{n} \lambda_i^2$$

In terms of the original adjacency matrix *A* we may also write:

$$\chi^2 (A) = \quad \sum_{i,j} a_{ij} \sum_{i=2}^{n} \lambda_i^2$$

We now see how to use the eigenvalues to determine how much of the "signal" is accounted for by a single eigenvector, or set of eigenvectors:[11]

$$\sum_{i=2}^{k} \lambda_k^2 \Bigg/ \sum_{i=2}^{n} \lambda_i^2$$

Another important result derives from the meanings of the coefficients of the eigenvectors. CA gives a specific weight to each of them. So far we have only shown them to have useful directions, but have not paid much attention to their length. In the geometric section we were concerned with the spaces and distances defined by the *eigenvectors*. Now we will turn things around and look at the spaces and distances defined by the *data* (in our case, the matrix *C* derived from the original adjacency matrix *A*).

We have discussed the eigenvectors of *M* without describing how to calculate them (except to suggest the Power Method with Deflation for obtaining the first few). Because *M* is not symmetric, we cannot use the fast, stable methods available for symmetric matrices. However we *can* restate the problem so that only symmetric matrices are involved, use the symmetric methods, and then transform the results back to get a solution for *M*. This is the idea behind the Singular Value Decomposition (SVD), and its extension, the Generalized SVD (GSVD) (Hoffman & Franke, 1986).

1. Correspondence Analysis

Correspondence Analysis begins by subtracting the expecteds from the matrix and the result in the χ metric by dividing the entries by the square root of the row and column sums. The problem is generally stated as:

$$\chi (C) = D_r^{-1/2} (C - rr^t) D_c^{-1/2} = Pd_\lambda N^t$$

and the Generalized Singular Value decomposition (GSVD) is used to find orthonormal matrices *P* and *N* ($P^t P = I$ and $N^t N = I$), where diagonal matrix D_λ has the singular values λ (in our case, the eigenvalues) along the diagonal. *P* and *N* are matrices of column eigenvectors of the rows and columns of $\chi(C)$.

In this formulation, the matrix of *expected* values rr^t is subtracted off [12], which eliminates the largest eigenvalue (the eigenvector that used to have the eigenvalue 1 now has the eigenvalue 0), but does not change the eigenvectors. We have kept this eigenvalue up to now to show its relationship to the largest eigenvalue of the Standard method. Since we know *C* to be symmetric, the row sums must equal the column sums, and we can write:

$$D_r^{-1/2} (C - rr^t) D_r^{-1/2} = PD_\lambda N^t$$

and we only use *P*, since *C* is symmetric. The CA solution is then expressed as:

$$G = D_r^{-1/2} PD_\lambda$$

that is, the *rows* of *P* are multiplied by $r^{-1/2}$, and the *columns* are multiplied by λ_j. The column multiplication weights the eigenvectors according to their contributions to $\chi(C)$ and affects the lengths of the eigenvectors, but not their directions.

The columns of *G* are eigenvectors[13] which are *proportional* to the eigenvectors of *M*. Each element of *G* has been assigned a specific value such that the rows are weighted to neutralize the *amount* contributed by each row

(this is the weighting by expected values), and each column has been weighted by the *significance* of contribution to the total chi-squared.

The first *n-1* columns of *G* are eigenvectors which describe the "signal" that remains after subtracting the "background" expected eigenvector, weighted by the eigenvalues. The last column of *G* is weighted by **0**, and is called the "trivial" eigenvector. As we said, it corresponds to the expected values, and plays no further part in CA.

In the previous section, we used $x^t D_r x = 1$ to derive the minima (and maxima) of total distances between connected nodes, whereas the CA matrix *G* satisfies $G^t D_r G = D_\lambda$, a diagonal matrix of the eigenvalues, where $1 > \lambda_j \geq 0$. The total distance *decreases* if we use the columns of *G* rather than the normalized (by $_D$) eigenvectors of *M*, since the constraints become $\lambda_j < 1$ for each dimension added.

2. The χ^2 Metric

Now we can talk about distances in the χ^2 (chi-squared) metric (Benzécri, 1969), which allows us to pull some most useful information out of the data. For example, we can compute the χ^2 distance between a pair of nodes in the network by examining their rows in the matrix *G*. The squared distance between rows *i* and *j* is

$$d_{ij}^2 = (g_i. - g_j.)^t (g_i. - g_j.)$$

where $g_i.$ and $g_j.$ are rows of *G*. This distance may also be expressed in terms of the original data matrix:

$$d_{ij}^2 = \sum_k \frac{(c_{ik}/r_i - c_{jk}/r_j)^2}{r_k} \tag{7}$$

or in terms of the matrix *M**:

$$d_{ij}^2 = (m^*_i. - m^*_j.)^t D_r^{-1} (m^*_i. - m^*_j.)$$

In particular, we have the very important Principle of Distributional Equivalence (Greenacre, 1984):

Any two rows having a zero distance between them may be replaced by a single row consisting of their sum.

While this principle suggests a method for reducing the size of any problem involving CA, it has a more important interpretation.

Any two nodes that have exactly the same pattern of connections will have chi-squared distance of zero (for a non-binary *A*, the chi-squared distance will be zero if the rows are *proportional*).

This can be seen from Equation 7, since the two rows *must* have the same row sums $r_i = r_j$, and $c_{ik} = c_{jk}$ for all *k* (for non-binary *A*, proportional rows have $c_{ik}/r_i = c_{jk}/r_j$ for all *k*).

As a consequence, rows with very similar patterns of connections will be very close in chi-squared space, and thus in any display of it.

This is another explanation for the good clustering properties of the Normal method.

Now we bring up the matter of eigenvalue signs again. The GSVD will always produce positive values along the diagonal of D_μ. For symmetric *C*, the matrices *P* and *N* will differ only in sign, and only if there are negative eigenvalues. This is another reason we prefer to work with *M*, which always has non-negative eigenvalues. We can always calculate the eigenvalues of *M** (and so get their signs). We can then decide whether we want to look at the eigenvectors that cluster, or those that color, and then order (and weight) the eigenvectors appropriately. Indeed, for some types of network analysis, we might be *most* interested in the coloring eigenvectors, as we would if we were interested in structural equivalence rather than cohesion-based cliques.

The plots in Figures 11 and 12 very clearly illustrates the difference between the eigenvectors associated with positive and negative eigenvalues. Both plots are pictures of the same network for which the first two eigenvalues are negative. Fig. 11 was drawn with the eigenvectors corresponding to the two largest *positive* eigenvalues (the "positive eigenvectors"), while Fig. 12 was drawn with the eigenvectors corresponding to the two largest *negative* eigenvalues (the "negative eigenvectors"). There is no clear structure visible in Fig. 11, but Fig. 12 shows that the network divides quite nicely into four blocks, each of which has most of its connections with one of the others, so that the network as a whole could be described as a pair of loosely connected bipartite networks.
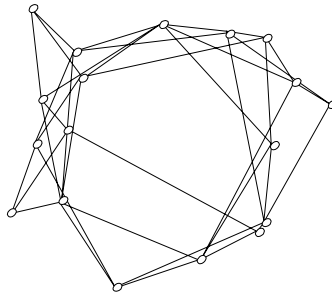

Figure 11. Done with M ($d = 0.5$), this figure shows the view determined by the two largest positive eigenvectors.
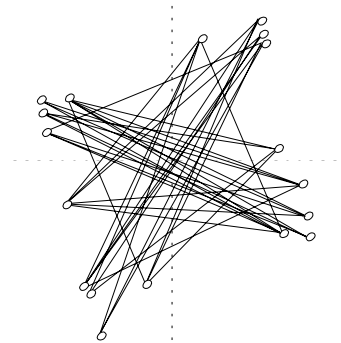
Figure 12. Done with M^* ($d = 0.0$), this figure shows the view determined by the two largest eigenvectors, both of which happen to be negative.

Figures 13 and 14 show a network in which the second eigenvalue is negative and the third and fourth ones are positive. Fig. 13 shows the results of ignoring the negative (2nd) eigenvector and using only the positive (3rd and 4th) ones gives a less informative picture of the network's structure than Fig. 14, which uses the largest two, one being negative and one positive.

We can also decide whether to work with M, M^* or some other matrix with a *different* diagonal weighting. So far, we have used the "meet-you-halfway" attributes of M, which is *.5(M^* + I)*. We could also construct "meet-you-3/4-way" matrices from *.75M^* + .25 I* and so on. From each of these it is easy to calculate the new distribution of eigenvalues (the eigenvectors do not change), and thus the different convergence properties. These different weightings correspond to different "damping" factors in the convergence[14].
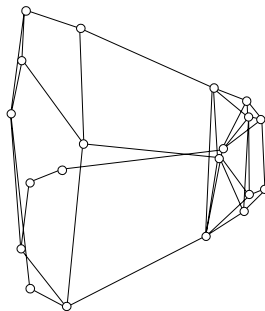

Figure 13. Done with M, this figure presents a misleading picture of the network's structure. It uses the largest two positive eigenvectors.
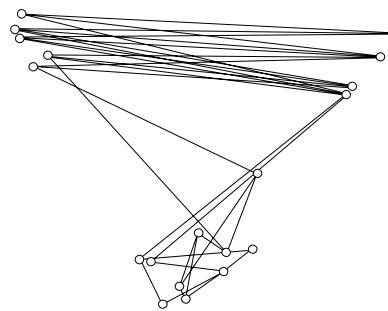
Figure 14. Using M^* to give a more informative view of the network shown in Fig. 7. The second eigenvector — the horizontal axis — was negative, while the third was positive.

We can go further and define a variable *d* (for damping) which allows us to deal with any diagonal weighting[15] *(1 - d) M* + d I*. Varying *d* will not change the eigenvectors, but it will change the eigenvalues as follows:

$$\delta_j = (1 - d)\, \lambda_j^* + d$$

Different values of *d* will control the order of the *emergence* of the fundamental eigenvectors and so will affect the behavior of position vector *p* when we iterate *p ← Mp*. We will consider this property of Convergence Analysis in the next section. Choosing *d < 0* to emphasize the negative λ_j^* results in a matrix which is no longer stochastic, and similarly for *d > 1*. We can also use *d* in the expression for the Convergence Model introduced in Part 1:

$$P_i{}' = \frac{1}{2}\,(P_i + \frac{\sum_{j \neq i} P_j A_{i,j}}{\sum_j A_{i,j}}) = d\,P_i + (1-d)\left(\frac{\sum_{j \neq i} P_j A_{i,j}}{\sum_j A_{i,j}}\right) \quad \text{when } d = .5$$

The Convergence Model as expressed by French (Equation 1) has individuals moving to the point of "equilibrium" halfway between the positions of the individuals, which implies a value of **.5** for *d*.

While varying *d* will *not* affect the eigenvectors of *(1 - d) M* + d I*, it *will* affect the CA matrix *G*. If the eigenvectors and eigenvalues of *M** are available, we can construct *G* for any value of *s* as follows:

1) construct *X*, with eigenvectors of *M** as columns, arranged by descending λ^* (we can ignore the "trivial" column belonging to $\lambda_1^* = 1$, so there will be n-1 columns).

2) for *j=2,n* calculate $\delta_j = (1 - d)\, \lambda_j^* + d$. The δ_j are the new eigenvalues, which, unlike λ_j^*, may not be ordered in decreasing size;

3) calculate XD_δ (i.e. multiply each column by δ_j)

4) if necessary, reorder the columns in descending δ_j;

5) add a last column of 0's for the expecteds.

This is *G* for the given value of *d*. Not all values of *d* result in a rearrangement of columns and a corresponding change in the emergent properties of the iteration *p = Mp*.

E. The Pragmatic view

"The perfect is the enemy of the good" - Anon.

In this section we will attempt to address some practical matters. For the large sparse arrays typical of social networks, it is not likely that we will want a complete eigen decomposition, for two reasons:

1) It would be extremely time (and space) consuming, even with more efficient methods (such as the Lanczos Algorithm (Parlett, 1979) for finding eigenvectors.

2) T*he network is very likely to be changing, with connections varying and even being created or deleted.*

1. Partial Decomposition

Complete eigen decomposition is $O(n^3)$, and thus impractical for large networks. What would we do with all the results? A complete decomposition would give *n*-1 eigenvectors and an equal number of eigenvalues. This will actually be more numbers than were in the original data! We already know that a few axes will probably contain most of the important information, so why not stop after 2 or 3 dimensions? χ^2 can be calculated once and used as a guide for how much of the structure has been explained.

We have suggested that Power Method with Deflation is a useful way to obtain the first few eigenvectors, and for large sparse arrays, and 2 or 3 dimensions, it's really not that bad. A few hundred iterations of $p' \leftarrow Mp - F_1$ gives the first dimension; a few hundred more of $p' \leftarrow Mp - F_1 - \lambda_2 F_2$ gives the second, and that may be all that is necessary to get a reasonably informative "global" view of the structure of the network. There are more clever, faster methods of finding the first few dimensions. However it is done, what kind of results are obtained? We give a few pictures of the kind of global structures that arise.
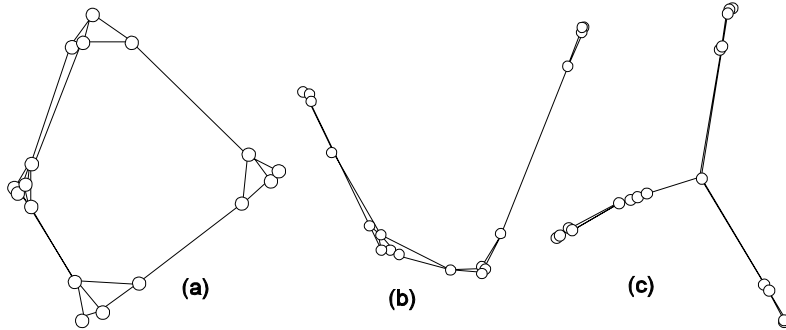

Figure 15. Network (a) is four cliques connected in a cycle. This network has positive eigenvectors which is evident from the figure because of the clear clustering of interconnected nodes. Network (b) is a long chain of clusters nodes in which all paths between the ones on the ends go through the ones in the middle. This is essentially a one-dimensional structure. Network (c) is three clusters connected to a central node.

2. Partial Iteration

We have used iteration of $p \leftarrow Mp$ to model the emergence of structure in a network, and have shown how the "latent" structure of eigenvectors determines the path that the elements of p follow. The fact that M is stochastic suggests that, in the long run, there will be a stable configuration. But in the short run, p will be most affected by the largest (in the senses explored in the last section) structural features. The graphs in Figure 16 show the squared correlations of p with the first four (non-trivial) eigenvectors.
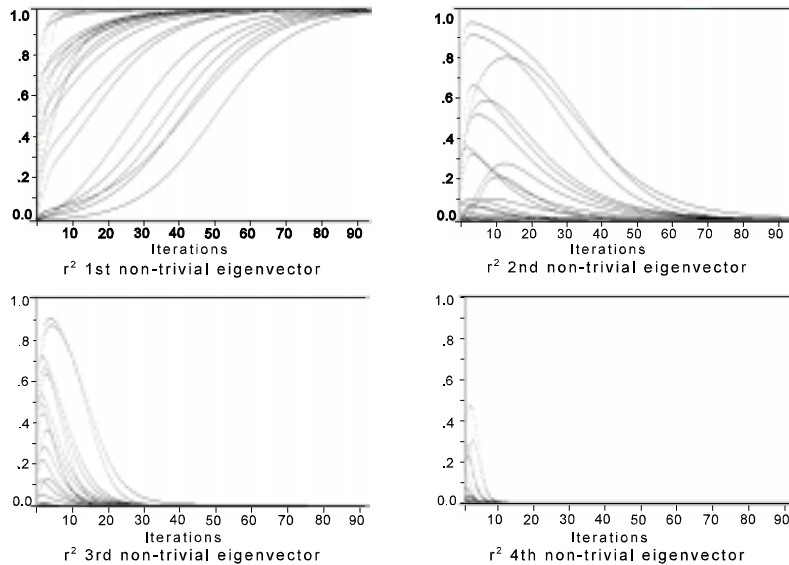

Fig. 16. Squared correlations of p with eigenvectors 2, 3, 4, 5; twenty random starting configurations of p.

There is a fairly wide range of variation evident in these graphs. With some starting configurations convergence is rapid, requiring only five to six iterations for all but the largest eigenvector to drop out of the

picture. With other configurations, a hundred or more iterations may be required. This variation is due to the fact that a random starting configuration may result in a *p* that points in the direction of one of the eigenvectors. If it points in the direction of the largest one, convergence will be rapid, while if it points in the direction of another, it will take more iterations. In fact, if *p* happens to be one of the other eigenvectors, iterating may not change the position vector at all.

To make partial iteration a useful tool, we propose choosing an initial *p* that is orthogonal to the largest eigenvector. Since the largest eigenvector will be the positive constant determined by the matrix of expecteds, a vector containing whole number values ranging from -**n/2** to **n/2** (for even **n**) or -**(n-1)/2** to **(n-1)/2** (for odd **n**) will be suitable. This initial *p* will have moderately large correlations with several eigenvectors, and will converge smoothly over a moderate number of iterations, as illustrated by Fig. 17.
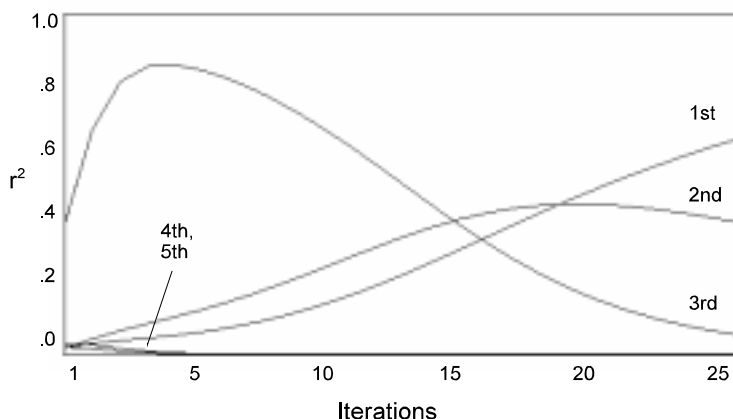

Fig. 17. Squared correlations of the five largest non-trivial eigenvectors with *p* over 25 iterations. From about the 10th to the 18th iteration, three eigenvectors make significant contributions to the total r^2.

We can use this idea to devise a "divide-and-conquer" strategy for quick analysis of the main network structures. The Spectral Decomposition of *M* shows that after a few *p* ← *Mp* iterations, the vector *p* will consist mostly of the contributions of the most "important" eigenvectors. In effect, we are projecting the *i*th iteration of *M* upon a single dimension. Now order *p* by the magnitudes of its components. Since each eigenvector induces a partition on *p*, we would expect to see the components of the ordered *p* changing rapidly where an important eigenvector induces a partition. We can make these partitions clearer by looking at the first difference of ordered *p*, as shown in Figure 18. A more sophisticated pattern-recognition approach, perhaps using edge detection techniques, would be useful. [16]
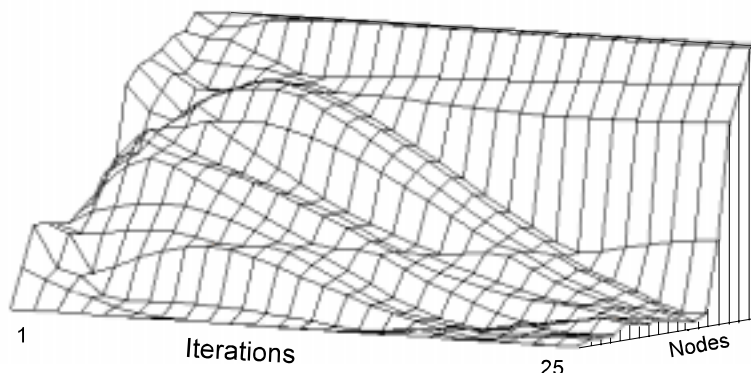
Fig 18. The more vertical the slope is, the farther apart the nodes are on *p*. Where the slope is horizontal, nodes have the same location on *p*. It is easy to see how clusters become visible and then disappear as the iterative process continues and more and more dimensions drop out of the picture as *p* converges towards the largest eigenvector. From about the 8th to the 15th iterations, two clusters are visible in the middle of the plot. By the 25th iteration, most of the nodes are clustered near the bottom.

Where the first difference of ordered p is relatively constant, we would expect to find a cluster, and now we can search the sparse array for the nodes defined by this cluster, and perform any tests we wish to define "closeness", "group membership", and so on. Each induced partition is examined in turn. Of course, we must subtract off the direction associated with the "trivial" eigenvector at each step of the iteration, but this is very easy to do even with a large sparse matrix. We can also introduce the damping factor d described earlier to emphasize the structural features associated with the positive and negative eigenvalues, and again, this is easy to do.

F. Conclusions

Part 1 explored the implications of Convergence theory and showed that they are more complex than they would seem to be on the surface. French's model seems to produce a system that converges to a point. We examined this "point" and saw that it was not such a simple point. In fact, it is a multidimensional structure, and it is dependent entirely on the pattern of connections among the nodes in the network. Surprisingly, the initial starting positions has no effect on the final state of the system except under certain rare conditions [17].

In Part 2 we explained the behaviour we saw in Part 1 by applying methods borrowed from various branches of mathematics. We introduced notation and results from linear algebra and Markov Processes to show that the Convergence Model can be analyzed as a Markov process; although we are not really interested in the long-term convergence to a single point. Rather, we examined the Spectral Decomposition of a network expressed as a stochastic matrix.

Using the Convergence Model, we introduced a variation to the Standard method of Algebraic Graph Theory. Because it uses a row-normalized form of the adjacency matrix and produces eigenvalues in the range 1.0 to -1.0, we called it the "Normal" method. We showed that we get results comparable to those of the Standard Method that appear to be better for both clustering and colouring.

We developed a geometric interpretation of the eigenvectors and considered an especially useful non-Euclidian coordinate system produced by the Normal Method which we called the D_r metric. We showed that in this metric the Normal Method minimizes (maximizes) distances between connected nodes, which explains its good partition properties. In the low rank approximation this metric can produce very useful displays of the structure of the network.

We applied methods of Correspondence Analysis to give quantitative meanings to the eigenvalues and the components of the eigenvectors: the Convergence Model is very closely related to Spectral Analysis of the χ^2 matrix. We do not generally work in the χ^2 metric, because it adds an extra level of complexity, although χ^2 distances are quite easy to calculate when needed. We showed the connection between the Convergence Model and the Normal method. Although the Normal method differs from the Standard method in that it, like the Convergence Model, divides each row by its row sum, the Normal method may be used to find both clustering and colourings, while the Convergence Model finds only clusterings.

The Convergence Model uses the augmented Markov matrix M and gives positive eigenvalues. Here, nodes with connections to one another tend to move towards one another. But we saw that the negative eigenvectors also contain important structural information. When we used M* (the Normal method), we were able to get both positive and negative eigenvectors. When we change the diagonal, as we do when going from M to M*, we move away from the convergence model; we have switched to an analytic method. Table 1, on the next page, summarizes the differences between Correspondence Analysis, the Normal Method, and the Standard Method.

Finally, we addressed a number of practical matters of interest to the researcher who wishes to apply the methods we are developing to large sparse matrices. Both partial decomposition and partial iteration can give very useful descriptions of network structure.

Table 1. Comparison of three approaches

Correspondence Analysis	Normal Method	Standard Method
any contingency table (not necessarily square or symmetric)	adjacency matrices (square symmetric non-negative matrices with zeros along the diagonal)	adjacency matrices (square symmetric non-negative matrices with zeros along the diagonal)
uses the χ^2 metric	uses the D_r metric	Euclidian metric
returns non-negative eigenvalues $0 \le \lambda < 1$	returns both positive and negative eigenvalues $-1 \le \lambda \le 1$	returns both positive and negative eigenvalues
uses unadjusted diagonal	diagonal is loaded with row sums weighted by the damping factor d	diagonal is either not loaded or is loaded with a constant
colouring vs. clustering difficult to determine without examining both left and right eigenspaces	positive eigenvalues give clustering; negative ones give colouring	same as Normal Method but not as direct
gives Convergence Model when the diagonal is loaded with row sums	gives Convergence Model when $d = 0.5$	N.A. (row sums constant only for regular graphs)
distributional equivalence applies	distributional equivalence applies	distributional equivalence *does not* apply
trivial dimension has eigenvalue 0.0	trivial dimension has eigenvalue 1.0	no trivial dimension, except for regular graphs
trivial dimension not used in analysis	trivial dimension is long-term convergence	no trivial dimension, except for regular graphs
eigenvectors are solutions for maximizing and minimizing total χ^2 distance of connected nodes	eigenvectors are solutions for maximizing and minimizing total D_r distance of connected nodes	eigenvectors are *not* solutions for maximizing and minimizing total distance of connected nodes

The Convergence Model describes stable, closed networks in which the pattern of connections doesn't change over time and there are no inputs from the environment. In large complex social systems, such as cities or countries in which there is a great deal of complexity and in which patterns of connections are constantly shifting, a statistical approach that ignores the micro-level details and focusses instead on the overall cognitive space will be more useful if the goal is to describe changes in cognitive structures. However, when the focus is on patterns of interaction or the flow of information or influence, it will be more useful to examine the structure of the network of connections. When interactions draw individuals towards one another, the Convergence Model will be both a useful description of the behavior of the system and a method for analyzing the patterns of relationships. When relationships differentiate individuals (e.g. relationships involving differential status or power) or when the relationship under consideration can only join individuals who belong to different categories (e.g. landlord/tenant, employer/employee), there is still convergence, but it is not the convergence described by the Convergence Model. Instead of bringing together individuals who are linked *to one another*, it brings together individuals who are linked *to similar sets of individuals*[18]. Here the Normal method is appropriate, both as a model of the system's behavior and as an analytic tool for investigating its structure. Convergence Analysis thereby provides a unified model that unites two seemingly disparate approaches to network analysis. We believe this eliminates the need to choose whether to focus on either cohesion or structural equivalence, thereby allowing more valid descriptions of networks.

References

Acton, F. (1990). *Numerical Methods That Work*. Harper and Row

Allen, R.K. (1970). *A Comparison of Communication Behaviors in Innovative and Non-Innovative Secondary Schools*. Ph.D. Thesis, East Lansing, Michigan State University.

Aspvall, B., Gilbert, J. (1984). "Graph Coloring using Eigenvalue Decomposition." *SIAM J. Alg. Disc. Math*. Vol. 5 No. 4 Dec. 1984 526-538

Back, K. (1951). "Influence through social communication." *J. Abnormal Social Psychology, 46*: 9-23.

Barnes, E. (1982). "An Algorithm for Partitioning the Nodes of a Graph." *SIAM J. Alg. Disc. Math*. Vol.3 No. 4 Dec. 1982 541-550

Barnett, G., (1993). "Correspondence Analysis: A Method for the Description of Communication Networks." In W. Richards, and G. Barnett (Eds.) *Progress in Communication Sciences XII.* Ablex 135-164

Barnett (1973). "On the nature of random structure". Department of Communication, Michigan State University, East Lansing.

Barnett, G.A. & D.L. Kincaid (1983). "Cultural covergence: A mathematical theory." In W. Gudykunst (Ed.) *International Intercultural Communicaiton Annual VII*. Beverly Hills: Sage.

Barnett, G.A. & Woelfel, J. (1988). *Readings in the Galileo System: Theory, Method and Applications*. Kendall/Hunt: Dubuque, IA.

Becker, C. (1993). *A Communication Based Theory of Cultural Reproduction and Transformation: The Case of Brazilian Immigrants in Japan*. Unpublished Ph. D. dissertation. Department of Communication: State University of New York at Buffalo.

Benzécri, L. (1969). "Statistical Analysis as a Tool to Make Patterns Emerge from Data." In S. Watanabe (Ed.) *Methodologies of Pattern Recognition*, Academic Press 35-74

Biggs, N. (1993). *Algebraic Graph Theory*. New York: Cambridge University Press

Courant, R. and Hilbert, D. (1966). *Methods of Mathematical Physics*. Interscience Publishers

Curtis, C. (1974). *Linear Algebra: An introductory Approach.* Allyn and Bacon

Cvetkovic, D., Doob, M., Gutman, I., Torgasev, A. (1988). *Recent Developements in the Theory of Graph Spectra*. North-Holland

Danowski, J. (1980). "Group attitude conformity and connectivity of organizational communication networks for production, innovation and maintenance content." *Human Communication Research*, *6*: 299-308.

Dean, J. & Brass, D. (1985). Social interaction and the perception of job characteristics in an organization. *Human Relations*, *38*, 571-582.

Dow, G. (1988). Configurational and coactivational views of organizational structure. *Academy of Management Review*, *12*(1), 53-64.

Erickson, B. (1988). "The relational basis of attitudes." in B. Wellman & S. Berkowitz (Eds.), *Social structures: A network approach*. New York: Cambridge University Press, 99-121.

Feller, W. (1950). *Probability Theory and its Applictions*. New York: John Wiley and sons

French, J.R.P. (1956). "A formal theory of social power." *Psychological Review 63*: 181-194.

French, J.R.P., and Snyder, R. (1957). "Leadership and interpersonal power." In D. Cartwright *et al. Studies in social power.* Ann Arbor: Institute for Social Research.

Garey, M., Johnson, D. (1979). *Computers and Intractability*. New York: W. H. Freeman

Greenacre, M., (1984). *Theory and Appication of Correspondence Analysis*. Academic Press

Hall, K. (1970). "An r-Dimensional Quadratic Placement Algorithm." *Management Science 17*: 3, 219-229

Heider, F. (1946). "Attitudes and cognitive organization." *Journal of Psychology, 21*: 107-112.

Hoffman, D., Franke, G. (1986). "Correspondence Analysis: Graphical Representation of Categorical Data in

Marketing Research." *J. Market Research 23*: 213-227

Hovland, C., and Weiss, W. (1952). "The influence of source credibility on communication effectiveness." *Public Opinion Quarterly, 15*: 635-650.

Johnson, J. C., & Miller, M. L. (1986). Behavioral and cognitive data: A note on the multiplexity of network subgroups. *Social Networks*, *8*, 65-77.

Jolliffe, I. T. (1986). *Principal Components Analysis*. Springer-Verlag

Kincaid, D. L. (1987). "The convergence theory of communication, self-organization, and cultural evolution." In D. L. Kincaid (Ed.), *Communication Theory: Eastern and Western Perspectives* (pp. 209-221). New York: Academic Press.

Kincaid, D. L., Yum, J. O., Woelfel, J., & Barnett, G. (1983). "The cultural convergence of Korean immigrants in Hawaii: An empirical test of a mathematical theory." *Quality and Quantity, 18*: 59-8.

Kincaid, D. L. (1988). "The convergence theory of communication: Its implications for intercultural communication." In Y. Y. Kim (Ed.), International and Intercultural Annual: Vol. 12. *Theoretical Perspectives on International Communication* (pp. 280-298). Beverly Hills: Sage.

Laumann, E. & Marsden, P. (1979). "The analysis of oppositional structures in political elites: Identifying collective actors." *American Sociological Review*, *44*: 713-732.

Lin, N. (1966). *Innovative Internalization in a Formal Organization.* Ph.D. Thesis, East lansing, Michigan State University.

Lippitt, P., Polanski, N., and Rosen, S. (1952). "The dynamics of power." *Human Relations, 5*; 37-64.

Newcomb, T.M. (1953). "An approach to the study of communicative acts." *Psychological Review, 60*: 393-404.

Parlett, B. (1980). *The Symmetric Eigenvalue Problem*. Prentice-Hall

Parlett, B. (1979). "The Lanczos Algorithm with Selective Orthogonalization" *Mathematics of Computation, 33*:145, 217-238

Press, W., Flannery, B., Teukolsky, S., Vetterling, W., (1986). *Numerical Recipes*. New York: Cambridge University Press

Richards, W.D. & Seary, A.J. *Eigen Analysis of Networks* (working title). Forthcoming. Newark: Hampton Press.

Rogers, E. M., & Kincaid, D. L. (1981). *Communication Networks: Toward a New Paradigm for Research*. New York: Free Press.

Seary, A.J. & Richards, W.D. (1995) "Partitioning Networks by Eigenvectors." Presented to European Netowrk Conference, London. Submitted for publication in *Social Networks*.

Senata, E. (1981). *Non-negative Matrices and Markov Chains.* Springer-Verlag

Shoemaker, F.F. (1971). *System Variables and Educational Innovativeness in Thai Government Secondary Schools.* Ph.D. Thesis, East Lansing, Michigan State University.

Tushman, M. & Romanelli, E. (1983). Uncertainty, social location and influence in decision making: A sociometric analysis. *Management Science*, *29*(1), 12-23.

Weller, S., Romney, A. (1990). *Metric Scaling: Correspondence Analysis.* Beverly Hills: Sage

Woelfel, J. & Saltiel, J. (1978). "Cognitive processes as motions in a multidimensional space." In F. Casmir (Ed.). *International and intercultural communicaiton*. New York: University Press.

Woelfel, J. & Fink, E.L. (1980). *The measurement of communication processes: Galileo theory and method*. New York: Academic Press.

Zaltman, G. and others (1973). *Innovation s and Organizations.* New York: Wiley-Interscience.

1.The matrix *A*, used throughout the chapter to illustrate various aspects of the discussion, is symmetric. The model of convergence described in Part 1 of the chapter works equally well for most asymmetric networks (for which the adjacency matrix will also be asymmetric). The exceptions are what French calls "weakly connected" networks —

directed networks in which there is at least one pair of nodes for which there is no path from one to the other (1956, p. 187), in which some nodes do not converge. We use a symmetric *A* because space limitations do not allow the more complex presentation that would otherwise be required. See our comments on this issue in Part 2. See also Seary, A. & Richards, W.D. (1995) and Richards, W.D. & Seary, A. (forthcoming) for a more complete treatment.

2. In a positive semi-definite matrix, all eigenvalues are non-negative.

3.The Lagrange multiplier is a standard method for finding maxima and minima of an unknown subjected to a constraint.

4. Equations of this form arise when we consider a set of objects with masses *r*, connected by springs with restoring forces *C*. This model is used in Barnett and Kincaid (1983) and Kincaid, *et al.* (1983). This research is discussed in Part 1.

5. The equality $\lambda_n = 0$ is reached only for bipartite graphs, which generally have the $\begin{vmatrix} 0 & B \\ B^T & 0 \end{vmatrix}$ form shown on the right,

where *B* is a matrix of **1**'s (and probably some **0**'s), and *0* is a matrix of **0**'s of appropriate size.

6. i.e. $p' = Mp,$
$p'' = Mp' = MMp = M^2 p,$
$p''' = Mp'' = MMMp = M^3 p \ \ldots$

7. Both Barnett (1989) and Kincaid (1993) do eigen decomposition of distance matrices in which the i,j entry is the number of steps in the shortest path from *i* to *j*. Barnett gives special meaning to negative eigenvalues. Sincee are *not* dealing with the geometry of distance matrices, his conclusions do not apply.

8. The amount of time required to perform the calculations is a function of n^3 and the amount of memory required varies as a function of n^2

9. Note that we get *M* by adding *row sums* to *A*, rather than some constant value as is usually done to make a matrix positive definite (Weller & Romney, p. 71, 1990; Acton, 204-220, 1990). Adding a constant to the diagonal changes the eigenvectors and thus gives a distorted picture of the structure of the network.

10. If D_r were the identity matrix — i.e. if we were working with the Euclidean metric — it *would* be the standard eigenvalue problem.

11.   NOTE: It is always worthwhile, especially with a large sparse matrix, to calculate χ^2 first to make sure there is some signal, and that it is significant.

12. This is an example of Deflation.

13. *G* is thus VD_λ where *V* was the matrix introduced in the previous section.

14. Damping is also considered in Kincaid, *et al.* (1983) and in Barnett and Kincaid (1983).

15. Yamaguchi (1994, p. 63-64) uses this approach to model actor's "self-interest" — their resistance to give up control. In the convergence model described in Part 1, it serves to dampen the movement of positions from iteration to iteration.

16. This is the procedure used in NEGOPY.

17. If the initial position vector *p* happens to correlate *perfectly* with one of the eigenvectors, *p* will be unchanged by iterating.

18. This is the goal of blockmodelling and structural equivalence, approaches used by many sociologists.