Sociology 375 Exam 1 Fall 2010 Prof Montgomery

Answer all questions. 220 points possible. You may be time-constrained, so you should allocate your time carefully.

1) [70 points] Consider relation R = {(1,4), (2,1), (2,3), (2,4), (4,3)} on S = {1, 2, 3, 4}.

a) Show how the relation R can be represented as

(i) an adjacency matrix A
(ii) a (directed) graph

b) Find (by computation or inspection)

(i) the number of 2-paths between each ordered pair of actors
(ii) the distance matrix
(iii) the reachability matrix

c) In general, given a relation represented as directed graph, how can you tell (using the graph) whether the relation is asymmetric? In the present example, is the relation R asymmetric? If not, list any violations of the asymmetry condition.

d) State the matrix test for transitivity, and then show whether A passes this test. If not, list any violations of the transitivity condition.

e) Is R a strict partial order? If it is, then draw the Hasse diagram. If not, compute the transitive closure matrix [= $(A + A^2 + A^3 + A^4)\#$], and then draw the Hasse diagram of the transitive closure.

2) [25 points] Consider an arbitrary adjacency matrix A which is $n \times n$.

a) State the formula for the reachability matrix (given adjacency matrix A).

b) For each of the properties below, indicate whether the reachability relation will *always*, *sometimes*, or *never* have this property. If you answer *sometimes*, then explain what property of the A matrix would determine the outcome.

(i) reflexive
(ii) symmetric
(iii) transitive

c) Given the answer to part (b), what condition(s) on the A matrix determine whether reachability is an equivalence relation? What do social network analysts call the equivalence classes of the reachability relation?

3) [20 points] There is a well-known mathematical result (first proven by Erdős) that describes how, fixing the number of nodes in a random graph, the expected proportion of nodes in the largest component changes as you vary the average degree of the nodes. Briefly describe this result. Why might it initially seem surprising?

4) [105 points] Consider a social relation on a set of actors {1, 2, 3, 4, 5} characterized by the adjacency matrix below. [HINT: It may be useful to draw the graph.]

A =

0	1	1	0	0
1	0	0	1	0
1	0	0	1	0
0	1	1	0	1
0	0	0	1	0

a) State the definition of an n-clique, then list all of the 2-cliques in this network. Is each 2-clique also a 2-clan? What additional test did you need to perform to decide?

b) For each of the following pairs of actors, find a maximal set of node-independent paths between the actors, and then state the local connectivity for this pair of actors.

(i) actors 2 and 3 (ii) actors 1 and 4 (iii) actors 1 and 5

c) What is the connectivity level of the full graph? Give a cutset. Are there any subgraphs with a higher connectivity level? If so, indicate the connectivity level, the actors in the subgraph, and give a cutset. If not, explain why.

d) Partition the set of actors into structural equivalence classes. Give the blockmodel.

e) What partition of actors is indicated by the result of the CONCOR algorithm below?

```
>> X = [A;A']; for i = 1:20; X = corrcoef(X); end; X
X =
   1.0000  -1.0000  -1.0000   1.0000  -1.0000
  -1.0000   1.0000   1.0000  -1.0000   1.0000
  -1.0000   1.0000   1.0000  -1.0000   1.0000
   1.0000  -1.0000  -1.0000   1.0000  -1.0000
  -1.0000   1.0000   1.0000  -1.0000   1.0000
```

f) For this example, the subsets in the CONCOR partition (in part e) are also regular equivalence classes. Report the appropriate matrix test to confirm this result, then give the blockmodel. [HINT: You may need to rearrange the rows and columns of A.]

g) Discuss the difference between the 1-blocks in the structural equivalence blockmodel (part d) and those in the regular equivalence blockmodel (part f).

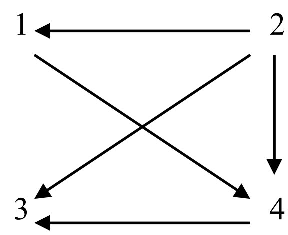
h) Find a partition of the set of actors that does *not* satisfy the regular equivalence test. Report the partition and the matrix test result. Then given one reason why this partition failed the regular equivalence test.

Soc 375 **Exam 1** **Fall 2010** **SOLUTIONS**

1a) [12 pts]

(i) $A = \begin{bmatrix} 0 & 0 & 0 & 1 \\ 1 & 0 & 1 & 1 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix}$

(ii)



b) [18 pts] (i) (ii) (iii)

$$A^2 = \begin{bmatrix} 0 & 0 & 1 & 0 \\ 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix} \qquad D = \begin{bmatrix} 0 & \infty & 2 & 1 \\ 1 & 0 & 1 & 1 \\ \infty & \infty & 0 & \infty \\ \infty & \infty & 1 & 0 \end{bmatrix} \qquad R = \begin{bmatrix} 1 & 0 & 1 & 1 \\ 1 & 1 & 1 & 1 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 1 & 1 \end{bmatrix}$$

c) [10 pts] The graph of an asymmetric relation must have no loops, and no edge between distinct nodes with arrows in both directions. R is asymmetric.

c) [15 pts] The matrix test for transitivity is $A \geq (A^2)\#$. In this case,

$$A - A^2\# = \begin{bmatrix} 0 & 0 & -1 & 1 \\ 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix}$$

Thus, R is not transitive. There is a 2-path from node 1 to node 3, but no 1-path.

e) [15 pts] No, R is not a strict partial order (because it is not transitive). The transitive closure of R is characterized by the adjacency matrix T, which generates a Hasse matrix H, which generates the Hasse diagram below.

$$T = \begin{bmatrix} 0 & 0 & 1 & 1 \\ 1 & 0 & 1 & 1 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix} \qquad H = T - T^2\# = \begin{bmatrix} 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix}$$

2a) [6 pts] Given an $n \times n$ adjacency matrix A, the reachability matrix is given by $R = (I + A + A^2 + \ldots + A^{n-1})\#$.

b) [13 pts] i) always; ii) sometimes (if A is symmetric then R is symmetric); iii) always

c) [6 pts] If A is symmetric then reachability is an equivalence relation (given that it is reflexive, symmetric, and transitive). In social network analysis, the equivalence classes of the reachability relation are called "weak cliques."

3) [20 pts] To restate the formal theorem: Consider a random graph with n nodes and e edges, so that average degree $k = 2e/n$. Let s denote the proportion of nodes in the largest component. Given n very large, the relationship between s and k is given by $s = 1-e^{-ks}$.

More intuitively, a random graph with a very small number of edges (relative to the fixed number of nodes) will have many small components. As the number of edges rises, the expected number of nodes in the largest component rises slowly until the number of edges is approximately equal to the number of nodes (i.e., $e \approx n$, which implies $k \approx 1$). When the number of edges rises beyond that level, the expected size of the largest component grows very rapidly. Additional edges are likely to be "bridges" connecting previously unconnected components. Perhaps surprisingly, this "bridging" occurs at relatively low network density, when the number of edges is still very low relative to the number of edges possible (= (1/2)n(n-1) given n nodes). To illustrate, given 100 nodes, "bridging" occurs at a network density of about $[100]/[(1/2)(100)(99)] \approx 2\%$.

4a) [18 pts] An n-clique is a maximal set of nodes with distance is less than or equal to n for every pair of nodes. For the present example, there are 2 different 2-cliques: {1, 2, 3, 4} and {2, 3, 4, 5}. Each of these 2-cliques is also a 2-clan because the distance remains less than or equal to 2 within the 2-clique subgraph.

b) [15 pts]	(i)	{(2,1,3), (2,4,3)}	connectivity(2, 3) = 2
	(ii)	{(1, 2, 4), (1, 3, 4)}	connectivity(1, 4) = 2
	(iii)	{(1, 2, 4, 5)} or {(1, 3, 4, 5)}	connectivity(1, 5) = 1

c) [15 pts] The full graph is a 1-component. The cutset is {4}. The subgraph with nodes {1, 2, 3, 4} is a 2-component. The cutset is {1, 4} or {2, 3}.

d) [10 pts] {{1}, {2, 3}, {4}, {5}}

0	1	1	0	0
1	0	0	1	0
1	0	0	1	0
0	1	1	0	1
0	0	0	1	0

e) [5 pts] {{1, 4}, {2, 3, 5}}

f) [16 pts]

Representing the partition in part (e) as matrix E, the matrix test is (AE)# = (EA)#.

$$E = \begin{matrix} 1 & 0 & 0 & 1 & 0 \\ 0 & 1 & 1 & 0 & 1 \\ 0 & 1 & 1 & 0 & 1 \\ 1 & 0 & 0 & 1 & 0 \\ 0 & 1 & 1 & 0 & 1 \end{matrix} \qquad (AE)\# = (EA)\# = \begin{matrix} 0 & 1 & 1 & 0 & 1 \\ 1 & 0 & 0 & 1 & 0 \\ 1 & 0 & 0 & 1 & 0 \\ 0 & 1 & 1 & 0 & 1 \\ 1 & 0 & 0 & 1 & 0 \end{matrix}$$

Rearranging the rows and columns of the A matrix (so that the ordering is now 1, 4, 2, 3, 5), the blockmodel is

$$\begin{array}{cc|ccc} 0 & 0 & 1 & 1 & 0 \\ 0 & 0 & 1 & 1 & 1 \\ \hline 1 & 1 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \end{array}$$

g) [6 pts] In the blockmodel based on structural equivalence, every element in a 1-block must be equal to 1. In the blockmodel based on regular equivalence, each column and each row of a 1-block must have at least one element equal to 1. Thus, regular equivalence is a weaker condition.

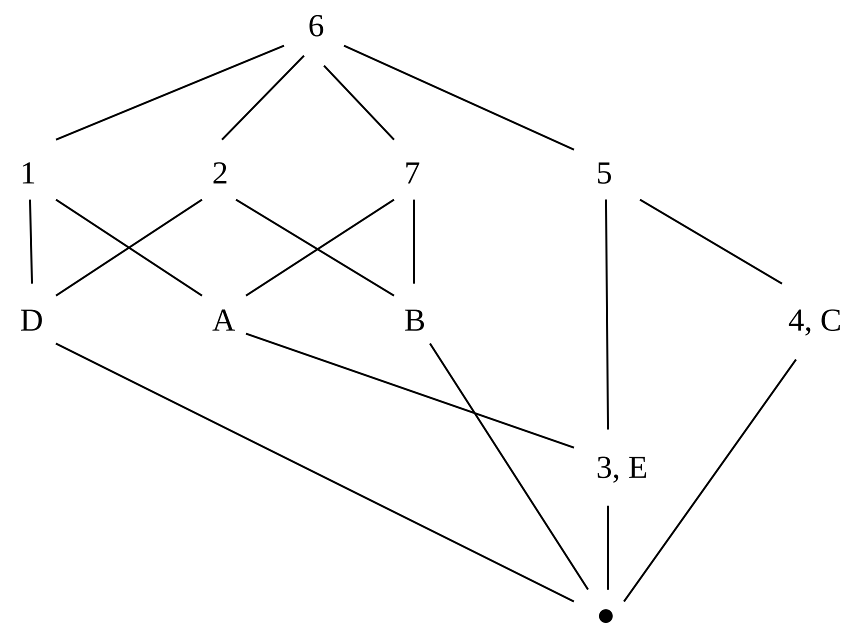
h) [20 pts] Many answers are possible. For instance, the partition {{1}, {2, 3, 4, 5}} fails the RE test. Actors 2, 3, 4, 5 are placed in the same class, but 2 and 3 are linked to someone in 1's class (in particular, they are linked to 1 himself) while 4 and 5 are not linked to someone like 1. Thus, 2 and 3 can't belong to the same class as 4 and 5.

$$E = \begin{matrix} 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 1 & 1 & 1 \\ 0 & 1 & 1 & 1 & 1 \\ 0 & 1 & 1 & 1 & 1 \\ 0 & 1 & 1 & 1 & 1 \end{matrix} \qquad (AE)\# = \begin{matrix} 0 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & 1 \\ 0 & 1 & 1 & 1 & 1 \\ 0 & 1 & 1 & 1 & 1 \end{matrix} \qquad (EA)\# = \begin{matrix} 0 & 1 & 1 & 0 & 0 \\ 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & 1 \end{matrix}$$

Sociology 375 Final Exam Fall 2010 Prof Montgomery

Answer all questions. 240 points possible.

1) [50 points] Seven students (labeled 1 to 7) took an exam with five questions (labeled A to E). Each student answered each question either correctly or incorrectly. The results are indicated by the following Galois lattice. (The lattice has been given reduced labeling. It is oriented so that the empty set of questions is at the bottom; the universal set of questions is at the top. Thus, the "easier" questions are placed closer to the bottom of the lattice.)



a) Using the lattice, find the 7 × 5 data matrix D summarizing the exam results (with D(i,j) = 1 if student i answered question j correctly, and D(i,j) = 0 otherwise).

b) The HICLAS procedure yielded a rank-2 approximation with the row-bundle matrix S and the column-bundle matrix P below. Use the S and P matrices to find the estimated matrix M and draw the rank-2 lattice.

$$S = \begin{bmatrix} 0 & 1 \\ 0 & 1 \\ 0 & 0 \\ 1 & 0 \\ 1 & 0 \\ 1 & 1 \\ 0 & 1 \end{bmatrix} \qquad P = \begin{bmatrix} 0 & 1 \\ 0 & 1 \\ 1 & 0 \\ 0 & 1 \\ 0 & 0 \end{bmatrix}$$

c) Comparing the data matrix D from part (a) to the estimated matrix M from part (b), how good is the approximation? Compute the relevant measure.

2) [30 points]

a) According to the Structure Theorem, there are two (logically equivalent) conditions that determine whether a signed graph is balanced. State these two conditions.

b) Similarly, there are two (logically equivalent) conditions that determine whether a signed graph is clusterable. State these two conditions.

c) For each of the following cases, state whether it is *possible* or *impossible* to construct a signed graph with the pair of properties given.

i) balanced and clusterable
ii) balanced and not clusterable
iii) clusterable and not balanced
iv) not clusterable and not balanced

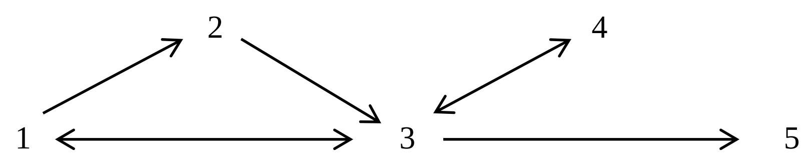
3) [70 points]

a) Directed graphs have 16 types of triads. Holland and Leinhardt referred to these types with the following labels (that have now become conventional in network analysis):

003	021U	021C	030C	111D	120U	120C	210
012	021D	030T	102	111U	120D	201	300

For each type of triad, draw the graph and indicate whether the triad is transitive or not transitive.

b) For the following graph, what is the number of dyads? Give the dyad census. What is the number of triads? List each triad (using the node labels) and then give its type (from the list of 16 types).



c) How many 003 triads appear in the graph above? How many would have been expected to appear "by chance" given the dyad census? (Answer the latter question under both the assumption that dyads are sampled with and without replacement.)

4) [40 points]

In Correspondence Analysis, the underlying data is a rectangular two-mode matrix P. (For concreteness, suppose P(i,j) denotes the number of species i at site j.) The P matrix is used to compute a square matrix. The eigenvectors and eigenvalues of this square matrix are then used to draw a two-dimensional scatterplot.

a) What is the formula for the square matrix? Be sure to define your notation. Explain how this formula can be motivated by the "reciprocal averaging" method.

b) Once you've computed the eigenvectors and eigenvalues of the square matrix, how do you determine the location of each point in the scatterplot? How many points are plotted? What is the interpretation of the scatterplot diagram?

5) [50 points]

Consider a kinship system in a society with 5 clans. The W and C matrices are given by

$$W = \begin{bmatrix} 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 & 0 \end{bmatrix} \qquad C = \begin{bmatrix} 0 & 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \end{bmatrix}$$

a) Draw the graph of the W and C relations on the set of clans. Use solid arrows for the W relation, and dotted arrows for the C relation. [HINT: It will be useful to arrange the nodes in a clockwise pattern.]

b) Find the mathematical group generated by the W and C matrices, listing the elements and computing the multiplication table for the group. [HINT: Looking at the graph in part (a), it should be obvious that $I = W^5$ and that $C = W^{-1} = W^4$.]

c) Is the above kinship system g-balanced? Explain why (or why not), giving any relevant computations/demonstrations. According to Boyd (*J Math Psych* 1969), what are the substantive interpretations of the g-balance conditions?

d) We might generalize this problem so that the society has n clans (where n is chosen arbitrarily), with $W(i,i+1) = C(i+1, i) = 1$ for all $i \in \{1, \ldots, n-1\}$, and $W(n,1) = C(1,n) = 1$. (Note that you've already considered the special case where $n = 5$.) Is the kinship system g-balanced for any $n \geq 2$? Briefly explain.

Soc 375 Exam 2 Fall 2010 Solutions

1a) [14 pts]

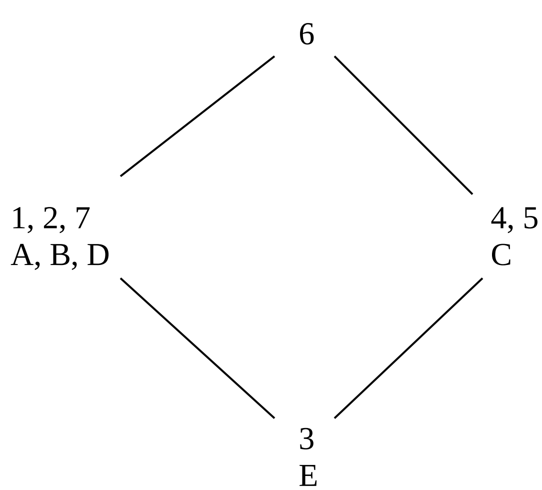
$$D = \begin{bmatrix} 1 & 0 & 0 & 1 & 1 \\ 0 & 1 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 & 1 \\ 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 0 & 0 & 1 \end{bmatrix}$$

b) [30 pts]

Using Matlab notation, M = ~(~S*P'). Conceptually, M(i,j) = 1 if row bundle S(i,:) weakly contains column bundle P(j,:). Here,

$$M = \begin{bmatrix} 1 & \underline{1} & 0 & 1 & 1 \\ \underline{1} & 1 & 0 & 1 & \underline{1} \\ 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 & \underline{1} \\ 0 & 0 & 1 & 0 & 1 \\ 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 0 & \underline{1} & 1 \end{bmatrix}$$

The lattice can be drawn as



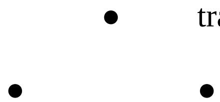
c) [6 pts] Comparison of the M and D matrices reveals 5 discrepancies (underlined in the M matrix).

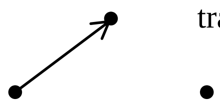
2a) [12 pts] (1) No negative cycle. (2) Nodes can be partitioned into 2 subsets (one possibly empty) such that there are no positive edges between subsets and no negative edges within subsets.

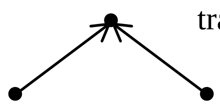
b) [10 pts] (1) No cycle with exactly one negative edge. (2) Nodes can be partitioned into $n \geq 2$ subsets ("clusters") such that there are no positive edges between clusters and no negative edges within clusters.

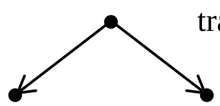
c) [8 pts] Balance implies clusterability. (Note that any balanced graph has 2 clusters.) Thus, (i) and (iii) and (iv) are possible, while (ii) is impossible.

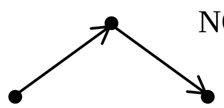
3a) [30 pts]

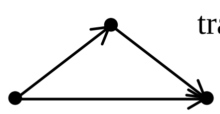
003  transitive

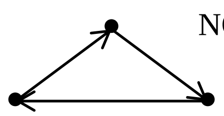
012  transitive

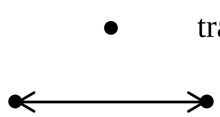
021U  transitive

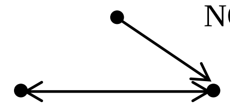
021D  transitive

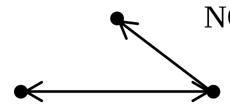
021C  NOT transitive

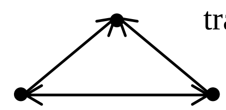
030T  transitive

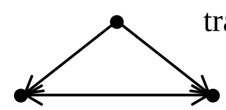
030C  NOT transitive

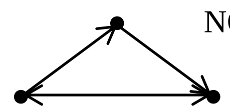
102  transitive

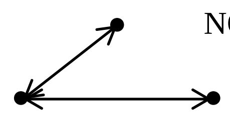
111D  NOT transitive

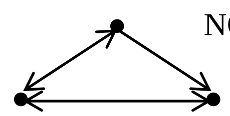
111U  NOT transitive

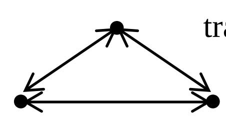
120U  transitive

120D  transitive

120C  NOT transitive

201  NOT transitive

210  NOT transitive

300  transitive

3b) [30 pts] In general, a graph with g nodes has g(g-1)/2 dyads. Given g = 5 nodes, there are 10 dyads. By inspection of the graph, the dyad census is M = 2, A = 3, N = 5.

In general, a graph with g nodes has g(g-1)(g-2)/6 triads. Given g = 5 nodes, there are 10 triads. I've listed each triad along with its type:

{1,2,3} 120C	{1,3,5} 111U	{2,4,5} 003
{1,2,4} 012	{1,4,5} 003	{3,4,5} 111U
{1,2,5} 012	{2,3,4} 111D	
{1,3,4} 201	{2,3,5} 021C	

3c) [10 pts] There are two 003 triads. Sampling dyads with replacement, the expected number would be (5/10)*(5/10)*(5/10)*(10 triads) = 1.25 triads. Sampling dyads without replacement, the expected number would be (5/10)*(4/9)*(3/8)*(10 triads) = .833 triads. Either way, the observed number of 003 triads is greater than what would be expected by chance.

4a) [20 pts] The formula for the square matrix is $C^{-1} P' R^{-1} P$ where C is a diagonal matrix giving the column sums of P (hence C^{-1} is a diagonal matrix giving the reciprocals of the column sums of P) and R is a diagonal matrix giving the row sums of P (hence R^{-1} is a diagonal matrix giving the reciprocals of the row sums of P).

Using reciprocal averaging, each species is given a species score which is a weighted average of site scores (with the weights given by the proportion of the species located at each site). Similarly, each site is given a site score which is proportional to a weighted average of species scores (with the weights given by the proportion of the site occupied by each species). Formally, this yields the pair of equations

$$u = R^{-1} P v \qquad \text{and} \qquad \lambda v = C^{-1} P^T u$$

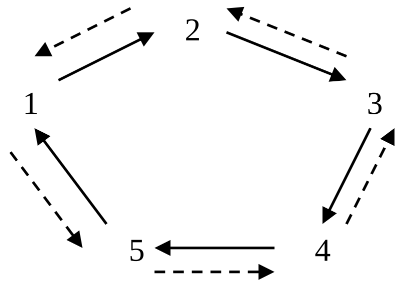
The factor $\lambda < 1$ is necessary for a non-trivial solution. (Given $\lambda = 1$, the equations yield the trivial solution where u(i) = v(j) = c for all i and j.) Substituting the first equation into the second, we obtain

$$\lambda v = [C^{-1} P^T R^{-1} P] v$$

Note that v is an eigenvector and λ is an eigenvalue of the square (bracketed) matrix.

b) [20 pts] Each site j is plotted at $\{v_2(j), v_3(j)\}$ where v_2 and v_3 are the eigenvectors corresponding to the second and third largest eigenvalues. (We ignore v_1 corresponding to the trivial solution where $\lambda = 1$.) Each species i is plotted at $\{u_2(i), u_3(i)\}$ where $u_k = R^{-1} P v_k$. In this way, we plot n species and m sites. On the scatterplot, the location of each species is a weighted average of the site locations. Species with more similar distributions over sites will thus appear nearer to each other.

5a) [6 pts]



b) [18 pts]

	I	W	W^2	W^3	C
I	I	W	W^2	W^3	C
W	W	W^2	W^3	C	I
W^2	W^2	W^3	C	I	W
W^3	W^3	C	I	W	W^2
C	C	I	W	W^2	W^3

Note that other labels might have been used for the elements (e.g., $W^3 = C^2$).

c) [8 pts] Yes, this system is g-balanced. Constructing the graph of the multiplication table (using the W and C columns of the table in part b), it is obvious that this graph is isomorphic to the graph of W and C on the set of clans (given in part a). Formally, to prove that an isomorphism exists, we need to construct a function ϕ that maps each node of one graph into a node of the other graph, and then show that edge (i,j) is present in the first graph if and only if edge $(\phi(i), \phi(j))$ is present in the second graph. In the present case, one obvious isomorphism is $\phi(1) = I$, $\phi(2) = W$, $\phi(3) = W^2$, $\phi(4) = W^3$, $\phi(5) = C$.

Boyd proved that g-balance holds if and only if every path between clans has the same sign (given by the multiplication table), and if and only if every cycle from a clan to itself has the sign of the identity element of the group. Boyd suggests that the "path" condition reflects "normative consistency" (given that members of one clan should have consistent normative obligations to members of another clan, regardless of the particular path of W and C relations connecting clans). Boyd suggests that the "cycle" condition reflects "consistent self-evaluation" (much like the cycle condition in balance theory reflects positive self-evalution).

d) [8 pts] Yes, the system would be g-balanced for any $n \geq 2$. The W and C matrices would generate a group with n elements: $\{I, W, W^2, W^3, \ldots, W^{n-2}, C\}$. Both the graph of W and C on the set of clans and the graph of the multiplication table will have the kind of "circular" pattern shown in part (a), and will thus be isomorphic.