Sociology 375 Exam 1 Fall 2007 Prof Montgomery

Answer all questions. 250 points possible.

1) [30 points] The antisymmetry condition can be written as

$$(xRy \wedge yRx) \rightarrow (x = y) \quad .$$

Using a truth table, determine whether the condition

$$\neg(x = y) \vee \neg(xRy) \vee \neg(yRx)$$

is equivalent to the antisymmetry condition. [HINT: Applying some elementary rules of logic, you could potentially answer this question without a truth table, but you must construct the truth table to receive credit on this question.]

2) [20 points] a) State the (set-theoretic) conditions for symmetry and asymmetry.

b) Consider the empty relation $R = \varnothing$ on a non-empty set S. Describe (in words) how this relation would appear if represented as a graph or as an adjacency matrix. Is this relation symmetric, asymmetric, both, or neither? Explain why. [HINT: Use the conditions that you stated in part (a).]

3) [90 points] Consider the relation R = {(1,2), (1,3), (3,2), (3,4), (4,2)} on the set S = {1, 2, 3, 4}.

a) Show how the relation R could be represented

i) as a (directed) graph
ii) as an adjacency matrix
iii) using infix notation

b) Determine (by computation or inspection)

i) the number of 2-paths between each pair of individuals
ii) the number of 3-paths between each pair of individuals
iii) the reachability matrix
iv) the distance matrix

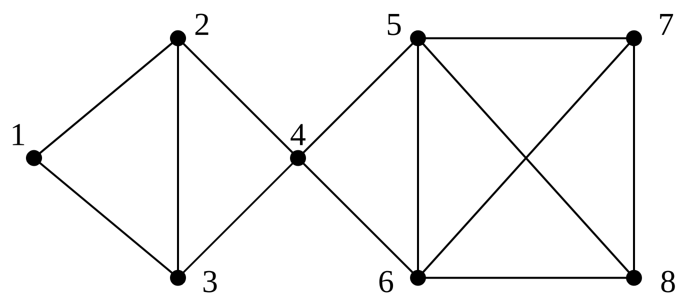
c) State the matrix test for transitivity. Using this test, determine whether the relation R is transitive. If not, list any violations of transitivity.

d) Give the transitive closure of R. Is the transitive closure of R an equivalence relation? If not, explain why. If so, give the equivalence classes. Is the transitive closure of R a strict partial order? If not, explain why. If so, draw the Hasse diagram.

4) [15 points] Suppose you are given a symmetric adjacency matrix A that generates the reachability matrix R. Now consider the matrix U composed of the unique rows of the reachability matrix. (Using matlab, this matrix is computed as U = unique(R, 'rows').) For purposes of social network analysis, what do we learn by counting the number of rows of U (i.e., by computing size(U,1) in matlab)? What do we learn by computing the row sums of U (i.e., by computing sum(U′)′ in matlab)? What answer would we get if we computed the *column* sums of U (i.e., sum(U) in matlab)? Briefly explain.

5) [25 points] Briefly describe the main findings in the two papers by Duncan Watts (*American Journal of Sociology* 1999, *Science* 2002) that we covered in lecture. [HINT: Two or three sentences about each paper should be sufficient. You might also want to use some diagrams to describe these findings. If so, make sure that they are properly labeled.]

6) [70 points] Consider the social relation A on the set of individuals {1,…, 8} which is represented by the following graph:



The A relation is also represented by the adjacency matrix A on the attached sheet, which also provides some relevant matlab computations.

a) Briefly interpret the R1, R2, R3, and R4 matrices from the attached sheet.

b) List the strong cliques (i.e., the 1-cliques) for the A relation.

c) List the 2-cliques for the A relation. How could you find these from the graph of the R2 relation?

d) List the 3-cliques for the A relation.

e) The con matrix on the attached sheet gives the (local) connectivity between each pair of individuals. Briefly discuss two implications that follow from con(4,8) = 2.

f) What is the (global) connectivity level k for the entire A graph? List the k-cutset(s).

g) Given the connectivity level k from part (f), are there any subgraphs of A that are k+1 connected? If not, explain why. If so, list the (nodes in the) k+1-component(s) and give one k+1-cutset for each k+1-component.

matlab computations for question 6

```
A  =

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

>> R1 = (eye(8) + A) > 0

R1 =

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

>> R2 = (eye(8) + A)^2 > 0

R2 =

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

>> R3 = (eye(8) + A)^3 > 0

R3 =

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

>> R4 = (eye(8)+A)^4 > 0

R4 =

     1     1     1     1     1     1     1     1
     1     1     1     1     1     1     1     1
     1     1     1     1     1     1     1     1
     1     1     1     1     1     1     1     1
     1     1     1     1     1     1     1     1
     1     1     1     1     1     1     1     1
     1     1     1     1     1     1     1     1
     1     1     1     1     1     1     1     1

>> con = [ ]; for i = 1:8; for j = 1:8; …
con(i,j) = connectivity(A,i,j); end; end; con

con =

   Inf   Inf   Inf     2     1     1     1     1
   Inf   Inf   Inf   Inf     1     1     1     1
   Inf   Inf   Inf   Inf     1     1     1     1
     2   Inf   Inf   Inf   Inf   Inf     2     2
     1     1     1   Inf   Inf   Inf   Inf   Inf
     1     1     1   Inf   Inf   Inf   Inf   Inf
     1     1     1     2   Inf   Inf   Inf   Inf
     1     1     1     2   Inf   Inf   Inf   Inf
```

Sociology 375 Exam 1 Fall 2007 Solutions

1) [30 pts] Applying some elementary rules of logic, $(xRy \wedge yRx) \rightarrow (x=y)$

is equivalent to its contrapositive $\neg(x=y) \rightarrow \neg(xRy \wedge yRx)$

which is equivalent to $(x=y) \vee \neg(xRy \wedge yRx)$

which is equivalent to $(x=y) \vee \neg(xRy) \vee \neg(yRx)$

which is *not* equivalent to the second condition given in the problem. (Note that the first term of that condition is $\neg(x=y)$ rather than $(x=y)$.)

To derive this answer using a truth table (as the problem requires), you might first note that there are 3 propositions {xRy, yRx, x=y} which might be true or false. Thus, the truth table will have $2^3 = 8$ rows. Given the 3 "input" propositions and the 2 conditions (compound propositions), we obtain the table below. Following convention, I have written the truth value of subexpressions (e.g., $xRy \wedge yRx$) underneath the connective (e.g., $\wedge$). The main columns giving the truth value of the entire condition are denoted (*) and (**). (Note that the order of the parenthesis for the second condition is arbitrary because $(P \vee Q) \vee R \leftrightarrow P \vee (Q \vee R)$.)

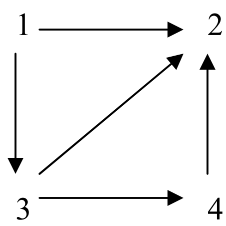
inputs			*antisymmetry condition*		*second condition*				
xRy	yRx	x=y	$(xRy \wedge yRx) \rightarrow (x=y)$		$(\neg(x=y) \vee \neg(xRy)) \vee \neg(yRx)$				
T	T	T	T	T	F	F	F	F	F
T	T	F	T	F	T	T	F	T	F
T	F	T	F	T	F	F	F	T	T
T	F	F	F	T	T	T	F	T	T
F	T	T	F	T	F	T	T	T	F
F	T	F	F	T	T	T	T	T	F
F	F	T	F	T	F	T	T	T	T
F	F	F	F	T	T	T	T	T	T
				(*)				(**)	

Because columns (*) and (**) are different, the two conditions are *not* equivalent.

2a) [8 pts] symmetry: $xRy \rightarrow yRx$ for all $x,y \in S$
asymmetry: $xRy \rightarrow \neg yRx$ for all $x,y \in S$

b) [12 pts] $R = \varnothing$ would generate a graph with no edges or an adjacency matrix with all 0's. Given $R = \varnothing$, the antecedent of both the symmetry and asymmetry conditions (xRy) is false for all $x,y \in S$. Thus, both the condition $xRy \rightarrow yRx$ and the condition $xRy \rightarrow \neg yRx$ is true for all $x,y \in S$. That is, the relation R is both symmetric and asymmetric.

3a) [15 pts] i)



ii) $R = \begin{bmatrix} 0 & 1 & 1 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 1 \\ 0 & 1 & 0 & 0 \end{bmatrix}$

iii) 1R2, 1R3, 3R2, 3R4, 4R2

b) [32 pts] i) $R^2 = \begin{bmatrix} 0 & 1 & 0 & 1 \\ 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}$ bii) $R^3 = \begin{bmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}$

biii) reachability(R) = $\begin{bmatrix} 1 & 1 & 1 & 1 \\ 0 & 1 & 0 & 0 \\ 0 & 1 & 1 & 1 \\ 0 & 1 & 0 & 1 \end{bmatrix}$ biv) distance(R) = $\begin{bmatrix} 0 & 1 & 1 & 2 \\ \infty & 0 & \infty & \infty \\ \infty & 1 & 0 & 1 \\ \infty & 1 & \infty & 0 \end{bmatrix}$

c) [18 pts] The matrix test for transitivity is $R - R^2\# \geq 0$.

$$\text{Here, } R - R^2\# = \begin{bmatrix} 0 & 1 & 1 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 1 \\ 0 & 1 & 0 & 0 \end{bmatrix} - \begin{bmatrix} 0 & 1 & 0 & 1 \\ 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix} = \begin{bmatrix} 0 & 0 & 1 & -1 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 1 & 0 & 0 \end{bmatrix}$$

and thus the relation R is *not* transitive. Note that 1R3 and 3R4 but not 1R4.

3d) [25 pts] The transitive closure of R is given by $T = (R + R^2 + R^3 + R^4)\#$

$$\text{Here, } T = \begin{bmatrix} 0 & 1 & 1 & 1 \\ 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 1 \\ 0 & 1 & 0 & 0 \end{bmatrix}.$$

T is not an equivalence relation because it is not reflexive and not symmetric. T is antireflexive, asymmetric, and transitive, and is thus a strict partial order. The Hasse relation is given by the transitivity test matrix

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

and thus the Hasse diagram is given by

(Because arrows must point downwards in Hasse diagrams, it is conventional to simply draw undirected edges.)

4) [15 pts] If A is symmetric, the reachability is an equivalence relation and will generate equivalence classes (called "weak cliques"). Thus, each unique row of R characterizes one of the weak cliques. Thus, the number of rows of U gives the number of weak cliques, and the rows sums give the number of individuals in each weak clique. Recognizing that equivalence classes partition the set of individuals, each individual belongs to exactly one weak clique. Thus, every column of U contains exactly one 1 (and the matlab command sum(U) would produce a row vector of 1's).

5) [25 pts] In his 1999 AJS paper, Watts showed how "small-world" graphs (which have a high clustering coefficient and low average distance) can be obtained by randomly "rewiring" some of the edges of a highly structured "caveman" graph (which is highly clustered with high average distance). Intuitively, a little random rewiring creates "shortcuts" across the caveman graph, greatly decreasing average distance but only slightly decreasing the clustering coefficient.

In the 2002 Science paper, Watts et al studied the conditions under which networks are "searchable." In the classic small-world experiment conducted by Travers and Milgram, some "starter" individuals were able to forward a package to a "target" individual through a chain of acquaintances, each of whom followed the simple rule ("greedy heuristic") of forwarding the package to an acquaintance similar to the target (e.g., in the same city or holding the same occupation). To determine the types of networks that can be "searched" in this fashion, Watts et al first posit a "cognitive hierarchy" (which allows them to measure "social distance" between individuals based on their characteristics), and then introduce a homophily parameter governing the extent to which individuals tend to form links to others with low social distance. Numerical experiments reveal that chains (between a randomly chosen starter and target) are most likely to be completed when homophily is moderately high (so that low social distance is associated with low network distance) but too high (because complete homophily would eliminate all paths between starters and targets with high social distance).

6a) [5 pts] Each Rt matrix gives reachability at step t

b) [12 pts] {1, 2, 3}, {2, 3, 4}, {4, 5, 6}, {5, 6, 7, 8}

c) [14 pts] {1, 2, 3, 4}, {2, 3, 4, 5, 6}, {4, 5, 6, 7, 8}. The 2-cliques of the A relation are the strong cliques (i.e., 1-cliques) of the R2 relation.

d) [8 pts] {1, 2, 3, 4, 5, 6}, {2, 3, 4, 5, 6, 7, 8}

e) [12 pts] First, you would need to remove 2 nodes to disconnect 4 and 8. (In particular, you would need to remove the nodes in the cutset {5, 6}.) Second, 4 and 8 are connected by 2 node-independent paths. (For instance, the paths 4-6-8 and 4-5-8.)

f) [7 pts] The global connectivity is given by the minimum element of the con matrix. Thus, the global connectivity of the graph of A equals 1. The (unique) 1-cutset is {4}.

g) [12 pts] There are two different 2-components: the 2-component {1, 2, 3, 4} has 2-cutset {2, 3}; the other 2-component {4, 5, 6, 7, 8} has 2-cutset {5, 6}.

Sociology 375 Final Exam Fall 2007 Prof Montgomery

Answer all questions. 250 points possible.

1) [60 points] Consider a social relation A on a set of 5 individuals. The corresponding adjacency matrix (along with its eigenvectors and eigenvalues) is given by

```
A =

     0     1     0     0     0
     1     0     1     0     0
     0     1     0     1     0
     0     0     1     0     1
     0     0     0     1     0
```

```
>> [eigvec, eigval] = eig(A)

eigvec =

    0.2887   -0.5000    0.5774    0.5000    0.2887
   -0.5000    0.5000    0.0000    0.5000    0.5000
    0.5774    0.0000   -0.5774    0.0000    0.5774
   -0.5000   -0.5000   -0.0000   -0.5000    0.5000
    0.2887    0.5000    0.5774   -0.5000    0.2887

eigval =

   -1.7321         0         0         0         0
         0   -1.0000         0         0         0
         0         0   -0.0000         0         0
         0         0         0    1.0000         0
         0         0         0         0    1.7321
```

a) Define structural equivalence. Given the A relation, are any (distinct) individuals structurally equivalent?

b) Give a non-trivial partition of the 5 individuals which satisfies the regular equivalence condition. Use the matrix test for regular equivalence to verify your solution. [NOTE: I'm looking for a *non-trivial* regular equivalence. You will receive zero points if you place each individual in his/her own subset, or place all individuals in the same subset.]

c) In practice, social network analysts often use equivalence "detectors" (i.e., algorithms like CONCOR or REGE) rather than the precise definitions of structural or regular equivalence when assigning individuals to "positions" in networks. Briefly explain why.

d) State the equation for the Bonacich centrality measure. Using the A matrix above, compute this measure given $\alpha = 1$ and $\beta = 0$. Given these parameter values, which individual(s) are the most "central"?

e) Given the A relation, what is the upper bound for β? Why can't β be set above this bound? What is the ranking of individuals as β approaches this upper bound? How do you know this?

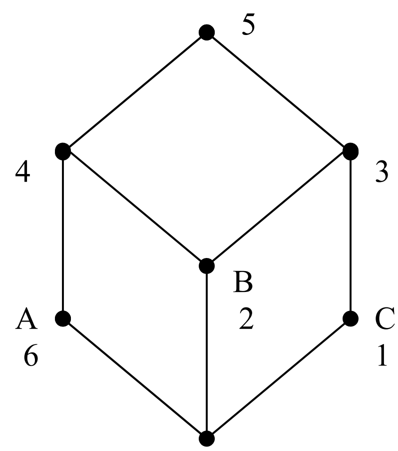
2) [60 points] Suppose that six actors (1, 2, 3, 4, 5, 6) attended three events (A, B, C), with the attendance pattern given by the matrix

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

a) In his paper on duality, Breiger (1974) shows how an actor-by-event matrix might be used to derive an actor-by-actor matrix and an event-by-event matrix. Give the formula for each matrix. Then, using the P matrix above, compute the actor-by-actor matrix and interpret the elements of this matrix.

b) We might say that actor i "contains" actor j if the set of events attended by i contains the set of events attended by j. Using the P matrix above, derive the adjacency matrix for the strict containment relation on the set of actors, and then draw the Hasse diagram. [HINT: There are simple matrix procedures for deriving the containment matrix and then the Hasse matrix. But to save time, you can simply complete each step by inspection (without doing all the matrix algebra) if you prefer.]

c) Below is a Galois lattice (oriented so that the empty set of events is at the bottom). Explain why this lattice could *not* have been generated by the P matrix above.



d) Construct the Galois lattice that *is* generated by the P matrix above. Use reduced labeling for both actors and events.

3) [40 points] Using the P matrix from question 2, we may compute the correlation coefficient matrix X (which results from comparing each pair of columns in P). The X matrix is given below, along with its eigenvectors and eigenvalues.

```
>> X = corrcoef(P)

X =

    1.0000         0         0
         0    1.0000   -0.5000
         0   -0.5000    1.0000
```

```
>> [eigvec, eigval] = eig(X)

eigvec =

         0    1.0000         0
   -0.7071         0   -0.7071
   -0.7071         0    0.7071

eigval =

    0.5000         0         0
         0    1.0000         0
         0         0    1.5000
```

a) Find the (3×1) column vector v such that the outer product vv^T gives the best least-squares approximation of the X matrix. [HINT: Recall that matlab scales eigenvectors so that their inner product is equal to 1. Thus, you may need to do some rescaling.]

b) Using the v vector from part (a), compute the "estimated" matrix $E = vv^T$, and then compute the "residual" matrix $R = X - E$. Use the X matrix to compute the total sum of squares (TSS), and then use the R matrix to compute the error sum of squares (ESS). The difference between TSS and ESS is the regression sum of squares (RSS). How is the RSS related to an eigenvalue? [HINT: If you're correct, you should be able to easily verify your answer from the matlab computations supplied above.]

4) [35 points] Consider a positive relation P and a negative relation N on some set of individuals S. Following the usual assumptions in balance theory, assume that P and N are symmetric (i.e., $P = P^T$, $N = N^T$) and disjoint (i.e., $P \cap N = \varnothing$) and that null ties are possible (i.e., neither iPj nor iNj holds for some pairs $(i,j) \in S \times S$). Using the P and N relations, we can generate compound relations such PP, PN, NP, NN, PPP, PPN, etc. [Formally, $PPN = \{(i,j) \in S\times S \mid iPh \wedge hPk \wedge kNj \text{ for some } h,k \in S\}$. Other compound relations are defined analogously.]

a) Suppose you construct the PPN and NPN relations and find that $PPN \cap NPN \neq \varnothing$. Is (the signed graph corresponding to) this pair of relations *balanced* or *not balanced* or is *either possible*? Briefly explain, drawing a graph to illustrate your answer.

b) Which of the following conditions must hold if the (signed graph corresponding to the) pair (P, N) is balanced? Which of the following conditions must hold if the pair (P, N) is clusterable? Briefly explain.

i) $PP \cap P = \varnothing$
ii) $PN \cap P = \varnothing$
iii) $NP \cap N = \varnothing$
iv) $NPNN \cap P = \varnothing$
v) $PPN \cap PPP = \varnothing$
vi) $PNNPN \cap N = \varnothing$

5) [55 points] The attached sheet gives adjacency matrices for a positive relation P and a negative relation N on a set of 20 individuals. Following the previous question, these relations are symmetric (in matlab notation, P = P′ and N = N′) and disjoint (in matlab notation, the matrix P&N would be a 20×20 matrix of 0s). However, in contrast to the previous question, I now assume that the signed graph of (P, N) is "complete" in the sense that every edge between distinct nodes is either positive or negative. I further assume that both P and N are antireflexive. Thus, the matrix P+N is a 20×20 matrix with 0s on the main diagonal and 1s everywhere else (as shown on the attached sheet).

a) Given 20 individuals, what is the total number of dyads in a dyad census? What is the total number of triads in a triad census?

b) Given the maintained assumptions on symmetry and completeness, there are 2 types of dyads and 4 types of triads in the signed graph of (P, N). Use the computations below to conduct a dyad census and then a triad census. [HINT #1: Do *not* attempt to use the P and N matrices directly. They are provided merely to help you conceptualize the problem. The dyad and triad censuses can be conducted using only the computations provided below.] [HINT #2: If you've correctly computed these censuses, the total number of dyads and triads should match your answers to part (a).]

c) Based on the dyad census, compute the *expected* number of triads of each type following the logic of Davis and Leinhardt. [HINT: To simplify your calculations, you can follow Davis and Leinhardt's assumption that sampling occurs *with* replacement.]

d) According to balance theory, which types of triads are prohibited? Comparing the actual numbers of triads to the expected numbers of triads, does this data set seem to strongly support balance theory? Briefly discuss.

matlab computations for problem 5

```
>> trace(P*P)

ans =   272

>> trace(N*N)

ans =   108

>> trace(P*P*P)

ans =        2496

>> trace(P*P*N)

ans =        1000

>> trace(P*N*N)

ans =   400

>> trace(N*N*N)

ans =   144
```

```
>> P
P =
    0    1    1    1    1    1    0    1    1    1    0    1    1    1    1    1    1    1    0    1
    1    0    1    0    1    1    1    1    1    0    0    1    1    1    1    0    0    1    0    0
    1    1    0    1    1    1    1    1    1    0    0    0    1    1    0    1    0    1    1    1
    1    0    1    0    1    1    1    0    1    0    1    0    1    1    0    0    0    1    1    1
    1    1    1    1    0    1    0    1    0    1    1    1    0    1    1    1    1    1    0    1
    1    1    1    1    1    0    0    1    1    1    0    0    1    1    1    0    0    1    1    1
    0    1    1    1    0    0    0    1    0    1    1    0    0    0    1    1    1    0    0    0
    1    1    1    0    1    1    1    0    1    0    0    1    1    1    1    1    1    1    1    1
    1    1    1    1    0    1    0    1    0    1    1    1    1    1    1    1    0    1    1    0
    1    0    0    0    1    1    1    0    1    0    1    1    1    1    1    1    0    1    1    1
    0    0    0    1    1    0    1    0    1    1    0    1    0    1    1    1    0    1    1    1
    1    1    0    0    1    0    0    1    1    1    1    0    1    1    1    0    1    1    1    1
    1    1    1    1    0    1    0    1    1    1    0    1    0    0    1    0    1    1    1    1
    1    1    1    1    1    1    0    1    1    1    1    1    0    0    1    0    0    0    1    1
    1    1    0    0    1    1    1    1    1    1    1    1    1    1    0    1    1    1    0    1
    1    0    1    0    1    0    1    1    1    1    1    0    0    0    1    0    1    1    1    0
    1    0    0    0    1    0    1    1    0    0    0    1    1    0    1    1    0    0    1    1
    1    1    1    1    1    1    0    1    1    1    1    1    1    0    1    1    0    0    1    0
    0    0    1    1    0    1    0    1    1    1    1    1    1    1    0    1    1    1    0    1
    1    0    1    1    1    1    0    1    0    1    1    1    1    1    1    0    1    0    1    0

>> N
N =
    0    0    0    0    0    0    1    0    0    0    1    0    0    0    0    0    0    0    1    0
    0    0    0    1    0    0    0    0    0    1    1    0    0    0    0    1    1    0    1    1
    0    0    0    0    0    0    0    0    0    1    1    1    0    0    1    0    1    0    0    0
    0    1    0    0    0    0    0    1    0    1    0    1    0    0    1    1    1    0    0    0
    0    0    0    0    0    0    1    0    1    0    0    0    1    0    0    0    0    0    1    0
    0    0    0    0    0    0    1    0    0    0    1    1    0    0    0    1    1    0    0    0
    1    0    0    0    1    1    0    0    1    0    0    1    1    1    0    0    0    1    1    1
    0    0    0    1    0    0    0    0    0    1    1    0    0    0    0    0    0    0    0    0
    0    0    0    0    1    0    1    0    0    0    0    0    0    0    0    0    1    0    0    1
    0    1    1    1    0    0    0    1    0    0    0    0    0    0    0    0    1    0    0    0
    1    1    1    0    0    1    0    1    0    0    0    0    1    0    0    0    1    0    0    0
    0    0    1    1    0    1    1    0    0    0    0    0    0    0    0    1    0    0    0    0
    0    0    0    0    1    0    1    0    0    0    1    0    0    1    0    1    0    0    0    0
    0    0    0    0    0    0    1    0    0    0    0    0    1    0    0    1    1    1    0    0
    0    0    1    1    0    0    0    0    0    0    0    0    0    0    0    0    0    0    1    0
    0    1    0    1    0    1    0    0    0    0    0    1    1    1    0    0    0    0    0    1
    0    1    1    1    0    1    0    0    1    1    1    0    0    1    0    0    0    1    0    0
    0    0    0    0    0    0    1    0    0    0    0    0    0    1    0    0    1    0    0    1
    1    1    0    0    1    0    1    0    0    0    0    0    0    0    1    0    0    0    0    0
    0    1    0    0    0    0    1    0    1    0    0    0    0    0    0    1    0    1    0    0

>> P+N
ans =
    0    1    1    1    1    1    1    1    1    1    1    1    1    1    1    1    1    1    1    1
    1    0    1    1    1    1    1    1    1    1    1    1    1    1    1    1    1    1    1    1
    1    1    0    1    1    1    1    1    1    1    1    1    1    1    1    1    1    1    1    1
    1    1    1    0    1    1    1    1    1    1    1    1    1    1    1    1    1    1    1    1
    1    1    1    1    0    1    1    1    1    1    1    1    1    1    1    1    1    1    1    1
    1    1    1    1    1    0    1    1    1    1    1    1    1    1    1    1    1    1    1    1
    1    1    1    1    1    1    0    1    1    1    1    1    1    1    1    1    1    1    1    1
    1    1    1    1    1    1    1    0    1    1    1    1    1    1    1    1    1    1    1    1
    1    1    1    1    1    1    1    1    0    1    1    1    1    1    1    1    1    1    1    1
    1    1    1    1    1    1    1    1    1    0    1    1    1    1    1    1    1    1    1    1
    1    1    1    1    1    1    1    1    1    1    0    1    1    1    1    1    1    1    1    1
    1    1    1    1    1    1    1    1    1    1    1    0    1    1    1    1    1    1    1    1
    1    1    1    1    1    1    1    1    1    1    1    1    0    1    1    1    1    1    1    1
    1    1    1    1    1    1    1    1    1    1    1    1    1    0    1    1    1    1    1    1
    1    1    1    1    1    1    1    1    1    1    1    1    1    1    0    1    1    1    1    1
    1    1    1    1    1    1    1    1    1    1    1    1    1    1    1    0    1    1    1    1
    1    1    1    1    1    1    1    1    1    1    1    1    1    1    1    1    0    1    1    1
    1    1    1    1    1    1    1    1    1    1    1    1    1    1    1    1    1    0    1    1
    1    1    1    1    1    1    1    1    1    1    1    1    1    1    1    1    1    1    0    1
    1    1    1    1    1    1    1    1    1    1    1    1    1    1    1    1    1    1    1    0
```

Sociology 375 Final Exam Fall 2007 Solutions

1a) [10 pts] Formally, actor i is structurally equivalent to actor j when

$$(iAk \leftrightarrow jAk) \wedge (kAi \leftrightarrow kAj) \text{ for all actors } k$$

Equivalently, i and j are structurally equivalent when rows i and j of the A matrix are the same, and columns i and j of the A matrix are the same. For the A matrix given in the problem, no distinct actors are structurally equivalent.

b) [20 pts] One possible answer is {{1, 3, 5}, {2, 4}}. Another is {{1, 5}, {3}, {2, 4}}. Using the first answer, this partition corresponds to the equivalence relation given by the matrix

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

To verify that E is a regular equivalence relation, we need to show that (AE)# = (EA)#. For this problem,

$$(AE)\# = (EA)\# = \begin{bmatrix} 0 & 1 & 0 & 1 & 0 \\ 1 & 0 & 1 & 0 & 1 \\ 0 & 1 & 0 & 1 & 0 \\ 1 & 0 & 1 & 0 & 1 \\ 0 & 1 & 0 & 1 & 0 \end{bmatrix}$$

c) [5 pts] For most real-world data sets, few pairs of actors are precisely structurally equivalent. Similarly, regular equivalence classes are often trivial (either all actors in the same class, or each actor in his/her own class). Thus, in practice, sociologists often assign actors to the same position when they are "approximately" equivalent, using algorithms like CONCOR or REGE.

d) [15 pts] The formula may be written $c = \alpha [A + \beta A^2 + \beta^2 A^3 + \ldots] 1$, where 1 is understood as a column vector of 1s. Given $\alpha = 1$ and $\beta = 0$, the centrality vector is simply given by the row sums of the A matrix. For the A matrix in this problem,

$$c(1,0) = \begin{bmatrix} 1 \\ 2 \\ 2 \\ 2 \\ 1 \end{bmatrix}$$

which implies that actors 2, 3, and 4 are most "central."

e) [10 pts] The upper bound for β is given by the reciprocal of the largest eigenvalue of the A matrix. For this problem, this upper bound is $1/1.7321 = 0.577$. If β were set above this value, the infinite sum $[A + \beta A^2 + \beta^2 A^3 + \ldots]$ would not converge. As β approaches its upper bound, the centrality vector is proportional to the leading eigenvector (given by the final column of the eigvec matrix). Thus, actor 3 is the most central (no longer tied with actors 2 and 4).

2a) [20 pts] The actor-by-actor matrix is PP^T; the event-by-event matrix is P^TP. For the present problem,

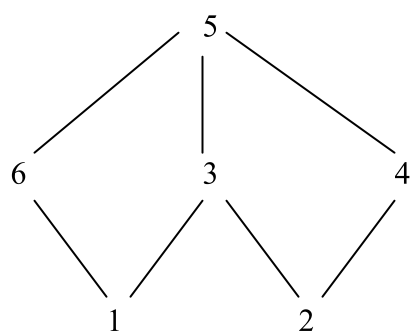
$$PP^T = \begin{bmatrix} 1 & 0 & 1 & 0 & 1 & 1 \\ 0 & 1 & 1 & 1 & 1 & 0 \\ 1 & 1 & 2 & 1 & 2 & 1 \\ 0 & 1 & 1 & 2 & 2 & 1 \\ 1 & 1 & 2 & 2 & 3 & 2 \\ 1 & 0 & 1 & 1 & 2 & 2 \end{bmatrix}$$

Element $(PP^T)(i,j)$ gives the number of events attended jointly by actors i and j.

b) [20 pts] The (strict) containment relation is

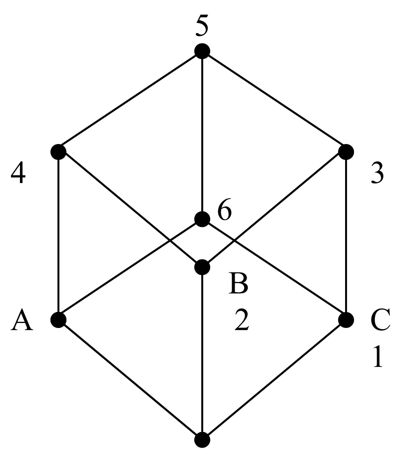
$$C = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 \\ 1 & 1 & 1 & 1 & 0 & 1 \\ 1 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}$$

which generates the Hasse diagram



c) [5 pts] This Galois lattice implies that actor 6 attends only event A (rather than events A and C).

d) [15 pts] The correct Galois lattice is



3a) [10 pts] The vector v is given by the leading eigenvector, multiplied by the square root of the corresponding eigenvalue. Thus,

$$v = \sqrt{1.5}\begin{bmatrix} 0 \\ -0.7071 \\ 0.7071 \end{bmatrix} = \begin{bmatrix} 0 \\ -0.866 \\ 0.866 \end{bmatrix}$$

b) [30 pts] The estimated matrix is thus given by $E = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0.75 & -0.75 \\ 0 & -0.75 & 0.75 \end{bmatrix}$

and the residual matrix is given by $R = X - E = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 0.25 & 0.25 \\ 0 & 0.25 & 0.25 \end{bmatrix}$.

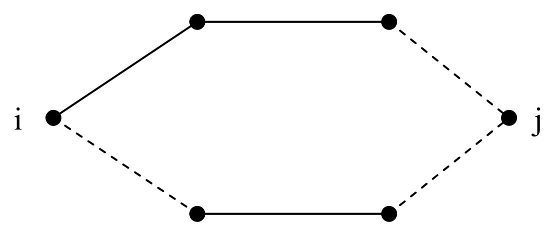
Squaring each element of the X matrix and summing over all elements, we obtain

$$TSS = 3*1^2 + 2*(-0.5)^2 = 3.5$$

Similarly, using the R matrix, we obtain $ESS = 1^2 + 4*(0.25)^2 = 1.25$

Thus, $RSS = 3.5 - 1.25 = 2.25$, which equals the square of the largest eigenvalue, $(1.5)^2$

4a) [15 pts] The signed graph corresponding to this pair of relations is *not balanced*. There must be some pair (i, j) such that iPPNj and iNPNj. Graphically,



This violates balance because there is both a positive and a negative path from i to j. Equivalently, there is a negative cycle in the graph.

b) [20 pts] If the graph is balanced, then conditions ii, iv, and v must hold. (Note that each of these conditions involves an odd number of Ns. Consequently, as we already saw in part (a), if any of these intersections was *not* empty, there must be both a positive and a negative path between some pair of nodes.)

If the graph is clusterable, then conditions ii and v must hold. (Note that each of these conditions involves exactly one N. Consequently, if either of these intersections was not empty, the graph must contain a cycle with exactly one negative edge.)

5a) [10 pts] There are 20*19/2 = 190 dyads and 20*19*18/6 = 1140 triads.

b) [20 pts] Each dyad is either positive or negative. (Recall that there are no null dyads.)

> number of positive dyads = trace(P*P)/2 = 136
> number of negative dyads = trace(N*N)/2 = 54
> (note that 136 + 54 = 190 as required)

The 4 types of triads are those originally identified by Heider: triads can have 3, 2, 1, or 0 positive edges (and hence 0, 1, 2, or 3 negative edges respectively).

> number of triads with 3 positive edges = trace(P*P*P)/6 = 416
> number of triads with 2 positive edges = trace(P*P*N)/2 = 500
> number of triads with 1 positive edge = trace(P*N*N)/2 = 200
> number of triads with 0 positive edges = trace(N*N*N)/6 = 24
> (note that 416 + 500 + 200 + 24 = 1140 as required)

c) [15 pts] Given the dyad census, we may construct triads "randomly" following the logic of Davis and Leinhardt (1972). As a first step, we determine the probability that each dyad is either positive or negative:

> p = probability of positive dyad = 136/190 = .716
> n = probability of negative dyad = 54/190 = .284

Using these probabilities, the expected probability of each type of triad is given by

> probability of triad with 3 positive edges = p^3 = .367
> probability of triad with 2 positive edges = $3p^2n$ = .437
> probability of triad with 1 positive edge = $3pn^2$ = .173
> probability of triad with 0 positive edges = n^3 = .023

Multiplying each of these probabilities by the number of triads, we obtain

> expected number of triads with 3 positive edges =.367*1140 = 418.4
> expected number of triads with 2 positive edges = .437*1140 = 498.2
> expected number of triads with 1 positive edge = .173*1140 = 197.2
> expected number of triads with 0 positive edges = .023*1140 = 26.2

d) [10 pts] In balance theory, the triads with 2 and 0 positive edges are prohibited. For this data set, the expected and actual numbers of triads are very close. (Given that I originally generated the P and N matrices randomly, this result was anticipated.) If balance theory was actually operating, the actual number of triads would presumably be much lower than the expected number of triads for the prohibited types of triads.