Sociology 375 Exam 1 Fall 2011 Prof Montgomery

Answer all questions. 250 points possible. Explanations can be brief.

1) Consider a set of students S = {Al, Beth, Carl, David, Ellen}. You learn that Al likes Beth, Al likes Carl, Beth likes Carl, Carl likes Beth, David likes Beth, Ellen likes Al, and Ellen likes Carl.

a) [20 points] Show how the relation R ('likes') on the set S could be represented

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

b) [25 points] Does the relation R on S satisfy each of the following conditions? If not, you should identify one violation. [HINT: You need to report only one violation.]

(i) reflexivity
(ii) antireflexivity
(iii) symmetry
(iv) antisymmetry
(v) transitivity

2) Consider a relation R on a set of four individuals S = {1, 2, 3, 4}. This relation is characterized by the following adjacency matrix:

$$A = \begin{matrix} 0 & 1 & 0 & 0 \\ 1 & 0 & 1 & 1 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 \end{matrix}$$

a) [40 points] Given the adjacency matrix A, compute the following (in matrix form):

(i) number of 2-paths
(ii) number of 3-paths
(iii) reachability
(iv) distance

b) [15 points] Is the reachability relation (from question 2.a.iii) of the following type? Briefly note why or why not.

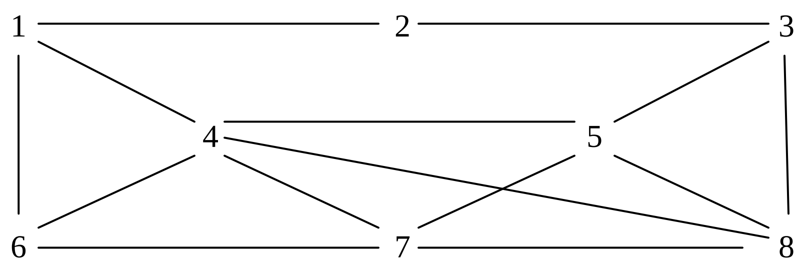
(i) equivalence relation (ii) quasiorder (iii) partial order

c) [10 points] State the condition for structural equivalence. Given the relation above (characterized by the adjacency matrix A), are there any individuals who are structurally equivalent to each other? If so, identify those individuals. If not, briefly explain why.

d) [20 points] State the condition for regular equivalence. Given the relation above (characterized by the adjacency matrix A), can you partition S into equivalence classes (characterized by an adjacency matrix E) such that individuals 3 and 4 are regularly equivalent to each other? If so, verify that E satisfies the regular equivalence condition. Otherwise, briefly explain why individuals 3 and 4 cannot be regularly equivalent.

3) [25 points] Briefly describe the experiment reported in Travers and Milgram (*Sociometry* 1969). What were their findings? From these findings, can you immediately know the distance (shortest path) between individuals? Can you immediately know the proportion of individuals who can reach each other (i.e., are connected by a path of any length)? Or is it more difficult to infer the answers to these questions from their experimental results? Briefly discuss.

4) Consider the social network below. This network is also characterized by the adjacency matrix A on the attached sheets, which contain a variety of Matlab computations that may (or may not) be useful for answering the following questions.



a) [25 points] Compute the clustering coefficient (C_i) for each node $i \in \{1, \ldots, 8\}$, and the clustering coefficient (C) for the graph overall. Conceptually, what do we learn from the clustering coefficient?

b) [30 points] List all of the cliques (i.e., 1-cliques) with at least 3 nodes. Then list all of the 2-cliques. Which of the 2-cliques are also 2-clans? What criterion did the 2-clans meet? Which of the 2-cliques are also 2-clubs? What criterion did the 2-clubs meet?

c) [10 points] What is the (global) connectivity of the entire graph? How can you determine this from the attached Matlab computations? Give a cutset to verify the global connectivity level.

d) [20 points] What is the (local) connectivity between nodes 3 and 6? Give a maximal set of node-independent paths to verify this local connectivity level. Then give a cutset to verify this local connectivity level.

e) [10 points] Given the connectivity level k of the entire graph (from part d), are there any (k+1)-components? If so, list them. Are there any (k+2) components? If so, list them.

Matlab computations for question 4

```
>> A

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

>> A^2

ans =
     3     0     1     1     1     1     2     1
     0     2     0     1     1     1     0     1
     1     0     3     2     1     0     2     1
     1     1     2     5     2     2     3     2
     1     1     1     2     4     2     2     3
     1     1     0     2     2     3     1     2
     2     0     2     3     2     1     4     2
     1     1     1     2     3     2     2     4

>> A^3

ans =
     2     4     2     8     5     6     4     5
     4     0     4     3     2     1     4     2
     2     4     2     5     8     5     4     8
     8     3     5    10    12     9    11    12
     5     2     8    12     8     5    11     9
     6     1     5     9     5     4     9     5
     4     4     4    11    11     9     8    11
     5     2     8    12     9     5    11     8

>> sum(A)

ans =

     3     2     3     5     4     3     4     4

>> sum(A) .* (sum(A)-1)

ans =

     6     2     6    20    12     6    12    12

>> sum(A).^2

ans =

     9     4     9    25    16     9    16    16
```

```
>> distance(A)

ans =

     0     1     2     1     2     1     2     2
     1     0     1     2     2     2     3     2
     2     1     0     2     1     3     2     1
     1     2     2     0     1     1     1     1
     2     2     1     1     0     2     1     1
     1     2     3     1     2     0     1     2
     2     3     2     1     1     1     0     1
     2     2     1     1     1     2     1     0

>> distance(A) <= 2

ans =

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

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

con =

   Inf   Inf     3   Inf     3   Inf     3     3
   Inf   Inf   Inf     2     2     2     2     2
     3   Inf   Inf     3   Inf     3     3   Inf
   Inf     2     3   Inf   Inf   Inf   Inf   Inf
     3     2   Inf   Inf   Inf     3   Inf   Inf
   Inf     2     3   Inf     3   Inf   Inf     3
     3     2     3   Inf   Inf   Inf   Inf   Inf
     3     2   Inf   Inf   Inf     3   Inf   Inf

>> con >= 3

ans =

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

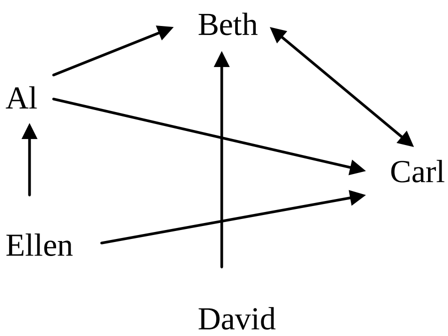
Soc 375 Exam 1 Fall 2011 Solutions

1) [45 points overall]

1a) [20 points]

R = {(Al, Beth), (Al, Carl), (Beth, Carl), (Carl, Beth), (David, Beth), (Ellen, Al), (Ellen, Carl)}

Al R Beth; Al R Carl; Beth R Carl; Carl R Beth; David R Beth; Ellen R Al; Ellen R Carl



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

b) [25 points]

(i) not reflexive: not Al R Al
(ii) antireflexive: not xRx for all $x \in S$
(iii) not symmetric: Al R Beth but not Beth R Al
(iv) not antisymmetric: Beth R Carl and Carl R Beth
(v) not transitive: Ellen R Al and Al R Beth but not Ellen R Beth

2) [85 points overall]

2a) [40 points]

(i) number of 2-paths given by $A^2 = \begin{bmatrix} 1 & 0 & 1 & 1 \\ 0 & 1 & 1 & 1 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}$

(ii) number of 3-paths given by $A^3 = \begin{bmatrix} 0 & 1 & 1 & 1 \\ 1 & 0 & 2 & 2 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 \end{bmatrix}$

(iii) reachability = $(A + A^2 + A^3)\# = \begin{bmatrix} 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 \\ 0 & 0 & 1 & 1 \\ 0 & 0 & 1 & 1 \end{bmatrix}$

(iv) distance = $\begin{bmatrix} 0 & 1 & 2 & 2 \\ 1 & 0 & 1 & 1 \\ \infty & \infty & 0 & 1 \\ \infty & \infty & 1 & 0 \end{bmatrix}$

b) [15 points] The reachability relation derived in 4.a.iii is reflexive and transitive. It is not an equivalence relation – it is reflexive and transitive but not symmetric. It is a quasiorder – it is reflexive and transitive. It is not a partial order – it is reflexive and transitive but not antisymmetric.

c) [10 points] Given a relation R on S, x and y are structurally equivalent if they send ties to exactly the same alters and receive ties from exactly the same alters. Equivalently, using the adjacency matrix A, x and y are structurally equivalent if row x is the same as row y, and column x is the same as column y. In the current case, there are no distinct individuals who are structurally equivalent to each other.

d) [20 points] Given a relation R (represented an adjacency matrix A) and an equivalence relation (represented as a matrix E), the individuals within each equivalence class are regularly equivalent to each other if

$$(AE)\# = (EA)\# \ .$$

Here, we can partition S into the subsets {1}, {2}, and {3,4} to form the matrix

$$E = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 1 \\ 0 & 0 & 1 & 1 \end{bmatrix} \quad \text{and then verify that} \quad (AE)\# = (EA)\# = \begin{bmatrix} 0 & 1 & 0 & 0 \\ 1 & 0 & 1 & 1 \\ 0 & 0 & 1 & 1 \\ 0 & 0 & 1 & 1 \end{bmatrix}$$

3) [25 points] The Travers and Milgram paper describes a famous experiment that introduced the “small-world” phenomenon. Travers and Milgram selected a “target” person (a Boston stockbroker) and “starting” persons drawn from various groups (Nebraska residents, Nebraska stockbrokers, Boston residents). Each starting person was given a letter intended for the target person, and was asked to forward the letter to someone who might know the target person on a first-name basis. Subsequent individuals in the chain were given the same instructions. Travers and Milgram observed the path followed by each letter, and whether the letter ultimately reached the target.

In their paper, Travers and Milgram report a variety of results. In class, I focused on two results: the probability of a completed chain (29% overall) and the average number of intermediaries in completed chains (5.2). As might be expected, chain completion and length varied somewhat for different starting individuals – chains originating from Nebraska were somewhat less likely to be completed and (conditional on completion) were somewhat longer.

It would be naïve to directly interpret these experimental results as reachability and distance between the starting and target persons. Presumably, many chains ended not because there was no possible path, but because some intermediary lacked the motivation to continue the chain. From theoretical studies of the small-world phenomenon (e.g., Watts AJS 1999), we should expect that most everyone belongs to the same "giant" component and hence can reach most everyone else. Further, completed chains are not necessarily the shortest possible paths, but rather the paths that happened to be found by the intermediaries (none of whom has full knowledge of the network). The experimental results merely give us an upper bound on distance (and further assumptions would be necessary to infer actual distance).

4) [95 points overall]

4a) [25 points] The clustering coefficient C_i for each node i is equal to the number of ties among i's immediate neighbors divided by the number of ties that could be present among i's immediate neighbors. Using the Matlab computations provided,

$$C_i = \frac{\text{number of ties among i's neighbors}}{\text{number of possible ties among i's neighbors}} = \frac{\text{(A^3)(i,i) / 2}}{\text{(sum(A) .* sum(A)-1)(i) / 2}}$$

Note that the 2's in the numerator and denominator cancel each other. Thus,

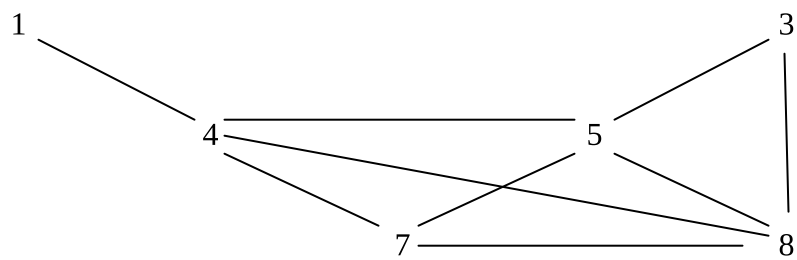
$C_1 = 2/6 = 1/3$	$C_4 = 10/20 = 1/2$	$C_7 = 8/12 = 2/3$
$C_2 = 0/2 = 0$	$C_5 = 8/12 = 2/3$	$C_8 = 8/12 = 2/3$
$C_3 = 2/6 = 1/3$	$C_6 = 4/6 = 2/3$	

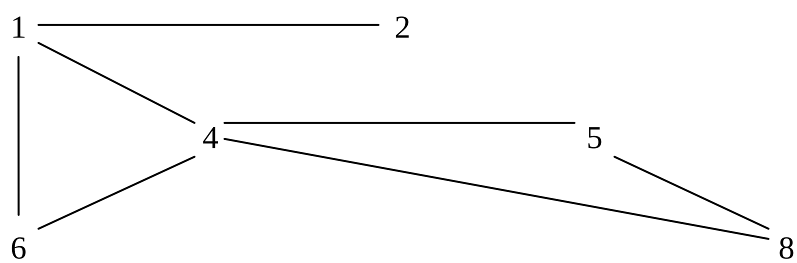
The clustering coefficient C for the graph is simply the mean of the clustering coefficients for the nodes. Thus, C = (23/6)/8 = 23/48 = .479. Conceptually, the clustering coefficient is one measure of the degree of transitivity (the level of "clustering") in the graph.

4b) [30 points] By inspection, the cliques are {1,4,6}, {4,6,7}, {4,5,7,8}, and {3,5,8}.

Using the distance matrix from the Matlab computations, we see that distance(2,7) = distance(3,6) = 3. Thus, 2 and 7 can't belong to the same 2-clique, and 3 and 6 can't belong to the same 2-clique. So we would need to remove the pair {2,3} or {2,6} or {7,3} or {7,6} from the set of individuals {1,…, 8} to form a 2-clique. Thus, the 2-cliques are {1,4,5,6,7,8}, {1,3,4,5,7,8}, {1,2,4,5,6,8}, and {1,2,3,4,5,8}.

The 2-cliques {1,2,3,4,5,8} and {1,4,5,6,7,8} are 2-clans. You need to check that the distance between each pair of nodes in the 2-clique is less than or equal to 2, using only paths within the 2-clique subgraph. As shown below, {1,3,4,5,7,8} is not a 2-clan because (once you remove nodes 2 and 6 from the original graph) the distance between nodes 1 and 3 is equal to 3; {1,2,4,5,6,8} is not a 2-clan because (once you remove nodes 3 and 7 from the original graph) the distance between nodes 2 and 5 is equal to 3.





The 2-cliques {1,2,3,4,5,8} and {1,4,5,6,7,8} are 2-clubs. You need to check that the 2-clique is a maximal subgraph with diameter less than or equal to 2 (i.e., you can't add any additional nodes to the subgraph without increasing the diameter of the subgraph).

4c) [10 points] The global connectivity level is 2. That's the minimum local connectivity level from the con matrix. The cutset is {1,3}.

4d) [20 points] Nodes 3 and 6 have a local connectivity level of 3. That implies you should be able to find a set of 3 node-independent paths between these nodes. For instance: (3,2,1,6), (3,5,4,6), and (3,8,7,6). That also implies you would need to remove 3 nodes from the graph to disconnect 3 and 6. One possible cutset is {2,5,8}.

4e) [10 points] The clique {4,5,7,8} is a 3-component. (Recall that any clique with n members is an (n-1)-component. Adding any additional node to this set, the connectivity level would fall below 3.) There are no 4-components.

Sociology 375　　　Exam 2　　　Fall 2011　　　Prof Montgomery

Answer all questions. 250 points possible. Explanations can be brief.

1) [40 points] Consider a sports league with 5 teams. Over the course of the season, each team played each other team 3 times. Last season's outcomes are given by the "beats" matrix below where B(i,j) is the number of times team i beat team j. Note that I've also computed the eigenvectors and eigenvalues of B, which may (or may not) be useful for answering the questions below.

```
B =

     0     2     2     3     0
     1     0     0     2     1
     1     3     0     1     1
     0     1     2     0     0
     3     2     2     3     0

>> [eigenvectors,eigenvalues] = eig(B)

eigenvectors =

   0.4310          -0.3474 - 0.0907i  -0.3474 + 0.0907i  -0.2207 + 0.4102i  -0.2207 - 0.4102i
   0.3089          -0.0456 + 0.4319i  -0.0456 - 0.4319i  -0.3061 + 0.0023i  -0.3061 - 0.0023i
   0.4441           0.5555             0.5555             0.3245 + 0.1739i   0.3245 - 0.1739i
   0.2337          -0.1185 - 0.4677i  -0.1185 + 0.4677i  -0.0452 - 0.3381i  -0.0452 + 0.3381i
   0.6834          -0.1805 + 0.3295i  -0.1805 - 0.3295i   0.6614             0.6614

eigenvalues =

   5.1216               0                  0                  0                  0
        0         -1.4102 + 1.9204i        0                  0                  0
        0               0            -1.4102 - 1.9204i        0                  0
        0               0                  0            -1.1506 + 0.8601i        0
        0               0                  0                  0            -1.1506 - 0.8601i
```

a) Compute the Bonacich centrality measure when $\alpha = 1$ and $\beta = 0$ (i.e., compute c(1,0)). According to this measure, how would you rank the teams?

b) What is the substantive interpretation of the β parameter in the Bonacich measure? In general, what formula determines the maximum possible value of β? Compute the maximum possible (numerical) value of β given the B matrix above. Using the Bonacich centrality vector with maximum β, how would you rank the teams? [HINT: You don't need to normalize the centrality vector to answer this part.]

c) Using Bonacich's suggested normalization, compute α for the centrality vector in part (c). Using this value of α, compute the normalized centrality vector. [HINT: Matlab normalizes eigenvectors so that their inner products are equal to 1.]

2) [100 points] Suppose that 4 actors {1, 2, 3, 4} attended 5 events {A, B, C, D, E} according to the participation matrix

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

where P(i,j) = 1 indicates that actor i participated in event j.

a) In his paper 1974 *Social Forces* paper on the duality of persons and groups, Breiger shows how the participation matrices (characterizing a relation from actors to events) can be used to construct an actors × actors matrix and also an events × events matrix. Give the equations for these matrices in generic form (i.e., for any P matrix). Then compute these matrices using the P matrix above. Give the interpretation for the elements of these matrices.

b) Using the P matrix above, construct a Galois lattice to show the containment relations among actors and events. You should use reduced labeling, and orient the lattice to that the set of all events (if you were using full labeling) corresponds to the top node.

c) Given the Galois lattice from part (b), it is apparent there is a rank-3 HICLAS approximation with no discrepancies (i.e., the "approximation" is the same as the data matrix). Report the (4×3) row-bundle matrix and the (5×3) column-bundle matrix for this approximation. Conceptually, what does the existence of a perfect rank-3 approximation suggest?

d) Suppose there was a 5th actor (not included in the participation matrix above) who participated in events A and E (but not events B, C, and D). Given the HICLAS approximation from part (c), is it possible to assign a row bundle to actor 5 without introducing any discrepancies between the data matrix and estimated matrix? If so, give the row bundle for actor 5. If not, give the row bundle for actor 5 that would minimize the number of discrepancies, and report the number of discrepancies introduced. [HINT: It may be helpful to use the lattice from part (b) to determine your answer.]

e) Suppose you had access to Matlab and wanted to perform correspondence analysis using the P matrix above. Describe (as precisely as possible) the steps you would take. What would be the final output? What would be its interpretation?

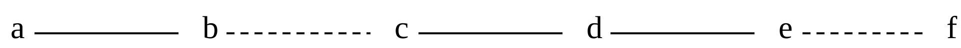
3) [50 points]

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

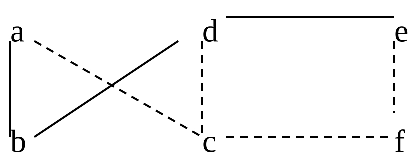
b) Similarly, there are two (logically equivalent) conditions that determine whether a signed graph is clusterable. State these two conditions. Is clusterability a stronger or weaker condition than balance? Briefly explain.

c) Determine whether each of the following graphs is balanced and/or clusterable. (A solid edge denotes a positive tie; a dotted edge denotes a negative tie.) If so, redraw the graph to demonstrate the result. If either condition does not hold, explain why.

(i)

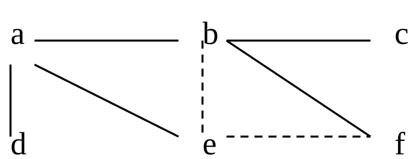


(ii)



(iii)

(iv)



4) [60 points] Consider a society with 4 clans, where the clan of a man's wife and children are given by

```
>> W

W =
     0     1     0     0
     1     0     0     0
     0     0     0     1
     0     0     1     0

>> C

C =
     0     1     0     0
     0     0     1     0
     0     0     0     1
     1     0     0     0
```

The Matlab computations on the next page may be useful for answering the questions below.

a) Find the group G (of permutation matrices) that is generated by the W and C matrices above. [HINT: G needs to be closed under multiplication.] How many elements are in G? Report the multiplication table. [HINT: Use the Matlab computations on the next page. To save time, you don't need to report the full table. You can just give the two columns corresponding to the generator elements W and C.]

b) Given the answer to part (a), is this society "generalized-balanced" (in the sense implied by Boyd's "group partition theorem")? If so, demonstrate why. If not, show failures of "evaluative consistency" and "self-consistency" in this society.

c) In class, we studied whether marriages could occur between various types of first cousins. But it is also possible to determine whether marriages could occur between more distant cousins. In each of the following cases, is it ever possible for ego to marry the type of individual described? If not, show why. If it could be possible, compute whether such a marriage is allowed (regardless of the man's clan) given the W and C matrices above.

i) mother's mother's sister's daughter's daughter
ii) mother's father's brother's daughter's daughter
iii) father's mother's sister's daughter's daughter
iv) mother's mother's sister's son's daughter

[HINT: You should simplify each compound relation as much as possible before using the multiplication table from part (a).]

matlab computations for question 4

```
>> W*W
ans =
    1    0    0    0
    0    1    0    0
    0    0    1    0
    0    0    0    1

>> W*C
ans =
    0    0    1    0
    0    1    0    0
    1    0    0    0
    0    0    0    1

>> C*W
ans =
    1    0    0    0
    0    0    0    1
    0    0    1    0
    0    1    0    0

>> C*C
ans =
    0    0    1    0
    0    0    0    1
    1    0    0    0
    0    1    0    0

>> W*W*W
ans =
    0    1    0    0
    1    0    0    0
    0    0    0    1
    0    0    1    0

>> W*W*C
ans =
    0    1    0    0
    0    0    1    0
    0    0    0    1
    1    0    0    0

>> W*C*W
ans =
    0    0    0    1
    1    0    0    0
    0    1    0    0
    0    0    1    0

>> W*C*C
ans =
    0    0    0    1
    0    0    1    0
    0    1    0    0
    1    0    0    0

>> C*W*W
ans =
    0    1    0    0
    0    0    1    0
    0    0    0    1
    1    0    0    0

>> C*W*C
ans =
    0    1    0    0
    1    0    0    0
    0    0    0    1
    0    0    1    0

>> C*C*W
ans =
    0    0    0    1
    0    0    1    0
    0    1    0    0
    1    0    0    0

>> C*C*C
ans =
    0    0    0    1
    1    0    0    0
    0    1    0    0
    0    0    1    0

>> W*C*W*W
ans =
    0    0    1    0
    0    1    0    0
    1    0    0    0
    0    0    0    1

>> W*C*W*C
ans =
    1    0    0    0
    0    1    0    0
    0    0    1    0
    0    0    0    1

>> W*C*C*W
ans =
    0    0    1    0
    0    0    0    1
    1    0    0    0
    0    1    0    0

>> W*C*C*C
ans =
    1    0    0    0
    0    0    0    1
    0    0    1    0
    0    1    0    0
```

Soc 375 Exam 2 Fall 2011 Solutions

1a) [10 pts] The formula for the Bonacich centrality measure can be written

$$c = \alpha (B + \beta B^2 + \beta^2 B^3 + \ldots) \mathbf{1}$$

where **1** is a column vector of 1's. Thus, given $\alpha = 1$ and $\beta = 0$, the Bonacich measure simplifies to $c = B \mathbf{1}$ which simply reports row sums of the B matrix (i.e., the total number of wins for each team). For this example,

$$c = \begin{bmatrix} 7 \\ 4 \\ 6 \\ 3 \\ 10 \end{bmatrix}$$

Ranking teams based on the number of wins, we obtain

team 5 > team 1 > team 3 > team 2 > team 4

1b) [20 pts] The parameter β reflects the amount of "weight" placed on indirect victories (i.e., being able to "reach" other terms through a chain of who beat whom). As we saw in part (a), the measure places no weight on indirect victories when $\beta = 0$. The maximum value of β is the reciprocal of the largest eigenvalue of the B matrix. For the B matrix given, this is 1 / 5.1216 = .1953. As β approaches this maximum value, the centrality vector converges to the leading eigenvector of B (i.e., the first column of the eigenvectors matrix reported on the exam) which generates the ranking

team 5 > team 3 > team 1 > team 2 > team 4

1c) [10 pts] Bonacich normalizes the centrality vector so that α^2 c'c = n where n is the number of actors. Equivalently, α = sqrt(n / c'c). Because Matlab normalizes eigenvalues to that the inner product c'c = 1, we obtain α = sqrt(n). Multiplying each element of the leading eigenvector by sqrt(5), we obtain

$$c = (2.236) \begin{bmatrix} .4310 \\ .3089 \\ .4441 \\ .2337 \\ .6834 \end{bmatrix} = \begin{bmatrix} .9637 \\ .6907 \\ .9930 \\ .5226 \\ 1.5281 \end{bmatrix}$$

Note that this normalization doesn't affect the rankings from part (b), but does scale the centrality scores so that high scores are greater than 1, and low scores are less than 1.

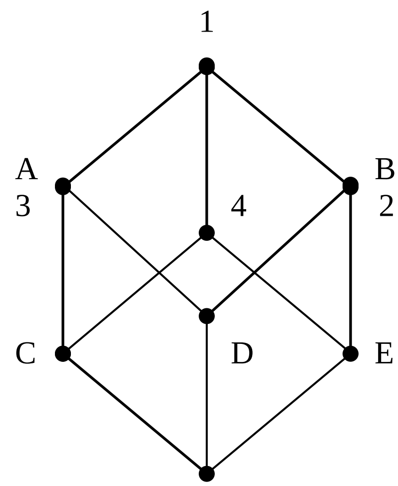
2a) [20 pts] PP′ generates an actor × actor matrix. P′P generates an event × event matrix. For the present example,

```
>> P*P'          % actor × actor matrix
ans =
     5     3     3     2
     3     3     1     1
     3     1     3     1
     2     1     1     2
```

```
>> P'*P          % event × event matrix
ans =
     2     1     2     2     1
     1     2     1     2     2
     2     1     3     2     2
     2     2     2     3     2
     1     2     2     2     3
```

PP′(i,j) gives the number of events attended jointly by actors i and j; P′P(i,j) gives the number of actors who attended both events i and j.

2b) [30 pts]



c) [15 pts] Note that the lattice in part (b) is the full rank-3 lattice. Thus, we can simply assign factor bundle 111 to the top-most node (actor 1), factor bundle 110 to node with event A and actor 3, factor bundle 101 to the node with actor 4, and so on. The row-bundle and column-bundle matrices are given below. I've also double-checked that these matrices produce an estimated matrix identical to the data matrix.

```
>> S    % row bundles
S =
     1     1     1
     0     1     1
     1     1     0
     1     0     1
```

```
>> P  % colmn bundles
P =
     1     1     0
     0     1     1
     1     0     0
     0     1     0
     0     0     1
```

```
>> ~(~S*P')  % estimate
ans =
     1     1     1     1     1
     0     1     0     1     1
     1     0     1     1     0
     0     0     1     0     1
```

The existence of a perfect rank-3 approximation suggests that there are 3 underlying unobserved "factors" possessed by actors and required by events which explain the observed participation matrix.

2d) [15 pts] Given the lattice in part (b), you could try placing actor 5 at each of the 8 nodes, and count the number of discrepancies generated. Given the orientation of the lattice, the estimate would assume that actor 5 participated in event j if and only if he can reach event j through a downward path. Note that discrepancies can be false positives (the estimate implies actor 5 participated in an event when he actually did not) or false negatives (the estimate implies actor 5 did not participate in an event when he actually did). We obtain the best estimate by assigning factor bundle 001 to actor 5 (i.e., placing actor 5 at the same node as event E) which generates 1 discrepancy (i.e., we estimated that actor 5 attended only event E when he actually attended events A and E).

You would have obtained this answer by inspection. But I've given the computations below to show how you can use Matlab to obtain this answer.

```
>> L  % all possible factor bundles

L =
    1     1     1
    1     1     0
    1     0     1
    0     1     1
    1     0     0
    0     1     0
    0     0     1
    0     0     0
```

```
>> for i = 1:8; m = ~(~L(i,:)*P'); d = sum(m ~= [1 0 0 0 1]); disp(['bundle ' num2str(L(i,:)) ' produced ' num2str(d) ' discrepancies']); end

bundle 1  1  1 produced 3 discrepancies
bundle 1  1  0 produced 3 discrepancies
bundle 1  0  1 produced 2 discrepancies
bundle 0  1  1 produced 3 discrepancies
bundle 1  0  0 produced 3 discrepancies
bundle 0  1  0 produced 3 discrepancies
bundle 0  0  1 produced 1 discrepancies
bundle 0  0  0 produced 2 discrepancies
```

2e) [20 pts] In brief, correspondence analysis obtains actor scores (u vector) from event scores (v vector) through the equation $u = R^{-1} P v$ (where R is a diagonal matrix of row sums of P) and obtains event scores from actor scores through the equation $\lambda v = C^{-1} P' u$ (where C is a diagonal matrix of column sums of P, and λ is a scaling factor). Combining these two equations, we obtain $\lambda v = [C^{-1} P' R^{-1} P] v$. Thus, the event scores are eigenvectors of the bracketed matrix. In particular, the eigenvectors corresponding to the second and third largest eigenvalues provide the coordinates for events in the scatterplot. We then use the equation $u = R^{-1} P v$ to obtain the coordinates for actors in the scatterplot.

Conceptually, correspondence analysis gives a visual representation of the two-mode participation matrix. In the scatterplot, events are positioned near similar events (i.e, events attended by many of the same actors), and each actor's position is a weighted average of the positions of the events in which they participate.

3a) [8 pts] A signed graph is balanced if (i) all cycles are positive or (equivalently) if (ii) the nodes can be partitioned into two subsets, with all positive edges within subsets and all negative edges between subsets.

b) [10 pts] A signed graph is clusterable if (i) there are no cycles with exactly one negative edge or (equivalently) if (ii) the nodes can be partitioned into (two or more) subsets, with all positive edges within subsets and all negative edges between subsets. Clusterability is a weaker condition than balance. Clusterability allows nodes to be partitioned into any number of subsets (not only 2). Clusterability only restricts cycles with exactly one negative edge (not all negative cycles).

c) [32 pts]

(i) Balanced (and hence clusterable). Can partition nodes into sets {a, b, f} and {c, d, e}.

(ii) Not balanced (given negative cycle defcd), but graph is clusterable. Can partition nodes into sets {a, b, d, e}, {c}, and {f}.

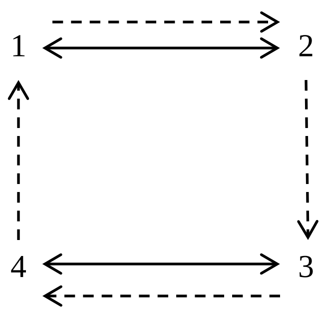
(iii) Balanced (and hence clusterable). Can partition nodes into sets {a, b, c, f, e} and {d, h, g}.

(iv) Not balanced and not clusterable (given cycle abea with exactly one negative edge).

4a) [21 pts] The group contains 8 elements which I've labeled {W, C, I, WC, CW, CC, WCW, WCC}. I've given the (first two columns of the) multiplication table below:

	W	C
W	I	WC
C	CW	CC
I	W	C
WC	WCW	WCC
CW	C	W
CC	WCC	WCW
WCW	WC	I
WCC	CC	CW

4b) [15 pts] The society is generalized-balanced when there's an isomorphism between the graph of W and C on the set of clans and the graph of the multiplication table of the group. For the present example, because there are 4 clans and the group has 8 elements, it's obvious that no such isomorphism exists and hence the society is not g-balanced. Using the graph of W and C on the set of clans (see below), it is easy to find failures of evaluative consistency. For example, you can get from clan 1 to clan 2 by both W and C edges, but $W \neq C$. You can also check that self-consistency fails because cycles do not have the sign of the identity element of the group. For example, starting from clan 1 and going clockwise, there's a cycle with sign $WCCC = CW \neq I$.



4c) [24 pts]

i) The clan of this type of cousin is given by the matrix $C^{-1}WC^{-1}WW^{-1}CW^{-1}C$, which reduces to I. The wife's clan is given by the matrix W. Because all kinship systems require $W \neq I$, the focal male cannot marry this type of cousin.

ii) $C^{-1}WC^{-1}CW^{-1}C$ reduces to I. Because $W \neq I$, marriage is not allowed.

iii) $C^{-1}C^{-1}WW^{-1}CW^{-1}C$ reduces to $C^{-1}W^{-1}C$. In principle, marriage might be allowed if $C^{-1}W^{-1}C = W$. However, in the present example, marriage is not permitted. The multiplication table shows $C^{-1} = WCW$. Thus, $C^{-1}W^{-1}C = WCWW^{-1}C = WCC$. Note that WCC and W are different elements of the group, hence $WCC \neq W$.

iv) $C^{-1}WC^{-1}WW^{-1}CC$ reduces to $C^{-1}WC$. In principle, marriage might be allowed if $C^{-1}WC = W$. However, in the present example, marriage is not permitted. The multiplication table shows $C^{-1} = WCW$ and $W^{-1} = W$. Thus, $C^{-1}WC = WCWW^{-1}C = WCC$. Again, WCC and W are different elements of the group, hence $WCC \neq W$.