ACTOR ALLEGIANCE AND BLOCKMODEL STRENGTH

John Rigsby and Dr. Jeff Solka
Contact-John Rigsby, RigsbyJT@NSWC.NAVY.MIL
17320 Dahlgren Road B10, Dahlgren, VA, 22448

Key Words: social network analysis, blockmodel, allegiance

Abstract:

This paper presents a new method to measure strength change in blockmodels as the partition size changes. This is done by measuring how much an actor supports his block as the number of partitions changes. This measure of an actor's block support is called allegiance.

1. Introduction

Blockmodeling is the social scientist way of grouping actors. A good measure of blockmodel strength does not exist; especially when you are trying to figure out how many partitions the blockmodel should have. We needed a way to establish a strength measure for block models dependent upon the number of partitions and arrangement. Out of this need, the concept of allegiance was born. Allegiance is a way to calculate changes in blockmodel strength at both the actor level and overall blockmodel level. This paper covers how to calculate allegiance, where the concept came from, and a way to estimate the proper number of partitions for a blockmodel.

2. Blockmodel

First we define a labeled, directed graph D with a set of vertexes (actors) called V and a set of edges (ties) called E. The actors can be subdivided into discrete, non-overlapping subsets called partitions where $P = \{P_1, ..., P_k\}$ with k partitions. Obviously, partitions are not unique, so one has to formulate a symbolic expression for an actor relative to the particular method one uses to accomplish a partition. A symbolic means to represent the actor-partition is to use the form $P(i, k)$, which refers to the partition that actor i is part of when the total number of partitions is k. This partitioning of actors determines up the blockmodel. The blockmodel acts as a grouping method for the analysis of social network data where $B = \{B_{1,1}, B_{1,2}, ..., B_{k,k-1}, B_{k,k}\}$ [4]. The block, $B_{x,y}$, is formed from the ties of actors in partition x, P_x, to the actors in partition y, P_y. The block $B_{x,y}$, when $x \neq y$, represents the ties from actors in partition x, P_x, to actors in vertical partition y, P_y. If x equals y then $B_{x,y}$ represents the internal ties of the actors within block.

Figure 1 illustrates a partition of a twelve actor graph into a blockmodel. The partitions are P_1, P_2, and P_3. The blockmodel is represented by the blocks $B_{1,1}$, $B_{1,2}$, ..., $B_{3,2}$, and $B_{3,3}$. The proper partition size, k, is 3. The possible values of k are between one and twelve. The actors i and j are in the same partition. The actor q is in a separate partition from i and j. The block, $B_{1,2}$, is formed by the overlapping of horizontal partition one, P_1, and vertical partition two, P_2. If I were mathematically referring to $P(i, 3)$, I would be referring to P_2 because P_2 is the partition occupied by actor i at $k = 3$. The relationship of actors i and j is part of block $B_{2,2}$ because they are both members of P_2, which mathematically is $P(i, 3)$ or $P(j, 3)$. The relationship of actor i to q is part of block $B_{2,3}$ because i is in horizontal overlapped P_2, and q is in vertical overlapped P_3. The relationship of actor q to i is part of block $B_{3,2}$ because q is in horizontal P_3, and i is in vertical P_2. The total number of actors is defined as N . $N_{P(i,3)}$ is the number of actors in the partition, $P(i, 3)$, which is four. The variables, i , j, and q, will be used for actors or indices of actors. We use the term G to denote groups of actors; the subscript of G will define the group. The group $G(i, k)$ is the group of actors that belong to the partition, $P(i, k)$.

3. The Road to Allegiance

We have found that there is a shortcoming in the formulation of the concept of the blockmodel. What is missing is a concept that we have termed allegiance. There are three shortcomings that the allegiance concept allow us to overcome. First, we wanted a way to measure blockmodel strength change when number of partitions can change. A formulation of a measure of blockmodel strength could help estimate the true number of partitions. Second, we needed a way to show the differences between time separated graphs and give a measure of how partition size and actor ties change. Third, we want to fuse all this information together and turn the time separated graphs into a single graph that portrays

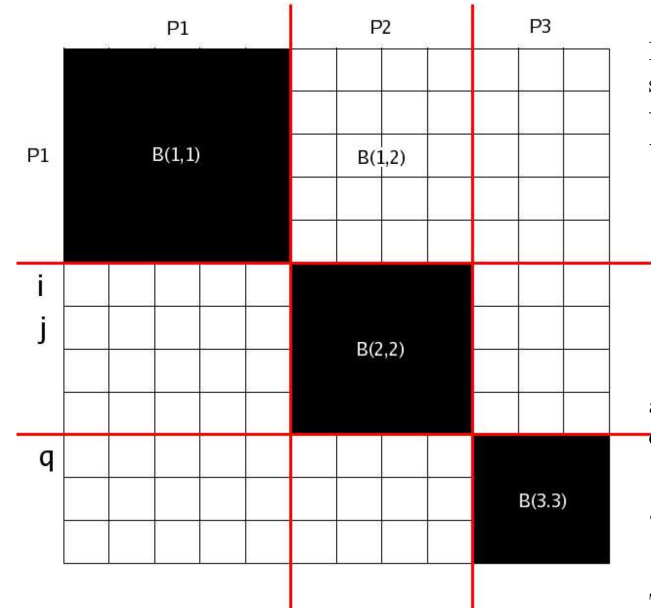

Figure 1: Illustration of Actors, Blocks, and Partitions

all the same information. Using allegiance to measure blockmodel strength change and to estimate the number of partitions will be the scope of this paper.

4. Allegiance

Allegiance measures the support that an actor provides for the structure of his block.

An actor supports his block by having internal block edges. A measure of this is the total number of internal block edges that an actor has; this will be called H_{int}, where $G(i,k)$ is the group of all actors belonging to the same partition, $P(i,k)$, as actor i at k partitions.

$$H_{int}(i) = \sum_{j \in G(i,k)} E(i,j) \quad (1)$$

where, $E(i,j)$ is the edge weight from actor i to actor j

An actor supports his block by not having external edges from the block. A measure of this is the total number of possible external edges minus the total number of existing external edges; this will be called H_{ext}. The total number of actors is defined as N. $N_{P(i,k)}$ is the total number of actors in the same partition, $P(i,k)$, as actor i at k partitions.

$$H_{ext}(i,k) = N - N_{P(i,k)} - \sum_{j \neq G(j,k)} E(i,j) \quad (2)$$

We define $A(i,k)$, allegiance, as the measure of how much an actor supports his block at a partition size of k. α works as a weight function. Allegiance works similarly to a betweenness measure for clusters.

$$A(i,k) = \alpha[H_{int}(i,k)] + (1-\alpha)[H_{ext}(i,k)] \quad (3)$$

where, $\alpha \in [0,1]$

For this paper α is one-half. Initially the data is all one partition, and $A(i,1)$ is simply the half the out degree of actor i.

5. Determination of the number of groups via allegiance

The allegiance of actors changes as the number of partitions change. If the overall allegiance, $O(A(k))$, is positive then a good partitioning was made. This process is iterated until additional partitioning no longer have positive effects.

$\Sigma_{A(k)}$ is the summation of allegiance for all actors at k partitions. The first cut divides the data into two partitions; the second cut divides the data into three partitions, and so on and so forth. Each new cut divides a former partition into two new partitions. $O(A(k))$, overall allegiance is the summation of total allegiance changes at the new partitioning, k, compared to $k-1$.

$$\Sigma_{A(k)} = \sum_{i=1}^{N} A(i,k) \quad (4)$$

$$O(A(k)) = \Sigma_{A(k)} - \Sigma_{A(k-1)}) \quad (5)$$

where, $(N) \geq k \geq 1$

$\Sigma_{A(1)}$ is the half of the sum total out degree for all the actors. The individual actor differences, $D_A(i,k)$, in allegiance at each partitioning shows how each actor is affected. The summation of actor allegiance differences, $\Sigma_{D_A}(k)$, shows the strength change in the block model at partitioning k. When $\Sigma_{D_A}(k)$ is negative, the block structure strength is decreased by this partitioning; The first negative value of $\Sigma_{D_A}(k)$ yields the maximum number of partitions. If this first negative value of $\Sigma_{D_A}(k)$ is significantly negative then the maximum partition size becomes one less.

$$D_A(i,k) = A(i,k) - A(i,k-1) \quad (6)$$

$$\Sigma_{D_A}(k) = \sum_{i=1}^{N} D_A(i,k) \quad (7)$$

Overall allegiance, $O(A(k))$, equals the summation of actor allegiance differences, $\Sigma_{D_A}(k)$. For clarity I will only use $\Sigma_{D_A}(k)$ throughout the rest of the paper.

6. Toy Example

This is a graph where the correct partition size is three, $k = 3$. We are using equivalence clustering to order the partition cuts; this is calculated using the equiv.clust function from the SNA [6] R [1] package. SNA stands for Social Network Analysis. Then we measure allegiances of affected actors as k, the number of partitions, increases.

Figure 2 shows that at $k = 4$ the data is a perfect blockmodel.

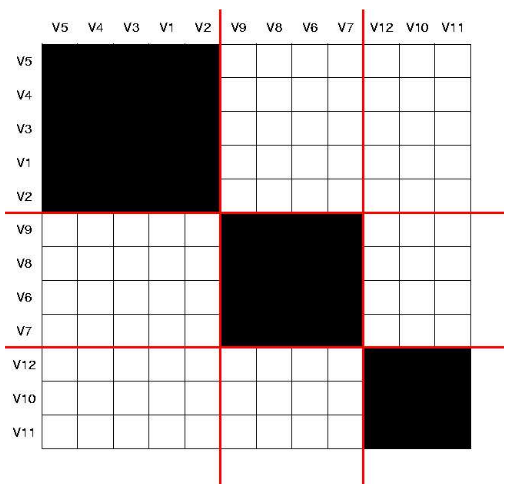

Figure 2: Perfect blockmodel of 12 actors into 3 partitions, $k = 3$ is the correct answer

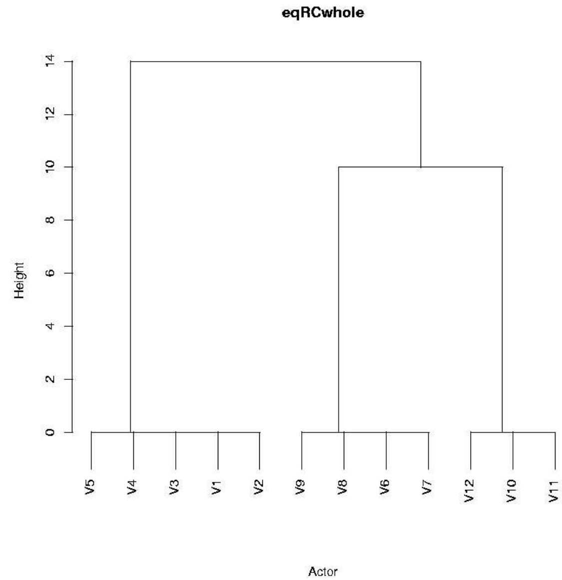

Figure 3: Equivalence Cluster Dendrogram

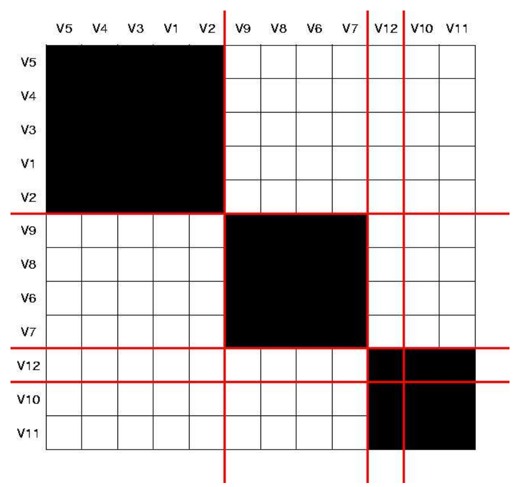

Figure 4: Blockmodel at $k = 4$

Figure 3 is a plot of the equivalence cluster tree. The equivalence cluster shows that the partition size is three. The height for all partition cuts with a number of partitions beyond three is zero. The partition cut ordering and equivalence cluster tree is calculated using the equiv.clust function from the SNA [6] package for R [1].

At four partitions and beyond, actors are placed in groups by themselves, therefore changes in allegiance between partition cuts is always negative. Figure 4 shows the shaving of actors into singleton groups, but an actual measure of block strength change would be more useful.

For each partitioning for the value of k, the allegiance of the actors is calculated and compared to the previous partitioning. Figure 5 is a plot of $D_A(i,k)$. At $k = 1$, $D_A(i,1) = A(i,1)$. $D_A(i,1)$ and $A(i,1)$ equal half the out degree of actor i. $D_A(i,k)$ shows the allegiance change for individual actors between partitionings.

The darkest squares represent positive allegiance change, and the maximum occurs at $k = 2$ for actors V1, V2, V3, V4, and V5 and obtains a value of 7. The brightest squares represent negative changes in allegiance. The most negative value, -4, occurs for actor 1 at $k = 9$. An allegiance change of 0 occurs in the majority of the plot; it is the dull gray. The allegiance of an actor does not change between cuts if he is not in the partition being cut.

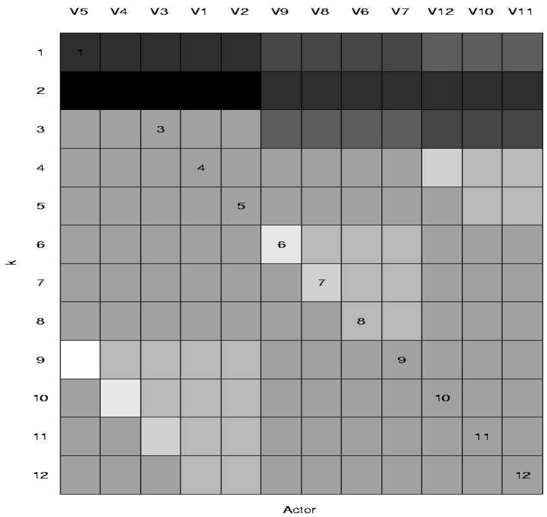

Figure 5: Plot of $D_A(actor, k)$ encoded in gray-scale, black = 7 and white = 4

The brighter squares at $k = 4$ represents the first negative cut. The change in allegiance for actors V10, V11, and V12 is negative at $k = 4$ and zero for the rest of the actors.

The sum of allegiance change at each partition value allows us to ascertain the correct number of groups. Figure 6 is a plot of $\Sigma_{D_A}(k)$. The x-axis is nodes(actors), and the y-axis is k, the number of partitions.

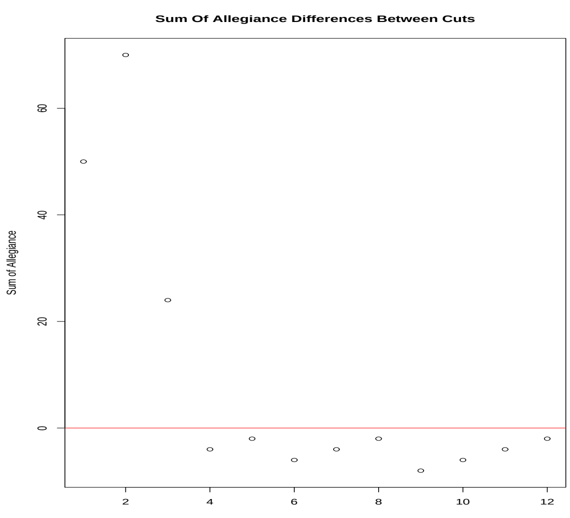

Figure 6: Sum of individual allegiance changes as a function of k, the number of partitions

As the number of partitions increases past three, the change in the allegiance becomes negative. Actors provide less and less block support as the number of partitions increases past three. Using allegiance as a measure yields the correct answer of a partition size of three, which is $k = 3$.

7. Application of Allegiance to the Analysis of Network Data

On a day to day basis, the data we are most interested in is computer network data for the purpose of network security. The data used in this analysis comes from a small office network that is part of our local area network [3]. The sensor collecting the data is a SHADOW [5] network sensor. SHADOW, Secondary Heuristic Analysis for Defensive Online Warfare, is a network intrusion detection system maintained and developed at the Naval Surface Warfare Center. This sensor employs standard TCPDUMP [3, 7] software in order to capture all the header packets that are generated by the machines in this local enclave. The ability to capture all of the header packets implies that we have full knowledge of all network transactions that are being completed by the machines. The typical data sets that are analyzed within a social network setting usually only have incomplete information regarding the interchange of information among the actors that are participating in the network [4]. The perfect knowledge that we have in our case makes our task a little easier. Figure 7 represents the equivalence cluster dendrogram for this network over a fourteen month period. A clear partition size is not self evident. The graph used in this section comes from previous work we have done. The graph is an actor relative co-membership graph. The edges represent relative actor co-membership weights [2] .

As the partition size increases allegiance goes up and down based on the grouping. Partitionings that lead to an increase in allegiance cause more block support and are considered good partitionings. Partitionings that lead to a decrease in allegiance cause non-supportive structure changes. Figure 8 is a plot of the $\Sigma_{D_A}(k)$, the sum of the change in allegiance between partitionings. The first negative summation of allegiance occurs at a partition size of seven. The first significant partition cut with a negative summation of allegiance occurs at a partition size of eight.

There is a dip in allegiance change at $k = 4$. Actor J is cut from the block containing J, I.J, I.I, S, I.G, I.F, Q, I.C, I.D, I.E, I.B, and R. The allegiance of actors I.J, I.I, S, I.G, I.F, Q, I.C, I.D, I.E, I.B, and R decreases because their one-way connections to J become edges to a non-block member. Actor J's allegiance increases because its lack of edges to its

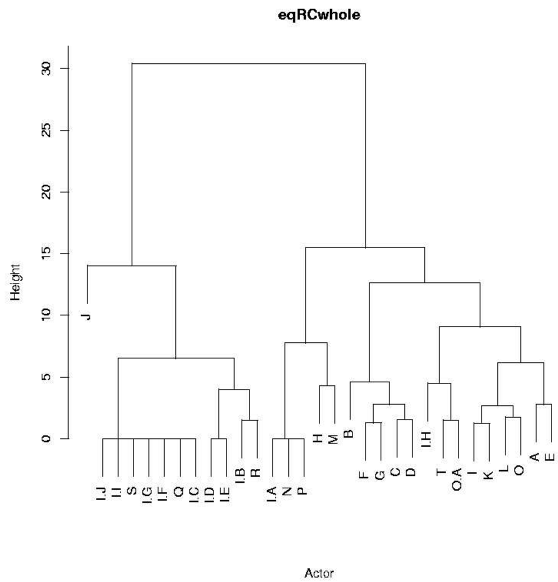

Figure 7: Equivalence Cluster Dendrogram for our local enclave data

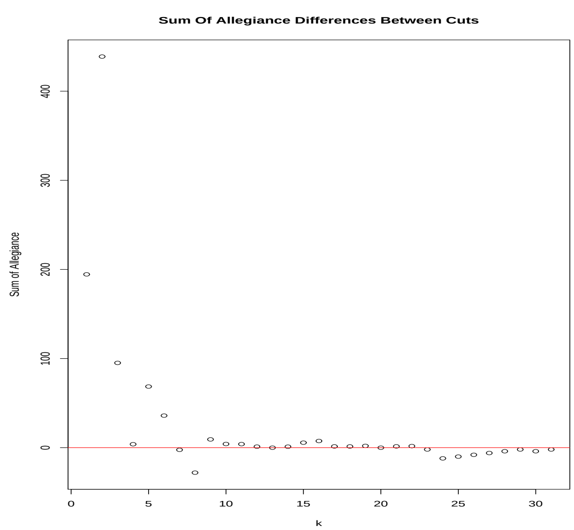

Figure 8: $\Sigma_{D_A}(k)$

old block members become positive assets as non-edges to non-block members. Actor J is a connector from block two, $B_{2,2}$ (I.J, I.I, S, I.G, I.F, Q, I.C, I.D, I.E, I.B, and R) to block four (B, F, G, C, D, I, K, L, O, A, and E), $B_{4,4}$. Figure 9 represents the blockmodel at $k = 4$.

A partition size of six, $k = 6$, holds the last positive value for the sum of allegiance change before the first significant negative sum of allegiance change at a partition size of eight, $k = 8$. This creates a

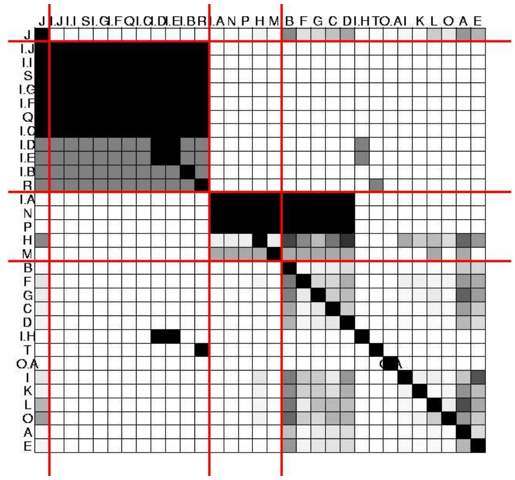

Figure 9: Blockmodel at $k = 4$

predicament because using allegiance as a measure for estimating partition size yields that the partition size is six or seven. At $k = 7$, an overall negative change in allegiance of -2.7, about 0.3%, occurs. This cut does not hurt or help allegiance.

At a partition size of seven, actors H and M are cut from the block containing I.A, N, P, H, and M. The difference in allegiance between partitionings six and seven for actors I.A, N, and P decreases. The difference in allegiance for H and M is positive. Block four (H and M), $B_{4,4}$, is a connector for block three(I.A, N, and P), $B_{3,3}$, to block one (J), $B_{1,1}$, and block seven (I, K, L, O, A, and E), $B_{7,7}$. By measure of allegiance this new partitioning causes a negative change in allegiance of -2.7 . This partitioning is deemed to neither help nor hurt allegiance. This is where blockmodeling and allegiance become less science and more art. Cut seven helps explain the role of actors H and M in the network as a connector group, but it does not help allegiance. In short, allegiance leads to good estimates of the maximum partition size. Figure 10 represents the blockmodel at $k = 7$.

Figure 11 represents the individual changes in allegiance as the number of partitions increases. The columns are actors, and the rows are different partition sizes. The individual color shades represent the change in allegiance for each actor at each partition size. At $k = 1$, allegiance is simply half the out degree of an actor. Black is the highest value of 19, which occurs at a partition size of three. White is the lowest value of -6, which occurs at partition size of 24. Most of the graph is zeroes because only allegiance of actors in the block being cut changes at each new partitioning. A visual zero reference is

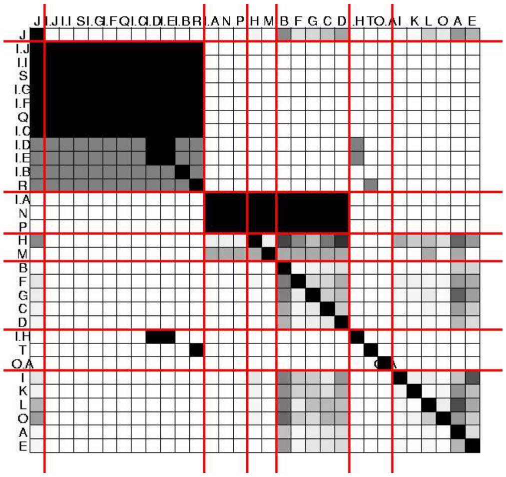

Figure 10: Blockmodel at $k = 7$

provided at row 20, $k = 20$, which is all zeroes. This new partitioning has absolutely no affect on allegiance; this is because actor O.A is isolate so he has allegiance only to himself. The first new partitioning that causes a significant negative change in allegiance occurs at $k = 8$. The change in allegiance for actors I.J, I.I, S, I.G, I.F, Q, I.C, I.D, I.E, I.B, and R is -4 at $k = 8$ yielding a total drop of -28; this is represented by the white bar at $k = 8$ from column position 2 to position 8. Figure 11 suggests a partition size of 7.

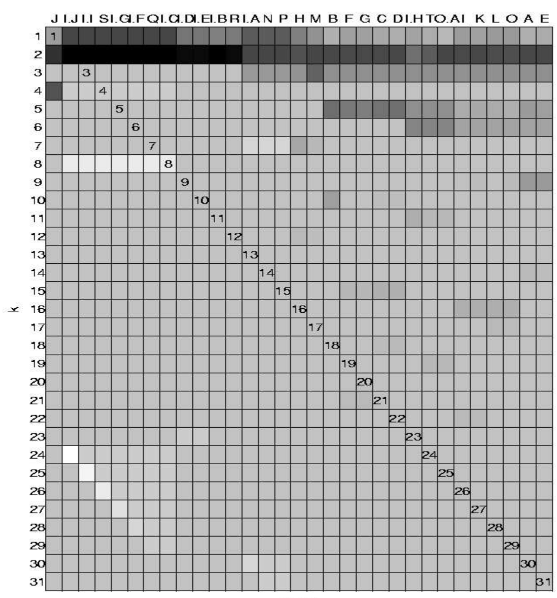

Figure 11: $D_A(actor, k)$ encoded in gray-scale where black = 19 and white = -6

8. Conclusions

We have introduced the concept allegiance. It provides us with a mechanism for evaluating the overall precedence of a proposed partitioning. Overall, changes in allegiance values between partition sizes can be used to estimate the correct partition size. It can also be used o estimate the maximum number of partitions. This the allegiance concept is a bridgework which allows us to determine how to partition blocks of actors so we cab infer important information about their behavior.

9. Thanks

We send many thanks to Dr. David Marchette of the Naval Surface Warfare Center and Johns Hopkins University for all of the hours of helpful discussions. We would also like to thank the Office of Naval Research In-house Laboratory Independent Research program for funding this work.

I would like thank the developers of the open source project, R. It is a great statistics tool. Finally, I would like to thank Dr. Carter Butts for writing the very useful SNA [6] package for R [1].

References

[1] Ihaka R. and Gentleman R. (1996). "R: A Language for Data Analysis and Graphics." *Journal of Computational and Graphical Statistics*.

[2] Rigsby J. and Solka J. (2003). "Computer Networks and Social Network Block Structures." *Proceedings of the Joint Statistics Meeting 2003*.

[3] Stevens, W. R. (1994). *The Protocols (TCP/IP Illustrated, Volume 1)* . Addison-Wesley

[4] Wasserman S. and Faust K. (1994). *Social Network Analysis: Methods And Applications*. Cambridge University Press, Cambridge, Uk

[5] SHADOW. Vers. 1.8. (April 2004). http://www.nswc.navy.mil/ISSEC/CID/index.html .

[6] SNA. Vers. 0.44. (May 2004). http://erzuli.ss.uci.edu/R.stuff/ .

[7] TCPDUMP. Vers. 3.8.3. (March 2004). http://www.tcpdump.org/ .