CHAPTER 9

PARTITIONING MULTIMODE NETWORKS

Martin G Everett[1], and Stephen P Borgatti[2]

[1] University of Manchester
[2] University of Kentucky

9.1 Introduction

Most networks examined so far involve connections between nodes all of the same type, known as one-mode networks. There are circumstances in which the nodes are of different types and the connections are only between different types of nodes, and not between nodes of the same type. We refer to these as multimode networks, some authors call these multiway networks. A simple example consists of nodes made up of authors and journals. An author is connected to a journal if they have published a paper in that journal. Since we have two types of nodes, authors and journals, this results in a 2-mode network. There are many examples of 2-mode networks such as people attending events [16], legislators being members of committees [38], directors serving on boards [13], companies collaborating on projects [39] etc. In principle there is no reason to limit the node types to two, we could have three or more. An example of a 3-mode dataset would be criminal by crime by victim. Such datasets are less common and we shall concentrate at first on 2-mode datasets and discuss general multi-mode approaches later in the chapter.

For clarity of exposition, when considering 2-mode data we shall refer to one mode as actors and the other mode as events. The resultant network will form a bipartite network, that is a network in which the nodes can be divided into two groups with edges only occurring between the groups and not within the groups (note that a one mode network can

Advances in Network Clustering and Blockmodeling,
March 4, 2018 00 : 24.

also be bipartite). If we have n actors and m events we can represent the data as an $n \times m$ affiliation matrix **A**, where $a(i, j) = 1$ if actor i attended event j, and 0 otherwise. If the data are valued, reflecting for example the time actor i was at event j, then we can replace the binary entry with the value. It is normal to ignore direction in 2-mode data since in most cases the direction is from one mode to the other; for example actors choose events and not the other way round. One can envisage examples in which this is not the case, for example heterosexual actors selecting members of the opposite sex indicating whom they would be interested in dating. However, there are few techniques or datasets of this type and so we will not discuss the issue further, but just mention it is probably worthy of additional research work.

There have been two distinct approaches to dealing with 2-mode data. The oldest method is to convert the data to one-mode and this is often referred to as projection. There are two possible projections for a 2-mode affiliation matrix one resulting in an actor by actor matrix the other in an event by event matrix. In these cases the relations are attended an event together and had an actor in common respectively. We can capture more information in our projections if we record the number of events each pair of actors attended and the number of actors each pair of events had in common. These are given by $\mathbf{AA}^T$ and $\mathbf{A}^T\mathbf{A}$ respectively where $\mathbf{A}^T$ represents the transpose of **A**. As Breiger [11] pointed out in his 1974 paper these should not be seen as two independent data matrices but as dual representations of the data. However, it became common practice to always dichotomise the data and in many applications only one projection was considered. Clearly reducing valued data to binary in any situation results in loss of information and this is compounded here by ignoring one of the projections. A consequence of this approach is that there is significant data loss resulting in an inferior analysis.

An alternative approach is to develop methods for analyzing the bipartite graph directly. The first systematic example of this approach was due to Borgatti [7] in 1989 and further developed by Borgatti and Everett [8]. They showed how to extend structural and regular equivalence to multimode data directly, this was later extended to generalized blockmodelling by Doreian et al [19]. In essence this was a simple matter of extending the known block structures for one mode data to block structures for multimode data. In practice these structures require very little modification and progress in the area of blockmodelling had until recently been entirely computational. In this chapter we shall apply both approaches to the same data set in order to gain some insight into how the techniques perform.

9.2 2-mode partitioning

At the heart of the blockmodeling approach described in the last paragraph is a need to optimize a cost function which captures the extent to which a given partition of the rows and columns of the data matrix corresponds to a blockmodel. The resulting combinatorial optimization problem is unlikely to have a polynomial time solution and hence more heuristic methods are required. There is no consensus on what methods will perform best but Rosmalen et al [42] examine five different techniques on simulated data in which they optimize using a Euclidean metric. Their test datasets were relatively small with a maximum value of n and m set at 120 and up to 7 clusters in the rows and columns. Their simulations indicated a 2-mode version of the k-means method (See the paper for details) had the best overall performance and they then validated this finding on some empirical data.

The study by Rosmalen et al. did not capitalize on the binary nature of any affiliation matrix (provided we have non-valued data). In [12] Brusco and Steinley propose an extension of variable neighbourhood search which did. Variable neighbourhood searches are meta-heuristic methods in which increasingly large neighbourhoods of the current best solution are explored. Overall the performance of their algorithm was very similar to the 2-mode k-means except in the situation that the block positions were known. In this case their algorithm was an improvement.

The methods discussed above are quite sophisticated and the articles contain pseudo-code which provide more details. Many of the applications apply fairly simple greedy algorithms that prioritize efficiency over the ability to avoid local minimum, but they can often find acceptable solutions by using many different starting positions. The techniques we have discussed so far are very general and can be applied to many different types of data. We now look at methods specifically designed for social network type data.

9.3 Community Detection

In a vain effort to bring some consistency to terminology we suggest that the term community detection is used for the partitioning of a network into groups such that actors within a group are more closely connected to each other than those in other groups, we shall refer to these groups as communities. If we allow actors to be in more than one group, so that groups overlap and do not insist that all actors are assigned to any group but still have highly connected groups, we shall call these groups cohesive subgroups. A consequence of these definitions is that community detection is a special case of blockmodeling. Blockmodeling does partition the actors but allows for more general forms of blocks which do not have to reflect closely connected sets of actors.

In considering 2-mode networks we shall consider the problem of partitioning both modes so that we find sets of actors and events. We will require the density of the submatrix containing these actors and events to be denser than the other submatrices containing either the actors or the events. It should be noted that some authors consider the event communities and the actor communities separately (see [24] for example). For single mode networks the most commonly used and accepted technique (although it has some well-known short-comings [22]) is Newman's community detection which optimizes modularity [36]. Barber [2] extended modularity to 2-mode data and developed an algorithm specifically for this type of data. We outline these ideas below.

First we give the formula for modularity for a single mode network in matrix form. Suppose a network with n nodes and m edges has adjacency matrix $\mathbf{A}$. Let $\mathbf{P}$ be a matrix of probabilities in which the i,j^{th} entry is the probability that actor i has an edge to actor j given that the edges are distributed at random (but with the expected degrees made to match those in $\mathbf{A}$). Given a partition of the nodes into c groups let $\mathbf{S}$ be the $n \times c$ indicator matrix in which the i,j^{th} entry is a 1 if actor i is a member of group j and 0 otherwise. Let $\mathbf{B} = \mathbf{A} - \mathbf{P}$ then the modularity Q is given by

$$Q = \frac{1}{2m}\mathrm{Tr}(\mathbf{S}^T\mathbf{B}\mathbf{S}) \tag{9.1}$$

where $\mathrm{Tr}(\mathbf{X})$ is the trace of matrix $\mathbf{X}$. In the 2-mode version the adjacency matrix is replaced by an affiliation matrix $\check{\mathbf{A}}$ and the $\mathbf{P}$ matrix adapted to take account of the bipartite structure to form $\check{\mathbf{P}}$, so that $\mathbf{B}$ is replaced by $\check{\mathbf{B}}$. In addition instead of a single $\mathbf{S}$ indicator matrix we need to have one matrix for each mode which we call $\mathbf{R}$ and $\mathbf{T}$.

```
                                     1   1 1 1 1
               1 2 3 4 5 6   7 8   9 1   0 2 3 4
              -----------------------------------
 1    EVELYN | 1 1 1 1 1 1 |   1 | 1   |         |
 2     LAURA | 1 1 1   1 1 | 1 1 |     |         |
 3   THERESA |   1 1 1 1 1 | 1 1 | 1   |         |
 4    BRENDA | 1   1 1 1 1 | 1 1 |     |         |
 5 CHARLOTTE |     1 1 1   | 1   |     |         |
 6   FRANCES |     1   1 1 |   1 |     |         |
              ------------------------------------
 7   ELEANOR |         1 1 | 1 1 |     |         |
10     VERNE |             | 1 1 | 1   |   1     |
 9      RUTH |         1   | 1 1 | 1   |         |
              ------------------------------------
 8     PEARL |           1 |   1 | 1   |         |
17    OLIVIA |             |     | 1 1 |         |
16   DOROTHY |             |   1 | 1   |         |
18     FLORA |             |     | 1 1 |         |
              ------------------------------------
11     MYRNA |             |   1 | 1   | 1 1     |
15     HELEN |             | 1 1 |   1 | 1 1     |
12 KATHERINE |             |   1 | 1   | 1 1 1 1 |
13    SYLVIA |             | 1 1 | 1   | 1 1 1 1 |
14      NORA |           1 | 1   | 1 1 | 1 1 1 1 |
              ------------------------------------
```

Figure 9.1: Group assignment maximizing modularity

The resultant formula has the form

$$Q = \frac{1}{m}\mathrm{Tr}(\mathbf{R}^T \check{\mathbf{B}} \mathbf{T}) \tag{9.2}$$

If we have c communities and our bipartite network has p actors and q events then $\mathbf{R}$ is $p \times c$, $\check{\mathbf{B}}$ is $p \times q$ and $\mathbf{T}$ is $q \times c$. Barber also proposes an algorithm, BRIM (Bipartite, Recursive Induced Modules) which uses the singular vectors of $\mathbf{B}$ to recursively partition both actors and events into groups. The algorithm does not however provide a method to find the maximum value for c the number of groups. To overcome this he suggests starting with $c = 1$, calculating Q and then keep doubling c until Q decreases. At this stage use bisection to find the value of c which maximises Q.

As an example we apply the technique to the Southern Women Data [16] and obtain 4 groups as given in the blocked affiliation matrix in Figure 9.1. In this data the rows correspond to 18 women and the columns to 14 social events attended by the women.

From the blocked affiliation matrix we can see that the top group of women were the main attendees of the first 6 events and the bottom group of women attended events 10,12,13 and 14. We see that most women attended the middle pair of events with Eleanor, Verne and Ruth all attending events 7 and 8 whereas Pearl, Olivia, Dorothy and Flora all attended event 9 with two of them attending event 11. These groupings are similar (but not exactly the same) to others found in this data [23].

Other authors have suggested alternative extensions for modularity see Guimera et al [24] and Murata [35] as examples as well as other measures [43].

9.4 Dual Projection

A common criticism of projection methods are that information is lost in the process. This is definitely true if the projection is dichotomised or if only one projection is used. However, some authors have claimed (without proof) that there is always data loss even if both projections are used in their un-dichotomized form [31]. Everett and Borgatti [21] challenge this assumption and provide evidence that it is not true. They argue that in the vast majority of cases given two projections it is possible to recover the original matrix and hence no information is lost. This issue has been further explored by Kirkland[26] where he shows as the size of the matrices increases then cases of data loss decrease. He also gives examples of when data loss does occur but these matrices are highly structured and are unlikely to occur in real data. Everett and Borgatti therefore suggest constructing both projections then use methods which are applicable to proximity matrices on both projections and finally combine these, preferably using the original data. They call this approach the dual projection approach and show how it can be used for blockmodeling, in particular core-periphery models and briefly centrality. In a later paper [32] Melamed explores the method as applied to community detection.

The dual projection works best when there are robust methods for analyzing the projected matrices. That is techniques that work well on proximity type data. We shall provide two examples. Our first is a core-periphery partition, this is a blockmodel but as we shall see the periphery is not well connected and so is not community detection. Borgatti and Everett [10] suggest a method for dividing a proximity matrix $\mathbf{P}$ into a core and periphery. This is a two stage process, the first is to find a vector $\mathbf{C}$ such that $||\mathbf{P} - \mathbf{C}\mathbf{C}^T||_2$, that is the Euclidean 2-norm, is minimized. This $\mathbf{C}$ then gives a core-periphery score for each object and we sort $\mathbf{C}$ to form $\mathbf{C}'$ so that its elements are in descending order. Let $\mathbf{I}_k$ be a vector in which the first k elements are 1 and the rest are 0. We next find the value of k for which the correlation of $\mathbf{I}_k$ and $\mathbf{C}'$ is the highest. We now assign the first k objects in $\mathbf{C}'$ to the core and the remainder to the periphery. In our dual projection method we use this on both projections and then map these back onto the affiliation matrix $\mathbf{A}$. Again using the Southern Women data we obtain the 2-mode core periphery structure shown in Figure 9.2.

Looking at the partition in Figure 9.2 we see that the core events were the most popular all with 8 or more attendees, whereas the peripheral events all had 6 or less. We see that peripheral actors attend core events and core actors attend peripheral events but not as much as core actors attending core events. The least dense area is peripheral actors attending peripheral events, which all gives some validation to the core-periphery structure.

We can also use dual projection to do community detection. In order to not lose any structural information we need to partition the proximity matrices (Note this is not the approach taken by Melamed in [32] as he uses a dichotomised projection). Guimera et al [24] do take this approach and use a simple extension to modularity to deal with the valued data. However, they only find clusters of women and attach the relevant events to the clustered women data. They also use the Davis data and report a straight split of the women into two groups namely {Evelyn, Laura, Theresa, Brenda, Charlotte, Frances, Eleanor, Ruth} and attach events 1 through 8 to this group using the weighted method. Clearly all the other women and events belong to the second group.

We partitioned both projections into two groups in order to obtain a comparison. Rather than use the modularity we used a fit function which used correlation between an ideal structure matrix of a 1 for within group interaction and a 0 between groups with a Tabu search. The results are given in Figure 9.3.

```
                                   1 1 1 1 1
               8 9 6 7 5   3 4 1 2 0 1 2 3 4
              -------------------------------
 1    EVELYN | 1 1 1   1 | 1 1 1 1           |
 2     LAURA | 1   1 1 1 | 1   1 1           |
 3   THERESA | 1 1 1 1 1 | 1 1   1           |
 4    BRENDA | 1   1 1 1 | 1 1 1             |
14      NORA |   1 1 1   |         1 1 1 1 1 |
 7   ELEANOR | 1   1 1 1 |                   |
 9      RUTH | 1 1   1 1 |                   |
13    SYLVIA | 1 1   1   |         1   1 1 1 |
             ---------------------------------
 6   FRANCES | 1   1   1 | 1                 |
 8     PEARL | 1 1 1     |                   |
10     VERNE | 1 1   1   |             1     |
12 KATHERINE | 1 1       |         1   1 1 1 |
11     MYRNA | 1 1       |         1   1     |
 5 CHARLOTTE |       1 1 | 1 1               |
15     HELEN | 1     1   |         1 1 1     |
16   DOROTHY | 1 1       |                   |
17    OLIVIA |   1       |           1       |
18     FLORA |   1       |           1       |
              -------------------------------
```

Figure 9.2: Dual Projection Core-Periphery of Southern Women

```
                                   1 1 1 1 1
               1 2 3 4 5 6 7 8 9   0 1 2 3 4
              -------------------------------
 1    EVELYN | 1 1 1 1 1 1   1 1 |           |
 2     LAURA | 1 1 1   1 1 1 1   |           |
 3   THERESA |   1 1 1 1 1 1 1 1 |           |
 4    BRENDA | 1   1 1 1 1 1 1   |           |
 5 CHARLOTTE |     1 1 1   1     |           |
 6   FRANCES |     1   1 1   1   |           |
 7   ELEANOR |         1 1 1 1   |           |
 8     PEARL |           1   1 1 |           |
 9      RUTH |         1   1 1 1 |           |
             ---------------------------------
10     VERNE |             1 1 1 |     1     |
11     MYRNA |               1 1 | 1   1     |
12 KATHERINE |               1 1 | 1   1 1 1 |
13    SYLVIA |             1 1 1 | 1   1 1 1 |
14      NORA |           1 1   1 | 1 1 1 1 1 |
15     HELEN |             1 1   | 1 1 1     |
16   DOROTHY |               1 1 |           |
17    OLIVIA |                 1 |   1       |
18     FLORA |                 1 |   1       |
              -------------------------------
```

Figure 9.3: Dual Projection Community Detection for the Davis data

```
                                1 1 1 1         1
                8 9 3 4 5 6 7   4 2 0 3   1 2   1
               ------------------------------------
  1    EVELYN | 1 1 1 1 1 1   |         | 1 1 |   |
  2     LAURA | 1   1   1 1 1 |         | 1 1 |   |
  3   THERESA | 1 1 1 1 1 1 1 |         |   1 |   |
  4    BRENDA | 1   1 1 1 1 1 |         | 1   |   |
  5 CHARLOTTE |     1 1 1   1 |         |     |   |
  6   FRANCES | 1   1   1 1   |         |     |   |
  7   ELEANOR | 1       1 1 1 |         |     |   |
  9      RUTH | 1 1     1   1 |         |     |   |
               ------------------------------------
 17    OLIVIA |   1           |         |     | 1 |
 18     FLORA |   1           |         |     | 1 |
               ------------------------------------
  8     PEARL | 1 1       1   |         |     |   |
 16   DOROTHY | 1 1           |         |     |   |
               ------------------------------------
 11     MYRNA | 1 1           |   1 1   |     |   |
 10     VERNE | 1 1         1 |   1     |     |   |
 15     HELEN | 1           1 |   1 1   |     | 1 |
 12 KATHERINE | 1 1           | 1 1 1 1 |     |   |
 13    SYLVIA | 1 1         1 | 1 1 1 1 |     |   |
 14      NORA |   1       1 1 | 1 1 1 1 |     | 1 |
               ------------------------------------
```

Figure 9.4: Dual Projection Community Detection for 4 groups

The results in Figure 9.3 are in close agreement with the results obtained by Guimera et al the only difference is we have placed Pearl into the first group together with event 9. We also applied the Louvain method [6] to both projections, while the method is local it does have the advantage of finding the optimum number of clusters. For the Women the method it reproduced the partition found by Guimera et al into two groups. It partitioned the events into 4 groups with events 1 to 6 in the first group, 9 to 14 in a second group and events 7 and 8 both in singleton clusters.

It should be noted that we cannot directly compare these methods with the 2-mode modularity of Barber. First, we can decide to have a different number of partitions in each mode. Secondly even if these are the same we do not necessarily have the groups defined by the diagonal blocks i.e. we do not have communities made up of both women and events but we partition these separately. To see this we repeat the analysis above but with 4 groups in each partition to obtain the results shown in Figure 9.4.

Examining Figure 9.4 we see that two of the diagonal blocks are zero and so this do not correspond to mixed mode communities. We need to examine the partitions and not the blocks, although having both to examine helps us understand the data. For example we can see that Olivia and Flora have been put together as they attended events 9 and 11 together and no other events.

9.5 Signed 2-mode networks

Heider's balance theory [25] has a long tradition in social networks and was formulated in network terms by Cartwright and Harary [14]. In the one-mode formulation the edges of the network are assigned either a positive or a negative sign reflecting positive or negative sentiment. In Heider's original formulation the actors showed positive or negative preferences to objects and so is more analogous to 2-mode data. In our formulation the actors attending events would see them as either positive or negative. A good example is a set of politicians or other actors voting on propositions or resolutions in which they can vote for it, against it or abstain. Data of this type was considered by Mrvar and Doreian [34] where the actors were supreme court judges and Doreian et al [20] where the actors were nation states voting in the UN. It follows that a 2-mode signed network is a 2-mode network of actors and objects (we use objects rather than events as this is more suggestive of the type of data that has been used) in which each edge has a positive or negative sign. In classic balance theory a balanced (one-mode) network can be partitioned into two sets with negative ties between the sets and positive ties within. Extended balance (Davis balance or clusterability) allows more than two clusters but still positive ties within clusters and negative ties between. Relaxed balance [18] again allows for any number of clusters but we now just require all the connections within a cluster to be of the same sign (positive or negative) and all the edges between pairs of clusters must also be of the same sign. In the 2-mode case Mrvar and Doreian retain the idea of relaxed balance, but of course given the structure of the data no longer have within cluster links. We now formalize these ideas but more details can be found in [20] and [18].

Let **A** be a signed affiliation matrix and suppose the rows are partitioned into k_1 clusters and the columns into k_2 clusters so that **A** is partitioned into k_1k_2 blocks. Then we say the clustering is ideal if none of the blocks contain both positive and negative ties. Given a partition of **A** we can measure how close it is to the ideal by simply counting the number of positive and negative violations there are in the blocks, call these P and N. These are used to produce a measure of inconsistency given by $\alpha N + (1-\alpha)P$. The α parameter allows us to weight either positive or negative violations more highly with a value of 0.5 weighting them equally. Unfortunately this function can always be made zero by placing each node in its own unique cluster. In fact Mrvar and Doreian prove a stronger result showing that this function is monotonically decreasing with k_1 and k_2. It is therefore necessary to find strategies in which the blocks remain sufficiently large. For fixed values of k_1 and k_2 Mrvar and Doreian propose a relocation algorithm which helps finding the minimum, but unfortunately it can easily get caught in local minima and needs many starts to reliably find a global minimum. As a consequence it becomes quite a challenge to partition such data particularly given the fact we have two parameters k_1 and k_2 to contend with and the computational complexity of the task. It is possible that other factors can help determine in more detail the structure of the various blocks. This is explored in [20] but is really only feasible because of the small size of one of the modes in the datasets, consisting of nine supreme court judges. Some further guidance on this issue and some ideas are further examined in [18] which has far larger mode sizes but no definitive approach is suggested. We conclude that while some early promising work has been done in this area there are many open issues worthy of further consideration.

9.6 Spectral methods

One technique that has been developed for networks but has been largely ignored in the social network community is bipartite spectral co-partitioning [17]. This technique has the added advantage that it can handle valued 2-mode networks. In this case we have an incidence matrix $\mathbf{A}$ in which $\mathbf{A}(i,j) = w$ indicates that i attended event j with weight w, where higher values represent a stronger association. An example would be that the actors are words, the events documents, and the entries in $\mathbf{A}(i,j)$ give then number of occasions word i was in document j. It was in this context that Dhillon proposed this method. We briefly outline the process but full details are given in his paper.

Given a weighted $n \times m$ affiliation matrix $\mathbf{A}$ then form $\mathbf{D}_1$ an $n \times n$ diagonal matrix with the row sums of $\mathbf{A}$ on the diagonal and $\mathbf{D}_2$ an $m \times m$ diagonal matrix with the column sums on the diagonal. The algorithm proceeds as follows.

1. Form $\mathbf{A}_n = \mathbf{D}_1^{-\frac{1}{2}} \mathbf{A} \mathbf{D}_2^{-\frac{1}{2}}$.
2. Compute the second singular vectors of $\mathbf{A}_n$, $\mathbf{u}_2$ and $\mathbf{v}_2$. That is the eigenvectors of $\mathbf{A}_n \mathbf{A}_n^T$ and $\mathbf{A}_n^T \mathbf{A}_n$ corresponding to the second largest eigenvalue.
3. Form $\mathbf{z}_2 = \begin{bmatrix} \mathbf{D}_1^{-\frac{1}{2}} \mathbf{u}_2 \\ \mathbf{D}_2^{-\frac{1}{2}} \mathbf{v}_2 \end{bmatrix}$.
4. Run k-means on $\mathbf{z}_2$ to bipartition the data.

This clearly partitions the rows and columns into two and we can recursively apply the procedure to obtain a finer partition. Alternatively Dhillon gives an extended version that allows us to find k groups by using additional singular vectors. Let $p = \lceil \log_2 k \rceil$ and instead of a vector $\mathbf{z}_2$ in step 3 create a matrix $\mathbf{Z}$ i.e.

Form $\mathbf{Z} = \begin{bmatrix} \mathbf{D}_1^{-\frac{1}{2}} \mathbf{U} \\ \mathbf{D}_2^{-\frac{1}{2}} \mathbf{V} \end{bmatrix}$.

Where $\mathbf{U} = [\mathbf{u}_2 : \mathbf{u}_3 : \ldots : \mathbf{u}_{p+1}]$ and $\mathbf{V}$ is defined similarly. Again use k-means to cluster the rows of $\mathbf{Z}$ with the first n rows giving the partition of the rows and the last m rows giving the partition of the columns.

We note that this method produces groups made up of nodes from both modes as in the Barber community detection discussed in section 3. As an example we do a two cluster split on the Davis data and the result is shown in Figure 9.5. As can be seen this is very similar to the dual projection community detection shown in Figure 9.3 with only event 9 moved to a different group and thus this partition agrees with that found by Guimera et al [24] and by the Louvain method in the dual projection as discussed above.

We also obtained a split into 4 groups which involves using two of the singular vectors to obtain the partition shown in Figure 9.6.

As we found groups of women and events we need to compare this result with the modularity result shown in Figure 9.1. In this case the event split is nearly the same with just event 9 moved from being with event 11 to the community containing evets 7 and 8. This leaves event 11 in a single cluster as found by the four split in the dual projection method. The first group of women is also similar but the effect of moving event 11 to a singleton cluster separates out Olivia and Flora into a community pair. This has a knock on effect on the way the rest of the women are partitioned, and although there are some

```
                                   1 1 1 1 1
               1 2 3 4 5 6 7 8   9 0 1 2 3 4
              -------------------------------
 1    EVELYN | 1 1 1 1 1 1   1 | 1           |
 2     LAURA | 1 1 1   1 1 1 1 |             |
 3   THERESA |   1 1 1 1 1 1 1 | 1           |
 4    BRENDA | 1   1 1 1 1 1 1 |             |
 5 CHARLOTTE |     1 1 1   1   |             |
 6   FRANCES |     1   1 1   1 |             |
 7   ELEANOR |         1 1 1 1 |             |
 8     PEARL |           1   1 | 1           |
 9      RUTH |         1   1 1 | 1           |
             ---------------------------------
10     VERNE |             1 1 | 1     1     |
11     MYRNA |               1 | 1 1   1     |
12 KATHERINE |               1 | 1 1   1 1 1 |
13    SYLVIA |             1 1 | 1 1   1 1 1 |
14      NORA |           1 1   | 1 1 1 1 1 1 |
15     HELEN |             1 1 |   1 1 1     |
16   DOROTHY |               1 | 1           |
17    OLIVIA |                 | 1   1       |
18     FLORA |                 | 1   1       |
              --------------------------------
```

Figure 9.5: Spectral Bisection of the Davis Data into two groups

```
                                     1 1 1 1   1
               1 2 3 4 5 6   7 8 9   0 4 2 3   1
              -----------------------------------
 1    EVELYN | 1 1 1 1 1 1 |   1 1 |         |   |
 2     LAURA | 1 1 1   1 1 | 1 1   |         |   |
 3   THERESA |   1 1 1 1 1 | 1 1 1 |         |   |
 4    BRENDA | 1   1 1 1 1 | 1 1   |         |   |
 5 CHARLOTTE |     1 1 1   | 1     |         |   |
 6   FRANCES |     1   1 1 |   1   |         |   |
 7   ELEANOR |         1 1 | 1 1   |         |   |
             -------------------------------------
 8     PEARL |           1 |   1 1 |         |   |
 9      RUTH |         1   | 1 1 1 |         |   |
10     VERNE |             | 1 1 1 |     1   |   |
15     HELEN |             | 1 1   | 1   1   | 1 |
16   DOROTHY |             |   1 1 |         |   |
14      NORA |           1 | 1   1 | 1 1 1 1 | 1 |
             -------------------------------------
13    SYLVIA |             | 1 1 1 | 1 1 1 1 |   |
11     MYRNA |             |   1 1 | 1   1   |   |
12 KATHERINE |             |   1 1 | 1 1 1 1 |   |
             -------------------------------------
17    OLIVIA |             |     1 |         | 1 |
18     FLORA |             |     1 |         | 1 |
              ------------------------------------
```

Figure 9.6: Spectral Bisection of the Davis Data into four groups

similarities there are also differences and it is difficult without additional information to decide which partition is the best.

It should be noted that the method can be extended to obtain different partitions of the rows and columns, see Kluger et al [28]. This requires different normalization and three normalization schemes are proposed as well as a more sophisticated technique for partitioning the rows and the columns separately. The first suggested normalization is to make all the rows have the same mean and all the columns to have the same mean (but not necessarily the same as the row mean). The second method suggests making all the row means the same as all the column means. The third method involves taking the log of the matrix $\mathbf{L}$ and then for each entry of $\mathbf{L}$, $\mathbf{L}(i, j) = \log \mathbf{A}(i, j)$, subtract off its row mean, its column mean and the overall mean of $\mathbf{L}$. Finally in step 3 do not form $\mathbf{Z}$ but run k-means on $\mathbf{AV}$ and $\mathbf{A}^T\mathbf{U}$ separately where $\mathbf{U}$ and $\mathbf{V}$ are constructed from the singular vectors by selecting subsets that are the best projections. We do not give the details here and so the interested reader should consult their paper for details.

One further approach needs to be mentioned and that is 2-mode stochastic block models proposed by Larremore et al [29]. In brief they assume a Poisson distribution and then search the likelihood space using a modified Kernigan-Lin algorithm. In its simple form this tends to sort out actors purely by degree but they use a degree correction procedure to counteract this. The degree correction explicitly takes into account the degree distribution of the data which allows for a wide variety of empirical degree distributions. When they use this technique on the southern women data they get exactly the same partition as the dual projection split shown in Figure 9.3.

9.7 Clustering

So far we have examined partitioning i.e. we have insisted that every node is placed uniquely in one group. We now relax this condition and allow nodes to be placed in more than one group and do not insist that actors are assigned to any group. We shall only consider clustering in which we are trying to find subsets of nodes which are well connected to each other. As mentioned before we shall refer to these as cohesive subgroups. The standard one mode definition of a clique as a maximal complete subgraph clearly can be generalized to a biclique as a maximally complete bipartite subgraph. Such structures have been considered in mathematics for over 400 years although not usually as subgraphs . The use of bicliques as cohesive subgraphs in social networks was probably due to Borgatti and Everett [9] in 1997. It is a very simple matter to extend standard clique algorithms to find bi-cliques and the same techniques used to deal with overlap can be applied. It is also a simple matter to extend concepts such as k-plexes, n-cliques, n-clubs, n-clans etc to the 2-mode case. We also note that since the projections result in proximity data then the vast number of clustering algorithms that have been developed for this type of data can be used. We should also comment that a common measure in 1-mode networks is the clustering coefficient (or transitivity index) that attempts to capture the extent to which a network is clustered. In this chapter we are concerned with uncovering structural patterns in terms of finding sets of nodes rather than providing graph invariants that try and capture the extent to which a network has a particular property. There have been a number of suggestions for extending the clustering coefficient (and other invariants) to 2-mode data and the interested reader should consult [30] and [37].

One area in which there has been some developments is that of k-cores and their extension to 2-mode. A k-core is an induced subgraph in which every node has degree k or more and was first proposed by Seidman [40] and extended to 2-mode in [1]. A k-core is not a cohesive subgroup but any cohesive subgroups must be wholly contained in a k-core. The concept was extended to generalized k-cores in [3] and then to 2-mode generalized cores by Cerinšek and Batagelj [15]. The idea of a generalized core is to extend the concept of degree to a property function defined on the nodes. For a network $N = (V, L, w)$ with node set V, edge set L and a weight function $w : L \to \mathbb{R}^+$ a property function $f(v,C) \in \mathbb{R}_0^+$ is defined for all $v \in V$ and $C \subseteq V$. A subset C induces the subnetwork to which the evaluation of the property function is restricted. We can now give a formal definition of a generalized 2-mode (p,q)-core. Let $N = ((V_1,V_2), L, (f,g), w)$, $V = V_1 \cup V_2$ be a finite 2-mode network – the sets V and L are finite. Let $P(V)$ be the power set of the set V. Let functions f and g be defined on the network N: $f, g : V \times P(V) \to \mathbb{R}_0^+$. A subset of nodes $C \subseteq V$ in a 2-mode network N is a generalized 2-mode core $C = \text{Core}(p,q; f,g)$, $p,q \in \mathbb{R}_0^+$ if and only if in the subnetwork $K = ((C_1,C_2), L|C, w)$, $C_1 = C \cap V_1$, $C_2 = C \cap V_2$ induced by C it holds that for all $v \in C_1 : f(v,C) \geq p$ and for all $v \in C_2 : g(v,C) \geq q$, and C is the maximal such subset in V. Algorithms for finding generalized 2-mode cores are relatively straight forward and efficient and are based on the simple idea of deleting nodes that do not satisfy the criteria. The paper provides some examples drawn from web of science data.

9.8 More complex data

So far we have considered 2-mode static data. If there are more than two modes then in some circumstances it is a straight forward matter to extend the 2-mode case to more modes. As already mentioned Borgatti and Everett [8] showed how to extend regular and structural equivalence to multimode data but they did not address the computational issues. Batagelj et al [4] suggest a dissimilarity measure for 3-mode structural equivalence and apply Ward's algorithm to partition the data. They demonstrate the effectiveness of their approach on a three way cognitive social structure.

If we have k-modes then we examine the $k(k-1)/2$ collection of 2-mode networks between every pair of modes. Let $\mathbf{A}_{(r,s)}$ denote the 2-mode affiliation matrix between mode r and mode s where $r < s$ and s runs from 1 to k. If we wanted to extend community detection then it is a simple matter to construct $\check{\mathbf{B}}_{(r,s)}$ and then define Q as

$$Q = \frac{1}{m} \sum_{i<j} \text{Tr}(\mathbf{S}_i^T \; \check{\mathbf{B}}_{(i,j)} \; \mathbf{S}_j) \tag{9.3}$$

Where S_i is the ith mode indicator matrix. One issue is that it is now not possible to use the spectral methods suggested described in section in section 3 in order to find a maximum for Q. Clearly we can use other optimization methods but these will probably not be as efficient. One solution to this problem is to simply construct an adjacency matrix $\mathbf{Z}$ from $\check{\mathbf{B}}_{(i,j)}$. We form $\mathbf{Z}$ (given in blocked form) as follows:

$$\mathbf{Z}(i,j) = \begin{cases} \check{\mathbf{B}}_{(i,j)} & \text{if } i < j \\ \check{\mathbf{B}}_{(i,j)}^T & \text{if } j < i \\ \mathbf{0} & \text{if } i = j \end{cases}$$

So that $\mathbf{Z}$ is a square adjacency matrix in which all the modes have been included. The fact there are not connections within the modes is captured by the zero blocks on the diagonal. In this case we have:

$$Q = \frac{1}{2m}\mathrm{Tr}(\mathbf{S}^T\mathbf{Z}\mathbf{S}) \tag{9.4}$$

Where m is the number of edges and $\mathbf{S}$ is an indicator matrix over all the modes. We can now use spectral partitioning to maximize Q. An example of this approach for a three-mode network is given in [33].

Looking at the other methods discussed above there does not seem any reason why spectral bipartitioning cannot be extended in the same way but this approach does not seem to have been explored. In this case the constructed adajacency matrix would not have $\check{\mathbf{B}}_{(i,j)}$ as the blocks but would use $\mathbf{A}_{(i,j)}$. The one exception for extending to more than two modes in this way is dual projection. In this instance we would not construct a large adjacency matrix but would project all pairs of $\mathbf{A}_{(i,j)}$. In this case when we have more than two modes then the same mode appears in a number of different projection matrices. As a consequence each mode would have a number of different partitions and this would not generally be of use unless the goal is to find different partitions for different pairs of modes.

We mention one further complication that is multi-mode data that involves a time element. Both data and techniques for dealing with such data are not common. However Tang et al [41] examine how communities evolve over time in a dynamic multi-mode framework. They consider time stamped data and they try and make a smooth transition in terms of community detection from one time stamp to the next. They present an algorithm and an example that uses the Enron email corpus [27].

9.9 Conclusion

In this chapter we have examined partitioning and clustering in multi-mode network data. We briefly mentioned that there are a number of techniques developed for dealing with non-binary data or more precisely non-network type data. We have not explored those methods in much detail as they are generally well described elsewhere. We have not discussed software but most of the articles referenced that develop methods discuss implementations and point to available tools. In addition we have mainly discussed methodological issues and have not discussed applications. There are an increasing number of application areas that are using these methods ranging from biology and information science through to sociology and political science. A good flavour of how to interpret some of these techniques can be gleaned from the examples in the book by Batagelj et al [5]. It should be noted that this is a very active area for research and new methods and ideas are constantly being explored particularly as new types of data emerge. The complexities of this type of data in terms of collecting, analyzing and interpreting remain both challenging and deeply fascinating.

Bibliography

1. A. Ahmed, V. Batagelj, X. Fu, S. Hong, D. Merrick, and A. Mrvar. Visualisation and analysis of the internet movie database. In *APVIS 2007, 6th International Asia-Pacific Symposium on Visualization 2007, Sydney, Australia, 5-7 February 2007*, pages 17–24, 2007. doi: 10.1109/APVIS.2007.329304.

2. M. J. Barber. Modularity and community detection in bipartite networks. *Phys. Rev. E*, 76(6):066102, 2007.

3. V. Batagelj and M. Zaveršnik. Fast algorithms for determining (generalized) core groups in social networks. *Advances in Data Analysis and Classification*, 5(2):129–145, 2011. doi: 10.1007/s11634-010-0079-y.

4. V. Batagelj, A. Ferligoj, and P. Doreian. *Indirect Blockmodeling of 3-Way Networks*, pages 151–159. Springer, Berlin, Heidelberg, 2007. ISBN 978-3-540-73560-1. doi: 10.1007/978-3-540-73560-1_14. URL `https://github.com/bavla/ibm3m`.

5. V. Batagelj, P. Doreian, A. Ferligoj, and N. Kejžar. *Understanding Large Temporal Networks and Spatial Networks: Exploration, Pattern Searching, Visualization and Network Evolution*. Wiley Series in Computational and Quantitative Social Science Series. John Wiley & Sons, New York, 2014. ISBN 9781118915370.

6. V. D. Blondel, J.-L. Guillaume, R. Lambiotte, and E. Lefebvre. Fast unfolding of communities in large networks. *J. Statistical Mechanics-theory Experiment*, page P10008, 2008.

7. S. Borgatti. Regular equivalence in graphs, hypergraphs, and matrices, 1989. Ph.D. dissertation.

8. S. P. Borgatti and M. G. Everett. Regular blockmodels of multiway, multimode matrices. *Social Networks*, 14(1-2):91–120, 1992.

9. S. P. Borgatti and M. G. Everett. Network analysis of 2-mode data. *Social Networks*, 19(3):243–269, 1997.

10. S. P. Borgatti and M. G. Everett. Models of core/periphery structures. *Social Networks*, 21(4):375–395, 1999.

11. R. L. Breiger. Duality of persons and groups. *Social Forces*, 53(2):181–190, 1974.

12. M. Brusco and D. Steinley. A variable neighborhood search method for generalized blockmodeling of two-mode binary matrices. *J. Math. Psychology*, 51(5):325–338, 2007.

13. R. S. Burt. Cooptive corporate actor networks - a reconsideration of interlocking directorates involving american manufacturing. *Administrative Science Quarterly*, 25 (4):557–582, 1980.

14. D. Cartwright and F. Harary. Structural balance - a generalization of heider theory. *Psychological Rev.*, 63(5):277–293, 1956.

15. M. Cerinšek and V. Batagelj. Generalized two-mode cores. *Social Networks*, 42: 80–87, 2015. URL `https://github.com/bavla/pqCores`.

16. A. Davis, B. Gardner, and M. Gardner. *Deep South: A Sociological Anthropological Study of Caste and Class*. University of Chicago Press, Chicago, 1941.

17. I. S. Dhillon. Co-clustering documents and words using bipartite spectral graph partitioning. In *Proceedings of the seventh ACM SIGKDD international conference on Knowledge discovery and data mining, San Francisco, CA, USA, August 26-29, 2001*, pages 269–274, 2001.

18. P. Doreian and A. Mrvar. Partitioning signed social networks. *Social Networks*, 31(1): 1–11, 2009.

19. P. Doreian, V. Batagelj, and A. Ferligoj. Generalized blockmodeling of two-mode network data. *Social Networks*, 26(1):29–53, 2004.

20. P. Doreian, P. Lloyd, and A. Mrvar. Partitioning large signed two-mode networks: Problems and prospects. *Social Networks*, 35(2):178–203, 2013.

21. M. G. Everett and S. P. Borgatti. The dual-projection approach for two-mode networks. *Social Networks*, 35(2):204–210, 2013.

22. S. Fortunato and M. Barthelemy. Resolution limit in community detection. *Proc. National Acad. Sciences United States Am.*, 104(1):36–41, 2007.

23. L. C. Freeman. Finding social groups: A meta-analysis of the southern women data. In R. Breiger, K. Carley, and P. Pattison, editors, *Dynamic Social Network Modeling and Analysis*, pages 39–97. The National Academies Press, Washington, DC, 2003.

24. R. Guimera, M. Sales-Pardo, and L. A. N. Amaral. Module identification in bipartite and directed networks. *Phys. Rev. E*, 76(3):036102, 2007.

25. F. Heider. *The psychology of interpersonal relations*. John Wiley & Sons, New York, 1958.

26. S. Kirkland. Two-mode networks exhibiting data loss. *Journal of Complex Networks*, page online, 2017.

27. B. Klimt and Y. Yang. Introducing the enron corpus. In *CEAS 2004 - First Conference on Email and Anti-Spam, July 30-31, 2004, Mountain View, California, USA*, 2004. URL `http://www.ceas.cc/papers-2004/168.pdf`.

28. Y. Kluger, R. Basri, J. T. Chang, and M. Gerstein. Spectral biclustering of microarray data: Coclustering genes and conditions. *Genome Research*, 13(4):703–716, 2003.

29. D. B. Larremore, A. Clauset, and A. Z. Jacobs. Efficiently inferring community structure in bipartite networks. *Phys. Rev. E*, 90(1):012805, 2014.

30. M. Latapy, C. Magnien, and N. Del Vecchio. Basic notions for the analysis of large two-mode networks. *Social Networks*, 30(1):31–48, 2008.

31. S. Lehmann, M. Schwartz, and L. K. Hansen. Biclique communities. *Phys. Rev. E*, 78 (1):016108, 2008.

32. D. Melamed. Community Structures in Bipartite Networks: A Dual-Projection approach. *Plos One*, 9(5):e97823, 2014.

33. D. Melamed, R. L. Breiger, and A. J. West. Community structure in multi-mode networks: Applying an eigenspectrum approach. *Connections*, 33(1):18–23, 2013.

34. A. Mrvar and P. Doreian. Partitioning Signed Two-Mode networks. *J. Math. Sociology*, 33(3):196–221, 2009.

35. T. Murata. Detecting communities from bipartite networks based on bipartite modularities. In *Proceedings of the 12th IEEE International Conference on Computational Science and Engineering, CSE 2009, Vancouver, BC, Canada, August 29-31, 2009*, pages 50–57, 2009. doi: 10.1109/CSE.2009.81.

36. M. E. J. Newman. Modularity and community structure in networks. *Proc. National Acad. Sciences United States Am.*, 103(23):8577–8582, 2006.

37. T. Opsahl. Triadic closure in two-mode networks: Redefining the global and local clustering coefficients. *Social Networks*, 35(2):159–167, 2013.

38. M. A. Porter, P. J. Mucha, M. E. J. Newman, and C. M. Warmbrand. A network analysis of committees in the US House of representatives. *Proc. National Acad. Sciences United States Am.*, 102(20):7057–7062, 2005.

39. T. Roediger-Schluga and M. J. Barber. R&d collaboration networks in the european framework programmes: Data processing, network construction and selected results. *International Journal of Foresight and Innovation Policy*, 4(3-4):321–347, 2008.

40. S. B. Seidman. Network structure and minimum degree. *Social Networks*, 5(3):269–287, 1983.

41. L. Tang, H. Liu, J. Zhang, and Z. Nazeri. Community evolution in dynamic multi-mode networks. In *Proceedings of the 14th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, Las Vegas, Nevada, USA, August 24-27, 2008*, pages 677–685, 2008. doi: 10.1145/1401890.1401972.

42. J. van Rosmalen, P. J. F. Groenen, J. Trejos, and W. Castillo. Optimization Strategies for Two-Mode partitioning. *J. Classification*, 26(2):155–181, 2009.

43. Y. Xu, L. Chen, B. Li, and W. Liu. Density-based modularity for evaluating community structure in bipartite networks. *Information Sciences*, 317:278–294, 2015.