Working with group memberships: 2-mode data

So far we have only considered data in which ties are given directly among a single set of actors. However, there are circumstances in which there are two or more different kinds of actors, and ties are collected only between the two kinds of actors, not within each type. [martin I really don't like this intro – you are implicitly thinking bipartite graph. Naïve reader would be puzzled by why there are no ties within groups and would assume you are talking about heterophily. I think it is better to start with the notion that sometimes you don't have ties among actors but you do know their memberships in groups or events, from which you can infer ties. After the reader has worked with an example, then you can introduce bipartite idea.]

A simple example would be a group of students and a set of classes. We could construct a network which ties students to classes. The relationship is "attends class." There would be no ties directly between students nor would there be ties between classes only ties connecting students to classes. We can represent this data by a special type of matrix called an affiliation matrix. This is rather like an adjacency matrix except the rows would represent one group and the students say, and the columns represent a different group, the classes. The matrix is no longer square and we refer to data of this type as 2-mode to reflect the fact there are two different types of nodes in the network.

Davis Gardner and Gardner collected data on the attendance at 14 society events by 18 southern women. There data is given in Figure 11.1 where the rows of the affiliation matrix are the women and the columns are the 14 events.

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

Figure 11.1

A one in row i column j indicates that woman i attended event j. Hence we see that Laura attended event 3 but not event 4.

This kind of data can be analyzed in a number of different ways. Two important ways are converting to 1-mode and converting to bipartite form. We consider both ways in this chapter.

11.1 Converting to 1-mode

One approach to dealing with data of this type is to convert it to 1-mode data – a new dataset in which a pair of actors is said to be tied to the extent that they share affiliations. This can be a relatively simple process. As an example, consider the Davis data above. We can construct a new 1-mode matrix in which both the rows and columns represent women, and the matrix cell values indicate the number of events the women with the relationship attended an event together. Mathematically, this can be done by post-multiplying the 2-mode matrix by its transpose, but more simply, for each pair of rows, we look at each column and count the number of times that both are 1. Hence Evelyn and Laura have a link in this new dataset since they both attended event 1 for example. On the other hand Flora did not attend any events with Laura and so they are not connected. In fact we can do more than simply construct a binary matrix we can form a proximity matrix in which the entries give the number of events each pair attended. These are sometimes called co-membership matrices. The co-membership matrix for the Davis data in Figure 11.1 is given in Figure 11.2.

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

Figure 11.2

We can see from the matrix that Brenda attended 6 events with Evelyn. Note that the diagonal elements give the number events each woman attended. We could also form a 1-mode matrix of the events instead of the women. This would result in an event by event matrix in which the entries would show how many women attended both events in common.

Once we have 1-mode data we can then apply all of the normal techniques used for analyzing valued network data. There is however one important caution when taking this approach. If the number of actors that attend each event is large, then we must be careful in interpreting co-membership. Two actors can attend large events in common – e.g., political demonstrations – and never even meet each other. In such cases, we might want to interpret the co-membership tie as a potential for interaction – the more events a pair of women attend in common, the greater the chance of meeting, establishing a relationship, etc. Or we may see it as a potential for activation. For example, suppose you and I are strangers but I would like to enlist you to join me in donating some money to a charitable cause. I may have an easier time of it if I can point out that we attended the same university, belong to the same country club, and so on.

A related issue with converting 2-mode data to 1-mode data is that large events create ties among many actors, even though the ties may be less meaningful that more intimate events. In most cases, we would consider it more significant if a pair of actors attended a number of small events together, than if they attended the same number of large events. Thus, when counting up the number of co-memberships per person, we might want to weight the events inversely by size so that co-membership in small events counts more than co-membership in large events. Programs like

UCINET offer an option for performing this weighting when converting to 1-mode data.

We might also be interested not just in the observed pattern of overlaps, but in the underlying tendency toward affiliation with certain actor rather others. For example, consider two actors who have no particular preference for each other or underlying commonalities. But they each attend many events – they are joiners. Then by chance alone, we should expect to find many instances in which they overlap. In contrast, consider two actors who are best of friends and do everything together. But they attend few events in a year. They will not overlap as much as the two described earlier, but as a percentage of the number of possible overlaps, they would be much higher. Thus, another possibility is to normalize the raw com-memberships by dividing by the maximums possible – given the number of events each attended overall – or by comparing with the expected values in a chance model, much like the chi-square test for independence in statistics. These options are also offered in programs like UCINET.

Finally, it is worth noting that the process of constructing normalized measures of overlaps can be viewed as measuring the similarity of rows (or columns) of the 2-mode matrix. Hundreds of similarity measures appropriate for binary data have been proposed in the literature, almost all of which can be used in this context as well.

11.2 Converting Valued 2-Mode Matrices to 1-Mode

So far we have only considered binary affiliation matrices. If the original 2-mode data were valued then we would need to take account of this when we construct our 1-mode datasets. As noted earlier, the matrix given in Figure 11.2 can be constructed from the matrix in Figure 11.1 by simply multiplying the original matrix by its transpose. Effectively, for each pair of rows we simply multiply the entries in each column and sum them up. Since a zero times anything is zero, the sum is only greater than zero when there are columns in which both values are 1s. We can use the same method for valued data. However this would mean the elements would be multiplied and then summed to give a value in the 1-mode data. This is a figure that is rather difficult to interpret, although clearly high values would indicate strong ties to the same events.

A more interpretable approach would be to take the minimum of the two cell values rather than the product. Hence if row i was (5,6,0,1) and row j was (4,2,4,0) then AA'(i,j) is 5x4+6x2+0x4+1x0=32 for the normal matrix product and min(5,4)+min(6,2)+min(0,4)+min(1,0)=6 for the minimum method. These would produce the same answers for binary data. To see why the minimum is more interpretable suppose the 2-mode dataset recorded how many hours each member of a consulting company spent on each client project. That is, the rows are persons and the columns are projects. Then constructing the person-by-person 1-mode matrix using the minimum method would yield the maximum possible time each pair of persons could have spent together.

11.2 Bipartite Networks

One of the problems with converting the affiliation matrix to 1-mode is that there is a loss of information. Two women could have the same degree of overlap as another pair of women, but through entirely different events. An alternative approach is to treat the affiliation matrix as if it were a piece of a much larger adjacency matrix in which the rows consist of both women and events, and the columns also consiste of both women and events, as shown in Figure xxxx. Note that ties exist only between women and events – there are no ties among women or among events. A dataset with this structure is known as a bipartite network. A graphical representation of the bipartite version of the Davis data is shown in Figure 11.3.

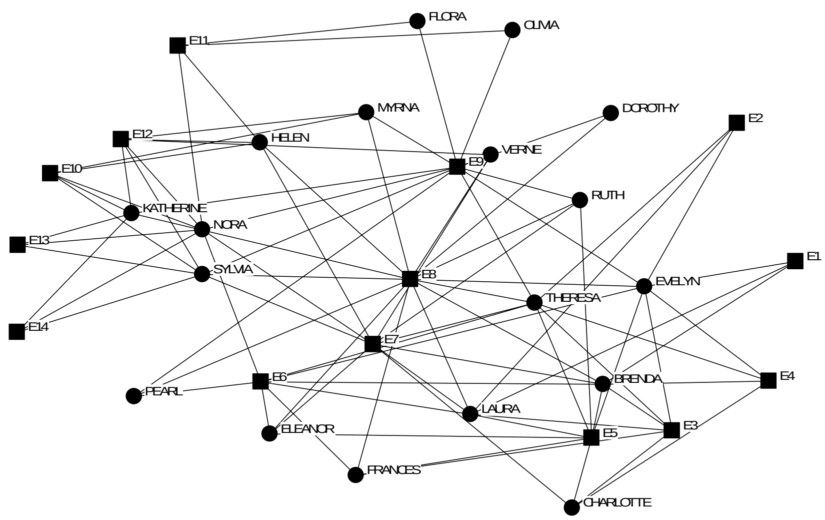

Figure 11.4. Circular nodes indicate women, square nodes indicate events.

An alternative visualization is to put the women on one side of the picture and the events on the other so that the edges only go across the page.

Since the bipartite network is simply a network then we can apply all of the normal network methods. However we need to be aware that since edges cannot occur within the two groups, this will affect our results. For example we cannot find any cliques, since the shortest possible cycle is of length 4. Also, standard normalizations of measures like centrality usually assume that all actors could in principle be connected to each other. Take as an example degree centrality. Ruth has degree 4 and so the normalized degree centrality would take the 4 and divide it by n-1, which is 31 in this case, to yield a normalized centrality of 13%. But Ruth could only attend a maximum of 14 events and so her normalized degree centrality should actually be 29%. We can run all of the standard centrality and centralization routines but we need to adjust the normalization scores to reflect the nature of the data.

We can also adjust the fit function for the optimization routines to take account of the special nature of 2-mode data. Hence we can search for 2-mode factions or 2-mode core periphery structures.