

Analyzing Affiliation Networks

Stephen P. Borgatti and Daniel S. Halgin

In social network analysis, the term “affiliations” usually refers to membership or participation data, such as when we have data on which actors have participated in which events. Often, the assumption is that co-membership in groups or events is an indicator of an underlying social tie. For example, Davis et al. (1941) used data provided by the society pages of a local newspaper to uncover distinct social circles among a set of society women. Similarly, Domhoff (1967) and others have used co-membership in corporate boards to search for social elites (e.g., Allen, 1974; Carroll et al., 1982; Galaskiewicz, 1985; Westphal and Khanna 2003). Alternatively, we can see co-participation as providing opportunities for social ties to develop, which in turn provide opportunities for things like ideas to flow between actors. For example, Davis (1991; Davis and Greve, 1997) studied the diffusion of corporate practices such as poison pills and golden parachutes. He found evidence that poison pills diffuse through chains of interlocking directorates, where board members who sit on multiple boards serve as conduits of strategic information between the different firms. An important advantage of affiliation data, especially in the case studying elites, is that affiliations are often observable from a distance (e.g., government records, newspaper reports), without requiring access to the actors.

In this chapter, we focus on issues involving the analysis of affiliation data, as opposed to the collection or the theoretical interpretation of affiliation data.

BASIC CONCEPTS AND TERMINOLOGY

Affiliation data consist of a set of binary relationships between members of two sets of items. For example, the well-known dataset collected by Davis et al. (1941) recorded which women attended which social events in a small southern town. Thus, there are two sets of items, women and events, and there is a binary relation that connects them, namely the “attended” relation. Figure 28.1 gives the Davis et al. (henceforth, DGG) data matrix in its original form. The rows correspond to the women and the columns are the events they attended.

In general, the kinds of binary relations we consider affiliations are limited to part/whole relations such as “is a member of” or “is a participant in” or “has” (in the sense of having a trait). Examples of affiliation data that have found their way into the social science literature include corporate board memberships (e.g., Mizruchi, 1983, 1992, 1996; Carroll et al., 1982; Davis, 1991; Lester and Canella, 2006; Robins and Alexander, 2004; Westphal, 1998), attendance at events (e.g., Davis et al., 1941; Faust et al., 2002), membership in clubs (e.g., McPherson, 1982; McPherson and Smith-Lovin, 1986, 1987), participation in online groups (Allatta, 2003, 2005), authorship of articles (e.g., Gmür, 2006; Lazer et al., 2009; Newman et al., 2001), membership in production teams (Uzzi and Spiro, 2005), and even course-taking patterns of high school students (e.g., Field et al., 2006). In addition, affiliation data are well known outside the social sciences, as in the

	E1	E2	E3	E4	E5	E6	E7	E8	E9	E10	E11	E12	E13	E14
EVELYN	1	1	1	1	1	1	0	1	1	0	0	0	0	0
LAURA	1	1	1	0	1	1	1	1	0	0	0	0	0	0
THERESA	0	1	1	1	1	1	1	1	0	0	0	0	0	0
BRENDA	1	0	1	1	1	1	1	0	0	0	0	0	0	0
CHARLOTTE	0	0	1	1	1	0	1	0	0	0	0	0	0	0
FRANCES	0	0	1	0	1	1	0	1	0	0	0	0	0	0
ELEANOR	0	0	0	0	1	1	1	0	0	0	0	0	0	0
PEARL	0	0	0	0	0	1	0	1	1	0	0	0	0	0
RUTH	0	0	0	0	1	0	1	1	1	0	0	1	0	0
VERNE	0	0	0	0	0	0	1	1	1	0	0	1	0	0
MYRNA	0	0	0	0	0	0	0	1	1	1	0	1	0	0
KATHERINE	0	0	0	0	0	0	0	1	1	1	0	1	1	1
SYLVIA	0	0	0	0	0	0	1	1	1	1	0	1	1	1
NORA	0	0	0	0	0	1	1	0	1	1	1	1	1	1
HELEN	0	0	0	0	0	0	1	1	0	1	1	1	0	0
DOROTHY	0	0	0	0	0	0	0	1	1	0	0	0	0	0
OLIVIA	0	0	0	0	0	0	0	0	1	0	1	0	0	0
FLORA	0	0	0	0	0	0	0	0	1	0	1	0	0	0

Figure 28.1 DGG women-by-events matrix

species-by-trait matrices of numerical taxonomy (Sokal and Sneath, 1973).

We can represent affiliations as mathematical graphs (Harary, 1969) in which nodes correspond to entities (such as women and events) and lines correspond to ties of affiliation among the entities. Figure 28.2 provides a representation of the DGG data. Affiliation graphs are distinctive in having the property of bipartiteness, which means that the graph's nodes can be partitioned into two classes, so that all ties occur only between classes and never within classes. We see in Figure 28.2 that there are only lines between women and the events that they attended. While all affiliation graphs are bipartite, in our view the reverse is not necessarily true. In empirical network data, graphs can be bipartite by chance alone, perhaps because of sampling error. What makes affiliation graphs different is that the two node sets are different kinds of entities, and the lack of ties within sets is by design, not happenstance. Formally, we define an affiliation graph as a bipartite graph $G(V_1, V_2, E)$, in which V_1 and V_2 are sets of nodes corresponding to different classes of entities, and E is an affiliation relation that maps the elements of V_1 to V_2 . The relation is typically conceived as a set of unordered pairs in which one element of each pair belongs to V_1 and the other belongs to V_2 . In contexts where we discuss multiple graphs, we use the notation $V_1(G)$ to indicate the V_1 node set in graph G , and $E(H)$ to refer to the ties in graph H .

Affiliation graphs or networks are often called "two-mode graphs." The terminology of "modes" refers to the number of different kinds of entities referenced in the rows and columns of a matrix. A one-mode matrix is square, and its rows and columns refer to the same set of entities—a single mode. An example, drawn from the famous Hawthorne studies (Roethlisberger and Dickson, 1939), is shown in Figure 28.3.¹

In contrast, a two-mode matrix is rectangular, and the rows and columns refer to two different sets of entities—two modes. For example, Figure 28.4 shows a two-mode, n-by-m person-by-group incidence matrix that is also based on the Hawthorne data. An incidence matrix has rows corresponding to nodes and columns corresponding to n-ary edges (also called hyperedges) that connect sets of nodes. In this case, the matrix indicates each individual's membership in each of five different groups.² The matrix clearly represents affiliations, and indeed all affiliation graphs can be represented as two-mode matrices, where the two modes correspond to the affiliation graph's two node sets.

It is important to note that while affiliation graphs can be represented by two-mode matrices, not all two-mode matrices are considered affiliation graphs. For example, a standard sociological case-by-variables matrix (e.g., person-by-demographics) might be seen as two-mode but would not normally be called an affiliation. "Affiliation" is reserved for the case when the data consist of

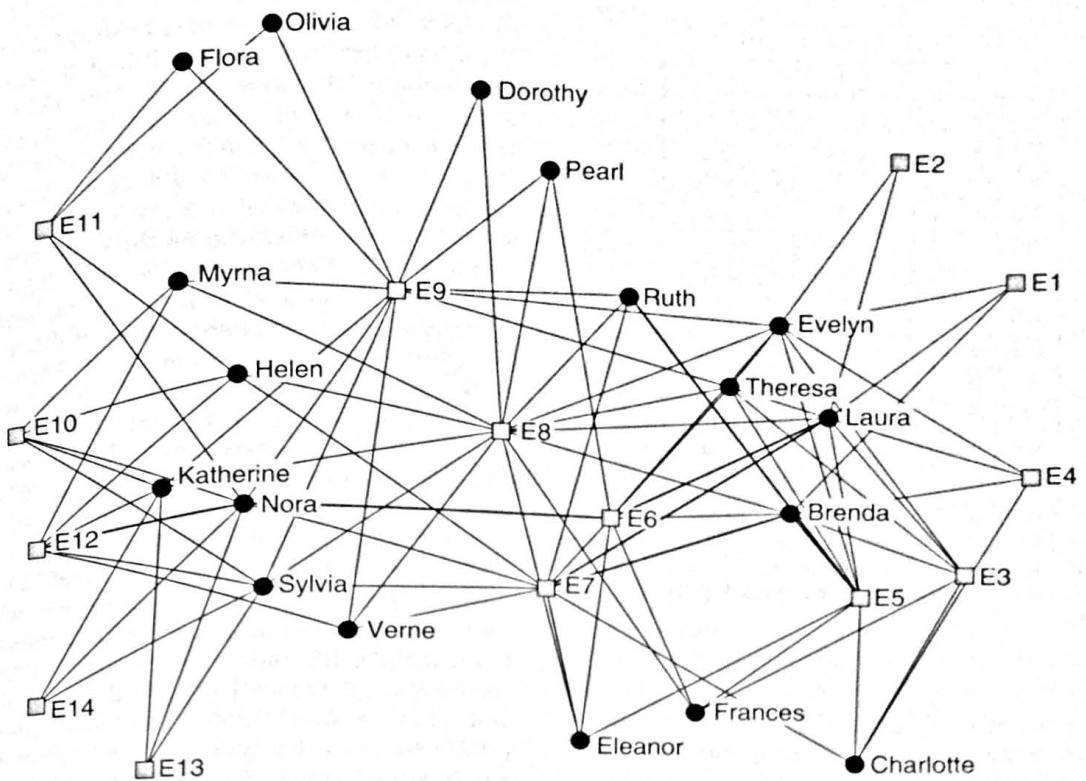


Figure 28.2 DGG women-by-events graph

	I1	I3	W1	W2	W3	W4	W5	W6	W7	W8	W9	S1	S2	S4
I1	0	0	0	0	1	0	0	0	0	0	0	0	0	0
I3	0	0	0	0	0	0	0	0	0	0	0	0	0	0
W1	0	0	0	0	1	1	0	0	0	0	0	1	0	0
W2	0	0	0	0	0	0	0	0	0	0	0	0	0	0
W3	1	0	1	0	0	1	0	0	0	0	0	1	0	0
W4	0	0	1	0	1	0	0	0	0	0	0	1	0	0
W5	0	0	0	0	0	0	0	0	0	0	0	0	0	0
W6	0	0	0	0	0	0	0	0	0	0	0	0	0	0
W7	0	0	0	0	0	0	0	0	0	1	1	1	0	0
W8	0	0	0	0	0	0	0	0	0	1	0	1	0	1
W9	0	0	0	0	0	0	0	0	0	1	1	0	0	1
S1	0	0	1	0	1	1	0	0	1	0	0	0	0	0
S2	0	0	0	0	0	0	0	0	0	0	0	0	0	0
S4	0	0	0	0	0	0	0	0	0	1	1	0	0	0

Figure 28.3 One-mode person-by-person positive relationship matrix

	Gr1	Gr 2	Gr 3	Gr4	Gr 5
I1	1	0	0	0	0
I3	0	0	0	0	0
W1	1	1	1	0	0
W2	1	1	0	0	0
W3	1	1	1	0	0
W4	1	1	1	0	0
W5	0	0	1	0	0
W6	0	0	0	1	0
W7	0	0	0	1	1
W8	0	0	0	1	1
W9	0	0	0	1	1
S1	0	1	1	0	0
S2	0	0	0	0	0
S4	0	0	0	0	1

Figure 28.4 Two-mode person-by-group matrix

some kind of participation or membership, as in people in events, projects, or groups.³ In this chapter we focus on affiliation data, but the techniques we discuss apply to two-mode data in general.

CO-AFFILIATION

In some cases, the purpose of collecting affiliation data is not to understand the pattern of ties between the two sets but to understand the pattern

of ties within one of the sets. It would seem perverse, in that case, to collect affiliation data, since by definition affiliation data do not include ties among members of either set. However, given affiliation data, we can in fact construct some kind of tie among members of a node set simply by defining *co-affiliation* (e.g., attendance at the same events, membership on the same corporate board) as a tie. For example, for the DGG dataset, we can construct a woman-by-woman matrix S , in which s_{ij} gives the number of events that woman i and woman j attended together (see Figure 28.5). If we like, we can then dichotomize so that there is a tie between two women if and only if they co-attended at least some number of events. Thus, affiliation data give rise to co-affiliation data, which constitute a kind of tie among nodes within a set.

One justification for relying on co-affiliation is the idea that co-affiliation provides the conditions for the development of social ties of various kinds. For example, the more often people attend the same events, the more likely it is they will interact and develop some kind of relationship. Feld (1981) suggests that individuals whose activities are organized around the same focus (e.g., voluntary organization, workplaces, hangouts, family, etc.) frequently become interpersonally connected over time. Physical proximity (which is simply co-affiliation with respect to spatial coordinates) is also clearly a major factor in enabling and, in the breach, preventing interaction (Allen, 1977). Another justification is almost the reverse of the first, namely that common affiliations can be the consequence of having a tie. For example, married couples attend a great number of events

	EVE	LAU	THE	BRE	CHA	FRA	ELE	PEA	RUT	VER	MYR	KAT	SYL	NOR	HEL	DOR	OLO	FLO
EVELYN	8	6	7	6	3	4	3	3	3	2	2	2	2	2	1	2	1	1
LAURA	6	7	6	6	3	4	4	2	3	2	1	1	2	2	2	1	0	0
THERESA	7	6	8	6	4	4	4	3	4	3	2	2	3	3	2	2	1	1
BRENDA	6	6	6	7	4	4	4	2	3	2	1	1	2	2	2	1	0	0
CHARLOTTE	3	3	4	4	4	2	2	0	2	1	0	0	1	1	1	0	0	0
FRANCES	4	4	4	4	2	4	3	2	2	1	1	1	1	1	1	0	0	0
ELEANOR	3	4	4	4	2	3	4	2	3	2	1	1	1	1	1	1	0	0
PEARL	3	2	3	2	0	2	2	3	2	2	2	2	2	2	2	1	0	0
RUTH	3	3	4	3	2	2	3	2	4	3	2	2	3	2	2	1	2	1
VERNE	2	2	3	2	1	1	2	2	3	4	3	3	4	3	3	2	1	1
MYRNA	2	1	2	1	0	1	1	2	2	3	4	4	4	3	3	2	1	1
KATHERINE	2	1	2	1	0	1	1	2	2	3	4	6	6	5	3	2	1	1
SYLVIA	2	2	3	2	1	1	2	2	3	4	4	6	7	6	4	2	1	1
NORA	2	2	3	2	1	1	2	2	2	3	3	5	6	8	4	1	2	2
HELEN	1	2	2	2	1	1	2	1	2	3	3	3	4	4	5	1	1	1
DOROTHY	2	1	2	1	0	1	1	2	2	2	2	2	1	1	2	1	1	1
OLIVIA	1	0	1	0	0	0	0	1	1	1	1	1	1	2	1	1	2	2
FLORA	1	0	1	0	0	0	0	1	1	1	1	1	1	2	1	1	2	2

Figure 28.5 DGG woman-by-woman matrix of overlaps across events

together and belong to a great number of groups together, and indeed they may come to share a great number of activities, interests, and beliefs. Thus, co-affiliation can be viewed as an observable manifestation of a social relation that is perhaps unobservable directly (such as feelings).

If either of these justifications is valid, then we may collect affiliation data simply because it is more convenient than collecting direct ties among a set of nodes. For example, if we are interested in studying relationships among celebrities, we could try to interview them about their ties with other celebrities, but this could be quite difficult to arrange. It would be easier to simply read celebrity news and record who has attended what Hollywood social event or who has worked on what project.

In deciding whether to use affiliation data as a proxy for social relations, it is useful to think about the conditions under which any of these justifications is likely to prove valid. One consideration is the size of affiliation events. For example, suppose we have a person-by-club matrix indicating who is a member of which club. If the clubs are small (like a board of directors), then our justifications seem, well, justifiable. But if the clubs are large (on the order of thousands of members), co-membership may indicate very little about the social tie between a given pair of members. Two people can be members of all the same (large) clubs or attend all the same (large) events, and yet they may not even be aware of each other's existence and never even meet.

It should also be noted that in adopting co-affiliations as a proxy for social ties, we confound the concept of social proximity with that of social similarity, which in other contexts are treated as competing alternatives (Burt, 1987; Friedkin, 1984). To see that co-affiliations are similarity data, consider the woman-by-woman co-affiliation network in Figure 28.5, constructed from the original two-mode woman-by-event attendance data. For each pair of women, we look at their respective rows in X , and count the number of times that they have 1s in the same places. This is simply an unnormalized measure of similarity of rows. In effect, for any pair of women we construct a simple 2-by-2 contingency table as shown in Figure 28.6 that shows the relationship between their pair of rows.

		Woman j		$a + b$
		1	0	
Woman i	1	a	b	
	0	c	d	$c + d$
		$a + c$	$b + d$	n

Figure 28.6 Contingency table

The quantity a gives the number of times that the pair of women co-attended an event. The quantity $a + b$ gives the total number of events that woman i attended, and $a + c$ gives the corresponding value for woman j . The quantity n is simply the number of events—the number of columns in matrix X . A simple way to bound a between 0 and 1 and promote comparability across datasets is to simply divide a by n , as shown in Equation 28.1.

$$a^* = \frac{a}{n} \quad (28.1)$$

Bounding a by the maximum possible score introduces the notion of other normalizations that take into account characteristics of the women, such as the number of events they attended. For example, if woman i and woman j attend three events in common, and woman k and woman l do as well, we would likely regard the two pairs as equally close. But if we knew that i and j each only attended 3 events, whereas k and l each attended 14 events, we would be more likely to conclude that the 100 percent overlap between i and j signals greater closeness than the 21 percent overlap between k and l .

Therefore, if we wanted to normalize the quantity a for the number of events that each woman attended, we might divide a by the minimum of $a + b$ and $a + c$, as shown in Equation 28.2. The resulting coefficient runs between 0 and 1, where 1 indicates the maximum possible overlap given the number of events attended by i and j . This approach takes into account that the number of overlaps between two women cannot exceed the number of events that either attended.

$$a_{ij}^* = \frac{a}{\min(a+b, a+c)} \quad (28.2)$$

Another well-known approach to normalizing a is provided by the Jaccard coefficient, which is described in Equation 28.3. It gives the number of events attended in common as a proportion of events that are “attendable,” as determined by the fact that at least one of the two women attended the event.

$$a_{ij}^* = \frac{a}{a+b+c} \quad (28.3)$$

Alternatively, we could take $a + d$ as a raw measure of social closeness. By including d , we effectively argue that choosing not to attend a

given event is as much of a statement of social allegiance as attending an event. A well-known normalization of $a + d$ is given by Equation 28.4, which is equal to the simple Pearson correlation between rows i and j of matrix X .

$$r_{ij} = \frac{1}{m} \sum_k x_{ik} x_{jk} - \bar{x}_i \bar{x}_j \quad (28.4)$$

Another approach, devised specifically for affiliation data, is provided by Bonacich (1972), who proposes normalizing the co-occurrence matrix according to Equation 28.5. Effectively, this measure gives the extent to which the overlap observed between i and j exceeds the amount of overlap we would expect by chance, given the number of events that i and j each attended.

$$a_{ij}^* = \frac{a - \sqrt{adbc}}{ad - bc}, \quad \text{for } ad \neq bc \quad (28.5)$$

All of these normalizations essentially shift the nature of co-affiliation data from frequencies of co-occurrences to tendencies or revealed preferences to co-occur. If we interpret frequencies of co-occurrences as giving the number of opportunities for interaction or flow of information or goods, then the raw, unnormalized measures

are the appropriate indices for measuring co-affiliation. In contrast, if the reason for studying co-affiliations is that co-affiliations reveal otherwise unseen relationships between people (e.g., sociometric preferences), the normalized measures are the most appropriate, as they essentially give us the tendency or preference for a pair of actors to co-occur while controlling for nuisance variables such as the number of times an actor was observed. The normalized measures tell us how often two actors are co-attending relative to the number of times they could have.

Consider the following hypothetical research project. Say that we are interested in analyzing connections between a group of 13 individuals based on their memberships in different social clubs (16 of them). Because we are interested in understanding relationships between the 13 individuals we convert the affiliation data (person-by-social club) into co-affiliations (person-by-person). We construct both a raw unnormalized co-affiliation matrix and a normalized co-affiliation matrix. Figure 28.7 is a graphical representation of the raw co-affiliation network using a standard graph layout algorithm. Individuals are labeled a through m . A line connecting two individuals indicates that they are members of at least two of the same social clubs. Node size varies by the number of social clubs that each individual is a member of; thus the larger the node, the more socially active the individual. Figure 28.8 is a depiction of Jaccard coefficients for each pair of individuals, such that a line connecting two individuals

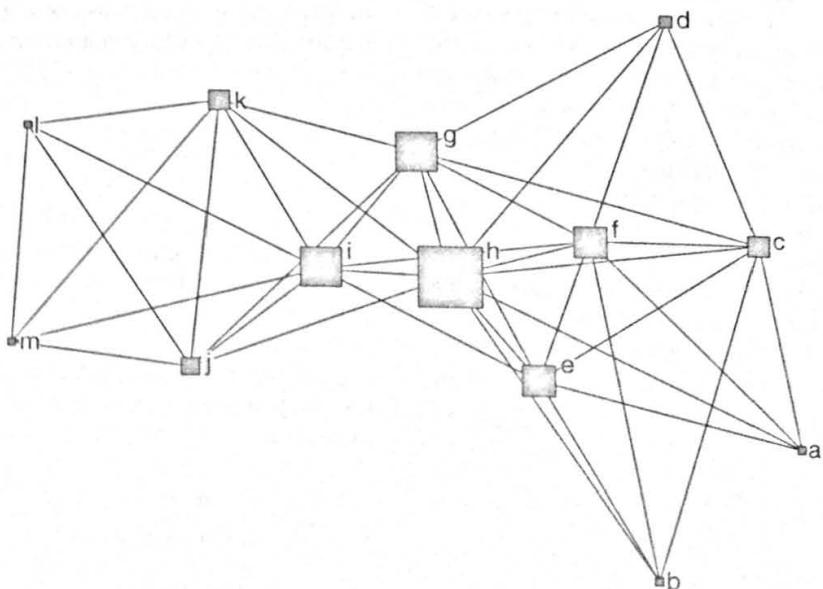


Figure 28.7 Co-membership in two or more social clubs. Node size is based on the number of social clubs that each individual is a member of

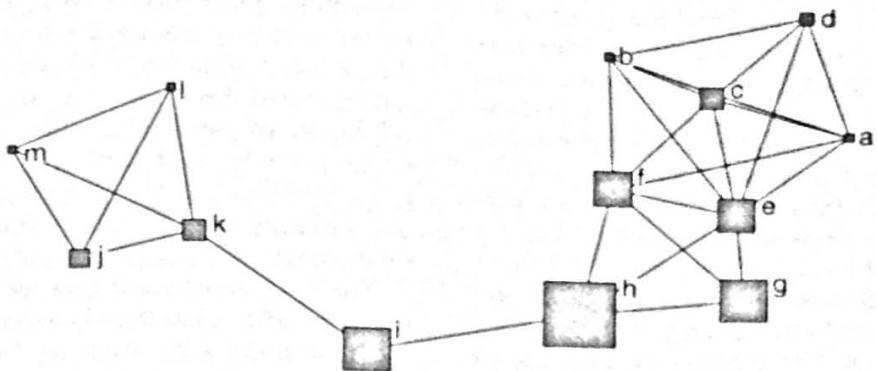


Figure 28.8 Spring embedding of Jaccard coefficients. An edge is shown if $c_{ij} > 0.38$. Node size is based on the number of social clubs that each individual is a member of

indicates that their social club membership profiles are correlated at greater than 0.38.

The raw co-affiliation network (presented in Figure 28.7) can be described as a core-periphery structure in that there is a set of core individuals who are members of multiple social clubs (persons *e*, *f*, *g*, *h*, *i*) and are surrounded by a collection of less connected individuals. We see that there are opportunities for interaction between many of the 13 individuals. However, the high social activity of the core individuals places them in the middle of the graph, which tends to obscure any subgrouping structure. Now consider the Jaccard similarity network (presented in Figure 28.8). This graph effectively highlights that there are two groupings of individuals with different membership profiles. The graph also effectively reveals the bridging role of individual *i*, which was not at all discernable when visualizing unnormalized co-occurrences among the individuals (see Figure 28.7).

Another kind of normalization worth mentioning has to do with the size of the events (or social clubs) that the individuals are affiliated with. If, in analyzing co-affiliation data, we are taking the point of view that greater co-affiliation creates more opportunities for social ties to develop, then when measuring person-to-person co-affiliations, we would probably want to take into account the relative sizes of different events. In the DGG data, for example, if two women co-attend an event that included just five people in total, it would seem that the likelihood of being aware of each other, of meeting, and indeed of changing their relationship is reasonably high. We would want to give that event a lot of weight. On the other hand if the same women co-attend an event in which thousands are present (such as a concert), we might want to weight that very little. An obvious approach, then, is to weight events inversely by

their size. Thus, referring back to Figure 28.6, the quantity n becomes the sum of weights of all events, and the quantity a is the sum of weights of the events that were co-attended by *i* and *j*. The measures described by Equations 28.1 to 28.4 can then be computed without modification.

Table 28.1 summarizes which normalization approaches are appropriate given one's attitude toward the nature of the co-affiliation data. For convenience, it is assumed that the two-mode affiliation data are actor-by-event and that we are interested in constructing the actor-by-actor co-affiliation matrix. As such, we refer to the actors/rows as "variables" and the events/columns as "cases." Therefore, the first kind of normalization discussed above can be referred to as "variable normalization" and the second as "case normalization."

Analysis of co-affiliation

Having constructed a co-affiliation matrix, we would typically want to analyze the data using all the tools of social network analysis—as with any other kind of tie. For the most part, this is unproblematic, aside from the caveats already voiced. The main issue we typically encounter is that the

Table 28.1 Appropriate normalizations by view of data

<i>Co-affiliation as opportunity</i>	<i>Co-affiliation as indicator</i>
<ul style="list-style-type: none"> No normalization (simple overlap counts) Case normalization (e.g., weighting inversely by event sizes) 	<ul style="list-style-type: none"> Variable normalization (e.g., Jaccard or Pearson correlations)

co-affiliation matrix is valued and many network-analytic techniques assume binary data – particularly those techniques with graph-theoretic roots. In those cases, the data will need to be dichotomized. Since the level of dichotomization is arbitrary, the normal procedure is to dichotomize at different levels and obtain measures for networks constructed with different thresholds for what is considered a tie. In other cases, there will be no need for dichotomization. For example, eigenvector centrality (Bonacich, 1972) and beta centrality (Bonacich, 1987, 2007) are quite happy to accept valued data, particularly when the values are “positive” in the sense that larger values can be interpreted as enhancing flows or coordination. Other centrality measures need to be modified to work with valued data. In general, measures based on lengths of paths, such as betweenness and closeness centrality, can easily be modified to handle valued data, provided the data can be sensibly transformed into distances or costs (Brandes, 2001). For example, the number of events co-attended by two women can be subtracted from the number of events in total and then submitted to a valued betweenness analysis.

Another possible difficulty with co-affiliation data is that, being similarity metrics, they tend to have certain mathematical properties that social network data may not. For example, most similarity metrics are symmetric so that $s(u,v) = s(v,u)$. We can construct nonsymmetric similarity measures, but these are rarely used and none of the ones we consider above are nonsymmetric. Similarity matrices such as Pearson correlation matrices have numerous other properties as well, such as being positive semi-definite (e.g., all eigenvalues are nonnegative). The main consequence is that the norms or baseline expectations for network measures on co-affiliation data should not be based on norms or expectations developed for sociometric data in general (cf. Wang et al., 2009).

At this point, we leave the discussion of co-affiliation data and focus entirely on visualizing and analyzing affiliation graphs directly without converting to co-affiliations.

DIRECT VISUALIZATION OF AFFILIATION GRAPHS

Affiliation graphs are typically visualized using the same graph layout algorithms used for ordinary graphs. In principle, certain algorithms, such as spring embedders or multidimensional scaling of path distances, should be less than optimal when applied to bipartite graphs because these

algorithms place nodes in space such that distances between them are loosely proportional to the path distances that separate them. Since nodes belonging to the same node set are necessarily a minimum of two links apart, we might expect some difficulty in detecting grouping in bipartite graphs. In practice, however, this is not a problem and ordinary graph layout algorithms work well on bipartite graphs.

The only adjustment that we typically have to make for affiliation data is to visually distinguish the two node sets, such as by using different colors and shapes for node symbols of different sets. For example, Figure 28.2 shows a visualization of the DGG dataset using the spring embedding procedure in NetDraw (Borgatti, 2002). Women are represented by circles, and events are represented by squares. In the figure, we can see a group of women on the far right together with a group of events (E1 through E5) that only they attend. On the left, one can see another group of women who also have their exclusive events (E10 through E14). In the middle of the figure are four events (E6 through E9) that are attended by both groups of women. The figure also makes clear that Olivia and Flora are a bit separate from the rest of the network and are structurally similar because they attended exactly the same events.

Another approach is to use a two-mode multivariate analysis technique such as correspondence analysis to locate nodes. Correspondence analysis delivers a map in which points corresponding to both the n rows and m columns of an n -by- m two-mode matrix are represented in a joint space. Computationally, correspondence analysis consists of a double-normalization of the data matrix to reduce the influence of variation in the row and column sums, followed by a singular value decomposition. The result is that, in the case of a woman-by-event matrix, two women will be placed near each other to the extent that they have similar event profiles, controlling for the sizes of the events, and two events will be near each other if they tend to have similar attendee profiles, controlling for the overall participation rates of the attendees. In the case of the DGG dataset, correspondence analysis gives the diagram shown in Figure 28.9. As a general rule, the advantage of correspondence representations is that, in principle, the map distances are meaningful and can be related precisely back to the input data. This is not the case with most graph layout algorithms, as they respond to multiple criteria such as avoiding the placement of nodes directly on top of each other or keeping line lengths approximately equal. The disadvantage of correspondence analysis layouts is that they can be less readable. For example, in Figure 28.9, Olivia is obscured by Flora, and the (accurate) portrayal of exactly how

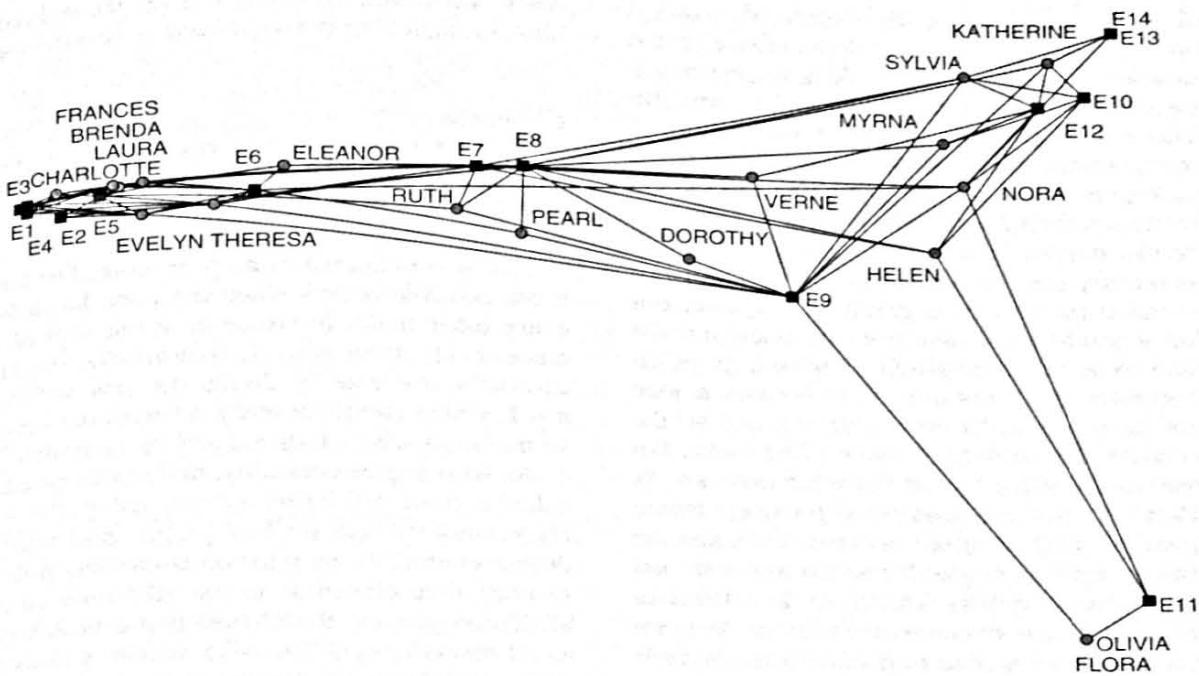


Figure 28.9 Correspondence analysis of two-mode DGG matrix

different Flora, Olivia, and Event 11 are from the rest makes the majority of the display very hard to read.

Direct analysis of affiliation graphs

There are several different approaches to analyzing affiliation data without converting to co-affiliations. Because affiliation graphs are graphs, an obvious approach is to simply use all the standard algorithms and techniques in the network analysis toolkit that apply to graphs in general. In doing this, we either effectively assume that the special nature of affiliation graphs will not affect the techniques, or we can pretend that ties within node sets could have occurred and just didn't. This approach works for a small class of methods, but by no means for all. A case where it does not work is measuring transitivity: calculating transitivity fails because transitive triples are impossible in bipartite graphs (all ties are between node sets, which means that if $a \rightarrow b$ and $b \rightarrow c$ then a and c must be members of the same class, and therefore cannot be tied, making transitivity impossible).

An alternative approach is to develop new metrics and algorithms specifically designed for the bipartite case (affiliation graphs), taking into account the fact that the observed network is not

just bipartite by happenstance but is by design – similar to the concept of structural zeros in log-linear modeling. This sounds like a great deal more work, but in practice it is often possible to adjust metrics designed for general graphs by simply applying an appropriate post hoc normalization. This is the strategy we shall take in applying centrality metrics to affiliation data. In other cases, a wholly different approach must be constructed. For example, for the case of measuring transitivity, we might redefine transitivity in terms of quadruples, such that a quad is called transitive if $a \rightarrow b$, $b \rightarrow c$, $c \rightarrow d$, and $a \rightarrow d$.

Centrality

As discussed elsewhere in this book (cf. Hanneman and Riddle's chapter), centrality refers to a family of properties of node positions. A number of centrality concepts have been developed, together with their associated measures (Borgatti and Everett, 2006). In this section, we consider the measurement of four well-known centrality concepts.

Degree

In ordinary graphs, degree centrality, d_i , is defined as the number of ties incident upon a node i . In the

affiliation case, of course, the degree of a node is the number of ties it has with members of the other node set. So in the DGG data, for women, it is the number of events they attended, and for events, it is the number of women who attended. If we represent affiliations as a bipartite graph, we can compute degree centrality as usual and obtain perfectly interpretable values, at least with respect to the raw counts. However, it is usual to normalize centrality measures by dividing by the maximum value possible in a graph of that size. For ordinary graphs, this value is $n - 1$, where n is the number of nodes in the graph. However, for affiliation graphs, this is not quite right because a node cannot have ties to its own node set, and so the value of $n - 1$ cannot be attained.⁴ The maximum degree is always the size of the other node set. In the DGG dataset, the maximum possible degree for a woman is the number of events (14), and the maximum possible degree for an event is the total number of women (18). Therefore, to normalize degree centrality in the case of affiliation data, we must apply two separate normalizations depending on which node set a node belongs to, as shown in Equation 28.6.

$$\begin{aligned} d_i^* &= \frac{d_i}{n_2}, \quad \text{for } i \in V_1 \\ d_j^* &= \frac{d_j}{n_1}, \quad \text{for } j \in V_2 \end{aligned} \quad (28.6)$$

The key benefit of normalizing degree centrality in this way is that we can not only assess the relative centrality of two women or two events but also whether a given woman is more central than a given event. Without such normalization, nodes with equal propensities to have ties could only have equal degrees if the node sets were the same sizes. However, while normalization handles the mathematical issues of comparability, the substantive interpretation of a woman's centrality relative to an event's is still an issue, and depends on the details of the research setting. For example, it may be that the events are open to all, and ties in the affiliation graph reflect only a woman's agency in choosing which events to attend. In this case, if a woman has greater degree than a given event, we might say that her gregariousness is greater than the event's attractiveness, although this implies that the degree centrality measurement does not measure the same thing for women as for events, which runs counter to the basic idea in the direct analysis of affiliation graphs. On the other hand, the events might be by invitation only, in which case both women and events have a kind of agency. In general, centrality measures in this context have the most straightforward interpretations

when the affiliations result from some kind of bilateral matching process, such as speed dating.

Closeness

In ordinary graphs, closeness centrality, c_i , refers to the sum of geodesic distances from node i to all $n - 1$ others in the network. As such, it is an inverse measure of centrality in which greater centrality is indicated by a lower score. The lowest score possible occurs when the node has a tie to every other node, in which case the sum of distances to all others is $n - 1$. To normalize closeness centrality, we usually divide the raw score into $n - 1$, which simultaneously reverses the measure so that high scores indicate greater centrality.⁵

As with degree centrality, raw closeness can be calculated in affiliation graphs using the same algorithms we use for any graph. But, also like degree centrality, we must do something different to normalize closeness in the affiliation case. In affiliation graphs, the closest that a node can be to all others is $n_2 + 2(n_1 - 1)$, which is distance 1 from all nodes in the other node set and distance 2 from all other nodes in its own set. Therefore, to normalize (and simultaneously reverse) closeness in the bipartite case, we divide the raw closeness of a node in V_1 into $n_2 + 2(n_1 - 1)$ and the raw closeness of a node in V_2 into $n_1 + 2(n_2 - 1)$, as shown in Equation 28.7 in which c_i represents raw closeness centrality, and n_1 and n_2 represent the number of nodes in each node set.

$$\begin{aligned} c_i^* &= \frac{n_2 + 2(n_1 - 1)}{c_i}, \quad \text{for } i \in V_1 \\ c_j^* &= \frac{n_1 + 2(n_2 - 1)}{c_j}, \quad \text{for } j \in V_2 \end{aligned} \quad (28.7)$$

Using the DGG dataset for illustration, we can see that the maximum number of nodes that can be distance 1 from a woman is 14 (since there are 14 events), and the maximum number of nodes that can be distance 2 from any of the 18 women is 17 (since there are 18 women). Thus, the theoretical minimum value of closeness centrality for a woman is $14 + 2 * (18 - 1)$, and the theoretical minimum value for an event is $18 + 2 * (14 - 1)$.

Betweenness

In any graph, betweenness centrality, b_i , refers to the "share" of shortest paths in a network that pass through a node i , as given by Equation 28.8.

$$b_k = \frac{1}{2} \sum_{i \neq k}^n \sum_{j \neq k, i}^n \frac{s_{ikj}}{s_{ij}} \quad (28.8)$$

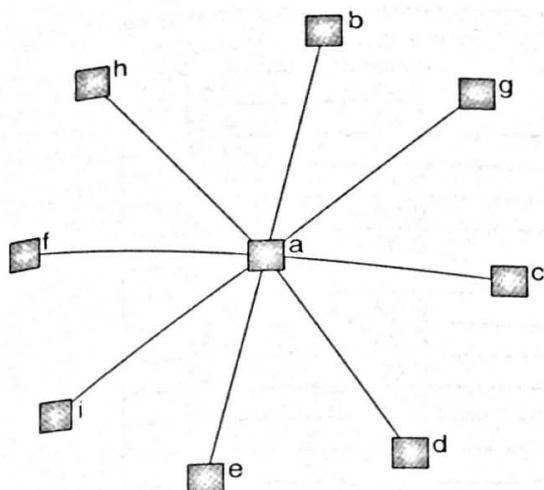


Figure 28.10 Star-shaped network

To normalize betweenness, we divide by the maximum possible value, which in the case of an ordinary graph is achieved by the center of a star-shaped network, as shown in Figure 28.10.

In the bipartite case, unless one node set contains just one node, an affiliation graph cannot attain that level of centralization. As a result, the maximum possible betweenness for any node in a bipartite graph is limited by the relative size of the two node sets, as given by Equation 28.9. To normalize betweenness, we simply divide b_i by either $b_{V_1 \max}$ or $b_{V_2 \max}$ (see Equation 28.9), depending on whether node i belongs to node set V_1 or V_2 .

$$\begin{aligned} b_{V_1 \max} &= \frac{1}{2} [n_2^2(s+1)^2 + n_2(s+1)(2t-s-1) \\ &\quad - t(2s-t+3)] \\ s &= (n_1 - 1) \text{ div } n_2, t = (n_1 - 1) \text{ mod } n_2 \\ b_{V_2 \max} &= \frac{1}{2} [n_1^2(p+1)^2 + n_1(p+1)(2r-p-1) \\ &\quad - r(2p-r+3)] \\ p &= (n_2 - 1) \text{ div } n_1, r = (n_2 - 1) \text{ mod } n_1 \end{aligned} \quad (28.9)$$

Eigenvector

Eigenvector centrality, e_i , is defined as the principal eigenvector of the adjacency matrix of a graph (Bonacich, 1972), as defined by Equation 28.10. In eigenvector centrality, a node's score is proportional to the sum of the scores of its neighbors. In a bipartite graph such as the one created by DGG, this means a woman's centrality will be proportional to the sum of centralities of the events she attends, and similarly the centrality of an event will be proportional to the sum of

centralities of the women who attend it. As a result, eigenvector centrality applied to the adjacency matrix of an affiliation graph is conceptually and mathematically identical to singular value decomposition (Eckhardt and Young, 1936) of the two-mode incidence matrix.⁶ In addition, both of these are equivalent to an eigenvector analysis of the simple co-affiliation matrix.

$$e_i = \lambda \sum_j a_{ij} e_j \quad (28.10)$$

where λ is the principal eigenvalue of A .

Empirical illustration of centrality measures

As an illustration, Figure 28.11 presents normalized centrality scores for all four types of centrality discussed above for the DGG bipartite graph presented in Figure 28.2. Note that three events (E8, E9, and E7) are more central than any of the women on all of the measures except for normalized degree centrality. It is also worth highlighting that E7 has 10 ties while Nora has only 8, but Nora has a slightly higher normalized degree centrality because there are fewer events than women, so here 8 represents a greater percentage of the possible ties.

COHESIVE SUBGROUPS

Cohesive subgroups refer to dense areas in a network that typically have more ties within a group than with the rest of the network. Affiliation data pose special problems for cohesive subgroup analysis because the area around any given node can never be very dense since none of a node's "friends" can be friends with each other. As a result, some traditional graph-theoretic methods of finding subgroups need to be modified for the bipartite case.

One of the most fundamental subgroup concepts is that of a clique (Luce and Perry, 1949). A clique is defined as a maximally complete subgraph, which means that every member of the clique has a tie to every other (a property known as completeness), and there is no other node that could be added to the subgraph's set of vertices without violating the completeness requirement (this is the property of maximality). Cliques of large size are rare in ordinary graphs, and they are impossible in bipartite graphs. As a result, applying ordinary clique algorithms to affiliation graphs is not useful.

One solution is to use the N-clique concept, which is a relaxation of the clique idea. In an N-clique, we do not require each member of the clique to have a direct tie with every other, but

	Node ID	Normalized degree	Normalized clustering	Normalized betweenness	Normalized mechanical
E8	14	0.78	0.85	0.24	0.51
E9	12	0.67	0.79	0.23	0.38
E7	10	0.56	0.73	0.13	0.38
Nora	8	0.57	0.80	0.11	0.26
Evelyn	8	0.57	0.80	0.10	0.33
Theresa	8	0.57	0.80	0.09	0.37
E6	8	0.44	0.69	0.07	0.33
Sylvia	7	0.50	0.77	0.07	0.28
Laura	7	0.50	0.73	0.05	0.31
Brenda	7	0.50	0.73	0.05	0.31
Katherine	6	0.43	0.73	0.05	0.22
E5	8	0.44	0.59	0.04	0.32
Helen	5	0.36	0.73	0.04	0.20
E3	6	0.33	0.56	0.02	0.25
Ruth	4	0.29	0.71	0.02	0.24
Verne	4	0.29	0.71	0.02	0.22
E12	6	0.33	0.56	0.02	0.20
Myrna	4	0.29	0.69	0.02	0.19
E11	4	0.22	0.54	0.02	0.09
Eleanor	4	0.29	0.67	0.01	0.23
Frances	4	0.29	0.67	0.01	0.21
Pearl	3	0.21	0.67	0.01	0.18
E4	4	0.22	0.54	0.01	0.18
Charlotte	4	0.29	0.60	0.01	0.17
E10	5	0.28	0.55	0.01	0.17
Olivia	2	0.14	0.59	0.01	0.07
Flora	2	0.14	0.59	0.01	0.07
E2	3	0.17	0.52	0.00	0.15
E1	3	0.17	0.52	0.00	0.14
Dorothy	2	0.14	0.65	0.00	0.13
E13	3	0.17	0.52	0.00	0.11
E14	3	0.17	0.52	0.00	0.11

Figure 28.11 Normalized centrality scores for the DGG affiliation graph

instead that the member be no more than distance n from every other. Choosing $n = 2$ gives us subgroups in which all pairs of members are within two links of each other. Applied to an ordinary graph, this yields subgroups that are “looser” than ordinary cliques, meaning that they are less than 100 percent dense. However, when applied to an affiliation graph, a two-clique can be regarded as complete, since all possible ties are present, due to the constraints of bipartite graphs. For this reason, Borgatti and Everett (1997) give two-cliques in affiliation graphs a name of their own, the *bi-clique*. Effectively, a bi-clique is to affiliation graphs what a clique is to ordinary graphs.

Because bi-cliques can be numerous and overlapping, it is often useful to perform a secondary analysis by constructing a node-by-clique matrix and by correlating the profiles of each node across

bi-cliques so that nodes that are members of many of the same bi-cliques will be given a high correlation. This correlation matrix can then be treated as a valued adjacency matrix and visualized using standard graph layout algorithms. Figure 28.12 shows the result of such an analysis. The results are striking in the way they differentiate between two groups of women tied to two distinct groups of events. In addition, the diagram clearly shows the separation of Flora and Olivia, and the bridging position of Ruth.

Structural equivalence

Structural equivalence refers to the extent that pairs of nodes have ties to the same third parties. In affiliation graphs such as the DGG dataset,

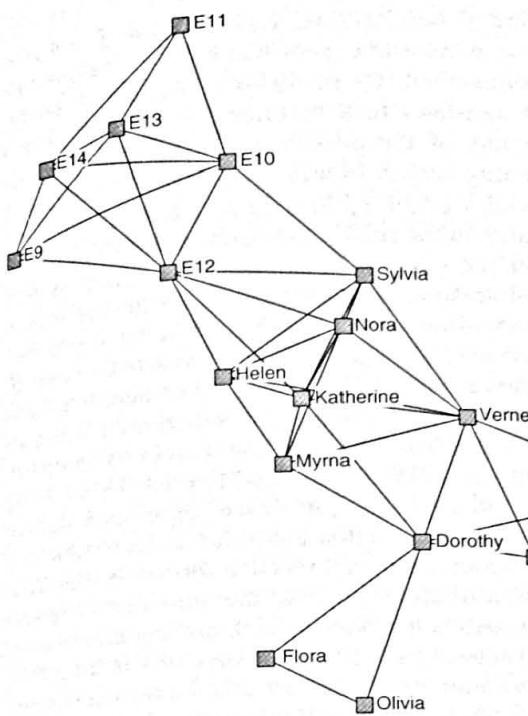


Figure 28.12 A tie indicates that the correlation between two nodes is greater than 0.60

actors are structurally equivalent to the extent that they attend the same events, and events are structurally equivalent to the extent that they are attended by the same actors. Strictly speaking, in affiliation graphs there can be no equivalence between nodes of different node sets, because they cannot have any nodes in common. As a result, structural equivalence analyses of affiliation graphs are virtually identical to analyses of the actor-by-actor and event-by-event co-affiliation matrices. For example, a standard approach to measuring structural equivalence in ordinary graphs is to correlate the rows (and columns) of the adjacency matrix, and then do a hierarchical cluster analysis of the correlation matrix to identify blocks of approximately equivalent nodes. If we take this approach to the $(n + m)$ by $(n + m)$ adjacency matrix of an affiliation graph, we are virtually guaranteed to find the two modes of the affiliations dataset as the dominant partition in the hierarchical clustering. The next partition will then split one of the two node sets, and so on. In the end, the results are essentially the same as if we had simply clustered each of the co-affiliation matrices separately.

An alternative approach to structural equivalence is blockmodeling (White et al., 1976). In ordinary graphs, blockmodeling refers to partitioning the rows and columns of the adjacency matrix in order for those corresponding to nearly

equivalent nodes to be placed in the same classes, as shown in Figure 28.13. Partitioning the rows and columns based on structural equivalence has the effect of partitioning the cells of the adjacency matrix into matrix blocks that have a characteristic pattern of homogeneity: either all of the cells in the block are 1s (called 1-blocks), or they all 0s (called 0-blocks). The job of a blockmodeling algorithm is to find a partitioning of the rows and columns that makes each matrix block as homogeneous as possible (Borgatti and Everett, 1992).

	A1	A2	A3	B1	B2	B3	B4	C1	C2	C3
A1	0	0	0	1	1	1	1	0	0	0
A2	0	0	0	1	1	1	1	0	0	0
A3	0	0	0	1	1	1	1	0	0	0
B1	1	1	1	0	0	0	0	1	1	1
B2	1	1	1	0	0	0	0	1	1	1
B3	1	1	1	0	0	0	0	1	1	1
B4	1	1	1	0	0	0	0	1	1	1
C1	1	1	1	1	1	1	1	0	0	0
C2	1	1	1	1	1	1	1	0	0	0
C3	1	1	1	1	1	1	1	0	0	0

Figure 28.13 Structural equivalence blockmodeling in an ordinary adjacency matrix

Applying this approach directly to affiliation graphs would mean partitioning the rows and columns of the $(n + m)$ -by- $(n + m)$ bipartite adjacency matrix B . This can be done, but the bipartite structure imposes certain constraints. For example, matrix blocks involving within-mode ties (e.g., woman-to-woman, event-to-event) are necessarily 0-blocks. In addition, the best two-class partition will almost certainly be the mode partition (except in trivial cases), and in general, all other partitions will be refinements of the mode partition (i.e., they will be nested hierarchically within the mode partition).

A more elegant (and computationally efficient) approach is to work directly from the two-mode incidence matrix X (Borgatti and Everett, 1992). To do this, we redefine the concept of a blockmodel to refer to not one but two independent partitions: one for the rows and one for the columns. We then apply an algorithm to find the pair of partitions that yields the most homogeneous matrix blocks. In other words, a structural equivalence blockmodeling of the two-mode incidence matrix is one in which row nodes are in the same class if they have similar rows, and column nodes are in the same class if they have similar columns. An example involving four classes of rows and three classes of columns is shown in Figure 28.14.

REGULAR EQUIVALENCE

In ordinary graphs, the idea of regular equivalence is that a pair of equivalent nodes is connected not necessarily to the same nodes (as in structural equivalence) but to equivalent nodes (White and Reitz, 1983). In other words if nodes u and v are

	E1	E2	E3	F1	F2	F3	F4	G1	G2	G3
A1	1	1	1	1	1	1	1	0	0	0
A2	1	1	1	1	1	1	1	0	0	0
A3	1	1	1	1	1	1	1	0	0	0
B1	1	1	1	0	0	0	0	0	0	0
B2	1	1	1	0	0	0	0	0	0	0
B3	1	1	1	0	0	0	0	0	0	0
B4	1	1	1	0	0	0	0	0	0	0
C1	0	0	0	1	1	1	1	0	0	0
C2	0	0	0	1	1	1	1	0	0	0
C3	0	0	0	1	1	1	1	0	0	0
D1	0	0	0	1	1	1	1	1	1	1
D2	0	0	0	1	1	1	1	1	1	1

Figure 28.14 Two-mode structural equivalence blockmodel

perfectly regularly equivalent, then if u has a friend p , we can expect v to have a friend q who is equivalent to p . In blockmodeling terms, this translates to a partitioning of the rows and columns of the adjacency matrix such that the resulting matrix blocks are either 0-blocks, or a special kind of 1-block in which every row and column in the matrix block has at least one 1.

In the case of structural equivalence, it was possible to apply the concept to the adjacency matrix of an affiliation graph, making it possible to use existing algorithms/programs to compute it. In the case of regular equivalence, there is a complication. Regular equivalence defines a lattice of partitions that all have the regularity property (Borgatti, 1989; Borgatti and Everett, 1989). Most standard regular equivalence algorithms deliver the maximum regular equivalence. Unfortunately, with affiliation graphs, the maximum regular equivalence is always trivial, placing all nodes in the same class. There are ways of handling this, but a better approach is to redefine regular equivalence for two-mode incidence matrices, as developed by Borgatti and Everett (1992). As they did with structural equivalence, Borgatti and Everett (1992) redefine the concept of a blockmodel to refer to not one but two independent partitions: one for the rows and one for the columns. Regular equivalence implies that we can section the matrix into rectangular blocks such that each block is a 0-block or a regular 1-block. For example, if the affiliation graph indicates which consumers visit which restaurants, the two-mode regular blockmodel shown in Figure 28.15 identifies four different types of consumers that visit three kinds of restaurants. Consumers of the same type do not necessarily visit the same restaurants, but they do visit the same kinds of restaurants. Thus all

	R1	R2	R3	R4	R5	R6	R7	R8	R9	R10
C1	1	0	1	0	1	1	0	0	0	0
C2	0	0	1	0	0	1	0	0	0	0
C3	0	1	0	1	1	0	1	0	0	0
C4	1	0	0	0	0	0	0	0	1	1
C5	1	0	1	0	0	0	0	1	1	0
C6	0	1	0	0	0	0	0	0	1	1
C7	0	1	1	0	0	0	0	1	0	1
C8	0	0	0	0	1	1	0	0	0	0
C9	0	0	0	0	0	1	0	0	0	0
C10	0	0	0	1	1	0	1	0	0	0
C11	1	0	1	1	0	1	0	0	1	1
C12	0	1	0	0	1	0	1	1	0	1

Figure 28.15 A two-mode regular equivalence blockmodel

consumers in the first class visit the first two kinds of restaurants, while all consumers in the second class visit only the first and third kinds of restaurants.

TWO-MODE RELATIONAL ALGEBRAS

In social network analysis, the term "relational algebra" is typically used very loosely to refer to the composition of relations. For example, if we measure both friendship and "teacher of" relations among a set of nodes, we can construct new, compound relations that link the actors, such as "friend of a teacher of" or "teacher of a friend of," as well as "friend of a friend" and "teacher of a teacher of." If the relations are represented as adjacency matrices, the composition relation can be equated to Boolean matrix multiplication⁷ of the adjacency matrices, so that if F represents the friendship relation and T represents the teacher of relation, then the Boolean matrix product FT represents the "friend of a teacher of" relation. Since the result of a composition is just another relation, we can construct compositions of compositions, yielding a long string of Boolean matrix products. For example, the string $FTTF'$ gives a relation in which, if u is tied to v via this relation, it indicates that v is liked by a student of someone who is the teacher of a friend of u . (Note that the transpose T' is used to represent the inverse relation "is taught by.")

Relational composition is also possible with affiliation data, provided the incidence matrices are conformable. For example, suppose we have a binary person-by-organization matrix M , indicating which persons are members of which organizations. Suppose we also have an organization-by-event matrix S , which indicates which organizations were sponsors of which events. Finally, suppose we have a person-by-event matrix (A), indicating which person attended which event. The product MS is a new matrix in which $MS(u,v) > 0$ indicates that person u belongs to at least one organization that sponsored event v . In a given research setting, we might use MS to explain matrix A —that is, test the hypothesis that people are more likely to attend events that are sponsored by their organizations.

Relational algebras can incorporate a mix of affiliation and ordinary networks. For example, if we also had a matrix F , indicating which persons were friends with which others, we could generate compositions such as FMS , in which $FMS(u,v) > 0$ indicates that a person u has a friend who is a member of an organization that sponsors an event v . Krackhardt and Carley (1998) use compositions of this type in their PCANS

model, which relates persons, tasks, and resources to each other, including person-to-person communications and task-to-task dependencies. For example if matrix A indicates which person is assigned to which task, and matrix P indicates which task precedes another, then the product AP relates each person u to each task v , indicating whether person u has a task that precedes task v . The triple product APA' relates each person u to each person v , indicating whether person u has a task that precedes a task that person v does—that is, it indicates whether person v is dependent on person u to get work done.

CONCLUSION

In this chapter we provide an introduction to the analysis of affiliation data. Two basic approaches are discussed: a conversion approach and a direct approach. The conversion approach consists of analyzing co-affiliations or similarities among elements of one node set with respect to their profiles across the other node set. The similarities are then treated as ties among the nodes. Co-affiliations are frequently analyzed to identify opportunities for interaction (e.g., the flow of goods or information) or unseen relationships between people (e.g., sociometric preferences). The direct approach consists of analyzing both node sets simultaneously, treating the elements of each on an equal footing. As discussed, the direct approach often requires the use of new metrics and algorithms specifically designed for bipartite graphs.

Our survey has focused on analysis, and within that, measurement of network concepts such as centrality, cohesive subgroups, structural equivalence, and regular equivalence. In doing so, we have ignored statistical modeling, such as the nascent field of exponential random graph models for affiliation data (see Robins's chapter in this volume for a more detailed discussion).

We close with suggestions for future analyses of affiliation data. One element that is underexplored in affiliation work is the temporal dimension. There are two important ways in which time can be brought into affiliation analysis. First, there is the case of affiliation graphs changing over time. We can conceptualize this as a series of person-by-organization matrices representing different slices of time, or a single three-mode affiliation network in which each tie links together a person, an organization, and a time period. Many of the direct analysis techniques discussed in this chapter can be generalized to this three-mode case (Borgatti and Everett, 1992).

The other important case is in the analysis of two-mode person-by-event data, where the events are ordered by time. For example, if we study Hollywood film projects, we typically have a data matrix that is actor by film, and the films are ordered by release date (or start date, etc.). If we are interested in how actors' previous collaboration ties affect the quality of a film project they are jointly engaged in, we need to construct the collaboration network continuously over time, since we would not want to predict film success based on collaborations that occur after the film was produced. Social network analysis programs such as UCINET (Borgatti et al., 2002) are just beginning to include tools for these kinds of analyses.

Another example of time-ordered affiliation data occurs in the study of career trajectories. Taking the three-mode approach we can examine how actors' co-location (in terms of both organization and time) ties affect their future careers. Or we can look at how individuals flow from organization to organization along directed paths. Here, the organizations can be ordered in time differently for each individual, although a key research question is whether an underlying ordering of the organizations (such as status) creates consistency in individual career moves.

NOTES

1 The node labels indicate whether the individual is an inspector (I), a worker (W), or a supervisor (S).

2 The groups were constructed by the present authors for illustrative purposes, based on a clique analysis.

3 This is not to imply that the data must be binary as we could have data in which persons have a degree of membership or participation in various groups or events.

4 Except for nodes that are in the only members of their special case where one vertex set contains just one node set.

5 Of course, this is a nonlinear transformation, unlike all other centrality normalizations. To maintain consistency we could instead divide raw closeness by its maximum and simply remember that it is a reverse measure.

6 In addition, singular value decomposition yields the measures of hubs and authorities proposed by Kleinberg (1999). Therefore, in affiliation data, eigenvector centrality and hubs and authorities are identical concepts, which is not true in ordinary graphs.

7 Boolean multiplication is simply ordinary matrix multiplication in which the resulting matrix is

dichotomized so that any value greater than 0 is assigned a 1.

REFERENCES

- Allatta, J.T. (2003) 'Structural analysis of communities of practice: An investigation of job title, location, and management intention', in M. Huysman, E. Wenger, and V. Wulf (eds), *Communities and Technologies*. Amsterdam: Kluwer Academic Publishers, pp. 23–42.
- Allatta, J.T. (2005) *Worker Collaboration and Communities of Practice*. Ph.D. dissertation, University of Pennsylvania.
- Allen, M. (1974) 'The structure of interorganizational elite cooptation: Interlocking corporate directorates', *American Sociological Review*, 39(3): 393–406.
- Allen, T. (1977) *Managing the Flow of Technology*. Cambridge, MA: MIT Press.
- Bonacich, P. (1972) 'Factoring and weighting approaches to status scores and clique identification', *Journal of Mathematical Sociology*, 2: 112–20.
- Bonacich, P. (1987) 'Power and centrality: A family of measures', *American Journal of Sociology*, 92: 1170–82.
- Bonacich, P. (2007) 'Some unique properties of eigenvector centrality', *Social Networks*, 29(4): 555–64.
- Borgatti, S.P. (1989) *Regular Equivalence in Graphs, Hypergraphs, and Matrices*. University of California, Irvine.
- Borgatti, S.P. (2002) *NetDraw: Graph Visualization Software*. Harvard: Analytic Technologies.
- Borgatti, S.P. and Everett, M.G. (1989) 'The class of all regular equivalences: Algebraic structure and computation', *Social Networks*, 11: 65–88.
- Borgatti, S.P. and Everett, M.G. (1992) 'Regular blockmodels of multiway, multimode matrices', *Social Networks*, 14: 91–120.
- Borgatti, S.P. and Everett, M.G. (1997) 'Network analysis of 2-mode data', *Social Networks*, 19(3): 243–69.
- Borgatti, S.P. and Everett, M.G. (2006) 'A graph-theoretic framework for classifying centrality measures', *Social Networks*, 28(4): 466–84.
- Borgatti, S.P., Everett, M.G. and Freeman, L.C. (2002) *UCINET for Windows: Software for Social Network Analysis*. Harvard, MA: Analytic Technologies.
- Brandes, U. (2001) 'A faster algorithm for betweenness centrality', *Journal of Mathematical Sociology*, 25(2): 163–77.
- Burt, R. (1987) 'Social contagion and innovation: Cohesion versus structural equivalence', *American Journal of Sociology*, 92(6): 1287–335.
- Carroll, W.K., Fox, J. and Ornstein, M.D. (1982) 'The network of directorate interlocks among the largest Canadian firms', *Canadian Review of Sociology and Anthropology*, pp. 245–68.
- Davis, G. (1991) 'Agents without principles? The spread of the poison pill through the intercorporate network', *Administrative Science Quarterly*, 36(4): 583–613.
- Davis, G. and Greve, H. (1997) 'Corporate elite networks and governance changes in the 1980s', *American Journal of Sociology*, 103(1): 1–37.

- pavis, A., Gardner, B. and Gardner, R. (1941) *Deep South*. Chicago: University of Chicago Press.
- Domhoff, W. (1967) *Who Rules America?* Englewood Cliffs, NJ: Prentice-Hall.
- Eckhardt, C. and Young, G. (1936) 'The approximation of one matrix by another of lower rank', *Psychometrika*, 1: 211–18.
- Faust, K., Willber, K., Rowlee, D. and Skvoretz, J. (2002) 'Scaling and statistical models for affiliation networks: Patterns of participation among Soviet politicians during the Brezhnev era', *Social Networks*, 24: 231–59.
- Feld, S. (1981) 'The focused organization of social ties', *American Journal of Sociology*, 86: 1015–35.
- Feld, S., Frank, K., Schiller, K., Riegler-Crumb, C. and Muller, C. (2006) 'Identifying positions from affiliation networks: Preserving the duality of people and events', *Social Networks*, 28(2): 97–186.
- Friedkin, N. (1984) 'Structural cohesion and equivalence explanations of social homogeneity', *Sociological Methods and Research*, 12: 235–61.
- Galaskiewicz, J. (1985) *Social Organization of an Urban Grants Economy*. New York: Academic Press.
- Gmür, M. (2006) 'Co-citation analysis and the search for invisible colleges: A methodological evaluation', *Scientometrics*, 57(1): 27–57.
- Harary, F. (1969) *Graph Theory*. Reading, MA: Addison-Wesley.
- Kleinberg, J. (1999) 'Authoritative sources in a hyperlinked environment', *Journal of the ACM*, 46(5): 604–32.
- Krackhardt, D. and Carley, K.M. (1998) 'A PCANS model of structure in organization', in *Proceedings of the 1998 International Symposium on Command and Control Research and Technology*. Monterey, CA. pp. 113–19.
- Lazer, D., Mergel, I. and Friedman, A. (2009) 'Co-Citation of Prominent Social Network Articles in Sociology Journals: The Evolving Canon', *Connections*, 29(1): 43–64.
- Lester, R. and Cannella, A. (2006) 'Interorganizational familiness: How family firms use interlocking directorates to build community-level social capital', *Entrepreneurship: Theory & Practice*, 30(6): 755–75.
- Luce, R. and Perry, A. (1949) 'A method of matrix analysis of group structure', *Psychometrika*, 14(2): 95–116.
- McPherson, J.M. (1982) 'Hypernetwork sampling: Duality and differentiation among voluntary organizations', *Social Networks*, 3: 225–49.
- McPherson, J.M. and Smith-Lovin, L. (1986) 'Sex segregation in voluntary associations', *American Sociological Review*, 51(1): 61–79.
- McPherson, J.M. and Smith-Lovin, L. (1987) 'Homophily in voluntary organizations: Status distance and the composition of face-to-face groups', *American Sociological Review*, 52(3): 370–79.
- Mizruchi, M. (1983) 'Who controls whom? An examination of the relation between management and boards of directors in large American corporations', *Academy of Management Review*, 8: 426–35.
- Mizruchi, M. (1992) *The Structure of Corporate Political Action: Interfirm Relations and Their Consequences*. Cambridge, MA: Harvard University Press.
- Mizruchi, M. (1996) 'What do interlocks do? An analysis, critique, and assessment of research on interlocking directorates', *Annual Review of Sociology*, 22: 217–98.
- Newman, M., Strogatz, H. and Watts, D. (2001) 'Random graphs with arbitrary degree distributions and their applications', *Physical Review E*, 64: 1–17.
- Robins, G. and Alexander, M. (2004) 'Small worlds among interlocking directors: Network structure and distance in bipartite graphs', *Computational & Mathematical Organization Theory*, 10(1): 69–94.
- Roethlisberger, F. and Dickson, W. (1939) *Management and the Worker*. Cambridge: Cambridge University Press.
- Sokal, R. and Sneath, P. (1973) *Numerical Taxonomy*. San Francisco: W.H. Freeman.
- Uzzi, B. and Spiro, J. (2005) 'Collaboration and creativity: The small world problem', *American Journal of Sociology*, 111(2): 447–504.
- Wang, P., Sharpe, K., Robins, G. and Pattison, P. (2009) 'Exponential random graph (p^*) models for affiliation networks', *Social Networks*, 31(1): 12–25.
- Westphal, J.D. and Khanna, P. (2003) 'Keeping directors in line: Social distancing as a control mechanism in the corporate elite', *Administrative Science Quarterly*, 48(3): 361–98.
- Westphal, J.D. (1998) 'Board games: How CEOs adapt to increases in structural board independence from management', *Administrative Science Quarterly*, 43: 511–37.
- White, H.C., Boorman, S.A. and Breiger, R.L. (1976) 'Social structure from multiple networks, I: Blockmodels of roles and positions', *American Journal of Sociology*, 81: 730–80.
- White, D. and Reitz, K. (1983) 'Graph and semigroup homomorphisms on networks of relations', *Social Networks*, 5: 193–224.