Identifying statistically significant edges in one-mode projections

Zachary Neal

Sociology Department, 316 Berkey Hall, Michigan State University, East Lansing, MI 48824-1111

Abstract

One-mode projections of two-mode data are typically valued, and therefore require dichotomization before they can be analyzed using many network analytic methods. The traditional dichotomization approach, in which a universal threshold is applied to all edge weights, can yield a binary one-mode projection with undesirable artifacts and requires the arbitrary selection of a threshold value. This paper proposes a method for identifying statistically significant edges in one-mode projections, which can be used to construct both binary and signed projections. The method is demonstrated using two-mode data on southern women's social event participation and U.S. Supreme Court justices' majority decision participation, and is compared to two alternative approaches for normalizing edge weights in one-mode projections.

Keywords: bipartite, dichotomize, one-mode projection, signed network, two-mode

1. Introduction

Two-mode networks have been examined in a variety of contexts including southern women attending social events (Davis, Gardner, & Gardner, 1941), supreme court justices joining majority opinions (Doreian, Batagelj, & Ferligoj, 2004; Mrvar & Doreian, 2009), world cities hosting branches of multinational firms (Taylor, 2001), individuals sitting on corporate boards (Mizruchi, 1996), and legislators sponsoring bills (Fowler, 2006). Although

Email address: zpneal@msu.edu (Zachary Neal)

several methods have been developed for the direct analysis of two-mode networks (e.g., Borgatti & Everett, 1997; Field, Frank, Schiller, Riegle-Crumb, & Muller, 2006), their one-mode projections are still frequently examined. Although one-mode projections contain less information then their two-mode sources (Latapy, Magnien, & Del Vecchio, 2008), their use may still be justified by an interest in only one of the two modes (e.g. in actors, but not events), or by a desire to utilize network analytic techniques not yet extended to the two-mode context.

Borgatti & Everett (1997) contend that "there is no need to develop any new techniques to analyze" one-mode projections "for which the full range of network analytic methods are available" (p. 246). However, the edges in one-mode projections are nearly always valued, which complicates their analysis. As Latapy et al. (2008) explain, "one transforms the problem of analyzing a bipartite structure into the problem of analyzing a weighted one, which is not easier" (p. 35). Most work addressing these issues has sought to develop either methods for examining two-mode network without the use of projection, or methods for examining weighted networks (e.g., Newman, 2004). This paper adopts a different approach, developing a method for identifying statistically significant edges in one-mode projections. By focusing only on statistical significant edges, a valued projection can be dichotomized in a principled way, and thus examined using the full range of classical network analytic techniques. However, this proposed method offers two additional benefits. First, through a test of edge weight statistical significance, it separates the structural "signal" from the "noise" introduced by projection. Second, when this test is used in a two-tailed form to identify edges with statistical significantly large weights *and* statistically significantly small weights, it permits not only the construction of a binary projection (i.e. actors either do or do not have a relationship), but also of a signed projection (i.e. actors have a positive relationship, a negative relationship, or no relationship).

I begin by formally defining one-mode projection and discussing issues with their analysis and interpretation. I then describe the proposed method for identifying statistically significant edges in one-mode projections. This method is illustrated using two example datasets: Davis et al.'s (1941) Deep South data and Doreian et al.'s (2004) U.S. Supreme Court data. It is also compared to two normalizations commonly applied to valued one-mode projections (Borgatti & Halgin, 2011). I conclude with a discussion of this approach's limitations, and directions for future research.

2. One-mode projections

Two-mode networks, which are also known as bipartite or affiliation networks, describe the affiliation patterns of nodes in one set with nodes in another. For example, Davis et al.'s (1941) classic *Deep South* data describes 14 southern women's (i.e. one set of nodes) participation in 18 social events (i.e. the other set of nodes). In the context of interlocking directorate research, two-mode networks are used to describe individuals' membership on corporate boards of directors (Mizruchi, 1996), while research on world city networks has used two-mode networks to describe patterns of major cities hosting branch offices of advanced producer service firms (Taylor, 2001). For simplicity, in this paper I use the terms *actor* and *event* to refer to the two sets of nodes, and the term *participation* to refer to actors' affiliation(s) with events. A two-mode network can be represented as a matrix, $\mathbf{T}$, in which $T_{ij} = 1$ if actor i participates in event j, and otherwise equals 0.

The one-mode projection, $\mathbf{P}$, of a two-mode network, $\mathbf{T}$, is defined by

$$\mathbf{P} = \mathbf{T}\mathbf{T}' \tag{1}$$

The diagonal cells, P_{ii}, equal the total number of events in which actor i participates and have the range

$$0 \leq P_{ii} \leq E$$

where E is the total number of events (i.e. the number of columns in $\mathbf{T}$). In the analysis of one-mode projections, as with the analysis of most networks, these diagonal values are typically ignored. The off-diagonal cells, P_{ij}, equal the number of events in which actor i and actor j both participate (Breiger, 1974), and have the range

$$min(P_{ii}, P_{jj}) - (E - max(P_{ii}, P_{jj})) \leq P_{ij} \leq min(P_{ii}, P_{jj})$$

These off-diagonal cells are typically the focus of analysis, driven by the assumption that when two actors participate in the same events, they are more likely to interact with one another. This approach to viewing social structure as arising from individuals' co-participation in events has its origins in Simmel's (1955) essay on *The Intersection of Social Circles*, and later formed the foundation of Feld's (1981) theory of *foci*.

Others have identified a number of challenges to the analysis of one-mode projections, including the loss of information in the projection transformation and the distortion of structural characteristics like density and clustering

(Latapy et al., 2008; Neal, 2012). Although these are important considerations, a more immediate challenge arises when treating a one-mode projection as a network reflecting actors' interactions with one another: the interpretation of the off-diagonal edge weights. Because P_{ij} explicitly indicates the number of events in which i and j both participated, one possibility is to interpret higher values as stronger evidence of i's interaction with j, or as evidence of a more intense interaction between them. Under this interpretation of edge weights, a valued projection can easily be transformed into a binary projection by using a *universal* threshold to dichotomize edge weights. Actors who co-participate in more than the threshold number of events are viewed as having a relationship, while those who co-participate in fewer than the threshold number of events are not. Using this approach, there are as many dichotomizations as there are values of P_{ij}, and which dichotomization is the 'right' one is often unclear. Frequently the threshold is set at 0 (i.e. any co-participation implies a relationship) or the mean of P_{ij} (i.e. an above-average number of co-participations implies a relationship), while others have explored selecting a threshold that yields a graph meeting certain conditions (e.g. amount of transitivity; Freeman, 1992).

Setting aside the selection of an appropriate value, any universal-threshold dichotomization of P_{ij} can be misleading. Suppose $P_{ij} = 10$ and $P_{xy} = 10$; this might be viewed as equally strong evidence that actors i and j interact and that actors x and y interact. However, further suppose that i and j each participated in 10 events, while x and y each participated in 1000 events. The fact that $P_{ij} = 10$ now appears as quite strong evidence of i and j's interaction because it indicates a 100% rate of co-participation; at every opportunity that i and j could be together, they actually were. In contrast, the fact that $P_{xy} = 10$ now appears as relatively weak evidence of x and y's interaction because it indicates only a 1% rate of co-participation; although x and y could have been together much more often, they were not. This example highlights a key shortcoming of applying universal thresholds to one-mode projections: the same number of co-participations may be noteworthily high for one dyad, but not for another. That is, when it comes to inferring relationships from levels of co-participation, one size does not fit all (dyads). The interpretation of P_{ij} as the likelihood or intensity of a relationship between i and j must take into account other factors, such as the total number of events in which each actor participated (P_{ii} and P_{jj}). Thus, the central question is: for a given dyad, what value of P_{ij} provides evidence that i and j have a relationship?

3. Identifying statistically significant edges

One answer to this question can be motivated by viewing event co-participation as arising from a type of market. Each actor has a finite stock of capital that enables him/her to participate in events. This 'participation capital' might take several forms, including free time, physical mobility, wealth, or gregariousness. Actors 'spend' their capital by participating in events, such that actors with more capital can participate in more events. As with any market, actors can chose how to spend their capital by attending some events but not others. However, their choices are constrained by their stock of capital and by the number of events on which they can spend it. For example, an assistant professor preparing for tenure has little free time, and thus her choices about event participation will be severely constrained. Likewise, an outgoing socialite living in a sleepy rural town has few events in which to participate, and thus will be similarly constrained.

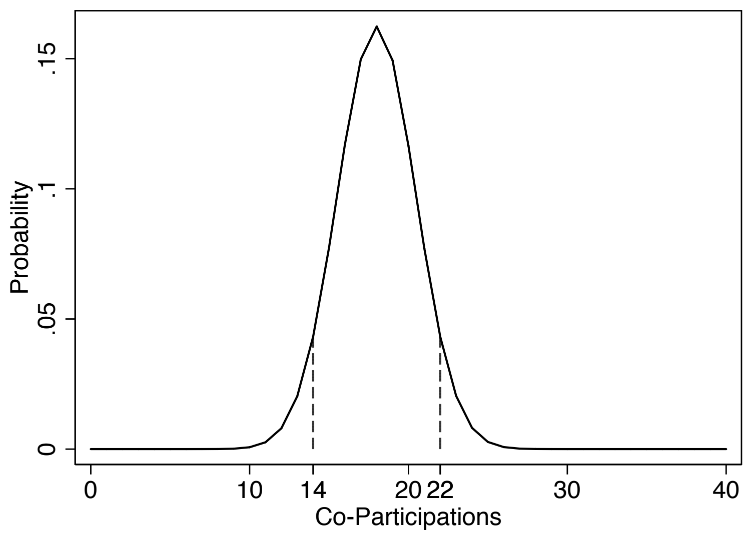
Adopting this market analogy, if actors' co-participation in events is to be interpreted as evidence of purposeful interaction rather than coincidence, it becomes necessary to consider the likelihood that i and j would spend their (potentially unequal) stocks of capital to participate in the same events coincidentally. Three different parameters are relevant to this consideration. The number of events open to i's and j's participation, E, is defined by the number of columns in $\mathbf{T}$. The number of events in which each is able to participate (i.e. their stock of capital) is unknown. However, the actual number of events in which each is observed to participate, P_{ii} and P_{jj} provides a plausible estimate, particularly if it is assumed that when one *can* participate in an event, one *does*. Given these parameters, if actors i and j participate in events (i.e. spend their capital) randomly, the probability that they co-participate in exactly C events is

$$Pr(P_{ij} = C) = \frac{\binom{E}{C}\binom{E-C}{P_{ii}-C}\binom{E-P_{ii}}{P_{jj}-C}}{\binom{E}{P_{ii}}\binom{E}{P_{jj}}} \tag{2}$$

The denominator captures the total number of ways the two actors could divide their stocks of participation capital among E events, while the numerator captures the the number of such divisions that yield exactly C co-participations.

A simple example serves to illustrate how equation (2) defines a useful probability mass function. Suppose a given setting offers 100 events in which i and j might participate. Among these events, i has sufficient time and money to participate in 45, while j has sufficient time and money to participate in only 40. Evaluating equation (2) for each possible value of C (i.e. the range of P_{ij}) yields the probability mass function shown in figure 1. Given these parameters, P_{ij} is statistically significant at the $\alpha = 0.10$ level if it is less than 14 or greater than 22.[1] In more substantive terms, it is likely that i and j will co-participate in between 14 and 22 events, even if both actors' select events for participation randomly, while there is a less than 10% chance that they would accidentally co-participate in fewer than 14 or more than 22 events.

Figure 1: Probability mass function ($E = 100, P_{ii} = 45, P_{jj} = 40$), with critical values ($\alpha = 0.10$)



Critical values obtained from the probability mass function defined by equation (2) can be used as dyad-specific thresholds in the transformation of a valued one-mode projection, **P**. First, they can be used as dichotomizing

[1]Throughout this paper, I use the somewhat liberal $\alpha = 0.10$ for the purposes of illustration; in practice, a more conservative α-level may be more appropriate.

thresholds to construct a binary one-mode projection, **B**, where

$$\begin{aligned} B_{ij} &= 1 \text{ if } P_{ij} > \text{upper threshold}_{ij}, \text{ otherwise} \\ &= 0 \end{aligned}$$

In this example, i and j are viewed (with 90% confidence) as having a relationship if they co-participated in more than 22 events, and otherwise are not. Second, they can be used as thresholds to construct a signed one-mode projection, **S**, where

$$\begin{aligned} S_{ij} &= 1 \text{ if } P_{ij} > \text{upper threshold}_{ij} \\ &= 0 \text{ if lower threshold}_{ij} \leq P_{ij} \leq \text{upper threshold}_{ij} \\ &= -1 \text{ if } P_{ij} < \text{lower threshold}_{ij} \end{aligned}$$

In the example, i and j's co-participation in more than 22 events suggests (with 90% confidence) that they selected events to *ensure* their co-participation and thus that they have a *positive* relationship. In contrast, their co-participation in fewer than 14 events suggests (again with 10% confidence) that they selected events in order to *avoid* co-participation and thus that they have a *negative* relationship. Notably, unlike a traditional approach to dichotomization that relies on a universal threshold, these transformations offer a more refined approach by using thresholds unique to each dyad and by using a probability mass function to selecting threshold values in a principled way. For the remainder of the paper, I focus on the more informative signed projection, **S**, of which the binary projection, **B**, is a special case.

4. Examples

In this section, I demonstrate the utility of this method by comparing the signed projections of Davis et al.'s (1941) Deep South data and Doreian et al.'s (2004) U.S. Supreme Court data to the known social structure among the actors.

4.1. Deep South

Davis et al. (1941) sought to understand the social structure of the deep south, focusing on differences by race and class, and seeking to explain the role that clique membership and mobility played in social life. To identify cliques among women in white society, they compiled data on 18 women's participation in 14 social events using interviews, participant-observation, guest lists, and newspapers' social columns. Using somewhat different terminology, they used a one-mode projection of these two mode data to infer interaction between them, explaining that the "relationships of any one person...to others in a group could be studied, as in the case of Mrs. Evelyn Jefferson. It will be noted that she participated six times with Miss Mandeville but only twice with Miss Liddell." (p. 149) This approach is rooted in the naive assumption noted above, that more co-participation is indicative of a more likely, or more intense, relationship between two individuals.

Under this assumption, the simplest way to uncover the social structure is to dichotomize the valued one-mode projection using a universal threshold. Figure 2 illustrates the one-mode projection dichotomized using a mean threshold, where two women are viewed as having a relationship if they engaged in more co-participations than average (i.e. $P_{ij} > 2.26$). This network displays many of the structural features identified by Davis et al. (1941) through interviews: a two-clique arrangement, Ruth as a bridge between the cliques, and Pearl as a fringe clique member. It also reproduces the clique memberships reported by Davis et al. (1941) with 83% accuracy, incorrectly identifying Dorothy, Olivia, and Flora as isolates rather than as members of the clique on the left.[2] But, is the mean number of co-participations the right threshold to distinguish pairs of women who are friends from those who are not? For example, are three co-participations high enough to suggest that Theresa and Sylvia are friends? Likewise, are two co-participations low enough to suggest that Verne and Dorothy are not?

The method described above in section 3 offers an alternative approach to uncovering structure in one-mode projections and a direct answer to such questions. Figure 3 illustrates a signed one-mode projection that includes only statistically significant ($\alpha = 0.10$) ties. Ties between actors that co-participated in more events than expected at random (i.e. positive ties) are shown as thick solid lines, while those between actors that co-participated in fewer events than expected at random (i.e. negative ties) are shown as thin dotted lines. The nodes are positioned by applying a stress minimization

[2] According to Davis et al. (1941), clique 1 included a core (Evelyn, Laura, Theresa, and Brenda), primary members (Charlotte, Francis, and Eleanor), and a secondary member (Pearl). Likewise, clique 2 included a core (Sylvia, Nora, and Helen), primary members (Myrna and Katherine), and secondary members (Ruth, Verne, *Dorothy*, *Olivia*, and *Flora*).

Figure 2: Stronger-than-average ties in the deep south

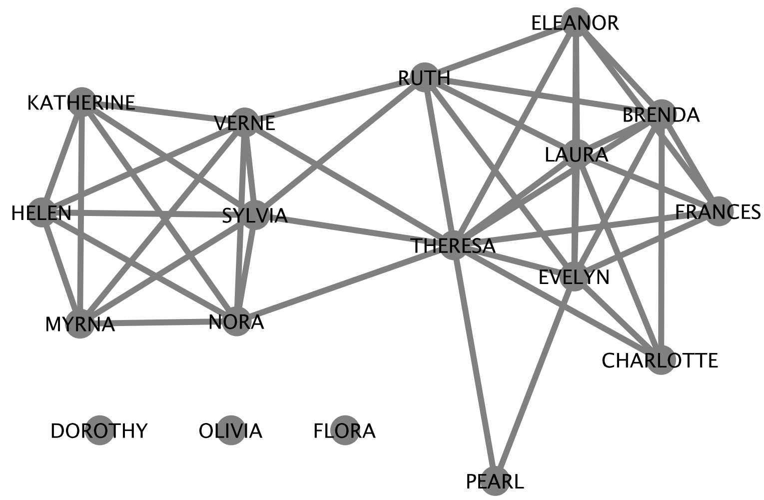
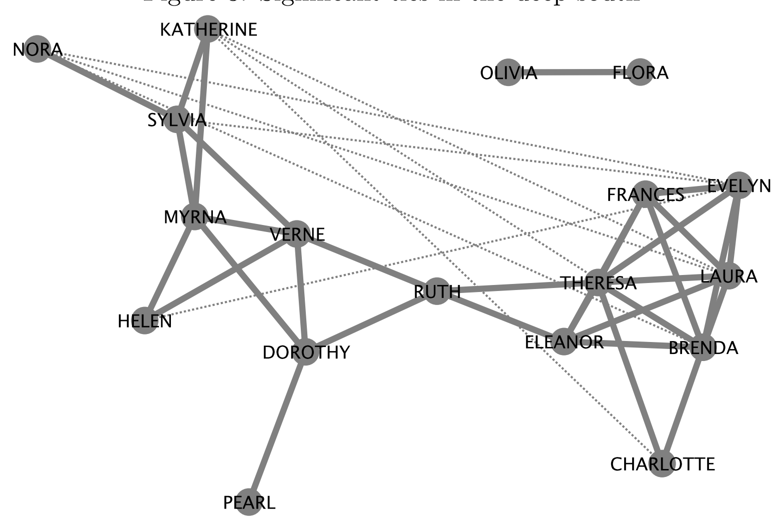


Figure 3: Significant ties in the deep south



algorithm to the positive ties only. The structure of positive ties in figure 3 is generally similar to figure 2, with the same two-clique structure and clique memberships described by Davis et al. (1941). However, comparison of these two networks with one another, and with Davis et al.'s (1941) observations, suggest that the proposed method effectively identifies important structural features that would otherwise be obscured.

In figure 3, the clique structure is more refined and more consistent with Davis et al.'s (1941) interview findings. Most notably, although the left clique appears dense and clustered in figure 2, it is relatively open in figure 3, mirroring the scattered pattern of its members' event participation. This difference highlights that when one-mode projections are dichotomized using a universal threshold, "even a random bipartite network – one that has no particular structure *built into it* at all – will be highly clustered" (Watts, 2003, p. 128). The apparent clustering in this clique in figure 2 is merely an artifact of the projection and dichotomization. Because the proposed method does not generate artificially high density, Ruth's position as a bridge between the two cliques is more clearly visible in figure 3, which better reflects Davis et al.'s (1941) observation that "Miss Ruth De Sand...was claimed by members of both cliques" (p. 151).

Differences in the presence or absence of positive relationships between specific women also make figure 3 more consistent with Davis et al.'s (1941) observed clique memberships. Such differences arise in two types of situations. In some cases, women who co-participated in a greater-than-average number of events and thus are connected in figure 2 are not connected in figure 3. Consider Theresa and Sylvia, who co-participated in three events. Because these women co-participated in more events than the average pair of women, a universal threshold dichotomization treats them as connected. However, equation (2) reveals that, because these women participated in about half of all events (Theresa in 8 of 14 and Sylvia in 7 of 14), they might co-participate in three events purely by chance. That is, Theresa and Sylvia's above-average number of co-participations is not enough to conclude they they have a relationship. Viewing Theresa and Sylvia as lacking a relationship is more consistent with Davis et al.'s (1941) observation that they were members of different cliques.

In other cases, women who co-participated in a less-than-average number of events and thus are not connected in figure 2 are connected in figure 3. Consider Verne and Dorothy, who co-participated in only two events. Because these women co-participated in fewer events than the average pair of women, a universal threshold dichotomization treats them as not connected. However, equation (2) reveals that because, Verne participated in only 4 of 14 events and Dorothy participated in only 2 of 14 events, they are unlikely to have co-participated in two events purely by chance. That is, Verne and Dorothy's number of co-participations, despite being below average, is still enough to conclude they they have a relationship. Viewing Verne and

Dorothy as having a relationship is more consistent with Davis et al.'s (1941) observation they they were members of the same clique.

A final difference between figures 2 and 3 lies in the latter's inclusion of *negative* ties. While dichotomization using a universal threshold distinguishes between present and absent relationships, the proposed method also allows relationships that are merely absent to be distinguished from those that suggest avoidance. Two features of the structure of negative ties in figure 3 closely mirror Davis et al.'s (1941) observations, highlighting the utility of examining a signed one-mode projection. First, the negative ties occur only *between* and never *within* cliques, suggesting that the cliques discerned through interviews with the women are genuinely distinct and cohesive social groups. Second, the negative ties occur primarily between women identified as 'core' members of their cliques, and never involve women identified as 'secondary' members, providing additional evidence of these women's respective roles in maintaining the social structure.

4.2. The U.S. Supreme Court

Examining the signed one-mode projection of Doreian et al.'s (2004) data on the U.S. Supreme Court offers a second opportunity to explore the proposed method's potential utility. These data describe the nine justices' 'participation' in the majority on 26 major cases heard during the court's 2000 – 2001 term. Figure 4 is a one-mode projection of these data dichotomized using a universal mean threshold, where two justices are viewed as connected if they both joined the majority on more cases than average (i.e. $P_{ij} > 10.6$). Figure 5 is a signed one-mode projection that includes only statistically significant ($\alpha = 0.10$) ties. In both cases, the nodes are positioned by applying a stress minimization algorithm to the positive ties, indicated by thick solid lines; negative ties are indicated by thin dotted lines. The universal threshold dichotomization (figure 4) suggests that Kennedy and O'Connor played the role of 'swing' justices bridging the court's conservative (Rehnquist, Scalia, and Thomas) and liberal (Breyer, Ginsburg, Stevens, and Souter) factions, which did not always agree. However, the signed projection (figure 5) challenges this view in two ways.

First, in figure 4, Kennedy and O'Connor occupy bridging positions simply because they are connected to most of the other justices – Kennedy to 7 (or 87.5%) and O'Connor to 6 (or 75%) – and thus by necessity are connected to justices in both the conservative and liberal factions. However, these justices' apparently high degrees do not reflect their ideologi-

Figure 4: Stronger-than-average ties in the U.S. Supreme Court

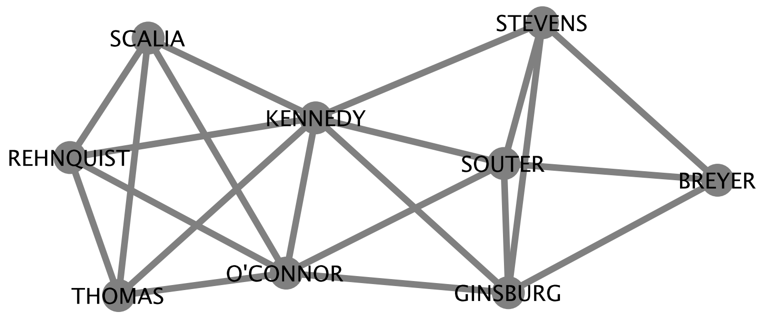
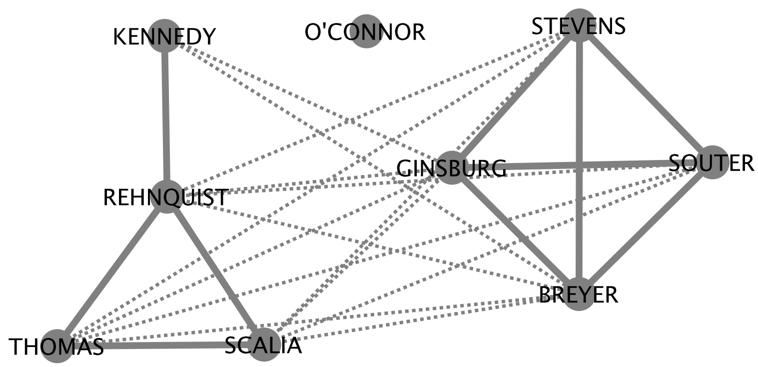


Figure 5: Significant ties in the U.S. Supreme Court



cal leanings, but rather the simple fact that they nearly always sided with the majority: Kennedy in 85% of the cases, and O'Connor in 73%. That is, they are not ideological bridges, but a yes-man and yes-woman. Their well-connectedness is another instance of the artificial density and clustering introduced by applying a universal dichotomization threshold to one-mode projections. Figure 5 more accurately captures Kennedy's and O'Connor's lack of ideological alignment by showing them with few, if any, connections to other justices.

Second, in figure 4, there is an absence of ties between justices in the conservative and liberal cliques. However, in figure 5, most of these ties are depicted as not merely absent but negative. In this context, a merely absent tie indicates two justices who maintain differing ideologies that lead them to sometimes agree, and sometimes disagree. In contrast, a negative tie indicates that two justices maintain directly oppositional ideologies that lead them to nearly always disagree. In the 2000 – 2001 court, for nearly every case in which the conservative justices were in the majority, the liberal justices were in the minority, and vice versa. Figure 5 more accurately captures this ideological divide by highlighting that the conservative and liberal factions are not simply ideologically different, but ideologically opposed.

5. Comparison to alternative approaches

The Deep South and Supreme Court examples suggest that the proposed method for identifying statistically significant edges in, and constructing binary and signed, one-mode projections offers a way to highlight structural features that would otherwise be obscured. However, it is worth considering whether this method represents an improvement over alternative approaches to handling one-mode projections. As noted earlier, the problem with applying a universal dichotomization threshold to $\mathbf{P}$ is that it ignores differences between dyads (e.g. actors' differing levels of event participation). One alternative involves first rescaling $\mathbf{P}$ based on such dyad-level factors, then applying a universal dichotomization threshold to the normalized projection.

Several methods of rescaling $\mathbf{P}$ have been suggested to control for dyadic characteristics, and thus to yield a normalized projection that shifts P_{ij} from indicating simple frequencies of co-participation to indicating "tendencies or revealed preferences to [co-participate]" (Borgatti & Halgin, 2011, pg. 422). Many of these normalizations begin by viewing each pair of actors' participations as a 2-by-2 contingency table (see table 1). For example,

Table 1: Contingency table of actors' event participation

		Actor j participates? yes	Actor j participates? no
Actor i participates?	yes	a	b
	no	c	d

Bonacich (1972) suggested that P_{ij} be normalized as

$$\begin{aligned} P'_{ij} &= 0.5 \text{ if } ad = bc \text{, otherwise} \\ &= \frac{ad - \sqrt{abcd}}{ad - bc} \end{aligned} \tag{3}$$

which rescales co-participation frequencies to range from 0, when i and j are engaged in the minimum possible number of co-participations, to 1, when they are engaged in the maximum possible number. Others have suggested that $a+d$ is a useful measure of social closeness, which is equal to the Pearson correlation between two actors' participation profiles when rescaled to range between -1 and 1 (Borgatti & Halgin, 2011). Thus, using this approach to normalization

$$\mathbf{P}'' = corr(\mathbf{T}') \tag{4}$$

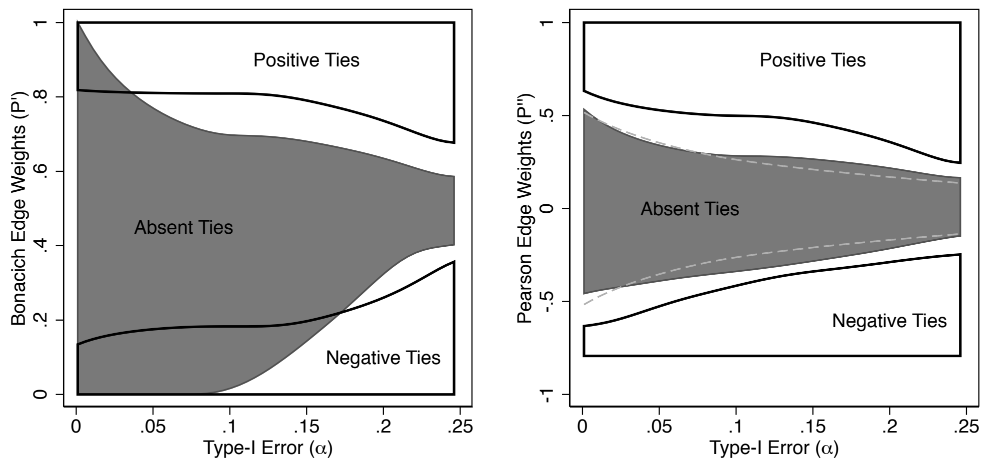
Can a universal dichotomization threshold, applied to a normalized one-mode projection ($\mathbf{P}'$ or $\mathbf{P}''$), reproduce the results of the proposed method?

Using Doreian et al.'s (2004) Supreme Court data, figure 6 compares edge weights obtained using the Bonacich ($\mathbf{P}'$, left panel) and Pearson ($\mathbf{P}''$, right panel) normalizations to signed edge weights obtained at different α-levels using the proposed method ($\mathbf{S}$).[3] In these plots, for a given α-level shown on the x-axis: (1) the outlined region at the top indicates the range of normalized edge weights that the proposed method classifies as positive ties, (2) the shaded region in the middle indicates the range of normalized edge weights that the proposed method classifies as absent ties, and (3) the outlined region at the bottom indicates the range of normalized edge weights that the proposed method classifies as negative ties. For example, in the Supreme Court network shown in figure 5 above, where $\alpha = 0.10$: (1) ties were identified as positive by the proposed method when $0.809 \leq P^{'}_{ij} \leq 1$ or when $0.498 \leq P^{''}_{ij} \leq 1$, (2) ties were identified as absent by the proposed method when $.0 \leq P^{'}_{ij} \leq 0.696$ or when $-0.310 \leq P^{''}_{ij} \leq 0.282$, and (3) ties were identified as negative by the proposed method when $0 \leq P^{'}_{ij} \leq 0.183$ or when $-0.793 \leq P^{''}_{ij} \leq -0.365$.

These comparisons highlight two reasons that these normalizations cannot be used to reproduce the signed one-mode projection ($\mathbf{S}$) constructed using the proposed approach. First, although they control for certain dyadic characteristics, using them to construct a binary or signed one-mode projection still requires the selection of threshold values. However, unlike the proposed method, neither normalization offers a principled way to select threshold values. Notably, the critical value of the Pearson correlation coefficient, illustrated in the right panel by a dashed line, approximates the threshold values necessary to distinguish positive, negative, and absent ties. However, the accuracy of this threshold approximation varies by network:

[3] Comparing these normalizations to the proposed method using the Deep South data yields similar results.

Figure 6: Comparison of normalized and signed edge weights in the Supreme Court



although it is quite accurate in the Supreme Court data, it is much less accurate in the Deep South data. Thus, in the absence of clear guidance on the selection of threshold values, equations (3) and (4) simply transform one valued projection ($\mathbf{P}$) into another ($\mathbf{P}'$ or $\mathbf{P}''$), and thus do not simplify its analysis.

Second, Bonacich normalized edge weights ($\mathbf{P}'$) are unable to distinguish absent ties from positive or negative ties at certain α-levels, indicated by the overlap of these regions in the left panel of figure 6. For example, when $\alpha = 0.10$, some edges with a normalized weight of 0 are identified by the proposed method as absent ties (e.g. between Kennedy and Souter), while other edges with a normalized weight of 0 are identified as negative ties (e.g. between Kennedy and Breyer). This occurs because equation (3) normalizes a dyad's edge weight based only its range of maximum and minimum possible co-participations, but does not incorporate the total number of attendable events as a parameter. Thus, low Bonacich edge weights do not necessarily indicate noteworthily low levels of co-participation, but only levels of co-participation that are near their minimum. For example, in a setting with 1000 attendable events in which actors i and j each participate in a *different* single event, $P'_{ij} = 0$ despite the fact that their lack of co-participation is virtually guaranteed. In this case, the Bonacich edge weight may lead one to view the relationship between i and j as negative, when it is more appropriately viewed as simply absent. Likewise, high Bonacich edge weights do not necessarily indicate noteworthily high levels of co-participation, but only levels of co-participation that are near their maximum. In a setting with 1000 attendable events in which actor i participates in 999 events and actor j participates in exactly 1 of these 999 events and no others, $P'_{ij} = 1$ despite the fact that their co-participation is virtually guaranteed. Here, the Bonacich edge weight may lead one to view the relationship between i and j as positive, when it is more appropriately viewed as absent. The ability to discriminate positive, absent, and negative ties by their edge weights does not appear to be a problem when using Pearson normalized edge weights, indicated by the fact that none of the regions overlap in the right panel of figure 6.

6. Conclusion

This approach is not without some limitations. First, the proposed method treats events as interchangeable. This is potentially problematic because co-participation in some events (e.g. small intimate gatherings) is more likely to indicate a relationship between two actors than co-participation in other events (e.g. large conventions attended by thousands). This limitation can be viewed as arising from the omission of an important dimension of the market analogy. Equation (2) controls for a setting-level factor (i.e. the number of events), and actor-level factors (i.e. each actor's stock of participation capital), but not for event-level factors (i.e. each event's popularity). Extensions of this method should investigate ways to incorporate such event-level factors.

Second, this method treats all events in the setting as equally open to participation from all actors. However, this may not be the case. For example, independent of the total number of events in which a given woman can participate, she may be blocked from participation in some of the events, perhaps because they are located too far away. Similarly, independent of the total number of cases in which a given justice can join the majority, he/she may be blocked from joining the majority in some of the cases, perhaps because a conflict of interest requires recusal. It is likely unwise to interpret one-mode projections constructed from two-mode data in which many of the actor actors cannot, in principle, participate in many events. However, an extension of this method should also consider ways to account for more moderate cases where some actors cannot participate in some events.

Despite these limitations, which point toward directions for future research, this paper has demonstrated how a market analogy can be used to consider the likelihood of two actors' co-participation in events, and offered a corresponding test to identify statistically significant edges in one-mode projections. The critical values that distinguish significant from non-significant edges can be used as thresholds for the construction of binary and signed one-mode projections, which can be more easily analyzed than valued projections. In the cases of Davis et al.'s (1941) Deep South and Doreian et al.'s (2004) Supreme Court data, the resulting signed projections are consistent with expectations about the underlying social and ideological structures, and highlight structural features that would be obscured in other one-mode projections of these data. Additionally, because this approach is grounded in a clearly defined probability mass function, from which threshold critical values for specified α-levels can be determined, it offers a more principled method of dealing with one-mode projections than alternative normalizations.

References

Bonacich, P. (1972). Technique for analyzing overlapping group memberships. *Sociological Methodology*, 4, 176 – 185.

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

Borgatti, S. P. & Halgin, D. S. (2011). Analyzing affiliation networks. Pp. 417 – 434 in *The Sage Handbook of Social Network Analysis*, Scott, J. & Carrington, P. J. (eds.). Thousand Oaks, CA: Sage.

Breiger, R. L. The duality of persons and groups. *Social Forces*, 53, 181 – 190.

Davis, A., Gardner, B. G., & Gardner, M. R. (1941). *Deep South: A Social Anthropological Study of Caste and Class.* Chicago: University of Chicago Press.

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

Feld, S. L. The focused organization of social ties. *American Journal of Sociology*, 86, 1015 – 1035.

Field, S., Frank, K. A., Schiller, K., Riegle-Crumb, C., & Muller, C. (2006). Identifying positions from affiliation networks: preserving the duality of people and events. *Social Networks*, 28, 97 – 123.

Fowler, J. H. (2006). Legislative cosponsorship networks in the U.S. house and senate. *Social Networks*, 28, 454– 465.

Freeman, L. C. (1992). The sociological concept of "group": An empirical test of two models. *American Journal of Sociology*, 98, 152 – 166.

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

Mizruchi, M. S. (1996). What do interlocks do? An analysis, critique, and assessment of research on interlocking directorates. *Annual Review of Sociology*, 22, 271 – 298.

Mrvar, A., & Doreian, P. (2009). Partitioning signed two-mode networks. *Journal of Mathematical Sociology*, 33, 196 – 221.

Neal, Z. (2012). Structural determinism in the interlocking world city network. *Geographical Analysis*, 44, 162 – 170.

Newman, M. E. J. Analysis of weighted networks. *Physics Review E*, 70, 056131.

Taylor, P. (2001). Specification of the world city network. *Geographical Analysis*, 33, 181 – 194.

Taylor, P. J. & Catalano, G. (2000). GaWC dataset 11. Retrieved on 31 May 2012 from http://www.lboro.ac.uk/gawc/datasets/da11.html.

Simmel, G. (1955). The web of group affiliations. Pp. 125 – 195 in *Conflict and the Web of Group Affiliations*, transl. Reinhard Bendix. Glencoe, IL: The Free. Press.

Watts, D. J. (2003). *Six Degrees: The Science of a Connected Age*. New York: W. W. Norton.