The backbone of bipartite projections: Inferring relationships from co-authorship, co-sponsorship, co-attendance and other co-behaviors

Zachary Neal*

Psychology Department, 316 Physics Road, Michigan State University, East Lansing, MI 48824, United States

**A R T I C L E   I N F O**

**Keywords:**
- Bill co-sponsorship
- Binarize
- Dichotomize
- Political network
- Projection
- Two-mode

**A B S T R A C T**

The analysis and visualization of weighted networks pose many challenges, which have led to the development of techniques for extracting the network's backbone, a subgraph composed of only the most significant edges. Weighted edges are particularly common in bipartite projections (e.g. networks of co-authorship, co-attendance, co-sponsorship), which are often used as proxies for one-mode networks where direct measurement is impractical or impossible (e.g. networks of collaboration, friendship, alliance). However, extracting the backbone of bipartite projections requires special care. This paper reviews existing methods for extracting the backbone from bipartite projections, and proposes a new method that aims to overcome their limitations. The stochastic degree sequence model (SDSM) involves the construction of empirical edge weight distributions from random bipartite networks with stochastic marginals, and is demonstrated using data on bill sponsorship in the 108th U.S. Senate. The extracted backbone's validity as a network reflecting political alliances and antagonisms is established through comparisons with data on political party affiliations and political ideologies, which offer an empirical ground-truth. The projection and backbone extraction methods discussed in this paper can be performed using the -onemode- command in Stata.

© 2014 Elsevier B.V. All rights reserved.

1. Introduction

Analyzing and visualizing weighted networks presents a number of challenges, which has lead to the development of methods for extracting the 'backbone' of these networks. Such backbone extraction methods aim to reduce the original, weighted network into a simpler, binary network that preserves only those edges whose weights are sufficiently large to suggest they are significant. The challenge lies in determining how strong an edge's weight must be before deeming it significant. Several techniques for assessing a edge's significance and thus achieving this reduction have been proposed, ranging from relatively simple methods like an unconditional threshold that retains the strongest edges to more sophisticated methods that compare observed edge weights to expectations from a null model (Serrano et al., 2009) or to empirical distributions (Forti et al., 2011).

Weighted networks arise in many different contexts, but are particularly common in the case of bipartite network projections, including networks where authors are linked by the number of papers they have co-authored (e.g. de Stefano et al., 2013) or actors are linked by the number of movies they have both appeared in (e.g. Watts and Strogatz, 1998). However, the backbone extraction methods developed for natively one-mode networks are not well suited for one-mode networks that have been obtained from bipartite data via projection. Such methods fail to incorporate information present in the original bipartite data into their decisions about whether a given edge should be preserved in the backbone network. Several alternative backbone extraction methods have been developed specifically for bipartite projections (e.g. Zweig and Kaufmann, 2011; Neal, 2013), but they are computationally complex and risk imposing too many or too few assumptions. The purpose of this paper is to review the existing methods for extracting the backbone from bipartite projections, then to propose and demonstrate a new method that aims to overcome some of the existing methods' limitations.

After defining bipartite networks and some key features of their projections, I briefly review the most commonly used methods for extracting the backbone of bipartite projections, noting their strengths and weaknesses. I then describe a new method – the stochastic degree sequence model (SDSM) – that involves building empirical probability distributions using a sample of random bipartite networks with stochastic row and column degree sequences. In Section 5, I provide a step-by-step demonstration of this method using data on bill sponsorship activities in the 108th U.S. Senate.

* Tel.: +1 517 432 1811.
  E-mail address: zpneal@msu.edu

http://dx.doi.org/10.1016/j.socnet.2014.06.001
0378-8733/© 2014 Elsevier B.V. All rights reserved.
to infer political alliances and antagonisms among senators. In this context, the SDSM involves asking whether two senators co-sponsored significantly more bills (suggesting an alliance) or significantly fewer bills (suggesting an antagonism) than they might have co-sponsored in plausible alternate worlds in which the senators randomly sponsored roughly the same number of bills and the bills were randomly sponsored by roughly the same number of senators. It yields a backbone network of political alliances and antagonisms that exhibits a high level of criterion validity when compared to expectations based on political party and ideology data, which offer an empirical ground-truth. The paper concludes with a discussion of the proposed method’s limitations and directions for future research on the analysis of bipartite projections.

2. Bipartite networks and projections

A bipartite network is composed of two mutually exclusive sets of nodes: edges may exist between nodes in different sets, but not between nodes in the same set. Also known as two-mode or affiliation networks, bipartite networks have been discussed in many different contexts including southern women attending social events (Davis et al., 1941), individuals sitting on corporate boards (Mizruchi, 1996), actors appearing in movies (Watts and Strogatz, 1998), world cities hosting branches of multinational firms (Taylor, 2001), supreme court justices joining majority opinions (Doreian et al., 2004), legislators sponsoring bills (Fowler, 2006a), and ingredients possessing flavor compounds (Abn et al., 2011). Formally defined, an m-by-n bipartite network, B, in which \( B_{ij} = 1 \) if there is an edge between i and j and otherwise is zero, can be projected onto an m-by-m unipartite or one-mode network, \( P \), as \( BB \) (Breiger, 1974). Using this approach, for example, a bipartite network that describes legislators’ sponsorship of bills is transformed into a unipartite or one-mode network of legislators linked to one another by their co-sponsorship of bills.

To facilitate a discussion of bipartite projections, some generic terminology is useful. Throughout this paper, I use the terms agent and artifact to describe the two sets of nodes in a bipartite network. Agents are represented as rows in B and are the primary nodes of interest. An agent’s degree is the row marginal of B, and indicates an agent’s total number of artifacts, for example, how many social events (the artifacts) a given person (an agent) attended or how many bills (the artifacts) a given legislator (an agent) sponsored. Artifacts are represented as columns in B and are instrumental in forging the linkages between agents in P, but are not of direct interest in the bipartite projection. An artifact’s degree is the column marginal of B, and indicates an artifact’s total number of agents, for example, how many people (the agents) attended a given social event (an artifact) or how many legislators (the agents) sponsored a given bill (an artifact).

Much has already been written about the mathematical properties of bipartite projections (see Latapy et al., 2008), however the nature of edge weights in bipartite projections is of particular concern in the methods discussed below, and thus warrants brief consideration. The weight of an edge in the projection, \( P_{ij} \), reflects the number of artifacts that agents i and j have in common (e.g. the number of bills two legislators both sponsored), Some have argued that bipartite projections are easier to analyze than the original bipartite network because they are one-mode networks, noting that “there is no need to develop any new techniques to analyze [bipartite projections]...for which the full range of network analytic methods are available” (Borgatti and Everett, 1997, p. 246).

However, because bipartite projections are nearly always weighted networks, their analysis is not as straightforward as this claim implies. Projecting a bipartite network into a one-mode network merely “transforms the problem of analysing a bipartite structure into the problem of analysing a weighted one, which is not easier” (Latapy et al., 2008, pp. 34–35). One important but rarely noted feature of these edge weights is their constrained range of possible values. The range of values an edge between agents i and j may take in a bipartite projection can be expressed as a function of these agents’ degrees (i.e. \( D_i \) and \( D_j \)) and the total number of artifacts (\( A \)):

\[
\min(D_i, D_j) - (A - \max(D_i, D_j)) \leq P_{ij} \leq \min(D_i, D_j)
\]

A simple example serves to illustrate. Suppose Tom attends 5 of 10 parties, and Jerry attends 7 of the same 10 parties. From Eq. (1), we know that Tom and Jerry must have co-attended at least 2 parties, and could not have co-attended more than 5 parties. A critical implication of this identity is that, ceteris paribus, higher-degree agents will necessarily have stronger edges than lower-degree agents.

Before turning to methods for dealing with these edge weights, it is also useful to consider why one would examine a bipartite projection at all. Indeed, the projection transformation involves the loss of information including the specific identity of the artifacts responsible for forging linkages between agents (Latapy et al., 2008), and methods are emerging for analyzing bipartite networks without requiring their projection (Borgatti and Everett, 1997; Agneessens and Everett, 2013). Nonetheless, bipartite projections remain an important methodological tool in research where the interest is in a natively one-mode network, but where measurement of this network is impossible or impractical. In developmental psychology research on peer relationships among youth, high non-response rates and challenges associated with obtaining parental permission required by Institutional Review Boards make the direct collection of one-mode social network data is difficult. As a solution, a method known in this literature as Social Cognitive Mapping uses bipartite projections in which children are linked by their co-participation in social groups to infer the unobserved social network of interest (e.g. Cairns and Cairns, 1994; Gest et al., 2007; Neal and Neal, 2013). Similarly, in political science research on relationships of political alliance and influence, politicians’ compelling strategic reasons for wanting to conceal their alliances makes direct collection of such data impossible. As a solution, some have turned to bipartite projections reflecting bill co-sponsorship or committee co-membership to infer the unobserved social network of interest (e.g. Porter et al., 2005; Fowler, 2006a). Finally, in geography research on global economic relations between cities, although some types of data exist on trade and foreign direct investment between countries, no such data exists at the city level. As a solution, a method known in this literature as the Interlocking World City Network Model uses bipartite projections in which cities are linked by the co-location of branches of advanced producer service firms (e.g. institutional banks, law firms, accounting agencies, etc.) to infer the unobserved economic network of interest (e.g. Taylor, 2001; Neal, 2008). In each case, a bipartite projection is used as a proxy to infer an unobserved, natively one-mode network of interest. When used as a proxy measurement tool, the methods for handling edge weights in bipartite projections must permit such inferences to be made in a principled way.

3. Existing methods for backbone extraction

All methods of network backbone extraction, whether applied to natively one-mode networks or to bipartite projections, involve the use of a threshold. Edges whose weights exceed the threshold value are retained in the backbone, while those whose weights are below the threshold value are omitted from the backbone. Backbone extraction methods vary, however, in how threshold values are identified. In this section, I review three broad approaches that can be distinguished by the information on which the selection of threshold values is conditioned. Table 1 summarizes examples of
these approaches, highlighting how the threshold value is selected in each case, and the key disadvantages. The -onemode- command in Stata, which can be downloaded and installed by typing \texttt{sac install onemode} in the Stata command line, will extract backbones from bipartite projections using each of these approaches.

### 3.1. Unconditional thresholds

The simplest and most widespread approach to extracting the backbone of bipartite projections has been through the application of an unconditional or global threshold. A single threshold is selected by the researcher and applied to all edges in the bipartite projection; edges are retained in the backbone network if their weight in the bipartite projection exceeds the threshold value. The most commonly used threshold value of 0 preserves all edges with a non-zero weight (Latapy et al., 2008), but others have used different thresholds including those set at a percentage of the maximum observed edge weight (e.g., Derudder and Taylor, 2005; Neal, 2008) or at the mean observed edge weight (e.g., Neal, 2013). Although it is widely used, the unconditional threshold approach suffers from three major shortcomings: arbitrariness, structural bias, and uniscalarity.

First, and perhaps most obviously, the structure of a backbone network extracted using an unconditional threshold critically depends on the specific threshold value chosen. For example, a low threshold value (e.g., 0, thus preserving all edges of any weight) will yield a large, dense network, while a high threshold will yield a relatively smaller, sparser network. As a consequence, different unconditional thresholds can yield networks with radically and often unpredictably different structural attributes (Butts, 2009). Although this may not be problematic on its own, it becomes problematic in the absence of a strong theoretical rationale for the selected threshold value. The application of an arbitrary unconditional threshold yields an equally arbitrary backbone network.

Second, many have noted that certain structural features of unconditional threshold backbones of bipartite networks are systematically biased. For example, commenting on affiliation networks like the Hollywood co-acting network that informed early research on small-world networks, Watts (2003) observed that

> “even a random bipartite network – one that has no particular structure built into it at all – will be highly clustered…random affiliation networks [extracted using an unconditional threshold] will always be small-world networks.” (p. 128)

One cannot claim to have ‘discovered’ small-world structure in an unconditional threshold backbone of a bipartite projection; such structures are simply products of the backbone extraction method.

As metrics that are closely associated with the clustering coefficient, density and transitivity are also inflated in unconditional threshold backbones of bipartite projections. Additionally, Neal (2012) has shown that the number of artifacts relative to the number of agents constrains the types of motifs that may be observed in an unconditional threshold backbone, while the number and degree of artifacts constrains the number and size of maximal subgraphs (i.e., cliques). These instances of structural bias arise because an artifact held by agents induces \((a(a-1)/2)\) edges in a backbone extracted using an unconditional threshold of zero, or put another way, in an unconditional threshold backbone all agents having a given artifact are defined as linked to one another (Latapy et al., 2008). Thus, when examining unconditional threshold backbones of bipartite projections, many interesting questions about network structure (e.g. is it a small-world? does it contain many cliques?) cannot be addressed or are not worth asking.

Finally, the use of an unconditional threshold for backbone extraction eliminates any multiscalar features of the network. The uniscalarity of unconditional threshold backbones was initially noted in the context of natively one-mode networks, where Serrano et al. (2009) observed that such an extraction method “would destroy the multiscale nature…where weights are locally correlated on edges incident to the same node” (p. 6483). This issue also arises in the case of bipartite projections where weights are locally correlated due to Eq. (1). Recall that the edge weight between two agents is a function of each of their number of artifacts, and thus agents with more artifacts necessarily have larger edge weights. Thus, using an unconditional threshold backbone of a bipartite projection to infer an edge between two agents would lead to inferences that agents with few artifacts are never (or rarely) linked, while those with many artifacts are always (or often) linked. Substantively, for example, this might mean inferring that people (the agents) who attend few parties (the artifacts) are never friends with others, while those who attend many parties are always friends with others. However, this ignores the fact that what counts as a large number of shared artifacts, and thus might be used as evidence for inferring an edge between two agents, will depend heavily on these agents’ degrees. Again, more substantively, what counts as ‘attending many of the same parties’ is different for people who attend few parties and those who attend many parties.

In certain contexts, the issues of arbitrariness, structural bias, and uniscalarity that arise in unconditional threshold backbones may not represent limitations. Consider, for example, using such a backbone to infer a one-mode network intended to capture individuals’ exposure to a virus through attendance at a party where an infected person is present. In this case, the number of event co-attendances may be irrelevant because even a single instance of
co-presence at an event creates an exposure. Likewise, the inflated density meaningfully captures the fact that a single infected person at a party exposes everyone else at the party to the virus. More broadly, if any instance of shared artifacts between agents is sufficient to infer that an edge exists between two agents (e.g., viral exposure through mere co-presence), then backbone extraction using an unconditional threshold is appropriate. If, however, inferring that an edge exists between two agents requires more than simply an instance of shared artifacts, unconditional threshold backbones are not suitable. For example, simply observing two children in the same social group is not particularly strong evidence for inferring they are friends, and simply observing two legislators co-sponsoring a bill is not particularly strong evidence for inferring they are political allies. Some children and legislators may be socio-political butterflies that participate in every social group or sponsor every bill that passes by their desk, and thus observing their high levels of co-participation or co-sponsorship with others tells us little about who their friends and allies are. In such cases, unconditional thresholds are not suitable for inferring inter-agent relationships because they fail to consider agents’ differing degrees.

3.2. Agent-degree conditioned thresholds

Agent-degree conditioned thresholds aim to overcome the arbitrariness associated with unconditional thresholds by using agents’ degree to select threshold values, and they aim to overcome issues of structural bias and unisalcity by allowing these threshold values to vary for agents with different degrees. There are at least three distinct varieties of agent-degree conditioned threshold approaches: (a) those developed for natively one-mode networks, (b) those that involve normalizing edge weights in the bipartite projection, and (c) exact tests.

Many agent-degree conditioned thresholds were initially developed to extract the backbone from natively one-mode weighted networks. Nystuen and Dacey (1961) proposed one of the earliest such agent-degree conditioned threshold, suggesting that the backbone network be composed of each node’s single strongest edge. More sophisticated approaches allow a broader scope by examining a node’s distribution of edge weights, and including in the backbone any of a node’s edges whose weights exceed values expected in a null model (Puebla, 1987; Serrano et al., 2009). For example, Serrano et al. (2009) computed the probability that an edge’s observed weight is larger than the value expected in a null model where the node’s total edge weight was randomly distributed across each of its edges as

$$Pr(O_{ij} > Null_{ij}) = 1 - (D_i - 1) \int_0^{p_{ij}} (1 - x)^{D_j - 2}$$  \hspace{1cm} (2)

where $O_{ij}$ is the observed weight of the edge between agents $i$ and $j$, $D_i$ is agent $i$’s unweighted degree and $p_{ij}$ is the proportion of $i$’s total weighted degree carried by $O_{ij}$. They suggest retaining an edge in the backbone if it is viewed as statistically significant (at a given $\alpha$-level) from the perspective of either node, that is, if either $Pr(O_{ij} > Null_{ij})$ or $Pr(O_{ij} > Null_{ij}) < \alpha$. Although these agent-degree conditioned threshold methods were initially developed for extracting backbones from natively one-mode networks where edge weights are directly observed, some have since been used to extract backbones from bipartite projections as well (e.g., Ahn et al., 2011). However, they are only effective in strongly disordered networks where there is substantial variation in a given node’s edge weights (Serrano et al., 2009).

A second variety of agent-degree conditioned thresholds involves normalizing the edge weights in the bipartite projection in a way that controls for agents’ differing numbers of artifacts and thus transforms the edge weights into measures that assess “tendencies or revealed preferences to co-occur” (Borgatti and Halgin, 2011, p. 422). Two commonly used normalizations begin by viewing each pair of agents’ artifact profiles as a 2-by-2 contingency table (see Table 2).

From the values in Table 2, Bonacich (1972) suggested that the normalized weight of the edge between agents $i$ and $j$ be defined as

$$P_{ij} = \frac{0.5 if ad = bc, \ otherwise = \frac{ad - \sqrt{abcd}}{ad - bc}}{= \frac{0.5 if ad = bc, \ otherwise = \frac{ad - \sqrt{abcd}}{ad - bc}}{ad - bc}$$  \hspace{1cm} (3)

This rescales edge weights in the bipartite projection to range from 0, when agents $i$ and $j$ share the minimum possible number of artifacts, to 1, when they share the maximum possible number or artifacts, where the minimum and maximum values are given by Eq. (1). Others have suggested that $a + d$ is a useful normalization, which treats the absence of a given artifact among two agents as equally important to the sharing of a given artifact by two agents. When this measure is rescaled to range between −1 and 1, it is equal to the Pearson correlation between two agents’ column vectors of artifacts (Borgatti and Halgin, 2011). More common in computer science, mutual information measures represent still a third widely used similarity index, which Song et al. (2012) have shown to perform similarly to correlation measures. Whatever similarity index is used, once the normalized edge weights have been computed, the backbone can then be extracted by applying an unconditional threshold to the normalized weights. When critical values for the chosen similarity index can be determined from a theoretical probability distribution (e.g. for a Pearson correlation coefficient), they may serve as a guide for the selection of an unconditional threshold value, but in cases where critical values are unknown, selection of a threshold value may still be arbitrary (Neal, 2013).

A final agent-degree conditioned threshold approach relies on an exact test of the probability of two agents’ observed number of shared artifacts, conditioned on the agents’ degrees. The probability that $i$ and $j$ share exactly $x$ artifacts in a random agents-to-artifacts matching process that is constrained by agents’ degrees is given by a hypergeometric distribution

$$Pr(P_{ij} = x) = \frac{A \binom{A - x}{D_i - x} \binom{A - D_i}{D_j - x}}{\binom{A}{D_i} \binom{A}{D_j}}$$  \hspace{1cm} (4)

where $A$ is the total number of artifacts, and $D_i$ and $D_j$ are agent $i$’s and $j$’s degrees, respectively (Sudarsananam et al., 2002; Goldberg and Roth, 2003; Neal, 2013, 2013). A given edge is deemed statistically significant and retained in the backbone if $\sum_{x=\max(D_i,D_j)}^\min(D_i,D_j)Pr(P_{ij} = x) < \alpha$. The test models a kind of hypothetical market outcome. Each agent has a stock of resources (e.g. $D_i$) with which to purchase some artifacts from an available pool of $A$ artifacts. For example, different children have different amounts of free time with which they may participate in social groups, while different legislators have different levels of credibility with which they may sponsor bills. The exact test computes the proportion of all possible ‘spending’ patterns in which two agents co-purchased at least as many

<table>
<thead>
<tr>
<th>Artifact held by agent $i$?</th>
<th>Artifact held by agent $j$?</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Yes</td>
<td>a</td>
<td>b</td>
</tr>
<tr>
<td>No</td>
<td>c</td>
<td>d</td>
</tr>
</tbody>
</table>

---

**Table 2**  
Agent-by-artifact contingency table.
of the same artifacts as they were observed to have actually co-purchased.

Despite some similarities, the agent-degree conditioned approaches described by Eqs. (2) and (4) differ in three important ways. First, they differ in the data to which they are applied. Eq. (2) was developed for extracting the backbone of natively one-mode networks, and thus in the bipartite context must be applied to edge weights in the one-mode bipartite projection. In contrast, Eq. (4) was developed specifically for extracting the backbone of bipartite projections and is applied using information in the original (i.e. unprojected) bipartite data. Second, they differ in their view of edge weight values. Because Eq. (2) was developed for one-mode networks where edge weights can, in principle, be continuous variables, it constructs a probability density function and relies on integration to compute probabilities. In contrast, because Eq. (4) was developed for bipartite projections where edge weights are discrete counts of shared artifacts, it constructs a probability mass function and relies on summation to compute probabilities. Finally, they differ in their unit of analysis. Eq. (2) is a nodewise test that identifies a unique edge weight threshold for each node; Serrano et al. (2009) suggest preserving edges in the backbone if they are considered significant by one or both nodes linked by the edge. In contrast, Eq. (4) is an edgewise test that identifies a unique threshold for each edge by simultaneously considering the degrees of both agents linked by the edge.

Although agent-degree conditioned thresholds represent an improvement over unconditional thresholds, they nonetheless fall short because they implicitly treat artifacts as interchangeable. Because agent-degree conditioned threshold methods originally developed for one-mode networks (e.g. Nystuen and Dacey, 1961; Serrano et al., 2009) are applied to edge weights in the bipartite projection, details about artifacts are already lost. Similarly, the contingency table-based similarity measures and the exact test computation ignore which artifacts underlie agents’ degree. However, often artifacts are not interchangeable. For example, one’s attendance at an intimate social gathering provides more insight into who one’s friends may be than attendance at a large concert, and likewise one’s sponsorship of an unpopular piece of legislation provides more insight into where one’s political alliances lie than sponsorship of an uncontroversial or routine motion. In such cases, agent-degree conditioned thresholds are not suitable for inferring inter-agent relationships because they fail to consider artifacts’ differing degrees.

3.3. Dual-degree conditioned thresholds

Overcoming the limitations of unconditional and agent-degree conditioned threshold approaches requires a null model that identifies the distribution of expected edge weights that would be observed if agents were linked to artifacts randomly, but where this linking process is conditioned on (or constrained by) both the agents’ and artifacts’ degrees. The most widely used such model is the Fixed Degree Sequence Model (FDSM), which compares observed projection edge weights to the distribution of possible edge weights that might be observed if all agents’ and all artifacts’ degrees were fixed at their values in the empirical data (e.g. Guillaume and Latapy, 2003; Zweig and Kaufmann, 2011; Stegbauer and Rausch, 2012; Horvat and Zweig, 2013). No exact test of an edge weight’s statistical significance under the FDSM exists, so in practice this approach relies on a conditional uniform graph test. In general, a conditional uniform graph test offers a way of testing the statistical significance of a network statistic in an observed network by comparing it to the distribution of the same network statistic in a sample of random graphs that have been constructed to preserve certain features present in the observed network. In this case, the network statistic under consideration is the weight of the edge between two agents, and the features to be preserved in the random graphs are the agent and artifact degrees.

Fig. 1 outlines the steps in a conditional uniform graph test using the FDSM. First, the observed bipartite network (B) is projected onto a one-mode network (P, where P = BB T) to obtain the observed edge weight (P ij) between agents i and j. Next, a random bipartite network (B*) is generated according to the FDSM, such that the agent degrees (i.e. the row marginals of B*) and the artifact degrees (i.e. the columns marginals of B*) exactly match these values in the observed bipartite network (B). Finally, the random bipartite network (B*) is projected onto a random one-mode network (P*, where P* = B*B T) to obtain a random edge weight (P* ij) that is conditioned on fixed agent and artifact degrees. A large number, N, of random bipartite graphs are created and projected in this manner to obtain a conditional distribution of random edge weights. In the case of extracting a binary backbone, the edge between agents i and j is deemed statistically significant at the specified α-level and counted as present the backbone if P* ij > P ij in fewer than αN random bipartite projections. In the case of extracting a signed backbone, the edge between agents i and j is deemed statistically significant at the specified α-level and counted as positive the backbone if P* ij > P ij in fewer than (α/2)N random bipartite projections, and counted as negative if if P* ij < P ij in fewer than (α/2)N random bipartite projections.

In principle, the FDSM yields a distribution of expected edge weights that is conditioned on both agents’ and artifacts’ degrees, but in practice this approach to backbone extraction encounters both methodological and conceptual problems. First, those recommending the approach often “assume that there is a method with which we can sample uniformly at random from [the population of bipartite networks with fixed row and column marginals]” (Zweig and Kaufmann, 2011, p. 193), but such an assumption may not be warranted. To be sure, several different algorithms exist, but “a truly usable algorithm remains elusive” (Bezakova, 2008, p. 31). A recently proposed simulated annealing algorithm is impractically complex with worst-case running times like O(n 11 log 5 n) for an n x n matrix (Bezakova et al., 2007), while a configuration model algorithm is more efficient but does not always yield valid results (e.g. Blanchet and Stauffer, 2012). In the case of Markov chain algorithms that involve swapping edges in the observed bipartite network to obtain a random bipartite network, the number of edge swaps necessary to ensure that graphs are sampled uniformly at random from all possible bipartite graphs remains unknown (Gionis et al., 2007). Second, the FDSM fixes the agent and artifact degrees of the random bipartite graphs at exactly the values in the observed bipartite network, but offers no conceptual rationale for such stringent conditioning. The logic underlying the conditional uniform graph test is that each random bipartite graph represents an alternate possible realization of agent-artifact linkages. The FDSM views it as possible for agents to have been linked to different artifacts, but not possible for agents to have been linked to different numbers of artifacts. However, this seems to go too far. For example, although a legislator may be observed to have sponsored 100 bills, in an equally plausible alternate world driven by the same political forces, the same legislator might have sponsored 99 bills, or 101 bills. Thus, even if uniformly randomly sampling from fixed-degree random bipartite networks were easy, doing so risks over-conditioning or imposing too many assumptions on the null model.

4. The stochastic degree sequence model (SDSM)

As Table 1 summarizes, unconditional thresholds, agent-degree conditioned thresholds, and dual-degree conditioned thresholds using the FDSM have been used throughout the literature to extract
the backbone of bipartite projections, but each suffers from its own set of disadvantages. In this section, I propose the stochastic degree sequence model (SDSM) as an alternative to the FDSM that overcomes many of these disadvantages. Specifically, it offers a backbone extraction approach that conditions on both agent and artifact degrees, but that is relatively computationally simple and relaxes the stringency of the conditioning. As Fig. 1 highlights, the SDSM offers an alternative to the FDSM for generating the relevant sample of conditionally random bipartite networks, but is still employed within a conditional random graph test framework.

Accordingly, most of the steps in the process remain the same; only the procedure used to obtain random bipartite networks is different (i.e. step 2). Notably, unlike the FDSM, an exact test of an edge weight’s statistical significance under the SDSM is possible, however such a test is computationally impractical in most cases; details are provided in Appendix A.

The SDSM views the observed bipartite network \( (B) \) as one of many possible outcomes of an unobserved, stochastic process of agent-to-artifact matching. This process is driven by a set of probabilities that describe the likelihood that a given agent will be linked...
to a given artifact (c.f. McCulloh et al., 2010). For example, there is some probability \( p_{ik} \) that agent \( i \) will be linked to artifact \( k \), and another probability \( p_{jk} \) that agent \( j \) will be linked to artifact \( k \), and so on for all agents and artifacts. Agents’ and artifacts’ degrees impose constraints on these probabilities. Thus, other things being equal, a higher-degree agent has a higher probability of being linked to any artifact, and a higher-degree artifact has a higher probability of being linked to any agent. In the set of bipartite networks that can arise from this stochastic process, each agent’s and artifact’s degree is not fixed, but rather lies within a distribution that is a product of the matching process.

The goal of the SDSM is to estimate these probabilities, then use them in the construction of random bipartite networks, which is outlined as a series of three steps in Fig. 1. In the first step of the SDSM (step 2A in the binary), a unary outcome model is used to predict the observed agent-artifact links as a function of the agent’s degree, the artifact’s degree, and their product (i.e. \( B_k = \beta D_i + \beta D_j + \beta D_k D_l \)). Next, the coefficients from the fitted model are used to compute the probability of a linkage \( p_{ij} \) between each agent-artifact pair. Finally, a random bipartite network \( (B^*) \) is constructed in which the value of \( B^*_{ij} \) is the outcome of a single Bernoulli trial with probability \( p_{ij} \) of success.

Many different binary outcome models might be used in step 2A to estimate the agent-artifact linkage probabilities, including the well known logit and probit models, and the less common scobit and complementary log–log models. The most appropriate model will be the one that yields random bipartite graphs where the agent and artifact degrees most closely match the observed values. Using the bill co-sponsorship data described in the next section, Fig. 2 compares observed agent and artifact degrees to the agent and artifact degrees in random bipartite network generated using agent-artifact linkage probabilities estimated by different binary outcome models. The solid line indicates a perfect match between degree in the observed and random bipartite networks; the preferred model should yield degree sequences with the least deviation from this line, which can be measured using the root mean squared error (RMSE). As these comparisons illustrate, none of the commonly used binary outcome models yield agent-artifact linkage probabilities that produce random bipartite networks in which agent and artifact degrees perfectly match their observed degrees. However, the scobit model yields consistent superior results (i.e. the smallest RMSE) and thus is the most appropriate estimation technique for the SDSM in these data; other binary outcome models may be more appropriate in other data. Although not widely used, scobit is simply a generalized form of the more traditional logistic model that relaxes the assumption that the independent variables have the greatest impact on cases with an initial probability of 0.5 (Nagler, 1994). The -onemode- command in Stata will extract the backbone from a bipartite projection using the SDSM with the optimal binary outcome model.

5. Example: bill co-sponsorship in the 108th U.S. Senate

Given the influence legislators have over our lives through their ability to make laws, there is substantial interest in the network of political alliances that exist among them. Political scientists, the media, and constituents want to know who influences whom when it comes to their elected officials. Direct collection of network data about political alliances is very challenging for many reasons, chief among which is the fact that to preserve their ability to maneuver in complex political negotiations, legislators may decline to provide such information or may provide only a carefully constructed version of reality. Thus, those interested in understanding networks of political alliance among legislators have needed to look elsewhere (Fowler et al., 2011). Bipartite projections have offered a promising methodological tool for inferring network ties between legislators using readily available information. The most widely adopted approach has been to examine legislators’ sponsorship of bills, although others have also considered legislators’ membership on committees (Porter et al., 2005).

As Fowler (2006a) explains, “Since 1967 in the U.S. House and the mid-1930s in the U.S. Senate, legislators have had an opportunity to express support for a piece of legislation by signing it as a co-sponsor” (p. 454). A range of technical distinctions can be made in this practice, for example between the bill’s sponsor who introduces it for consideration and its cosponsors who subsequently sign on to the bill in support, or in the order in which cosponsors sign on. For the sake of simplicity, I do not make such distinctions in this demonstration: all legislators who formally express support for a given bill are viewed as sponsors. Those adopting the bipartite projection as a proxy measure assume that observing two legislators sponsoring the same bill provides some evidence for the existence of a relationship of alliance between them. The challenge, however, lies in determining how many bills a given pair of legislators must co-sponsor before an inference that they are political allies is warranted. This is a clear case when unconditional thresholds fail us: setting the threshold too low leads us to conclude that all legislators have alliances with all other legislators, while setting it too high leads us to conclude that there are no political alliances. Two key pieces of information must be taken into account when inferring the existence of a political alliance from bill co-sponsorship activity: the total number of bills each legislator has sponsored (i.e. agents’ degrees), and the total number of legislators that have sponsored each bill (i.e. artifacts’ degrees).

To demonstrate the proposed backbone extraction method, I turn to data on the 108th U.S. Senate (2003–2005), initially analyzed by Fowler (2006a) and available at http://jhfowler.ucsd.edu/cosponsorship.htm. These data offer a useful test case for several reasons. First, aspects of their network structure have been repeatedly analyzed in the pages of Social Networks and other journals (e.g. Fowler, 2006a,b; Zhang et al., 2008; Cho and Fowler, 2010). Second, these data are publicly available, allowing for replication of the results shown below. Finally, and perhaps most importantly, legislators’ political party affiliations and ideological alignments provide an exogenous ground-truth against which the extracted backbone networks can be tested for accuracy. Specifically, political alliances should be most common between senators who belong to the same political party and/or share ideological stances, and should be least common between senators who belong to different parties and/or have different ideological stances.

5.1. Step 1: the observed bipartite network

The bipartite network describes 100 senators’ sponsorship of 7804 bills. There is substantial variation in senator’s degrees \( M = 363, sd = 162 \). For example, Sen. Clinton (D-NY) sponsored the most bills at 827, while Sen. Shelby (R-AL) sponsored the fewest at 126. Likewise, there is substantial variation among bills’ degrees \( M = 5, sd = 9 \). For example, every senator sponsored S.Res. 110 “A resolution honoring Mary Jane Jenkins Ogilvie, wife of former Senator Chaplin, Reverend Dr. Lloyd John Ogilvie”, while over 3000 bills including S.Con.Res 105 “A concurrent resolution designating the second week of March 2005 as Extension Living Well Week” were sponsored by only one senator.

From the observed bipartite network, a bipartite projection can be obtained. In this case, the bipartite projection is a one-mode network among senators in which edges are weighted by the number of bills co-sponsored by a pair of senators. For example, Sen. Clinton and Sen. Shelby co-sponsored (i.e. both signed as sponsors) 28 bills, and thus are connected in the bipartite projection by an edge
Fig. 2. Observed vs. stochastic degrees.
Table 3
Scobit coefficient estimates for 108th U.S. Senate.

<table>
<thead>
<tr>
<th>Coefficient</th>
<th>Estimate</th>
</tr>
</thead>
<tbody>
<tr>
<td>Agent degree</td>
<td>-0.0098</td>
</tr>
<tr>
<td>Artifact degree</td>
<td>0.0004</td>
</tr>
<tr>
<td>Agent degree × artifact degree</td>
<td>0.0014</td>
</tr>
<tr>
<td>Constant</td>
<td>-1.4421</td>
</tr>
<tr>
<td>α-Parameter</td>
<td>0.0317</td>
</tr>
</tbody>
</table>

with a weight of 28. The goal of backbone extraction is to determine how large this weight must be to infer that Sen. Clinton and Sen. Shelby are politically antagonistic. To make such decisions, it is important to consider both senators' degrees and bill's degrees. There are a few clear-cut cases. A political alliance is likely to exist between a pair of senators who sponsor just a few bills each, but where those few co-sponsored bills were very unpopular. These are senators who are proverbially 'sticking their neck out' for each other. Similarly, it is likely that a political antagonism exists between a pair of senators who each sponsor many bills, but where they are only seen co-sponsoring uncontroversial bills that all other senators co-sponsored as well. These are senators that may jump on the bandwagon of popular bills, but otherwise avoid one another. The SDSM aims to assist with inferences in cases that lie between these two extremes.

5.2. Step 2: generating a random bipartite network using the SDSM

The construction of a random bipartite network under the SDSM begins by fitting a binary outcome model. In this context, each senator-bill pairing is a separate observation. The binary dependent variable, whether the senator sponsored (1) or did not sponsor (0) the bill, is predicted using three independent variables: the senator's degree, the bill's degree, and the product of these two (i.e. an interaction term). Although a logit model is most commonly used to predict binary outcomes, as Fig. 2 demonstrated, the related scobit model is more appropriate here. The coefficients estimated by the scobit model are shown in Table 3, which includes a coefficient for each of the independent variables as well as a constant term and an α-parameter coefficient that is unique to the scobit model. The values and statistical significance of these coefficients are not of direct interest.

Instead, these coefficients are 'plugged' into the scobit equation along with the values of the independent variables for a specific senator-bill pairing to obtain the estimated probability that the given senator will sponsor the given bill, conditional on (i.e. controlling for) the senator's and bill's degree. An example from these data serves to illustrate. Sen. Clinton's degree is 827 (she sponsored 827 bills), S.Res 110's degree is 100 (all 100 senators sponsored this bill), and the product of these two independent variables is 82,700. Placing these values (shown in plainface) and the estimated coefficients (shown in boldface) back into the scobit equation

\[ 1 - \frac{1}{(1 + \exp(z))^{0.0317}} \cdot \text{where} \quad z = -1.4421 + (-0.0098 \times 827) + (0.0004 \times 100) + (0.0014 \times 82,700) \]

(5)
gives a probability of 0.96. That is, taking into account Sen. Clinton's propensity to sponsor bills, and taking into account senators' propensities to sponsor S.Res 110, there is a 96% chance that Sen. Clinton will sponsor S.Res. 110. Repeating Eq. (5) for each senator-bill pairing yields a series of estimated probabilities. For example, based on the scobit-estimated coefficients, we find that the high-degree Sen. Clinton almost certainly will sponsor the popular S.Res 110 (Pr = 0.9688), while the low-degree Sen. Shelby is somewhat less likely to sponsor it (Pr = 0.3708). Similarly, we find that Sen. Clinton likely will not sponsor the unpopular S.Con.Res 105 (Pr = 0.02119), but Sen. Shelby almost certainly will not sponsor it (Pr = 0.0081).

Finally, these senator-bill sponsorship probabilities are used to generate a random bipartite network. In the random bipartite network, whether or not a given senator sponsors a given bill is the result of a single Bernoulli trial (e.g. a single coin flip) where the probability of success (e.g. the probability of 'heads') is the probability estimated using the procedure illustrated in Eq. (5). Thus, for example, in the construction of the random bipartite network, there is a 96% chance that Sen. Clinton will sponsor S.Res 110, and a 4% probability she will not.

Fig. 3 illustrates the degrees of two senators and the degrees of two bills in 10,000 random bipartite networks generated using this approach. The dashed lines in each plot show the degree in the observed bipartite network (i.e. the actual number of bills sponsored and actual number of senators sponsoring). Senate Bill S.B. 1157 was a widely sponsored bill called the “National Museum of African American History and Culture Act” that was introduced by Sen. Brownback (R-KS); it passed in the Senate but not in the House of Representatives. Senate Bill 2271 was a bill with more limited sponsorship called the “Clean Cruise Ship Act of 2004” that was introduced by Sen. Durbin (D-IL); it died in committee and has been re-introduced by Sen. Durban multiple times since. These plots highlight the proposed method’s stochastic degree characteristics. Rather than fixing each senator’s and bill’s degree at its value in the observed bipartite network, these values vary within distributions that are roughly symmetric and centered on their observed values. For example, although Sen. Clinton in fact sponsored 827 bills, the same stochastic process she and other senators followed to select bills to sponsor might have led her to sponsor, say, 800 bills instead. Recall, however, that although each degree is allowed to vary, as Fig. 2 demonstrated, in the aggregate they closely approximate the observed valued.
5.3. Step 3: project the random bipartite network

The bipartite projection of one of these random bipartite networks yields a random one-mode network in which the weight of the edges indicate the number of bills a pair of senators co-sponsored in an alternate world. This alternate world is still a plausible one, however, because in it each senator still sponsors about the same number of bills as the observed world, and each bill is sponsored by about the same number of senators as the observed world. That is, these alternate worlds have been conditioned on senators’ and bills’ degrees. Examining the weight of the edge between a pair of senators in a large number of these random bipartite projections provides a probability distribution of co-sponsorships for the pair of senators, under a null model that has been conditioned on senators’ and bills’ degrees. That is, it indicates how many and how few bill co-sponsorships we might expect for a given pair of senators, if the senators’ bill sponsorship activities were conditionally random. Fig. 4 illustrates these probability distributions for three different pairs of senators; the dashed lines in each plot shows the pair’s actual number of co-sponsored bills. For example, in the observed data, Sen. Clinton and Sen. Schumer (D-NY) actually co-sponsored 407 bills, while in the alternate worlds described by the random bipartite projections, they tended to co-sponsor only between 200 and 275 bills.

5.4. Step 4: hypothesis testing

In the final step, these probability distributions provide the information necessary to apply a conventional hypothesis testing approach to decide whether a pair of senators co-sponsored enough bills for a political alliance to be inferred, or if they co-sponsored few enough bills for a political antagonism to be inferred. In the left panel of Fig. 4, it is clear that Sen. Clinton and Sen. Schumer (D-NY) co-sponsored dramatically more bills (407) than might be expected at random (i.e. than they co-sponsored in the random projections). Indeed, in only 1 of the 10,000 alternate worlds described by the random bipartite networks did they co-sponsor more than 407 bills. Because the number of co-sponsorships between Sen. Clinton and Schumer in the random networks \( P_h \) was greater than their observed number of co-sponsorships \( P_h \) in fewer than 2.5% (i.e. 250 = (0.05/2)10,000) of the random projections, their observed number of bill co-sponsorships (i.e. their edge weight in the bipartite projection) is statistically significantly greater than random at the \( \alpha = 0.05 \) level. Substantively, we might infer from this that Sen. Clinton and Sen. Schumer have a political alliance, which is not surprising given that both senators are democrats from New York.

In the center panel, although Sen. Shelby and Sen. McConnell (R-KY) co-sponsored only 26 bills, this was still dramatically greater than would be expected at random, and following a similar logic provides strong evidence of a political alliance between them. Again, such an inference is not surprising given that both are Republican senators from traditionally conservative Southern states, and that McConnell was the Majority Whip responsible for gathering votes on key issues. Comparing the Clinton–Schumer and Shelby–McConnell cases highlights the danger of unconditional thresholds: inferring political alliances via an unconditional threshold would likely identify the Clinton–Schumer alliance with its large number of co-sponsorships, but would likely miss the Shelby–McConnell alliance with its small number of co-sponsorships.

The right panel illustrates a notably different case, and. Sen. Clinton and Sen. Shelby are seen co-sponsoring a significantly smaller number of bills (28) than would be expected at random. This provides strong evidence that there is not a political alliance between these two senators. Again, this is an unsurprising inference given that one is a liberal Northern democrat and the other is a conservative Southern republican. However, beyond simply suggesting the absence of a political alliance, this might be viewed as evidence of a political antagonism and worth preserving as a negative edge in the backbone network. From a formal hypothesis testing perspective, this conclusion is warranted because the number of co-sponsorships between Sen. Clinton and Sen. Shelby in the random networks \( P_h \) was smaller than their observed co-sponsorships \( P_h \) in fewer than 2.5% (i.e. 250 = (0.05/2)10,000).

5.5. The backbone of the 108th U.S. Senate

Using edge weight probability distributions like those in Fig. 4, a complete backbone of the bipartite projection can be extracted by retaining only edges that are statistically significant at a given \( \alpha \)-level. Fig. 5 shows the backbone of the 108th U.S. Senate extracted at the \( \alpha = 0.05 \) level. In the figure, only edges whose weights were statistically significantly greater than random expectations are shown, thus the network shows positive edges indicating political alliances. The nodes are positioned using the multi-scale layout developed by Harel and Koren (2002) and implemented in NodeXL (Smith et al., 2010); parallel edges have been bundled for visual clarity. Although the choice of a specific value for \( \alpha \) is arbitrary (0.05 and 0.01 are common disciplinary conventions), here it offers a very precise interpretation of the backbone network: a pair of senators is linked in this backbone network if there is a less than 2.5% chance (because this is a two-tailed test of statistical significance) that they co-sponsored more bills than expected if their bill sponsorship activities were constrained only by their degrees and the bills’ degrees, but were otherwise random. Thus, the backbone edges indicate pairs of senators between whom a political alliance might reasonably be inferred to exist. As Fig. 6 highlights, there is no direct relationship between an edge’s weight in the bipartite projection
and its inclusion in the backbone network: edges deemed statistically significantly greater than random and thus included have weights ranging from 22 to 407, while those deemed not statistically significantly different from random have weights ranging from 13 to 244, and those deemed statistically significantly less than random have weights ranging from 13 to 173. This confirms that using an unconditional threshold to extract a backbone would not have been appropriate in this case; sometimes a large number of co-sponsorships was significant, and sometimes not.

The backbone network exhibits a high degree of face validity, clearly highlighting the 108th Senate’s even split between 48 Democrats clustered on the left and 51 Republicans clustered on the right. Comparison of the backbone with two ground-truth indicators of political alliance – party membership and DW-NOMINATE ideology scores – further confirm its validity. Members of the same political party should be more likely to form political alliances than members of opposing political parties, and thus the community structure of a network reflecting political alliances should closely match political party membership (Zhang et al., 2008). Confirming this expectation, a modularity-maximizing partition places the senators into two primary groups that match political party memberships in about 90% of cases (Modularity = 0.317; Girvan and

Fig. 5. Backbone of the 108th U.S. Senate ($p < .025$).
Mismatches may occur because either the backbone network or political party membership does not accurately reflect political alliances, but closer inspection suggests the backbone network is more accurate. For example, Susan Collins was a Republican senator from Maine, but is placed in the predominantly Democrat network community, reflecting her status as “the last survivors of a once common species of moderate Northeastern Republican” (Tumulty and Newton-Small, 2009). Similarly, during the 108th Senate (2003–2005), Lincoln Chafee was a Republican senator from Rhode Island, but is also placed in the predominantly Democrat network community, foreshadowing his shift in political affiliation to Independent in 2007, and eventually to Democrat in 2013.¹

Political ideology, which influences but is nonetheless distinct from political party, is also likely to influence the formation of political alliances between legislators. Specifically, political alliances should be more likely among those with similar ideology, and less likely among those with different ideologies. Thus, in a network reflecting senatorial political alliances, the geodesic distance between two senators should be related to the dissimilarities in their ideological stances. As a binary network, geodesic distances are straightforward to compute in the backbone network, and here range from 1 (i.e. pairs of senators with a political alliance) to 3 (i.e. pairs of senators who lack political alliances with a common third senator). Legislators’ ideological positions are commonly measured using DW-NOMINATE scores, which identify legislators’ positions along two ideological dimensions through iterative multidimensional scaling of roll-call voting data. The first dimension is often interpreted as reflecting liberal-conservative ideology, and thus legislators’ distance along this dimension provides a useful measure of their ideological dissimilarity (Poole and Rosenthal, 2007).²

As expected, senators that are closer in the backbone network have more similar ideologies: senators that are directly linked have a mean ideological dissimilarity of 0.20, those separated by two steps in the network have a mean ideological dissimilarity of 0.51, and those separated by three steps in the network have a mean ideological dissimilarity of 0.79 (mean differences between these groups are all significant at p < 0.001). These political party and ideology comparisons suggest that the proposed method accurately extracts relationships that capture likely political alliances in the U.S. Senate, and thus that the extracted backbone can be used as a proxy for a one-mode network of political alliance that is not directly measurable.

5.6. Negative ties and political antagonisms

The discussion above has focused on identifying edges with weights that are significantly greater than values expected if agents were matched to artifacts randomly, controlling for their degrees. In the legislative context, this amounts to extracting a backbone of political alliances. However, because the statistical test used in the proposed backbone extraction method is two-tailed, it can also be used to identify edges with weights that are significantly less than random expectations. In the legislative context, such edges might capture political antagonisms, like the one illustrated in the right panel of Fig. 4, between Sen. Clinton and Sen. Shelby. The political party and ideology comparisons above can be used to evaluate the validity of these antagonistic backbone ties also.

Community detection algorithms aim to partition nodes into groups such that the majority of edges are within, rather than between, groups. However, a much more robust notion of communities that takes into account edge valence would aim to partition nodes into groups such that the majority of positive edges are within groups, and the majority of negative edges are between groups (e.g. Traag and Bruggeman, 2009). In the legislative context, such a partitioning would reflect political alliances among those with shared ideologies, but also political antagonisms among those with conflicting ideologies. Table 4 summarizes the organization of political alliances and antagonisms inferred by the backbone extraction method, by the two primary communities discussed above. Recall that the community detection algorithm used here only considers political alliances. Thus, the first row of this table simply confirms the high modularity value reported above, and highlights that the community partition obtained through modularity maximization does indeed place most political alliances within communities and relatively few between communities. The values in the bottom row concern ties of inferred political antagonism, which were not used to detect communities, but which do exhibit the expected pattern: antagonistic relations exist primarily between communities, not within them. Returning to the DW-NOMINATE scores, it is also possible to compare the ideological similarity of senators with and without politically antagonistic relationships. As expected, pairs of senators linked by political antagonism in the backbone network have a significantly greater mean ideological dissimilarity (0.71) than pairs of senators that are not linked by political antagonism (0.36). Together, these findings suggest that proposed method also accurately extracts relationships that capture likely political antagonisms in the U.S. Senate, and thus that the extracted backbone can be used as a proxy for a one-mode network of political antagonism that is not directly measurable.

6. Conclusion

There are many contexts where analysis of a one-mode network among a set of agents could help explain complex phenomena, but where collection of this type of data is impractical or impossible (e.g. social relationships among children, economic relationships among cities, political relationships, among legislators). One promising solution involves inferring these types of relationships from bipartite projections, which can be constructed from data that is often much easier to collect (e.g. group participation, firm location, bill sponsorship). To be viable, however, a method for deciding when such inferences are warranted – that is, of deciding when an edge in the bipartite projection should be retained in the backbone – is necessary. Several methods have been proposed, but each has shortcomings that limit its usefulness. For example, unconditional thresholds distort the structure of the network, while the fixed degree sequence model (FDSM) is computationally complex and risks over-conditioning the null model.

The goal of this paper has been twofold: to briefly summarize and compare the existing methods for extracting the backbone of bipartite projections, and to propose a new method that overcomes

<table>
<thead>
<tr>
<th>Alliance</th>
<th>Within</th>
<th>Between</th>
</tr>
</thead>
<tbody>
<tr>
<td>Antagonism</td>
<td>1201</td>
<td>46</td>
</tr>
<tr>
<td></td>
<td>99</td>
<td>431</td>
</tr>
</tbody>
</table>

¹ Similar stories can be told about other party–community mismatches. Vermont senator James Jeffords identified as an Independent but appears in the Democrat network community; he chose to caucus with Senate Democrats. Georgia senator Zell Miller identified as a Democrat but appears in the Republican network community; he supported George Bush in the 2004 presidential election and sided with Republicans so often that the Associated Press asked in 2004 “Has Zell Miller lost his mind?”
² In the past, the second dimension has captured differences on civil rights issues between the North and South, but is no longer viewed as important. All DW-NOMINATE data are available online at http://voteview.com/s.
their primary limitations. The proposed stochastic degree sequence model (SDSM) offers a useful alternative to the FDSM that relaxes the requirement that agents’ and artifacts’ degrees be strictly fixed in the null model, and in doing so simplifies the construction of the random bipartite graphs used in the conditional uniform graph test. By allowing agents’ and artifacts’ degrees to vary, the null model more closely mirrors variations that might emerge from stochastic matching processes. This model can be used to extract not only a binary backbone, but also a signed backbone that captures edge valence. Thus, it can be used to infer both positive and negative relationships between agents from bipartite data. Applying this method to data on bill sponsorship activities in the 108th U.S. Senate yields a backbone network of political alliances and antagonisms that, based on a close correspondence to a ground-truth furnished by political party affiliations and ideological positions, exhibits a high level of criterion validity.

Although this method offers an improvement over existing unconditional threshold, agent-degree conditioned threshold, and dual-degree conditioned backbone extraction methods, two challenges remain in the use of bipartite projections to infer one-mode networks. First, these methods do not account for the possibility that some observed agent-to-artifact matchings (or non-matchings) in the bipartite data may be fixed. For example, perhaps a given senator is legally prohibited from sponsoring a given bill due to a conflict of interest (i.e. a structural 0 in the bipartite data), or a given firm is legally required to maintain an office in a given city under its articles of incorporation (i.e. a structural 1 in the bipartite data). In such cases, ad-hoc adjustment of the estimated agent-by-artifact probabilities estimated in step 2b of the SDSM may be sufficient, but such adjustments will alter the agent and artifact degree distributions in the sampled bipartite networks. Second, because these methods can in principle be applied to any rectangular binary matrix, its use to infer a network of relationships among agents requires a strong theoretical rationale. For example, there are good theoretical reasons to believe that the co-sponsorship of bills hints at political alliance, and that firm co-location in cities hints at the spatial movement of information and capital. But, there are likely no good theoretical reasons to believe that co-coloration would hint at relationships between fruits; apples and cherries may both be red, but they likely do not have a ‘relationship.’ However, in cases where there is a theoretical rationale for shared artifacts reflecting an underlying relationship, and where each agent could at least in principle have been matched to each artifact, the SDSM proposed in Section 4 provides a promising way to measure one-mode networks in contexts where such data would otherwise be unavailable.

Appendix A. Exact test of edge weight probabilities under the SDSM

Whether using the FDSM or SDSM in a conditional uniform graph test, one of the most computationally intensive steps is the projection of each of the random bipartite graphs, which runs in $O(n^2m)$ time. Thus, a method for directly identifying edge weight distributions and their dual-degree conditioned thresholds, without the need for projecting each randomly sampled graph, could be useful. In principle, direct identification of exact SDSM thresholds from the probability matrix is possible, but in practice is more computationally intensive than simply performing the projections. Nonetheless, for the sake of completeness, I sketch the outlines of an exact test corresponding to the SDSM.

Let $p_{ik}$ and $p_{jk}$ be the probability that agent $i$ and agent $j$, respectively, are linked to the $k$th artifact under the SDSM. These probabilities are obtained using the fitted binary outcome model coefficients described above. The probability that two agents, $i$ and $j$, share no artifacts in common is

$$
(1 - p_{iz}p_{jz})(1 - p_{iz}p_{jy})(1 - p_{iz}p_{jy}) = \prod_{i=1}^{k}(1 - p_{ik}p_{jk})
$$

(6)

and the probability that two agents share all artifacts in common is

$$
(p_{iz}p_{jz}) \cdot (p_{iz}p_{jy}) \cdot \cdots \cdot (p_{iz}p_{jy}) = \prod_{i=1}^{k}(p_{ik}p_{jk})
$$

(7)

Computing the probability that two agents share some exact number of artifacts is more complicated. For example, the probability that two agents, $i$ and $j$, share exactly one artifact in common, when there are a total of three artifacts, is

$$
(p_{iz}p_{jz})(1 - p_{iz}p_{jy})(1 - p_{iz}p_{jy}) + (1 - p_{iz}p_{jz})(p_{iz}p_{jy}) + (1 - p_{iz}p_{jz})(1 - p_{iz}p_{jy})(p_{iz}p_{jy})
$$

(8)

Here, the first line captures the probability that they share artifact 1, but not 2 or 3; the second line captures the probability that they share artifact 2, but not 1 or 3; and the third line captures the probability that they share artifact 3, but not 1 or 2. Expressed in more general form:

$$
\sum_{i=1}^{x}(p_{iz}p_{jz}) \prod_{i=1}^{k}(1 - p_{ik}p_{jk}) \quad \text{where } x \neq k
$$

(9)

Likewise, the probability that two agents, $i$ and $j$, share exactly two artifacts in common, when there are a total of three artifacts, is

$$
(p_{iz}p_{jz})(p_{iz}p_{jy}) + (1 - p_{iz}p_{jz})(p_{iz}p_{jy}) + (1 - p_{iz}p_{jz})(1 - p_{iz}p_{jy})(p_{iz}p_{jy})
$$

(10)

Here, the first line captures the probability that they share artifacts 1 and 2, but not 3; the second line captures the probability that they share artifact 1 and 3, but not 2; and the third line captures the probability that they share artifact 2 and 3, but not 1. Expressed in more general form:

$$
\sum_{i=1}^{x}\sum_{j=1}^{y}(p_{iz}p_{jz})(p_{iz}p_{jy}) \prod_{i=1}^{k}(1 - p_{ik}p_{jk}) \quad \text{where } x < y \text{ and } x \neq y \neq k
$$

(11)

In general, computing the probability that two agents share exactly $n$ of $A$ artifacts involves summing $A \choose n$ sets of products. In a bipartite network with even a relatively small number of artifacts, the number of operations required to identify exact threshold values for every pair of agents is extremely large.

References


