Network Science (2023), 1–14
doi:10.1017/nws.2023.3

CAMBRIDGE UNIVERSITY PRESS

RESEARCH ARTICLE

The duality of networks and groups: Models to generate two-mode networks from one-mode networks

Zachary P. Neal

Michigan State University, East Lansing, MI, USA
Email: zpneal@msu.edu

Action Editor: Ulrik Brandes

Abstract

Shared memberships, social statuses, beliefs, and places can facilitate the formation of social ties. Two-mode projections provide a method for transforming two-mode data on individuals' memberships in such groups into a one-mode network of their possible social ties. In this paper, I explore the opposite process: how social ties can facilitate the formation of groups, and how a two-mode network can be generated from a one-mode network. Drawing on theories of team formation, club joining, and organization recruitment, I propose three models that describe how such groups might emerge from the relationships in a social network. I show that these models can be used to generate two-mode networks that have characteristics commonly observed in empirical two-mode social networks and that they encode features of the one-mode networks from which they were generated. I conclude by discussing these models' limitations and future directions for theory and methods concerning group formation.

Keywords: bipartite; Blau space; generative model; group formation; organizations; projection; team formation

1. Introduction

A natural question in the social networks literature has been: *Where do social networks come from?* The answers have been diverse, and contributions have taken the form of both theoretical propositions for underlying mechanisms such as homophily (e.g., McPherson et al., 2001) and statistical frameworks for testing these propositions (e.g., Robins et al., 2007; Snijders et al., 2010). Some have proposed that social networks come from groups such as parties or clubs that present opportunities for individuals to meet and form ties by focusing social activity (Feld, 1981). However, this raises the obvious question: *Where do such groups come from?*

There is a duality of social networks and groups, such that networks can emerge from groups, but groups can also emerge from networks. A sketch of this duality was already present in the initial articulation of focus theory (Feld, 1981). However, most subsequent work has examined how networks emerge from groups, while neglecting how groups emerge from networks. In this paper, I aim to elaborate the second half of this duality. Drawing from a range of disciplinary contexts, I develop three models for how groups can emerge from social networks: as teams (Guimera et al., 2005), as clubs (Backstrom et al., 2006; Schaefer et al., 2022), and as organizations (McPherson, 2004). While these models offer insight into how groups can emerge from networks, they also contribute to the methodological literature as two-mode network generative models, which currently "are practically non-existent" in the literature (Filho & O'Neale, 2020a, p. 3).

The remainder of the paper is organized in five sections. In Section 2, I briefly review the theories and methods available for understanding how networks and groups co-evolve. Then, in

© The Author(s), 2023. Published by Cambridge University Press. This is an Open Access article, distributed under the terms of the Creative Commons Attribution licence (http://creativecommons.org/licenses/by/4.0/), which permits unrestricted re-use, distribution and reproduction, provided the original article is properly cited.

https://doi.org/10.1017/nws.2023.3 Published online by Cambridge University Press

Table 1. The evolution and formation of networks and groups

Process	Representative theories and methods
Network → Network	Balance, Homophily, Preferential attachment, Reciprocity, Transitivity (Fuhse & Gondal, 2022; Yap & Harrigan, 2015); Temporal exponential random graph models (TERGM; Krivitsky & Handcock, 2014) and Stochastic actor oriented models (SOAM; Snijders et al., 2010)
Groups → Groups	Diffusion of innovation theory (Rogers, 2003); Bipartite ERGM (BERGM; Wang et al., 2013) and SOAM (Schaefer et al., 2022)
Groups → Network	Focus theory (Feld, 1981); Two-mode projections (Breiger, 1974) and backbones (Neal, 2014)
Network → Groups	Mentioned but not developed in focus theory (Feld, 1981); "generative models are practically non-existent" (Filho & O'Neale, 2020a, p. 3)

Section 3 I introduce three models for how groups can emerge from social networks. In Section 4, I use simulations to show that these models can be used to generate two-mode networks that have characteristics commonly observed in empirical two-mode social networks and illustrate how the generated two-mode networks encode features of the one-mode network from which they were generated. Finally, in Section 5 I conclude by considering these models' limitations and their potential applications for building and testing both theories and methods.

2. Background

Wellman (1988) warned that "the world is composed of networks, not groups" (p. 37). This claim may have gone too far, but did highlight that networks and groups are different. A group is a collection of individuals who might be socially cohesive or not; its internal social structure is unspecified. In contrast, a network is a social structure among individuals who might cluster into discrete sets or not; its members' categorical affiliations are unspecified. Accordingly, it is possible to study just networks, just groups, or how they influence each other. Table 1 shows how the formation of networks and groups has been studied and highlights where the present study fits among these lines of research.

Within the field of (social) network analysis, perhaps the most widely studied formative process involves the evolution of networks (Network → Network). Many mechanisms have been hypothesized to explain how network ties form, and how ties that are present or absent at time 1 impact the presence or absence of ties at time 2 (Fuhse & Gondal, 2022; Yap & Harrigan, 2015). For example, ties may form and networks may evolve through a process of preferential attachment, such that new ties tend to be formed with already well-connected others. Ties may also form through processes of balance that promote friendship cycles (i.e., A → B → C → A) or of status seeking that prohibit them. Among the most intuitive tie formation processes are those that unfold when individuals share something in common. Ties can form through homophily when individuals share an interest or demographic characteristic, through propinquity when they share a space, or through transitivity when they share a set of common friends. While theories of tie formation are well developed, so too are formal statistical methods for modeling the evolution of networks, which include temporal exponential random graph models (TERGM; Krivitsky and Handcock, 2014) and stochastic actor oriented models (SOAM; Snijders et al., 2010).

The evolution of groups has also been well-studied (Groups → Groups). Groups can evolve in several ways, including expanding by merging with other groups or when new members join and shrinking by splintering into smaller groups or when existing members leave. Because the process of groups at time 1 evolving into groups at time 2 does not explicitly implicate networks, research on group evolution often does not draw on network theories or methods. One notable exception might be diffusion of innovation theory, which seeks to explain how and when members of the

https://doi.org/10.1017/nws.2023.3 Published online by Cambridge University Press

non-adopter group become members of the adopter group, which can depend in part on networks (Rogers, 2003). Because group memberships can be represented using two-mode networks, extensions to statistical methods for modeling network evolution have proven useful for modeling group evolution (Wang et al., 2013; Schaefer et al., 2022).

Much attention has been devoted to how networks or groups evolve over time, but it is also possible to study how one emerges from the other. Focus theory hypothesized that networks emerge from individuals' shared membership in groups or other "foci," which are "social, psychological, legal, or physical entity around which joint activities are organized" (Groups → Network; Feld, 1981, p. 1016). The interactions that take place in these groups "bring people together in a mutually rewarding situation" because they are focused on something that is shared, and therefore, these interactions are "positively valued" (Feld, 1981, p. 1017). Through these positively valued interactions, the participants "develop positive sentiments toward each other" and thus positive affective ties (Feld, 1981, p. 1026). Feld (1981) summarized the process by explaining that "As a consequence of interaction associated with their joint activities, individuals whose activities are organized around the same focus will tend to become interpersonally tied and form a cluster" (Feld, 1981, p. 1016). Two-mode projections (Breiger, 1974) and backbones (Neal, 2014) represent generative models that formalize this hypothesis and explicitly show how weighted and unweighted networks, respectively, can emerge from groups by transforming a two-mode network into a one-mode network.

Although focus theory is traditionally viewed as explaining how networks emerge from groups, Feld (1981) also acknowledged that groups can emerge from networks, noting that "Once there is a tie between two individuals, these individuals will tend to find and develop new foci around which to organize their joint activity" (p. 1019; Network → Groups). Indeed, his diagram of the dynamics of the focus model is cyclical, with groups creating ties, which in turn create new groups. Schaefer et al. (2022) recently provided empirical evidence of this process, finding that direct influence from friends was the single most important exogeneous predictor of whether a high school student would form or join a new extracurricular activity. However, none of focus theory's twenty propositions deal with how or when groups emerge from networks, and corresponding "generative models are practically not-existent" (Filho & O'Neale, 2020a, p. 3). It is these gaps that the present work seeks to address, thereby filling in the missing second half of focus theory and developing the complement to two-mode projection.

3. Groups from networks

Focus theory (Feld, 1981) and two-mode projection (Breiger, 1974; Neal, 2014) already offer a detailed description of how networks might emerge from groups. In this section, I propose three models for how a group might emerge from a network. Each model represents a simplified implementation of a theory about the formation of a specific type of group: teams, clubs, and organizations. For each model, I first present the motivating theory, then describe the model, and provide a concrete illustration of a group forming from a network according to the model.

3.1 Teams model

The *teams* model derives from an existing model of team formation. Guimera et al. (2005) suggested that the individuals who form teams in a given setting are embedded in a "complex network [that is] the medium in which future collaborations will develop" (p. 697–8). That is, teams emerge from an existing social network. In their original model, all teams had a fixed size m. Each of the m positions on a newly forming team were filled based on probabilities p and q. Specifically, a position was filled with (a) a new person joining the setting from an unlimited pool of outsiders with probability $1 - p$, (b) a person who is already a member of the setting with probability $p(1 - q)$,

https://doi.org/10.1017/nws.2023.3 Published online by Cambridge University Press

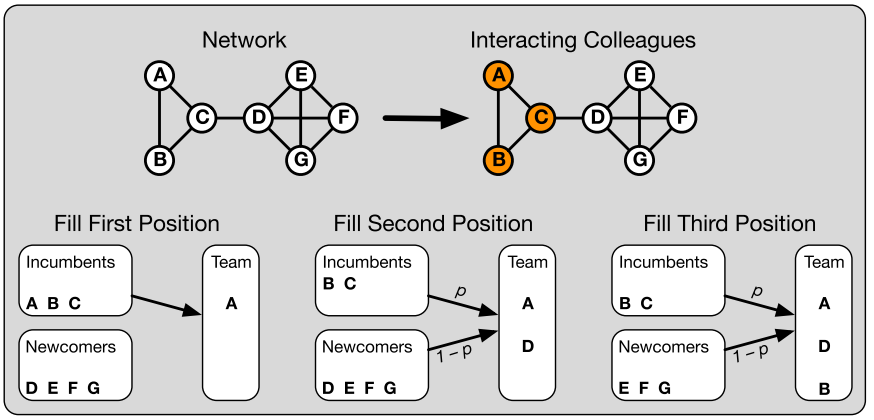

Figure 1. Example of group formation via the teams model. A new team emerges from a set of interacting colleagues {A,B,C}. The first position on the new team is filled by a random incumbent, here A. The second position is filled by either a random incumbent (with probability p) or a random newcoming (with probability $1 - p$), and here is filled by newcomer D. The third position is also filled by either a random incumbent or newcomer, and here is filled by incumbent B, yielding the new team {A,D,B}.

or (c) a person who is already a member of the setting and is connected in the social network to individuals on the new team with probability pq. Their model was dynamic because outsiders join the setting over time and because each new team contributes to the network that influences the formation of future teams. It is also complex because it is parameterized by three values m, p, and q.

The teams model is a modification of Guimera et al.'s (2005) complex dynamic model and allows teams of varying size to emerge from a static network based on a single parameter p. Given an existing social network, cliques represent sets of colleagues who all know or interact with one another, and who therefore might form a team. Each new team emerges from one of these cliques, but can involve changes in membership. Because some of the clique's members (i.e., incumbents) may be unavailable or lack the necessary skills for the newly forming team's task, they may be replaced by others (i.e., newcomers). The model outcome depends on a parameter that specifies the probability with which incumbents are retained (p), rather than replaced by newcomers on new teams ($1 - p$). Accordingly, the parameter p controls how closely the memberships of new teams will match the memberships of cliques in an existing social network. When $p = 1$, where incumbents are always retained, the teams model reduces to the model described by Guillaume & Latapy (2004), where teams are equivalent to cliques. Additionally, generating a two-mode network under this model where $p = 1$ *and* the generated network contains the minimum number of teams necessary to perfectly reproduce the original social network is equivalent to solving the NP-hard "clique cover problem" (Karp, 1972). A pseudocode algorithm of the teams model is provided in the *supplementary material* at https://osf.io/eyua4/.

Figure 1 provides a concrete example. Suppose the network on the left is a network among colleagues in an academic department, and the clique {A,B,C} represents a set of colleagues who know each other, perhaps because they worked together on a grant proposal. A new three-member team is emerging from this group to submit a new proposal. Because they are the ones initiating the new team's formation, the first position on the new team must be filled by one of the group's incumbents {A,B,C}. In this example, the first position is filled by incumbent A. The remaining two positions on the team are filled by selecting incumbents with probability p, and selecting newcomers with probability $1 - p$. In this example, the second position is filled by newcomer D, while the third position is filled by incumbent B. The new team {A,D,B} could be the outcome of a situation in which newcomer D replaced incumbent C to address reviewers' concerns with the earlier proposal.

https://doi.org/10.1017/nws.2023.3 Published online by Cambridge University Press

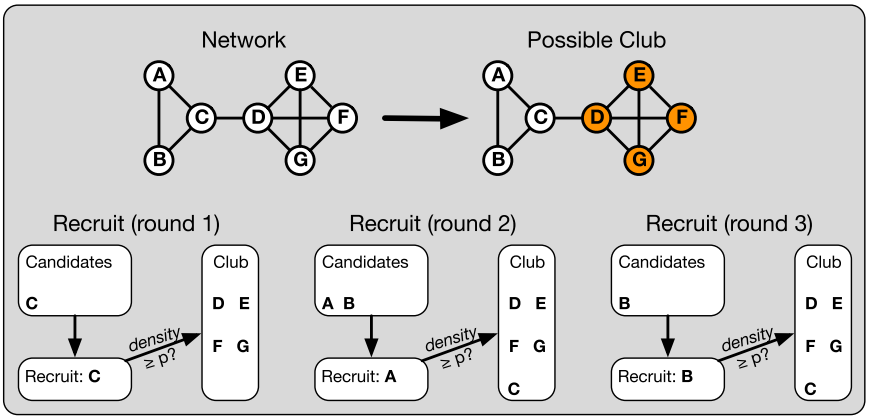

Figure 2. Example of group formation via the clubs model. A new club grows as members of a possible club {D,E,F,G} try to recruit additional members. In the first round, C is a friend of existing members and so is a candidate for recruitment. C joins because doing so maintains a minimum density of 70% among club members. In the second and third rounds, A and B are candidates for recruitment, but neither joins because doing so would reduce the within-club density below 70%. This yields a new club of {C,D,E,F,G}.

3.2 Clubs model

The *clubs* model is informed by findings about how social groups such as clubs form in both online and offline social networks. Backstrom et al. (2006) examined 19 characteristics of the group and potential joiner in two online social networks (LiveJournal and DBLP), while Schaefer et al. (2022) considered eight mechanisms that drive high school students to join extracurricular activities. Both studies found that the probability of joining a group depends on the number of friends one already has in the group. Additionally, Backstrom et al. (2006) found that the probability of joining a group also depends on the proportion of friends in the group who are friends with each other.

While these studies focused on individuals joining existing groups, their findings also have implications for the density of a newly forming group. The fact that i tends to join a group when she already has many friends j in the group increases the group's density by increasing the likelihood of i–j edges. Additionally, the fact that i tend to join a group when her friends j in the group are friends with each other increases the group's density by increasing the likelihood of j–j edges. Therefore, groups whose initial formation is guided by the conditions identified by Backstrom et al. (2006) and Schaefer et al. (2022) will be cohesive and have a relatively higher density than the overall network. From this implication, the clubs model views clubs as forming via an agglomeration process: a clique serves as the seed of a potential club. Then, seeking to establish a viable club, members recruit their friends, who join on the condition that the club would still have a minimum density p. Accordingly, p functions as a parameter that controls the club formation process. When $p = 1$, where new members join only if the new club would be a clique, the clubs model reduces to the model described by Guillaume & Latapy (2004), where groups are equivalent to cliques. Additionally, generating a two-mode network under this model where $p = 1$ *and* the generated network contains the minimum number of groups necessary to perfectly reproduce the original social network is equivalent to solving the NP-hard "clique cover problem" (Karp, 1972). A pseudocode algorithm of the clubs model is provided in the *supplementary material* at https://osf.io/eyua4/.

Figure 2 provides a concrete example, where $p = 0.7$. Suppose the network on the left is a friendship network, within which a group of friends {D,E,F,G} (a randomly selected clique) wishes to start a book club. To make their book club viable, they must recruit other friends to participate. The challenge is that these friends are socially anxious and only feel comfortable in group settings where at least 70% of the members are friends with each another. Initially, C is the only candidate

https://doi.org/10.1017/nws.2023.3 Published online by Cambridge University Press

because they are friends with existing book club members. The book club attempts to recruit C, and C decides to join because doing so would result in a book club in which 70% of the members are friends with each other. Once C is a member, A and B become candidates for recruitment. The book club attempts to recruit A first; however, A declines to join because doing so would yield a book club in which only 53% of members are friends with each other. The book club's attempt to recruit B is unsuccessful for the same reason. Thus, the new book club's members are {D,E,F,G,C}.

3.3 Organizations model

The *organizations* model mirrors the Blau space model of organizational recruitment (McPherson, 1983, 2004). Blau space is a multidimensional space within which individuals are located based on their sociodemographic characteristics. As McPherson (2004) explains, Blau space has two important properties: it "at once organizes the social interactions among individuals, and structures the opportunities for the formation of social entities that are associated with individuals in that space" (p. 267). First, it organizes social interactions because individuals who are sociodemographically similar are located nearby in the space and, according to the principle of homophily (McPherson et al., 2001), are therefore more likely to interact with each other. This implies that network ties will tend to be local within Blau space. Second, it structures the formation of social entities because organizations recruit members from specific regions in this space, known as niches (Popielarz & Neal, 2007). For example, a youth yachting league might recruit its members from the region located at the lower end of the age dimension, but the upper end of the family wealth dimension.

The organizations' model does not attempt to formalize all aspects of niche or organizational ecology theories (Popielarz & Neal, 2007; Shi et al., 2017), but instead is a simplification that incorporates only two central elements: individuals' positions in an unobserved Blau space derived from their distances in a social network and organizations' recruitment of members from niches in this space. Individuals' locations in Blau space can be estimated by embedding network geodesic distances in a d-dimensional space (Freeman, 1983; Péli & Bruggeman, 2006). While d can take any value between 1 and $N-1$, where N is the number of nodes in the network, I use a two-dimensional space because social networks tend to have low dimensionality (Freeman, 1983; Bonato et al., 2014), because many dimensions of social distinction are highly correlated (e.g., income and education), and because Blau space analysis is typically performed on low dimensionality spaces (Genkin et al., 2018). Organizations have d-dimensional circular niches within this space that reflect the type of member they seek to recruit (Péli & Bruggeman, 2006; Suh et al., 2017). Organizations' niche sizes vary, however most organizations are narrow-niche specialists, while a few are wide-niche generalists (Carroll, 1985). An organization's success at recruiting members depends on whether the prospective members are inside its niche (with probability p) or outside its niche (with probability $1-p$; Popielarz & McPherson, 1995). Accordingly, p serves as a parameter that controls the importance of niche location in individuals' joining behavior. A pseudocode algorithm of the organizations model is provided in the *supplementary material* at https://osf.io/eyua4/.

Figure 3 provides a concrete example. Suppose the network on the left is a social network of neighborhood friends. The geodesic distances between individuals in the network can be used to embed them in a 2-dimensional space via multidimensional scaling. Friends (e.g., C & D) are close together in this space, while friends-of-friends (e.g., A & D) are further apart in this space, and friends-of-friends-of-friends (e.g., A & F) are furthest apart. The sociodemographic characteristics described by these two dimensions are unknown, but perhaps they are income and education; notice the two dimensions are highly correlated. A multi-level marketing company selling beauty products aims to recruit sales associates; its niche is people who have less income and education, which includes four people. It recruits each person inside this niche with probability p, and in this example successfully recruits A, B, and D. Because its niche included four people, it aims to still recruit a fourth sales associate. It attempts to recruit those nearest the niche first, with probability

https://doi.org/10.1017/nws.2023.3 Published online by Cambridge University Press

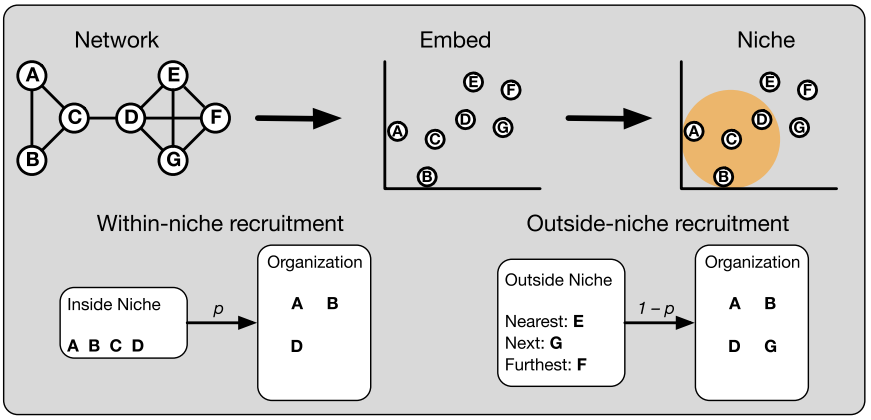

Figure 3. Example of group formation via the organizations model. A new organization grows by recruiting members depending on their positions in a sociodemographic space, which are inferred from the network. Individuals inside the organization's sociodemographic niche are recruited with probability p, which here leads to the recruitment of A, B, and C. Additional individuals are recruited from outside the organization's niche with probability $1 - p$, starting with those nearest the niche, which leads to the recruitment of G. This yields a new organization of {A,B,D,G}.

$1 - p$. In this example, it fails to recruit E, but successfully recruits G, at which point recruitment ends. This yields a neighborhood sales team of {A,B,D,G}.

4. Two-mode generative models

The models introduced in Section 3 each describe how one new group might emerge from an existing social network. However, if they are applied repeatedly on the same social network, they can also be viewed as generative models because they can generate two-mode networks representing group memberships from one-mode networks representing social networks. Many one-mode network generative models already exist, including the Erdős-Rényi model for generating random graphs (Erdős & Rényi, 1959), the Watts-Strogatz model for generating small-world graphs (Watts & Strogatz, 1998), and the Barabási-Albert model for generating scale free graphs (Barabási & Albert, 1999). However, as Filho & O'Neale (2020a) observe, "when it comes to bipartite networks [including two-mode networks]—a class of network frequently encountered in social systems, among others—generative models are practically non-existent" (p. 3). Their claim may have gone too far because there are methods for generating random two-mode networks (e.g., Newman et al., 2002), for randomizing existing two-mode networks (e.g., Jasny, 2012; Neal et al., 2021), or for generating bipartite (but not necessarily two-mode) networks using latent space (Filho & O'Neale, 2020a) or Bayesian (Caron, 2012) methods. However, none of these methods generate two-mode networks *from* one-mode networks and thus do not attempt to model how group memberships might emerge from social networks.

Generative models are not designed to simulate actual processes in the world, but instead are designed to reproduce observed empirical patterns using simple mechanisms. For example, the Watts-Strogatz model simply involves randomly re-wiring edges in a regular lattice. While this does not simulate how social networks actually form (e.g., people do not randomly swap friends), it does generate networks with characteristics observed in empirical social networks (e.g., clustering). Likewise, as two-mode generative models, these are not designed to simulate actual group formation processes, which are likely quite complex. Instead, they are designed to generate two-mode networks that have characteristics observed in empirical two-mode social networks and that encode features of the one-mode networks from which they were generated. In this section, I explore the extent to which they achieve these goals.

https://doi.org/10.1017/nws.2023.3 Published online by Cambridge University Press

The generative models are implemented in the `incidence.from.adjacency()` function in the `incidentally` package for R (Neal, 2022b). The code necessary to reproduce the results reported in this section is available at *https://osf.io/eyua4/*.

4.1 Reproducing empirical patterns

One way to evaluate these generative models involves examining whether they generate two-mode networks that have characteristics commonly observed in empirical two-mode social networks. Although much attention has been devoted to identifying the typical or universal properties of social networks (e.g., clustering, degree distributions; Watts & Strogatz, 1998; Barabási & Albert, 1999), relatively little work has examined the typical or universal properties of two-mode social networks. However, three characteristics are commonly observed: positively skewed agent degree distributions, positively skewed group degree distributions, and short cycles.

In a two-mode network generated by these models, the agent degree distribution captures the number of groups with which each agent is associated. Across many empirical contexts, this degree distribution tends to be positively skewed because most people are associated with just a few groups, while some people are associated with many groups. For example, most students participate in just a few extracurricular activities while some participate in many (Schaefer et al., 2022), most legislators sponsor just a few bills while some sponsor many (Neal, 2020), most women attend just a few parties while some attend many (Davis et al., 1941), and most authors write just a few papers while some write many (Filho & O'Neale, 2020b).

The group degree distribution captures the number of agents associated with each group. Again, across many empirical contexts this degree distribution tends to be positively skewed because most groups have just a few members, while some groups have many members. For example, most extracurricular activities have just a few participants while some have many (Schaefer et al., 2022), most bills are sponsored by just a few legislators while some are sponsored by many (Neal, 2020), most parties have just a few attendees while some have many (Davis et al., 1941), and most papers have just a few authors while some have many (Filho & O'Neale, 2020b).

Finally, empirical two-mode networks typically contain more four-cycles than would be expected at random. A four-cycle occurs when two nodes of one type are both connected to the same two nodes of another type, or in this context, two people are both members of the same two groups. Filho & O'Neale (2020b) demonstrated this pattern in three author-paper networks and one member-board network, arguing that it helps explain the strong ties observed in social networks due to shared groups. Drawing on this empirical pattern, Schaefer et al. (2022) explicitly hypothesized observing the formation of four-cycles through a mechanism they called "co-member influence," whereby high school students join the same new extracurricular activities as co-members of their existing extracurricular activities. Indeed, this is such an important property of two-mode networks that Saracco et al. (2015) count and control four-cycles (calling them X-motifs) in their null models.

Figure 4 illustrates how I examine whether the two-mode networks generated by these models have these empirically common characteristics. First, I generate a small-world network containing 50 nodes and 150 undirected edges. I use a small-world network because it has properties that are observed in many real-world social networks (e.g., clustering, small mean distance). Second, I use the clubs model, with $p = 0.95$ to generate a two-mode network containing 50 groups. I choose to generate 50 groups because it keeps the experiment a manageable size, but large enough that each agent could be a member of a singleton group. In this generated two-mode network, most agents belong to just a few groups, and thus, the agent degree distribution is positively skewed (skewness = 1.08, using Fisher's moment coefficient of skewness; Joanes & Gill, 1998). Likewise, most groups have just a few members, and thus, the group degree distribution is also positively skewed (skewness = 1.80). Finally, I use the fastball algorithm (Godard & Neal, 2022) to generate a random two-mode network with the same degree sequences, comparing the number of four-cycles

https://doi.org/10.1017/nws.2023.3 Published online by Cambridge University Press

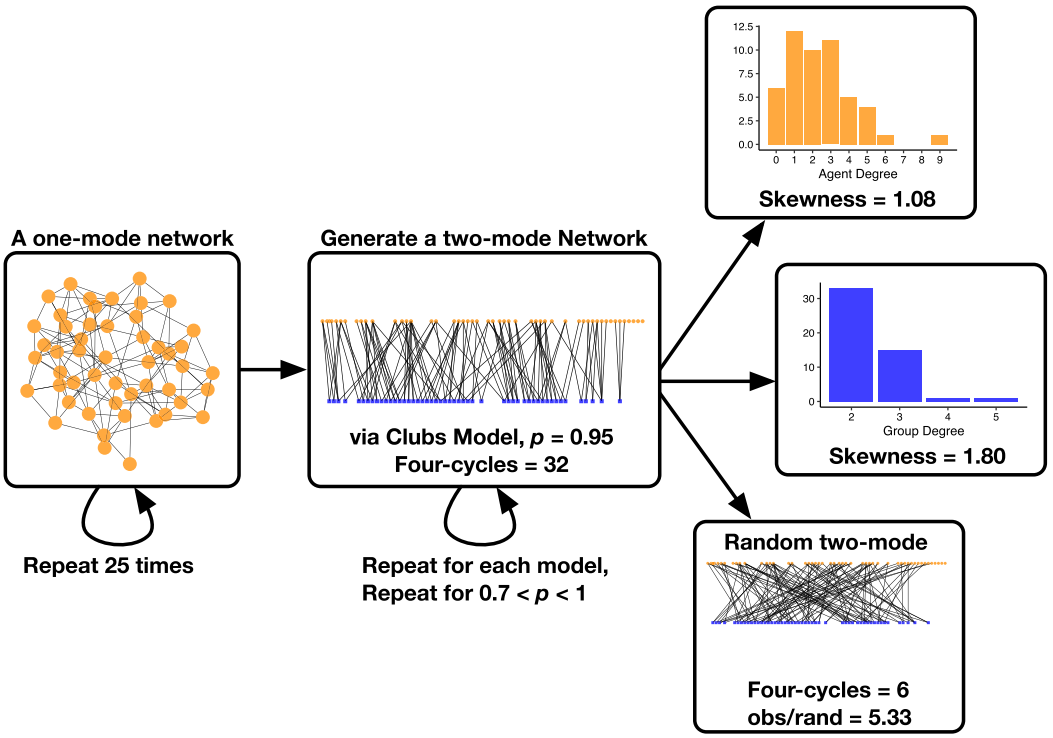

Figure 4. Evaluating a generated two-mode network. Given a one-mode network, a two-mode network is generated using one of the models (here, the Clubs Model is shown). The generated two-mode network is summarized by the skewness of its agent degrees, the skewness of its group degrees, and its over-representation of four-cycles relative to a random two-mode network. In this example, the Clubs Model with $p = 0.95$ generates a two-mode network with three properties commonly observed in empirical two-mode networks: positively skewed agent degrees, positively skewed group degrees, and an over-representation of four-cycles.

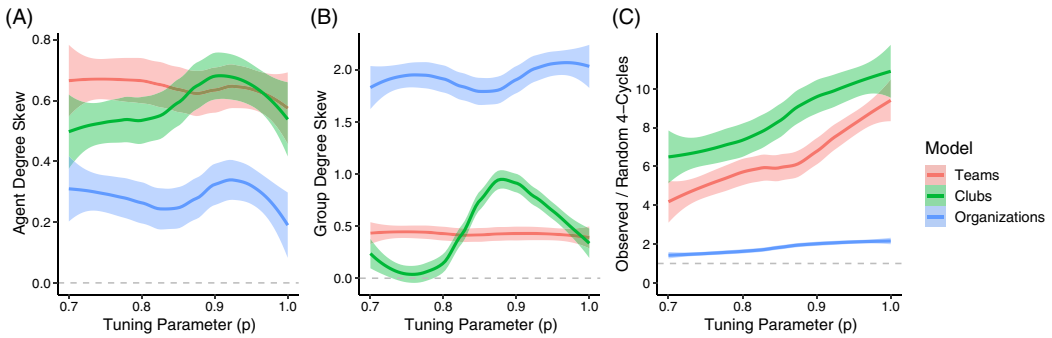

Figure 5. Experimental evaluation of generative models. (A) All models yield networks with positively skewed agent degrees. (B) Models usually yield networks with positively skewed group degrees. (C) All models yield networks with an over-representative of four-cycles.

in the generated and random networks. In this example, the generated network contains 5.33 times more four-cycles than a corresponding random network. Thus, in this example, the clubs model generated a two-mode network with all three expected properties.

Figure 5 shows the results of repeating this evaluation process 25 times, for each generative model, and for each parameter p between 0.7 and 1 in 0.025 intervals. Within each panel, the solid

https://doi.org/10.1017/nws.2023.3 Published online by Cambridge University Press

lines (red = teams model, green = clubs, blue = organizations) report averages over 25 replications, while the shaded bands indicate the 95% confidence interval. Panel A illustrates that for all models and all values of p, the generated two-mode networks have a positively skewed agent degree distribution. Panel B illustrates that except for two-mode networks generated using the clubs model with low values of p, all generated networks also have a positively skewed group degree distribution. Finally, panel C illustrates that for all models and all values of p, the generated two-mode networks have more four-cycles than a corresponding random network. Thus, this experiment demonstrates that under a broad set of circumstances, these models generate two-mode networks that have characteristics commonly observed in empirical two-mode social networks.

4.2 Encoding one-mode networks

The generative models all yield two-mode networks that have characteristics commonly observed in empirical two-mode social networks. However, the generative models should also yield two-mode networks that encode features of the particular one-mode networks from which they were generated. To evaluate this, I examine how well the original one-mode network can be recovered from the generated two-mode network.

Using the Zachary (1977) karate club network as the input, I use each model to generate a two-mode network of 1000 groups, with $p = 0.8$ (see Figure 6). Setting $p = 0.8$ ensures that the generated two-mode networks contain a fair amount of noise. Generating a large number of groups mirrors what a researcher might encounter when attempting to collect data in the field: an inability to directly observe the network of interest, but the ability to observe many instances of small events (e.g., Neal et al., 2022). For example, while it may be impossible to directly observe the karate club's social network, a researcher might be able to observe who participates in many small practice sessions and social events.

From each of the generated two-mode networks, I extract the backbone of its two-mode projection using the stochastic degree sequence model (SDSM; Neal, 2014, 2022a), then compute the similarity between this backbone and the original network. The simple matching coefficients (97%–80%) indicate that which dyads are (dis)connected in the backbone extracted from the generated two-mode networks closely matches which dyads are (dis)connected in the original one-mode network. More conservative similarity indices—correlation (0.85–0.33) and Jaccard coefficient (0.76–0.27)—are expectedly lower, but are still positive and generally large.

Variation in the correspondence between the original network and the backbone of the projection of a generated two-mode network may be driven by the model used to generate the two-mode network (i.e., teams, clubs, organizations), by the model used to extract the backbone (here, the SDSM), or both. Understanding the circumstances under which a given one-mode network can be recovered is an important direction for future research. However, the present analysis illustrates that the two-mode networks generated by these models are not simply random two-mode networks with empirically common features, but in fact are two-mode networks that encode features of the specific one-mode networks from which they were generated.

5. Discussion

Over a century ago, Simmel (1922) sketched the close association between individuals and groups. Building on these early ideas, Breiger (1974) demonstrated a method for deriving an interpersonal social network from individuals' group memberships, while Feld (1981) proposed focus theory to explain how social ties emerge from shared groups. Together, these methodological and theoretical contributions have facilitated research on how networks emerge from groups.

https://doi.org/10.1017/nws.2023.3 Published online by Cambridge University Press

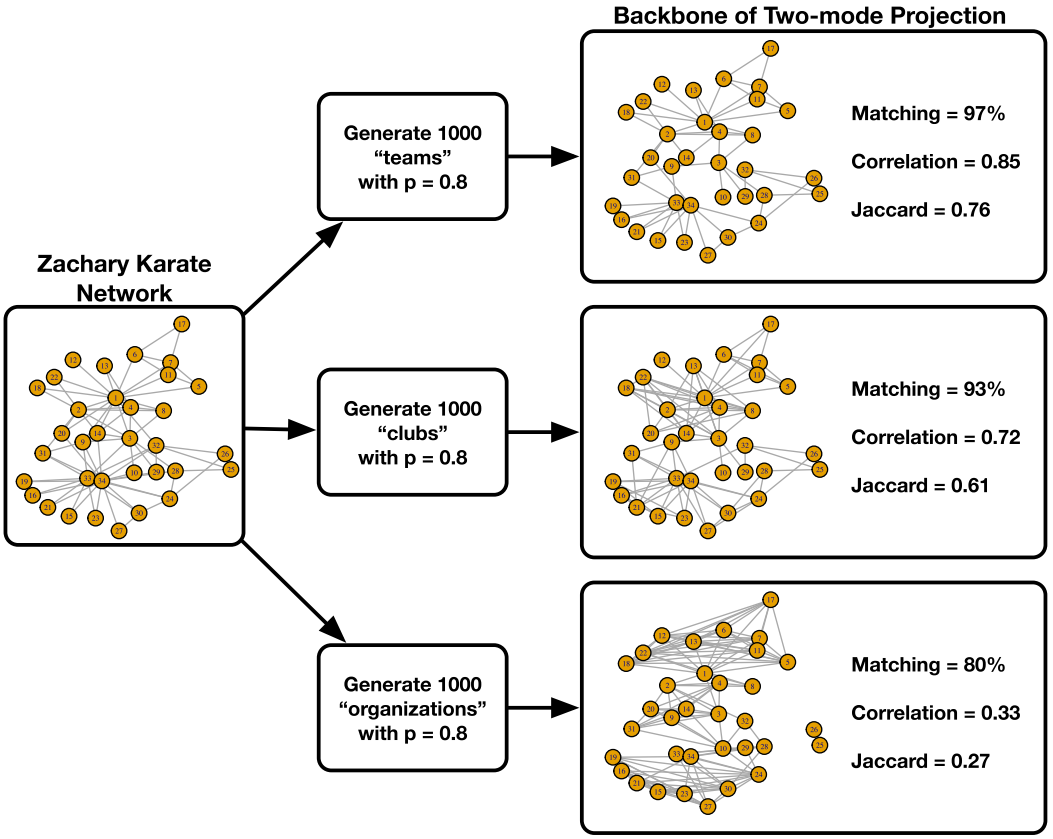

Figure 6. Recovering a one-mode network. Starting from the Zachary Karate Club network, a two-mode network is generated using each of the three models with $p = 0.8$. The backbone of the projection of the generated two-mode network is extracted and then compared to the original network. The positive and large similarity indices indicate that the generated two-mode networks encode features of the one-mode network from which they were generated.

While prior work has provided the theoretical and methodological tools for understanding how groups lead to networks, less is known about the opposite process: *how do networks lead to groups?* In this paper, building on ideas already present in focus theory and drawing on related theories of team (Guimera et al., 2005), club (Backstrom et al., 2006; Schaefer et al., 2022), and organization (McPherson, 1983) recruitment, I proposed three simple models for how a new group might emerge from an existing social network. In the *teams* model, a new team is formed from incumbents of, and newcomers to, network cliques. In the *clubs* model, a new club emerges as members of a network clique attempt to recruit friends. Finally, in the *organizations* model, a new organization recruits members from the interior and periphery of a sociodemographic niche.

These models can be viewed as two-mode network generative models, which are controlled by a tuning parameter p that adjusts how closely the generated groups match the social network. A series of simulations demonstrated that these models generate two-mode networks that have characteristics commonly observed in empirical two-mode social networks: positively skewed agent degrees, positively skewed group degrees, and an over-representation of four-cycles. Additionally, an example using the Zachary (1977) karate club network illustrated that the generated two-mode networks encode features of the one-mode network from which they were generated.

These models represent a theoretical contribution to the literature on networks and groups because they elaborate the missing second half of focus theory (Feld, 1981). Specifically, while focus theory hypothesized that groups (i.e., foci) lead to networks, and networks in turn lead to

https://doi.org/10.1017/nws.2023.3 Published online by Cambridge University Press

new groups, nearly all applications and extensions have focused on the first process, while neglecting the second process. To be sure, these models are simplified implementations of theories about group formation, and therefore are highly stylized. However, they provide a formalized starting point for further theoretical elaboration of focus theory and of the co-evolution of networks and groups.

These models also represent a methodological contribution to the literature on network generative models. One-mode generative models—for example, the Erdős-Rényi (Erdős & Rényi, 1959), Watts-Strogatz (Watts & Strogatz, 1998), and Barabási-Albert (Barabási & Albert, 1999) models—have played a critical role in understanding the properties of networks and are frequently used as null models against which observed networks are evaluated. However, "when it comes to bipartite networks [including two-mode networks]. . .generative models are practically non-existent" (Filho & O'Neale, 2020a, p. 3). The generative models developed here, which yield two-mode networks with empirically common features and that encode features of one-mode networks, begin to fill that gap. Like existing generative models, they can be used to explore the properties of social two-mode networks and can be used as null models against which observed two-mode networks are evaluated.

5.1 Limitations and future directions

These models and results are subject to some limitations, which highlight possible directions for future research. First, each model describes the emergence of a group solely from a network (i.e., network $\rightarrow$ group) and therefore does not allow individuals' participation in one group to influence their participation in future groups. More complex future models may allow groups to emerge not only as a function of the network but also as a function of already existing groups (i.e., $\substack{\text{network}\\\text{existing groups}}\rightarrow$ new group). Second, each model represents only a simplified implementation of a theory and therefore does not attempt to incorporate all of the theory's mechanisms. For example, the organizations model is a significantly reduced form of organizational ecology, but provides a framework for future versions to incorporate additional elements such as niche carrying capacities (Popielarz & Neal, 2007) or competition (Shi et al., 2017). Finally, the evidence that these models generate two-mode networks that contain features commonly observed in empirical two-mode networks is restricted to two types of features: degree distributions and cycles. As future research identifies other common features of empirical two-mode networks, the simulations described in Section 4.1 can be replicated to evaluate whether the generated two-mode networks also display these features.

6. Conclusions

Theories and methods have long acknowledged that individuals' group memberships can facilitate the formation of social ties. However, it is equally plausible that individuals' social ties can facilitate the formation of new groups. In this paper, I have sketched three models that describe how this might happen and formalized them as two-mode generative models. These models have the potential to advance theories of how groups emerge from networks as well as to provide methods for understanding and evaluating observed social two-mode networks. Moreover, as theoretically informed but simple models, they also offer a starting point for the development of more complex and realistic models.

Competing interests. None.

Funding statement. This work was supported by the National Science Foundation (#2016320 and #2211744).

https://doi.org/10.1017/nws.2023.3 Published online by Cambridge University Press

Data availability statement. Pseudocode algorithms and the code to replicate these analyses are available at https://osf.io/eyua4/.

References

Backstrom, L., Huttenlocher, D., Kleinberg, J., & Lan, X. (2006). Group formation in large social networks: membership, growth, and evolution. In *Proceedings of the 12th ACM SIGKDD international conference on Knowledge discovery and data mining* (pp. 44–54).

Barabási, A.-L., & Albert, R. (1999). Emergence of scaling in random networks. *Science*, *286*(5439), 509–512.

Bonato, A., Gleich, D. F., Kim, M., Mitsche, D., Prałat, P., Tian, Y., . . . Hayasaka, S. (2014). Dimensionality of social networks using motifs and eigenvalues. *PLOS One*, *9*(9), e106052.

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

Caron, F. (2012). Bayesian nonparametric models for bipartite graphs. In Pereira, F., Burges, C., Bottou, L., & Weinberger, K. (Eds.), *Advances in neural information processing systems* (Vol. *25*). Montreal: Curran Associates, Inc.

Carroll, G. R. (1985). Concentration and specialization: Dynamics of niche width in populations of organizations. *American Journal of Sociology*, *90*(6), 1262–1283.

Davis, A., Gardner, B. B., & Gardner, M. R. (1941). *Deep south: A social anthropological study of caste and class*. Chicago, IL: University of Chicago Press.

Erdős, P., & Rényi, A. (1959). On random graphs. *Publicationes Mathematicae*, *6*, 290–297.

Feld, S. L. (1981). The focused organization of social ties. *American Journal of Sociology*, *86*(5), 1015–1035.

Filho, D. V., & O'Neale, D. R. (2020a). Latent space generative model for bipartite networks. In *International Conference on Network Science* (pp. 3–16). Cham: Springer.

Filho, D. V., & O'Neale, D. R. (2020b). Transitivity and degree assortativity explained: The bipartite structure of social networks. *Physical Review E*, *101b*(5), 052305.

Freeman, L. C. (1983). Spheres, cubes and boxes: Graph dimensionality and network structure. *Social Networks*, *5*(2), 139–156.

Fuhse, J. A., & Gondal, N. (2022). Networks from culture: Mechanisms of tie-formation follow institutionalized rules in social fields. *Social Networks*. 10.1016/j.socnet.2021.12.005.

Genkin, M., Wang, C., Berry, G., & Brashears, M. E. (2018). Blaunet: An R-based graphical user interface package to analyze Blau space. *PLOS One*, *13*(10), e0204990.

Godard, K., & Neal, Z. P. (2022). fastball: A fast algorithm to sample binary matrices with fixed marginals. *Journal of Complex Networks*, *10*(6), cnac049.

Guillaume, J.-L., & Latapy, M. (2004). Bipartite structure of all complex networks. *Information Processing Letters*, *90*(5), 215–221.

Guimera, R., Uzzi, B., Spiro, J., & Amaral, L. A. N. (2005). Team assembly mechanisms determine collaboration network structure and team performance. *Science*, *308*(5722), 697–702.

Jasny, L. (2012). Baseline models for two-mode social network data. *Policy Studies Journal*, *40*(3), 458–491.

Joanes, D. N., & Gill, C. A. (1998). Comparing measures of sample skewness and kurtosis. *Journal of the Royal Statistical Society: Series D (The Statistician)*, *47*(1), 183–189.

Karp, R. M. (1972). Reducibility among combinatorial problems. In Miller, R. E., Thatcher, J. W., & Bohlinger, J. D. (Eds.), *Complexity of computer computations* (pp. 85–103). Boston, MA: Springer.

Krivitsky, P. N., & Handcock, M. S. (2014). A separable model for dynamic networks. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, *76*(1), 29–46.

McPherson, M. (1983). An ecology of affiliation. *American Sociological Review*, *48*(4), 519–532.

McPherson, M. (2004). A blau space primer: Prolegomenon to an ecology of affiliation. *Industrial and Corporate Change*, *13*(1), 263–280.

McPherson, M., Smith-Lovin, L., & Cook, J. M. (2001). Birds of a feather: Homophily in social networks. *Annual Review of Sociology*, *27*(1), 415–444.

Neal, J. W., Neal, Z., & Durbin, C. E. (2022). Inferring signed networks from preschoolers' observed parallel and social play. *Social Networks*, *77*, 80–86.

Neal, Z. P. (2014). The backbone of bipartite projections: Inferring relationships from co-authorship, co-sponsorship, co-attendance and other co-behaviors. *Social Networks*, *39*, 84–97.

Neal, Z. P. (2020). A sign of the times? Weak and strong polarization in the US congress, 1973–2016. *Social Networks*, *60*, 103–112.

Neal, Z. P. (2022a). backbone: An R package to extract network backbones. *PLOS One*, *17a*(5), e0269137.

Neal, Z. P. (2022b). incidentally: An R package to generate incidence matrices and bipartite graphs. *OSF Preprints*. https://doi.org/10.31219/osf.io/ectms

Neal, Z. P., Domagalski, R., & Sagan, B. (2021). Comparing alternatives to the fixed degree sequence model for extracting the backbone of bipartite projections. *Scientific Reports*, *11*(1), 23929.

https://doi.org/10.1017/nws.2023.3 Published online by Cambridge University Press

Newman, M. E., Watts, D. J., & Strogatz, S. H. (2002). Random graph models of social networks. *Proceedings of the National Academy of Sciences*, *99*(suppl_1), 2566–2572.
Péli, G., & Bruggeman, J. (2006). Networks embedded in n-dimensional space: The impact of dimensionality change. *Social Networks*, *28*(4), 449–453.
Popielarz, P. A., & McPherson, J. M. (1995). On the edge or in between: Niche position, niche overlap, and the duration of voluntary association memberships. *American Journal of Sociology*, *101*(3), 698–720.
Popielarz, P. A., & Neal, Z. P. (2007). The niche as a theoretical tool. *Annual Review of Sociology*, *33*(1), 65–84.
Robins, G., Pattison, P., Kalish, Y., & Lusher, D. (2007). An introduction to exponential random graph (p*) models for social networks. *Social Networks*, *29*(2), 173–191.
Rogers, E. M. (2003). *Diffusion of innovation*. New York: Free Press.
Saracco, F., Di Clemente, R., Gabrielli, A., & Squartini, T. (2015). Randomizing bipartite networks: The case of the World Trade Web. *Scientific Reports*, *5*(1), 10595.
Schaefer, D. R., Khuu, T. V., Rambaran, J. A., Rivas-Drake, D., & Umaña-Taylor, A. J. (2022). How do youth choose activities? Assessing the relative importance of the micro-selection mechanisms behind adolescent extracurricular activity participation. *Social Networks*. 10.1016/j.socnet.2021.12.008.
Shi, Y., Dokshin, F. A., Genkin, M., & Brashears, M. E. (2017). A member saved is a member earned? The recruitment-retention trade-off and organizational strategies for membership growth. *American Sociological Review*, *82*(2), 407–434.
Simmel, G. (1955 [1922]). The web of group affiliations. In Wolff, K. H. (Ed.), *Conflict and the web of group affiliations* (pp. 127–195). New York: Simon and Schuster.
Snijders, T. A., Van de Bunt, G. G., & Steglich, C. E. (2010). Introduction to stochastic actor-based models for network dynamics. *Social Networks*, *32*(1), 44–60.
Suh, C. S., Shi, Y., & Brashears, M. E. (2017). Negligible connections? The role of familiar others in the diffusion of smoking among adolescents. *Social Forces*, *96*(1), 423–448.
Wang, P., Pattison, P., & Robins, G. (2013). Exponential random graph model specifications for bipartite networks—A dependence hierarchy. *Social Networks*, *35*(2), 211–222.
Watts, D. J., & Strogatz, S. H. (1998). Collective dynamics of small-worldnetworks. *Nature*, *393*(6684), 440–442.
Wellman, B. (1988). Structural analysis: From method and metaphor to theory and substance. In Wellman, B., & Berkowitz, S. D. (Eds.), *Social structures: A network approach* (pp. 19–61). Cambridge: Cambridge University Press.
Yap, J., & Harrigan, N. (2015). Why does everybody hate me? Balance, status, and homophily: The triumvirate of signed tie formation. *Social Networks*, *40*, 103–122.
Zachary, W. W. (1977). An information flow model for conflict and fission in small groups. *Journal of Anthropological Research*, *33*(4), 452–473.

Cite this article: Neal Z. P. The duality of networks and groups: Models to generate two-mode networks from one-mode networks. *Network Science*. https://doi.org/10.1017/nws.2023.3

https://doi.org/10.1017/nws.2023.3 Published online by Cambridge University Press