Network Science **7** (2): 141–159, 2019
doi:10.1017/nws.2019.2

CAMBRIDGE
UNIVERSITY PRESS

ORIGINAL ARTICLE

Clustering ensembles of social networks

Tracy M. Sweet[1*], Abby Flynt[2] and David Choi[3]

[1]Department of Human Development and Quantitative Methodology University of Maryland College Park, MD, USA, [2]Department of Mathematics Bucknell University Lewisburg, PA, USA and [3]Heinz College Carnegie Mellon University Pittsburgh, PA, USA
*Corresponding author. Email: tsweet@umd.edu

Action Editor: Stanley Wasserman

Abstract

Recently there has been significant work in the social sciences involving ensembles of social networks, that is, multiple, independent, social networks such as students within schools or employees within organizations. There remains, however, very little methodological work on exploring these types of data structures. We present methods for clustering social networks with observed nodal class labels, based on statistics of walk counts between the nodal classes. We extend this method to consider only non-backtracking walks, and introduce a method for normalizing the counts of long walk sequences using those of shorter ones. We then present a method for clustering networks based on these statistics to explore similarities among networks. We demonstrate the utility of this method on simulated network data, as well as on advice-seeking networks in education.

Keywords: clustering, graph kernel, random walks, advice networks, education

Author ORCID. Tracy M. Sweet, https://orcid.org/0000-0001-9929-7518

1. Introduction

There are many contexts in the social sciences that involve ensembles of social networks, that is, multiple, independent networks such as networks of students in different classes or schools (Gest and Rodkin, 2011; Harris et al., 2009; Paluck and Shepherd, 2012), networks of teachers in different schools (Frank et al., 2004; Hopkins et al., 2015; Spillane et al., 2017), or employee networks at different companies (Sarkar et al., 2010). Recently, a small body of research has focused on multilevel statistical network models; there has been some recent work on multilevel social selection models (Snijders and Kenny, 1999; Sweet et al., 2013; Zijlstra et al., 2006), influence models (Frank et al., 2014), and models that disentangle the two (Snijders and Baerveldt, 2003; Snijders et al., 2008).

Still, there has been very little collective attention paid to exploratory methods for analyzing multiple social networks. The work that does exist tends to focus on a substantive analysis (e.g., Sinani et al., 2008) or another methodological issue altogether (e.g., Sweet and Zheng, 2017; Traud et al., 2011). One possible exception is Faust and Skvoretz (2002), who used parameter estimates from network models in an exploratory comparative analysis.

We argue that exploratory methods for multiple network data are needed, given the increasing availability of data and renewed focus on both replication and generalizability in substantive research. In addition, we argue that these methods both scale to accommodate larger datasets and allow analysts to incorporate content knowledge in their exploration.

We propose using cluster analysis to cluster social networks based on tie structure among nodes with certain attributes. While cluster analysis has largely been applied to individual networks to determine communities of nodes, we aim to cluster entire networks into groups that are similar in terms of how nodes of some a certain type interact. Our pedagogical example focuses on advice-seeking networks across 14 elementary schools. School staff members have various assigned roles: teacher, department chair, instructional coach, and administrator. Thus, our method aims to cluster schools based on the differing patterns of advice-seeking ties in each school, between staff of different roles.

Towards this aim, we highlight a data-mining concept known as a graph kernel (Kashima et al., 2003; Ralaivola et al., 2005; Vishwanathan et al., 2010), in which each network is implicitly represented by a vector of attributes that are derived from its adjacency matrix, which for particular choices of attributes can be efficiently manipulated by a technique known as the "kernel trick" (Friedman et al., 2001). For example, Ralaivola et al. (2005) consider settings where nodal group labels are included with the network data and propose to represent each network by its walk counts between the nodal groups. For this choice of attributes, very high dimensional attribute vectors—i.e., involving many different walk patterns—can be tractably utilized, which potentially may improve the richness of the network representation.

In this work, we propose extensions to the methodology of Ralaivola et al. (2005) which may improve the interpretability of the walk counts, and also their suitability for sparse networks. Specifically, we propose a method for normalizing the walk count statistics. Just as the skewness of a distribution normalizes by variance and is a better descriptor than the unnormalized third moment, our walk counts will be similarly normalized in hopes of better representing the network. Additionally, we propose counting only the non-backtracking walks in a network, which has been shown to improve performance for sparse graphs in other settings (Krzakala et al., 2013; Martin et al., 2014).

In our advice-seeking network data, we conjecture that the usage of walk counts between staff members in a formal leadership position (coach, chair, and administrator) and staff members without a formal leadership position may be informative. As information might propagate in a network by "walking" along its edges, we believe that the walk count statistics will capture aspects of the school networks that are relevant to the flow of information between teachers and formally designated school leaders.

2. Related work: data analysis for ensembles of networks

When analyzing data from multiple networks, it is common to characterize the networks as independent replications or samples from some super population of networks; that is, these networks are isolated, independent, but similar enough to one another to be analyzed collectively (Lazega and Snijders, 2015; Sweet et al., 2013). Thus in our discussion of exploratory methods for multiple network data, we generally assume that networks are comparable; network ties represent the same relationship across networks, and ties are measured in the same way.

Our discussion will focus specifically on studies involving cluster analysis, but we note that this is not a comprehensive review of the literature. One notable work is Faust and Skvoretz (2002), who clustered 42 different networks based after fitting network models. Given each fitted network model, they calculated a pairwise similarity measure based on predicted tie probabilities for one network based on the parameter estimates from another network. They used correspondence analysis to organize these similarities. It is worth noting that this is the only study in which the nodes are a different species across each network.

Related to cluster analysis on an ensemble of networks, existing work has also considered the classification task. For example, Saigo et al. (2009) considered classifiers which in principle could use all possible network subgraphs as the set of attributes to describe each network. As this is a computationally intractable set to enumerate, a computational technique known as column

https://doi.org/10.1017/nws.2019.2 Published online by Cambridge University Press

$$\begin{pmatrix} 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 1 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \end{pmatrix}$$

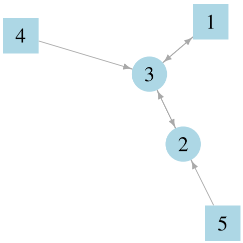

Figure 1. A binary, directed social network with six nodes can be represented by a sociomatrix (left) or a sociogram (right). Leaders are represented as circles and followers are shown as squares in the sociogram (right).

generation was used to greedily select subgraphs to be added to the set of network attributes under consideration. In Vogelstein et al. (2013), a classifier was developed where the network attributes are a subset of all $\binom{n}{2}$ dyads; this subset is selected by testing for each dyad the hypothesis that its probability of occurrence is equal across classes. In addition, classifiers using graph kernels have been considered in works such as Ralaivola et al. (2005), Vishwanathan et al. (2010), and Kashima et al. (2003). Our paper extends the methodology for random walk kernels (e.g., Ralaivola et al. (2005)), by proposing a normalization for the walk counts and by considering the usage of non-backtracking walks.

3. Using walk counts to cluster networks

In this section, we present the walk count statistics used in our graph kernel approach. We also give extensions to improve interpretability and suitability for sparse networks.

3.1 Counting walks in a network

To apply the walk count statistics of Ralaivola et al. (2005), we require a social network with nodal class labels. That is, we require an $n \times n$ sociomatrix Y, such that Y_{ij} is the value of the relationship from node i to node j, along with a categorical label for each node. For example, in Section 5 we consider advice networks between teachers, where a tie indicates that a teacher goes to another teacher for advice regarding instruction on a regular basis, and where the data contain a categorical label for each teacher, designating whether or not they hold a formal leadership position in the school, such as instructional coach or principal.

Relationships are measured as binary, ordinal, or continuous, and network ties may be directional. For example, advice-seeking networks are directed so that Y is not necessarily symmetric. For simplicity, our examples will involve binary-valued network ties, such that $Y_{ij} = 1$ indicates the presence of a tie from i to j; however, our walk count formulae will also be applicable to multigraphs and nonnegative weighted graphs without modification.

A toy example network is given in Figure 1, which shows the binary sociomatrix Y for the six nodes and the corresponding sociogram (network plot). In addition, our sociogram uses squares and circles to denote followers and leaders, respectively. Note that being a leader or a follower is a node-level attribute determined independently from the network structure; for example, in a school, some staff members are leaders based on their formal position. Further, this designation could be generalized to a different node-level attribute.

We will use the example network in Figure 1 throughout this section. We define a walk as a sequence of connected nodes (respecting tie directionality) which can include the same node or edge multiple times. For example, a possible walk from node 1 to node 2 in our toy network (Figure 1) is 1–3–2, where we start at node 1, then go to node 3, and then to node 2. Another longer walk from node 1 to node 2 is 1–3–1–3–2. We can also specify the length of the walk; the walk 1–3–2 has length 2 since it traverses two edges in the network, whereas the walk of 3–1–3–2 has length 3.

https://doi.org/10.1017/nws.2019.2 Published online by Cambridge University Press

Table 1. Number of walks of length 1 and 2 of the toy network Figure 1

ll	*lf*	*fl*	*ff*
2	1	3	0

lll	*llf*	*lfl*	*lff*	*fff*	*flf*	*fll*	*ffl*
2	1	1	0	0	1	3	0

In our toy example (Figure 1), there are 2 length 1 walks from leaders to leaders (2–3 and 3–2) and 1 length 1 walk from leaders to followers (3–1). As examples of length 2 walks, 2–3–1 is the only length 2 walk in the network from leader to leader to follower, while there exist 2 different length 2 walks involving leaders only, namely 3–2–3 and 2–3–2. For our example network (Figure 1), Table 1 shows the possible walk configurations and the corresponding length 1 and 2 walk counts; there are four possible sequences corresponding to length 1 walks (or "length 1 walk types") and eight possible length 2 walk sequences. Of course, walks of any length can be enumerated; for our toy example, there is at most 1 walks of length 3 for a given configuration, so we did not include these counts or counts of longer walks in Table 1.

To compute these counts as well as counts of longer walks, define D_l as a diagonal matrix such that $[D_l]_{ii} = 1$ if node i is a leader and 0 otherwise. Similarly, D_f is defined as the diagonal matrix such that $[D_f]_{ii} = 1$ for follower nodes and 0 otherwise. Then we can calculate the number of walks by taking products of Y, D_l, and D_f. This is perhaps best illustrated by an example. Let W_{lf} denote the number of length 1 walks from leaders to followers, then

$$W_{lf} = \sum_{i:leader} \sum_{j:follower} Y_{ij} = 1^T D_l Y D_f 1$$

In this formula, one can easily see that we are creating a matrix $D_l Y D_f$ that has a 1 for every tie from a leader to a follower and 0 otherwise. Then we pre- and post-multiply by the vector 1 to compute the sum of the entries in $D_l Y D_f$.

For longer walks, we can extend the formula to capture additional ties. For example,

$$W_{llfl} = \sum_{i:leader} \sum_{j:leader} \sum_{k:follower} \sum_{l:leader} Y_{ij} Y_{jk} Y_{kl} = 1^T D_l Y D_l Y D_f Y D_l 1$$

counts the number of walks which follow the sequence *llfl*, and for a general sequence $c_1 c_2 \cdots c_\ell$ where $c_i \in \{l, f\}$ for $i = 1, \ldots, \ell$, it holds that

$$W_{c_1 \cdots c_\ell} = 1^T D_{c_1} \left(\prod_{i=2}^{\ell} Y D_{c_i} \right) 1 \tag{1}$$

Given a network and a list of walk types, Equation (1) can be evaluated to count the number of walks of each type, generating a vector of descriptive statistics for each network.

3.2 Modifications of walk counts

In this section, we propose two modifications to computing walk count vectors, which may improve practical performance. For later use, we define W_f and W_l as the number of leaders and followers in the network. These are given by

$$W_f = 1^T D_f 1$$

$$W_l = 1^T D_l 1$$

https://doi.org/10.1017/nws.2019.2 Published online by Cambridge University Press

Standardizing the walk counts. Clustering the vector of unnormalized walk counts can lead to a degraded and unreliable solution due to two artifacts of counting walks, both of which are partially illustrated in our example with length 1 and 2 walks (see Section 6). First, longer walks are exponentially more numerous than shorter ones, so the feature vectors are poorly scaled. Second, the walk counts are highly correlated with one another. For example, if there are a lot of edges between leaders, then there will be a lot of walks of length 2 between leaders as well.

Thus to compare feature vectors across networks and ultimately cluster them, we need to address the scaling issue. To standardize the walk counts, we rescale counts based on edge probabilities. That is, we scale the count of each walk type by the number of ways to choose that combination of connectivity and count the fraction of possible walk combinations that are connected. As an example, the number of leader to follower walks, W_{lf}, is rescaled to

$$W^{\star}_{lf} = \frac{W_{lf}}{W_l W_f}$$

The length 2 walk W_{lff} is rescaled to

$$W^{\star}_{lff} = \frac{\frac{W_{lff}}{W_l W_f W_f}}{\left(\frac{W_{lf}}{W_l W_f} \cdot \frac{W_{ff}}{W_f W_f}\right)} \tag{2}$$

Notice that the numerator is the same form as for the length 1 walk, whereas the denominator is the product of the individual probabilities for the two events that need to occur to have a length 2 walk of lff.

We can extend this normalization to an arbitrary walk type $c_1, c_2, \ldots, c_\ell$, where $c_i \in \{l, f\}$ for $i = 1, \ldots, \ell$:

$$W^{\star}_{c_1 c_2 \cdots c_\ell} = \frac{\frac{W_{c_1 c_2 \cdots c_\ell}}{W_{c_1} W_{c_2} \cdots W_{c_\ell}}}{\left(\frac{W_{c_1 c_2}}{W_{c_1} W_{c_2}} \cdots \frac{W_{c_{\ell-1} c_\ell}}{W_{\ell-1} W_\ell}\right)} \tag{3}$$

Non-backtracking walks. In our toy network (Figure 1), there are 2 *llll* walks, 2–3–2–3 and 3–2–3–2. We call these *backtracking walks* because they involve traveling along an edge and then immediately backtracking by following the same edge in the opposite direction.

For sparse networks, a significant proportion of long walks will involve repeated edges. In such cases, we found that the normalization formula Equation (2) was less successful. Our intuition for this observation is that the denominator of Equation (2) is a product of independent edge probabilities, which is not valid when edges are repeated. For this reason, we need to modify our walk count vector so that only walks with non-repeated edges are counted; however, this is a computationally difficult task. As an approximation to counting only walks with non-repeating edges, we instead will count only non-backtracking walks, which can be computed efficiently by replacing the sociomatrix Y with its Hashimoto non-backtracking matrix (Hashimoto, 1989). This matrix has been used in a number of network analyses for both community detection (Krzakala et al., 2013) and measures of centrality (Martin et al., 2014).

The non-backtracking Hashimoto matrix H is an $m \times m$ matrix, where m is the number of potential ties in the original network Y. Each entry in H provides information about the next step one can take on the graph given the prior step just taken. We can mathematically define H as follows:

$$H_{k\to\ell, i\to j} = \begin{cases} 1 & \text{if } j = k \text{ and } i \neq \ell \\ 0 & \text{otherwise} \end{cases}$$

https://doi.org/10.1017/nws.2019.2 Published online by Cambridge University Press

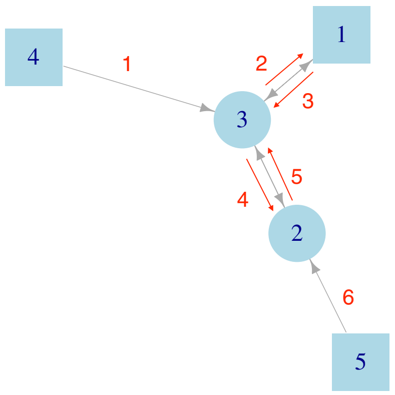

Figure 2. Our toy network with the edges relabeled to create a Hashimoto non-backtracking matrix. Walks can now be represented by a sequence of edges.

For example, let us enumerate the six directed ties in Figure 1; we can label each directed tie with a number from 1 to 6 (see Figure 2). Edge 1 is one walk or step that we can take along the network. Having followed along Edge 1, we can then go to edge 2 or edge 4; the other edges are not possible routes. Then H_{21} and H_{41} would be 1, and the rest of the first column of H would be 0.

The entire Hashimoto non-backtracking matrix for our toy network is given as

$$H = \begin{pmatrix} 0 & 1 & 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \end{pmatrix}$$

The walk along nodes 4–3–2 is represented by the edge sequence 1–4. Now using H instead of Y in our formula for walk counts, we can calculate W_{lff}, for example, as the number of non-backtracking walks from leaders to followers to followers, which would be $H_{32} + H_{12} = 2$.

3.3 Clustering feature vectors

Given a set of feature vectors for each network, we are now interested in exploring how networks vary in their walk counts. In particular, we explore network structure using statistical clustering to determine which networks have feature vectors that are similar to each other and different from other networks in some meaningful way. There are many methods of clustering available for this type of analysis, including k-means clustering (MacQueen , 1967), hierarchical clustering (Ward, 1963), and model-based clustering (Fraley and Raftery, 2002; Wolfe, 1963). With the feature vectors measuring walk counts within a network, it is natural to think about differences in networks by comparing the “distances” between the walk counts. Based on this, and extensive testing via simulation, agglomerative hierarchical clustering with Ward’s linkage and Euclidean distance was the method able to most accurately recover the simulated cluster structure of the networks.

In this clustering method, observations begin as singletons in their own clusters and at each step, clusters are merged together based on some distance-based criterion, until all observations are in one cluster. This creates a set of hierarchical or nested clusters that are easily visualized using a dendrogram (see Figure 3), regardless of the dimensionality of the data. Ward’s linkage merges pairs of clusters based on the minimum increase in the total within-cluster error sum of squares. That is, two clusters will be merged (over a different two clusters) when they produce a

https://doi.org/10.1017/nws.2019.2 Published online by Cambridge University Press

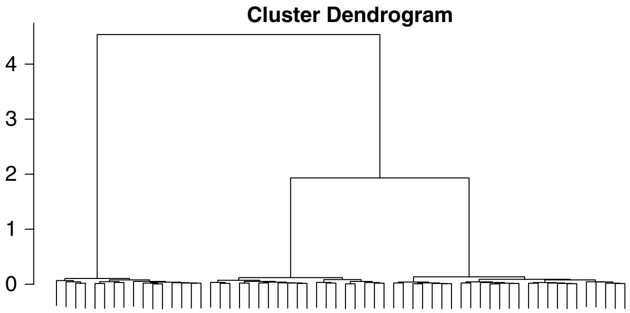

Figure 3. Example dendrogram illustrating a hierarchical clustering solution. Each leaf represents one network. The tree structure illustrates the merges made at each agglomerative step, where the height of the merge is proportional to the increase in within-cluster error sum of squares.

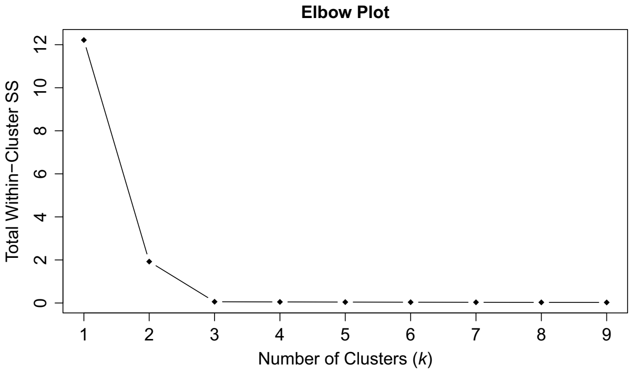

Figure 4. The within-cluster sum of squares versus the number of clusters in a clustering solution shows a clear elbow (turning point) indicating $k = 3$ clusters as a solution.

single cluster with the smallest total within-cluster sum of squared deviations of each observation to the centroid of the cluster. The height of the merge (y-axis) is proportional to the increase in within-cluster error sum-of-squares. In other words, it is a measure of dissimilarity between the clusters. Note that the actual values on the y-axis will be reliant on the data and choice of distance/dissimilarity measure, and therefore should not be compared across different sets of feature vectors.

One disadvantage of hierarchical clustering is that the user must choose where to "cut the tree." In other words, the user must decide how many clusters to choose for the final clustering solution. There is no substitute for expert knowledge so, a researcher should use their content knowledge in addition to the following heuristics to make a practical choice for the final number of clusters, k (Everitt et al., 2011). Visually, one can try to ascertain a reasonable cut based on the dendrogram (Figure 3), by looking for a large vertical distance between branches. Similarly, one can look at an elbow plot of the total within sum of squares versus the number of clusters, with k taken to be the elbow or turning point in the line plot, where the reduction in sum of squares seems to drop off (Figure 4).

Tibshirani et al. (2001) formalized a procedure for finding the "elbow" of such plots using a gap statistic. A gap statistic compares changes in the within-cluster dispersion with that expected under a null distribution, often chosen as a uniform distribution. The number of clusters, k, is chosen as the value where the difference in the within-cluster sum of squares between the clustering solution and reference distribution is maximized, while also taking the standard error into account. Specifically, the number of clusters k is chosen as the smallest k such that $\text{Gap}(k) \geq \text{Gap}(k+1) - se_{k+1}$, where se_{k+1} is the standard error of the gap statistic for $k + 1$ clusters. When there are moderate to large numbers of observations being clustered, the gap statistic method works well for choosing k. When there are fewer observations, however, the gap statistic cannot be used to find k due to the large standard errors of the gap statistic at each k.

In our simulation studies, we find the gap statistic to be reliable with at least 25 simulated networks. For the example in Figure 3 with 60 networks (see Section 4.1), the gap statistic correctly chooses $k = 3$. Because this section produces aggregate results over 50 replications of each scenario, using the gap statistic to choose the optimal number of clusters for each solution is necessary. However, in our application to school network data (see Section 5), we have only 14 schools and so the choice of k needs to be done visually using the dendrogram and expert knowledge.

4. Examples using simulated data

We have proposed clustering feature vectors of walk counts as a method of exploratory data analysis for ensembles of networks. To illustrate the feasibility of this method, we present three simple examples using simulated data. We use simple examples both for understanding and to show how one can recursively generate increasingly more complicated network structures.

To build intuition for the walk count features, our aim is to create separate synthetic network data that cluster by standardized counts of length 1, length 2, and length 3 walks. Continuing with our school example, we generate data such that each node is defined *a priori* to be either a leader or a follower; in a school system, these roles would be defined based on position—a principal would be a leader for example.

We use standard stochastic blockmodels (Holland et al., 1983) to generate network data for our first example and stochastic blockmodels with clique structure for our latter examples. The feature vectors summarizing counts of lengths of walks will then serve as the variables for our clustering algorithm, where each observation is a different network. The algorithm will merge clusters of networks together based on the similarity of the feature vectors, measured using Ward's linkage and Euclidean distance.

4.1 Clustering based on walks of length 1

To generate networks that differ with respect to the normalized walks of length 1, we use a binary nodal attribute indicating whether or not that node is in a leadership position. To generate networks that would then differ in the relative number of walks among leaders and followers (*ll*, *lf*, *fl*, and *ff*), we use a stochastic blockmodel as our generative model. These models provide a natural way of generating ties among and between groups with differential probabilities.

In this example, we simulated networks using a stochastic blockmodel with one of three B matrices that we term highly assortative, less assortative, and highly disassortative. There are three clusters of 20 networks generated for a total of 60 networks. The generative model for a single network is given as

$$Y_{ijk} \sim \text{Bernoulli}(g_{ik}^T B_k g_{jk}), \;\; i, j = 1, \ldots, 50 \tag{4}$$

such that B_k is one of three matrices, $\begin{pmatrix} 0.6 & 0.05 \\ 0.05 & 0.6 \end{pmatrix}$, $\begin{pmatrix} 0.3 & 0.1 \\ 0.1 & 0.3 \end{pmatrix}$, or $\begin{pmatrix} 0.05 & 0.6 \\ 0.6 & 0.05 \end{pmatrix}$, depending on the true network cluster assignment. Furthermore, the distribution of leaders to followers in each network was randomly selected as either 30/30, 20/40, or 40/20; this distribution was independent of cluster assignment.

An example of a simulated network from each of the three clusters is shown in Figure 5; the nodes that are leaders are shown as circles and the followers are shown as squares. In the highly assortative network (left), leaders tend to have ties with only leaders and followers tend to have ties with only followers, and this network is much more dense within leaders and followers than between them. The less assortative network (center) shows this phenomenon to a lesser degree, whereas the highly disassortative network (right) shows an increase in ties between leaders and followers and a decrease in ties among leaders and among followers compared with the highly

https://doi.org/10.1017/nws.2019.2 Published online by Cambridge University Press

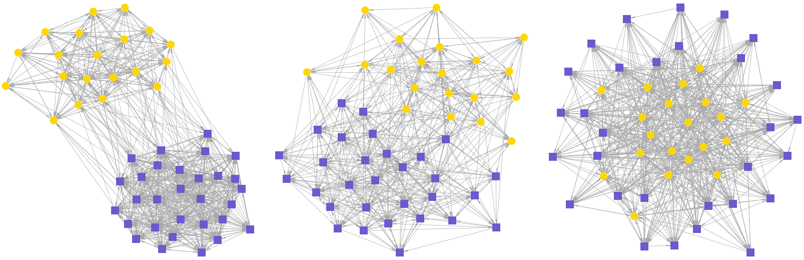

Figure 5. A simulated network from each true network type shows the differences in length 1 walk counts between leaders (circles) and followers (squares).

assortative network. Across these three networks, it is visually apparent that the standardized walk counts of *ll*, *lf*, *fl*, and *ff* would be different across these network types.

For each of the 50 replications of 60 networks, we obtained a clustering solution using hierarchical clustering with Ward's linkage and Euclidean distance. We used the R function clusGap from the cluster package to compute the gap statistic for our simulated networks (Maechler et al., 2018), using 500 bootstrap resamples from a uniform reference distribution. Additionally, we tested solutions between 1 and 8 clusters.

The accuracy of the clustering solution was measured against the true simulated clusters using the adjusted Rand index (ARI; Hubert and Arabie, 1985). This is an index that measures the amount of agreement between any two hard partitions of data, where an ARI of 1 indicates perfect agreement and a value near 0 indicates no agreement. We calculated the average ARI based on 50 replications.

For all 50 replications, the gap statistic chose $k = 3$, which was also confirmed visually, and the resulting clustering solutions were in perfect agreement with the truth, resulting in an ARI of 1 for each replication. Because length 1 walks measure the density of ties within a network with block structure, it is not surprising that the cluster algorithm was able to differentiate between the three classes of block models. When we clustered the length 2 feature vectors within each replication, no cluster structure was found and the average ARI was 0.

4.2 Clustering based on walks of length 2

To create networks that differ based on normalized walks of length 2 and length 3, our aim was to generate networks that differ on longer walk counts even when adjusting for walks of shorter lengths; that is, our goal was to generate networks that have relatively different numbers of walks, for example, *lll* versus *llf* ties, despite having similar numbers of *ll* and *lf* ties.

To generate networks with differences in longer walk counts, we created a network-generating model that would ensure that either triads or pairs of edges with an adjacent node would appear regularly in the network by generating networks with clique structure. A clique is set of nodes that all share ties, so networks with cliques would have nonzero counts of longer walks.

Thus, our data-generating model for one cluster of networks was a random graph model with tie probability 0.2. The second cluster of networks had added clique structure among its leaders and the third cluster had added clique structure among its followers. That is, we selected a subset of the leaders (or followers) and generated ties with higher tie probability among a smaller subset of nodes. Note that all networks generated had equal expected density of 0.2.

In addition, we varied network size, which was sampled from a truncated normal distribution centered at 50 with a variance of 25 and minimum value 20, and the number of leaders, which we sampled from a normal distribution centered at 35% of the network size with a variance of 4. The tie probability within the clique among leaders and followers in network clusters 1 and 2, respectively, is 0.25, so the added clique structure was fairly weak.

https://doi.org/10.1017/nws.2019.2 Published online by Cambridge University Press

Table 2. Summary statistics for the ARI across the 50 replications, for clustering length 2 walks

	Min.	First Quartile	Median	Mean	Third Quartile	Max.
Length 2	0.776	0.950	1.000	0.970	1.000	1.000

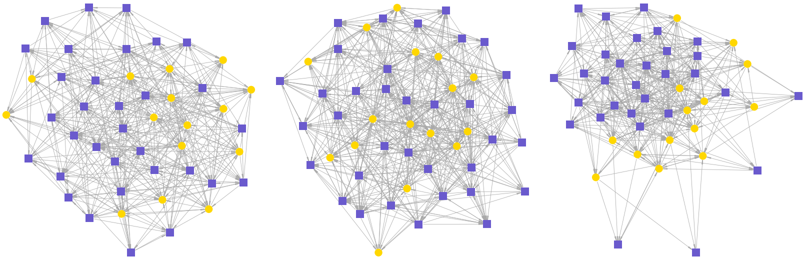

Figure 6. To simulate networks to cluster based on length 2 walks, two of the true groups generated clique structure among leaders (left) and followers (right).

Figure 6 shows examples of the simulated networks, one from each of the three groups. The network on the left shows additional clique structure among the leaders, the network in the center is a random graph, and the network on the right shows clique structure among the followers. Note that all three networks had an expected density of 0.2. Although it is less obvious, Figure 6 does convey differences in some walk counts of length 2, namely *lll* and *fff*, that would differentiate clusters.

When clustering walks of length 2, the gap statistic correctly chose three clusters for each replication, and the clustering solutions produced an average ARI of 0.971. Summary statistics for the ARI across replications are shown in Table 2. Additionally, we clustered the length 1 walks for this simulation, and the gap statistic found no cluster structure in any of the 50 replications for an average ARI of 0.

4.3 Clustering based on walks of length 3

To simulate network data with longer walks, we used the same generating model as in Section 4.2: one group had clique structure among the leaders, the second group had clique structure among the followers, and the third group was a random graph. Note that in Section 4.2, we generated cliques with slightly higher tie probability of 0.25, only slightly higher than the overall density of 0.2. To simulate networks with nonzero counts of *llll* for example, our networks required clique structure that was more dense and in this simulation, the tie probability among the subset of leaders (or followers) was 0.90. Again, we varied network size and proportion of leaders as we did in the previous simulation. We generated 60 networks (20 of each group) and repeated the simulation 50 times.

Figure 7 shows examples of the simulated networks, and the strong clique structure is quite apparent, particularly in the network on the right, whose generating model incorporated dense clique structure among a subgroup of followers. The network on left shows high clique density among leaders; there are a small number of squares that are well connected. The network in the middle is again a random graph.

For this simulation, the gap statistic chooses 1 (2 replications), 4, 5, or 6 clusters for each replication. For the solutions with four or more clusters, a simulated network type is being divided into a greater number of smaller clusters. Due to the nature of the ARI counting pairs of observations in the same class, this still produces a clustering solution that has high agreement with the true simulated classes. The average ARI for clustering walks of length 3 is 0.809. Summary statistics for the ARI across the 50 replications are shown in Table 3.

https://doi.org/10.1017/nws.2019.2 Published online by Cambridge University Press

Table 3. Summary statistics for the ARI across the 50 replications, for clustering length 3 walks

	Min.	First Quartile	Median	Mean	Third Quartile	Max.
Length 3	0.000	0.820	0.832	0.809	0.860	0.900

Table 4. Summary statistics for the ARI across the 50 replications, for clustering length 1–3 walks

	Min.	First Quartile	Median	Mean	Third Quartile	Max.
Length 1–3	0.800	0.827	0.839	0.846	0.864	0.900

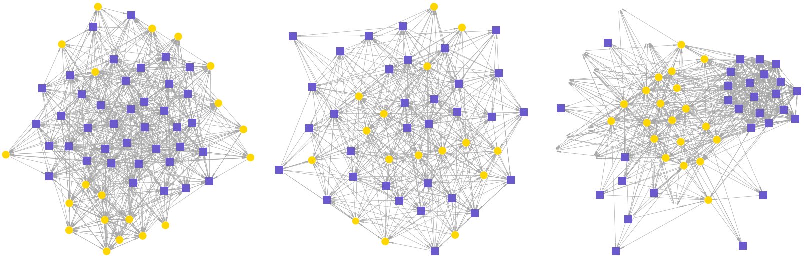

Figure 7. To simulate networks to cluster based on length 3 walks, more extreme clique structure is added. Two of the true clusters generated clique structure among leaders (left) and followers (right) such that expected clique density is 0.9.

Rather than pulling out the length 3 walks, we can cluster the networks using all of their walks from length 1 to length 3. We find similar clustering solutions when compared to those for just the length 3 walks. The average ARI for clustering networks based on walks of length 1, 2 and 3 is 0.846, and the summary statistics are shown in Table 4.

5. Application

We now apply our method of clustering networks to real-world network data, again focusing on length 1, 2, and 3 walks. These data come from a longitudinal study of a midwestern elementary school district consisting of 14 schools that have been used in a number of studies in education research (e.g., Spillane and Hopkins, 2013; Spillane et al., 2016, 2017) and are a part of a larger study on distributed leadership (www.distributedleadership.org).

Teachers were surveyed each spring in 2010–2013 and 2015, with response rates of 81%, 95%, 94%, 94%, and 96%. Due to low response rates in 2010 and the issues that arise in how one deals with missing data (e.g., Žnidaršič et al., 2017), we will be using the 2011–2015 data in this analysis. In addition to survey items regarding their formal position in the school—teacher, administrator, instructional coach, etc.—teachers were asked to nominate those staff members to whom they go to for advice or information around instruction. These nominations were recorded for a number of subjects, where we focus on the advice-seeking ties around mathematics.

We also classified nodes as *leaders* or *followers* (i.e., non-leaders) based on their formal assigned role in the school; principals, assistant principals, department/grade-level chairs, and instructional coaches were all coded as leaders and other school staff were coded as followers. Note that we are using the term "followers" for consistency in this manuscript and are not attempting to characterize non-leaders in a particular way.

As an example, Figure 8 shows the mathematics networks from 2015 where leaders and followers, as defined by their formal staff position, are discerned by color and shape; the orange circles represent the formal leaders and the purple squares depict the followers. There is both variability

https://doi.org/10.1017/nws.2019.2 Published online by Cambridge University Press

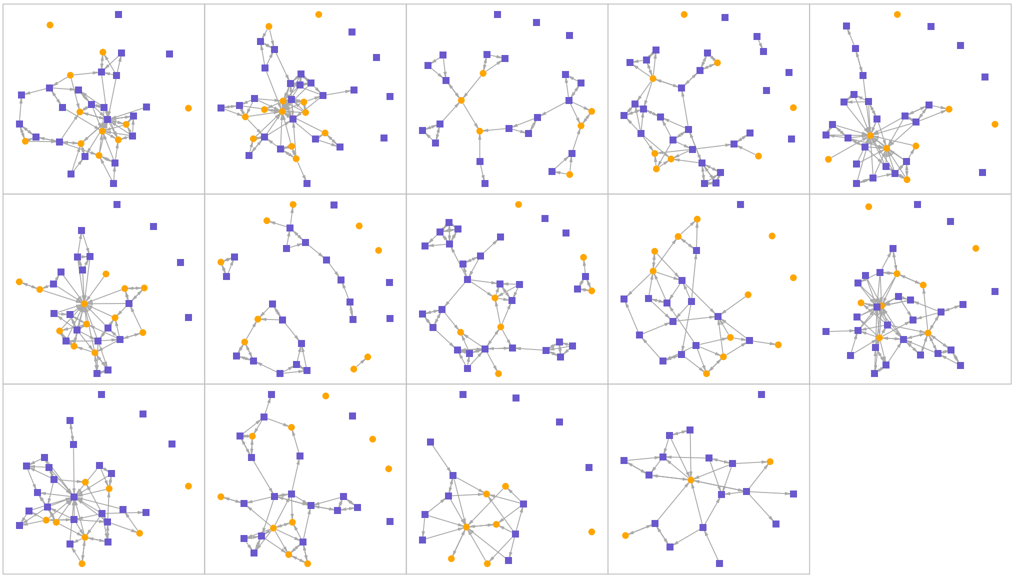

Figure 8. Advice-seeking networks among staff in 14 elementary schools. Nodes are colored and shaped based on formal assignment; orange circle nodes are leaders and purple square nodes are followers (non-leaders).

in network structure across these schools as well as variability in the placement of leaders within their networks. There are some schools in which leaders appear to be quite central and other schools where some leaders appear somewhat marginalized.

Our goal then is to see if these networks cluster based on the relative counts of length 1, 2, and 3 walks among leaders and followers. Understanding whether and how schools differ based on how leaders and followers interact is relevant to current research on distributed leadership (Harris, 2009; Harris and Spillane, 2008; Spillane, 2012; Spillane et al., 2001). Further, we know of some major policy changes that took place during these years. The entire district began implementing a new mathematics curriculum in 2010 and to aid in this transition, several schools were assigned instructional coaches. Two schools received coaches in 2011, five additional schools gained coaches in 2013, and one school received a coach in 2015.

Schools vary in terms of network size as well as the proportion of formal leaders in the school. It is surprising that schools assigned an instructional coach (a formal leadership position) do not have higher proportions of leaders than schools without coaches. Of the seven schools reporting 30% or higher formal leaders, only two are schools that employed an instructional coach. Similarly, the schools who received coaches do not appear to differ in terms of network size.

Using our standardized, non-backtracking walk-counting algorithm described in Sections 3.1 and 3.2, we created feature vectors based on the mathematics advice-seeking networks in 2011–2013, and 2015. We then used hierarchical clustering (see Section 3.3) to produce the dendrograms shown in Figures 9, 10, and 11 for the length 1, 2, and 3 walks, respectively. Furthermore, the final clustering solutions were based on choosing reasonable cut points in the dendrograms.

Length 1 walk counts (standardized). Figure 9 shows the network cluster assignments based on length 1 walks among leaders and followers. Note that the clustering algorithm results in many solutions since one could cut the dendrogram at any height, but a two-cluster solution does appear most likely in years 2011, 2013, and 2015, along with a two-cluster solution in 2012 after removing School 1. What is particularly interesting is that schools with coaches tend to cluster together. For example, 2011 (Figure 9, top left) was the first year Schools 1 and 2 had coaches and these were

https://doi.org/10.1017/nws.2019.2 Published online by Cambridge University Press

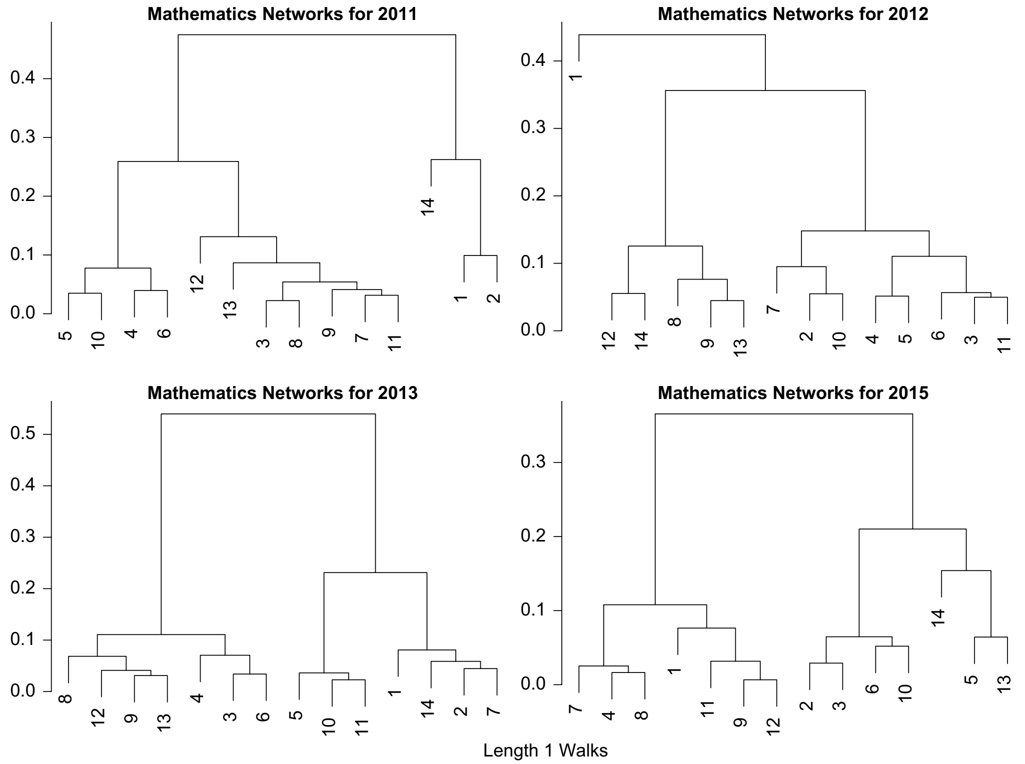

Figure 9. Dendrograms of Mathematics networks clustering length 1 walks.

the only schools with coaches. Schools 1, 2, and 14 appear to be quite different from the other 11 schools that year and in fact this cluster differs from the other cluster of schools in the relative counts of walks from followers to leaders and leaders to leaders.

Schools 1 and 2 were again the only schools with coaches in 2012 (Figure 9, top right); however, the clustering solution does not reflect this information. School 1 differs from the other 13 schools in the relative numbers of *fl* and *ll* walks. When the cluster of 13 schools is broken down again, the cluster on the left has fewer *ll* ties and more *ff* ties.

In 2013 (Figure 9, bottom left), Schools 5, 6, 9, 10, and 11 were in their first year of coaches, and some of these schools tend to cluster together but the more obvious two-cluster solution groups schools with coaches and schools without coaches together. The cluster on the left has fewer *fl* and *ll* walks than the cluster of schools on the right.

In 2015 (Figure 9, bottom right), School 1's coach went back to School 1 as a teacher; Schools 2, 5, 6, 10, 11, and 13 had coaches. Again, the two-cluster solution appears to group Schools 2, 5, 6, 10, and 13 together, which are five of the six schools with coaches that year. The cluster with more coaches also had higher relative counts of *fl* walks, whereas *ff* interactions were relatively more frequent in the cluster without coaches.

Although there is variability across years in cluster solutions, there are several patterns to note. First, schools with coaches do tend to cluster together and these clusters generally have higher counts of ties from followers to leaders. This means that followers are seeking advice from leaders in these schools, most likely because teachers are going to coaches for advice around mathematics. The fact that a single instructional coach is associated with structural differences between networks (as related to walks from followers to leaders) suggests that instructional coaches may be quite influential in how leadership is organized within a school.

https://doi.org/10.1017/nws.2019.2 Published online by Cambridge University Press

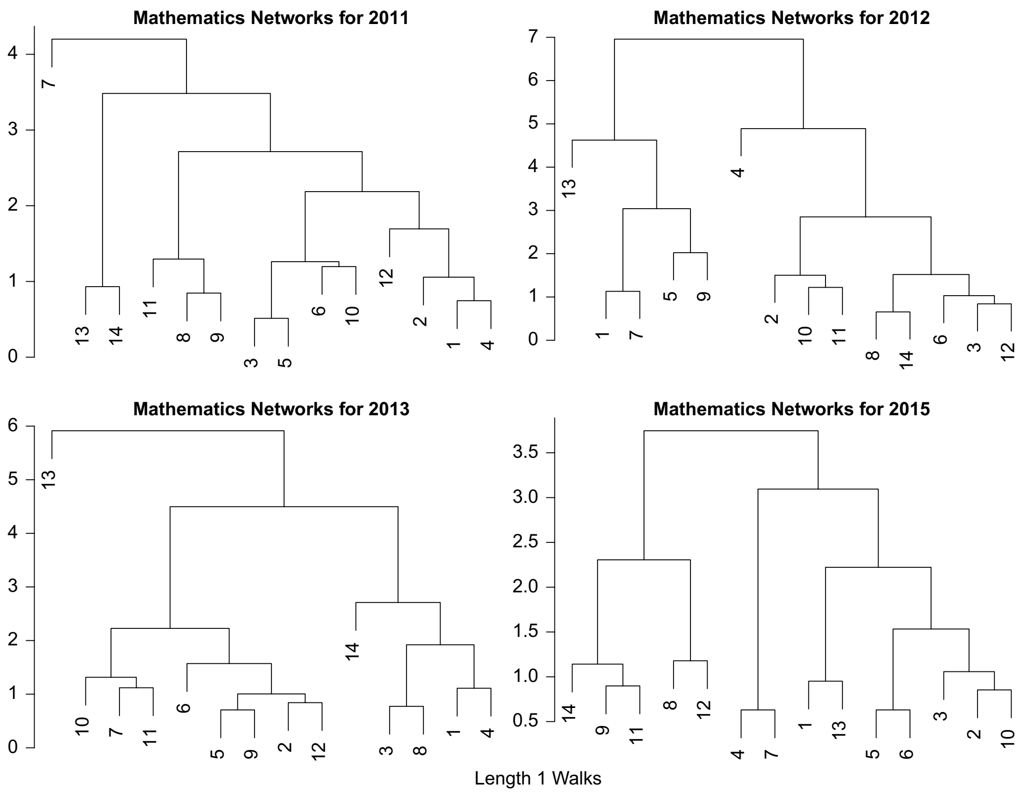

Figure 10. Dendrograms of Mathematics networks clustering length 2 walks.

Length 2 walk counts (standardized). Figure 10 shows the dendrograms for the length 2 walk clustering solutions in the years 2011–2013 and 2015. Here the cluster analysis results do not appear to align with the presence or absence of coaches. Additionally, the dendrograms for years 2011–2012 and 2015 do not suggest a clear choice for the number of clusters or where to cut the tree. Intuitively, this seems consistent with the nature of the networks: because coaches tend to advice providers and because these are directed networks, walks of length 2 involving a coach may be less common and less likely to differentiate networks.

In 2011 (Figure 10, top left), Schools 1 and 2 appear in the same final cluster although several clustering solutions are possible. Assuming a three-cluster solution, differences across the three final clusters involve counts of triads involving leaders and followers (e.g., *lfl*, *flf*). In 2012 (Figure 10, top right), the cluster on the left has higher relative counts of *llf* and *lll* walks than the cluster on the right, and in 2013 (Figure 10, bottom left), School 13 has more *llf* and *flf* than the other schools. In 2015, there are arguably anywhere between 2 and 6 clusters (Figure 10, bottom right). If we compare the two groups on the left with the four on the right, the biggest differences are in the relative counts of *llf*.

Recall that in Section 4.2, we show in simulations that networks with varying amounts of clique structure can cluster well on length 2. In our current data example, the lack of "interesting" clustering results for the standardized walks of length 2 seems to agree with the intuition that advice-seeking is not a relationship that is associated with cliques among elementary school teachers; that is, unlike friendship or collaboration, advice-seeking ties are probably less likely to result in dense subgroup structure among leaders or followers. In fact, Spillane et al. (2015) found under a variety of conditions that advice-seeking occurs primarily when the advice provider is in a leadership position.

https://doi.org/10.1017/nws.2019.2 Published online by Cambridge University Press

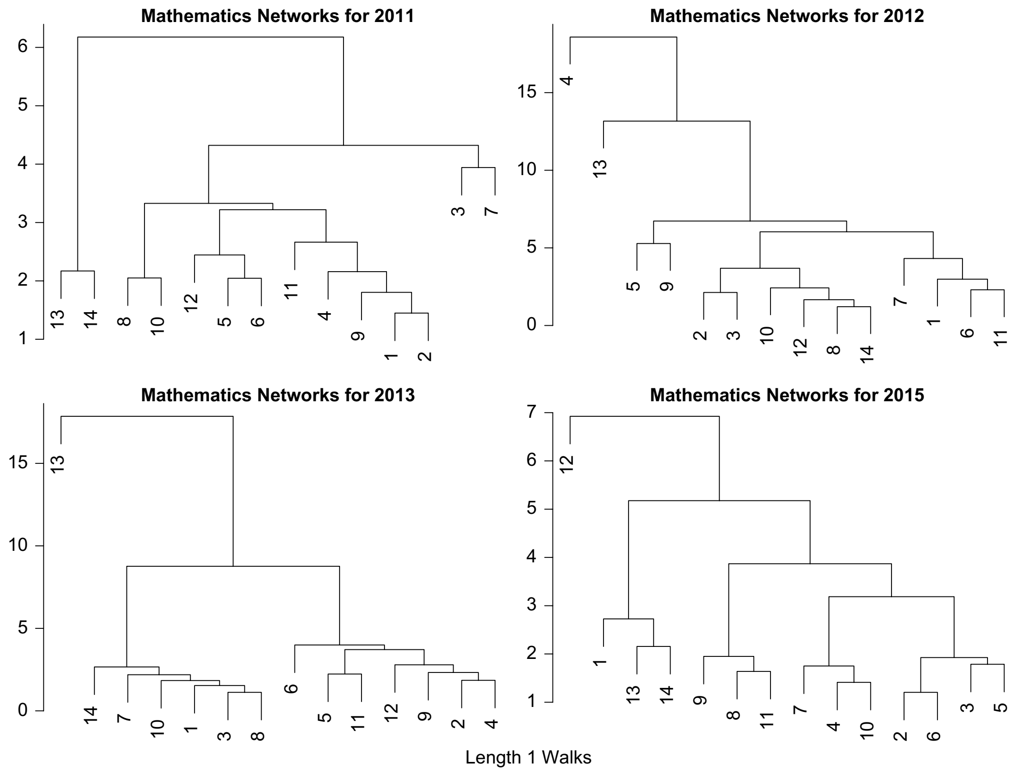

Figure 11. Dendrograms of Mathematics networks clustering length 3 walks.

Length 3 walk counts (standardized). Given the lack of cluster structure from length 2 walks, we again do not see strong cluster structure based on length 3 walk counts (Figure 11). In fact, the dendrograms for length 3 walks resemble the dendrograms for the length 2 walks (Figure 10). Any obvious separation among schools does not appear to be related to coaches in the schools and in fact the strongest cluster solutions result in 1 or 2 schools separated from the rest of the schools.

In 2011, Schools 13 and 14 had a relatively high walk counts of *lffl* and *lfff* and low counts of most other types of walks. In 2012, School 4 has high counts of *lfll*, *ffll*, and *lflf*, and School 13 had high counts of *fllf* and *lllf* compared to the other 12 schools. School 13 was different from other schools in 2013 in that it had much higher counts of almost every type of walk, and in 2015, School 12 had higher counts of *lfll* and *llfl* than the other schools.

The purpose of this application was to illustrate and provide a motivating example for our walk-counting algorithm which now allows researchers to explore differences among schools based on how leaders and non-leaders interact. That we found differences in how leaders and non-leaders interact in schools with coaches is of particular interest to those who study leadership within schools, and this methodology in general will be of interest to researchers interested in comparing organizations based on interactions among certain types of people. For example, our walk-counting algorithm can be used to count interactions among individuals of different races, and researchers can cluster organizations with respect to how people of certain races interact.

6. Considering unstandardized walk counts

One of the contributions of this paper is our focus on counting standardized, non-backtracking walks for which we developed efficient algorithms. Further, we argue that standardized and non-backtracking walk counts should be used for clustering social network data. The argument for

https://doi.org/10.1017/nws.2019.2 Published online by Cambridge University Press

non-backtracking walk counts may be obvious since backtracking allows any undirected or bidirectional edge between two nodes to count towards increasingly long walk counts; that is, we found increasing counts of walks of increasing length, a phenomenon we would not expect in a social network.

The argument for standardizing the counts of non-backtracking walks, however, is less obvious, and in this section, we explore, through an example why it is necessary to standardize walk counts to properly cluster social networks. Using the mathematics advice-seeking networks among school staff introduced in Section 5, we construct new feature vectors for the 2011–2013 and 2015 networks using the *unstandardized* walk counts and employ the same methods to cluster the 14 schools based on length 1, length 2, and length 3 walks separately.

Figure 12 shows the cluster solutions for the 14 schools when clustering based on length 1 walks (left), length 2 walks (center), and length 3 walks (right) for each of the four years of network data. Note the similarity in cluster solutions using walks of different length. Most of the small clusters formed are the same. In fact, the cluster solutions differ by at most two schools. This was not the case in the standardized count cluster solutions shown in Figures 9–11.

What these similarities in clustering solutions suggest is that the unstandardized walk counts of length 1, 2, or 3 capture largely the same information. When comparing Figure 12 with the standardized solutions given in Figures 9–11, we find that the unstandardized walk counts appear to be perhaps a combination of the standardized clustering solutions. Thus, without standardizing walk counts, one cannot separate networks based specifically on walks of length k. Treating walks of any count as equivalent is much less useful because it disqualifies further exploration into why some networks are clustering together and what the differences within and between clusters are in terms of specific walks between leaders and followers in our example, or among nodes of various attributes more generally.

7. Discussion

We introduced a new method for exploring differences in network structure across ensembles of social networks by clustering feature vectors. Specifically, this paper motivated and introduced algorithms for counting walks in a network: a method for standardizing walk counts based on node attributes and a method for counting only non-backtracking walks. We applied these methods to count walks among leaders and followers in various social network ensembles, where network sizes vary and both counts and proportions of leaders varied.

In particular, we demonstrated how one can cluster networks using feature vectors of standardized non-backtracking walk counts for walks of length 1, 2, and 3, and we illustrated the feasibility of this method through a series of simulated data examples in which clusters based on counts of walks were recovered with high accuracy. In addition, we also applied this method to real-world data to illustrate how social networks could be clustered in practice. We found evidence that clusters based on length 1 walks align with the presence or absence of instructional coaches since schools with coaches tend to have relatively more ties from followers to leaders than schools without coaches. Clusters based on length 2 or length 3 walks did not have a similar contextual explanation.

We demonstrated with the use of the mathematics advice-seeking networks that clustering unstandardized walk counts will not produce unique solutions based on the length of the walk count. This is meaningful if we want to relate counts of walks of a certain length with network differences, and this is particularly important in networks with different numbers of nodes and different distributions of nodal attributes.

In addition to our methodological contributions, we also present this paper as a general proof of concept of exploratory methods for ensembles of social networks. That is, cluster analysis based on feature vectors of walk counts is simply one example of such exploration. Given the abundance of network structures that may be of interest, counts of other features could be used

https://doi.org/10.1017/nws.2019.2 Published online by Cambridge University Press

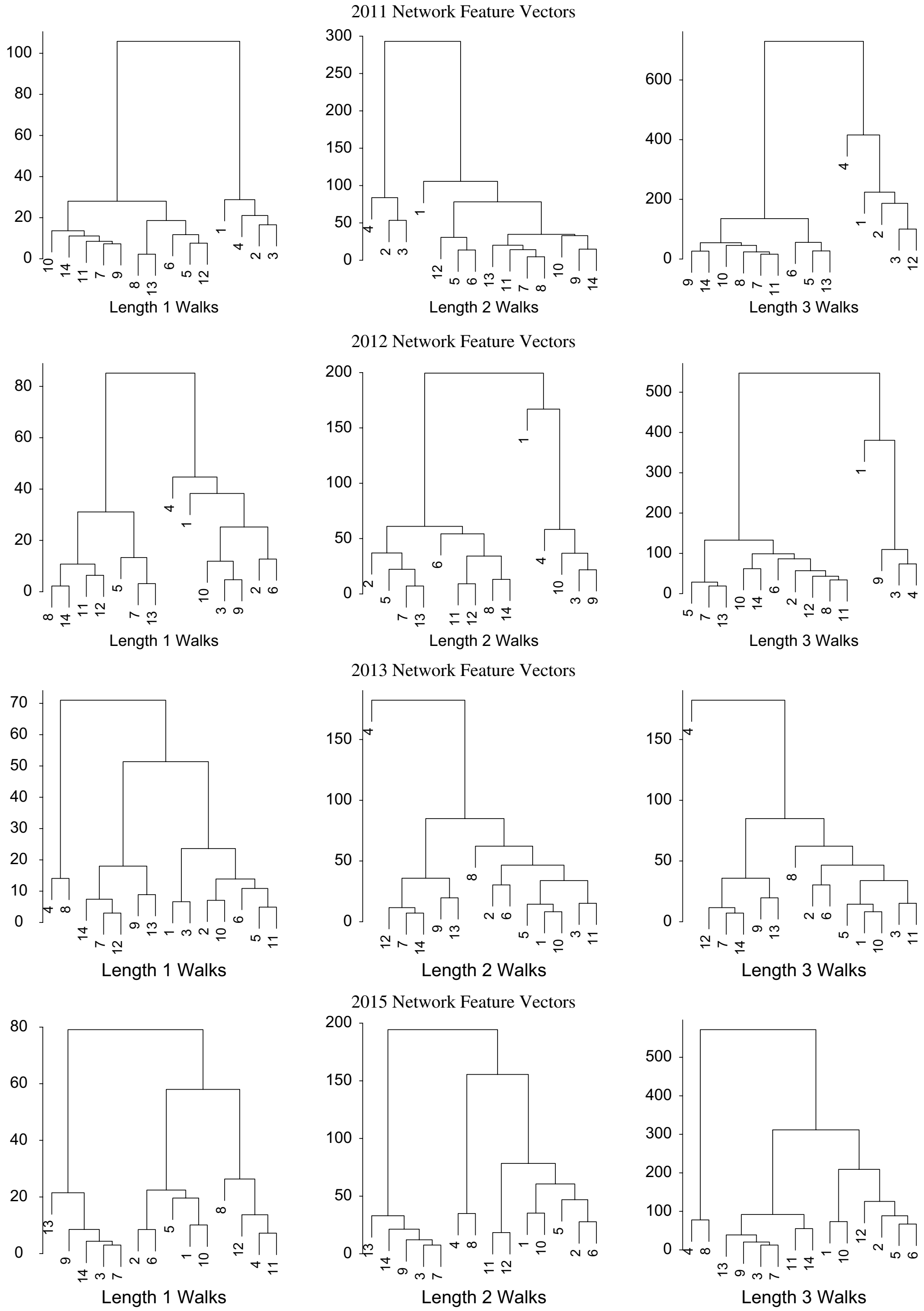

Figure 12. Dendrograms of Mathematics networks clustering unstandardized length 1–3 walks.

for feature vectors such as reciprocated dyads, triads, cliques, or k-stars, assuming sufficient attention is paid to both standardizing these counts and potentially high correlations among counts of these structures.

Further, given feature vectors for each network, there are a number of other classification methods that can be applied. For example, model-based clustering fits a mixture model to the data and produces a likelihood value allowing for the number of clusters to be selected using a measurement like the Bayesian Information Criterion.

Finally, we argue that future research on exploratory methods on ensembles of social networks in particular is both interesting and warranted. First, there are types of structures found in social networks not seen in other networks. Social networks vary in size, structure, as well as nodal attributes. In addition, there are specific structure/phenomena observed in social networks (e.g., homophily, small world, preferential attachment) that make exploratory analyses across networks even more complicated and rewarding.

Acknowledgments. We would like to acknowledge Jim Spillane and the Distributed Leadership Studies (http://www.distributedleadership.org) at Northwestern University, including the NebraskaMATH study, for the use of their data in this work. This work was supported in part by the US Department of Education, Institute of Education Sciences, under grant no. R305D150045. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the granting agencies.

Conflict of interest. The authors have nothing to disclose.

References

Everitt, B., Landau, S., Leese, M., & Stahl, D. (2011). *Cluster analysis*. Wiley Series in Probability and Statistics. Hoboken: John Wiley & Sons.

Faust, K., & Skvoretz, J. (2002). Comparing networks across space and time, size and species. *Sociological Methodology*, *32*, 267–299.

Fraley, C., & Raftery, A. E. (2002). Model-based clustering, discriminant analysis, and density estimation. *Journal of the American statistical Association*, *97*, 611–631.

Frank, K. A., Lo, Y.-J., & Sun, M. (2014). Social network analysis of the influences of educational reforms on teachers practices and interactions. *Zeitschrift für Erziehungswissenschaft*, *17*, 117–134.

Frank, K. A., Zhao, Y., & Borman, K. (2004). Social capital and the diffusion of innovations within organizations: The case of computer technology in schools. *Sociology of Education*, *77*, 148–171.

Friedman, J., Hastie, T., & Tibshirani, R. (2001). *The elements of statistical learning* (vol. 1). New York: Springer.

Gest, S. D., & Rodkin, P. C. (2011). Teaching practices and elementary classroom peer ecologies. *Journal of Applied Developmental Psychology*, *32*, 288–296.

Harris, A. (2009). Distributed leadership: What we know. In *Distributed leadership* (pp. 11–21). Dordrecht: Springer.

Harris, A., & Spillane, J. (2008). Distributed leadership through the looking glass. *Management in Education*, *22*, 31–34.

Harris, K., Halpern, C., Whitsel, E., Hussey, J., Tabor, J., Entzel, P., & Udry, J. (2009). The national longitudinal study of adolescent health. Research design. Retrieved from http://www.cpc.unc.edu/projects/addhealth/design (September 2011).

Hashimoto, K.-i. (1989). Zeta functions of finite graphs and representations of p-adic groups. *Automorphic Forms and Geometry of Arithmetic Varieties*, *15*, 211–280.

Holland, P., Laskey, K., & Leinhardt, S. (1983). Stochastic blockmodels: First steps. *Social Networks*, *5*, 109–137.

Hopkins, M., Lowenhaupt, R., & Sweet, T. M. (2015). Organizing instruction in new immigrant destinations: District infrastructure and subject-specific school practice. *American Educational Research Journal*, *52*, 408–439.

Hubert, L. & Arabie, P. (1985). Comparing partitions. *Journal of Classification*, *2*, 193–218.

Kashima, H., Tsuda, K., & Inokuchi, A. (2003). Menlo Park: AAAI Press. Marginalized kernels between labeled graphs. In *Proceedings of the 20th international conference on machine learning (ICML-03)* (pp. 321–328).

Krzakala, F., Moore, C., Mossel, E., Neeman, J., Sly, A., Zdeborová, L., & Zhang, P. (2013). Spectral redemption in clustering sparse networks. *Proceedings of the National Academy of Sciences*, *110*, 20935–20940.

Lazega, E., & Snijders, T. A. (2015). *Multilevel network analysis for the social sciences: Theory, methods and applications* (vol. 12). Berlin, Germany: Springer.

MacQueen, J., (1967). Some methods for classification and analysis of multivariate observations. In *Proceedings of the fifth Berkeley symposium on mathematical statistics and probability*, Oakland, CA, USA (vol. 1, pp. 281–297). Berkeley: University of California Press.

Maechler, M., Rousseeuw, P., Struyf, A., Hubert, M., & Hornik, K. (2018). *cluster: Cluster analysis basics and extensions*.

https://doi.org/10.1017/nws.2019.2 Published online by Cambridge University Press

Martin, T., Zhang, X., & Newman, M. (2014). Localization and centrality in networks. *Physical Review E*, *90*, 052808.

Paluck, E. L., & Shepherd, H. (2012). The salience of social referents: A field experiment of collective norms and harassment behavior in a school social network. *Journal of Personality and Social Psychology*, *103*, 899–915.

Ralaivola, L., Swamidass, S. J., Saigo, H., & Baldi, P. (2005). Graph kernels for chemical informatics. *Neural Networks*, *18*, 1093–1110.

Saigo, H., Nowozin, S., Kadowaki, T., Kudo, T., & Tsuda, K. (2009). gBoost: A mathematical programming approach to graph classification and regression. *Machine Learning*, *75*, 69–89.

Sarkar, A., Fienberg, S., & Krackhardt, D. (2010). Predicting profitability using advice branch bank networks. *Statistical Methodology*, *7*, 429–444.

Sinani, E., Stafsudd, A., Thomsen, S., Edling, C., & Randøy, T. (2008). Corporate governance in Scandinavia: Comparing networks and formal institutions. *European Management Review*, *5*, 27–40.

Snijders, T., & Kenny, D. (1999). The social relations model for family data: A multilevel approach. *Personal Relationships*, *6*, 471–486.

Snijders, T. A., & Baerveldt, C. (2003). A multilevel network study of the effects of delinquent behavior on friendship evolution. *Journal of Mathematical Sociology*, *27*, 123–151.

Snijders, T. A., Steglich, C. E., Schweinberger, M., & Huisman, M. (2008). *Manual for SIENA version 3.2.* Department of Sociology, ICS, University of Groningen, Groningen, The Netherlands.

Spillane, J., Hopkins, M., & Sweet, T. (2015). Intra- and inter-school instructional interactions: Exploring conditions for instructional knowledge production within and between schools. *American Journal of Education*, *122*, 71–110.

Spillane, J. P. (2012). *Distributed leadership* (vol. 4). San Francisco: John Wiley & Sons.

Spillane, J. P., Halverson, R., & Diamond, J. B. (2001). Investigating school leadership practice: A distributed perspective. *Educational Researcher*, *30*, 23–28.

Spillane, J. P., & Hopkins, M. (2013). Organizing for instruction in education systems and school organizations: How the subject matters. *Journal of Curriculum Studies*, *45*, 721–747.

Spillane, J. P., Hopkins, M., & Sweet, T. M. (2016). Exploring the relationship between teachers' instructional ties and teachers' instructional beliefs: Trying not to 'put the cart before the horse'. *American Journal of Education*, *122*, 71–110.

Spillane, J. P., Shirrell, M., & Sweet, T. M. (2017). The elephant in the schoolhouse: The role of propinquity in school staff interactions about teaching. *Sociology of Education*, *90*, 149–171.

Sweet, T., & Zheng, Q. (2017). A mixed membership model-based measure for subgroup integration in social networks. *Social Networks*, *48*, 169–180.

Sweet, T. M., Thomas, A. C., & Junker, B. W. (2013). Hierarchical network models for education research: Hierarchical latent space models. *Journal of Educational and Behavioral Statistics*, *38*, 295–318.

Tibshirani, R., Walther, G., & Hastie, T. (2001). Estimating the number of clusters in a data set via the gap statistic. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, *63*, 411–423.

Traud, A. L., Kelsic, E. D., Mucha, P. J., & Porter, M. A. (2011). Comparing community structure to characteristics in online collegiate social networks. *SIAM Review*, *53*, 526–543.

Vishwanathan, S. V. N., Schraudolph, N. N., Kondor, R., & Borgwardt, K. M. (2010). Graph kernels. *Journal of Machine Learning Research*, *11*, 1201–1242.

Vogelstein, J. T., Roncal, W. G., Vogelstein, R. J., & Priebe, C. E. (2013). Graph classification using signal-subgraphs: Applications in statistical connectomics. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *35*, 1539–1551.

Ward Jr., J. H. (1963). Hierarchical grouping to optimize an objective function. *Journal of the American Statistical Association*, *58*, 236–244.

Wolfe, J. H. (1963). *Object cluster analysis of social areas*, Ph.D. thesis, University of California.

Zijlstra, B., van Duijn, M., & Snijders, T. (2006). The multilevel p2 model. *Methodology: European Journal of Research Methods for the Behavioral and Social Sciences*, *2*, 42–47.

Žnidaršič, A., Ferligoj, A., & Doreian, P. (2017). Actor non-response in valued social networks: The impact of different non-response treatments on the stability of blockmodels. *Social Networks*, *48*, 46–56.

Cite this article: Sweet T. M., Flynt A., and Choi D. (2019). Clustering ensembles of social networks. *Network Science* **7**, 141–159. https://doi.org/10.1017/nws.2019.2