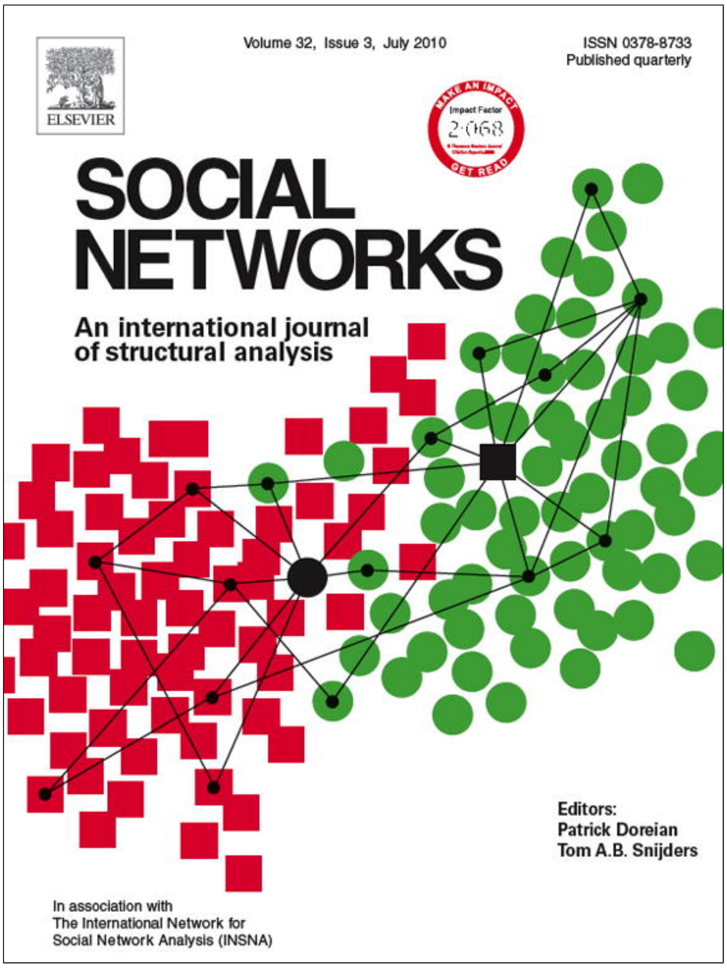
Volume 32, Issue 3, July 2010
ISSN 0378-8733
Published quarterly
ELSEVIER
SOCIAL NETWORKS
An international journal of structural analysis
Editors:
Patrick Doreian
Tom A.B. Snijders
In association with
The International Network for
Social Network Analysis (INSNA)

Social Networks 32 (2010) 234–244

Contents lists available at ScienceDirect

Social Networks

journal homepage: www.elsevier.com/locate/socnet

A new method for finding hierarchical subgroups from networks

Mo-Han Hsieh [a,*], Christopher L. Magee [b]

[a] Engineering Systems Division, Massachusetts Institute of Technology, Cambridge, MA 02139, USA
[b] Engineering Systems Division & Department of Mechanical Engineering, Massachusetts Institute of Technology, Cambridge, MA 02139, USA

ARTICLE INFO

Keywords:
Positional analysis
Hierarchical subgroups
Automorphic equivalence
Decomposability
k-Means method
Generalized Blockmodeling

ABSTRACT

We present a new method for decomposing a social network into an optimal number of hierarchical subgroups. With a perfect hierarchical subgroup defined as one in which every member is automorphically equivalent to each other, the method uses the REGGE algorithm to measure the similarities among nodes and applies the *k*-means method to group the nodes that have congruent profiles of dissimilarities with other nodes into various numbers of hierarchical subgroups. The best number of subgroups is determined by minimizing the intra-cluster variance of dissimilarity subject to the constraint that the improvement in going to more subgroups is better than a network whose *n* nodes are maximally dispersed in the *n*-dimensional space would achieve. We also describe a decomposability metric that assesses the deviation of a real network from the ideal one that contains only perfect hierarchical subgroups. Four well known network data sets are used to demonstrate the method and metric. These demonstrations indicate the utility of our approach and suggest how it can be used in a complementary way to Generalized Blockmodeling for hierarchical decomposition.

1. Introduction

In sociology and anthropology, viewing society as a hierarchical set of social classes, castes, or strata has a long tradition (Lynd and Lynd, 1937; Warner and Lunt, 1941). To study people, groups, or organizations, they are not ranked on a continuous scale of prestige but are often classified into a limited set of discrete ranks, for example, upper class, middle class, and lower class, and each class is further subdivided into smaller classes related to occupation (Saunders, 1990).

The initiation of hierarchy theory is often attributed to Simon (1962), and some consider his contribution to be in the research tradition of general systems theory (Bertalanffy, 1950). Since then, two major views for understanding hierarchy have been developed. The first takes the ontological position and deals with the "ultimate reality" of structure (Salthe, 1985). The second takes the stance that there may be an external reality, but it is not relevant to the discourse because we only have access to subjective experience (Ahl and Allen, 1996). In this research, we follow the ontological approach and study not only a narrowly defined hierarchy, in which every subordinate reports to only one superior, but also directed networks in which the relations among subordinates are pervasive and complex.

In the social networks literature, two lines of research have been developed to understand the hierarchical properties of social relations. These are: (1) the local approach that uses local structure to infer the overall nature of the network (Johnsen, 1985); and (2) the global approach that decomposes the network according to a certain set of pre-specified criteria (Doreian et al., 2000, 2005). The local approach continuously develops the concept of transitivity (Holland and Leinhardt, 1971) for explaining social structures, and the global approach has Generalized Blockmodeling (Doreian et al., 2005) as its latest development, where the hierarchical structure of the networks can be described by a variety of blocks in specific patterns.

In this study, we take the global approach and examine the notion of hierarchical subgroups within the broader research tradition of positional and role analysis (White et al., 1976). We propose a new definition for hierarchical subgroups of a network that utilizes the concept of automorphic equivalence (Everett et al., 1990; Pattison, 1982, 1988; Winship, 1988; Winship and Mandel, 1983). Automorphism occurs when a re-labeling of nodes in a network preserves the network's structure; thus, two actors are automorphically equivalent if and only if there is an automorphism that maps one of the nodes to the other.[1] We define a *perfect* hierarchical subgroup as one whose members are all automorphically equivalent to each other. Moreover, we partition a network into hierarchical subgroups that place automorphically equivalent or

* Corresponding author at: 77 Massachusetts Avenue, E38-454, Cambridge, MA 02139-4307, USA. Tel.: +1 857 234 6570.
E-mail addresses: mohan76@mit.edu (M.-H. Hsieh), cmagee@mit.edu (C.L. Magee).

[1] More information about automorphic equivalence can be found in Wasserman and Faust (1994).

0378-8733/$ – see front matter © 2010 Elsevier B.V. All rights reserved.
doi:10.1016/j.socnet.2010.03.005

nearly automorphically equivalent nodes in the same subgroup. Since in reality perfect hierarchical subgroups rarely exist, the method developed in this study is aimed at finding hierarchical subgroups from real directed and acyclic networks,[2] whose structures often deviate with different extent from that of ideal networks.

In, 2008, the authors of this paper proposed a method and a decomposability metric for partitioning a network into an optimal number of structurally equivalent classes (Hsieh and Magee, 2008). We used the k-means method (Lloyd, 1982; MacQueen, 1967) to determine the best decomposition of the network for various numbers of subgroups. The best number of subgroups into which to decompose a network is determined by minimizing the intra-cluster variance of similarity subject to the constraint that the improvement in going to more subgroups is better than a random network would achieve. In this paper, we modified the method and the decomposability metric to extend its applicability and applied it to partition a network into hierarchical subgroups.

Our method for finding hierarchical subgroups of a network is complementary to Generalized Blockmodeling. By using the approach of Generalized Blockmodeling, the type of blocks for the subgroups and between subgroups has to be defined, and, most important of all, the number of subgroups has to be specified in advance. This requires the use of context information in forming hypotheses and gives a criterion function (i.e. inconsistencies) that measures the fit of a specified blockmodel to the actual data. In this regard, our method has the capability of automatically determining the best number of subgroups. For the purpose of hypotheses formation, the decomposability metric generated by our method can be used to evaluate the appropriateness of decomposing a network into hierarchical subgroups or other choices.

This paper is organized into the following sections. In Section 2, we present the new method for finding hierarchical subgroups and its application to an ideal network that contains only perfect hierarchical subgroups. In Section 3, we discuss the use of decomposability for assessing the deviation of a network from the ideal networks that contain only perfect hierarchical subgroups. Application of our method including the decomposability metric to four known social networks is presented in Section 4. Brief concluding remarks are provided in Section 5.

2. A new method for finding hierarchical subgroups

The method proposed here for analyzing hierarchical subgroups is based on the broader research tradition of positional and role analysis (White et al., 1976). In other words, we view the hierarchical subgroups as one specific type of *position* within the networks, and the *roles* of these hierarchical subgroups are manifested by their asymmetric relationships with each other.

As an example of potential ambiguity in determining hierarchical subgroups, we consider the hierarchical organization of an artificial national retailer. At the top level of such a hierarchy, there can be five regional managers and one or more office directors reporting to the chief operations officer. All direct reports to the top level can be considered as one hierarchical group. However, since each of the five regional managers oversees a 500-people organization in his/her region and the office director supervises a team of a few special assistants, it can be more logical to group the regional managers into the same hierarchical subgroup because these managers are functionally *interchangeable* with each other. Therefore, interchangeability can be argued as the key attribute in determining whether a group of people should belong to the same hierarchical subgroup.

Using interchangeability as the criterion for identifying hierarchical subgroups leads us to define a *perfect* hierarchical subgroup as the one whose members are all automorphically equivalent to one another. By contrast, the use of *rank* is less helpful for understanding the hierarchical structure of a network because it only incorporates the information of how many layers a node is from the source or sink nodes which depends on the scheme of placing nodes. With this definition of a perfect hierarchical subgroup, we propose that a hierarchical subgroup should be the one whose nodes are either automorphically equivalent or nearly automorphically equivalent to each other.

Although there is no known fast algorithm that guarantees identification of automorphically equivalent nodes in all graphs (Everett et al., 1990), it has been pointed out that automorphically equivalent nodes are identical with respect to all graph theoretic properties (Borgatti and Everett, 1992). With this important insight, we propose to use a node's profile of similarities with all other nodes as a proxy for its graph theoretic properties, where the nodal similarities are measured by the REGGE algorithm proposed by White (1985) and formulated in the paper by Žiberna (2008). The rationale of our proposal is based on the fact that REGGE algorithm uses an iterative procedure in which estimates of the similarities between pairs of nodes are adjusted in light of the similarities of the nodes adjacent to and from members of the pair. Furthermore, if a pair of nodes has the same REGGE similarities with every other node, their in-degree and out-degree must be the same, and there must be a one-to-one match for every other node with the same number of hops from each of the pair that has the same in-degree and out-degree, unless the node is a common node sitting at the branches of the pair. Since in this case the pair of nodes cannot be differentiated structurally from each other (i.e. an automorphism), by definition, they are automorphically equivalent.

It should be noted that, although the REGGE algorithm was originally designed to measure the extent to which a pair of nodes is regular equivalent (White and Reitz, 1985), when incorporated into our method, it is every node's distribution of REGGE similarities with every other nodes that is used for finding hierarchical subgroups, not the pair-wise single measure of REGGE similarities. However, we recognize that a formal proof of whether two nodes having the same profile of REGGE similarities leads to their being automorphically equivalent is an outstanding problem that merits further attention.

Our method for finding hierarchical subgroups starts with the calculation of pair-wise similarities among nodes using the REGGE algorithm. The normalized similarity, E_{ij}^{t+1}, for two nodes, i and j, at iteration $t+1$ is given by the following equation (Žiberna, 2008):

$$E_{ij}^{t+1} = \frac{\sum_{k=1}^{n}\left(\max_{m=1}^{n}(E_{km}^{t}\ {}_{ij}M_{km}) + \max_{m=1}^{n}(E_{km}^{t}\ {}_{ji}M_{km})\right)}{\sum_{k=1}^{n}(r_{ik}+r_{ki}+r_{jk}+r_{kj})}, \tag{1}$$

where n is the number of nodes, ${}_{ij}M_{km}$ equals $\min(r_{ik}, r_{jm})$ plus $\min(r_{ki}, r_{mj})$, with r_{ij} being the value of the tie from node i to node j. In this study, we use the size of network as the number of iterations so that sufficient differentiation of the similarities among nodes can be obtained and it would not be too large to drive the similarities between nodes that are regularly equivalent into zeroes. We note that the number of iterations used for REGGE in our method is an important issue and merits further attention in the future.

The REGGE algorithm generates meaningful similarity measures among nodes when the network follows the basic acyclic requirement of a hierarchy. However, there are some restrictions on applying the algorithm to other networks. Generally, the algorithm does not work for non-directional networks, networks that have self-ties, and networks in which each node is involved in at least one reciprocated tie (Wasserman and Faust, 1994). Because find-

[2] Because finding hierarchical subgroups from a cyclic network is not meaningful, we restrict our method to deal with only directed and acyclic networks.

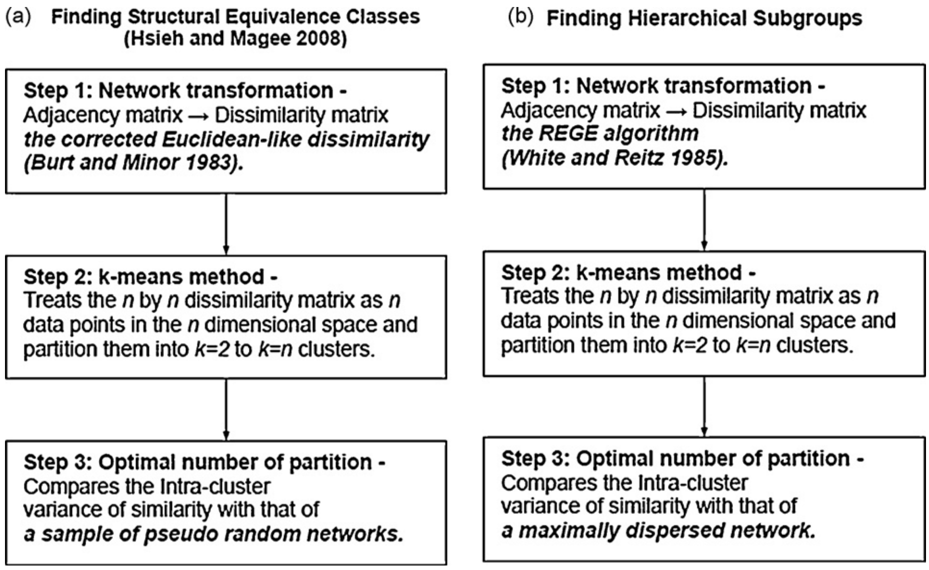

Fig. 1. Flow charts of the algorithms for finding (a) structural equivalence classes (Hsieh and Magee, 2008), and (b) hierarchical subgroups introduced in this study.

ing hierarchical subgroups from a non-directional network is not meaningful, for a cyclical directed network, we suggest that such networks be reduced to the level of *strong components*[3] before measuring their nodal similarities with the REGGE algorithm. By doing so, the reduced network becomes acyclic and the members of the same condensed strong component are pre-classified into the same subgroup.[4]

In a previous study (Hsieh and Magee, 2008), we have presented an algorithm and a metric for network decomposition from similarity matrices. Fig. 1 shows the flow chart of the algorithm used in this study to decompose a network into hierarchical subgroups compared with the one in our previous work that decomposes a network into structural equivalence classes. While both algorithms start with transforming the adjacency matrix of a network into normalized dissimilarity matrix (i.e. Step 1 of both Fig. 1(a) and (b)), the one that finds structural equivalence classes makes the transformation by the corrected Euclidean-like dissimilarity (Burt and Minor, 1983), and the one introduced in this study makes it by the REGGE algorithm described previously. Notice that the dissimilarity matrix of a network is obtained by simply subtracting the similarity matrix (measured by the REGGE algorithm) from a unity matrix.

As shown in Fig. 1(a) as Step 2, for an n-node network, with its known n by n dissimilarity matrix, our previously developed algorithm treats the matrix as n data points in the n-dimensional space and repeatedly applies the method of k-means clustering (MacQueen, 1967) to partition the n data points into $k = 2$ to $k = n$ clusters. In this study, our method for finding hierarchical subgroups applies the same procedure (i.e. Step 2 in Fig. 1(b)), and Lloyd's k-means algorithm (Lloyd, 1982) was used. Lloyd's algorithm begins with a set of k reference points which are randomly selected from the data set. All of the data points are partitioned into k clusters by assigning each point to the cluster of its closest reference point. In each following iteration, the centroid for each cluster is calculated. A partition is then made using the newly calculated centroids as reference points for all of the data points. Because the algorithm applied is a heuristic algorithm, different initial reference points can generate different partitions. We use multiple sets of initial points to evaluate whether the obtained partition has approached its global minimum sum of intra-cluster distances.

It should be noted that, compared with, for example, Jancey's algorithm (Jancey, 1966) that begins with an initial partition and then relocates objects among clusters until no further reduction in the within-cluster points-to-centroid distances is possible, the solution generated by Lloyd's algorithm is not locally-optimal with respect to all possible relocations of any point from its current cluster to any other cluster. Moreover, although Lloyd's algorithm applied by our method does not guarantee a globally optimal solution, for networks with smaller size, there are exact algorithms that guarantee a global optimum. These exact methods have been proposed by Koontz et al. (1975), du Merle et al. (2000) and Brusco (2006).

The readers should not be confused by the fact that we are using the non-hierarchical procedure of k-means clustering method to identify hierarchical subgroups. Since the hierarchical relationships among the subgroups are identified in Step 1, applying the k-means method in Step 2 contributes only to partitioning the nodes with the similar hierarchical relationships into the same subgroups.

Finally, as shown in Fig. 1(a) as Step 3, our previously developed algorithm stops further dividing a network into additional subgroups if the decrease of its n nodes' sum of within-cluster points-to-centroid distances, D_k, from k to $k + 1$ is less than that of a sample of random networks, D_k^{random}. Because a sample of random networks was found to be biased for acyclic directed networks, we pursued a new and we think in the end more powerful approach in this work. Since the rationale of using random networks for comparison is their lack of structure, we propose to compare the actual network's sum of within-cluster points-to-centroid distances with that of a network whose n nodes are *maximally* dispersed from each other in the normalized n-dimensional space.[5] We require

[3] In graph theory, a directed graph is strongly connected if it contains a directed path from u to v and a directed path from v to u for every pair of vertices u, v. The maximal strongly connected sub-graphs in a graph are referred to as the strong components.

[4] We realize the use of strong components as a processing step of our method might compromise the integrity of a network's structure for identifying automorphically equivalent subgroups. However, the method is primarily designed to find hierarchical subgroups, which is only meaningful when the network is directed and acyclic. Therefore, we propose to reduce the networks into strong components whenever they are cyclic and to call attention to possible compromises in these cases.

[5] A normalized n-dimensional space has the coordinates of all of its dimensions ranging from zero to one.

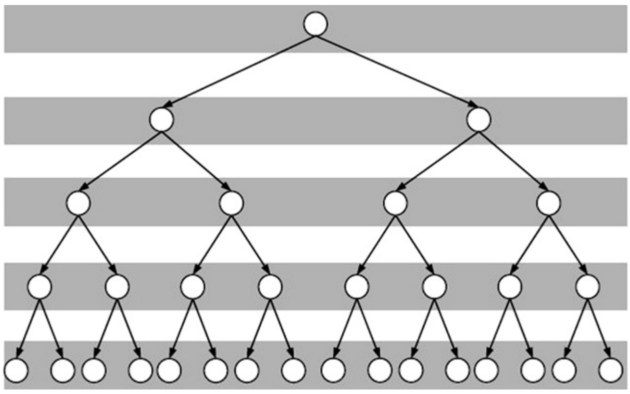

Fig. 2. The 31-node five-layer binary tree.

the n-dimensional space to be a normalized one because the actual network's n-dimensional coordinates of its n nodes come from the normalized dissimilarity matrix. When n nodes are maximally dispersed in the normalized n-dimensional space, they all have the same maximum distances from each other and thus are in-differentiable from one another spatially. In essence, such a network is *structureless*, and becomes the benchmark for the actual network that we intend to partition.[6]

Based on the observable results for two and three dimensional space, we assume that for n nodes to be maximally dispersed from each other in the normalized n-dimensional space, every node will have the same distance, $\sqrt{2}$, from each other. One trivial spatial arrangement of the n nodes in the normalized n-dimensional space that achieves maximum dispersion is to place each of the n nodes at n different vertexes adjacent to the origin. For example, for three nodes in the normalized 3-dimensional space, placing the three nodes at the coordinates of (1, 0, 0), (0, 1, 0), and (0, 0, 1) respectively could achieve maximum dispersion. For the aforementioned trivial spatial arrangement of the n nodes divided into k clusters, the sum of intra-cluster points-to-centroid distance, $D_{\text{max-dispersion}(n,k)}$, is approximated as

$$D_{\text{max-dispersion}(n,k)} = \{1 - 1/(n/k)\}^2 \cdot n. \tag{2}$$

With $D_{\text{max-dispersion}(n,k)}$ available in closed form, no sample of networks is required as is the case of Step 3 in Fig. 1(a), and we stop further dividing a network into additional subgroups if the decrease of its n nodes' sum of within-cluster points-to-centroid distances, D_k, from k to $k+1$ is less than that of the *structureless* network, $D_{\text{max-dispersion}(n,k)}$ (i.e. Step 3 in Fig. 1(b)).

To find hierarchical subgroups, with the dissimilarity matrix generated from using the REGGE algorithm, we can partition a network into an optimal number of hierarchical subgroups by proceeding with the procedures just described. We use a simple example to demonstrate the method of identifying hierarchical subgroups in a directed network. The artificial network shown in Fig. 2 is a 31-node five-layer binary tree. Its five hierarchical subgroups (i.e. the layers) can easily be identified by observation.

Using the new method proposed here to find the hierarchical subgroups of this network, we first apply the REGGE algorithm to the network and obtain the similarity matrix of the nodes. The dissimilarity matrix of the network is obtained by subtracting the similarity matrix from a unity matrix. By reading each of the 31 rows of the dissimilarity matrix as a specific data point that has a 31-dimensional coordinates, we repeatedly apply Lloyd's k-means algorithm to partition the 31 data points into $k=2$ to $k=31$ clusters. The sum of intra-cluster points-to-centroid distances, D_k, as a function of k is obtained and shown in Fig. 3(a) as dark gray bars. By using Eq. (2), the value of $D_{\text{max-dispersion}(n,k)}$ for n equals 31 and k ranges from 2 to 31 is shown in Fig. 3(a) as light gray bars. Fig. 3(b) shows the fitness index generated by subtracting D_k of the tree network from that of the structureless network.

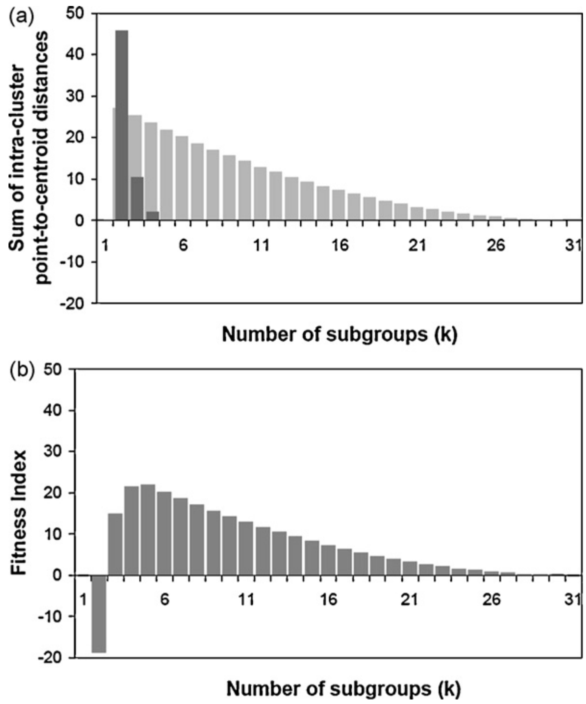

Fig. 3. (a) The sum of intra-cluster point-to-centroid distances of the 31-node five-layer binary tree network (dark gray bars) and that of the *structureless* network whose n nodes are maximally dispersed from each other in the n-dimensional space (light gray bars) and (b) the fitness index of the tree network as a function of the number of subgroups.

The peak of fitness index is at five subgroups because further subdivision only reduces D_k of the tree network at a rate equal to or less than a structureless network. The nodes belonging to the five subgroups form the most appropriate hierarchical subgroups of the tree network, and they correspond to the five layers of the network as shown in Fig. 2 marked with gray color blocks.[7]

3. Hierarchical decomposability

As introduced in our previous study (Hsieh and Magee, 2008), to differentiate among networks with various levels of deviation from the networks having ideal structures, the decomposability of a network divided into k subgroups is defined as

$$Q = 1 - D_k / D_{\max(n,k)}, \tag{3}$$

where D_k is the sum of intra-cluster points-to-centroid distance when the network is divided into k subgroups by the k-means method, and $D_{\max(n,k)}$ is the maximum sum of intra-cluster points-to-centroid distance for networks having n nodes and k clusters.

In the previous study, though $D_{\max(n,k)}$ was mathematically defined, the significance of the network manifested by it was not clear. As a consequence, the *decomposability* thus defined needed the relationship between a network's decomposability and its percentage of linkage perturbation from ideal networks to calibrate

[6] We note that such a network represents only the theoretical boundary for the possible configuration of the REGGE dissimilarities but may not have real existence.

[7] All other simple perfect hierarchical structures can also be derived automatically by the method described here.

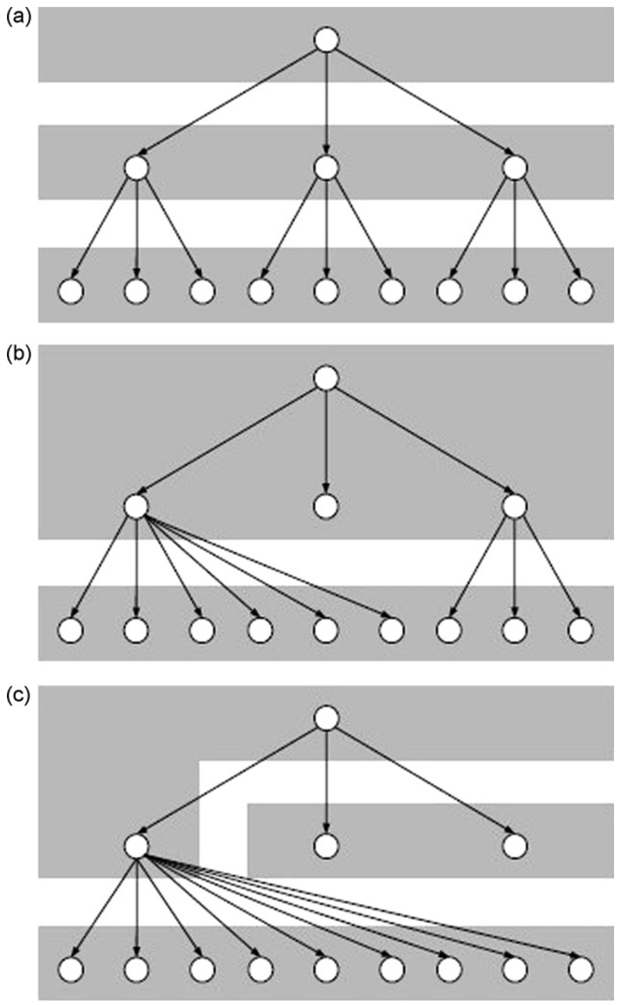

Fig. 4. Three artificial tree networks decomposed into hierarchical subgroups. The hierarchical subgroups suggested by our method are marked with gray color blocks.

its meaning. In this study, we propose to remedy this drawback by redefining the decomposability as

$$Q = 1 - D_k/D_{\text{max-dispersion}(n,k)}, \quad (4)$$

where D_k is the same as defined in Eq. (3), and $D_{\text{max-dispersion}(n,k)}$ is defined previously and can be calculated by Eq. (2). The sum of intra-cluster points-to-centroid distance of zero is achieved only by ideal networks, and it generates decomposability of one. Decomposability of zero means that the network has maximum dispersion, $D_{\text{max-dispersion}(n,k)}$, and such networks have *nodes* that cannot be differentiated from one another spatially. The decomposability thus defined is bounded between the ideal networks and networks that are structureless.

We use the following example to demonstrate the change of decomposability as a network deviates from the one having only perfect hierarchical subgroups.[8] Fig. 4 shows three artificial networks that have the same number of 13 nodes. While the network of Fig. 4(a) is a tree with a branching ratio of three, the network of Fig. 4(b) has the same tree structure but moves the three leaves in the middle branch to the left branch, and the network of Fig. 4(c) follows the same tree structure but has all of the leaves connected to the left branch.

By applying our method of finding hierarchical subgroups to these networks, the results suggests that each of the three networks should be divided into either two or three subgroups as marked with gray blocks in the figure. The decomposability of the network of Fig. 4(a), (b), and (c) are 1.00, 0.72, and 0.94, delineating that the latter two networks have different levels of deviation from the former network that has only perfect hierarchical subgroups. For the network of Fig. 4(b), it can be confirmed with visual inspection that both identified subgroups are imperfect hierarchical subgroups. Compared with the network of Fig. 4(b), the network of Fig. 4(c) has only one imperfect hierarchical subgroup that contains the root node and the leftmost node in the second layer. With the network of Fig. 4(a) having the highest decomposability and that of Fig. 4(b) the lowest, the decomposability confirms what visual inspection tells us; the network of Fig. 4(c) is closer to the ideal network than that of Fig. 4(b).

To demonstrate the usefulness of decomposability as an objective measure to compare the quality of different decompositions, we use the same three networks as shown in Fig. 4 and decompose them into structural equivalence classes (Lorrain and White, 1971). Two nodes are said to be structurally equivalent if they link to and are linked by exactly the same set of other nodes in the network. Compared with automorphic equivalence, structural equivalence is more stringently defined. Nodes that are structurally equivalent are also automorphically equivalent, but the reverse is not necessarily true. To decompose the networks into structural equivalence classes, we apply the same method introduced in this study, with the only difference that the similarities among nodes are now measured by the corrected Euclidean-like dissimilarity (Burt and Minor, 1983) and normalized between zero and one. The decomposition results of our method are shown in Fig. 5, in which different structural equivalence classes are marked with gray blocks.

As shown in the figure, the 13 node network of Fig. 5(a) is decomposed into four structural equivalence classes. With visual inspection, the three identified classes that together form the third layer of the tree network are structural equivalent classes; our method partitions the rest of the four nodes into the same class, where each of the four nodes is a structural equivalence class of its own. Our method indicates that the network of Fig. 5(b) should be divided into four structural equivalence classes. With visual inspection, three of the four classes are structural equivalence classes, and each of the three nodes in the class that sits at the upper right corner is a structural equivalence class of its own. Similarly, the network of Fig. 5(c) is divided into three structural equivalence classes suggested by our method. Visual inspection again tells us that two of the identified classes are structural equivalence classes, and each of the three nodes in the upper-right class is a structural equivalence class of its own.

Calculated with our method, the decomposability for the networks of Fig. 5(a), (b) and (c) are 0.51, 0.80 and 0.91. By comparing the decomposability generated for the structural equivalence classes with that generated for the hierarchical subgroups, we can evaluate the quality of the two decompositions for the same network. For the same network of Figs. 4(a) and 5(a), the higher decomposability of 1.00 for partitioning the network into three hierarchical subgroups, compared with that of 0.51 for two structural equivalence classes, indicates that partitioning the network into the former structure is much more effective. For the same network of Figs. 4(b) and 5(b), the decomposability of 0.80 for structural equivalence classes is higher than that of 0.72 for hierarchical subgroups, indicating that the network is more similar to the ideal network having only structural equivalence classes. Finally, for the same network of Figs. 4(c) and 5(c), the decomposability of 0.94 for hierarchical subgroups is higher than that of 0.91 for structural equivalence classes, indicating that decomposing the network into

[8] These deviations give rise to the hierarchical ambiguity of human organizations discussed in Section 2.

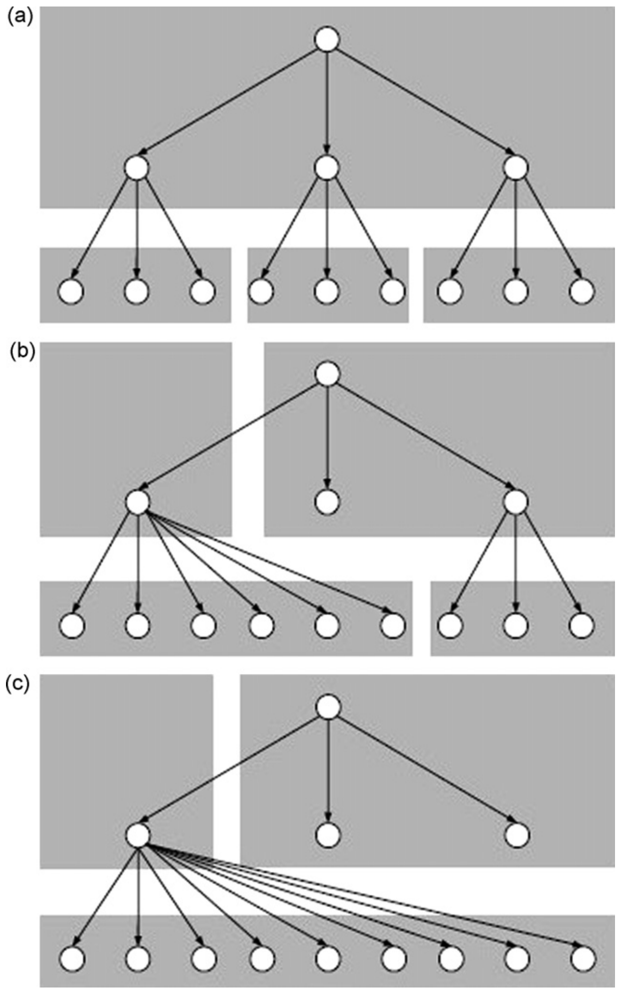

Fig. 5. Three artificial tree networks decomposed into structural equivalence classes. The structural equivalence classes suggested by our method are marked with gray color blocks.

hierarchical subgroups is a slightly better choice for understanding the structure of the network.

As we have shown, for the same network of Figs. 4(a) and 5(a) and Figs. 4(c) and 5(c), partitioning the network into hierarchical subgroups has higher decomposability than into structural equivalence classes; partitioning the network of Figs. 4(b) and 5(b) into structural equivalence classes has higher decomposability than into hierarchical subgroups. These examples clearly show the merit of using the decomposability as an objective criterion for choosing the type of decomposition for understanding the structure of the networks. In the following section, we use four real examples to further demonstrate our method and the decomposability metric.

4. Application of the method

In the previous section, we proposed a method for decomposing a network into hierarchical subgroups and a decomposability metric for measuring a network's deviation from an ideal network that contains only perfect hierarchical subgroups. In this section, we test our method and decomposability metric with four examples of real networks.

Our first example is the authoritative relationship among the personnel in an office reported by Thurman (1979). Though the original data is comprised of the social relationship and the authoritative relationship among the 15 workers, we analyze only the authoritative relationship among the workers as shown in Fig. 6.

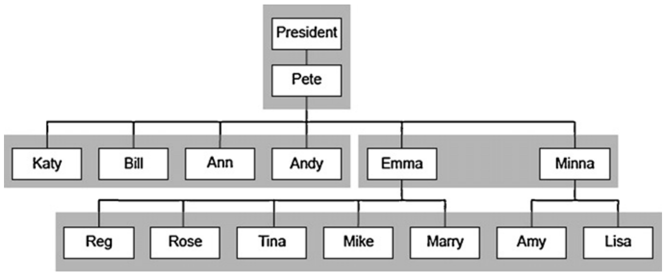

Fig. 6. The authoritative relationship of Thurman (1979) office data.

By applying our method to the authoritative network, the results suggest that the network be divided into four subgroups denoted in Fig. 6 with gray blocks. These hierarchical subgroups are: (1) the President and Pete, (2) Katy, Bill, Ann, and Andy, (3) Minna, and Emma, and (4) Peg, Rose, Tina, Mike, Marry, Amy, and Lisa. Though this partition does not match the nominal four hierarchical layers of the organization, it does identify the four hierarchical subgroups as the optimal number of subgroups that deviates least from the requirement of being automorphic equivalence classes.

It is worth mentioning that, the *k*-means clustering step of our method does identify the seven perfect hierarchical subgroups of the network (i.e. {President}, (Saunders), {Katy, Bill, Ann, Andy}, {Emma}, {Minna}, {Reg, Rose, Tina Mike}, and {Amy, Lisa}). However, the stopping criterion of our method suggests that further dividing the network into more than four subgroups contributes to less decrease of its 15 node's sum of within-cluster point-to-centroid distance than that achieved with the structureless network. With the stopping rule, the only perfect hierarchical subgroup identified by our method is the one that contains Katy, Bill, Ann, and Andy.

With the network size equals to 15 and the number of subgroups equals to four, we calculate $D_{\text{max-dispersion}(15,4)}$ from Eq. (2) as 8.07. According to Eq. (4), the decomposability for the authoritative network is calculated as 0.88. This high value of decomposability indicates that the authoritative network is fairly similar to an ideal network in terms of its hierarchical subgroups and thus can reliably be discussed in terms of this structure.

Our next example is the network of "liking ties" for children in a classroom studied by Jennings (1948). To present the basic structure of the liking ties for the children in the classroom, Doreian et al. (2005) used Generalized Blockmodeling to obtain a three-subgroup ranked-cluster model as a decomposed representation of the network. The image matrix of the three-subgroup ranked-cluster model has symmetric blocks in the diagonal, null blocks in the upper-triangle, and null or one blocks in the lower-triangle. Fig. 7 shows the adjacency matrix of the network and the decomposition results of their analysis marked with the grid lines.

Doreian et al. (2005) used *inconsistency* to evaluate the quality of the decomposition, and the partition shown in Fig. 7 has 54 inconsistencies, including 16 in the diagonal blocks that violate the symmetric requirement and 38 in the lower-triangle blocks that deviate from the null requirement. The total inconsistencies can be reduced if the pre-specified blockmodel is a hierarchical structure where the requirement of the lower-triangle blocks was relaxed to allow for column or row regular blocks. In this case, the partition shown in Fig. 7 has 17 inconsistencies, including still 16 in the diagonal blocks but now only one in the lower-triangle blocks.

We analyze the same network with the method introduced in this study and decompose it into hierarchical subgroups. Since the network is directed but not acyclic, we first find the strong components of the network and reduce the network into the level of

	23	28	29	33	3	6	8	12	13	14	15	16	17	18	19	20	21	25	26	27	30	34	1	2	4	5	7	9	10	11	22	24	31	32	35
23-Jacque	0	1	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
28-Jack	1	0	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
29-Wolf	1	1	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
33-Nero	1	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
3-Lily	0	0	0	0	0	0	0	0	0	0	1	1	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
6-Carmen	0	0	0	0	0	0	1	1	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
8-Ethel	0	0	0	0	0	1	0	1	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
12-Lisa	0	0	0	0	0	0	0	0	1	0	0	1	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
13-Nina	0	0	0	0	0	0	0	1	0	1	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
14-Thela	0	0	0	0	0	0	0	1	1	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
15-Clara	0	0	0	0	1	0	0	0	0	1	0	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
16-Janice	0	0	0	0	1	0	0	0	0	0	1	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
17-Pat	0	0	0	0	0	0	0	0	0	0	1	1	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
18-Alice	0	0	0	0	0	0	0	1	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
19-Angel	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
20-Teresa	0	0	0	0	0	1	0	0	0	0	0	0	0	0	1	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
21-Rea	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
25-Guy	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
26-Davan	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
27-Dalti	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
30-Joe	0	1	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
34-Mick	0	1	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
1-Edna	0	0	0	0	0	0	0	0	0	0	1	1	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0
2-Paula	0	0	0	0	0	0	0	0	0	0	1	1	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0
4-Claire	0	0	0	0	0	1	0	0	0	0	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
5-Ella	0	0	0	0	0	1	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0
7-Edith	0	0	0	0	1	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0
9-Lorine	0	0	0	0	0	0	0	0	1	1	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
10-July	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	1	0	0	0	0	0
11-Lucia	0	0	0	0	0	0	0	1	1	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
22-Sylvia	1	1	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
24-Lionel	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0
31-Frank	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	1	0
32-Marc	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	1	0	0
35-Peter	0	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0

Fig. 7. The network of the liking ties for children of a classroom and the partition generated by Doreian et al. (2005).

	#23	#3	#25	30	34	7	9	11	#1	4	5	10	22	24	#31	35
#23-Jacque	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
#3-Lily	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
#25-Guy	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
30-Joe	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
34-Mick	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
7-Edith	0	1	0	0	0	0	1	0	0	0	0	0	0	0	0	0
9-Lorine	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0
11-Lucia	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0
#1-Edna	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0
4-Claire	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0
5-Ella	0	1	0	0	0	1	0	0	0	0	0	0	0	0	0	0
10-July	0	1	0	0	0	0	1	1	0	0	0	0	0	0	0	0
22-Sylvia	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0
24-Lionel	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0
#31-Frank	1	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0
35-Peter	1	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0

#23-Jacque: { 23-Jacque, 28-Jack, 29-Wolf, 33-Nero }
#3-Lily: { 3-Lily, 6-Carmen, 8-Ethel, 12-Lisa, 13-Nina, 14-Thela, 15-Clara, 16-Janice, 17-Pat, 18-Alice, 19-Angel, 20-Teresa, 21-Rea }
#25-Guy: { 25-Guy, 26-Davan, 27-Dalti }
#1-Edna: { 1-Edna, 2-Paula }
#31-Frank: { 31-Frank, 32-Marc }

Fig. 8. The dichotomized adjacency matrix of the liking ties for children in a classroom reduced into the level of its strong components. The partition result of this network generated by our method is shown with the grid lines.

strong components.[9] To reduce a network into the level of strong components, members of the same strong component are represented by one node in the network, and their individual arcs to the same other node in the network are combined into one valued arc (with counts of the arcs combined). We dichotomize the reduced network so that only the presence or absence of the ties among children and strong components was presented. Fig. 8 shows the adjacency matrix of the reduced network that is dichotomized and removed of self-loops; the partition of three hierarchical subgroups resulted in applying our method is shown in the figure with grid lines.

As shown in Fig. 8, the "#" signs placed in front of some of the name labels indicate their being sets of children that form

[9] As noted, this procedure compromises the decomposition into hierarchical groups and thus this example is not fully consistent with the procedure not using strong components.

	23	28	29	33	3	6	8	12	13	14	15	16	17	18	19	20	21	25	26	27	30	34	7	9	11	1	2	4	5	10	22	24	31	35	35
23-Jacque	0	1	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
28-Jack	1	0	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
29-Wolf	1	1	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
33-Nero	1	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
3-Lily	0	0	0	0	0	0	0	0	0	0	1	1	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
6-Carmen	0	0	0	0	0	0	1	1	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
8-Ethel	0	0	0	0	0	1	0	1	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
12-Lisa	0	0	0	0	0	0	0	0	1	0	0	1	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
13-Nina	0	0	0	0	0	0	0	1	0	1	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
14-Thela	0	0	0	0	0	0	0	1	1	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
15-Clara	0	0	0	0	1	0	0	0	0	1	0	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
16-Janice	0	0	0	0	1	0	0	0	0	0	1	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
17-Pat	0	0	0	0	0	0	0	0	0	0	1	1	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
18-Alice	0	0	0	0	0	0	0	1	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
19-Angel	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
20-Teresa	0	0	0	0	0	1	0	0	0	0	0	0	0	0	1	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
21-Rea	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
25-Guy	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
26-Davan	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
27-Dalti	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
30-Joe	0	1	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
34-Mick	0	1	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
7-Edith	0	0	0	0	1	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0
9-Lorine	0	0	0	0	0	0	0	0	1	1	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
11-Lucia	0	0	0	0	0	0	0	1	1	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
1-Edna	0	0	0	0	0	0	0	0	0	0	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0
2-Paula	0	0	0	0	0	0	0	0	0	0	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0
4-Claire	0	0	0	0	0	1	0	0	0	0	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
5-Bla	0	0	0	0	0	1	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0
10-July	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	0	0	0	0	0	0	0	0	0	0
22-Sylvia	1	1	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
24-Lionel	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0
31-Frank	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	1	0
32-Marc	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	1	0	0
35-Peter	0	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0

Fig. 9. The network of the liking ties for children of a classroom and the partition generated by our method.

strong components in the network. For example, "#23-Jacque" represents the group of Jacque, Jack, Wolf, and Nero that form a strong component. It should be noted that, by reducing the network into the level of strong component, we have already assigned members of the same strong components to the same hierarchical subgroups. This "pre-decomposition" would later be expanded into the optimal partition of hierarchical subgroups generated by our method.

By populating the hierarchical subgroups identified from the reduced network with the nodes of their strong components, Fig. 9 shows the partition of the children in the classroom generated by our method. This result and that of Doreian et al. (2005) both have three subgroups, but in other ways are very different. However, if we fit our partition result with the three-subgroup ranked-clusters model used by Doreian et al. (2005) mentioned previously, our partition has the same 54 inconsistencies, among which 13 are in the diagonal blocks that violate the symmetric requirement (where Doreian et al. had 16 inconsistencies) and 41 are in the lower-triangle blocks that deviate from the null requirement (where Doreian et al. had 38).

If we again relax the null or one requirement for the lower-triangle blocks of the ranked-cluster model and allow for column or row regular blocks, the total inconsistencies of our partition can be reduced to 16, including 13 in the diagonal blocks (where Doreian et al. had 16) and three in the lower-triangle blocks (where Doreian et al. had one). The 16 inconsistencies generated by fitting our partition result with the relaxed ranked-clusters model compared with the 17 inconsistencies generated by Doreian et al. (2005) does not indicate superiority of our method. It should be emphasized that our partition can easily be obtained by the approach of Generalized Blockmodeling with other search or optimization algorithms since the underlying blockmodel is the same. However, it does indicate our method can generate an effective result that is comparable with those found through Generalized Blockmodeling while automatically identifying the best number of subgroups.

The Ragusan family marriage network presented by Krivošić (1990) is our third example. Two matrices that describe the marriage networks of the Ragusan noble families in the 16th century and in the 18th century and the beginning of the 19th century were constructed by Krivošić. Batagelj (1996) analyzed the marriage networks of both periods and used conventional blockmodeling to obtain a two-cluster partition of the families that follows the core-periphery model. Doreian et al. (2000) used the marriage network of the second period to demonstrate the method of symmetric-acyclic decomposition and Generalized Blockmodeling. In this study, we use the marriage networks of both periods to demonstrate the usefulness of our method for finding hierarchical subgroups in the networks.

Fig. 10 shows adjacency matrices of the Ragusan family marriage networks in the 16th century and in the 18th and early 19th century, where the rows represent the families of the groom and the columns represent the houses of the bride. While the original networks were valued (with counts of marriages between families), we dichotomized the networks so that only the presence or absence of marriage ties between families were presented.

Since both networks are directed but not acyclic, to apply our method, we first have to find the strong components of the networks and reduce them into a single pseudo-node. We then dichotomize the reduced networks and apply our method to find hierarchical subgroups. The results indicate that both networks should be divided into three hierarchical subgroups.

Fig. 11 shows the Ragusan family marriage networks reduced into the level of strong components, where Fig. 11(a) is the network of the 16th century, and Fig. 11(b) is that of the 18th and early 19th century. The three hierarchical subgroups for each network identified by our method are marked by gray blocks. As shown in the figure, the "#" sign placed in front of the family label indicates its being the set of families that form a strong component in the network. The members of the strong component sets are shown in the figure. As mentioned previously, by reducing the network into the level of strong components, we have already assigned

(a)

		14	20	1	3	7	10	12	19	22	23	2	8	9	6	13	11	16	21	4	5	15	17	18
Natali	14	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Slatarich	20	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Basilio	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0
Bona	2	0	0	0	0	1	0	0	0	0	0	0	0	0	0	1	1	0	1	0	0	0	0	0
Bonda	3	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0
Caboga	6	1	0	0	0	1	0	0	1	0	0	0	0	0	0	1	0	0	1	0	0	0	0	0
Cerva	7	0	0	0	0	0	0	0	0	0	0	1	0	0	1	1	0	0	1	0	0	0	0	0
Georgi	8	0	0	0	1	1	0	0	0	0	1	1	0	0	0	0	0	1	0	0	0	0	0	0
Ghetaldi	9	1	0	0	0	0	1	0	1	0	0	1	1	0	0	0	0	0	0	0	0	0	0	0
Gondola	10	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0
Goze	11	0	0	1	1	0	0	1	0	1	0	1	0	0	0	0	0	0	1	0	0	0	0	0
Gradi	12	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	1	0	0	0	0	0
Menze	13	0	0	0	0	0	0	0	0	0	1	0	0	1	1	0	0	0	0	0	0	0	0	0
Poza	16	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	1	0	1	0	0	0	0	0
Saraca	19	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0
Sorgo	21	0	1	0	1	0	0	0	0	0	1	1	0	0	0	0	1	0	0	0	0	0	0	0
Tudisi	22	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0
Zamagna	23	0	0	0	0	0	0	0	0	0	0	1	0	1	0	0	0	0	1	0	0	0	0	0
Bosdari	4	0	0	0	0	0	0	0	0	0	1	1	0	0	0	0	0	0	0	0	0	0	0	0
Bucchia	5	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0
Pauli	15	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0
Ragnina	17	0	0	0	1	0	0	0	0	0	0	1	0	0	0	0	1	0	1	0	0	0	0	0
Resti	18	0	0	0	0	0	0	1	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0

(b)

		2	4	6	15	21	3	5	7	8	9	10	11	12	13	14	16	19	20	22	24	1	17	18	23
Basilio	2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Bocignolo	4	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Bucchia	6	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Palmota	15	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Saraca	21	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Benessa	3	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	1	0	0	0	1	0	0	0	0
Bona	5	0	0	0	1	0	1	1	0	0	1	0	1	1	1	1	0	0	0	0	1	0	0	0	0
Caboga	7	0	0	1	0	0	0	0	1	0	0	0	1	0	0	1	0	0	0	0	0	0	0	0	0
Crieva	8	0	0	0	0	0	0	1	0	1	0	1	1	1	0	0	0	0	1	0	1	0	0	0	0
Georgio	9	0	0	0	0	1	0	1	1	0	1	0	0	0	0	0	0	0	0	1	0	0	0	0	0
Gondola	10	0	0	0	0	0	0	0	0	0	0	0	1	1	0	0	0	0	0	1	0	0	0	0	0
Goze	11	0	1	0	0	0	1	1	1	1	0	1	1	0	0	0	0	1	1	0	0	0	0	0	0
Gradi	12	0	0	0	0	0	0	0	0	0	0	0	1	1	0	0	0	0	0	1	0	0	0	0	0
Lucari	13	0	0	0	0	0	0	0	0	1	0	0	1	1	0	0	0	0	0	1	0	0	0	0	0
Menze	14	1	0	0	0	0	0	1	1	0	0	0	1	0	0	0	0	0	0	1	0	0	0	0	0
Poza	16	0	0	0	0	0	0	1	0	0	1	0	0	0	0	0	1	0	0	0	0	0	0	0	0
Ragnina	19	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0
Resti	20	0	0	1	0	0	0	1	0	1	1	0	0	0	0	0	0	0	1	0	0	0	0	0	0
Sorgo	22	0	1	0	0	0	1	0	0	0	0	0	1	1	1	1	0	0	1	1	0	0	0	0	0
Zamagna	24	0	0	0	0	0	0	1	0	0	1	1	1	0	0	0	0	0	0	1	0	0	0	0	0
Babalio	1	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0
Proculo	17	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0
Prodanelo	18	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0
Tudisio	23	0	0	0	0	0	0	0	0	0	0	0	0	1	0	1	0	0	0	0	0	0	0	0	0

Fig. 10. The adjacency matrices of the Ragusan family marriage networks in (a) the 16th century and (b) the 18th and early 19th century. The rows represent the families of the groom and the columns represent the houses of the bride.

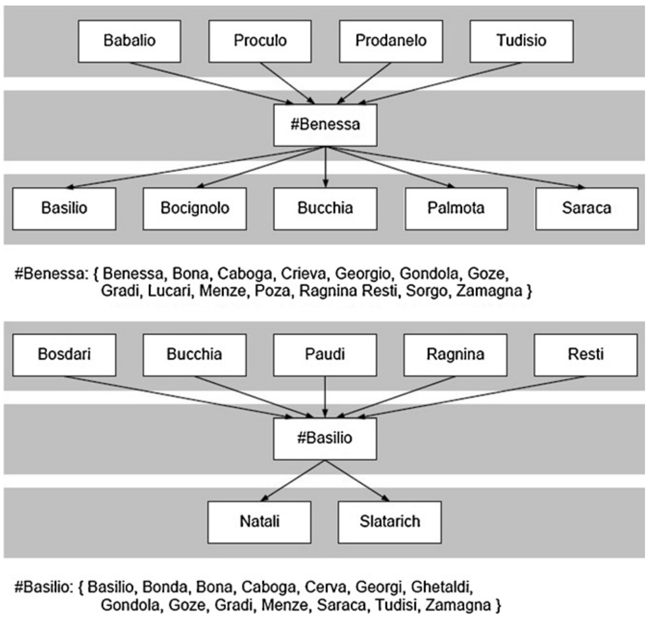

Fig. 11. The Ragusan family marriage networks in (a) 16th century and (b) 18th and early 19th century reduced into the level of its strong components. The hierarchical subgroups identified by our method are marked with the gray blocks.

members of the same strong components to the same hierarchical subgroups.

For both reduced networks, the three identified hierarchical subgroups clearly have their own distinct characteristics. They are the families that provide only grooms, families that provide only brides, and families that provide both grooms and brides and often have marriage bonds with each other. Since it can be verified with visual observation that the identified subgroups in both reduced networks are all automorphic equivalence classes, both reduced networks are ideal networks that contain only perfect hierarchical subgroups. Therefore, the decomposability for each reduced network is 1.0.

By assigning the members of strong components to their corresponding hierarchical subgroups identified from the reduced network, we obtain the partition of families shown in Fig. 10 with grid lines. In fact, our method generates the same hierarchical subgroups as the clusters generated by Doreian et al. (2000) for the 18th and early 19th century network. The result again demonstrates that our method for finding hierarchical subgroups is equally effective as compared with either the symmetric-acyclic decomposition or Generalized Blockmodeling. It is perhaps interesting that a case containing perfect automorphic groups (in the reduced matrix) is the only one leading to the same solution by the two methods.

Our fourth example is the dominance hierarchies in *Leptothorax* ants reported by Cole (1981). In the social organization of Leptothorax allardycei, the workers form linear dominance hierarchies characterized by routine displays of dominance, avoidance behavior, and even fighting. By observing interactions among ants over an extended period of time, Cole constructed two sets of dominance hierarchies, one for the queenright colony with 16 ants and the other for the queenless colony with 13 ants. While the original data were valued (with the number of times an interaction occurred between two ants), they were dichotomized so that only the presence or absence of interactions between ants is analyzed.

Fig. 12 shows the dominance relation between ants for the two colonies, where the rows represent the subordinate ants and the columns represent the dominant ants. For demonstration purpose, we permute the original network of the queenright colony to place the ant labeled by “da” next to the ant labeled by “oe”.

By applying our method to find the hierarchical subgroups of the two networks, maximization of the fitness index indicates that the best decomposition for the queenright colony has four hierarchical subgroups and that for the queenless colony has two. The partitions of the two networks are shown in Fig. 12 by the grid lines.

If we fit our partition results to the blockmodel of hierarchical structure whose image matrix has regular, complete, or null blocks in diagonal, null blocks in the upper-triangle, and column or row regular blocks in the lower-triangle, the four-subgroup partition of the queenright colony has three inconsistencies, with two in the diagonal blocks and one in the lower-triangle blocks that all deviate from the regular requirement. Similarly, the two-subgroup partition of the queenless colony has one inconsistency in the diagonal blocks that deviates from the regular requirement. Since the two colonies are decomposed into a different number of hierarchical subgroups, it is less meaningful to compare the effectiveness of their decompositions using the criterion of blockmodel inconsistency.

To deal with this situation, we compare the decomposability of the two partitions to make an objective evaluation. In this case, the decomposability of the queenright colony's network partitioned into four hierarchical subgroups is 0.88, and that of the queenless colony's network partitioned into two hierarchical subgroups is 0.78. With the queenright colony's decomposability higher than

(a)

		1	9	2	3	4	5	6	7	8	10	11	12	13	14	15	16
oe	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
da	9	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
he	2	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
po	3	1	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0
sa	4	1	0	1	1	0	0	0	0	0	0	0	0	0	0	0	0
lo	5	1	0	1	1	1	0	0	0	0	0	0	0	0	0	0	0
mi	6	1	0	1	1	1	1	0	0	0	0	0	0	0	0	0	0
lp	7	1	0	1	1	1	0	1	0	0	0	0	0	0	0	0	0
pa	8	0	0	0	1	1	0	0	0	0	0	0	0	0	0	0	0
ch	10	1	0	1	1	1	0	1	0	0	0	0	0	0	0	0	0
fa	11	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0
do	12	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0
li	13	0	0	1	0	0	0	1	0	0	0	0	0	0	0	0	0
bl	14	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
le	15	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
ri	16	1	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0

(b)

		1	2	3	4	5	6	7	8	9	10	11	12	13
ca	1	0	0	0	0	0	0	0	0	0	0	0	0	0
me	2	1	0	0	0	0	0	0	0	0	0	0	0	0
st	3	1	1	0	0	0	0	0	0	0	0	0	0	0
po	4	1	1	1	0	0	0	0	0	0	0	0	0	0
cl	5	1	1	1	1	0	0	0	0	0	0	0	0	0
cll	6	1	0	1	1	1	0	0	0	0	0	0	0	0
sq	7	0	0	0	1	1	1	0	0	0	0	0	0	0
ta	8	1	1	0	0	0	1	0	0	0	0	0	0	0
u	9	0	1	1	0	1	1	0	0	0	0	0	0	0
ro	10	0	0	0	0	1	1	0	0	0	0	0	0	0
an	11	0	0	1	0	1	0	0	0	0	0	0	0	0
ni	12	1	0	0	1	1	1	0	0	0	0	0	0	0
sl	13	0	0	0	0	0	1	0	0	0	0	0	0	0

Fig. 12. Dominance hierarchy of *Leptothorax* ants in (a) queenright colony and (b) queenless colony. The rows represent the subordinate ants and the columns represent the dominant ants.

that of the queenless colony, we conclude that the four-subgroup partition of the queenright colony has better quality than the two-subgroup partition of the queenless colony in terms of their conformity to the ideal networks having only perfect hierarchical subgroups. In other words, the queenright colony's network is *more hierarchical* than the queenless colony's network. This result may provide useful information for researchers who study the social organization of Leptothorax allardycei and is a clear demonstration that our method can be use for objectively comparing the quality of decompositions among different networks.

5. Conclusion and discussion

In this study, we developed a new method for finding hierarchical subgroups in directed networks. By defining the hierarchical subgroups as subsets of nodes that are automorphically or nearly-automorphically equivalent, the partition found by our method matches very well with that obtained by the ranked-clusters model proposed by (Doreian et al., 2005). One important advantage of our method is that we do not have to specify the number of hierarchical subgroups in advance. The stopping criterion of our method yields the best number of subgroups for the particular network.

The approach developed in this study can be used in a complementary way to Generalized Blockmodeling. When studying a new set of network data, one could first apply our approach to find the possible decomposition inductively. Refinement of the decomposition can then be achieved by studying the context of the data so that new blockmodels can be developed for hypothesis testing. As shown in this paper, applying both methods seem to be appropriate in all cases because the results in Section 4 indicate that sometimes they can deliver slightly different and yet interesting decompositions. Moreover, the examples also show the usefulness of using the decomposability metric to objectively assess the quality of the decompositions.

We would like to point out some computational issues of the method presented in this paper for finding hierarchical subgroups. Because the method applies the REGGE algorithm to transform the adjacency matrix of the network into the dissimilarity matrix among nodes, the computational complexity of $O(n^5)$ of the REGGE algorithm is a serious bottleneck that prevents the method from being applicable for large-scale networks. One possible alternative is to use the *CATREGE algorithm* (Borgatti and Everett, 1993), which has a computational complexity of $O(n^3)$ but works only for binary or nominal data. Future research can focus on finding more efficient algorithms to measure the nodal similarities that form the basis of our method for identifying automorphically equivalent nodes. As to the *k*-means method that forms the backbone of our method, though few meaningful theoretical bounds on the worst-case running time are known, the method is well known for its *observed speed* (Arthur and Vassilvitskii, 2006). In fact, many experimental studies have been conducted on the running time of *k*-means algorithm. For example, Har-Peled and Sadri (2005) use experimental data to show that the *k*-means algorithm terminates quickly even on large data sets. Thus, we are confident that the method presented in this study is capable of handling very large problems without being limited by its use of the *k*-means algorithm.

Although the method as demonstrated in this paper is for finding the hierarchical subgroups of a network, the new algorithm outlined in Fig. 1(b) can be used as a general approach for exploring the structure of networks. While Generalized Blockmodeling requires the pre-specification of different blockmodels, with any definition that can transform the selected type of network structure into the dissimilarity among nodes, our method can be applied to obtain an optimal number of subgroups or classes that best represent the appropriate decomposition of the network. Moreover, even for a set of elements without an adjacency matrix but only the dissimilarity measured among the elements, we can still apply our general approach to find an optimal partition for the set of elements. In this case, the discovered subgroups or classes can be examined with the context of the system (i.e. the collection of elements) so that the significance of the dissimilarity measurement can be mapped to a specified type of structure.

In the network analysis literature, other than seeking equivalence classes in a network, the other important line of research for finding meaningful subgroups is to identify cohesive subgroups. The well-known method by Newman and Girvan (2004) and methods by others (Blondel et al., 2008) have established approaches for finding cohesive classes without a priori specification of numbers of communities. These kinds of algorithms have become popular enough that some apparently think that cohesion is the only meaningful basis for describing structure in networks. As has long been known in the social networks literature, such thinking is far too limited. With this paper and the previous one (Hsieh and Magee, 2008), we now also have methods for finding equivalence and hierarchical groupings without a priori specification of numbers of communities. Thus, a given network can be analyzed by all of the approaches to identify the different interesting structures inherent in it.

It would be valuable to have a means to compare the appropriateness of the two major types of decompositions for specific networks by the use of decomposability metrics. This is currently not possible because the methods are different enough not to have congruent decomposability metrics. In this regard, we feel our general approach of decomposing a network from a pre-defined similarity matrix stands a very good chance to bridge the two lines of research. The crucial breakthrough would come if researchers can transform the structural requirements of "cohesiveness" into a similarity measure among nodes of a network. With the valid transformations, the effectiveness of decomposing a network into either cohesive subgroups or various kinds of equivalence classes can objectively be compared to each other, and thus benefit the empirical study of social networks.

References

Ahl, V., Allen, T.F.H., 1996. Hierarchy Theory, a Vision, Vocabulary and Epistemology. Columbia University Press, New York.

Arthur, D., Vassilvitskii, S., 2006. How slow is the *k*-means method? In: SCG'06, Sedona, Arizona, USA.

Batagelj, V., 1996. Ragusan families marriage networks. In: Ferligoj, A., Kramberger, A. (Eds.), Developments in Data Analysis. FDV, Ljubljana, Slovenia.

Bertalanffy, L. von, 1950. An outline of general systems theory. British Journal for the Philosophy of Science 1, 134–165.

Blondel, V.D., Guillaume, J.-L., Lambiotte, R., Lefebre, E., 2008. Fast unfolding of community hierarchies in large networks. *arXiv*:0803.0476.

Borgatti, S.P., Everett, M.G., 1992. The notion of position in social network analysis. In: Marsden, P. (Ed.), Sociological Methodology. Basil Blackwell, London.

Borgatti, S.P., Everett, M.G., 1993. Two algorithms for computing regular equivalence. Social Networks 15, 361–376.

Brusco, M.J., 2006. A repetitive branch-and-bound procedure for minimum within-cluster sums of squares partitioning. Psychometrika 71, 357–373.

Burt, R.S., Minor, M.J., 1983. Applied Network Analysis: A Methodological Introduction. Sage Publications, Beverly Hills.

Cole, B.J., 1981. Dominance hierarchies in leptothorax ants. Science 212, 83–84.

Doreian, P., Batagelj, V., Ferligoj, A., 2005. Generalized Blockmodeling. Cambridge University Press, Cambridge, U.K.; New York.

Doreian, P., Batagelj, V., Ferligoj, A., 2000. Symmetric-acyclic decompositions of networks. Journal of Classification 17, 3–28.

du Merle, O., Hansen, P., Jaumard, B., Mladenović, N., 2000. An interior point algorithm for minimum sum-of-squares clustering. SIAM Journal on Scientific Computing 21, 1485–1505.

Everett, M.G., Boyd, J.P., Borgatti, S.P., 1990. A testing example for positional analysis techniques. Social Networks 12, 253–260.

Har-Peled, S., Sadri, B., 2005. How fast is the *k*-means method? Algorithmica 41, 185–202.

Holland, P.W., Leinhardt, S., 1971. Transitivity in structural models of small groups. Comparative Group Studies 2, 107–124.

Hsieh, M.-H., Magee, C.L., 2008. An algorithm and metric for network decomposition from similarity matrices: application to positional analysis. Social Networks 30, 146–158.

Jancey, R.C., 1966. Multidimensional group analysis. Australian Journal of Botany 14, 127–130.

Jennings, H.H., 1948. Sociometry in Group Relations: A Manual for Teachers. Glenwood, Westport, CT.

Johnsen, J.C., 1985. Network macrostructure models for the Davis-Leinhardt set of empirical sociomatrices. Social Networks 7, 203–224.

Koontz, W.L.G., Narendra, P.M., Fukunaga, K., 1975. A branch and bound clustering algorithm. IEEE Transaction on Computers C-24, 908–915.

Krivošić, S., 1990. Stanovnistvo Dubrovnika i demografske promjene u proslosti.Population of Dubrovnik and Demographic Changes in Its Past, Jugoslavenska akademija znanosti i umjetnosti u Zagrebu[[nl]]Yugoslav Academy of Sciences and Arts, Dubrovnik.

Lloyd, S.P., 1982. Least-squares quantization in PCM. IEEE Transactions on Information Theory 28, 129–137.

Lorrain, F., White, H.C., 1971. Structural equivalence of individuals in social networks. Journal of Mathematical Sociology, 49–80.

Lynd, R.S., Lynd, H.M., 1937. Middletown in Transition: A Study in Cultural Conflicts. Harcourt, Brace and Company, New York.

MacQueen, J.B., 1967. Some methods for classification and analysis of multivariate observations. In: Proceedings of the 5th Berkeley Symposium on Mathematical Statistics and Probability. University of California Press, Berkeley.

Newman, M.E.J., Girvan, M., 2004. Finding and evaluating community structure in networks. Physical Review E 69.

Pattison, P.E., 1982. The analysis of semigroups of multirelational systems. Journal of Mathematical Psychology 25, 87–118.

Pattison, P.E., 1988. Network models: some comments on papers in this special issue. Social Networks 10, 383–411.

Salthe, S., 1985. Evolving Hierarchical Systems: Their Structure and Representation. Columbia University Press, New York.

Saunders, P., 1990. Social Class and Stratification. Routledge, London.

Simon, H.A., 1962. The architecture of complexity. Proceedings of the American Philosophical Society 106, 467–482.

Thurman, B., 1979. In the office—networks and coalitions. Social Networks 2, 47–63.

Warner, W.L., Lunt, P.S., 1941. The Social Life of a Modern Community. Yale University Press, New Haven.

Wasserman, S., Faust, K., 1994. Social Network Analysis: Methods and Applications. Cambridge University Press, Cambridge; New York.

White, D.R., 1985. Doug White's Regular Equivalence Program. In http://eclectic.ss.uci.edu/~drwhite/REGGE/REGGE.FOR.

White, D., Reitz, K.P., 1985. Measuring role distance: structural, regular and relational equivalence. Unpublished manuscript, University of California, Irvine.

White, H.C., Boorman, S.A., Breiger, R.L., 1976. Social structure from multiple networks. I. Blockmodels of roles and positions. American Journal of Sociology, 730–779.

Winship, C., 1988. Thoughts about roles and relations: an old document revisited. Social Networks 10, 209–231.

Winship, C., Mandel, M., 1983. Roles and positions: a critique and extension of the blockmodeling approach. In: Leinhardt, S. (Ed.), Sociological Methodology 1983–1984. Jossey-Bass, San Francisco, pp. 314–344.

Žiberna, A., 2008. Direct and indirect approaches to blockmodeling of valued networks in terms of regular equivalence. Journal of Mathematical Sociology 32, 57–84.