

Social Networks

The Correspondence Analysis of Two-Mode Networks Revisited

--Manuscript Draft--

Manuscript Number:	SON-D-24-00266
Article Type:	VSI: Duality in social networks
Keywords:	Correspondence Analysis; Generalized Relational Similarity; Duality; Two-mode networks; Centrality
Corresponding Author:	Omar Lizardo, Ph.D. University of California Los Angeles Los Angeles, CA UNITED STATES
First Author:	Omar Lizardo, Ph.D.
Order of Authors:	Omar Lizardo, Ph.D.
Abstract:	<p>This paper reconsiders the use of Correspondence Analysis (CA) in analyzing two-mode network data, highlighting aspects that have not been previously emphasized, going beyond the use of CA as a visualization tool. It argues that CA can be used to compute a dual centrality score on both modes, related to but not mathematically equivalent to the Bonacich (1991) two-mode centrality. This "reflective" centrality score connects the use of CA in two-mode network analysis to its use in other disciplines as a method of ordination. Additionally, I show CA can extract latent positional information on people and groups, with similarities to recent work on "generalized relational similarity" in two-mode networks. Thus, CA can be used for indirect community or subgroup identification in two-mode networks, connecting to its use in some disciplines as a clustering method. The paper demonstrates these applications of CA, comparing them with the Bonacich dual centrality scores, and provides guidance on using CA for structural analysis in network analysis.</p>
Suggested Reviewers:	<div>John Levi Martin jlmartin@uchicago.edu</div> <div>Stephen Borgatti sborgatti@uky.edu</div> <div>Katherine Faust kfaust@uci.edu</div> <div>John Roberts jmrob@uwm.edu</div>
Opposed Reviewers:	

Highlights

- This paper revisits the use of Correspondence Analysis (CA) in analyzing two-mode network data, highlighting its potential beyond data visualization.
- CA can compute a dual centrality score on both modes of a two-mode network, related to but not mathematically equivalent to the Bonacich (1991) two-mode centrality.
- CA can extract latent positional information on people and groups, with similarities to recent work on generalized relational similarity” in two-mode networks.

The Correspondence Analysis of Two-Mode Networks Revisited

Omar Lizardo^{a,*,1}

^aUniversity of California, Los Angeles, 264 Haines Hall, Los Angeles 90095 CA, USA

ARTICLE INFO


Keywords:

Correspondence Analysis
Generalized Relational Similarity
Duality
Two-mode networks
Centrality

ABSTRACT

This paper reconsiders the use of Correspondence Analysis (CA) in analyzing two-mode network data, highlighting aspects that have not been previously emphasized, going beyond the use of CA as a visualization tool. It argues that CA can be used to compute a dual centrality score on both modes, related to but not mathematically equivalent to the ? two-mode centrality. This “reflective” centrality score connects the use of CA in two-mode network analysis to its use in other disciplines as a method of ordination. Additionally, I show CA can extract latent positional information on people and groups, with similarities to recent work on “generalized relational similarity” in two-mode networks. Thus, CA can be used for indirect community or subgroup identification in two-mode networks, connecting to its use in some disciplines as a clustering method. The paper demonstrates these applications of CA, comparing them with the Bonacich dual centrality scores, and provides guidance on using CA for structural analysis in network analysis.

*Corresponding author

 olizardo@soc.ucla.edu (O. Lizardo)

 <http://olizardo.bol.ucla.edu/> (O. Lizardo)

ORCID(s): 0000-0002-5405-3007 (O. Lizardo)

¹This paper was written while the author was partially supported by National Science Foundation grant HNSD-R #2214215.

Highlights

The Correspondence Analysis of Two-Mode Networks Revisited

- This paper revisits the use of Correspondence Analysis (CA) in analyzing two-mode network data, highlighting its potential beyond data visualization.
- CA can compute a dual centrality score on both modes of a two-mode network, related to but not mathematically equivalent to the Bonacich (1991) two-mode centrality.
- CA can extract latent positional information on people and groups, with similarities to recent work on generalized relational similarity" in two-mode networks.

The Correspondence Analysis of Two-Mode Networks Revisited

ARTICLE INFO

Keywords:

Correspondence Analysis
Generalized Relational Similarity
Duality
Two-mode networks
Centrality

ABSTRACT

This paper reconsiders the use of Correspondence Analysis (CA) in analyzing two-mode network data, highlighting aspects that have not been previously emphasized, going beyond the use of CA as a visualization tool. It argues that CA can be used to compute a dual centrality score on both modes, related to but not mathematically equivalent to the Bonacich (1991) two-mode centrality. This “reflective” centrality score connects the use of CA in two-mode network analysis to its use in other disciplines as a method of ordination. Additionally, I show CA can extract latent positional information on people and groups, with similarities to recent work on “generalized relational similarity” in two-mode networks. Thus, CA can be used for indirect community or subgroup identification in two-mode networks, connecting to its use in some disciplines as a clustering method. The paper demonstrates these applications of CA, comparing them with the Bonacich dual centrality scores, and provides guidance on using CA for structural analysis in network analysis.

ORCID(s):

1. Introduction

While not particularly common, using Correspondence Analysis—hereafter CA—in analyzing two-mode network data has had a somewhat rocky career in the social networks literature. Initially criticized by Borgatti and Everett (1997) as a relatively limited and perhaps even inapplicable tool, CA has found various proponents who see it as an important component of the SNA arsenal for two-mode network data analysis, particularly regarding its ability to economically provide synoptic (e.g., “joint”) graphical representations of structural connectivity patterns across the two-modes (Roberts Jr, 2000; Breiger, 2000; Faust, 2005), with the primary aim being to use CA—or its variants like Multiple Correspondence Analysis (MCA)—to “visually explore” such networks (D’Esposito, De Stefano and Ragozini, 2014). This paper contributes to the stream of previous work applying CA to analyze two-mode data. Nevertheless, it departs from the already-mentioned previous efforts in highlighting aspects of CA for two-mode data analysis in SNA that have not been emphasized or treated in detail before, moving beyond the focus on data visualization.

Particularly, I show that CA can be thought of as computing a kind of *dual centrality* score on both modes—one related to but not mathematically equivalent to the Bonacich (1991) two-mode centrality—an approach that, as noted by van Dam, Dekker, Morales-Castilla, Rodríguez, Wichmann and Baudena (2021), has been recently re-invented in some corners of network science under the guise of the “economic complexity index” (Hidalgo and Hausmann, 2009; Mealy, Farmer and Teytelboym, 2019). With van Dam et al. (2021), I argue that this is a rather restricted interpretation of a more general centrality score connecting the use of CA in two-mode network analysis to way CA features in some disciplines also concerned with two-mode data (e.g., ecology) as a method of *ordination*—e.g., the discovery of latent one-dimensional orderings among a set of entities. In social network analysis, this applies to conceptualizing the venerable duality between people and the groups they join (Breiger, 1974).

In addition to its capacity to reveal latent ordinal structure in two-mode data, I also argue that CA can also be used to extract latent *positional* information on people and groups. Particularly, there is a suggestive similarity between the scores obtained from the first dimension of the CA of the two-mode affiliation matrix (the aforementioned dual centrality score) and recent work on “generalized relational similarity” in two-mode networks (Kovács, 2010; Lizardo, 2024). As such, CA recovers clusters of entities (e.g., people) linked by their similar connectivity patterns to similar entities in the other mode (e.g., groups). Thus, CA can be used as an approach to indirect “community” identification in two-mode networks connecting to how CA is used in some disciplines (e.g., computer and information science) as a *clustering* method (Zha, He, Ding, Simon and Gu, 2001), namely, the discovery of sociometrically similar subsets of entities in affiliation networks. In this way, CA emerges as a tool central to the usual network-analytic tasks and not just as a visualization or data-summarization tool.

1.1. Organization of the Paper

The rest of the paper is organized as follows. In the next section, I motivate the use of CA as producing a type of Bonacich-style dual centrality via weighted iteration across the two modes of the affiliation matrix. This leads naturally to an abbreviated way of computing the same scores via eigenvector decomposition of the (inverse degree-weighted) one-mode projection of the affiliation matrix for each set of nodes. This approach is computationally and mathematically related but *not* equivalent to the eigenvalue decomposition of the one-mode projection matrices, which results in the usual Bonacich dual centrality scores. I then draw on recent work by van Dam et al. (2021) to show the links between the ordering of nodes along the first CA dimension and Kovacs’s generalized relational similarity, and the distance of nodes in the space formed by the first two dimensions and the connectivity-similarity between nodes across modes. I contrast this ordering to that provided by the Bonacich dual centrality scores, which in contrast to the CA ordering—which reveals a dual community partition—recovers the core-periphery structure of the two-mode network instead. In closing, I provide some pointers on how CA can be used as a method for structural (e.g., centrality and positional) analysis, independently of the usual emphasis on joint graphical displays and visual exploration.

2. Reflective Centralities in Two-Mode Networks

In a highly cited piece, Hidalgo and Hausmann (2009) motivated what they saw at the time as a novel way of ranking nodes in a two-mode network—effectively computing a version of two-mode centrality—using what they called at the time a “reflective” approach. Hidalgo and Hausmann’s original empirical application was to the two-mode country-by-product networks, hence the original baptizing of their approach as one geared to extracting the “economic complexity” of countries in the world system (and dually the complexity of given products). Subsequent work shows that there is no logical connection between the formal method and this particular application since the approach proposed can be

used to analyze any two-mode data matrix. As such, I introduce it here using the more intuitive—and classical from a social network analysis perspective—case of the duality of persons and groups (Breiger, 1974).

If we are going to compute the centrality of nodes in a two-mode network, the most natural place to start is with the good old *degree centrality* (Faust, 1997). A two-mode network composed of a set of people P and their affiliation relations to a set of groups G can be represented by an affiliation matrix A of dimensions $|P| \times |G|$ with people along the rows and groups across the columns, where $|P|$ is the cardinality of the people set and $|G|$ is the cardinality of the group set, with cell entries $a_{pg} = 1$ if person p is affiliated with group g and $a_{pg} = 0$ otherwise.

Given this, the degree-centrality of people is given by:

$$C_p^R(1) = \sum_g a_{pg} \quad (1)$$

And for groups:

$$C_g^R(1) = \sum_p a_{pg} \quad (2)$$

That is, the first-order centrality of people is the row sum of the entries in the affiliation matrix A , and the column sums of the same matrix give the first-order centrality of groups. As noted by Hidalgo and Hausmann (2009), the key to the reflective approach is the observation that, once we have these first-order quantities, it is possible to compute “second-order centralities” $C^R(2)$ for both people and groups using the (averaged) first-order centralities of the entities in the other mode they are connected to.

For people, these are given by:

$$C_p^R(2) = \frac{1}{C_p^R(1)} \sum_g a_{pg} C_g^R(1) \quad (3)$$

And for groups:

$$C_g^R(2) = \frac{1}{C_g^R(1)} \sum_p a_{pg} C_p^R(1) \quad (4)$$

Equation 3 says “people are more central when the average sum of the size of the groups they belong to is large” (e.g., whenever $a_{pg} = 1$ and $C_g^R(1)$ is a big number). Equation 4 says “groups are more central when the average activity of their members is large” (e.g., whenever $a_{pg} = 1$ and $C_p^R(1)$ is a big number). Of course, we can keep on going and define third-order reflections.

For the people, these are given by:

$$C_p^R(3) = \frac{1}{C_p^R(1)} \sum_g a_{pg} C_g^R(2) \quad (5)$$

And for groups:

$$C_g^R(3) = \frac{1}{C_g^R(1)} \sum_p a_{pg} C_p^R(2) \quad (6)$$

Equation 5 says something like “people are more central when the average sum of the average activity of the members of the groups they belong to is large” (e.g., $a_{pg} = 1$ and $C_g^R(2)$ is a big number). Equation 6 says, “groups are more central when the average sum of the average size of the groups their members belong to is large.” (e.g., $a_{pg} = 1$ and $C_p^R(2)$ is a big number).

Table 1
Southern Women Data.

	E1	E2	E3	E4	E5	E6	E7	E8	E9	E10	E11	E12	E13	E14
EVELYN	1	1	1	1	1	1	0	1	1	0	0	0	0	0
LAURA	1	1	1	0	1	1	1	1	0	0	0	0	0	0
THERESA	0	1	1	1	1	1	1	1	1	0	0	0	0	0
BRENDA	1	0	1	1	1	1	1	1	0	0	0	0	0	0
CHARLOTTE	0	0	1	1	1	0	1	0	0	0	0	0	0	0
FRANCES	0	0	1	1	1	1	0	1	0	0	0	0	0	0
ELEANOR	0	0	0	1	1	1	1	1	0	0	0	0	0	0
RUTH	0	0	0	1	1	0	1	1	1	0	0	0	0	0
VERNE	0	0	0	0	0	0	1	1	1	0	0	1	0	0
MYRA	0	0	0	0	0	0	0	1	1	1	0	1	0	0
KATHERINE	0	0	0	0	0	0	0	1	1	1	0	1	1	1
SYLVIA	0	0	0	0	0	0	1	1	1	1	0	1	1	1
NORA	0	0	0	0	0	0	1	0	1	1	1	1	1	1
HELEN	0	0	0	0	0	0	1	1	0	1	1	1	0	0
OLIVIA	0	0	0	0	0	0	0	0	1	0	1	0	0	0
FLORA	0	0	0	0	0	0	0	0	1	0	1	0	0	0
PEARL	0	0	0	0	0	1	0	1	1	0	0	0	0	0
DOROTHY	0	0	0	0	0	0	0	1	1	0	0	0	0	0

Note: Rows ordered according to the generalized blockmodeling solution of Doreian et al. (2004, Table 4).

Note that for the people, the even-numbered reflection $C_p^R(2)$ assigns centrality based on a formal feature of the *groups* they belong to (in this case, the group sizes). On the other hand, the odd-numbered reflection $C_p^R(3)$ assigns centrality based on a formal feature of the *members of the groups* they belong to (in this case, the average size of the groups they belong to). In the same way, for the groups, the even-numbered reflection $C_g^R(2)$ assigns centrality based on a formal feature of the *people* who belong to them (in this case, their activity). On the other hand, the odd-numbered reflection $C_g^R(3)$ assigns centrality based on a formal feature of the *other groups their members belong to* (in this case, their average group size). While these are distinct metrics, in practice, the ordering of the nodes in each mode ends up being identical across even and odd centralities after the ranks “freeze” past a given number of iterations (proportional to the network size).

More generally, Hidalgo and Hausmann (2009) show that we can define a series of reflective quantities for people and groups (whose verbal and substantive interpretation becomes more complex as the number of iterations increases).

For people, these are given by:

$$C_p^R(q) = \frac{1}{C_p^R(1)} \sum_g a_{pg} C_g^R(q-1) \quad (7)$$

And for groups:

$$C_g^R(q) = \frac{1}{C_g^R(1)} \sum_p a_{pg} C_p^R(q-1) \quad (8)$$

Equation 7 says that the reflective centrality of a person p at iteration q is the sum of the $q-1$ centralities of the groups they belong to ($C_g^R(q-1)$) divided by their number of memberships $C_p^R(1)$. Equation 8 says that the q^{th} group reflective centrality is the sum of $q-1$ centralities $C_p^R(q-1)$ of their members, divided by the number of group members $C_g^R(1)$.

2.1. Empirical Example

Figure 1 shows the trajectory of the HH reflective centralities for persons (top panel) and groups (bottom panel) in the Southern Women data (Davis, Gardner and Gardner, 1941), shown in Table 1. The rank order of people and groups in the q^{th} centrality is plotted on the y-axis, and the centrality iteration q is plotted on the x-axis.

As the bottom panel of the figure shows, *NORA*, *FLORA*, *CHARLOTTE*, and *EVELYN* are the top-ranked actors when it comes to $C_p^R(2)$: The average number of members of the groups they belong to. However, their fates in this reflective metric diverge at higher reflections, with *NORA* and *FLORA* maintaining their top positions but *CHARLOTTE* and *EVELYN* tumbling down the ranks, suggesting that the members of the groups they belong to affiliate with smaller groups than the members of the groups *NORA* and *FLORA* belong to (and so on for higher reflections). Notably, the reflective centrality rankings after freezing ($q \geq 18$) recover the Doreian et al. (2004) block partition (see Table 1), but this time with Dorothy and Pearl in the middle separating the two largest blocks of women.

The bottom panel of Figure 1 shows the corresponding bump chart for the events. Just like for these actors, the equilibrium reflective centralities recover the ordering of the columns according to Doreian et al. generalized blockmodel with events 1-6 separated from events 10-15 by events 7-9 (see also Kovács (2010)). Figure 1 shows that *E14* experiences the most dramatic improvement in status as we move to higher reflections. Relatively low ranked when it comes to the average number of memberships of its members, it increases in standing when considering the average of the average number of memberships of its members (and so forth).

3. Correspondence Analysis, Dual Centrality, and Communities in Two-Mode Networks

The reader may ask what the point of going through all these reflective centralities, since Bonacich (1991) already developed a dual conception of centrality in two-mode networks based on a very similar idea: Define the centralities of entities in one mode based on the centralities of entities in the other mode to which they are connected. In that paper, Bonacich also considered CA in passing but dismissed its application to centrality rankings. Instead, Bonacich noted that, in analyzing the Southern Women data shown in Table 1, “[r]ather than centrality, what they [CA scores] seem to capture is membership in the two cliques that attended two different sets of events” (1991, 164).

As we will see, Bonacich was only half right. Indeed, the CA scores capture a version of “clique” (more accurately, community) membership, but they—particularly the first dimension—also capture a version of centrality in the most general sense of ranking nodes based on a meaningful criterion (Borgatti and Everett, 2006). In fact, they capture the limit ($q \rightarrow \infty$) of the reflective centralities discussed in the previous section (Mealy et al., 2019).

To see this, recall that the solutions to the following system of linear equations give the Bonacich (C^B) two-mode centralities:

$$AC_p^B = \lambda C_g^B \quad (9)$$

$$A^T C_g^B = \lambda C_p^B \quad (10)$$

Where A is the original affiliation matrix. Equations 9 and 10 have the typical form of a linear eigensystem, which means the unknown C_g^B and C_p^B scores can be obtained from the row and column eigenvectors corresponding to the largest eigenvalue λ , obtained from the singular value decomposition (SVD) of the rectangular affiliation matrix A . Importantly, as Faust (1997, 170) notes, it is possible to express Equations 9 and 10 in terms of person and group-specific centralities.

For people, these are given by:

$$C_p^B = \frac{1}{\lambda} \sum_g a_{pg} C_g^B \quad (11)$$

And for groups:

$$C_g^B = \frac{1}{\lambda} \sum_p a_{pg} C_p^B \quad (12)$$

Where λ is the (first) eigenvalue corresponding to the eigenvector containing the C_p^B scores. These equations make clear that in the Bonacich two-mode, the centrality of people is proportional to the centralities of the groups they join,

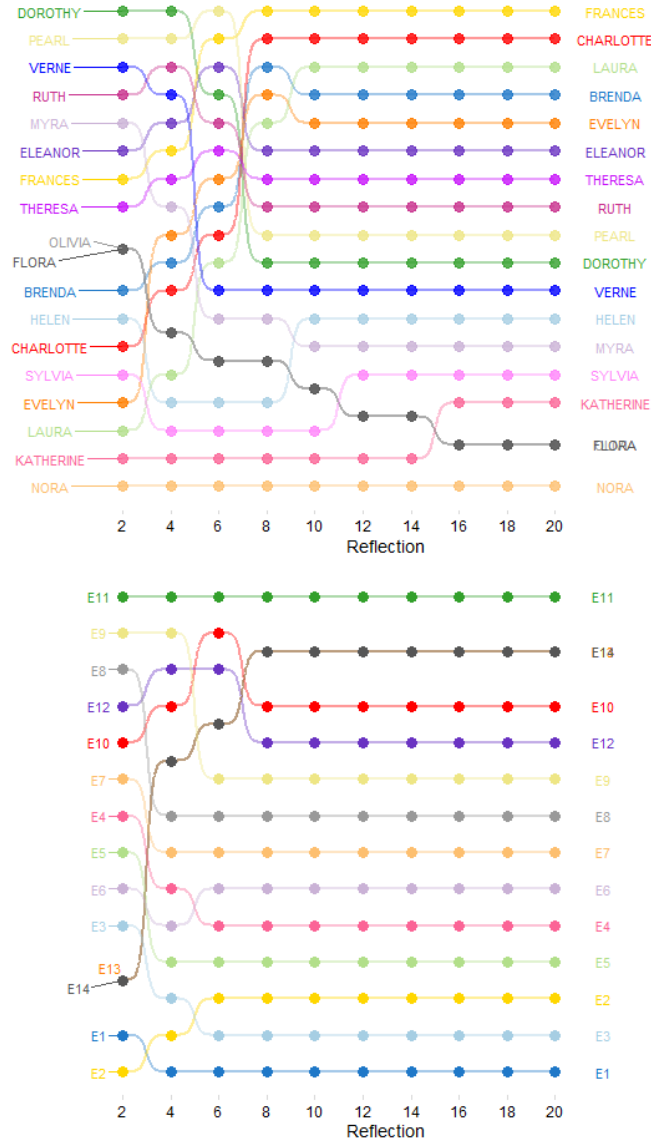


Figure 1: Reflective centrality trajectories of persons and groups (even reflections).

and the centrality of groups is proportional to the centralities of their members. These capture the duality property since the centralities of nodes in each mode are given by aggregating their connections to nodes in the other mode. Note the formal similarity between equations 11 and 12 and equations 7 and 8. The main difference is that in the reflective equations, a person's centrality is proportional to the *activity-weighted* centralities of the groups they join, and a group's centrality is proportional to the *group size weighted* centrality of the people who are members. We will return to this crucial point later.

3.1. The Bonacich Two-Mode Centrality as a Reflective Centrality

To see the connection between the reflective and Bonacich two-mode centralities more clearly, we can motivate the Bonacich two-mode centrality using the same “reflective” approach we used to introduce the HH reflective centrality in Section 2. Admittedly, this is a somewhat unorthodox way of presenting the eigenvector centrality for two-mode networks (Bonacich, 1991), but it will help clarify the similarities and differences between the HH and the Bonacich

approaches. Accordingly, starting with Equations 1 and 2, we can define a second-order “Bonacich-reflection” on both the persons and groups using the formulas:

$$C_p^B(2) = \sum_g a_{pg} C_g^R(1) \quad (13)$$

$$C_g^B(2) = \sum_p a_{pg} C_p^R(1) \quad (14)$$

Equation 13 says that people are central when the sum of the number of members of the groups they belong to is large. Equation 14 says groups are central when the sum of the number of memberships of the people who belong to them is large.

As before, we can keep on going and define a third-order reflection using the formulas:

$$C_p^B(3) = \sum_g a_{pg} C_g^R(2) \quad (15)$$

$$C_g^B(3) = \sum_p a_{pg} C_p^R(2) \quad (16)$$

Equation 15 says that people are central when the sum of the sum of the number of memberships held by the people who belong to the groups they belong to is large. Equation 16 says that groups are central when the sum of the sum of the number of members of the groups their members belong to is large. Once again, we can keep going and define fourth order, fifth order, and higher reflections $C_p^R(4), C_p^R(4) \dots C_p^R(q)$, where the centralities of nodes in one mode are based on the sums, of the sums, of the sums, of the centralities of nodes in the other mode.

More generally, and in parallel with equations 7 and 8, the reflective Bonacich centralities for persons and groups are given by:

$$C_p^B(q) = \sum_g a_{pg} C_g^R(q-1) \quad (17)$$

$$C_g^B(q) = \sum_p a_{pg} C_p^R(q-1) \quad (18)$$

This just says that the Bonacich reflective centrality at step q is just the sum of the group centralities at step $q-1$ (for people) and the sum of the person centralities at step $q-1$ (for groups).

Of course, it is evident that these sum of sums would diverge to a bigger and bigger quantity at each step q . To prevent this and guarantee convergence, we normalize the vector of reflective Bonacich centralities for persons and groups at each step $q > 1$ before calculating the subsequent sum at step $q+1$ as follows:

$$C_p^B(q) = \frac{C_p^B(q)}{\|C_p^B(q)\|_2} \quad (19)$$

$$C_g^B(q) = \frac{C_g^B(q)}{\|C_g^B(q)\|_2} \quad (20)$$

Where the denominator in 19 and 20 is the Euclidean vector norm.¹ The normalization will prevent divergence of the sum of centralities for persons and groups, formalizing the weaker assumption of *proportionality* between the

¹For any vector \mathbf{x} of length n , the L_2 norm is given by: $\|\mathbf{x}\|_2 = \sqrt{\sum_i^n x_i^2}$.

centrality of each set of nodes and the sum of the centralities of the nodes in the other mode to which they are connected rather than the stronger assumption of strict equivalence (Bonacich and Lloyd, 2001). Furthermore, the normalization guarantees convergence and the “freezing” of the sums around steady values after a few iterations. These values will be equivalent (up to rounding error) to the (absolute value) of the dominant row and column eigenvectors of \mathbf{A} as given in 9 and 10.² In fact, the iterative (normalized sum of sums) approach is one way of computing the leading eigenvectors of a rectangular matrix (e.g., the “power” method of Mises and Pollaczek-Geiringer (1929)).

This exercise establishes that there is more than a superficial similarity between the HH “reflective” centralities and the Bonacich two-mode centralities. In fact, both can be seen as instantiating an underlying reflective model of how centrality is distributed in the two-mode network, with the Bonacich approach presuming that centrality points are distributed equally across nodes in each mode regardless of their own centrality (both large and small degree nodes distribute the same “amount” of centrality to others) and the HH reflective approach normalizing by the centrality of nodes in each mode, so that nodes distribute a given centrality quantum that is proportional to their first-order degree centrality, with large degree nodes having less centrality to distribute than low degree nodes.

3.2. HH Reflective Centrality as an Eigenvector Centrality

Since iterating through normalized sums of sums is one way of obtaining the Bonacich two-mode centralities, it would be surprising if the equilibrium values of the HH reflective iterations were not themselves the solution to some eigenvalue decomposition problem. As has been noted recently by Mealy et al. (2019) and van Dam et al. (2021), the Hidalgo and Hausmann’s (2009) reflective centralities can indeed be obtained directly (without iterations) as the solution to an eigenvalue decomposition problem.

To see this, recall that, as Bonacich (1991, 157) notes, we can solve for C_g^B in 10 and C_p^B in 9 and substitute the respective solutions into 9 and 10. Matrix-algebraic reduction of the resulting equations would show that the Bonacich dual centralities can also be obtained as solutions to the eigensystem:

$$(AA^T) C_p^B = \lambda^2 C_p^B \quad (21)$$

$$(A^T A) C_g^B = \lambda^2 C_g^B \quad (22)$$

This shows that the dual centrality Bonacich scores for persons and groups are equivalent to the eigenvectors of the respective one-mode projection matrices corresponding to the first (largest) eigenvalue.

Now, consider the $|P| \times |P|$ matrix Dp , which contains the “first order” reflective centralities of each person $C^R(1)_p$ (activity) along the diagonals and zeroes in every other cell. In the same way, consider the $|G| \times |G|$ matrix Dg , which contains the “first order” (degree) centralities of each group $C^R(1)_g$ (size) along the diagonals and zeroes in every other cell. It can be shown (van Dam et al., 2021), that in the limit ($q \rightarrow \infty$), the iterative HH reflective centralities can be obtained as the solution of the eigensystem:

$$(Dp^{-1} A Dg^{-1} A^T) C_p^R = \lambda C_p^R \quad (23)$$

$$(Dg^{-1} A^T Dp^{-1} A) C_g^R = \lambda C_g^R \quad (24)$$

Note the formal similarity (and key differences) between these equations and the Bonacich two-mode centralities in equations 21 and 22. Both extract individual and group centralities as eigenvectors of the one-mode projection of the original affiliation matrix: AA^T for people and $A^T A$ for groups. The difference is that the reflective centralities pre-multiply the affiliation matrix and its transpose by the inverse of the first-order centralities of the nodes in each mode before computing the eigenvalue decomposition, essentially normalizing the one-mode projection by the degrees of both sets of nodes.

²Note that dividing by any vector norm—e.g., the L_1 or max norm—will prevent divergence and return scores perfectly correlated to the Bonacich eigenvector approach. Dividing by the L_2 norm returns scores that are *exactly* the same, save for rounding error, as the absolute value of the dominant left and right eigenvectors of the affiliation matrix.

This can be clearly seen if we express $(Dg^{-1}A^TDp^{-1}A)$ in 23 in terms of each cell entry (Mealy et al., 2019, eq. 4):

$$a_{pp'} = \sum_g \frac{a_{pg}a_{p'g}}{C_p^R(1)C_g^R(1)} = \frac{1}{C_p^R} \sum_g \frac{a_{pg}a_{p'g}}{C_g^R(1)} \quad (25)$$

In equation 25, the numerator is equal to one when person p and person p' share membership in a group g . Summed across groups, this gives the number of memberships that p and p' have in common (Breiger, 1974). As noted, the Bonacich dual centralities are obtained from the eigenvector corresponding to the first eigenvalue of this matrix of shared memberships (for people) and shared people (for groups). The reflective centralities, on the other hand, are given by the eigenvectors of a weighted version of the same matrix, where the weights are the sizes of each of the groups p shares with each other person summed across groups and divided by the total number of p 's memberships.

The same reasoning applies to groups in 24, whose entries are given by:

$$a_{gg'} = \sum_p \frac{a_{pg}a_{pg'}}{C_p^R(1)C_g^R(1)} = \frac{1}{C_g^R} \sum_p \frac{a_{pg}a_{pg'}}{C_p^R(1)} \quad (26)$$

Now, as has been noted by two-mode data analysts in other contexts (e.g., Faust, 2005) it turns out that this “double-pre-weighting” of each cell entry by the degrees of each mode (e.g., the row and column sums of the original affiliation matrix) is precisely that used for the CA of a two-mode matrix.³ And indeed, the limit reflective centralities obtained from equations 23 and 24—which will be given by the eigenvectors corresponding to the *second* largest eigenvalue of the resulting solution—will be equivalent to the row and column scores corresponding to the first non-trivial dimension (for people and groups respectively) obtained from a simple CA of the original affiliation matrix (Fouss, Saerens and Shimbo, 2016, 398, eq. 9.17).⁴ In fact, the method of “reflections” is just a re-discovery of the older idea of reciprocal averaging (Hill, 1973), which is yet another way of computing row and column scores for elements in a two-mode matrix that are substantively equivalent to those obtained by (the first non-trivial dimension of) CA (Mealy et al., 2019).

Degree-pre-weighting does not radically alter the nature of the data. Indeed, it corresponds to moving from sums to averages, thus “adjusting” for the influence of person-activity and group size—a seemingly perennial issue in two-mode data analysis (Bonacich, 1991, 159ff). This can be seen by the fact that Equations 25 and 26 show that, substantively, what the degree pre-weighting does is that, for people, co-memberships count for more in determining interpersonal similarity when the group in question is small than when it is large. On the group side, shared members who do not have many affiliations count more in determining intergroup similarity than those with many affiliations.

Accordingly, it would make no sense to call the Bonacich scores C_p^B and C_g^B from equations 21 and 22 “centrality scores” but fail to hold that designation from C_p^R and C_g^R , given that the mathematics are not only “similar” (Bonacich, 1991, 162) but formally identical. In the Bonacich case, an eigendecomposition of the one-mode projection of the two-mode network; in the reflective case, an eigendecomposition of the degree-weighted one-mode projection of the same network. Accordingly, the CA of a two-mode network will return—along the first dimension—a rank-ordered score for nodes in each mode that meets all the conditions for qualifying as a centrality measure for two-mode networks. Particularly, the CA centrality retains the duality property (Faust, 2005, 128), as the (average) centrality of people is a function of the (average) centrality of the groups they belong to. The (average) centrality of groups, in turn, is a function of their members’ (average) centrality. The person and group score ranks given by the first CA dimension are thus better thought of as *activity and group-size normalized* versions of the familiar Bonacich (1991) dual centralities for two-mode networks.

4. Illustrative Analysis of the Southern Women Data

4.1. Row and Column Re-ordered Affiliation Matrices

Figures 2(a) and 2(b) illustrate the key differences between the two versions of dual centrality C^R and C^B . Each panel shows the original Southern Women affiliation matrix but with rows and columns re-ordered according to the

³More accurately, cells are weighted by the inverse of the square root of the product of the row, and column sums (e.g., Faust, 2005, 124).

⁴As Faust (2005, 126) notes, the first non-trivial CA dimension is also given by the eigenvector corresponding to the second-largest eigenvalue of $D_p^{-1/2}AD_g^{-1/2}$.

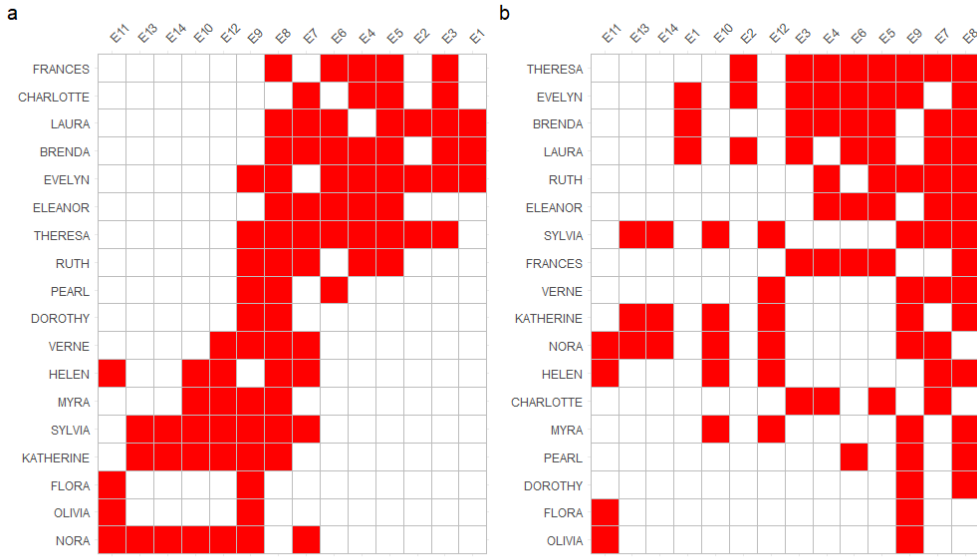


Figure 2: Southern Women Affiliation matrix with re-ordered rows and columns: (a) row and columns re-ordered according to the person and group score on the first CA dimension, (b) row and columns re-ordered according to person and group Eigenvector centrality score.

first CA dimension in (a) and the Bonacich (1991) centrality in (b). In the plot, a cell entry is colored red if it has a one in the original affiliation matrix and is white when the corresponding entry is zero. As we can see, the two row-column-reshuffled matrices have appreciably distinct structures, with the C^R re-ordered affiliation matrix having a block-diagonal structure and the C^B re-ordered affiliation matrix having a triangular structure. Accordingly, the C^R centrality rankings reveal a dual community partition between groups of women who selectively attend two groups of events (on the top-right and bottom-left of the plot). The traditional eigenvector centrality re-ordering, on the other hand, reveals a classic core-periphery partitioning (Borgatti and Everett, 2000), with a group of highly active women (on the top-left) who selectively co-participate in highly attended events (on the top right-hand side of the plot) and less active women (in the bottom half) who go to less well-attended events (on the left side). The two centralities extract qualitatively different information from the two-mode network, with C^R geared toward community partitioning and C^B more focused on finding a core of well-connected people and groups.

4.2. Eigenvector Plot

We can confirm that the first CA axis points toward the underlying community partitioning of the two-mode network by looking at Figure 3(a) and 3(c), which shows the C^R score of persons and groups on the x-axis against the score's rank order on the y-axis. If a two-mode network has no discernible community structure, the first dimension CA scores would be distributed as a continuous logistic curve; when community structure is present, we should observe discernible breaks in this distribution (van Dam et al., 2021). The Southern Women data clearly belong to the second category. In the people mode, we have [*Frances, Laura, Brenda, Charlotte, Evelyn, Theresa, Eleanor, Ruth, Pearl*] on one side, [*Dorothy, Nora, Katherine, Sylvia, Flora, Olivia, Myra, Helen, Verne*] on the other. Among events, we have [*E1, E2, E3, E4, E5, E6*] on side [*E9, E10, E11, E12, E13, E14*] on the other and [*E7, E8*] in an ambiguous middle position. Note that this is roughly the same partitioning obtained by (Doreian, Batagelj and Ferligoj, 2004) and reproduced by Kovács (2010) and Lizardo (2024) using a generalized relational similarity approach. This indicates that persons and groups receive similar scores in the first dimension of CA only when they have similar connectivity similar to *similar* groups, where group similarity is defined dually as having members in common who are themselves similar.

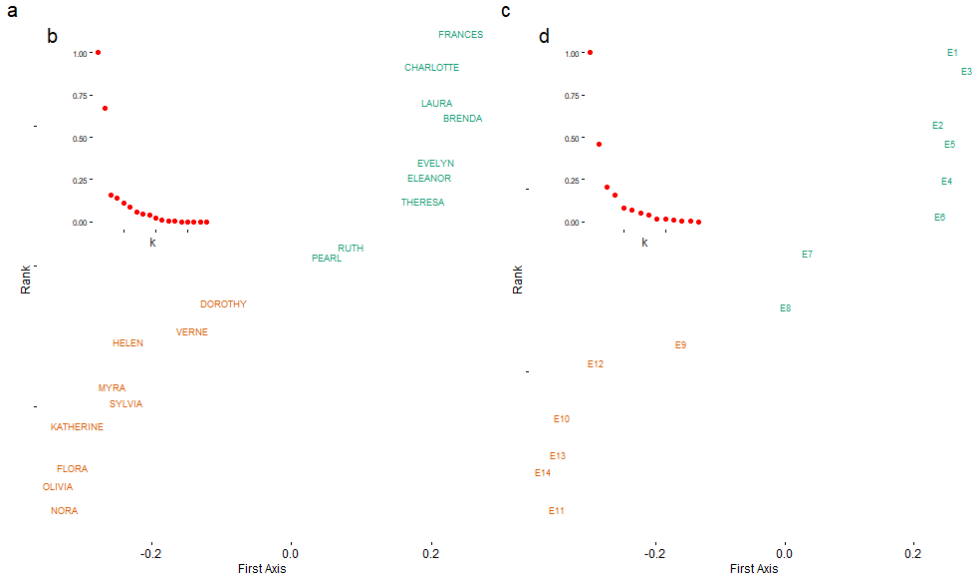


Figure 3: CA Eigenvector plot for persons (a) and groups (c). The first CA score is on the x-axis, and the rank order of persons and groups is on the y-axis. Insets show the eigenvalue orderings on the x-axis (k) for persons (b) and groups (d).

4.3. Normalized Similarity Plot

To further appreciate the links between CA, reflective centralities, and community partitioning in two-mode networks, note that $S_p = ADg^{-1}A^T$ in 23 can be thought of as giving the similarity between pairs of people weighted by the size of the groups they belong to (so that people are more similar when they share smaller memberships).⁵ The same interpretation can be given to $S_g = A^T Dp^{-1}A$ in 24, which is the similarity of groups weighted by the activity of the people in them (so that groups are more similar when they share members who do not belong to many other groups). As van Dam et al. (2021) show, the eigenvectors corresponding to (the Laplacian⁶ of) these similarity matrices are equivalent to C^R . This means that the relative spread of the eigenvalues of the degree-weighted similarity matrix provides information concerning the quality of the resulting community partition. These are shown as inset plots in Figure 3 for both people and groups. Note that in both cases, the first two eigenvalues separate from the rest, strongly indicating a dual-community structure in the Southern Women two-mode network.

To fix the intuition, the top-left panel (a) of Figure 4 shows a heatmap representation of the group-size weighted similarity matrix S_p (for people); the top-right panel (b) shows the same plot for the activity-weighted similarity matrix S_g (for groups). In both figures, darker squares indicate more similar node pairs, and closer to white indicate less similarity. Moreover, the rows of each matrix are sorted according to C^R for both people and groups. As we can see, for both people and groups, the re-ordered matrix recovers maximally similar communities of nodes, where the similarity is based on their shared (degree-weighted) connections to entities in the other mode.

4.4. CA Correspondence Plot

The bottom panel of Figure 4 shows the usual correspondence plot of the first two CA dimensions. People and groups are colored according to a six-cluster k-means solution based on the first nine eigenvectors of their respective similarity matrices. The plot reveals that the distances between nodes in the standard correspondence plot—usually taken to be a low-dimensional representation of the *original* affiliation matrix (Borgatti and Everett, 1997)—are best thought of as low-dimensional representations of the (other mode’s) degree-normalized similarity network

⁵Note that Newman (2001, eq.2) proposes a slightly modified version of this two-mode similarity score for people—in Newman’s case, scientists—except that the one mode projection is normalized as $A(Dg - I)^{-1}A^T$; namely, the size of the “group”—the number of co-authors on a scientific paper—minus one. Substantively, this is unlikely to make much difference as the rank order of dyads by similarity between the two scores will be identical.

⁶For the similarity matrix S , the Laplacian is defined as $D - S$ where D is the matrix containing the degrees of either people or groups in the diagonal and zeroes in every other cell.



Figure 4: CA similarity plots for persons (a) and groups (b) on the left and right top panel. The bottom panel (c) shows the correspondence plot for persons and groups with scores corresponding to the first dimension on the horizontal axis and scores corresponding to the second dimension on the vertical axis.

across people and groups. People with similar values in S_p —such as [*Olivia, Flora*], [*Katherine, Sylvia, Nora*], [*Dorothy, Verne*], and [*Brenda, Laura, Evelyn*]⁷—appear closer in the correspondence plot (and are assigned to the same cluster by the k-means algorithm). The same goes for events, those with similar values in S_g —such as [*E13, E14*] and [*E1, E2, E3*]⁷—appear closer in the correspondence plot, while those with dissimilar values are placed far apart. The distances between same-mode entities in the correspondence plot will thus be a function of their (inverse) degree-normalized similarity.

Note that this differs from the usual interpretation of the CA correspondence plot, which is usually taken to bring together nodes with “similar” connectivity patterns, where similarity is presumed to be a function of their raw row profiles (for people) or column profiles (for groups). This interpretation implies (for instance) that two people who attend many of the same events or two groups with many members will appear close in the plot. But this (still common) interpretation is off the mark. What the CA correspondence plot distance captures is, instead, people and groups that are *surprisingly* similar (e.g., from the point of view of a suitable null model, like independence) after taking people’s activity levels and the sizes of the groups they belong to into account.⁷ Thus, people who share memberships in small groups will be closer in the diagram than people who share memberships in big groups. In the same way, groups

⁷Note that this, finding “surprising” similarities in terms of the indirect paths linking nodes in a network after the main effects of node-connectivity are considered, is the same rationale given by Leicht, Holme and Newman (2006) for weighting the Katz (1953) similarity matrix by the degrees of the corresponding nodes (see Fouss et al., 2016, 68, eqs. 2.13 and 2.14). As we have seen, this is precisely the key contrast between the Bonacich and the CA dual centrality measures for two-mode networks.

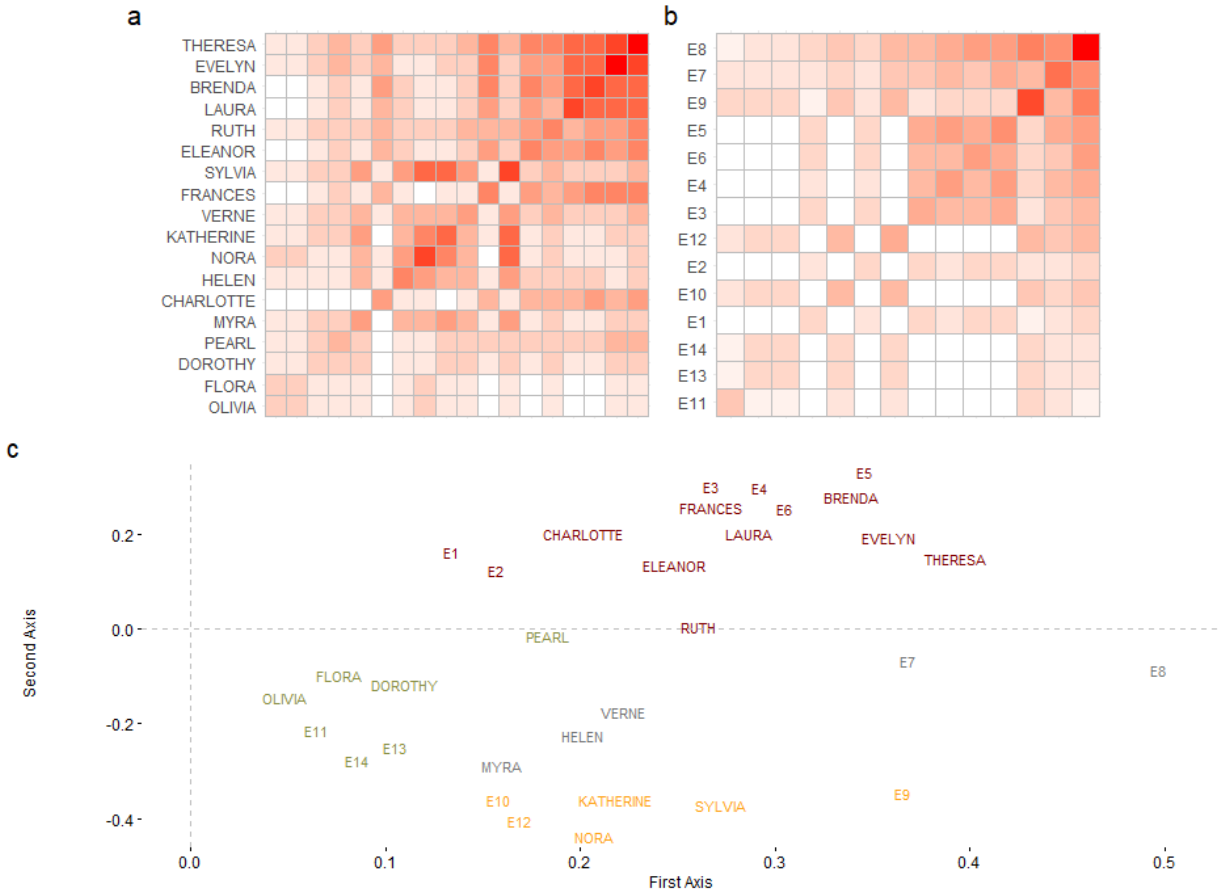


Figure 5: Bonacich dual centrality similarity plots for persons (a) and groups (b) on the left and right top panel. The bottom panel (c) shows the Bonacich eigenvector centrality correspondence plot for persons and groups with scores corresponding to the standard Bonacich eigenvector centrality on the horizontal axis and scores corresponding to the second largest eigenvector of the affiliation matrix on the vertical axis.

that share people with few memberships will be closer in the diagram than those sharing people with many other memberships.

4.5. One-Mode Projection Matrix and Eigenvector Plot

This ambiguity in interpretation may stem from the fact that CA is usually seen as a technique to generate a plot that provides a “low-dimensional approximation to the input data” (Faust, 2005, 125), where the “input data” is presumed to be the original affiliation matrix \mathbf{A} . But as we have seen, this is *not* what CA is designed to do. Instead, CA is meant to provide a low-dimensional approximation of a *transformed* version of the input data, where the transformation is meant to adjust for people’s activity levels and group sizes. Notably, if a low-dimensional representation of the original “input data” (\mathbf{A}) is what we are after, this may be more closely approximated by the first two eigenvectors of the usual one-mode projections of the matrix (where the first is, of course, the standard Bonacich dual centrality score).

To illustrate, Figures 5(a) and 5(b) shows the “raw” (unweighted by other-mode degree) similarity scores for persons ($a_{pp'} = \sum_g a_{pg} a_{p'g}$) and groups ($a_{gg'} = \sum_p a_{pg} a_{p'g'}$), with the rows and columns sorted by the first eigenvector of the matrix, which is the usual Bonacich centrality score. Both similarity matrices reproduce the triangular, core-periphery structure we observed in the re-ordered affiliation matrix in Figure 2(b). Figure 5(c) plots C^B on the x-axis against the second eigenvector of the unweighted similarity matrix—a relatively unusual but not substantively unmotivated practice (Iacobucci, McBride and Popovich, 2017).⁸ We can see that the plot of the first two eigenvectors does a

⁸Nodes are colored to a four-cluster k-means solution using the first six eigenvectors of the respective similarity matrices.

good job of recovering the raw connectivity structure of the Southern Women affiliation network, partitioning the core persons and groups (on the upper-right) from the more peripheral ones (on the lower-left).

Moreover, if all we are interested in is capturing a low-dimensional representation of which people have particular affinities for which events (regardless of people's activity levels or group sizes), then Figure 5(c) does a better job of that than the usual CA correspondence plot in Figure 4(c). For instance, *[Flora, Olivia]* do have a special affinity for *[E11]* and *[Nora, Katherine, Sylvia]* for *[E10, E12]*. In the same way, core events like *[E8, E9, E10]*—shown on the lower half of the plot—are different from core events *[E3, E4, E5, E6]*—shown in the upper half. The former are more inclusive of peripheral members while the latter are more “cliquish,” including only core members.

This “preferential attachment” (Barabási and Albert, 1999) of core people to core events and peripheral people to peripheral events seems to be captured by the second dimension, left over after considering each node's Bonacich eigenvector centrality. For instance, *[E1, E2]*, although as poorly attended as other peripheral events, tends to include core members and thus appear closer to other core actors in the upper half of the plot. Similarly, the main difference between *[Ruth]* and *[Sylvia]*, despite their comparable C^B scores, is that the former's (closer to the upper half of the plot) event profile is composed mainly of core events. In contrast, peripheral events dominate the latter's attendance profile (shown in the bottom half of the plot), accordingly, *[Ruth]* and *[Sylvia]* are assigned to distinct clusters by the k-means algorithm.

4.6. Correspondence Analysis and Generalized Relational Similarity

As noted, there seems to be a relationship between the ordering of persons and groups produced by CA of a two-mode network and that produced by previous work using a “generalized relational similarity” (GRS) strategy. Recall that for objects (let us say people) to be similar according to the GRS criterion, they must have overlapping connections, and those links should go to objects in the other mode *that are themselves similar*, where objects' similarities are given by their pattern of connections to objects in the other mode. This recursive definition of similarity thus recalls the classic distinction between structural and regular equivalence (Everett and Borgatti, 1994). In the context of two-mode networks, Kovács (2010) proposed one such approach to defining a GRS for nodes in one and two-mode networks using a modified version of the correlation distance.⁹

An earlier effort to define a GRS for persons and groups in two-mode networks, one more relevant for a direct comparison with CA and the “reflective” approaches already considered, can be found in Jeh and Widom (2002). In that work, the authors dubbed their similarity measure “SimRank.” In the context of two-mode network analysis, the goal is to compute a matrix of similarities for each of the two-node sets, where the similarity of people is a function of the groups they belong to, and the similarity of groups is a function of the people who belong to them, making the similarity of persons and groups “mutually reinforcing notions” (Jeh and Widom, 2002, 540). Thus,

- People are *similar* if they belong to *similar* groups.
- Groups are *similar* if they share *similar* members.

Which is consistent with a GRS approach (see Kovács, 2010; Lizardo, 2024). To accomplish this, Jeh and Widom (2002, 540, eq. 2 and eq. 3) propose that we define the similarity of each pair of people $S(p, p')$ as given by:

$$S(p, p') = \frac{\alpha}{C^{R(1)}_p C^{R(1)}_{p'}} \sum_{i=1}^{C^{R(1)}_p} \sum_{j=1}^{C^{R(1)}_{p'}} S(g(i)_{i \in N(p)}, g(j)_{j \in N(p')}) \quad (27)$$

Where everything is as before, and $g(i)_{i \in N(p)}$ is the i^{th} group in the set of groups that person p belongs to, $g(j)_{j \in N(p')}$ is the j^{th} group in the set of groups that person p' belongs to, and α is a free parameter obeying the restriction: $0 > \alpha < 1$. By construction, $S(p, p) = 1$, for all p . Thus, equation 27 says that the SimRank similarity between two people is a function of the sum of the similarities of each unordered pair of groups they both belong to, weighted by the reciprocal of the products of their number of memberships multiplied by α .

Likewise, for groups, the SimRank similarities are given by:

⁹Lizardo (2024) shows the connection between Kovacs's idea of generalized relations similarity and the two-mode network projection.

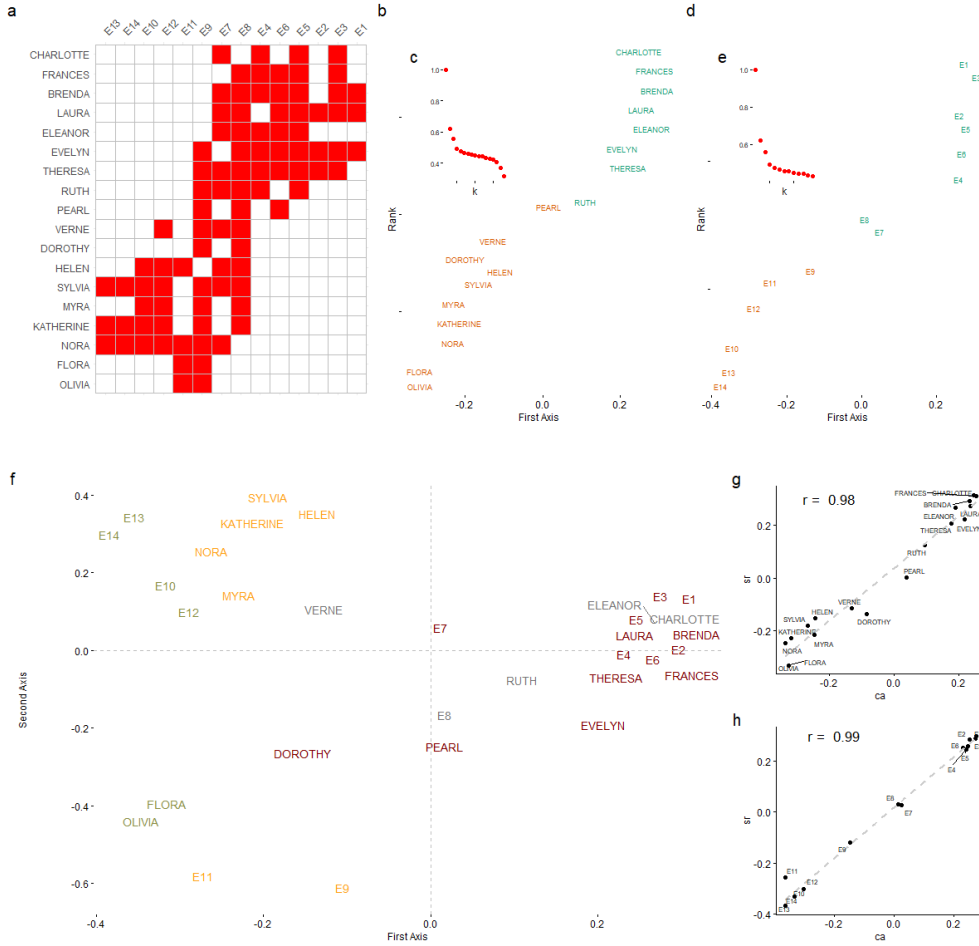


Figure 6: Simrank versus CA comparison. Panel (a) shows the affiliation matrix re-ordered according to the first eigenvector of the SVD of the SimRank similarity matrix. Panels (b) and (d) show the corresponding eigenvector and eigenvalue—panels (c) and (e)—plots for persons and groups, respectively, of the SimRank similarity matrix. Panel (f) shows the correspondence plot based on the eigenvector decomposition of the SimRank similarity matrix. Panels (g) and (h) show the Pearson correlation between the scores corresponding to the first eigenvector of the SimRank similarity matrix (on the y-axis) and the scores corresponding to the first CA dimension (on the x-axis).

$$S(g, g') = \frac{\alpha}{C^{R(1)}_g C^{R(1)}_{g'}} \sum_{i=1}^{C^{R(1)}_g} \sum_{j=1}^{C^{R(1)}_{g'}} S(p(i)_{i \in N(g)}, p(j)_{j \in N(g')}) \quad (28)$$

Where $p(i)_{i \in N(g)}$ is the i^{th} person in the set of members of group g , and $p(j)_{j \in N(g')}$ is the j^{th} person in the set of members of group g' , and $S(g, g) = 1$. Thus, the SimRank similarity between two groups is a function of the sum of the similarities of each pair of people who belong to both groups, weighted by the reciprocal of the products of the number of members of each group multiplied by α .

In two-mode networks, SimRank scores for each pair of nodes across the two sets can be estimated via a simple algorithm, in which we first estimate $S(p, p')$ in equation 27 using baseline values. Hence, only groups that two people share contribute to the initial values of $S(p, p')$ since only $S(g, g) > 0$ at the outset. We then plug those values into equation 28, then loop back to equation 27 with the resulting $S(g, g')$ values, and continue iterating until convergence—generally achieved after five iterations (Jeh and Widom, 2002), which is confirmed here for the Southern Women data.

Note that equations 27 and 28 share a formal similarity with HH's "method of reflections" discussed earlier, in particular, the fact that both compute quantities based on nodes in the other mode averaged by the degree of nodes in the focal mode. The key difference is that SimRank works directly with pairwise comparisons between node dyads (Jeh and Widom, 2002). Nevertheless, this double weighting by degree should make us suspect that the results of the SimRank similarity analysis would not be too far from those obtained via CA, given the mathematical equivalence of CA and the method of reflections (Mealy et al., 2019; van Dam et al., 2021).

And indeed, they are not. Figure 6(a) shows the original affiliation matrix, this time with rows and columns rearranged according to the values of the main non-trivial eigenvector (corresponding to the second-largest eigenvalue) of the SimRank similarity matrices—with α set to 0.8—for the row and column nodes. Like in Figure 2(a), this reshuffling reveals the block-diagonal structure separating the two communities of persons and groups revealed by CA, suggesting that CA and SimRank uncover a similar underlying partitioning of the two-mode network's community structure.

Figure 6(b) and Figure 6(d) show a plot similar to that shown in Figures 3(a) and Figure 3(b) but this time with the main eigenvector of SimRank similarity matrices of both persons and groups on the x-axis and the corresponding rank on the y-axis. Looking at the plots from left to right, we can see that the partitioning of the node sets on the first informative eigenvector is substantively equivalent to those revealed by the first CA eigenvector, separating similar blocks of people and events.

Figure 6(f) shows the correspondence plot of the first and second eigenvectors of the SimRank similarity matrices for both persons and groups. Once again, this plot is substantively equivalent (in terms of the internode groupings and distances) to that shown in Figure 4(c), suggesting that the first-CA dimensions of the two-mode network recover clusters of "similar" entities in the two node sets, where internode similarity is a generalized relational similarity as defined earlier. Thus, [Flora, Olivia] appear close because they choose similar groups, as do [Sylvia, Katherine, Helen]. In the same way, groups [E9, E11] are close in the plot because they are chosen by similar people, as are events [E10, E12, E13, E14]. Figures 6(g) and (h) show a regression plot of the first CA eigenvector for persons (top) and groups (bottom) on the x-axis and the first eigenvector of the SimRank similarity matrix for both persons and groups on the y-axis. As we can see, the CA and SimRank ordering of nodes along the first dimension agree quite closely ($r = 0.98$ for persons and $r = 0.99$ for groups), confirming that these first eigenvectors capture the same (GSR) information across the two approaches.

5. Discussion and Concluding Remarks

This paper reconsiders the role of CA in the analysis of two-mode network data. We began with an accidental "rediscovery" of CA in the analysis of two-mode networks via a "reflective" approach (Hidalgo and Hausmann, 2009), showing that the method of reflections leads to an eigenvector-style solution that is equivalent to simple CA of the affiliation matrix. Working backward from this, we also clarified the linkages between CA and the more commonly used eigenvector approach for calculating dual centralities in two-mode data due to Bonacich (1991). I showed that just like the reflective centralities that converge to the CA scores of the affiliation matrix, we could also think of the Bonacich centralities as the equilibrium solution to reflective iterations through the two-mode matrix, with the main difference being that the CA reflections deal with sums of averages weighted by the degrees of nodes in each mode, while the Bonacich approach works with unweighted (but normalized) sums. This exercise clarifies the links between CA and the dual two-mode eigenvector centrality in a more coherent and systematic way, an issue that was left open and somewhat unclear in Bonacich's (1991) classic paper. Thus, one conclusion that emerges from this analysis is that CA computes a kind of centrality for two-mode networks, equivalent to Hidalgo and Hausman's "reflective" centrality (van Dam et al., 2021).

But CA does more than reveal a latent dual ordering of nodes in the two-mode network. In addition to doing this, CA can be shown to reveal *groupings* of nodes based on some conception of the similarity of their connections to nodes in the other mode. This is evident once we use the scores of the first CA dimension to re-order the rows and columns of the original affiliation matrix. When the CA scores are used, a clear (and now well-known) dual partition between persons and groups emerges in the classic Southern Women data. This partition is substantively distinct from that which would emerge if we use the same approach to re-ordering the original data using the Bonacich centralities, which instead uncover a latent core-periphery partition between actors and events.

This is another way in which CA and the Bonacich dual centrality approach systematically differ; one is geared to uncovering blocks of actors with similar connections to events in the other mode (and vice versa), while the other reveals blocks of actors who are the most active and who attend the largest events. Both are, of course, legitimate ways

of analyzing the structure of a two-mode network. Still, they differ in terms of the structural patterns that they are sensitive to, with the CA approach closer to a community partition where nodes that are surprisingly similar end up in the same clusters—where “surprising” means similarity based on deviations from a suitable null model based on random mixing given their first-order degrees.

But what kind of similarity is the clustering based on the CA of the two-mode network sensitive to? Here we argued that CA recovers a type of generalized relational similarity (GRS). That is, actors who have similar patterns of linkage to similar events are deemed similar, while events that have similar patterns of connectivity to similar actors are also deemed similar. Using a well-known iterative method to recover such generalized similarities from two-mode networks (Jeh and Widom, 2002), we saw that the scores from the first dimension of CA end up being a fairly accurate approximation to the resulting partition from the generalized relational similarity approach. Thus, we can clarify that two-mode network CA reveals latent groupings of nodes based on a GRS criterion.

Overall, the preceding has shown that CA can be upgraded from a method designed to generate joint plots and visualization of two-mode data to one that can be seen as more “central” to the usual social-network-analytic tasks, like ordering the nodes in the two sets according to some substantively meaningful rank—centrality analysis—or finding sets of similar actors in the network (subgroup or community detection). At the very least, it seems like the Bonacich style dual centrality based on the eigenvector decomposition of the raw affiliation matrix should not be the default “reflective” centrality approach unless the analyst has the explicit analytic goal of exploring center-periphery partitioning in the network.

A better approach, similar to the one exemplified here, would be to present a comparison of the reflective centralities obtained by CA and Bonacich side-by-side to see whether the underlying display a substantively interesting similarity partitioning in addition to any core-periphery ordering. Of course, suppose the analyst is more interested in such a “subgroup” analytic approach. In that case, the Bonacich dual centrality approach is less helpful (even if multiple dimensions are considered), and CA should be the first line of attack.

References

- Barabási, A.L., Albert, R., 1999. Emergence of scaling in random networks. *science* 286, 509–512.
- Bonacich, P., 1991. Simultaneous group and individual centralities. *Social networks* 13, 155–168.
- Bonacich, P., Lloyd, P., 2001. Eigenvector-like measures of centrality for asymmetric relations. *Social networks* 23, 191–201.
- Borgatti, S.P., Everett, M.G., 1997. Network analysis of 2-mode data. *Social networks* 19, 243–269.
- Borgatti, S.P., Everett, M.G., 2000. Models of core/periphery structures. *Social networks* 21, 375–395.
- Borgatti, S.P., Everett, M.G., 2006. A graph-theoretic perspective on centrality. *Social networks* 28, 466–484.
- Breiger, R.L., 1974. The duality of persons and groups. *Social forces* 53, 181–190.
- Breiger, R.L., 2000. A tool kit for practice theory. *Poetics* 27, 91–115.
- van Dam, A., Dekker, M., Morales-Castilla, I., Rodríguez, M.Á., Wichmann, D., Baudena, M., 2021. Correspondence analysis, spectral clustering and graph embedding: applications to ecology and economic complexity. *Scientific reports* 11, 8926.
- Davis, A., Gardner, B.B., Gardner, M.R., 1941. *Deep South: A Social Anthropological Study of Caste and Class*. University of Chicago Press, Chicago.
- Doreian, P., Batagelj, V., Ferligoj, A., 2004. Generalized blockmodeling of two-mode network data. *Social networks* 26, 29–53.
- D’Esposito, M.R., De Stefano, D., Ragozini, G., 2014. On the use of multiple correspondence analysis to visually explore affiliation networks. *Social Networks* 38, 28–40.
- Everett, M.G., Borgatti, S.P., 1994. Regular equivalence: General theory. *Journal of mathematical sociology* 19, 29–52.
- Faust, K., 1997. Centrality in affiliation networks. *Social networks* 19, 157–191.
- Faust, K., 2005. Using correspondence analysis for joint displays of affiliation networks. *Models and methods in social network analysis* 7, 117–147.
- Fouss, F., Saerens, M., Shimbo, M., 2016. *Algorithms and models for network data and link analysis*. Cambridge University Press.
- Hidalgo, C.A., Hausmann, R., 2009. The building blocks of economic complexity. *Proceedings of the national academy of sciences* 106, 10570–10575.
- Hill, M.O., 1973. Reciprocal averaging: an eigenvector method of ordination. *The Journal of Ecology* , 237–249.
- Iacobucci, D., McBride, R., Popovich, D., 2017. Eigenvector centrality: Illustrations supporting the utility of extracting more than one eigenvector to obtain additional insights into networks and interdependent structures. *Journal of Social Structure* 18, 1–22.
- Jeh, G., Widom, J., 2002. Simrank: a measure of structural-context similarity, in: *Proceedings of the eighth ACM SIGKDD international conference on Knowledge discovery and data mining*, pp. 538–543.
- Katz, L., 1953. A new status index derived from sociometric analysis. *Psychometrika* 18, 39–43.
- Kovács, B., 2010. A generalized model of relational similarity. *Social Networks* 32, 197–211.
- Leicht, E.A., Holme, P., Newman, M.E., 2006. Vertex similarity in networks. *Physical Review E* 73, 026120.
- Lizardo, O., 2024. Two-mode relational similarities. *Social Networks* 76, 34–41.
- Mealy, P., Farmer, J.D., Teytelboym, A., 2019. Interpreting economic complexity. *Science advances* 5, eaau1705.
- Mises, R., Pollaczek-Geiringer, H., 1929. Praktische verfahren der gleichungsauflösung. *ZAMM-Journal of Applied Mathematics and Mechanics/Zeitschrift für Angewandte Mathematik und Mechanik* 9, 58–77.

- Newman, M.E., 2001. Scientific collaboration networks. ii. shortest paths, weighted networks, and centrality. *Physical review E* 64, 016132.
- Roberts Jr, J.M., 2000. Correspondence analysis of two-mode network data. *Social Networks* 22, 65–72.
- Zha, H., He, X., Ding, C., Simon, H., Gu, M., 2001. Bipartite graph partitioning and data clustering, in: *Proceedings of the tenth international conference on Information and knowledge management*, pp. 25–32.