The Correspondence Analysis of Two-Mode Networks Revisited

ARTICLE INFO

Keywords:
Correspondence Analysis
Bonacich Centrality
Duality
Two-mode networks
Clustering

ABSTRACT

This paper reconsiders the status of Correspondence Analysis (CA) as a tool for analyzing two-mode networks, comparing it with the Bonacich dual centrality approach and revealing the mathematical linkages between them as eigenvector-based methods. While Bonacich centrality identifies core-periphery structures and is helpful for clustering nodes based on the criterion of similarity via structural equivalence, CA is best at detecting subsets of actors and events based on a generalized relational similarity criterion, thus coming closer to clustering via regular equivalence. Ultimately, both CA and Bonacich centrality prove to be valuable yet distinct strategies for the dual projection analysis of two-mode networks, highlighting the duality between actors and events.

ORCID(s):

1. Introduction

Despite being a frequently used tool for analyzing two-mode network data, Correspondence Analysis (hereafter CA) has had a somewhat rocky yet distinctive 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, 2014b).

This paper reconsiders the use of CA as a tool in the analysis of two-mode networks, clarifying its similarities and differences with other widely used approaches, such as Bonacich's (1991) dual centrality scoring. I show that both CA and Bonacich centrality analysis are closely related mathematical cousins, and can be viewed as eigenvector-based decomposition strategies of proximity matrices derived from the one-mode projection of the original two-mode affiliation matrix. The key differences between CA and the Bonacich approach concern the type of underlying structure in the two-mode network that they are sensitive to, with the Bonacich approach more useful for detecting primary and secondary core-periphery partitions—and when used as a blocking tool revealing clustering of nodes based on structural equivalence—and CA more useful for revealing clusters of similar nodes, where similarity relations obey a logic closer to that of regular equivalence, or "generalized relational similarity" (Kovács, 2010). In this respect, both Correspondence Analysis and Bonacich dual centrality analysis emerge as equally useful yet distinct strategies within the extended toolkit of options for analyzing two-mode networks from a "dual projection" perspective (Everett and Borgatti, 2013).

1.1. Organization of the Paper

The rest of the paper is organized as follows. In Section 2, I review previous SNA work dealing with the uses of CA to analyze two-mode network data. Then, in Section 3, I consider a relatively recent re-invention of CA (the economic complexity index) used to produce a type of Bonacich-style dual scoring via iterative scoring across the two modes of the affiliation matrix. This leads naturally to an abbreviated method for computing the same scores — CA and Bonacich — via spectral decomposition of the one-mode projection of the affiliation matrix for each set of nodes, weighted by node degree in the case of CA and unweighted in the case of Bonacich. Section 4 uses the classic Southern Women dataset to show the payoff of this approach, focusing on the distinct substantive insights extracted by CA and Bonacich scoring in a venerable dataset, showing that the CA-blocked affiliation matrix reveals a dual block partition between the two node sets (similar to that recovered by previous analyses of the same data using other methods), while the Bonacich-blocked affiliation recovers a core-periphery structure instead. In addition, I show that considering higher dimensions of the Bonacich-style decomposition facilitates the identification of secondary core-periphery structures in the two-mode network, while also yielding a meaningful clustering of nodes based on structural equivalence. Section 5 considers the hanging question of the exact substantive interpretation of the node clusters revealed by CA-style blocking, which are shown to be fairly close to those obtained by a seemingly computationally unrelated approach based on the idea of "generalized relational similarity," suggesting that the CA blocks are distinct from those obtained following the logic of structural equivalence. Finally, Section 6 closes with a summary of key points in the argument and suggestions for future development of the linkages between CA and positional and centrality analysis in two-mode networks.

2. Previous uses of CA in Two-Mode Network Data Analysis

For reasons of space and relevance, my discussion of previous work is circumscribed in three ways, focusing on (1) pieces discussing primarily *methodological* or interpretive aspects of CA, (2) regarding the analysis of *two-mode network* data or affiliation networks,[1] (3) analyzing the usual *binary* affiliation matrix measuring a single relation

[1]There are, of course, various substantive applications of CA for the analysis of two-mode networks (e.g., Breiger, 2000; Faust, Willert, Rowlee and Skvoretz, 2002; Schweizer, 1991; Serino, Picardi and Ragozini, 2024; Ragozini, Serino and D'Ambrosio, 2018) as well as substantive and methodological applications of CA to the analysis of one-mode network data (e.g., Noma and Smith, 1985; Kumbasar, Rommey and Batchelder, 1994; Lizardo and Taylor, 2020); these will not be part of the discussion here.

between persons and groups at one point in time, where links are observed only between persons and groups,[2] using *simple CA* of the affiliation matrix.[3]

One of the earliest considerations of CA as a data-analytic tool for examining duality relations in two-mode networks was provided by Bonacich (1991, 163-165) in his classic paper on dual centralities for affiliation networks. In that paper, Bonacich (1991) noted that since the usual affiliation matrix is a "cross-classification of people by the groups to which they belong," then it would seem that CA would be a relevant analytic technique. According to Bonacich (1991), the basic goal of correspondence analysis is to create a multidimensional space encoding geometric "distances" (scare quotes in the original) between row and column categories, with the distances capturing the best low-dimensional representation—in the least-squared error sense—of the similarities between row and column objects. Bonacich's discussion of CA came after his influential introduction of the Eigenvector centrality approach for two-mode network analysis. Intriguingly, Bonacich (1991, 163) noted that "[a]lthough the goals of correspondence analysis and centrality analysis are different, their mathematics are similar," because both rely on a version of the spectral decomposition of the affiliation matrix—namely, the Singular Value Decomposition (SVD) to obtain scores for the row and column objects in the two-mode network. As we will see in section 3.4, CA and centrality analysis are mathematically linked in a second way not noted by Bonacich, as they are also based on the spectral decomposition of the dual one-mode *projections* of the original rectangular affiliation matrix. In this particular respect, the mathematical connection between the Eigenvector approach to centrality analysis and CA is more similar than even Bonacich appreciated.

For Bonacich, the main differences between CA and Eigenvector centrality analysis stem from "the different ways in which the cross-classifications are treated before the SVD is applied." For Bonacich, the "most powerful and useful dimensions" of CA are designed to highlight and describe *similarities* in row or column patterns of affiliation and membership, and therefore "[b]efore the SVD is used, the cross-classification is modified so that these dimensions of similarity will be the most powerful." Nevertheless, Bonacich is not quite clear as to what this pre-modification amounts to, noting that given the fact that less "powerful" dimensions are usually discarded, it could be that among these "there may be a dimension corresponding to centrality," but "correspondence analysis is not designed to highlight this dimension." Interestingly, Bonacich went on to compute both the Eigenvector and CA scores for the canonical Southern Women data—which we will also do later in this paper—noting that "[r]ather than centrality what…[the CA scores] seem to capture is membership in the two cliques that attended different sets of events," with the magnitude of each woman and event score in the main dimension seeming to indicate the "purity" of their membership in each of the two cliques. For events, on the other hand, the magnitude of the CA scores seems to indicate the likelihood of being attended by their most representative members (Bonacich, 1991, 164).

In an influential paper on centrality in two-mode networks, Faust (1997) returned to the issue of the connection between eigenvector centrality, CA, and dual relations between actors and events in two-mode networks. Faust noted that one thing both techniques had in common was that they both produced scores that truly exploited the duality of the two modes in the data. In the eigenvector case, this duality is strict, as the eigenvector centrality of a given person is the sum of the eigenvector scores of the event she attends; for events, their scores are given by the sum of the eigenvector centralities of their members. Importantly, Faust (1997) noted that an exact parallel dual additive relationship was obtained between the CA scores corresponding to persons and groups. The CA score of a given person on a given dimension is the (weighted) sum of the scores in the same dimension of the events she attends. Similarly, for groups, the score on a given CA dimension is the weighted sum of the scores of the women who attend that event, a classic case of duality by mutual definition.

Borgatti and Everett (1997) provided another early and influential in-depth consideration of CA in the context of two-mode network data analysis with an emphasis on the visualization aspect of actors and events in a single space. The discussion was placed in a section of the paper titled "visual representations of 2-mode data," and naturally focused almost exclusively on the uses of CA as a visualization technique. According to Borgatti and Everett (1997,

[2] As such, work dealing with the application of CA or other factoring approaches to time-varying, multiplex, multilevel, or valued two-mode network data (Ragozini, De Stefano and D'Ambrosio, 2014; Ragozini, De Stefano and D'Esposito, 2015; Zhu, Kuskova, Wasserman and Contractor, 2016) will also be outside the scope of my discussion.

[3] There has been a spate of work focusing on the uses and advantages of using Multiple Correspondence Analysis or MCA—a variant of CA applied to a disjunctively-coded, doubly-centered version of the original affiliation matrix) to analyze and visualize two-mode networks (e.g., Dramalidis and Markos, 2016; D'Esposito et al., 2014b; Ragozini et al., 2014). Because much of what I have to say deals with the uses and interpretation of the scores obtained from simple CA applied directly to the affiliation matrix (or its one-mode projections), my discussion will not be directly relevant to this line of work, except for a paper directly comparing the substantive interpretation of the χ^2 distances in simple CA and MCA with regard to their substantive interpretation as indexes of structural similarity, which bears more directly to some of the subsequent discussion (D'Esposito, De Stefano and Ragozini, 2014a).

246), CA "can be seen as a method for representing points in a metric space such that distances between the points are meaningful." More specifically, Borgatti and Everett (1997), also using the Southern Women data as illustration, claimed that in the CA map, points representing the women will appear closer if the women attend "mostly the same events." Conversely, points representing the events will be placed near one another if the events were attended by "mostly the same women," and that women and events would appear closer in the plot if "those women attended those events." Borgatti and Everett (1997, 247-249) then went on to focus on various limitations of CA as a visualization tool, noting that given the limiting range of dichotomous network data, two-dimensional representations "will almost always be severely inaccurate or misleading." They also noted that since CA is a technique designed to deal with valued frequency data, its use with binary network data would result in dimensional scores that would be either "meaningless" or "difficult to interpret." Finally, they noted that since distances in the CA biplot are not Euclidean, they would mislead interpreters who would try to impose such a metric to make sense of them.

In a paper published shortly thereafter, Roberts Jr (2000) responded to some of the criticisms Borgatti and Everett (1997) levied at the use of CA for analyzing two-mode networks. First, Roberts Jr (2000) noted that many examples could be found using CA to analyze binary non-frequency data and that, therefore, the uses of CA are not constrained by the nature or provenance of the numerical quantities recorded in each cell of a two-mode network affiliation matrix. Roberts Jr (2000) also challenged the contention that the two-dimensional representation of the data was unduly limiting in its capacity to produce a faithful representation of the original data, noting a high correlation between the distances between the women and the event points in the two-dimensional representation and one that used all thirteen dimensions.[4] Finally, Roberts Jr (2000) also noted that the Euclidean distance interpretation of the separation of women and events points in the joint display is partially appropriate, as these Euclidean distances as actually low-dimensional approximations (to a high degree of accuracy in the first few dimensions) to the $\mathcal{X}^2$ distances between row profiles (for the women) and column profiles (for events) where the row profiles for a given woman is equal to the entry in the original affiliation matrix divided by the square root of the product of the number of events attended by that woman and the number of women who attended that event (and similarly for column profiles corresponding to events). Thus, the point distances are indeed interpretable as a kind of *degree-weighted similarity* between objects in each mode. This is a point that we will return to later.

Perhaps the most thorough discussion of the uses of CA as a tool for the dual "joint display" of affiliation networks is provided by Faust (2005). According to Faust (2005, 118), previous debates regarding the usefulness—or lack thereof—of CA as a visualization tool were marred by imprecision and lack of "formal specification" regarding the exact substantive meaning of the scores obtained via the procedure and a related lack of sustained discussion as to the exact "relationship between the model and the input data." Faust's basic point is that CA provides a way to place actors and events in an affiliation network in a joint space, such that the inter-point distances are interpretable and the link to the original data (e.g., the affiliation network matrix) is transparent. Along the way, Faust (2005) established several important points. First, Faust specifies the exact matrix that is subject to the singular value decomposition when CA is applied directly to the affiliation matrix of a two-mode network $\mathbf{A}$; namely, a scaled version where $\mathbf{A}$ is pre-multiplied by a diagonal matrix containing the reciprocal of the square root of its row sums and post-multiplied by a diagonal matrix containing the reciprocal of the square root of its column sums. This is equivalent, as both Faust (1997) and Roberts Jr (2000) had noted before, to dividing every non-zero entry of a_{pg} by the square root of the product of the number of memberships of person p and the number of members of group g before applying the SVD. Second, Faust specifies the exact relationship linking the resulting row and column CA scores to the original affiliation matrix input, providing the "reconstruction equation" that takes us from the left and right singular vectors and eigenvalues to the former. Thirdly, as she did in the 1997 paper, Faust clarifies the duality linking the row and column scores, noting that the scores for the people are (activity) weighted sums of the scores that the groups they join, and the scores for the groups are (group size) weighted sums of the scores of their members. Finally, Faust (2005, 128ff) clarifies the link between the CA scores—and different normalizations thereof—assigned to persons and events in an affiliation network on each dimension and (same node-set) inter-point distance in a joint space. Like Roberts Jr (2000) before her, Faust specifies that in the low (usually two) dimensional representation, inter-point distances are best (least squares) approximations of the chi-square distance between the membership profiles of pairs of people (or groups), where the

[4] In CA, the maximum number of dimensions is one less than the rank of the affiliation matrix, which in this case is fourteen, corresponding to the number of events.

profile of a given person-row or group-column is equal to the non-zero entries of the affiliation matrix divided by the corresponding row or column sum (people's activities and group sizes).[5]

D'Esposito et al. (2014a) compare CA and MCA as tools for representing the structural similarity between nodes in a two-mode network (actors in their case). Specifically, they analyze the significance of the typical χ^2 distances defined from the scores obtained from the simple CA of the affiliation matrix, as compared to the MCA of the indicator matrix of persons by groups, with events treated as disjunctively coded binary variables measured over the actors (as described earlier). The focus of the analysis is the extent to which the χ^2 distance of actor profiles obtained from the affiliation matrix **A** or the disjunctively coded indicator matrix **Z** best approaches a criterion for computing approximate structural equivalence between nodes (such as the Euclidean distance computed from the binary entries in the affiliation matrix).[6] That is, D'Esposito et al. (2014a) seek to ascertain whether CA or MCA is the best analytic tool to identify sets of *structurally equivalent* actors in a two-mode network.

Interestingly, D'Esposito et al. (2014a, 115) note that by examining the respective formulas defining the profile distances in CA versus MCA, we can verify that CA is not a good tool for revealing clusters of structurally equivalent actors. As D'Esposito et al. (2014a) put it, the reason is that the CA profile distance "adopts a peculiar and more complex weighting system that could lead to a distorted representation of the actual relational patterns" because the CA "distance between actors (events) depends, not only on the pattern of participation (attendance)" but also on the actor's *degree* and the event *sizes*. In addition, they note that because of this weighting scheme, in CA, "small size events...are associated to larger weights...[and therefore] differences...related to such small size events have [a] stronger impact on the overall...distance than differences related to larger size events." D'Esposito et al. (2014a) also point out that, in contrast, the MCA definition of inter-profile distance between actors based on **Z** is not affected by the degree-weighting aspect, and that in this respect, "[I]t turns out that MCA [distances] closely resembles the Euclidean distance," which is a straightforward criterion for structural equivalence. This is a conclusion that is supported by computational experiments conducted later in the paper, and which leads D'Esposito et al. (2014a) to conclude that the MCA defined χ^2 distances "better reproduces the *actual* relational pattern embedded in affiliation networks" (122, italics added).

2.1. Summary and Open Questions

As we can see from this brief and necessarily selective review, considerations of CA for two-mode network analysis have generally intertwined two basic themes. One beginning with Bonacich (1991) and continuing through Faust (1997, 2005) has commented on the formal commonalities and contrasts between "centrality analysis" and "correspondence analysis," noticing striking similarities in the mathematical formulations of the traditional dual (eigenvector) score and the scores obtained from CA, as both come from the eigendecomposition of the affiliation matrix, resulting in scores that meet the strict criterion for duality, as the scores of nodes in one mode are usually additive functions of the scores of nodes in the other mode. The other stream, which began with characterizations of CA as a visualization tool, focuses on the substantive meaning of the (low-dimensional representations of) the distances defined over the multidimensional space returned by the CA procedure.

In this last case, multiple interpretations have been offered, from the idea that CA returns clique-like structures (Bonacich, 1991), to intimations that the closeness in the CA dimensions implies a more structured similarity closer to the idea of structural equivalence (e.g., such that actors that attend the same events and events that the same actors attend will be closer in the space). This approach supports the notion that the primary objective of CA is to provide a parsimonious representation of the original "input data" (Faust, 2005), where this is understood to mean the original affiliation matrix. Yet, more recently, this line of interpretation has been controverted by (D'Esposito et al., 2014a) who have noticed that whatever the CA similarities are designed to do, they are a less-than-optimal representation of structural equivalence (and thus presumably the original input data), given the specific weighting by the degree of nodes in each mode that is built into their construction. As such, the exact relationship that the CA representation bears to the original affiliation matrix data, as well as the similarities and contrasts between correspondence analysis

[5]Note that because the chi-square distances are defined on these degree-weighted entries, they define a kind of *degree-weighted similarity* between pairs of people and pairs of groups, but not an exact similarity score as would be obtained, for instance, by using a criterion such as the proportion of matches in the affiliation matrix (Everett and Borgatti, 2013, 208).

[6]In the case of MCA applied to the binary affiliation matrix, the latter is transformed into a new matrix **Z** with the same number of rows (persons) as the original and with twice the number of columns representing either membership ($z_{pg+} = 1$) or non-membership ($z_{pg-} = 1$) in that a group (disjunctive coding). The MCA scores are obtained via SVD of a centered and doubly normalized—with respect to person activity and group sizes—version of **Z** (D'Esposito et al., 2014b).

and centrality analysis, and the relationship between these two approaches to traditional positional analysis in social networks, remains to be specified and clarified.

In one of the earliest and most comprehensive efforts introducing CA for the analysis of the composition and structure of relational data, Wasserman, Faust and Galaskiewicz (1990, 19-21) noted that the history of CA in the social and behavioral sciences was one of invention, collective forgetting, and subsequent re-invention and independent rediscoveries, happening both within disciplines over time and across various fields of scholarly investigation at roughly contemporaneous intervals throughout the twentieth century.[7] In the rest of the paper and starting in the next section, I show how a re-invention of CA for the analysis of "economic complexity"—two-mode networks of countries/regions by the products they are most adept at producing—in the now well-established field of "Network Science" in the twenty-first century, allows us to revisit controversies and open questions as to the meaning of CA scores in two-mode network data, and the relationship between correspondence analysis, centrality analysis, and positional analysis (e.g., the discovery of structurally similar clusters of actors in two-mode networks). Along the way, we will be able to shed light on the link between CA scores and traditional eigenvector centrality scores for two-mode networks, as well as ascertain the exact kind of similarity that is learned by estimating CA (and traditional eigenvector centrality) scores for nodes in two-mode networks along multiple dimensions.

3. Reflective Node Scores in Two-Mode Networks

In a highly cited piece, Hidalgo and Hausmann (2009) motivated what they saw at the time as a novel approach to assigning meaningful scores to nodes in a two-mode network, using what they called a "reflective" approach at the time. Hidalgo and Hausmann's original empirical application was to the two-mode country-by-product networks. Hence, they referred to their approach as a way to extract the "economic complexity" of countries in the world system (and dually the complexity of given products).[8] 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). This is consistent with the spirit of the original approach, which has been described by Hidalgo (2021, 92) as centered on the "*duality* between economic inputs and outputs" (italics mine).

If we are going to assign substantively meaningful scores to 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 $\mathbf{A}$ 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:

$$C^R_p(1) = \sum_g a_{pg} \tag{1}$$

And for groups:

$$C^R_g(1) = \sum_p a_{pg} \tag{2}$$

That is, the first-order centrality of people is the row sum of the entries in the affiliation matrix $\mathbf{A}$, 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 to computing the "complexity" of countries and products is the observation that, once we have these first-order quantities, it is possible to compute "second-order" quantities $C^R(2)$ for both people and groups using the *averaged* first-order centralities of the entities in the other mode they are connected to.

[7] As we will see in the next section, the pattern of CA reinvention and independent rediscovery continues into the twenty-first century, now involving disciplines outside the social and behavioral sciences narrowly conceived, including computer science, information science, and network science.

[8] For a state-of-the-art review of the various substantive applications of the economic complexity approach see Hidalgo (2021).

For people, these are given by:

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

And for groups:

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

While Equation 1 assigns a high value to people who belong to a lot of groups, Equation 3 assigns a high value to people who, on average, belong to large groups (e.g., whenever $a_{pg} = 1$ and $C_g^R(1)$ is a big number). In the same way, while Equation 2 assigns a high value to groups that have lots of members, Equation 2 assigns a high value to groups that, on average, have members who themselves have lots of memberships (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:

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

And for groups:

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

As we saw, while for people, Equation 3 measured the average size of the groups they join, Equation 5 assigns a high value to people who, on average, belong to groups who are themselves attended by highly active members (e.g., whenever $a_{pg} = 1$ and $C_g^R(2)$ is a big number). In the same way, while Equation 4 assigns a high value to groups whose members have lots of memberships, Equation 6 assigns a high value to groups that, on average, have members who themselves (also on average) belong to large groups (e.g., $a_{pg} = 1$ and $C_p^R(2)$ is a big number).

Note that for people, the even-numbered reflection $C_p^R(2)$ assigns scores 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 scores 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 scores 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 scores 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 principle, in practice, the ordering of the nodes in each mode ends up being identical across even and odd scores after their rank ordering "freezes" past a given number of iterations (proportional to the network size).[9]

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:

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

[9] This means that at least in the initial iterations, there is a formal resemblance between the reflection scores on the odd iterations and Bonacich's dual centrality idea, namely, that central people are those who belong to central groups, where the centrality of groups is defined by the centrality of people who belong to them (and vice versa). And, indeed, we find that the absolute value of the Pearson product-moment correlation between the odd reflection scores and the Bonacich eigenvector centrality is fairly high for persons ($r = 0.80$) in the third iteration but then settles to a lower value after the equilibrium is reached past the 20th iteration for the Southern Women data ($r = 0.54$). The even iterations for persons, on the other hand, begin with a relatively low correlation with the Bonacich centrality at the second iteration ($r = 0.30$) but then settle into the same equilibrium correlation with the Bonacich centrality as the odd iterations after the 20th reflection. I thank an anonymous reviewer for pointing out this pattern.

And for groups:

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

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

	E1 (6-27)	E2 (3-2)	E3 (4-12)	E4 (9-26)	E5 (2-25)	E6 (5-19)	E7 (3-15)	E8 (9-16)	E9 (4-8)	E10 (6-10)	E11 (2-23)	E12 (4-7)	E13 (11-21)	E14 (8-3)
W1 (Evelyn)	1	1	1	1	1	1	0	1	1	0	0	0	0	0
W2 (Laura)	1	1	1	0	1	1	1	1	0	0	0	0	0	0
W3 (Theresa)	0	1	1	1	1	1	1	1	1	0	0	0	0	0
W4 (Brenda)	1	0	1	1	1	1	1	1	0	0	0	0	0	0
W5 (Charlotte)	0	0	1	1	1	0	1	0	0	0	0	0	0	0
W6 (Frances)	0	0	1	0	1	1	0	1	0	0	0	0	0	0
W7 (Eleanor)	0	0	0	0	1	1	1	1	0	0	0	0	0	0
W8 (Pearl)	0	0	0	0	0	1	0	1	1	0	0	0	0	0
W9 (Ruth)	0	0	0	0	1	0	1	1	1	0	0	0	0	0
W10 (Verne)	0	0	0	0	0	0	1	1	1	0	0	1	0	0
W11 (Myra)	0	0	0	0	0	0	0	1	1	1	0	1	0	0
W12 (Katherine)	0	0	0	0	0	0	0	1	1	1	0	1	1	1
W13 (Sylvia)	0	0	0	0	0	0	1	1	1	1	0	1	1	1
W14 (Nora)	0	0	0	0	0	1	1	0	1	1	1	1	1	1
W15 (Helen)	0	0	0	0	0	0	1	1	0	1	1	1	0	0
W16 (Dorothy)	0	0	0	0	0	0	0	1	1	0	0	0	0	0
W17 (Olivia)	0	0	0	0	0	0	0	0	1	0	1	0	0	0
W18 (Flora)	0	0	0	0	0	0	0	0	1	0	1	0	0	0

Figure 1: Southern Women affiliation data matrix with rows and columns ordered as in the original DGG table.

3.1. Empirical Example

3.1.1. Southern Women Data

The following analyses use the classic "Southern Women" data (Davis, Gardner and Gardner, 1941) (DGG) as the running example. This data is doubly appropriate for showcasing the relevance of methods that exploit the duality between two sets of entities, as it was used in Breiger's classic paper (1974). Moreover, Southern Women is a dataset that, according to Freeman (2003, p.), "has become something of a touchstone for comparing analytic methods in social network analysis," having been analyzed "again and again," and reappearing "whenever any network analyst wants to explore the utility of some new tool for analyzing data." Freeman himself would on go to review twenty-one such attempts. The list has only gotten much longer in the last two decades since Freeman's methodological "meta-analysis" (e.g., Doreian, Batagelj and Ferligoj, 2004; Field, Frank, Schiller, Riegle-Crumb and Muller, 2006; Roffilli and Lomi, 2006; Wang, Sharpe, Robins and Pattison, 2009; Kovács, 2010; Brusco, 2011; Borgatti and Halgin, 2014; Everett and Borgatti, 2013; Lerner and Lomi, 2022; Batagelj, 2022; Lizardo, 2024, among many others). Despite the extensive list of studies on the Southern Women data, we will see that novel insights can still be gleaned from this venerable data source.

Despite its undisputed pedigree as a methodological touchstone in studies of duality and dual constitution in two-mode networks, a key source of confusion in the mini-literature surrounding the Southern Women data concerns the order of presentation of the rows and columns. The reason for this is that while the row mode (people) is "nominal" in both the scaling and the literal senses (being composed of women's names), the column mode (events) is more ambiguous. It *could* be treated as nominal; however, the original DGG table includes event *dates*, which means that it could also be treated as having a natural ordering, which could be reflected in the way we list the columns. As Freeman (2003) notes, in the canonical original table included in Davis et al. (1941), both the row and column modes had been re-ordered in a presentation to reveal what DGG ascertained (using intuitive methods and in-depth knowledge of the case) was the group structure among the women; this means that the events were *not* presented in chronological order from left to right across the columns.

To add to the confusion, the original DGG table added arbitrary ordered numerical designators to both people and events, such that people came to refer to "W1" or "W11" and "E1" or "E13" based on the original DGG ordering, which may not have mattered for the women, but it mattered for the events because "E14" in the original DGG data is not the last chronological even among the fourteen columns. In fact, as shown in Figure 1, it is the event that occurred on August 3rd, and three events occurred after that (E4, E8, and E13) in the DGG numbering.[10] To further add to the confusion, some early analyses of the data, including Breiger's (1974) classic article and Doreian (1979) used a *different* ordering of the women—taken from Homans (1950, 83)—from that of the original DGG table reproduced in Freeman (2003), and ordered the events from left to right in the affiliation matrix *chronologically*, which, as we have seen, also differs from the original DGG Table reproduced in Freeman (2003). This means that the arbitrary ordinal labeling of the rows (e.g., "W1," "W2," "W3," etc.) and the columns (e.g., "E1," "E2," "E3," etc.) in these (and other) publications—see, e.g., (Doreian, 1979, table 1)—does not correspond to the DGG ordering reproduced in Freeman (2003) and more recent publications (e.g., Borgatti and Halgin, 2014; Batagelj, 2022); in all, different versions of the dataset reproduce different versions of the labeling.

Of course, the specific labeling of the nodes does not really matter since we are usually interested in the structural properties of the rows and columns. Event ordering, however, matters for analyses that try to exploit the temporal properties of the data (e.g., Everett et al., 2018; Lerner and Lomi, 2022). Less substantively, but still importantly, the labeling of both the rows and columns does matter when trying to reconcile claims across publications from the narrative included in various papers, since "W1" or "E8" in one publication (e.g., Freeman, 2003) may not correspond to "W1" or "E8" in the other (e.g., Doreian, 1979). To avoid confusion, I use the numeric labeling corresponding to the original DGG data table, as reproduced in (among others) Freeman (2003, Figure 1) and Borgatti and Halgin (2014, Figure 28.1), and shown in Table 1. My presentation includes both the arbitrarily ordered labels (e.g., "W1" or "E1") and both the woman's name and the event date in parentheses, with "Myrna" (W11) rebaptized to her given name of "Myra" to correspond to the original DGG table (Freeman, 2003; Batagelj, 2022).

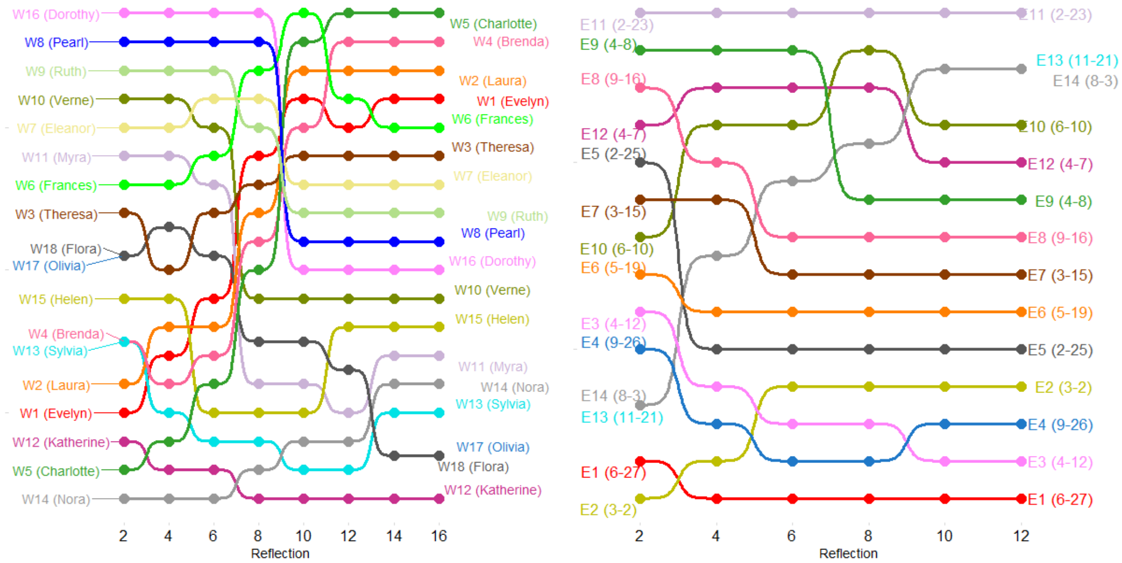
3.1.2. Reflective Score Trajectory

Figure 2a shows the trajectory of the HH reflective scores for persons, using a bump chart to plot the rank order trajectory of persons across reflections. 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 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).

In the exhaustive methodological meta-analysis mentioned earlier, Freeman (2003) noted that "the consensus of the analytic procedures is to assign women 1 through 9 to one group and women 10 through 18 to the other." Notably, the reflective score rankings after freezing ($q \geq 18$) recover this same dominant partition reading from top to bottom. Figure 2b shows the corresponding bump chart for the events. Clearly, $E14$ experiences the most dramatic improvement in status as we move to higher reflections. It ranks relatively low in terms of the average number of memberships among its members; however, its standing improves when considering the average of the average number of memberships

[10] Using the DGG labeling, the chronologically ordered set of events are {E11, E5, E2, E7, E12, E9, E3, E6, E10, E1, E14 E8, E4, E13} (Everett, Broccatelli, Borgatti and Koskinen, 2018, p. 68).

among its members, and so forth. Notably, the equilibrium reflective scores recover and the ordering of groups columns that re-appears across many articles that have analyzed these data with events 1-6 separated from events 10-14 by events 7-9 (see e.g., Doreian et al., 2004; Kovács, 2010; Lizardo, 2024; Everett and Borgatti, 2013).



(a) Reflective score trajectories of persons.

(b) Reflective score trajectories of groups.

Figure 2: Reflective scores for persons and groups (even reflections)

3.2. CA and Dual Centrality Scoring in Two-Mode Networks

The reader may ask what the point of going through all this reflective scoring is since Bonacich (1991) already developed a dual approach to scoring nodes in two-mode networks based on a very similar idea: Define the scores of entities in one mode based on the scores of entities in the other mode to which they are connected. As we have already noted, Bonacich also considered CA in passing as a possible technique to assign scores to nodes in a two-mode network but (perhaps rightly) dismissed it as producing scores that could be interpreted as a kind of "centrality," noting that it was better at capturing the clique structure of women who attended different events (1991, 164).

While using CA to capture patterns of similarity in attendance between persons and groups makes sense (although, as we will see, the exact patterns of similarity captured by CA are not reducible to "cliques" as traditionally understood), CA can also be used to assign meaningful scores to nodes in a social network, which in some applications may even be interpreted as a type of centrality (or at least have a specifiable mathematical relation to scores that are usually unambiguously interpreted as centrality measures). In fact, as I will show in a bit, in the two-mode case, the first CA dimension captures the limit ($q \to \infty$) of the reflective scores discussed in the previous section (Mealy, Farmer and Teytelboym, 2019). This makes the connection between the Bonacich and the CA scoring approaches more intimate than Bonacich thought.

To begin to see the linkages between CA and the Bonacich approach, recall that the solutions to the following system of linear equations give the Bonacich (C^B) two-mode centralities:

$$\mathbf{A}C^B_p = \lambda C^B_g \tag{9}$$

$$\mathbf{A}^T C^B_g = \lambda C^B_p \tag{10}$$

Where $\mathbf{A}$ is the original affiliation matrix. Equations 9 and 10 have the typical form of a linear eigensystem, which means the unknown C^B_g and C^B_p scores can be obtained from the row and column eigenvectors corresponding to the largest eigenvalue λ_1, after performing a singular value decomposition (SVD) of the rectangular affiliation matrix $\mathbf{A}$.

Importantly, as Faust (1997, 170) notes, it is possible to express Equations 9 and 10 in terms of person and group-specific scores; for people, these are given by:

$$C_p^B = \frac{1}{\lambda_1} \sum_g a_{pg} C_g^B \tag{11}$$

And for groups:

$$C_g^B = \frac{1}{\lambda_1} \sum_p a_{pg} C_p^B \tag{12}$$

Where λ_1 is the largest (first) eigenvalue corresponding to the eigenvector containing the C_p^B scores. These equations make it clear that in the Bonacich dual-scoring approach, the scores assigned to people are proportional to the sum of the scores of the groups to which they belong, and the scores assigned to groups are proportional to the sum of the scores simultaneously assigned to their members. This approach thus nicely captures the duality property, as the scores of nodes in each mode are determined by aggregating their connections to nodes in the other mode.

Note, however, the formal similarity between equations 11 and 12 and equations 7 and 8. The main difference is that in HH's reflective equations, a person's centrality is proportional to the *group-degree-weighted* centralities of the groups they join, and a group's centrality is proportional to the *person-degree-weighted* centrality of the people who are members. We will return to this crucial point later.

3.3. Bonacich Reflections

We can motivate the Bonacich eigenvector-based approach using the same "reflective" exposition we used to introduce the HH reflective approach in Section 3. Admittedly, this is a somewhat unorthodox way of presenting the eigenvector centralities for two-mode networks; 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:

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

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

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:

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

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

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:

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

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

This says that the Bonacich reflection 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, these sums of sums would diverge to a larger and larger quantity at each step q. To prevent this and guarantee convergence, we can 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:

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

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

Where the denominator in 19 and 20 is the Euclidean vector norm.[11] The normalization will prevent divergence of the sum of centralities for persons and groups, formalizing the weaker assumption of *proportionality* between the 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 of 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.[12] 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" scores and the Bonacich two-mode eigenvector centralities. In fact, both can be seen as instantiating an underlying "prismatic" model—in the sense of Podolny (2001)—of how centrality is distributed in the two-mode network. The Bonacich approach, based on sums of sums, favors overall *volume*—or two-mode "total effects" in the sense of Friedkin (1991)—thus, groups receive more centrality points from their more central members and fewer centrality points from their less central members. Likewise, people receive more centrality points from their more central memberships and fewer centrality points from their less central memberships, and so on through every iteration. The HH reflective approach (and CA by implication) favors *averages* rather than volume; thus, a person with a few memberships can contribute as much centrality to the group they belong to as a person with many memberships if the person with fewer memberships belongs to (on average) more central groups and the person with more memberships belongs to (on average) less central ones. In the same way, a person can receive as much centrality from a small group they belong to compared to a large group, as long as the small group (on average) has more central members in it, and the large group (on average) has less central members.

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

[12] Note 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.

3.4. HH Reflections as Correspondence Analysis

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 by Mealy et al. (2019) and van Dam, Dekker, Morales-Castilla, Rodríguez, Wichmann and Baudena (2021), the Hidalgo and Hausmann's (2009) reflective scores 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:

$$\left(\mathbf{A}\mathbf{A}^T\right) C_p^B = \lambda^2 C_p^B \tag{21}$$

$$\left(\mathbf{A}^T\mathbf{A}\right) C_g^B = \lambda^2 C_g^B \tag{22}$$

This shows that the dual centrality Bonacich scores for persons and groups are equivalent to the eigenvectors of the respective one-mode "Breiger" (1974) projection matrices corresponding to the first (largest) eigenvalue (which is equivalent to the singular value λ from equations 9 and 10 squared). This linkage qualifies the Bonacich scores as a "dual projection" method for analyzing two-mode networks in the sense of Everett and Borgatti (2013).

We can motivate CA using the same dual projection strategy. Typically, when CA is presented as the solution to an eigenvalue problem, this is done by showing how the relevant scores can be obtained via the singular value decomposition (SVD) of the rectangular affiliation matrix (Borgatti and Everett, 1997; Bonacich, 1991; Faust, 1997, 2005). This gives the impression that CA is exclusively a "direct" rather than a dual projection method. Here, we will show that CA can also be motivated, and scores obtained via a similar "dual" projection approach just like the Bonacich eigenvector scores, except that these are scores associated with a *degree-weighted* projection of the original affiliation matrix.

To see this, consider the $|P| \times |P|$ matrix $\mathbf{D}p$, containing the "first order" reflective scores 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 $\mathbf{D}g$, which contains the "first order" (degree) centralities of each group $C^R(1)_g$ (size) along the diagonals and zeroes in every other cell. Using these matrices, we can compute *row stochastic* versions of the affiliation matrix and its transpose.[13] For the people, we can do this by taking the original affiliation matrix and pre-multiplying it by $\mathbf{D}p^{-1}$ (which now contains the *inverse* of the first-order centralities of each person $C_p^R(1)$ along the diagonals), yielding the row-stochastic matrix $\mathbf{P}_{pg}$ of dimensions $|P| \times |G|$:

$$\mathbf{P}_{pg} = \mathbf{D}p^{-1}\mathbf{A} \tag{23}$$

We can do the same with the groups, except that we pre-multiply the *transpose* of the original affiliation matrix by $\mathbf{D}g^{-1}$ (which now contains the *inverse* of the first-order centralities of each group $C_g^R(1)$ along the diagonals), yielding the row-stochastic matrix $\mathbf{P}_{gp}$ of dimensions $|G| \times |P|$:

$$\mathbf{P}_{gp} = \mathbf{D}_G^{-1}\mathbf{A}^T \tag{24}$$

Both $\mathbf{P}_{pg}$ and $\mathbf{P}_{gp}$ are just centrality-weighted versions of the original affiliation matrix and its transpose; accordingly, nothing is stopping us from proceeding with the usual next step—due to Breiger (1974)—of calculating the degree-weighted *projections* for people and groups from these matrices.

For the people, the degree-weighted projection is:

$$\mathbf{P}_{pp} = \mathbf{P}_{pg}\mathbf{P}_{gp} \tag{25}$$

[13] Recall that a matrix is row stochastic if its rows sum to one

	W1	W2	W3	W4	W5	W6	W7	W8	W9	W10	W11	W12	W13	W14	W15	W16	W17	W18
W1 (Evelyn)	0.19	0.16	0.14	0.15	0.14	0.12	0.08	0.09	0.07	0.04	0.04	0.03	0.02	0.03	0.01	0.08	0.04	0.04
W2 (Laura)	0.14	0.18	0.12	0.13	0.1	0.12	0.11	0.07	0.07	0.04	0.02	0.01	0.02	0.03	0.03	0.04	0	0
W3 (Theresa)	0.14	0.13	0.16	0.12	0.16	0.12	0.11	0.09	0.09	0.06	0.04	0.03	0.04	0.04	0.03	0.08	0.04	0.04
W4 (Brenda)	0.13	0.13	0.1	0.17	0.16	0.12	0.11	0.07	0.07	0.04	0.02	0.01	0.02	0.03	0.03	0.04	0	0
W5 (Charlotte)	0.07	0.06	0.08	0.09	0.16	0.07	0.06	0	0.06	0.03	0	0	0.01	0.01	0.02	0	0	0
W6 (Frances)	0.06	0.07	0.06	0.07	0.07	0.12	0.08	0.07	0.05	0.02	0.02	0.01	0.01	0.02	0.01	0.04	0	0
W7 (Eleanor)	0.04	0.06	0.05	0.06	0.06	0.08	0.11	0.07	0.07	0.04	0.02	0.01	0.02	0.03	0.03	0.04	0	0
W8 (Pearl)	0.03	0.03	0.03	0.03	0	0.05	0.05	0.09	0.04	0.04	0.04	0.03	0.02	0.03	0.01	0.08	0.04	0.04
W9 (Ruth)	0.03	0.04	0.05	0.04	0.06	0.05	0.07	0.05	0.09	0.06	0.04	0.03	0.04	0.02	0.03	0.08	0.04	0.04
W10 (Verne)	0.02	0.02	0.03	0.02	0.03	0.02	0.04	0.05	0.06	0.11	0.08	0.05	0.06	0.04	0.07	0.08	0.04	0.04
W11 (Myra)	0.02	0.01	0.02	0.01	0	0.02	0.02	0.05	0.04	0.08	0.13	0.09	0.07	0.06	0.09	0.08	0.04	0.04
W12 (Katherine)	0.02	0.01	0.02	0.01	0	0.02	0.02	0.05	0.04	0.08	0.13	0.2	0.17	0.14	0.09	0.08	0.04	0.04
W13 (Sylvia)	0.02	0.02	0.03	0.02	0.03	0.02	0.04	0.05	0.06	0.11	0.13	0.2	0.18	0.15	0.11	0.08	0.04	0.04
W14 (Nora)	0.03	0.03	0.04	0.03	0.03	0.03	0.06	0.07	0.05	0.09	0.11	0.19	0.17	0.2	0.14	0.04	0.17	0.17
W15 (Helen)	0.01	0.02	0.02	0.02	0.03	0.02	0.04	0.02	0.04	0.08	0.11	0.07	0.08	0.09	0.16	0.04	0.12	0.12
W16 (Dorothy)	0.02	0.01	0.02	0.01	0	0.02	0.02	0.05	0.04	0.04	0.04	0.03	0.02	0.01	0.01	0.08	0.04	0.04
W17 (Olivia)	0.01	0	0.01	0	0	0	0	0.03	0.02	0.02	0.02	0.01	0.01	0.04	0.05	0.04	0.17	0.17
W18 (Flora)	0.01	0	0.01	0	0	0	0	0.03	0.02	0.02	0.02	0.01	0.01	0.04	0.05	0.04	0.17	0.17

(a) People.

	E1	E2	E3	E4	E5	E6	E7	E8	E9	E10	E11	E12	E13	E14
E1 (6-27)	0.14	0.09	0.07	0.07	0.05	0.05	0.03	0.03	0.01	0	0	0	0	0
E2 (3-2)	0.09	0.13	0.07	0.06	0.05	0.05	0.03	0.03	0.02	0	0	0	0	0
E3 (4-12)	0.14	0.13	0.17	0.16	0.13	0.1	0.07	0.06	0.02	0	0	0	0	0
E4 (9-26)	0.09	0.08	0.11	0.16	0.08	0.05	0.05	0.03	0.02	0	0	0	0	0
E5 (2-25)	0.14	0.13	0.17	0.16	0.19	0.13	0.12	0.09	0.04	0	0	0	0	0
E6 (5-19)	0.14	0.13	0.13	0.1	0.13	0.19	0.08	0.1	0.06	0.03	0.03	0.02	0.04	0.04
E7 (3-15)	0.1	0.09	0.11	0.13	0.15	0.1	0.19	0.11	0.07	0.09	0.08	0.12	0.09	0.09
E8 (9-16)	0.14	0.13	0.13	0.1	0.16	0.17	0.15	0.22	0.18	0.15	0.05	0.17	0.1	0.1
E9 (4-8)	0.04	0.08	0.04	0.06	0.06	0.09	0.09	0.15	0.27	0.14	0.28	0.16	0.14	0.14
E10 (6-10)	0	0	0	0	0	0.02	0.05	0.05	0.06	0.18	0.08	0.15	0.14	0.14
E11 (2-23)	0	0	0	0	0	0.02	0.03	0.01	0.09	0.06	0.33	0.05	0.04	0.04
E12 (4-7)	0	0	0	0	0	0.02	0.07	0.07	0.08	0.18	0.08	0.19	0.14	0.14
E13 (11-21)	0	0	0	0	0	0.02	0.03	0.02	0.04	0.09	0.03	0.07	0.14	0.14
E14 (8-3)	0	0	0	0	0	0.02	0.03	0.02	0.04	0.09	0.03	0.07	0.14	0.14

(b) Groups.

Figure 3: Degree-weighted projection matrices of the Southern Women data.

And for the groups:

$$\mathbf{P}_{gg} = \mathbf{P}_{gp}\mathbf{P}_{pg} \tag{26}$$

Both $\mathbf{P}_{pp}$ and $\mathbf{P}_{gg}$ are square ($|P| \times |P|$ and $|G| \times |G|$) respectively) degree-normalized projection matrices linking people and groups. Figure 3 shows the $\mathbf{P}_{pp}$ and $\mathbf{P}_{gg}$ matrices for the Southern Women Data. As we can see, like their constituent matrices, both $\mathbf{P}_{pp}$ and $\mathbf{P}_{gg}$ are row-stochastic (rows constrained to sum to 1.0), positive definite, but not symmetric, as can be seen in Figure 3. Interestingly, one way to interpret the entries in the degree-weighted projection matrices for people and groups is by giving the *probabilities* that a random walker starting at the row person (group) and, following any Markovian sequence of $person - group - person' - group'$ hops, will reach the column person (group), and where the probability of jumping from any person to a group they belong is given by the inverse of the number of groups they belong to (and similarly for groups) (Deng, Lyu and King, 2009, 240). Thus, higher values indicate an affinity or proximity between the people based on indirect connections in the two-mode network (and the same for groups in the corresponding matrix). The reason for the asymmetry is that, depending on their pattern of connections to other groups, person i may have more ways of indirectly reaching person j than vice versa (unless i and j are connected to the same groups).

It can be shown (van Dam et al., 2021), that in the limit ($q \rightarrow \infty$), the iterative HH reflective scores can be obtained as a solution of the eigensystem:

$$\mathbf{P}_{pp} C_p^R = \lambda_2 C_p^R \tag{27}$$

$$\mathbf{P}_{gg} C_g^R = \lambda_2 C_g^R \tag{28}$$

With the C^R scores for persons and groups obtained from the eigenvectors corresponding to the *second* largest eigenvalue of the respective degree-weighted projection matrices, which are the same as the CA scores and the limit HH reflective scores (van Dam et al., 2021).[14] These scores will also 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 via SVD (Fouss, Saerens and Shimbo, 2016, 398, eq. 9.17). This exact correspondence can be verified in Figure 4, which shows a scatterplot and associated linear regression line of the standardized scores computed via the iterative method of reflections ($q = 26$) on the y-axis against those obtained from the second eigenvector of the degree-weighted projection matrices for persons and groups on the x-axis ($r = 1.0$).[15]

Because $\mathbf{P}_{pp} = \mathbf{D}p^{-1}\mathbf{A}\mathbf{D}g^{-1}\mathbf{A}^T$ and $\mathbf{P}_{gg} = \mathbf{D}g^{-1}\mathbf{A}^T\mathbf{D}p^{-1}\mathbf{A}$ there is a formal equivalence in the duality relations implied by equations 27 and 28 and that implied by the Bonacich two-mode centralities in equations 21 and 22. Both extract individual and group scores as eigenvectors of the one-mode projection of the original affiliation matrix, $\mathbf{A}\mathbf{A}^T$ and $\mathbf{P}_{pp}$ for people and $\mathbf{A}^T\mathbf{A}$ and $\mathbf{P}_{gg}$ for groups. The key difference is that in CA, we 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.

This can be clearly seen if we express $\mathbf{P}_{PP} = \mathbf{D}g^{-1}\mathbf{A}^T\mathbf{D}p^{-1}\mathbf{A}$ in terms of each cell entry (Mealy et al., 2019, eq. 4):

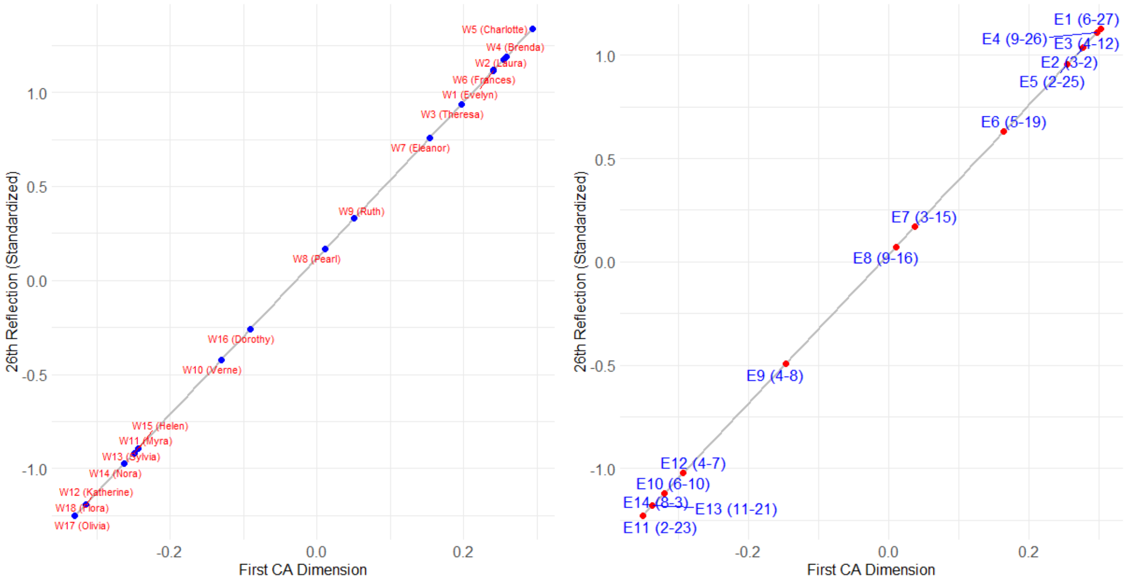
$$p_{pp'} = \sum_g \frac{a_{pg}a_{p'g}}{C_p^R(1)C_g^R(1)} = \frac{1}{C_p^R}\sum_g \frac{a_{pg}a_{p'g}}{C_g^R(1)} \tag{29}$$

In equation 29, 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 HH reflective scores, 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 the matrix $\mathbf{P}_{gg} = \mathbf{D}g^{-1}\mathbf{A}^T\mathbf{D}p^{-1}\mathbf{A}$, whose entries are given by:

[14] As we have seen, both equations 25 and 26 produce row stochastic matrices; this means that their largest (dominant) eigenvalue will also be 1.0, which will be associated with a trivial first eigenvector containing constant values for both the people and the groups.

[15] Formally, the eigendecomposition of the degree-weighted projection (for people) is given by the solution to the equation $\mathbf{P}_{pp} = \mathbf{U}\mathbf{V}\mathbf{U}^{-1}$, where $\mathbf{U}$ is a $|P| \times |P|$ matrix whose columns are the eigenvectors of $\mathbf{P}_{pp}$ and $\mathbf{V}$ is a $|P| \times |P|$ matrix whose diagonals are the eigenvalues of $\mathbf{P}_{pp}$ (with zeros in every other cell). Columns two through $|G| - 1$ of matrix $\mathbf{U}$ contain the standard CA scores for the people along the complete set of $|G| - 1$ dimensions; the same goes for $\mathbf{P}_{gg}$ in the case of groups. Note that both $\mathbf{P}_{pp}$ and $\mathbf{P}_{gg}$ have the same set of non-zero eigenvalues, with cardinality equal to the rank of the original affiliation matrix $\mathbf{A}$ minus one or $|G| - 1$; thus only the first $|G| - 2$ eigenvectors of both $\mathbf{P}_{pp}$ and $\mathbf{P}_{gg}$ are informative in the Southern Women data, since $|G| < |P|$.



(a) (b)

Figure 4: Scatter plot of the scores computed by the method of reflections ($q = 26$) on the y-axis and the first dimension of the Correspondence Analysis of the two-mode affiliation matrix on the x-axis, for both persons (a) and groups (b).

$$p_{gg'} = \sum_p \frac{a_{pg}a_{pg'}}{C_p^R(1)C_g^R(1)} = \frac{1}{C_g^R} \sum_p \frac{a_{pg}a_{pg'}}{C_p^R(1)} \tag{30}$$

Now, as we saw earlier (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.[16] In fact, as Mealy et al. (2019) have noted, the method of "reflections" is just a re-discovery of the older idea of computing CA scores via reciprocal averaging (Hill, 1973).

As D'Esposito et al. (2014a) have noted, degree-pre-weighting does alter the original input data, which means that CA is *not* meant to reproduce a low dimensional approximation of the original affiliation matrix. 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 29 and 30 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, thus adjusting similarity by group size so that co-memberships in groups everyone belongs to counts for less (Ragozini et al., 2014). In the same way, on the group side, shared members who do not have many affiliations count more in determining intergroup similarity than those with many affiliations.

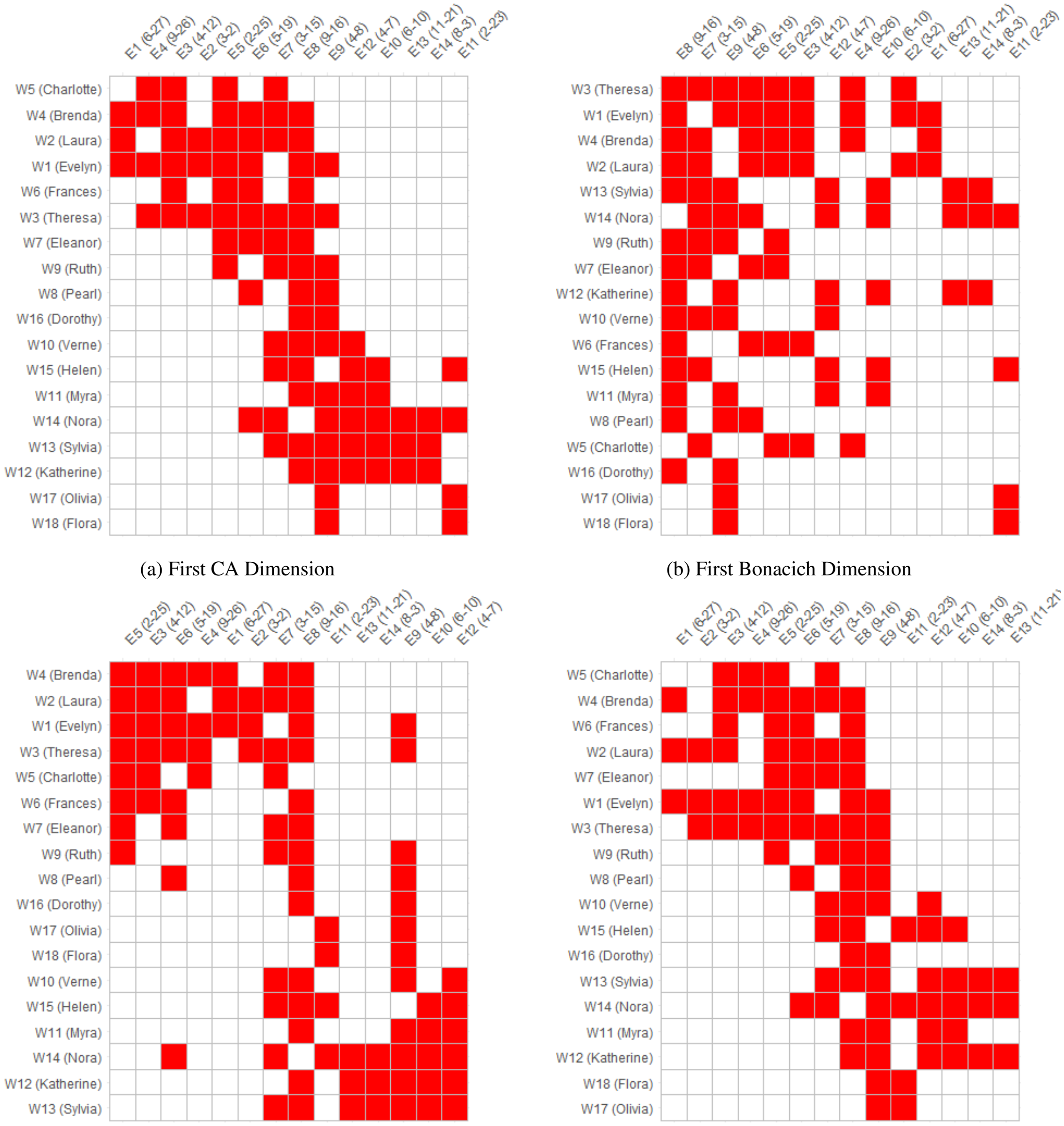
4. Comparing Bonacich and CA Dual Scoring in the Southern Women Data

4.1. Row and Column Re-ordered Affiliation Matrices

Figures 5a and 5b illustrate the key differences between the two versions of dual two-mode scoring C^R and C^B. Each panel shows the original Southern Women affiliation matrix but with rows and columns blocked according to the first CA dimension in (a) and the Bonacich (1991) centrality in (b).[17] In the plot, a cell entry is colored red if it

[16] More 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); accordingly, the first non-trivial CA dimension for persons and groups given by the right and left eigenvectors corresponding to the second-largest eigenvalue from the SVD of the matrix $\mathbf{D}_p^{-1/2}\mathbf{A}\mathbf{D}_g^{-1/2}$ (Faust, 2005, 126).

[17] This, in effect, a two-mode version of constructing "a posteriori" (exploratory or data-driven) blockmodels, in the sense of Wasserman and Anderson (1987), using scores from CA and Bonacich scoring to re-order the original matrix, as they did in that paper for one-mode networks.



(a) First CA Dimension

(b) First Bonacich Dimension

(c) Second Bonacich Dimension

(d) Main SimRank Eigenvector

Figure 5: Southern Women Affiliation matrix with re-ordered rows and columns: (a) rows and columns re-ordered according to the person and group score on the first CA dimension, (b) rows and columns re-ordered according to person and group Eigenvector centrality score, (c) rows and columns re-ordered according to person and group score on the second Bonacich dimension, (d) rows and columns re-ordered according to the second eigenvector of the SimRank similarity matrix.

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 (Wasserman et al., 1990, 34), and the C^B re-ordered affiliation matrix having a core-periphery structure.

Accordingly, the C^R reflective scores reveal a dual-block partition between groups of women who selectively attend two clusters of events, located at 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 half) who selectively co-participate in highly attended events (on the left-hand side) and less active women (on the bottom half) who go to less well-attended events (on the right-hand side). In fact, as Everett and Borgatti (2013, p. 206) have noted, computing the eigenvector scores for both sets of nodes in the mode network from the classic Breiger (1974) projections—as in Bonacich (1991)—and using them to re-order the rows and columns of the original affiliation matrix is a way to approximate using continuous scores an ideal discrete core-periphery partitioning of the two-mode data matrix, cheaper and more efficiently compared to discrete optimization methods. Accordingly, the two-block core-periphery partitioning of people and groups suggested by Figure 5b—separating the first eight women starting from the top from the rest and the first five events starting from the left from the rest) is substantively identical to that shown in Everett and Borgatti (2013, 206, Table 2), who previously extracted a core-periphery partitioning of the same data using a similar approach (we will discuss Figures 5c and 5d in subsequent sections).

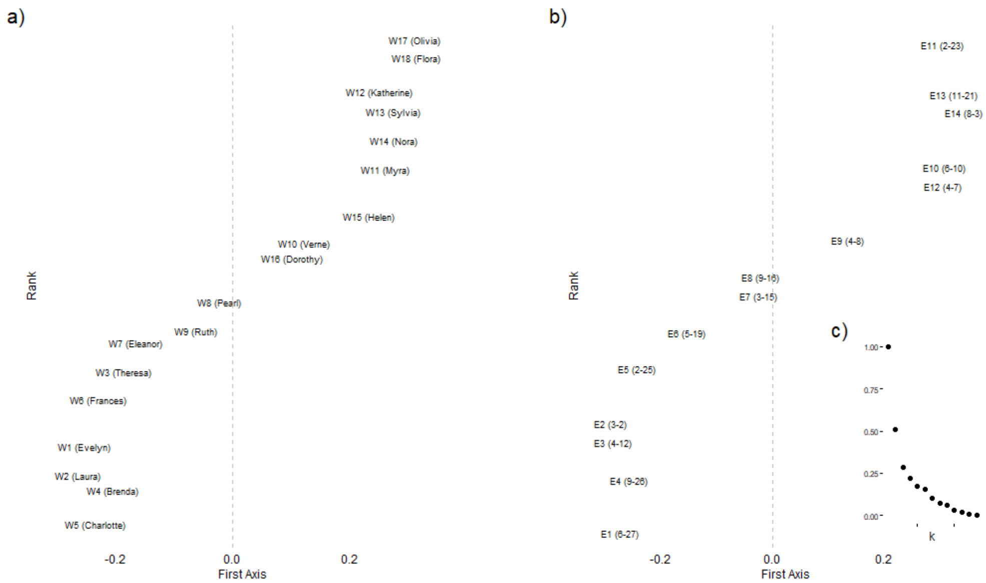


Figure 6: CA Eigenvector plot for persons (a) and groups (b). The first dimension CA score is on the x-axis, and the rank order of persons and groups on the same score is on the y-axis. The inset plot (c) shows the eigenvalue orderings on the x-axis (k) for persons and groups, showing a separation between the first three eigenvalues and the rest.

4.2. Eigenvector Plot

We can confirm that the first CA axis points toward the underlying similarity-based partitioning of the two-mode network by looking at Figure 6(a) and 6(c), which shows the $\boldsymbol{C}^R$ 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 block structure, the first dimension CA scores would be distributed as a continuous logistic curve; when block structure is present, we should observe discernible breaks in this distribution (van Dam et al., 2021). The Southern Women Data clearly belong in the second category. In the people mode, we have {Frances, Laura, Brenda, Charlotte, Evelyn, Theresa, Elanor, 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 the canonical consensus partitioning (for the people) highlighted by Freeman (2003) across twenty-one different methods and subsequently reproduced (for both persons and groups) by Kovács (2010) and Lizardo (2024) using a generalized relational similarity (GRS) 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; we will return to this point in Section 5.

4.3. Degree-Weighted Projections and CA Correspondence Plot

Panels (a) and (b) of Figure 7 shows a heatmap plot of the same weighted projection matrices from Figures 3a and 3b, but this time with the rows and columns re-ordered according to the scores corresponding to the first CA

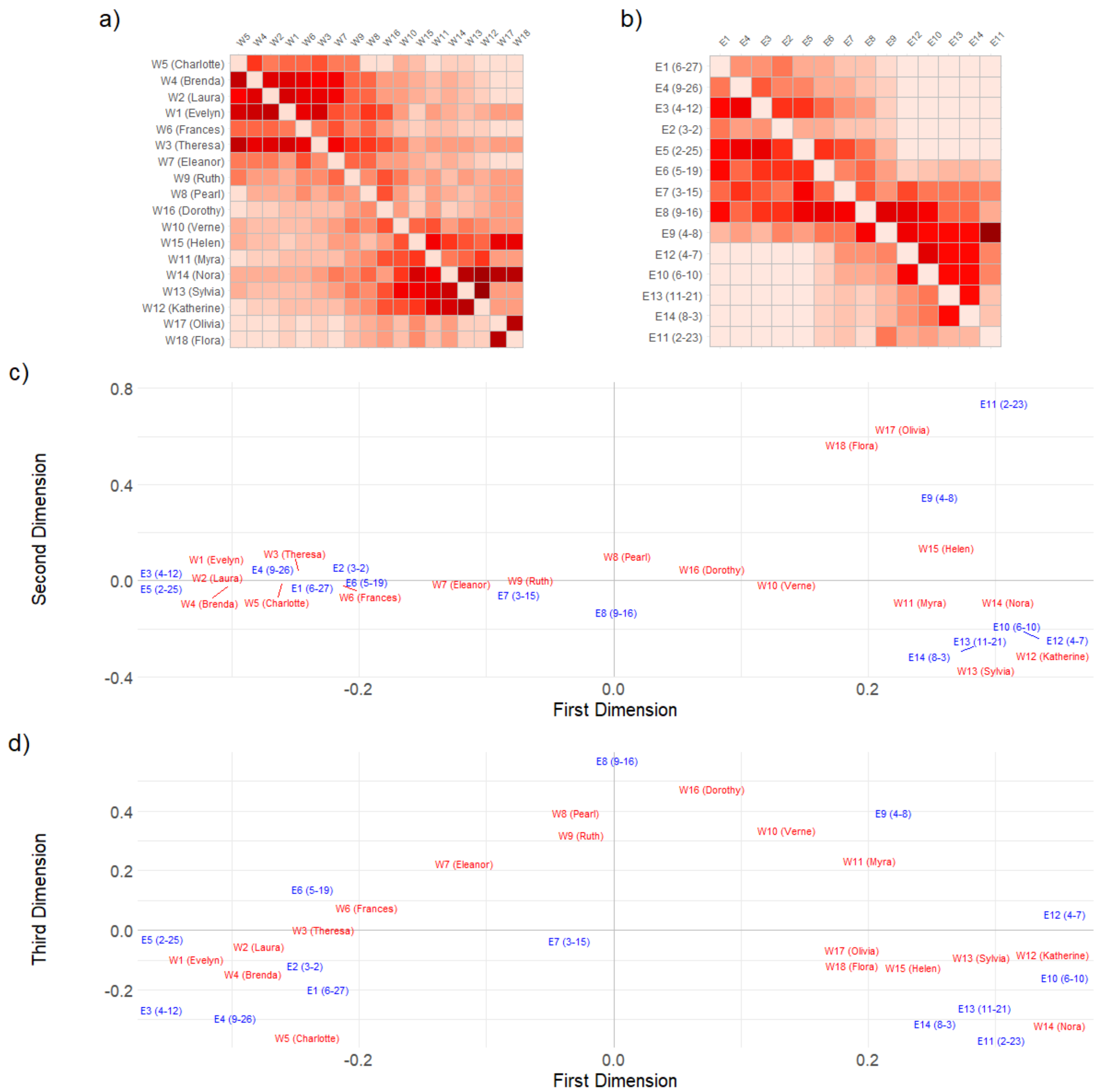


Figure 7: The top panels show the degree-weighted one-mode projection heatmap plots for (a) persons and (b) groups in the Southern Women Data with rows and columns re-ordered according to the first CA dimension. Darker red cells correspond to high entry values (normalized by the maximum) in the $\mathbf{P}_{pp}$ and $\mathbf{P}_{gg}$ matrices. The bottom panels show the correspondence plot for persons and groups with scores corresponding to (c) the first dimension on the horizontal axis with the second dimension on the vertical axis, and (d) the first dimension on the horizontal axis with the third dimension on the vertical axis.

dimension. In the plot, darker red indicates greater similarity between pairs of persons and groups, and gray indicates less similarity. The bottom panels of the same figure show the usual correspondence plot of the first three CA dimensions, with the first dimension on the x-axis against the second dimension on the y-axis in panel (c) and the first dimension plotted against the third dimension on the y-axis in panel (d).

This figure illustrates several important points. First, as the heatmap plots in panels (a) and (b) show, distances along the first CA axis correspond to degree-weighted similarities between persons and groups recorded in the $\mathbf{P}_{pp}$ and

$\mathbf{P}_{gg}$ matrices. Pairs of people and groups with the largest entries in the matrix have minimal differences in the first CA dimension score, appearing close to one another when we reshuffle the rows and columns of the weighted projection matrix. This also means that the distances between pairs of people and groups 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 respective *degree-normalized one-mode projections*. Pairs of people with large values in the projection matrix P_{pp}—such as {Olivia, Flora}, {Katherine, Sylvia, Nora}, {Dorothy, Verne}, and {Brenda, Laura, Evelyn}— appear closer in the correspondence plot. The same goes for pairs of groups; those with large values in P_{gg} matrix—such as {E13, E14} and {E1, E2, E3, E4}—appear closer in the correspondence plot, while those with small weighted similarity values are placed far apart. Distances between same-mode entities in the correspondence plot are thus a function of their (inverse) degree-normalized similarity.

Note that this differs from the usual interpretation of the CA correspondence plot, which is typically 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), a criterion closer to structural equivalence (D'Esposito et al., 2014a), or "first order" similarity (Kovács, 2010). This interpretation implies (for instance) that two people who attend many of the same events or groups with many members in common 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. This type of degree-normalization can be the preferred approach in some applications since people with many memberships and groups with many members will appear to be similar to many other people and groups if we go with the raw number of other-mode objects shared.[18] This implies that the similarities between pairs of people and groups will be weighted by the degrees of the other-mode objects that are shared between them. 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 that share people with few memberships will be closer in the diagram than those sharing people with many other memberships (D'Esposito et al., 2014a).

Another thing to note is that since the CA row and column scores in each dimension are obtained from the eigendecomposition and subsequent low-rank approximation of the degree-normalized similarity matrices $\mathbf{P}_{pp}$ and $\mathbf{P}_{gg}$, each dimension provides independent information on the similarity clustering of persons and groups.[19] Thus, as shown in panels (c) and (d) of Figure 7, as we have already seen, the first horizontal dimension recovers the main partition separating the two dominant clusters of persons and groups (Freeman, 2003). The second and third dimensions, by contrast, recover secondary and tertiary similarity clusters of people and groups independent of the main two-block partition. Thus, the second dimension (the y-axis in panel (c) of Figure 7) separates {Flora, Nora} and {E9, E11} from the rest of persons and groups, consistent with the strong similarity between these nodes in the original weighted projection matrix. We know from the original affiliation matrix in Figure 1 that {Flora, Nora} are maximally similar—they are structurally equivalent—and *only* attend events {E9, E11}, which makes these events similar in the second dimension given they are attended by the two people who are also most similar in the same dimension (this also accounts for the high similarity between the two events shown in the heatmap plot of the degree-weighted projection in Figure 7(b)).

The third CA dimension, shown in the y-axis in panel (d) of Figure 7 against the first dimension separation on the x-axis, by contrast, separates {Eleanor, Pearl, Ruth, Dorothy, Verne, Myra} and {E8, E9} from the rest. We know from Figure 5b that event E8, along with E7, is one of the most popularly attended events and thus contributes little to differentiate between the main partition of people and events on the first correspondence plot in panel (c), hence its location near the plot's origin. The pattern of attendance of these six women is one characterized by affiliation with these popular events, combined with sparser attendance at the events that most serve to differentiate the two primary clusters. In his original paper, Bonacich (1991, 164) noted that this third dimension "corresponded closely to the centrality analysis pattern," an observation likely driven by the affinity between the actors who load highly on this dimension and two of the most central events by raw popularity, {E8, E9}. What we can see instead is that the third CA

[18] Note 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.

[19] We know from inset plot (c) of Figure 6 that the eigenvalues corresponding to the first three dimensions exhibit strong separation from the rest, suggesting that these provide a good low-rank approximation of the original degree-weighted projection matrix for persons and groups.

dimension separates actors that are most "peripheral" to the two main clusters according to Freeman (2003); namely, {Eleanor, Pearl Ruth} on one side and {Dorothy, Verne, Myra} on the other, from core members of the two main blocks, who are placed in the bottom of the correspondence plot in panel (c)—low scores in the third dimension—and toward the extreme left and right ends of the plot, indicating very high absolute value scores in the first dimension.

4.4. Breiger Projections and the Bonacich Correspondence Plot

Earlier, we discussed how CA is sometimes considered a technique useful to generate a correspondence 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 **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 (D'Esposito et al., 2014a). In fact, CA is better thought of as a low-dimensional approximation of the degree-weighted projection of the affiliation matrix, as given by equations 27 and 28, which represents a degree-adjusted proximity matrix—or a Markov transition matrix under a different interpretation (Mealy et al., 2019, 2)—between pairs of people and pairs of groups after accounting for their respective activity and sizes, respectively.[20]

Here, once again, the contrast with the traditional dual Bonacich scores can prove instructive. Figures 8(a) and 8(b) shows the "raw" (unweighted by degree) similarity scores for persons ($a_{pp'} = \sum_g a_{pg} a_{p'g}$) and groups ($a_{gg'} = \sum_p a_{pg} a_{pg'}$)—the standard Breiger (1974) projections—with the rows and columns re-ordered according by the first eigenvector of the matrix, which is equivalent to the usual Bonacich eigenvector dual centrality score for two-mode networks (see equations 21 and 22). Both similarity matrices reproduce the core-periphery structure we observed in the re-ordered affiliation matrix in Figure 5(b). Figure 8(c) plots $\boldsymbol{C^B}$ on the x-axis against the second eigenvector of the unweighted similarity matrix.

This suggests that the eigenvector centralities do encode a kind of similarity between pairs of people and pairs of groups; it just happens to be based on the unweighted similarity scores given by the raw number of shared memberships (for people) and shared members (for groups). In this respect, the dual centrality eigenvector score mixes an ordinal rank criterion with a similarity criterion since people with lots of memberships will end up being "similar" to more people on the standard projection matrix by the simple fact of having more memberships (and the same for large groups with respect to members). Persons and groups high in eigenvector centrality can be thought of as being the most similar to most others, condensing the information in the Breiger projection matrix into a single ordinal rank. Also, just like with CA and the degree-weighted projections, higher-order axes of the Bonacich centrality eigendecomposition of the Breiger projection matrices should encode forms of similarity between pairs of people and pairs of groups, net of the main axis dominated by overall person activity and group size.

While plotting the Bonacich centrality against higher eigenvectors is not necessarily a substantively unmotivated practice (see Iacobucci, McBride and Popovich, 2017, for discussion in the one-mode network case), the exact meaning of the mysterious second eigenvector remains unclear in the two-mode case. We can use the same "a posteriori blockmodeling" approach we used before to contrast CA and the Bonacich centrality analysis to shed light on this issue (Wasserman and Anderson, 1987). Figure 5c shows the affiliation matrix blocked according to the *second* Bonacich dimension, and Figure 8 (c) shows the "Bonacich Correspondence Plot" with the usual eigenvector centrality score on the horizontal x-axis plotted against the second eigenvector of the Breiger projection matrix for both persons and groups on the vertical y-axis.

We can see from Figure 5c that the second Bonacich dimension displays a *dual core-periphery* structure, revealing two secondary cores and peripheries not evident in the blocking according to either the first CA or standard eigenvector scores. Interestingly, this partition is almost equivalent to that obtained in Everett and Borgatti (2013, Table 5) by clustering a version of the projection matrices of the Southern Women data where the proximities were designed to detect *structural equivalence*.[21] In Everett and Borgatti's (2013, 208) clustering solution, "[t]he women are then split into (a) those that attend four or fewer events and consequently attend very few of the peripheral events (Pearl, Ruth, Verne, Myrna [sic], Dorothy, Olivia, Flora), and (b) those that attend core events and the first group of peripheral events

[20]In fact, as noted by Martin and Porter (2012, 838), because the transformed version of the affiliation matrix is factorized using the Singular Value Decomposition (SVD), CA is guaranteed to provide the *best* low dimensional approximation, in the sense of least squares, otherwise known as the Eckart-Young Theorem.

[21]In that application, the cell entry in each projection matrix was the *proportion of matches* between each pair of people and groups. As Everett and Borgatti (2013, 208) define it, "[p]roportion matching is the number of times two profiles have the same entry (either zeros or ones) divided by the length of the profile."

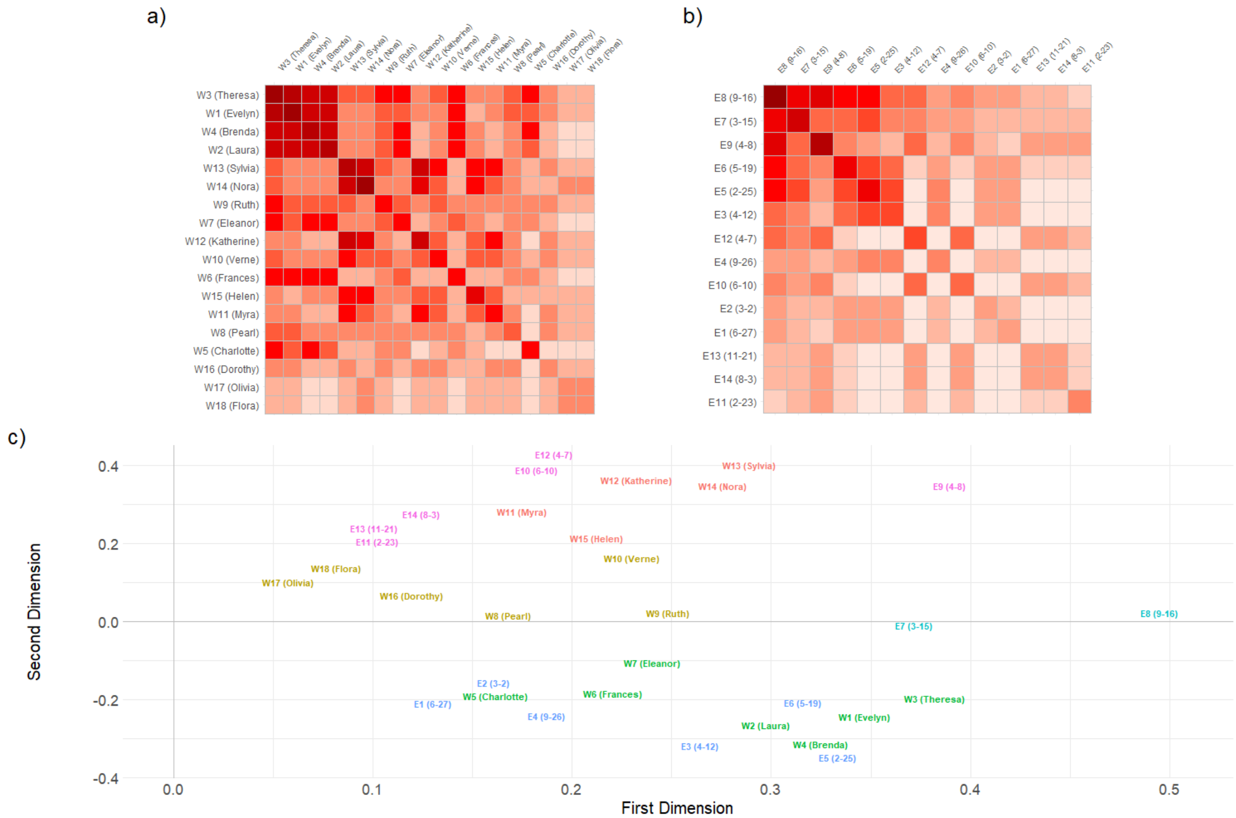


Figure 8: Bonacich dual centrality similarity plots for persons (a) and groups (b) on the left and right top panel. The bottom panel (c) shows the Bonacich eigenvector centrality correspondence plot for persons and groups with scores corresponding to the standard Bonacich eigenvector centrality on the horizontal axis and scores corresponding to the second largest eigenvector of the projection matrices on the vertical axis, points are colored according to a k-means three-cluster solution.

(Evelyn, Laura, Theresa, Brenda, Charlotte, Francis, Eleanor), and (c) those that attend core events and the second group of peripheral events (Helen, Katherine, Sylvia, Nora)." Everett and Borgatti (2013, 208) noted at the time that this was a partition "different from all of the other published blockings for these data." Note that the only disagreement between Everett and Borgatti's partition and the three-block solution suggested by Figure 5c is Myrna/Myra's position, who is assigned to Borgatti and Everett's third group in the a posteriori blocking obtained from Bonacich's second dimension. This may be partly due to the discrepant position of "core event" E9 along the columns of the blocked matrix, which is grouped with the second group of "peripheral" events {E10, E11, E12, E13, E14} in Figure 5c.

This analysis points to a new way to interpret the second Bonacich dimension, namely, as a way to reveal secondary core-periphery partitions net of a main core-periphery partition in the two-mode network (if any exists). Furthermore, the similarities between persons and groups defined over the two-dimensional Bonacich space, shown in Figure 8 (c), can thus be used to reveal *structurally equivalent* clusters of persons and events similar to those obtained by Everett and Borgatti (2013, Table 5). This is suggested by the coloring of the persons and groups in Figure 8(c), which comes from a k-means cluster analysis using the first two Bonacich dimensions to define initial cluster centroids, revealing a three-cluster solution almost equivalent to that obtained by Everett and Borgatti, with the exception of Myrna/Myra's and event E9's assingment as discussed earlier. This suggests that the Bonacich approach to similarity clustering is distinct from CA in being useful for revealing structural equivalence, thus making the Bonacich dimensions a more faithful basis for a low-rank reconstruction of the original "input data" than CA if that is the goal of the analysis (D'Esposito et al., 2014a).

	W1	W2	W3	W4	W5	W6	W7	W8	W9	W10	W11	W12	W13	W14	W15	W16	W17	W18
W1 (Evelyn)	0.19	0.16	0.14	0.15	0.14	0.12	0.08	0.09	0.07	0.04	0.04	0.03	0.02	0.03	0.01	0.08	0.04	0.04
W2 (Laura)	0.14	0.18	0.12	0.13	0.1	0.12	0.11	0.07	0.07	0.04	0.02	0.01	0.02	0.03	0.03	0.04	0	0
W3 (Theresa)	0.14	0.13	0.16	0.12	0.16	0.12	0.11	0.09	0.09	0.06	0.04	0.03	0.04	0.04	0.03	0.08	0.04	0.04
W4 (Brenda)	0.13	0.13	0.1	0.17	0.16	0.12	0.11	0.07	0.07	0.04	0.02	0.01	0.02	0.03	0.03	0.04	0	0
W5 (Charlotte)	0.07	0.06	0.08	0.09	0.16	0.07	0.06	0	0.06	0.03	0	0	0.01	0.01	0.02	0	0	0
W6 (Frances)	0.06	0.07	0.06	0.07	0.07	0.12	0.08	0.07	0.05	0.02	0.02	0.01	0.01	0.02	0.01	0.04	0	0
W7 (Eleanor)	0.04	0.06	0.05	0.06	0.06	0.08	0.11	0.07	0.07	0.04	0.02	0.01	0.02	0.03	0.03	0.04	0	0
W8 (Pearl)	0.03	0.03	0.03	0.03	0	0.05	0.05	0.09	0.04	0.04	0.04	0.03	0.02	0.03	0.01	0.08	0.04	0.04
W9 (Ruth)	0.03	0.04	0.05	0.04	0.06	0.05	0.07	0.05	0.09	0.06	0.04	0.03	0.04	0.02	0.03	0.08	0.04	0.04
W10 (Verne)	0.02	0.02	0.03	0.02	0.03	0.02	0.04	0.05	0.06	0.11	0.08	0.05	0.06	0.04	0.07	0.08	0.04	0.04
W11 (Myra)	0.02	0.01	0.02	0.01	0	0.02	0.02	0.05	0.04	0.08	0.13	0.09	0.07	0.06	0.09	0.08	0.04	0.04
W12 (Katherine)	0.02	0.01	0.02	0.01	0	0.02	0.02	0.05	0.04	0.08	0.13	0.2	0.17	0.14	0.09	0.08	0.04	0.04
W13 (Sylvia)	0.02	0.02	0.03	0.02	0.03	0.02	0.04	0.05	0.06	0.11	0.13	0.2	0.18	0.15	0.11	0.08	0.04	0.04
W14 (Nora)	0.03	0.03	0.04	0.03	0.03	0.03	0.06	0.07	0.05	0.09	0.11	0.19	0.17	0.2	0.14	0.04	0.17	0.17
W15 (Helen)	0.01	0.02	0.02	0.02	0.03	0.02	0.04	0.02	0.04	0.08	0.11	0.07	0.08	0.09	0.16	0.04	0.12	0.12
W16 (Dorothy)	0.02	0.01	0.02	0.01	0	0.02	0.02	0.05	0.04	0.04	0.04	0.03	0.02	0.01	0.01	0.08	0.04	0.04
W17 (Olivia)	0.01	0	0.01	0	0	0	0	0.03	0.02	0.02	0.02	0.01	0.01	0.04	0.05	0.04	0.17	0.17
W18 (Flora)	0.01	0	0.01	0	0	0	0	0.03	0.02	0.02	0.02	0.01	0.01	0.04	0.05	0.04	0.17	0.17

(a) People.

	E1	E2	E3	E4	E5	E6	E7	E8	E9	E10	E11	E12	E13	E14
E1 (6-27)	0.14	0.09	0.07	0.07	0.05	0.05	0.03	0.03	0.01	0	0	0	0	0
E2 (3-2)	0.09	0.13	0.07	0.06	0.05	0.05	0.03	0.03	0.02	0	0	0	0	0
E3 (4-12)	0.14	0.13	0.17	0.16	0.13	0.1	0.07	0.06	0.02	0	0	0	0	0
E4 (9-26)	0.09	0.08	0.11	0.16	0.08	0.05	0.05	0.03	0.02	0	0	0	0	0
E5 (2-25)	0.14	0.13	0.17	0.16	0.19	0.13	0.12	0.09	0.04	0	0	0	0	0
E6 (5-19)	0.14	0.13	0.13	0.1	0.13	0.19	0.08	0.1	0.06	0.03	0.03	0.02	0.04	0.04
E7 (3-15)	0.1	0.09	0.11	0.13	0.15	0.1	0.19	0.11	0.07	0.09	0.08	0.12	0.09	0.09
E8 (9-16)	0.14	0.13	0.13	0.1	0.16	0.17	0.15	0.22	0.18	0.15	0.05	0.17	0.1	0.1
E9 (4-8)	0.04	0.08	0.04	0.06	0.06	0.09	0.09	0.15	0.27	0.14	0.28	0.16	0.14	0.14
E10 (6-10)	0	0	0	0	0	0.02	0.05	0.05	0.06	0.18	0.08	0.15	0.14	0.14
E11 (2-23)	0	0	0	0	0	0.02	0.03	0.01	0.09	0.06	0.33	0.05	0.04	0.04
E12 (4-7)	0	0	0	0	0	0.02	0.07	0.07	0.08	0.18	0.08	0.19	0.14	0.14
E13 (11-21)	0	0	0	0	0	0.02	0.03	0.02	0.04	0.09	0.03	0.07	0.14	0.14
E14 (8-3)	0	0	0	0	0	0.02	0.03	0.02	0.04	0.09	0.03	0.07	0.14	0.14

(b) Groups.

Figure 9: SimRank projection matrices of the Southern Women data.

5. 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.

In this section, I show that the contrast between the CA and Bonacich similarities can be understood along the same lines, with the Bonacich decomposition being more in line with a first-order similarity (structural equivalence) partition of nodes in the network (person and groups are similar to the extent that they connect to the *same* groups and persons respectively), and the CA clustering producing similarities more congruent with a GRS criterion (thus closer to regular equivalence), where persons (groups) are similar to the extent that they connect to groups (persons) that are themselves similar.

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.

This definition 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:

$$S(p,p') = \frac{\alpha}{C^R(1)_p C^R(1)_{p'}} \sum_{i=1}^{C^R(1)_p} \sum_{j=1}^{C^R(1)_{p'}} S\left(g(i)_{i \in N(p)}, g(j)_{j \in N(p')}\right) \tag{31}$$

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 31 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:

$$S(g,g') = \frac{\alpha}{C^R(1)_g C^R(1)_{g'}} \sum_{i=1}^{C^R(1)_g} \sum_{j=1}^{C^R(1)_{g'}} S\left(p(i)_{i \in N(g)}, p(j)_{j \in N(g')}\right) \tag{32}$$

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 31 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 32, then loop back to equation 31 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 31 and 32 share a formal similarity with HH's "method of reflections" discussed earlier, in particular, the fact that both compute quantities based on information from 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 noted earlier. Beyond that, the SimRank approach can be thought of as yet another way of obtaining one-mode projections from an original two-mode network, in line with Everett and

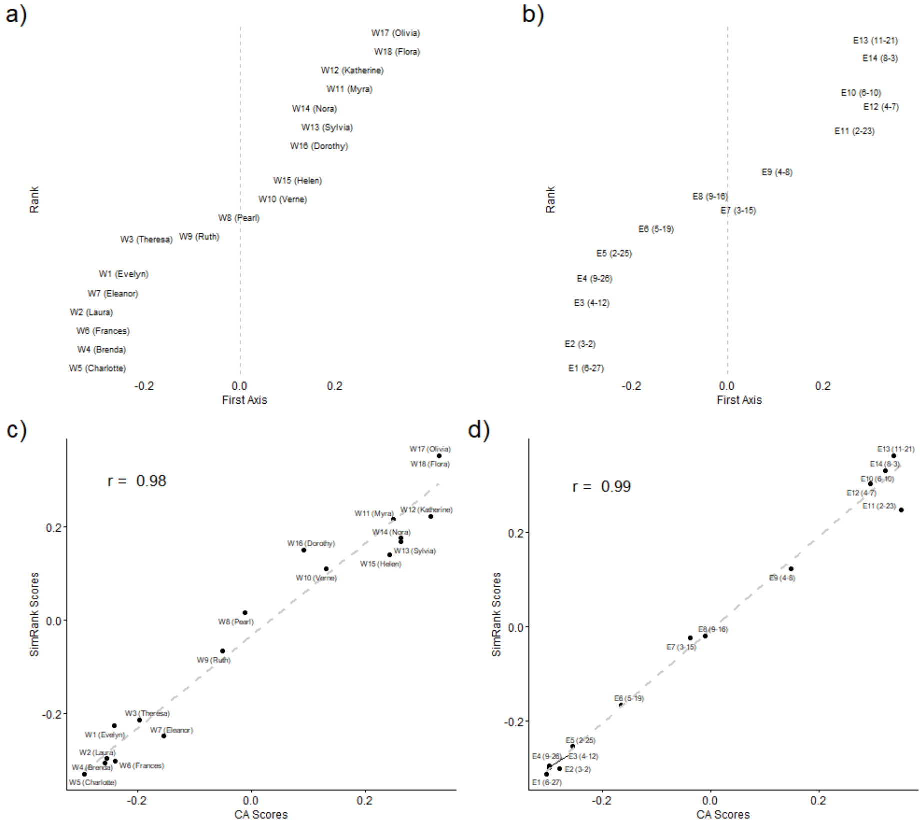


Figure 10: Simrank versus CA comparison. Panels (a) and (c) show the final SimRank similarity matrices for persons and groups (respectively) with rows and columns ordered according to the first CA dimension scores. Panel (b) shows the Pearson correlation between the scores corresponding to the second eigenvector of the SimRank similarity matrix (on the y-axis) and the scores corresponding to the first CA dimension (on the x-axis) for persons, panel (d) shows the same correlation for groups.

Borgatti's "dual projection" approach (Everett and Borgatti, 2013). The SimRank projection matrices for the Southern Women data—obtained after successful convergence with α set to 0.8—are shown in Figure 9.

I compare the results obtained from the SimRank analysis to those obtained via CA in three different ways. First, Figure 5d shows the original affiliation matrix, this time with rows and columns re-arranged according to the main non-trivial eigenvector (corresponding to the second-largest eigenvalue) of the SimRank projection matrices for persons and column groups. Like in Figure 5(a), this reshuffling displays a block-diagonal structure separating the same two blocks of persons and groups as that revealed by CA, suggesting that CA and SimRank uncover a similar underlying partitioning of the two-mode network's community structure.

Second, Figure 10(a) and Figure 10(b) show an eigenvector plot similar to that shown in Figures 6(a) and Figure 6(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 most informative eigenvector of the SimRank similarity matrix is substantively equivalent to those revealed by the first CA eigenvector, separating similar blocks of people and events.

Finally, Figures 10(c) and (d) show a scatter plot and associated regression line with the first CA dimension scores for persons and groups (respectively) on the x-axis and the scores extracted from the main non-trivial 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 they capture substantively identical similarity information for both persons and groups.

6. Discussion and Concluding Remarks

This paper reconsiders the role of CA in the analysis of two-mode network data. After considering the various ways CA has been dealt with in the two-mode network analysis literature, I examined 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 a simple CA of the affiliation matrix. Working backward from this, I also clarified the linkages between CA and the more commonly used eigenvector approach for calculating dual scores in two-mode data due to Bonacich (1991). I showed that just like the reflective scores 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. At the same time, the Bonacich approach works with unweighted (but normalized) sums. This exercise clarifies the links between CA and the dual two-mode eigenvector centrality more coherently and systematically, an issue that was left open and somewhat unclear in Bonacich's (1991) classic paper.

As Bonacich (1991) suspected in his original paper, CA can be shown to reveal a *clustering* of nodes based on 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 a posterior blocking approach to re-ordering the original data using the Bonacich centralities, which instead uncover a latent core-periphery partition between actors and events (Everett and Borgatti, 2013).

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 blocking 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.

Nevertheless, the contrast between the Bonacich approach as one geared toward "centrality" analysis and CA as one geared toward "positional analysis," is much too simple. After all, as many have noted before, the Bonacich-style eigendecomposition of a proximity matrix yields dimensions beyond the first eigenvector, which are sensitive to mesolevel structure in the network, including that linked to group and community partitioning (Iacobucci et al., 2017). I show that this is indeed the case in the two-mode case, with the second eigenvector of the Bonacich decomposition revealing a secondary dual core-periphery partition, and with a clustering that incorporates that dimension, revealing structurally equivalent blocks of persons and events (Everett and Borgatti, 2013).

This left open the question of what kind of similarity the clustering based on the CA of the two-mode network is sensitive to. I showed that, in contrast to the higher-order Bonacich analysis, 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), I showed 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 a two-mode network CA reveals latent groupings of nodes based on a GRS criterion (although in the Southern Women data, both he structural equivalence and GRS partitionings will tend to agree pretty closely).

Overall, the preceding has shown that it may be time to reconsider the place of CA in the pantheon of two-mode network analysis methods. As we saw at the outset, CA has either been thought of as a "data reduction" method designed to provide low-dimensional visualizations of the original affiliation matrix, or contrasted individually with methods—such as eigenvector centrality analysis—designed to produce scores for ranking the two sets of nodes in the affiliation network. This paper shows that neither of these CA considerations is quite on track.

Minimally, the previous analysis shows that CA can be upgraded from being a method geared exclusively to visualization to one that can be seen as more "central" to the usual social-network-analytic tasks, like finding blocks of structurally similar actors in the two-mode network, thus bringing the use of CA within the SNA community closer to how it is used in link analysis, information science, and network science, where it is seen primarily as a spectral clustering method (van Dam et al., 2021; Yen, Saerens and Fouss, 2010; Mealy et al., 2019). 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—out of the many options available (Yang, Aronson, Odabas, Ahn and Perry, 2022)—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 blocking produced via the reflective scores obtained by CA and the Bonacich approach side by side to see whether the underlying display suggests a substantively interesting similarity partitioning in addition to any core-periphery ordering. Of course, suppose the analyst is interested in a blocking or clustering approach. In that case, the contrast should shift to the possibly distinct, possibly overlapping similarity clustering that can be obtained via CA or the traditional Bonacich approach, which, as we have seen, is keyed to analytically distinct conceptions of the ways nodes can be similar in a two-mode network. In that respect, once the commonalities and differences in the underlying mathematical machinery are clarified, both CA and Bonacich centrality analysis are equally worthy but distinctive weapons in the arsenal of the "dual projection" (Everett and Borgatti, 2013) approach to the analysis of two-mode networks.

References

Batagelj, V., 2022. Analysis of the southern women network using fractional approach. Social Networks 68, 229–236.

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., Halgin, D.S., 2014. Analyzing affiliation networks, in: Carrington, J.S.P.J. (Ed.), The Sage handbook of social network analysis. Sage Publications, pp. 417–433.

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.

Brusco, M., 2011. Analysis of two-mode network data using nonnegative matrix factorization. Social Networks 33, 201–210.

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.

Deng, H., Lyu, M.R., King, I., 2009. A generalized co-hits algorithm and its application to bipartite graphs, in: Proceedings of the 15th ACM SIGKDD international conference on Knowledge discovery and data mining, pp. 239–248.

D'Esposito, M.R., De Stefano, D., Ragozini, G., 2014a. A comparison of χ 2 metrics for the assessment of relational similarities in affiliation networks, in: Analysis and Modeling of Complex Data in Behavioral and Social Sciences, Springer. pp. 113–122.

D'Esposito, M.R., De Stefano, D., Ragozini, G., 2014b. On the use of multiple correspondence analysis to visually explore affiliation networks. Social Networks 38, 28–40.

Doreian, P., 1979. On the evolution of group and network structure. Social Networks 2, 235–252.

Doreian, P., Batagelj, V., Ferligoj, A., 2004. Generalized blockmodeling of two-mode network data. Social networks 26, 29–53.

Dramalidis, A., Markos, A., 2016. Subset multiple correspondence analysis as a tool for visualizing affiliation networks. Journal of Data Analysis and Information Processing 4, 81–89.

Everett, M.G., Borgatti, S.P., 1994. Regular equivalence: General theory. Journal of mathematical sociology 19, 29–52.

Everett, M.G., Borgatti, S.P., 2013. The dual-projection approach for two-mode networks. Social networks 35, 204–210.

Everett, M.G., Broccatelli, C., Borgatti, S.P., Koskinen, J., 2018. Measuring knowledge and experience in two mode temporal networks. Social Networks 55, 63–73.

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.

Faust, K., Willert, K.E., Rowlee, D.D., Skvoretz, J., 2002. Scaling and statistical models for affiliation networks: patterns of participation among soviet politicians during the brezhnev era. Social networks 24, 231–259.

Field, S., Frank, K.A., Schiller, K., Riegle-Crumb, C., Muller, C., 2006. Identifying positions from affiliation networks: Preserving the duality of people and events. Social networks 28, 97–123.

Fouss, F., Saerens, M., Shimbo, M., 2016. Algorithms and models for network data and link analysis. Cambridge University Press.

Freeman, L.C., 2003. Finding social groups: A meta-analysis of the southern women data, National Academies Press.

Friedkin, N.E., 1991. Theoretical foundations for centrality measures. American journal of Sociology 96, 1478–1504.

Hidalgo, C.A., 2021. Economic complexity theory and applications. Nature Reviews Physics 3, 92–113.

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.

Homans, G.C., 1950. The human group. Harcourt, Brace.

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.

Kumbasar, E., Rommey, A.K., Batchelder, W.H., 1994. Systematic biases in social perception. American journal of sociology 100, 477–505.

Leicht, E.A., Holme, P., Newman, M.E., 2006. Vertex similarity in networks. Physical Review E 73, 026120.

Lerner, J., Lomi, A., 2022. A dynamic model for the mutual constitution of individuals and events. Journal of Complex Networks 10, cnac004.

Lizardo, O., 2024. Two-mode relational similarities. Social Networks 76, 34–41.

Lizardo, O., Taylor, M.A., 2020. Correspondence analysis. SAGE Publications Ltd.

Martin, C.D., Porter, M.A., 2012. The extraordinary svd. The American Mathematical Monthly 119, 838–851.

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.

Noma, E., Smith, D.R., 1985. Scaling sociomatrices by optimizing an explicit function: correspondence analysis of binary single response sociomatrices. Multivariate Behavioral Research 20, 179–197.

Podolny, J.M., 2001. Networks as the pipes and prisms of the market. American journal of sociology 107, 33–60.

Ragozini, G., De Stefano, D., D'Ambrosio, D., 2014. Correspondence analysis with doubling for two-mode valued networks, in: Book of abstracts of the 1st EUSN conference, EUSN.

Ragozini, G., De Stefano, D., D'Esposito, M.R., 2015. Multiple factor analysis for time-varying two-mode networks. Network Science 3, 18–36.

Ragozini, G., Serino, M., D'Ambrosio, D., 2018. On the analysis of time-varying affiliation networks: The case of stage co-productions, in: Studies in Theoretical and Applied Statistics: SIS 2016, Salerno, Italy, June 8-10, Springer. pp. 119–129.

Roberts Jr, J.M., 2000. Correspondence analysis of two-mode network data. Social Networks 22, 65–72.

Roffilli, M., Lomi, A., 2006. Identifying and classifying social groups: A machine learning approach, in: Data Science and Classification. Springer, pp. 149–157.

Schweizer, T., 1991. The power struggle in a chinese community, 1950-1980: A social network analysis of the duality of actors and events. Journal of Quantitative Anthropology 3, 19–44.

Serino, M., Picardi, I., Ragozini, G., 2024. Mapping epistemic pluralism: A network analysis of discursive practices in communities promoting refused knowledge about healthcare and wellbeing. Poetics 107, 101929.

Wang, P., Sharpe, K., Robins, G.L., Pattison, P.E., 2009. Exponential random graph models for affiliation networks. Social Networks 31, 12–25.

Wasserman, S., Anderson, C., 1987. Stochastic a posteriori blockmodels: Construction and assessment. Social networks 9, 1–36.

Wasserman, S., Faust, K., Galaskiewicz, J., 1990. Correspondence and canonical analysis of relational data. Journal of Mathematical Sociology 15, 11–64.

Yang, K.C., Aronson, B., Odabas, M., Ahn, Y.Y., Perry, B.L., 2022. Comparing measures of centrality in bipartite patient-prescriber networks: A study of drug seeking for opioid analgesics. Plos one 17, e0273569.

Yen, L., Saerens, M., Fouss, F., 2010. A link analysis extension of correspondence analysis for mining relational databases. IEEE Transactions on Knowledge and Data Engineering 23, 481–495.

Zhu, M., Kuskova, V., Wasserman, S., Contractor, N., 2016. Correspondence analysis of multirelational multilevel networks. Multilevel Network Analysis for the Social Sciences: Theory, Methods and Applications , 145–172.