Chapter 7
Correspondence Analysis of Multirelational Multilevel Networks

Mengxiao Zhu, Valentina Kuskova, Stanley Wasserman, and Noshir Contractor

Introduction

Social network analysis is concerned not only with social relations (Wellman 1988), but also more generally with attributes across pairs of social actors, which are referred to as *dyadic attributes* (Borgatti and Everett 1987, p. 243). These dyadic attributes range from shared affiliations to distances between cities to similarities in respondents' answers to items on a questionnaire. While most network studies have investigated one-mode networks (Borgatti and Everett 1987), social network approaches are easily extended to two-mode data, such as the relationship between employees and work teams with which they are affiliated (Wasserman and Faust 1994). In two-mode networks, different types of nodes (e.g., employees and teams) are represented as different modes. Unlike typical affiliation networks (for a primer on affiliation networks, please refer to Wasserman and Faust 1994;

M. Zhu (✉)
Educational Testing Service, Princeton, NJ, USA
e-mail: mzhu@ets.org; mengxiao.zhu@gmail.com

V. Kuskova
National Research University Higher School of Economics, Moscow, Russian Federation
e-mail: vkuskova@hse.ru

S. Wasserman
National Research University Higher School of Economics, Moscow, Russian Federation

Indiana University, Bloomington, IN, USA
e-mail: stanwass@indiana.edu

N. Contractor
Northwestern University, Evanston, IL, USA
e-mail: nosh@northwestern.edu

© Springer International Publishing Switzerland 2016

E. Lazega, T.A.B. Snijders (eds.), *Multilevel Network Analysis for the Social Sciences*, Methodos Series 12, DOI 10.1007/978-3-319-24520-1_7

Borgatti et al. 2013; Newman 2010), where a second mode is just a subset of the first, we focus on the data with emergent properties of the second mode (e.g., teams are more than just subsets, because they perform additional emergent functions that go beyond uniting individual employees together). The relations in this network are the links between nodes of different modes. Over the years, a variety of tools and techniques have been developed for displaying, analyzing, and interpreting such data (e.g., Borgatti and Everett 1987; Doreian et al. 2004; Latapy et al. 2008; Roberts 2000).

However, most current two-mode network analysis techniques focus exclusively on the links between different modes, without considering dyadic attributes such as attribute similarity or nodes nested in networks at different levels. Despite recent advancements in network analysis methods, including extensions of multiple correspondence analysis (e.g., D'Esposito et al. 2014), there is clearly a need to further extend many of these approaches to multiple levels. The multilevel approach, recently popularized in fields such as management, combines the unit of observation with higher levels in which the focal unit is embedded (individuals in teams, teams in units, units in organizations; e.g., Phelps et al. 2012). This multiple level approach has been considered, for example, in studies on innovation (Berends et al. 2011) and knowledge management (Zhao and Anand 2013). Methods such as canonical correlation analysis have been used to evaluate two sets of variables simultaneously with the contingency table as input. For example Parkhe (1993) utilized this approach to study the relationships between a set of performance variables and a set of payoffs (e.g., Parkhe 1993). While it is a very useful technique and can be applied to multiway contingency tables (Gilula and Haberman 1988), canonical correlation analysis is not designed for the study of network structure and composition variables, and this limitation is often acknowledged in studies using this technique (e.g., Berends et al. 2011; Payne et al. 2011; Zhao and Anand 2013). As these researchers noted, some research questions can neither be asked nor answered, because of the lack of methodology for analyzing relational data at multiple levels.

Consider, for example, individual actors nested within teams. Due to existing organizational structures or other a priori arrangements, actors are often nested within multiple teams that share one or more members, thus giving rise to affiliation data (Wageman et al. 2012). In such an arrangement, the actors are the lower level, the teams are the upper level, and the actors can be in more than one team. These types of data are multilevel and can be complex if the actor-nesting is not mutually exclusive. In many teams, especially self-assembled teams, individuals participate in more than one team (e.g., Denton 1997; Kauffeld 2006). In the meantime, individuals are socially connected to each other through previous collaborations or communications with each other. The dependencies among overlapping teams create an additional level of complexity in the analysis, with the presence of the aggregate effects of team members' interactions with one another (Klein and Kozlowski 2000). Further, there are theoretical reasons for understanding the effects of multiple team membership on productivity and learning at the individual and team level (O'Leary et al. 2011).

The need for a robust method for visualizing and modeling multilevel relational data becomes even more challenging as data sets become richer and larger. Consider, for example, big data, and the need to have exploratory and data reduction tools to deal with it. According to McAffee and Brynjolfsson (2012), about 2.5 exabytes of data are created every day, and this number doubles approximately every 40 months (an exabyte is 1000 times a petabyte, which is equivalent to about 20 million cabinets' worth of text. Walmart alone generates approximately 2.5 petabytes of data an hour from customer transactions). While insights gathered from analyzing big data can allow companies to substantially outperform the competition (McAfee and Brynjolfsson 2012), these insights can only be unleashed when we have a better understanding of how to use analytics and methodological tools for data reduction (LaValle et al. 2011). In addition to the size, the most notable thing about big data is the embedded relationality: patterns of connections between individuals, groups of people, relationships between them, or just the structure of information, such as presence of latent factors within the set (Boyd and Crawford 2011). Network approaches are a logical choice to discern these insights, and in this regard, multiple correspondence analysis can prove to become a useful tool, especially when multilevel relationships are embedded within the data. In this regard, the approach we propose can be used as a preliminary data analysis tool, allowing us to look at the structure of the data for the purpose of determining more advanced models that could fit that structure.

To address some of the issues outlined above, this study provides two examples, which model teams of individuals using network methodology. The most straightforward way to represent team membership is to use one-mode networks, where nodes represent individuals, and links among individuals indicate joint participation in one or more teams. This one-mode representation captures the overlapping team membership, but unfortunately, fails to preserve the team structures. For instance, links in a one-mode network between A, B and C fail to convey information about whether A, B, and C were together on one team or if A and B were on one team, C on another team, and B and C on yet another team. Instead, we use affiliation networks to represent teams and individuals, with links representing team membership. There are many representations of an affiliation network, including as an affiliation network matrix, bipartite graph, hypergraph, or simplicial complex (Borgatti and Halgin 2011; Faust 2005; Skvoretz and Faust 1999). Each of these representations contains exactly the same information, and any one representation can be derived from the other. This study uses bipartite networks, in which the two types of nodes represent teams and individuals. The social relations between individuals can then be easily represented using one-mode networks with nodes representing individuals and links representing relations between individuals.

Understanding the associations between and among variables in complex social systems, which exist, for example, in multiple nested groups, can be difficult, and there is a paucity of theoretical models that yield precise hypotheses. Hence we argue for the use of exploratory network analysis techniques as a useful theoretical preamble (and/or data reduction strategy) to more confirmatory approaches such as specifying *p**/ERGM models (Holland and Leinhardt 1981; Wasserman and Robins 2005). This exploration is especially useful in narrowing the potentially unwieldy combinatorial space for model specification. Hence we propose a method for exploration of multilevel relational data that distills the number of possible hypotheses and generative mechanisms that can be subsequently tested using confirmatory methods such as *p**/ERGM.

We use correspondence analysis (Wasserman and Faust 1994) and its extension, multiple correspondence analysis (Greenacre 2010), which enable us to analyze multiple relations and attributes at both individual and team levels. Correspondence analysis incorporates the interactions among observations and can be extended to more than two sets of variables; here, we show how it can be used on a much larger number. The results from correspondence analysis can also be presented graphically, using a plot rather than numbers alone. Relations among various variables as well as observed raw data can be shown in the same plot, essentially providing a much richer graphical output. Because of these advantages, correspondence analysis can be used as an important exploratory tool to examine the features of the dataset and the relations among variables of interest.

In this article, we present a brief history of this exploratory network analysis approach, provide theory for extending correspondence analysis to multiple levels, and then provide two illustrative examples from individuals playing in teams in massively multiplayer online games (MMOG). The first example is of combat teams made up of individuals from United States playing the MMOG EverQuest II. We explore the impact of among various individual-level and team-level attributes on team performance, while considering team affiliations and social relations among individuals. We use this example to show how multiple correspondence analysis can be used for hypotheses generation, and later for confirmatory testing with more advanced methods. The second example considers another MMOG, Dragon Nest, with individuals, this time in China, playing in multiple combat teams. We use the example to demonstrate the utility of our methodology to discern cultural differences in team assembly and performance, showing how comparative analysis of large datasets could unearth cultural differences.

Methodology

Existing Methods for Analyzing Affiliation Network Data

Bipartite networks are one way to represent affiliation networks as including two types of nodes as well as relationships between those two types of nodes (Wasserman and Faust 1994). Figure 7.1 illustrates a bipartite network of six children and the three parties they attend. The affiliations between individuals and teams are in-between links. For example, a link between *Allison* and *Party 1* indicates that *Allison* went to *Party 1*. This relationship system can have other

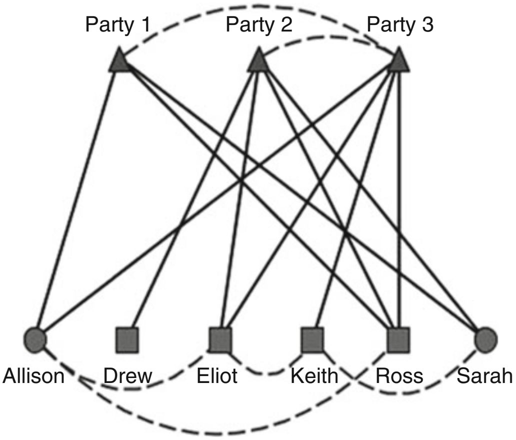
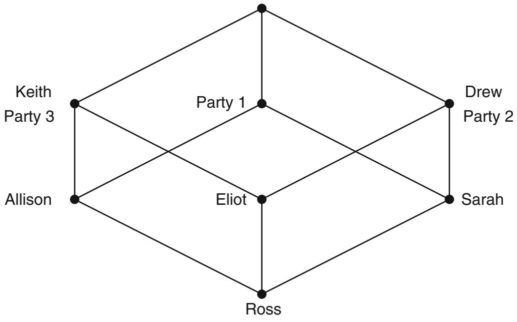


Fig. 7.1 Bipartite graph as example representation for multilevel networks (From Wasserman and Faust 1994, with modifications)

relations as well as attributes, such as one-mode relational ties: If we label two levels as *A* and *B*, we can have AA and BB ties, in addition to AB ties. Also, there could be more than two levels. In the example shown, we can add a third level, alliances (e.g., multiple parties sharing the same theme) and include relational ties within and between all levels (it is not shown on the graph, but can be easily inferred as one more level for alliances between parties). For such structures, there are two types of variables: Q composition (or attribute) and R structure (or relational) variables. Such a dichotomy of composition and structure variables is rather common in data analysis, where one or more response variables are predicted as a function of a collection of explanatory variables. In the example in Fig. 7.1, $g = 6$ and refers to the number of children; $h = 3$ refers to the number of parties. We consider the variables measured on each of the $g(g - 1)/2$ dyads in case of a one-mode network, or the $g * h$ dyads (in the case of a two-mode network with two levels). The number of dyads changes corresponding to the number of modes and levels, and we consider *N* pairs of possible inter-actor relationships (where *N* is, for example, equal to $g(g - 1)/2$ or $g * h$) as the rows of a matrix and consider the variables that are measured on the *N* rows. As a result, these dyadic pattern matrices have *N* rows and the number of columns equal to the levels of composition variables taken together.

Ideally, representation of multilevel data should facilitate the visualization of all three kinds of patterning described above: the AB structure, the A level structure, and the B level structure. While simplicial complexes and hypergraphs (for details, please see Wasserman and Faust 1994) provide representations, a common approach is to convert the two-mode network into two one-mode networks: one shows how actors are linked to each other in terms of events, and the other, how events are linked together in terms of actors. However, neither provides an overall picture of the total A, B, and AB structures. Bipartite graphs display the AB structure but they do not provide a clear image of the associations among A actors or B actors.

This limitation is resolved with Galois lattices (Wasserman and Faust 1994), which meet all three requirements in a clear visual model (Fig. 7.2 presents the



Drew: {Party 2}
Keith: {Party 3}
Sarah: {Party 1, Party 2}
Eliot: {Party 2, Party 3}
Allison: {Party 1, Party 3}
Ross: {Party 1, Party 2, Party 3}

Fig. 7.2 Galois lattice as example representation for multilevel networks (From Wasserman and Faust 1994)

previous example as a Galois lattice). The nodes in Galois lattices are points in a multidimensional space, each representing a subset of actors and events. Reading from the bottom up, there is a line or sequence of lines ascending from a child to a party if he or she attended that party. If two children's ascending lines reach the same party node, then they both attended that party. On the other hand, if two children's ascending lines can only reach the top (null) node, then it means that they did not attend any party together, (e.g., Keith and Drew in the example in Fig. 7.2). In a similar way, the relations between the parties can be read through the descending lines that reach children.

Despite the obvious clarity, focus on subsets, and ability to display complementary relationships between the A's and the B's, the Gallois lattice method has a number of disadvantages. First, the usual display becomes complex as the number of actors and events increases. Second, there is no unique best visual. The vertical dimension represents degrees of subset inclusion relationships among points, but the horizontal dimension is arbitrary. As a result, constructing good measures is problematic. Third, unlike graph theory, properties and analyses of Galois lattices are not well developed.

Correspondence Analysis and Multiple Correspondence Analysis

Correspondence analysis and its extension, multiple correspondence analysis, were originally developed as a multivariate statistical technique for analysis of categorical data (Greenacre 2010; Wasserman et al. 1990). In the 1980s, researchers started to apply this method to analyze the one-mode and two-mode social network structures. For a detailed review and comparison with other statistical methods, such as canonical analysis, refer to, e.g., Borgatti and Everett 1987; Faust 1997; Wasserman and Faust 1994; Wasserman et al. 1990. Many applications of correspondence analysis and multiple correspondence analysis have focused on visualizing and displaying structures in networks (e.g., D'Esposito et al. 2014; Roberts 2000). Recently (D'Esposito et al. 2014), extensions of correspondence analysis have allowed studies to take into account the nature of the relational data (e.g., structural (dis)similarity of actors or events) and nodes' attributes. In this study, we apply correspondence analysis and especially multiple correspondence analysis to multiple and multilevel networks by focusing on the relations between attribute variables at both the same and different levels, while considering and controlling for the network relations between nodes from different modes.

For a two-mode network with A actors as one mode and B actors as another mode, correspondence analysis is a method for visually representing both the rows and the columns of a two-mode matrix in a map, where points representing the A actors are placed together if they are tied to the same B actors; points representing the B actors are placed close together if they are tied to the same A actors; and A points and B points are placed together if those A's are tied to those B's. Correspondence analysis includes an adjustment for marginal effects. As a result, A's are placed closed to B's to the extent that these B's were tied to few other A's, and these A's are tied to few other B's. One of the advantages of this method is that it allows studying correlations between the scores in rows and columns. Using reciprocal averaging, a score for a given row is the weighted average of the scores for the columns, and the weights are the relative frequencies of the cells (Wasserman and Faust 1994).

Mathematically, a bipartite network B with g nodes on the first mode and h nodes on the second mode can be represented using a $g \times h$ matrix $\mathscr{M}$, where $m_{ij} = 1$ if node i from the first mode is affiliated with node j from the second mode and $m_{ij} = 0$ otherwise. For example, in Table 7.1, the bipartite network is represented as follows.

Correspondence analysis takes the affiliation matrix as an input and represents the relations between nodes in both modes as well as the relations in a low-dimensional map. Results then identify multiple factors, which help to cluster nodes from both modes based on the affiliation relations. Nodes from each mode are assigned a set of scores, which can be used to cluster these nodes. For a bipartite network of individuals and teams, the results of the correspondence analysis summarize relations among individuals and teams in a dimension much lower than the dimension of the network itself.

Given a $g \times h$ matrix $\mathscr{M}$ as input to the correspondence analysis, a set of $W = \min(g-1, h-1)$ scores are generated for each row and column, with a set of W

Table 7.1 An example of affiliation matrix for bipartite network

	Team 1	Team 2	Team 3
Person a	1	1	0
Person b	1	0	1
Person c	1	0	1
Person d	0	1	0

numbers measuring the correlation between the rows and columns (Wasserman and Faust 1994). The scores satisfy the following relations:

$$\eta_k u_{ik} = \sum\nolimits_{j=1}^{h} \frac{m_{ij}}{m_{i+}} v_{jk} \tag{7.1}$$

$$\eta_k v_{jk} = \sum\nolimits_{i=1}^{g} \frac{m_{ij}}{m_{+j}} u_{ik} \tag{7.2}$$

where, u_{ik} is the row score for row i on dimension k; v_{jk} is the column score for column j on dimension k. η_k^2 is called the *principal inertia*, and the u and v scores are called the *principal coordinates*.

Standard coordinates $\tilde{u}$ and $\tilde{v}$ (Greenacre 1984) can then be calculated by rescaling the principal coordinates:

$$\tilde{u}_{ik} = u_{ik}/\eta_k \tag{7.3}$$

$$\tilde{v}_{jk} = v_{jk}/\eta_k \tag{7.4}$$

These scores are called standardized scores because the weighted mean is equal to 0 and the weighted variance is equal to 1.

The results from the correspondence analysis are often presented by plotting the coordinates using just the first two dimensions (Nenadic and Greenacre 2007; Wasserman and Faust 1989). Each mode is represented in this two dimensional space, with points for each level of both modes. Let us label the first mode as *actors* and the second mode as *teams*. Actors are placed close to each other in this space if they are similar. Teams are placed close to each other in this space if the teams are similar. Specific actors and teams are placed close to each other if those actors involved are closely related to those teams. As an example, Fig. 7.3 shows the plotted correspondence analysis for the bipartite person-team network data as shown in Table 7.1. It can be observed from this plot that Person b and Person c are close to each other, and the data show that both of them are members of both Team 1 and Team 3.

In addition to demonstrating the structural features of the system of teams and individuals, correspondence analysis can also be used to study the relations of attributes of teams and individuals. This requires an extension of the standard correspondence analysis, to what is called *multiple correspondence analysis* (Nenadic and Greenacre 2007; Wasserman et al. 1990). To include attribute variables from many modes, the nonzero relations in the original $g \times h$ matrix $\mathscr{M}$ are represented as dyads and rows in the new matrix. All attribute variables need to be transformed to categorical variables and be represented by indicator vectors as columns in the new matrix. This new matrix is called the *multiple indicator matrix* (Wasserman et al. 1990). Multiple indicator matrices can be developed in a similar way when the relations studied in the models are one-mode networks.

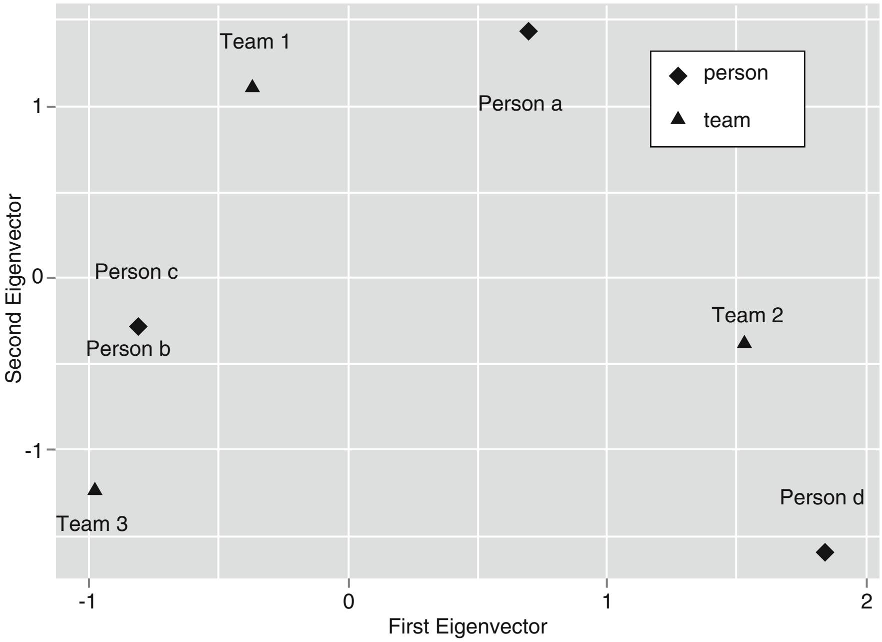


Fig. 7.3 Correspondence analysis for the bipartite network data

Table 7.2 Example of multiple indicator matrix

Dyads	Person gender female	Person gender male	Team performance high	Team performance medium	Team performance low
(a, 1)	0	1	1	0	0
(b, 1)	1	0	1	0	0
(c, 1)	0	1	1	0	0
(a, 2)	0	1	0	1	0
(d, 2)	0	1	0	1	0
(b, 3)	1	0	0	0	1
(c, 3)	0	1	0	0	1

Consider the multiple indicator matrix created based on the network where individuals belong to different teams (Table 7.2). In this example, *Person a*, *Person c*, and *Person d* are males, and *Person b* is female. Performance of *Team 1* is the highest, performance of *Team 2* is midlevel, and performance of *Team 3* is the lowest.

The correspondence analysis on this newly constructed, multiple indicator matrix is called *multiple correspondence analysis*. The results can be interpreted similarly as for simple correspondence analysis.

Illustrative Examples

Example 1: EverQuest II

Data and Sample

Data for this example were obtained from the Massively Multiplayer Online Game (MMOG) EverQuest II (EQ2). It is a fantasy-based game, centered on performance of combat teams, with multiple players nested within teams. The choice of the data is important for several reasons. First, EverQuest is a large dataset, which highlights the relevance of our proposed method in light of the discussion of "big data" above; indeed, large datasets can be analyzed using multilevel correspondence analysis. Second, the nesting feature is important because it indicates that there is more than one level in this example. Server records for the game include player attributes, activities, and relations. Data were collected in two stages: We used relational data among players using the US-based game server between September 5–11, 2006. We collected attribute data such as gender and affiliation with an in-game organization called a guild for the same set of individuals at around 6 p.m. on September 4, 2006. Data from the EQ2 game world data is extremely large and hence analytically intractable. However, the game world is partitioned into smaller island/continent zones. The zones are relatively independent of each other, and, over a short time period, there are no significant player transfers between them. To make the analysis more feasible, teams were identified and sampled by zone. Figure 7.4 shows Zone Everfrost in the EverQuest II game world. The dataset used in this study contains 192 players in 189 teams.

Figure 7.5 contains the bipartite graph of teams and individuals, from the Zone Everfrost, with spring embedding layout (Borgatti et al. 2002). The illustration demonstrates that the individuals and teams in the Zone Everfrost form one big connected subset and several smaller ones.

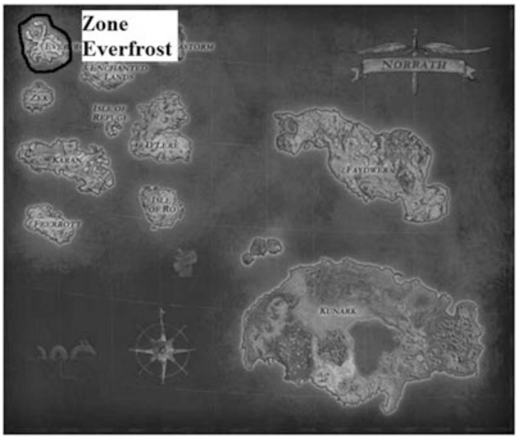


Fig. 7.4 Zone Everfrost in EverQuest II

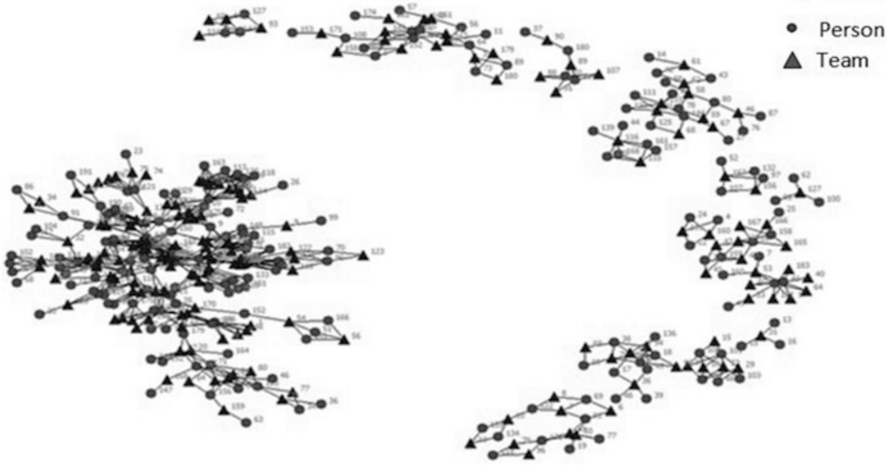


Fig. 7.5 Affiliation network of individuals and teams in Zone Everfrost with spring embedding layout

Variables

There are several types of variables collected in this sample: performance variables, individual- and team-level attributes, and relational variables.

Individual-level attributes include gender (avatar gender, not necessarily the user gender; male/female), age (user age), level (measure of general ability in the game), guild affiliation (whether or not a player belongs to one of the in-game organizations), and expertise. In EverQuest II, there are four prototype classes: Fighter, Priest, Mage, and Scout. They each have special expertise and serve different roles in a team.

Team-level attributes include team size (number of players in a team, with a minimum of two and maximum of six); team level (system-calculated level for the team, representing the general ability of the team); team life span (the length of time the team has been active). There are also expertise diversity, age diversity (coefficient of variation), gender diversity, and guild diversity. In most analyses used as examples in this study, we reported results on guild diversity.

This measure is calculated using the Blau's index (Blau 1977) for each team,

$$D = 1 - \sum\nolimits_{i=1}^{n} p_i^2$$

where n is the total number of guilds and p_i represents the percentage of team members who are in the ith guild.

Team performance variables are derived from the original six built-in metrics in the game: number of monsters killed, number of encounters, earned experience points, total level gain, number of quests completed, and the number of deaths (a negative measure of performance). *F1* is the short-term performance; it captures the number of monsters killed, number of encounters, earned experience points, and total level gain. *F2* is the long-term performance variable; it captures the number of quests completed by the team. *F3* is the negative performance variable; it captures the number of deaths of team members. Performance variables are categorized into three levels: high, mid, and low indexed as 3, 2, and 1 respectively and used as a suffix to the type of performance metric. Thus, for example, in the indicator matrix, F13 indicates high short-term performance, F21 indicates low-level long-term performance, and F31 indicates low-level negative performance.

Other relational independent variables. There are two other player-to-player relations available in the dataset. One is the communication relation constructed by using the one-on-one message exchange activities (AA1). Another is previous collaboration relations, constructed using data from the previous month's log record on who played with whom on the same team (AA2).

We present three example analyses in this study. The first two are simple, but illustrative, models to predict performance from team-level attributes (specifically guild diversity) and individual-level attributes (specifically guild affiliation) respectively. Greater variation in attribute variables implies greater ability to "explain" performance. The third example explores the association between individual attributes (specifically guild affiliation) and social relations (specifically prior communication and collaboration).

Analytic Method

Correspondence analysis was done using R package *ca* (Nenadic and Greenacre 2007) for the bipartite network of Zone Everfrost. Standardized scores from the first two dimensions in Fig. 7.6 above were plotted using R package *ggplot2* (Wickham 2012). The results show several clusters: one big cluster and several smaller ones. This is consistent with the observation of the bipartite network depicted in Fig. 7.5. The correlations and distances among these clusters are measured more mathematically in the correspondence analysis. Among the 12 variables, results from two variables and the performance factors are shown here as examples.

Results

To demonstrate interpretation of results from correspondence analysis, we present three examples using this dataset. The first two examples uses team performance as dependent variable and explanatory variables related to the in-game organization guilds. Two measures, one at the team level and one at the individual level, are analyzed with three measures of performance. The first case is demonstrated with two plots, one with and one without the observed raw relational data, to illustrate how the relational data can be included in the plots. All other plots in this chapter will not include such data. The third example uses multiple network relations

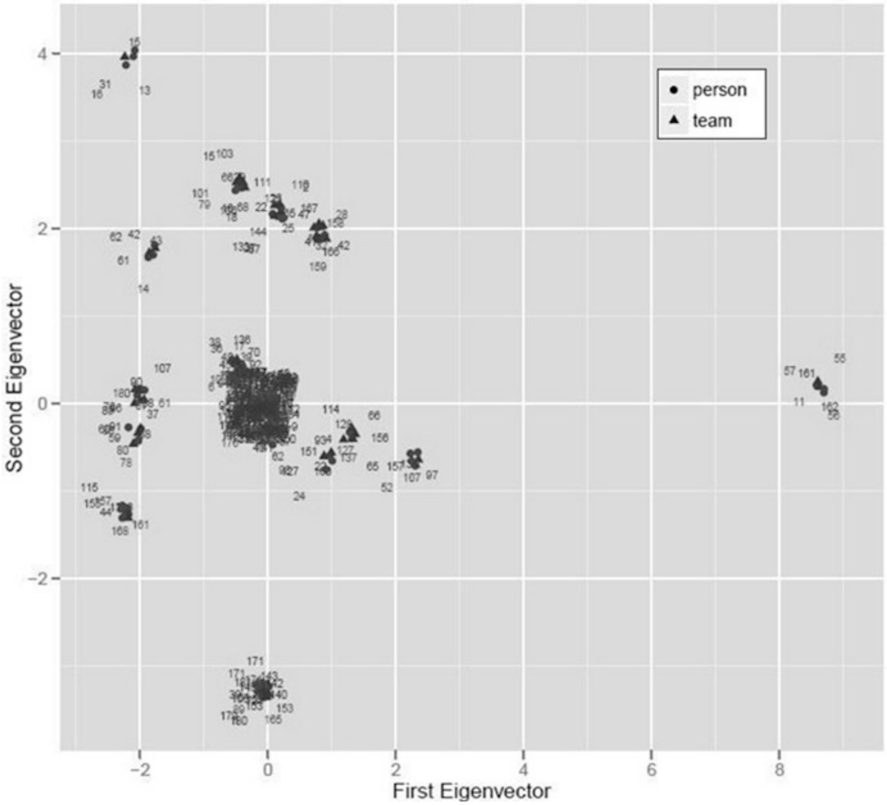


Fig. 7.6 Correspondence analysis for the bipartite network from Zone Everfrost

(i.e., previous collaboration and chat) and an individual-level variable indicating if a player is affiliated with a guild or not. The data are analyzed to discover potential relationships between guild affiliation and social interactions. Results from three analysis models are reported: one model for each network (e.g., collaboration and chat) and one that includes both networks. The results are demonstrated in plots without inclusion of the raw original data.

Team Performance and Team Guild Diversity: Fig. 7.7 shows the plot of the first two dimensions from the results of the multiple correspondence analysis with the original data. Figure 7.8 shows the same results without the original individual-level data, where the relations among the variables are easier to see. Later in the chapter, we show figures similar to Fig. 7.8, but for each of them, a figure similar to Fig. 7.7 can be created. The circles are the raw data of team affiliations, triangles are attribute variables, and the squares are the performance variables.

Table 7.3 shows the numerical breakdown of the analysis. The first two dimensions explained 45.9 % of the total observed variance. The locations of the data points and the locations of the variables show the relations among them. When data points or variables are near to each other, it indicates closer relations.

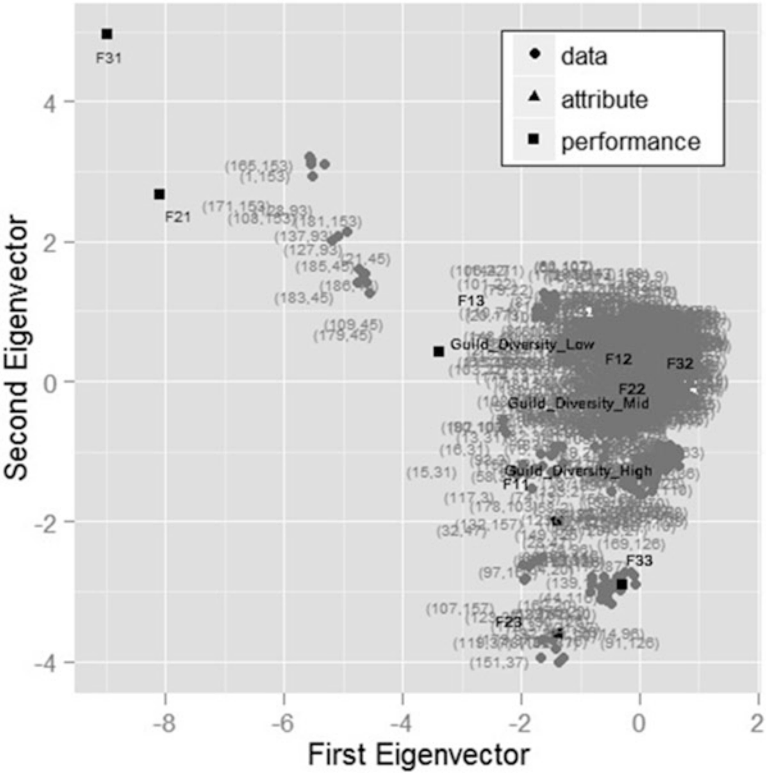


Fig. 7.7 Multiple correspondence analysis of guild diversity and performance with ties data

From the plot, most data points are clustered around the midlevel guild diversity and all three performance factors are at midlevel (F12, F22, and F32). Some teams have low long-term performance, shown by the small cluster near F21, or high long-term performance, shown by the cluster near F23.

Low guild diversity is closer to high and medium short-term performance (F13 and F12), rather than low short-term performance (F11). Low guild diversity is also closer to medium-level long-term performance (F22), rather than low long-term performance (F21). High guild diversity is also close to medium short-term performance (F12). Taken together these results suggest that, when the dependencies among observations are considered, teams with members from diverse guilds have higher short-term performance but lower long-term performance. These exploratory insights should stimulate theoretical explorations leading to hypotheses generation as well as offer opportunities for data reduction.

Team Performance and Individual Guild Affiliation: Next we explored the association of an individual level attribute, specifically guild affiliation, and team performance. Each dyad in the multiple correspondence matrix includes one node from each mode. Attributes at the individual level can be included in the same way

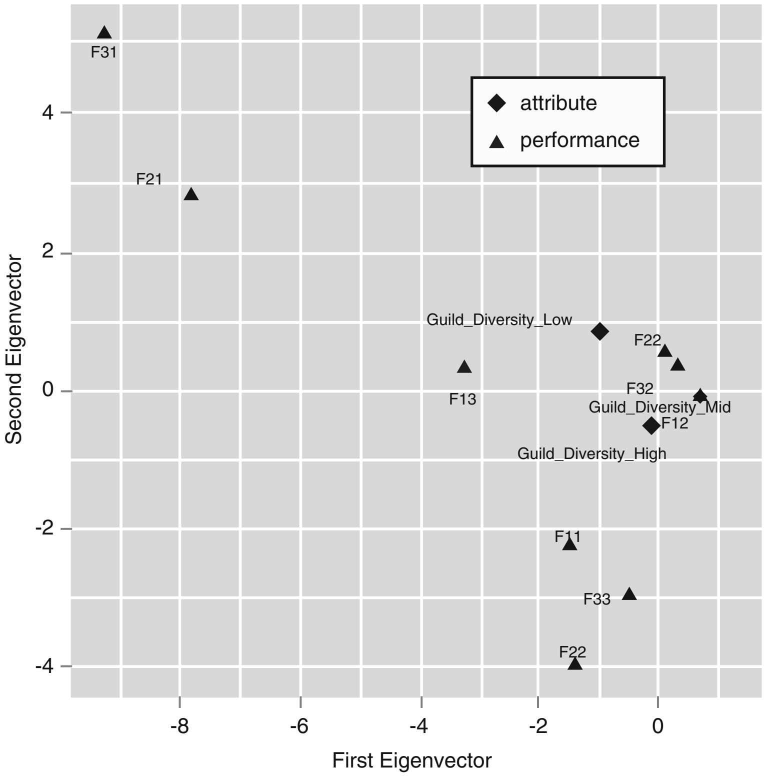


Fig. 7.8 Multiple correspondence analysis of guild diversity and performance

Table 7.3 Multiple correspondence analysis of guild diversity and performance; principal inertias (eigenvalues)

Dimension	Value	Percent variance explained	Cumulative percent variance explained	Screen plot
1	0.13	23.70	23.70	*************************
2	0.12	22.10	45.90	***********************
3	0.08	14.70	60.60	**************
4	0.07	13.00	73.60	************
5	0.05	9.80	83.50	********
6	0.04	7.60	91.00	*****
7	0.03	5.30	96.30	**
8	0.02	3.70	100.00	

as the team-level variables were in the previous example. Figure 7.9 shows the plots of the first two dimensions from the correspondence analysis; Table 7.4 shows the numerical breakdown of the analysis. The first two dimensions explain 50.7 % of the total variance.

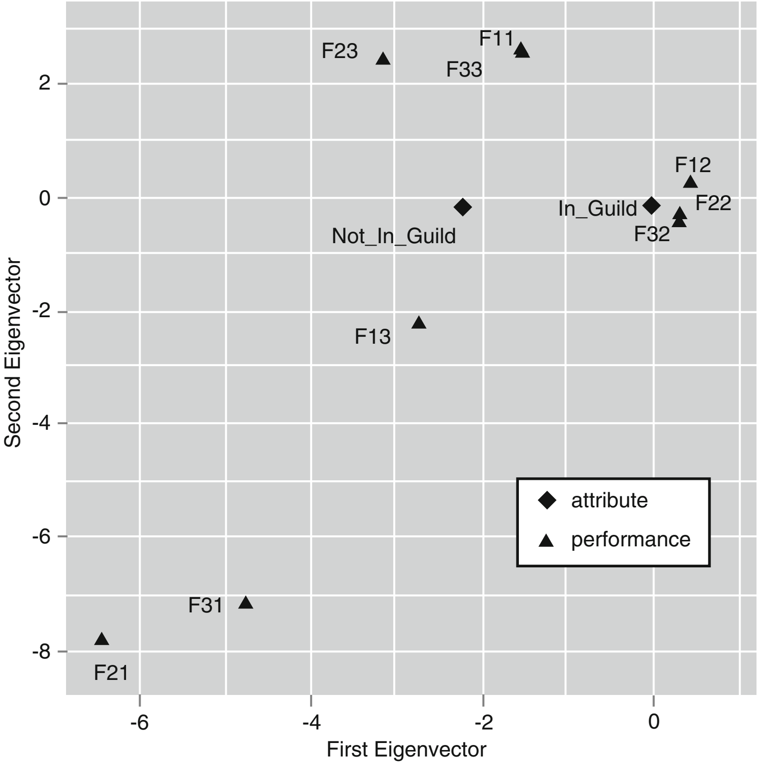


Fig. 7.9 Multiple correspondence analysis of guild affiliation and performance

Table 7.4 Multiple correspondence analysis of individual guild affiliation and performance; principal inertias (eigenvalues)

Dimension	Value	Percent variance explained	Cumulative percent variance explained	Screen plot
1	0.13	26.9	26.9	*************************
2	0.11	23.8	50.7	**********************
3	0.07	14.8	65.5	************
4	0.06	12.5	78.0	*********
5	0.05	11.3	89.3	********
6	0.03	6.5	95.8	***
7	0.02	4.2	100.0	

As shown in both Fig. 7.9, belonging to a guild is associated with mid-level short-term, long-term, and negative performance. Individuals not in a guild face more uncertainty. They may have either very high or very low short-term performance (about equal distance to F11 and F13) and they usually encounter more deaths (closer to F33 than F31).

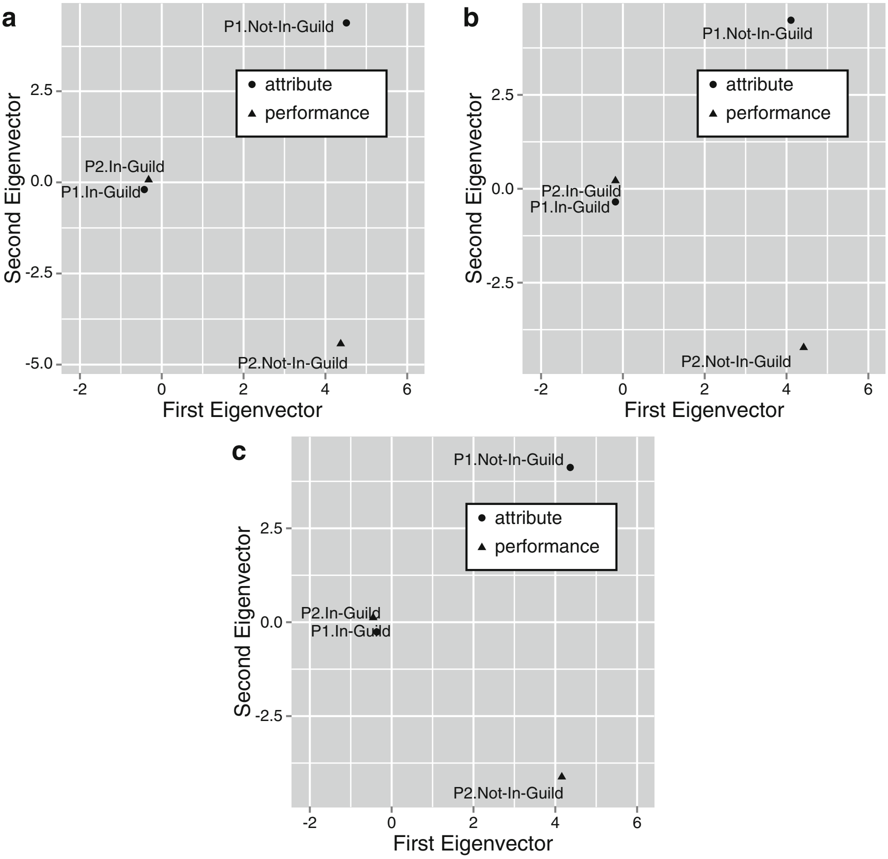


Fig. 7.10 Multiple network analysis results. (**a**) Guild membership and chat relation; (**b**) Guild membership and collaboration relation; (**c**) Guild membership and both relations

Multiple Network Analysis: Our third example with this dataset demonstrates the application of correspondence analysis to multiple networks. We use examples with the two-mode network team affiliation network (AB), two one-mode network relations (chat relation [AA1] and collaboration relation [AA2]), and two attribute variables, an individual level attribute (guild membership) and a team level attribute (team performance). We first explore the association of guild membership and the two one-mode network relations taken individually and then together. As shown in Fig. 7.10a, players in a guild tend to chat with other players in guilds; players not in a guild are not likely to chat with each other. Circles denote Node *i*'s attributes in the Node *i* and Node *j* dyad; triangles denote Node *j*'s attributes in Node *i* and Node *j* dyad.

Similar results are observed with the collaboration relation, presented in Fig. 7.10b. Players in guilds tend to collaborate with others in guilds. Players who do not belong to a guild are not likely to collaborate with each other. When both

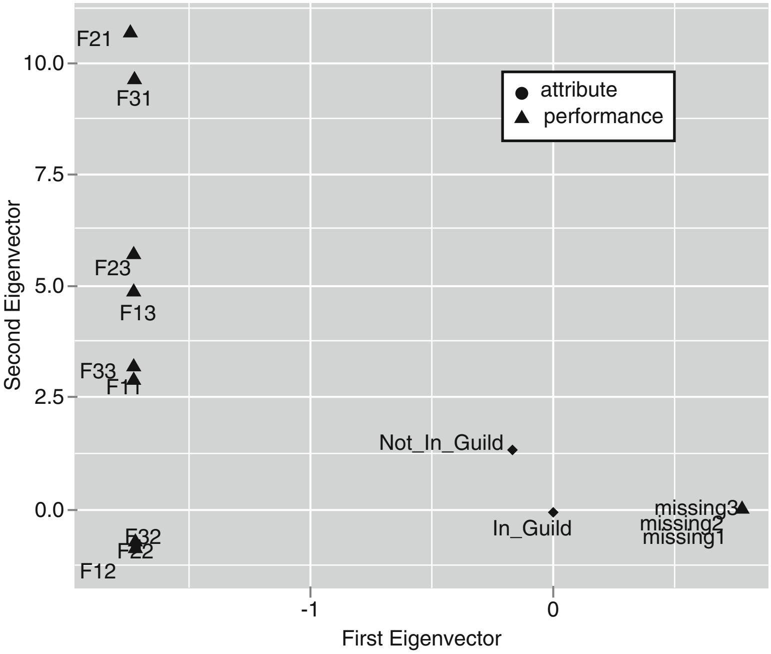


Fig. 7.11 Multiple correspondence analysis of guild affiliation and performance controlling for multiple network relations

relations are included together with guild membership, we found (see Fig. 7.10c) that they point in the same direction as the effects for each relation (i.e., guild members collaborate with and chat with others in guilds). In other words, the effects of each of the two relations while controlling for the other are similar to the effects of the other relation. Using the analogy of regression models, it means that the interaction effects are in the same direction as the main effects. For all three analyses, the first two dimensions explain 100 % of the total variance, because there are only two levels for one of the variables, and additional dimensionality is not possible. Hence, we omit the tables for principal inertias/eigenvalues.

Next, we conducted a multiple correspondence analysis with the team affiliation network (AB), the chat relation (AA1), the collaboration relation (AA2) and the two attribute variables, individual level guild affiliation and team level performance. Figure 7.11 shows the plot of the first two dimensions; and Table 7.5 shows the numerical breakdown of the analysis. The first two dimensions explain 66.3 % of the total variance. It is instructive to compare the association of guild affiliation and performance in Fig. 7.11 controlling for the two one-mode relations (chat and collaboration) with the association of guild affiliation and performance, shown in Fig. 7.9, that did not control for the two one-mode relations. The associations of guild affiliation and performance observed in Fig. 7.9 disappear in Fig. 7.11, when controlling for the two one-mode relations. This suggests that at least part of the

Table 7.5 Multiple correspondence analysis of individual guild affiliation and performance controlling for multiple network relations; principal inertias (eigenvalues)

Dimension	Value	Percent variance explained	Cumulative percent variance explained	Screen plot
1	0.56	54.3	54.3	*************************
2	0.12	12.0	66.3	**********************
3	0.11	10.9	77.2	************
4	0.07	6.6	83.8	*********
5	0.06	5.7	89.5	********
6	0.06	5.6	95.1	***
7	0.03	3.0	98.1	
8	0.02	1.9	100	

previously observed relation between guild affiliation and mid-level performance are captured by the frequent chat and collaboration between individuals belonging to guilds. Thus, if we control these relations, the previously observed relationship disappears.

It is important to consider the stability of the results. One source of instability is sampling variability. We have no reason to suspect this being a source of lack of stability of the solution because it is highly unlikely, given the data, that any one point contributes substantially greater to the solution. Further, we did not perform hypothesis testing, so that source of the lack of stability can also be eliminated. Sampling variability could also occur because, given the large size of the overall dataset, we randomly sampled from a wider population (Greenacre 2010). However, we repeated the analysis several times with different subsamples and obtained very similar results. For parsimony, we are not reporting the results here, but we have confirmed that the results are consistent. Yet another source of stability stems from the potential inadequacy of using the two-dimension plot to capture the association among the variables of interests. As suggested in Greenacre and Hastie (1987) and Roberts (2000), the quality of these two-dimension plots are generally "pessimistic," especially for multiple correspondence analysis, despite the fact that sometimes, the first two dimensions account for a seemingly not very high portion of the total inertia (such as 50.7 %, as in Table 7.4).

Developing Hypotheses from Preliminary Results

Multiple correspondence analysis, as described above, is a useful tool for hypotheses generation and subsequent testing with other more sophisticated confirmatory methods, such as p*/ERGM models (Robins et al. 2007a; Wang et al. 2013; Wasserman and Pattison 1996). Here its use is demonstrated on the example of the results in Fig. 7.9, relating individual guild affiliation with team performance.

As shown in Fig. 7.9, guild affiliation is associated with mid-level short-term and long-term performance (F12, F22), and also with mid-level negative performance (F32). When not in guilds, individuals can exhibit very high or very low short-term performance, and they usually show higher negative performance (face more deaths). In other words, belonging to a guild is associated with a mid-level performance, without any extremes of excellent results or higher deaths. Theoretically this might suggest that affiliating with an organization (in this case, a guild) serves to moderate a freelancer's performance. Affiliation buffers against very poor performance by leveraging the benefits of coordinating with, and gaining insights from, other guild members. But it might also stymie very high performance because of the costs incurred in coordinating with others. Therefore, a hypothesis deduced from this curvilinear reasoning would be as follows:

Hypothesis 1: Belonging to a guild reduces the chances of extreme high or low performance.

To test this hypothesis, we fitted exponential random graph models (*p** models, Wasserman and Pattison 1996) using MPNet software package. The essence of p* modeling is comparing the network under consideration with a series of random networks generated on the same set of nodes, to evaluate which network statistics are statistically more or less likely to result in a distribution of networks in which the observed network is very likely to occur. Model coefficients are logit-coefficients, with the dependent variable indicating the log odds of a tie between two existing nodes. Positive coefficients indicate that an attribute is more likely to appear in the observed network than could be expected by chance; negative coefficients indicate the opposite (Robins et al. 2007b).

We fit three models separately for the interaction between the individual attribute, guild affiliation, and the short-term team performance measure. Results are presented in Table 7.6, and we focus on interpreting results related to attribute variables. It can be seen from the table that, across all three performance levels, guild affiliation has positive effects, i.e., individuals with guild affiliation are more likely to join teams than those without guild affiliation. This effect needs to be controlled when considering the interaction effects of individual guild affiliation and team performance. We also control the effects of team performance in all three models using the three team performance effect terms. After controlling the structure effects (XEdge and XStar2A, i.e., density of the affiliation network and the stars in the affiliation network), the converged p* model results show that, across all three performance levels, there are negative and significant relationships (at $p < .1$ level) between guild affiliation and performance. A negative relationship, as explained above, means that the tie is less likely in the observed network than in a randomly generated network with the same nodes. Across the three performance levels, we find, that the association of guild affiliation with high-level and low-level performance are less likely to be observed than with mid-level performance (with parameters of and for low- and high-level performance vs. for mid-level performance). These results indicate that belonging to a guild reduces the chances of extreme performance on either end, supporting, in effect, Hypothesis 1.

The relationships explored and confirmed may not always seem intuitive. In fact, common sense indicates that guild affiliation – belonging to a group – may

Table 7.6 p* model results: guild affiliation and performance

Effects	Estimates	Standard errors
Low-level performance		
XEdge	−3.41	0.25
XStar2A	−0.20	0.05
Individual in-guild effect	0.18	0.20
Team low-level performance effect	0.61	0.37
In-guild low-level performance matching effect	−0.57	0.35
Mid-level performance		
XEdge	−3.63	0.36
XStar2A	−0.20	0.05
Individual in-guild effect	0.21	0.35
Team middle-level performance effect	0.31	0.30
In-guild middle-level performance matching effect	−0.09	0.31
High-level performance		
XEdge	−3.39	0.27
XStar2A	−0.21	0.05
Individual in-guild effect	0.21	0.22
Team high-level performance effect	0.19	0.31
In-guild high-level performance matching effect	−0.56	0.31

Note: t-statistics = (observation − sample mean)/standard error; SACF (sample autocorrelation)

increase one's performance. In our example, this is clearly not the case. Multiple correspondence analysis results, presented in Fig. 7.9, prompted us to test a more nuanced claim about the nature of the relationship between guild affiliation and performance. We then considered possible theoretical explanations and formulated this reasoning into a hypothesis and tested it with a more sophisticated confirmatory network analytic technique. This illustrates how the use of MCA as an exploratory tool to assist hypothesis formulation, especially with large datasets or multiple attributes, substantially simplifies the analysis process. While we conducted the confirmatory test using the same data sample, a stronger claim could be made if the results were tested on different samples and found to be consistent. On the other hand differences across data sets might prompt a second level of theorizing about explanations for potential differences. The second example reported below illustrates exploration of those differences.

Example 2: Dragon Nest

The second example helps us explore the impact of cultural differences on how teams form and perform. We begin by reviewing theoretical considerations of cultural differences that can inform an exploratory analysis; next, we utilize the analytical methods described above, to explore cultural differences by comparing the predominantly US-based EQ2 data with the predominantly Chinese data collected from the MMOG Dragon's Nest.

Hinds et al. (2011) lament that cultural differences are often ignored in the studies of global collaboration and recommend a more explicit and nuanced inclusion of cultural differences in studies of teams. They suggest that studies of collaboration in teams could benefit from consideration of cultural differences. A cultural dimension potentially influencing how teams form and perform is tightness-looseness (Gelfand et al. 2011). This dimension distinguishes cultures on the basis of the degree to which they have many social norms and low tolerance for deviant behavior (tight) versus few social norms and higher tolerance for deviant behavior (loose). Forming teams based on social norms are more likely to occur in tight cultures, such as China, with many social norms and low tolerance for deviant behavior than among those in loose cultures, like the US, where individuals might be more likely to form teams driven by performance considerations.

Data and Sample

Data for this example were obtained from the MMOG Dragon's Nest (DN), a fantasy game where players form teams in order to advance their characters and travel into dungeons. Unlike the EQ2 dataset, where players were predominantly from US, players in DN were predominantly from China. This offered the opportunity to uncover the influence of cultural differences on how teams form and perform.

This dataset has the same characteristics as EQ2 – a sample of "big data," with a nesting of levels reflecting players being members in multiple teams. The data were collected during 1 week, January 24–30, 2011, from Zone 101 of the game, and the dataset contained information on 304 persons from 217 teams. One distinguishing feature of this dataset is that teams are smaller, 2–4 people, so selecting a partner is perhaps more strategic than in EQ2.

Variables

Variable selection was very similar to the way it was carried out in the first example. Variables of interest included guild diversity (categorical, three-level variable), individual guild affiliation (binary – either in-guild or not-in-guild), team performance (categorical, three levels, equivalent to the short-term performance in the first example). The only network relation examined here was the person-team two-mode network.

Analytic Method

Correspondence analysis was performed in the same manner as in the first example, using R packages *ca* (Nenadic and Greenacre 2007), *ggplot2* (Wickham 2012).

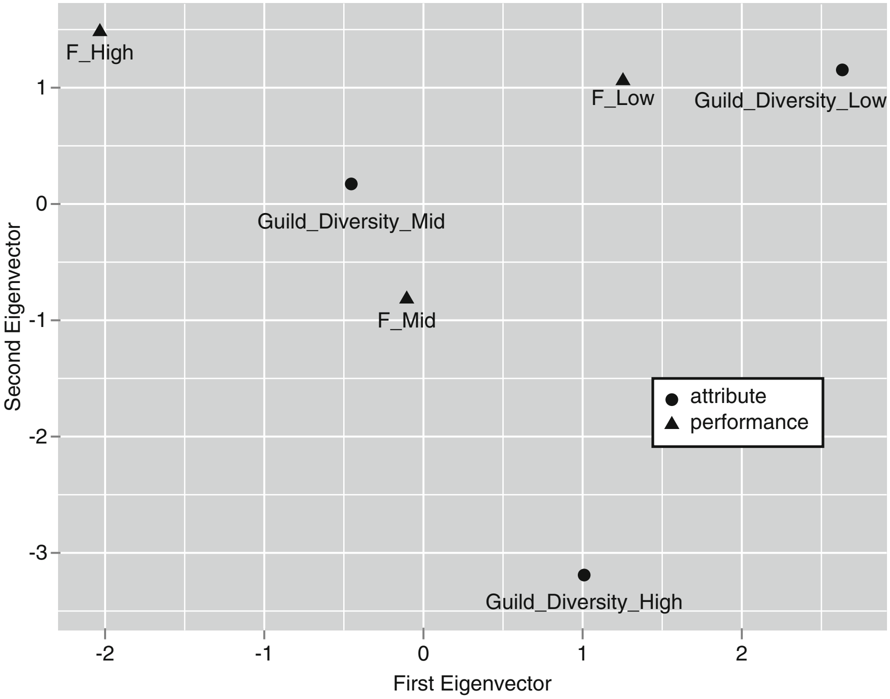


Fig. 7.12 Multiple correspondence analysis of guild diversity and performance

Because the primary purpose of this example was to demonstrate the cultural differences that could be discovered using multiple correspondence analysis, we focus more on the comparison with the first example than with the explanation of the analytic techniques already described previously.

Results

Team Performance and Guild Diversity: As in the first example, we plotted the results of the first two dimensions obtained using multiple correspondence analysis, presented below in Fig. 7.12. Triangles are attribute variables, and the squares are the performance variables. The first two dimensions explained 66.4 % of the total observed variance. Similar to the first example, the locations of the variables show the relations among them; when variables are near to each other, it indicates closer relations.

From the plot it is clear that unlike the EQ2 example, where most data points were clustered around the mid-level guild diversity, and all three performance factors were at mid-level, the data in this example is more evenly distributed between different performance levels. In other words, something other than performance alone keeps people affiliated with their respective guilds. This could serve as a demonstration of the cultural context described above: people in tight cultures

Table 7.7 Multiple correspondence analysis of guild diversity and performance; principal inertias (eigenvalues)

Dimension	Value	Percent variance explained	Cumulative percent variance explained	Screen plot
1	0.37	36.30	36.30	************************
2	0.31	30.10	66.40	******************
3	0.20	19.00	85.40	*****
4	0.15	14.60	100.00	

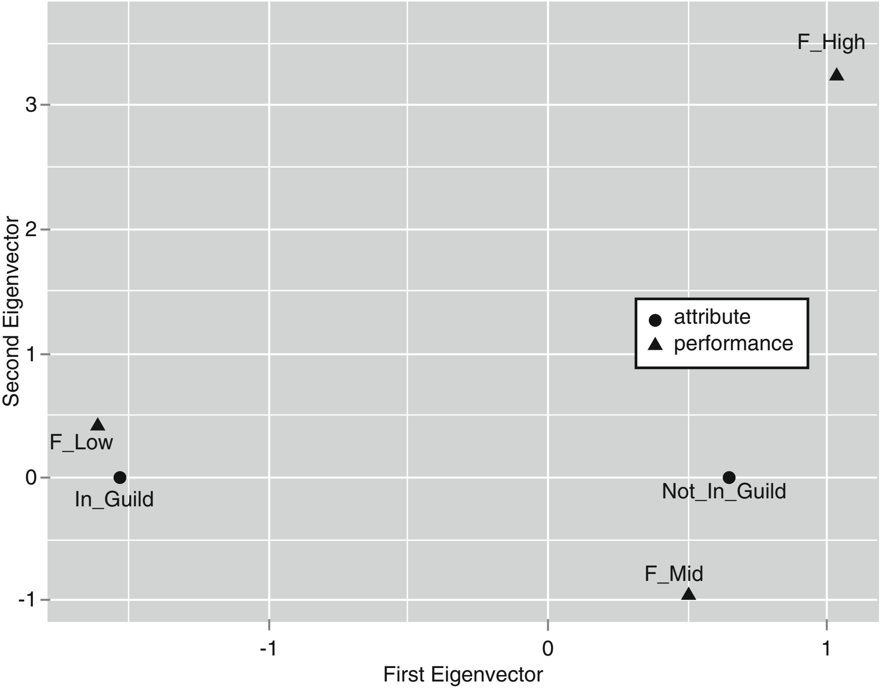


Fig. 7.13 Multiple correspondence analysis of guild affiliation and performance

(as this is an example comprised mostly of Chinese players) tend to form teams based on social norms, rather than expectations of performance typical for loose cultures (as in the US-based EQ2 example).

Table 7.7 shows the numerical breakdown of the analysis. The first two dimensions explained 66.4 % of the observed variance, and unlike the EQ2 example with eight dimensions, there were only four dimensions in total, which, again, could possibly account for cultural differences between tight and the loose cultures.

Team Performance and Individual Guild Affiliation: Similar to the EQ2 example, attributes at the individual level were included in the analysis in the same way as the team-level variables. Figure 7.13 shows the plot of the first two dimensions from the correspondence analysis; Table 7.8 shows the numerical breakdown of the analysis.

Table 7.8 Multiple correspondence analysis of guild affiliation and performance; principal inertias (eigenvalues)

Dimension	Value	Percent variance explained	Cumulative percent variance explained	Screen plot
1	0.44	54.50	54.60	************************
2	0.25	31.20	85.70	**********
3	0.11	14.30	100.00	

Again, as in the previous example, it is clear that there are pronounced differences in the way that people form teams, and that this influences how they perform. Here, the affiliation with the guild actually results in lower performance, with a mid-level performance associated more closely with not being in a guild. As results in Table 7.8 show, there are actually only three dimensions to the data, with the first two dimensions explaining over 85 % of the variance. Here, again, it is apparent that social norms dictate how people form into teams, which is quite different, culturally, from the EQ2 example where performance dictated how people formed teams.

As a result, we were able to use multiple correspondence analysis as a preliminary data analysis tool to explore cultural differences between how teams form and perform in the US and China. More specifically, using MCA, one can explore large datasets before engaging in time-consuming tasks of testing more complex models with control variables.

Conclusions

In this study, we illustrate the advantages of using correspondence analysis as an exploratory method for analyzing relational data at multiple levels using examples from two massively multi-player online games. Correspondence analysis incorporates relations among the observations and includes both the relational ties and attributes at multiple levels. The results from correspondence analysis can also be visually represented in easily accessible plots. Relations among various variables as well as observed raw data can be shown in the same plot although in some cases it might be useful to suppress visualizing the raw data. The plots display more information than just means and standard errors as seen in regression model results. With these advantages, correspondence analysis can be used as an important exploratory tool to examine the features of multilevel network datasets and the relations among variables of interest. The insights drawn from this exploratory technique serve as a theoretical and data reduction preamble for further analysis that can then be carried out using other more sophisticated confirmatory methods, such as p^*/ERGM models (Frank and Strauss 1986; Robins and Pattison 2005; Wasserman and Pattison 1996). As we have demonstrated with our examples, generating hypotheses from MCA results and testing them by fitting p^* models is much more streamlined with the use of MCA. Ideally, with the proliferation of big data, the confirmatory tests can be analyzed using separate but similar data sources. The proliferation of big data also increases the opportunity of using MCA to explore differences such as data gathered across different cultures.

This method also has some limitations. For instance, correspondence analysis requires categorical (or frequency) data. Some information is unavoidably lost during the transformation. The plots created using the scores from the first two dimensions are a projection of the higher dimensional data, which can lead to misinterpretation by the human eye. Furthermore, the magnitude of distances between points in the display does not indicate connections in the relation network. As with many exploratory approaches, the visualization of the results can be viewed in different ways by different people. Given the preliminary nature of this study, we recognize these limitations. Our goal is not to draw conclusions about the relationships of the study variables, but to use correspondence analysis as a way of developing hypotheses and models that can then be tested using subsequent techniques, such as p*/ERGM.

Acknowledgements The authors wish to acknowledge the Army Research Laboratory W911NF-09-2-0053 and the Army Research Office W911NF-12-1-0176 that contributed to the development of materials presented in this chapter. The views and conclusions contained in this document are those of the authors and should not be interpreted as representing the official policies, either expressed or implied, of the Army Research Laboratory or the Army Research Office or the U.S. Government. The U.S. Government is authorized to reproduce and distribute reprints for Government purposes notwithstanding any copyright notation here on. The book chapter was also prepared within the framework of a subsidy granted to the Higher School of Economics, Moscow, by the Government of the Russian Federation for the implementation of the Global Competitiveness Program.

References

Berends, H., van Burg, E., & van Raaij, E. M. (2011). Contacts and contracts: Cross-level network dynamics in the development of an aircraft material. *Organization Science, 22*, 940–960.

Blau, P. (1977). *Inequality and heterogeneity*. New York: Free Press.

Borgatti, S. P., & Everett, M. G. (1987). Network analysis of 2-mode data. *Social Networks, 19*, 243–269.

Borgatti, S. P., & Halgin, D. S. (2011). Analyzing affiliation networks. In P. Carrington & J. Scott (Eds.), *The Sage handbook of social network analysis* (pp. 417–433). London: Sage Publications.

Borgatti, S., Everett, M. G., & Freeman, L. C. (2002). *Ucinet for Windows: Software for social network analysis*. Harvard: Analytic Technologies.

Borgatti, S. P., Everett, M. G., & Johnson, J. C. (2013). *Analyzing social networks*. Thousand Oaks: SAGE Publications Limited.

Boyd, D., & Crawford, K. (2011). *Six provocations for big data*. Retrieved from http://papers.ssrn.com/sol3/papers.cfm?abstract_id=1926431

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

Denton, H. G. (1997). Multidisciplinary team-based project work: Planning factors. *Design Studies, 18*, 155–170.

Doreian, P., Batagelj, V., & Ferligoj, A. (2004). Generalized blockmodeling of two-mode network data. *Social Networks, 26*, 29–53.

Faust, K. (1997). Centrality in affiliation networks. *Social Networks, 19*, 157–191.

Faust, K. (2005). Using correspondence analysis for joint displays of affiliation networks. In P. J. Carrington, J. Scott, & S. Wasserman (Eds.), *Models and methods in social network analysis* (pp. 117–147). New York: Cambridge University Press.

Frank, O., & Strauss, D. (1986). Markov graphs. *Journal of the American Statistical Association, 81*(395), 832–842.

Gelfand, M., et al. (2011). Differences between tight and loose cultures: A 33 nation study. *Science, 332*, 1100–1104.

Gilula, Z., & Haberman, S. J. (1988). The analysis of multivariate contingency tables by restricted canonical and restricted association models. *Journal of the American Statistical Association, 83*(403), 760–771.

Greenacre, M. (1984). *Theory and applications of correspondence analysis*. London: Academic.

Greenacre, M. (2010). *Correspondence analysis in practice*. London: CRC Press.

Greenacre, M., & Hastie, T. (1987). The geometric interpretation of correspondence analysis. *Journal of the American Statistical Association, 82*, 437–447.

Hinds, P., Liu, L., & Lyon, J. (2011). Putting the global in global work: An intercultural lens on the practice of cross-national collaboration. *The Academy of Management Annals, 5*, 135–188.

Holland, P. W., & Leinhardt, S. (1981). An exponential family of probability distributions for directed graphs (with discussion). *Journal of the American Statistical Association, 76*, 33–65.

Kauffeld, S. (2006). Self-directed work groups and team competence. *Journal of Occupational and Organizational Psychology, 79*(1), 1–21.

Klein, K. J., & Kozlowski, S. W. (Eds.). (2000). *Multilevel theory, research, and methods in organizations: Foundations, extensions, and new directions*. San Francisco: Jossey-Bass.

Latapy, M., Magnien, C., & Del Vecchio, N. (2008). Basic notions for the analysis of large two-mode networks. *Social Networks, 30*(1), 31–48.

LaValle, S., Lesser, E., Shockley, R., Hopkins, M. S., & Kruschwitz, N. (2011). Big data, analytics and the path from insights to value. *MIT Sloan Management Review, 52*(2), 21–31.

McAfee, A., & Brynjolfsson, E. (2012). Big data: The management revolution. *Harvard Business Review, 90*(10), 60–66.

O'Leary, M., Mortensen, M., & Woolley, A. (2011). Multiple team membership: A theoretical model of its effects on productivity and learning for individuals and teams. *Academy of Management Review, 36*, 461–478.

Nenadic, O., & Greenacre, M. (2007). Correspondence analysis in R, with two- and three-dimensional graphics: The ca package. *Journal of Statistical Software, 20*(3). Retrieved from http://www.jstatsoft.org/v20/i03/

Newman, M. (2010). *Networks: An introduction*. Oxford: Oxford University Press.

Parkhe, A. (1993). Strategic alliance structuring: A game theoretic and transaction cost examination of interfirm cooperation. *Academy of Management Journal, 36*, 794–829.

Payne, G. T., Moore, C. B., Griffis, S. E., & Autry, C. W. (2011). Multilevel challenges and opportunities in social capital research. *Journal of Management, 37*, 491–520.

Phelps, C., Heidl, R., & Wadhwa, A. (2012). Knowledge, networks, and knowledge networks a review and research agenda. *Journal of Management, 38*(4), 1115–1166.

Roberts, J. M. (2000). Correspondence analysis of two-mode network data. *Social Networks, 22*(1), 65–72.

Robins, G., & Pattison, P. (2005). Interdependencies and social processes: Dependence graphs and generalized dependence structures. In P. J. Carrington, J. Scott, & S. Wasserman (Eds.), *Models and methods in social network analysis*. New York: Cambridge University Press.

Robins, G. L., Snijders, T., Wang, P., Handcock, M. S., & Pattison, P. E. (2007a). Recent developments in exponential random graph (p*) models for social networks. *Social Networks, 29*(2), 192–215.

Robins, G., Pattison, P., Kalish, Y., & Lusher, D. (2007b). An introduction to exponential random graph (p*) models for social networks. *Social Networks, 29*(2), 173–191.

Skvoretz, J., & Faust, K. (1999). Logit models for affiliation networks. *Sociological Methodology, 29*, 253–280.

Wageman, R., Gardner, H., & Mortensen, M. (2012). The changing ecology of teams: New directions for teams research. *Journal of Organizational Behavior, 33*(3), 301–315.

Wang, P., Robins, G., Pattison, P., & Lazega, E. (2013). Exponential random graph models for multilevel networks. *Social Networks, 35*(1), 96–115.

Wasserman, S., & Faust, K. (1989). Canonical analysis of the composition and structure of social networks. *Sociological Methodology, 19*, 1–42.

Wasserman, S., & Faust, K. (1994). *Social network analysis: Methods and applications*. Cambridge: Cambridge University Press.

Wasserman, S., & Pattison, P. E. (1996). Logit models and logistic regressions for social networks: I. An introduction to Markov graphs and p*. *Psychometrika, 61*(3), 401–425.

Wasserman, S., & Robins, G. (2005). An introduction to random graphs, dependence graphs, and p*. In P. J. Carrington, J. Scott, & S. Wasserman (Eds.), *Models and methods in social network analysis* (pp. 148–161). New York: Cambridge University Press.

Wasserman, S., Faust, K., & Galaskiewicz, J. (1990). Correspondence and canonical analysis of relational data. *Journal of Mathematical Sociology, 1*, 11–64.

Wellman, B. (1988). Thinking structurally. In B. Wellman & S. D. Berkowitz (Eds.), *Social structures: A network approach*. Cambridge: Cambridge University Press.

Wickham, H. (2012). *ggplot2* [Computer software]. Retrieved from http://ggplot2.org/

Zhao, Z. J., & Anand, J. (2013). Beyond boundary spanners: The ‘collective bridge’ as an efficient interunit structure for transferring collective knowledge. *Strategic Management Journal, 34*, 1513–1530.