On the use of Multiple Correspondence Analysis to visually explore affiliation networks

Maria Rosaria D'Esposito a, Domenico De Stefano b, *, Giancarlo Ragozini c

a Department of Economics and Statistics, University of Salerno, Via Ponte don Melillo, 84084 Fisciano (SA), Italy
b Department of Economics, Ca' Foscari University of Venice, Cannaregio 873, 30121 Venezia, Italy
c Department of Political Sciences, University of Naples Federico II, Via Rodinò, 22, 80138 Napoli, Italy

A R T I C L E   I N F O

Keywords:
Affiliation networks
Two-mode structural equivalence
Correspondence Analysis
Doubling
Actor and event covariates

A B S T R A C T

In this paper we discuss the use of Multiple Correspondence Analysis to analyze and graphically represent two-mode networks, and we propose to apply it in a Greenacre's doubling perspective. We discuss how Multiple Correspondence Analysis: (i) properly takes into account the nature of relational data and the intrinsic asymmetry of actors/events in two-mode networks; (ii) allows a proper graphical appraisal of the underlying relational structure of actors or events; (iii) makes it possible to add actor and event attributes to the analysis in order to improve results interpretation; and (iv) gives different results with respect to the usual Simple Correspondence Analysis.

© 2014 Elsevier B.V. All rights reserved.

1. Introduction

Analysis of the collaboration and affiliation structures arising among different subjects is now more and more of interest in both social and economic sectors. In general, collaboration and affiliation networks are characterized by a set of actors and a set of events in which the actors are involved. These elements give rise to so-called two-mode networks, which contrast with the more widely used and well known one-mode networks.

In the analysis of two-mode networks, Borgatti and Halgin (2011) identify two main approaches: the conversion approach and the direct approach. In the former, a two-mode network is converted into two one-mode networks and the analysis focuses on one mode at time, though this may result in the loss of some information (Borgatti and Everett, 1997; Everett and Borgatti, 2013). The direct approach, which we adopt in this paper, considers instead the two-mode network as it is, and the two modes are jointly analyzed.

Dealing with a two-mode network, one issue is to visualize it and make a graphical appraisal of the underlying relational structure. With this aim in mind, the direct approach mainly makes use of the bipartite graph and spring embedding (Borgatti and Halgin, 2011) or of factorial methods for qualitative data such as Correspondence Analysis (CA) (Benzécri, 1973; Greenacre, 1984). This latter derives low dimensional spaces in which it is possible to represent points corresponding to the actors and the events in order to evaluate similarities of participation/attendance patterns (Wasserman et al., 1989; Faust, 2005; Borgatti and Halgin, 2011).

CA can be applied if we consider the affiliation matrix, corresponding to a two-mode network, either as a two-way contingency table, or as a two-way case-by-variable matrix. The first approach has often been adopted and results in a simple correspondence analysis of the affiliation matrix. This use has given rise to a debate on its pros and cons (Bonacich, 1991; Borgatti and Everett, 1997; Roberts, 2000; Borgatti and Halgin, 2011; Borgatti et al., 2013). In the present paper, we focus on the second approach, i.e. we look at the affiliation matrix as a case-by-variable matrix. Within this framework, in our opinion a proper approach is Multiple Correspondence Analysis (MCA), the extension of CA to the case of many categorical variables (Blasius and Greenacre, 1994, 2006). Given the characteristics of two-mode networks, in this paper we propose to apply MCA by using the complete disjunctive coding in Greenacre’s doubling perspective (Greenacre, 1984) and we discuss how this version of MCA can be used to analyze and graphically represent two-mode networks. The interest in this technique is motivated thus: (i) MCA is designed to treat case-by-variable matrices, which are similar in structure to affiliation matrices; (ii) MCA assigns a

* Corresponding author. Tel.: +39 040 558 3041; fax: +39 040 558 7033.
E-mail addresses: mdesposi@unisa.it (M.R. D’Esposito),
domenico.destefano@unive.it (D. De Stefano),
giragoz@unina.it (G. Ragozini).

http://dx.doi.org/10.1016/j.socnet.2014.01.003
0378-8733/© 2014 Elsevier B.V. All rights reserved.
different role to actors and to events, thus allowing distinct features to be highlighted in each mode; (iii) MCA makes it possible to add covariates to the analysis in order to improve results interpretation; and (iv) MCA within the doubling perspective affords visualizations that can be easily interpreted in terms of similarities among actors/events network relations.

The paper is organized as follows. In Section 2 we present and discuss our approach to the use of MCA for affiliation networks, while in Section 3 we highlight the features of MCA useful for two-mode networks. A discussion on how MCA could incorporate external information in the analysis is given in Section 4. In Section 5 we discuss, through an analytical and experimental analysis, how MCA makes it possible to represent the degree of similarity of actor/event profiles and the differences existing with respect to simple CA. Section 6 offers some concluding remarks.

2. Multiple Correspondence Analysis for affiliation networks

Let \( U \) be an affiliation network. It consists of two sets of relationally connected units, actors and events, and can be represented by a triple \( (V_1, V_2, \mathcal{R}) \) composed of two disjoint sets of nodes, \( V_1 \) and \( V_2 \) of cardinality \( n \) and \( m \), and one set of edges or arcs, \( \mathcal{R} \subseteq V_1 \times V_2 \). By definition \( V_1 \cap V_2 = \emptyset \); the set \( V_1 = \{ a_1, a_2, \ldots, a_n \} \) represents the set of \( n \) actors whereas \( V_2 = \{ b_1, b_2, \ldots, b_m \} \) represents the set of \( m \) events. The edge \( r_{ij} = (a_i, b_j) \), \( r_{ij} \in \mathcal{R} \), is an ordered couple, and indicates whether an actor \( a_i \) attends an event \( b_j \). The set \( V_1 \times V_2 \) can be fully represented by the binary affiliation matrix \( F = (f_{ij}) \), \( i = 1, \ldots, n, j = 1, \ldots, m \), with \( f_{ij} = 1 \) if \((a_i, b_j) \in \mathcal{R}\) and 0 otherwise. Given \( F \), the row and column marginals \( f_{i.} = \sum_j f_{ij} \) and \( f_{.j} = \sum_i f_{ij} \) coincide with the degree \( d_i \) of the ith actor and the size \( s_j \) of the jth event, respectively, i.e. \( f_{i.} = d_i \) and \( f_{.j} = s_j \). To the best of our knowledge, the first attempt to apply MCA in a two-mode setting dates back to 1984 (Bourdieu, 1988). Bourdieu used Multiple Correspondence Analysis to analyze the universities, professional affiliations and background variables of Parisian professors. However, he did not treat his data matrix as an affiliation matrix because he considered the affiliations of professors as properties of the individuals and not as relations (De Nooy, 2003). After this unintentional use to analyze affiliation matrices, a particular version of MCA (Carroll et al., 1986) has been used in social network analysis (Wasserman et al., 1989; Faust, 2005). The CA algorithm is here applied to an “edges by actors and events matrix”, where the edges \( r_{ij} = (a_i, b_j) \) are assumed as units of the analysis and the affiliation matrix \( F(n \times m) \) is transformed into a multiple indicator matrix with \( n + m \) columns (one out of \( n \) for each actor and one out of \( m \) for each event) and as many rows as the edges, i.e. the rows correspond to the 1’s in the affiliation matrix and amount to the total number of ties \( L \) (Faust, 2005). However, the scaling proposed by Carroll et al. (1986) did not receive a general consensus and has been questioned in the statistical literature (Carroll et al., 1987, 1989; Greenacre, 1989).

In the present paper, we look at actors as observational units and at the participation in events as dichotomous categorical variables. The \( F \) matrix is, therefore, a multivariate case-by-variable data matrix (Gower, 2006), in which a different status is assigned to the rows and columns – in line with the duality perspective (Faust, 2005; Breiger, 1974).

We propose to perform MCA by applying the usual CA algorithm – SVD of the doubled normalized and centered profile matrix – to the multiple indicator matrix, or simply indicator matrix, \( Z \) derived from \( F \) through a full disjunctive coding.

In MCA for measurement data, each row of \( Z \) represents an observation and each column represents one modality of each categorical variable, with \( z_{ij} = 1 \) if the jth modality of the qth categorical variable is present in the ith case, 0 otherwise. In the case of Q categorical variables the indicator matrix is \( Z = \{ Z_1, \ldots, Z_Q \} \). If the qth variable has \( J_q \) categories, \( Z_q \) is \( I \times J_q \), and \( J = \sum Q_{q=1}^{Q} J_q \) is the total number of categories. In this case the row margins are constant and equal to \( Q \), while the column margins correspond to the category frequencies.

In the case of affiliation matrix \( F \), in order to apply the full disjunctive coding, we think of each event \( e_i \) as a dichotomous variable with categories \( e_i^1 \) and \( e_i^2 \). In the full disjunctive coding, each category is describing by a dummy variable, i.e. \( e_i^j \) is a dummy variable coding the participation in the event, and \( e_i^j \) is a dummy variable coding the non-participation. \( Z \) will then contain two orthogonal columns for each \( e_i \) (Fig. 1).

The \( Z \) matrix is an \( n \times 2m \) matrix of the form: \( Z = [F^+, F^-] \), where \( F^+ = (e_i^1) \) and \( F^- = (e_i^2) \) is \( I \times F^- = I \), where \( I \) is an \( n \times m \) all-ones matrix.

The indicator matrix \( Z \) turns out to be a doubled matrix, and the MCA we are proposing is equivalent to the CA of this doubled matrix. This allows us to interpret results in a doubling perspective. The \( Z \) row marginals \( z_i \) are constant and equal to the number of events \( m \), while the column marginals \( s_j \) equal to the event size \( s_j \), when associated to \( e_i^1 \), or to \( n - s_j \), when associated to \( e_i^2 \).

The profile matrix \( P \) and weight matrices \( D_a \) and \( D_e \) proper of the CA algorithm and involved in the normalization are:

\[
P = Z/ nm, \quad D_a = \text{diag} \left( \frac{1}{n}, \ldots, \frac{1}{n} \right), \quad D_e = \text{diag} \left( \frac{z_1}{nm}, \ldots, \frac{z_m}{nm} \right).
\]

The centered and doubled normalized \( Z \) is the matrix \( S \)

\[
S = D_a^{-1/2} (Z/ nm - D_a 1^T D_e) D_e^{-1/2} = \sqrt{n} \left( Z/ nm - D_a 1^T D_e \right) D_e^{-1/2},
\]

where \((1/n)1\) is the vector of actor weights and \( 1^T D_e \) is the vector of the event weights.

The SVD of \( S \) gives:

\[
S = U \Sigma V^T
\]

where \( U \) is the diagonal matrix of singular values, and \( U, V \) are the matrices of the left and right singular vectors, respectively.

---

2 We choose this approach, instead of the mathematically equivalent CA algorithm applied to the squared super-matrix of cross-tables (the so-called Burt matrix), in order to highlight the role of actors and to emphasize the relation of this approach with the doubling perspective. The idea behind doubling is the allocation of two complementary sets of data to a given rating scale, one labeled as the “negative” pole of the scale and the other as the “positive” pole (Greenacre, 1984).
The principal coordinates for the row and column categories, respectively, are defined as:

\[ \Phi = D_\alpha^{-1/2} \Sigma = \sqrt{n}U \Sigma \]  
(1)

\[ \Psi = D_\varepsilon^{-1/2} \Sigma \]  
(2)

and the standard coordinates for the row and column categories, respectively, are defined as:

\[ \Gamma = \sqrt{n}U \]  
(3)

\[ \Delta = D_\varepsilon^{-1/2} \varepsilon \]  
(4)

Usually, the first two columns of \( \Phi \) and \( \Psi \), or of \( \Gamma \) and \( \Delta \), are used to construct two-dimensional maps in which the data are graphically represented. A more thorough understanding of what MCA can visually render may be reached by looking at the geometry behind the method and at the interpoint distances between rows and between columns in the original spaces and in the projected two-dimensional maps. In the original spaces, the distances between actor profiles

\[ p_{ij} = \left[ \frac{z_{ij}}{m}, \ldots, \frac{z_{ij}}{m} \right], \]  
(5)

and event profiles

\[ p_{ej} = \left[ \frac{z_{ej}}{s_j}, \ldots, \frac{z_{ej}}{s_j} \right] \forall e_j \]  
(6)

\[ p_{ej} = \left[ \frac{z_{ej}}{(n-s_j)}, \ldots, \frac{z_{ej}}{(n-s_j)} \right] \forall e_j \]  
(7)

in the usual \( \chi^2 \) metric proper of CA, are respectively given by:

\[ \delta^2_{\text{MCA}}(p_{ij}, p_{ij}) = (p_{ij} - p_{ij})^T D_\alpha^{-1}(p_{ij} - p_{ij}) = \sum_{j=1}^{2m} \left( \frac{z_{ij}}{m} - \frac{z_{ij}}{m} \right)^2 \sum_{j=1}^{nm} \frac{n}{m} \sum_{j=1}^{n} \left( \frac{z_{ij}}{s_j} - \frac{z_{ij}}{s_j} \right)^2 \]  
(8)

\[ \delta^2_{\text{MCA}}(p_{ej}, p_{ej}) = (p_{ej} - p_{ej})^T D_\varepsilon^{-1}(p_{ej} - p_{ej}) = \sum_{j=1}^{n} \left( \frac{z_{ej}}{s_j} - \frac{z_{ej}}{s_j} \right)^2 \]  
(9)

where \([\cdot]\) is the Iverson bracket such that:

\[ [P] = \begin{cases} 
1 & \text{if } P \text{ is true}; \\
0 & \text{otherwise}. 
\end{cases} \]

The distance between two actor profiles is equal to zero when the actors participate in the same events. The distance is greater than zero when the participation patterns differ. The distance between two event profiles is equal to zero when the events are attended by the same actors. The distance is greater than zero when the attendance patterns are different.

A particular case of this distance occurs when comparing the events \( e_j \) and \( e_j' \). It is easy to show that Eq. (9) becomes:

\[ \delta^2_{\text{MCA}}(p_{ej}, p_{ej}) = \frac{n^2}{s_j(n-s_j)} \]  
(10)

This implies that the distance between \( e_j \) and \( e_j' \) is inversely related to the event size. This distance reaches its minimum value when \( s_j = n - s_j \), i.e. when half of the actors participate in the event \( e_j \) and half do not participate. The \( \chi^2 \) distances adhere to the distributional equivalence principle (Benzécri, 1973; Fichet, 2009). Two distributionally equivalent profiles will correspond to two coincident points in the original space, and vice versa. On the other hand, when profiles are not distributionally equivalent their \( \chi^2 \) distances are greater than zero.

In the two-dimensional maps derived through the first two principal coordinates, the distances between actors and between events optimally approximate the ones in the original spaces (Greenacre, 2006). Hence the analysis in the associated factorial plane provides insights with respect to the relative positions of actor/event profiles.

3. Some features of the MCA for affiliation networks

By exploiting the non-dyadic characteristic of two-mode networks, we can focus the analysis either on the actors represented as points in a space (event space) spanned by the principal axes having coordinates \( \Phi \) (Eq. (1)) or on the events represented as vectors in a space (actor space) spanned by the principal axes with coordinates \( \Psi \) (Eq. (2)). In particular, the event representation in the actor space affords special features and interpretations thanks to the adopted doubling principle.\(^\dagger\) Moreover, thanks to the analytical relation between the actor space and the event space, we can perform a joint representation of actors and events in a common space in order to analyze relationships between the two modes of the network.

In addition, by using the modalities of the attributes as supplementary information, MCA makes it possible to add actor or event attributes to the analysis in order to investigate the association between the relational patterns and the attributes. It is also possible to represent the position of groups of actors (or events) on factorial planes if cluster membership labels are available.

In the following, we exemplify the use of MCA for the visualization of an affiliation matrix \( F \) through the analysis of the southern women dataset. This is a well known dataset collected by Davis, Gardner and Gardner in their study of the social relationships of southern women (Davis et al., 1941). It is undoubtedly the dataset most commonly used by network analysts to explore the utility of new tools for analyzing affiliation networks (Breiger et al., 1975; Doreian, 1979; Bonacich, 1991; Freeman and White, 1993; Roberts, 2000; Freeman, 2003; Doreian et al., 2004). The study by Davis, Gardner and Gardner aimed at “examining the correspondence between people's social class levels and their patterns of informal interaction and . . . how much the informal contacts made by individuals were established solely (or primarily) with others at approximately their own class levels” (Freeman, 2003). To this aim, the authors collected data on the social activities of 18 women and their patterns of informal contacts. They focused on the attendance at 14 informal social events that took place over a nine-month period.

The affiliation matrix \( F \), which describes the data, is an \( 18 \times 14 \) individual-by-event matrix (Table 1).

3.1. Representing actors in the event space

In MCA applied to measurement data, even if each row has a profile and a position in the variable space, the representation of rows as points in such a space is not of interest. This is because rows are usually respondents to a questionnaire and there is no interest in identifying them. In our case, on the contrary, rows correspond to actors and it is thus a relevant issue to represent

\(^\dagger\) In MCA applied to a general indicator matrix, the meaning of the interpoint column distances has been questioned (Greenacre, 2006; Gower, 2006). In our case, such criticisms are overcome thanks to the equivalence of our approach with the doubling one.
Table 1
The 18 × 14 affiliation matrix describing the southern women data (Davis et al., 1941), with actor degrees, event sizes, and total number of ties (1) added.

<table>
<thead>
<tr>
<th></th>
<th>E1</th>
<th>E2</th>
<th>E3</th>
<th>E4</th>
<th>E5</th>
<th>E6</th>
<th>E7</th>
<th>E8</th>
<th>E9</th>
<th>E10</th>
<th>E11</th>
<th>E12</th>
<th>E13</th>
<th>E14</th>
<th>Degree</th>
</tr>
</thead>
<tbody>
<tr>
<td>Evelyn</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>8</td>
</tr>
<tr>
<td>Laura</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>7</td>
</tr>
<tr>
<td>Theresa</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>8</td>
</tr>
<tr>
<td>Brenda</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>7</td>
</tr>
<tr>
<td>Charlotte</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>4</td>
</tr>
<tr>
<td>Frances</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>4</td>
</tr>
<tr>
<td>Eleanor</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>4</td>
</tr>
<tr>
<td>Pearl</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>3</td>
</tr>
<tr>
<td>Ruth</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>4</td>
</tr>
<tr>
<td>Verne</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>4</td>
</tr>
<tr>
<td>Myra</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>4</td>
</tr>
<tr>
<td>Katherine</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>6</td>
</tr>
<tr>
<td>Sylvia</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>7</td>
</tr>
<tr>
<td>Nora</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>8</td>
</tr>
<tr>
<td>Helen</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>5</td>
</tr>
<tr>
<td>Dorothy</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>2</td>
</tr>
<tr>
<td>Olivia</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>2</td>
</tr>
<tr>
<td>Flora</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>2</td>
</tr>
<tr>
<td>Event size</td>
<td>3</td>
<td>3</td>
<td>6</td>
<td>4</td>
<td>8</td>
<td>8</td>
<td>10</td>
<td>14</td>
<td>5</td>
<td>4</td>
<td>6</td>
<td>3</td>
<td>3</td>
<td>1 = 89</td>
<td></td>
</tr>
</tbody>
</table>

them in a two-dimensional map by using the first two columns of the principal coordinates \( \Phi \) (Eq. (1)). In this map, each actor corresponds to a point and the interpoint distances best approximate the \( \chi^2 \) distances among actor profiles in the original space, and represent the actors' relative positions in the network. Two actors corresponding to two close points in the map have similar participation patterns. Actors corresponding to points close to the axes origin have a common participation habit (as they are close to the average actor profile located in the axes origin), while actors with corresponding points far from the center have an unusual

Fig. 2. MCA map of the southern women (dots) in the event space. The variance explained by the first axes is 40.2%, and by the second axis is 16.4%.
participation pattern. Isolated actors and group of actors can also be spotted. In Fig. 2 we note that Olivia and Flora collapse on the same point. This is due to the fact that they participate in exactly the same set of events. By looking at the relative positions of Katherine, Nora, and Sylvia, we observe they are close to each other as they show very similar attendance patterns.

Moreover, we note that two well separated actor subgroups appear on the first principal axis, with an intermediate position of Pearl, Ruth and Dorothy. This is in line with what is well-known in the literature (Freeman, 2003). It should moreover be noted that, if we take a detailed look at the MCA representation, the two groups may be split into subgroups on the basis of the coordinates on the second principal axis.

3.2. Representing events in the actor space

Due to the doubling approach, each event \( e_i \) in the actor space is represented by two opposite vectors, corresponding to the two poles \( e_i^+ \) and \( e_i^- \), lying on the same direction and passing through the origin. The length of each segment joining the two poles is proportional to the inverse of the product between the attendance and the non-attendance rates (Eq. (10)). The distances \( \delta_{ij}^+ \) and \( \delta_{ij}^- \) of the two poles from the origin are given by:

\[
\delta_{ij}^+ = \left(1 - \frac{s_{ij}}{n}\right); \quad \delta_{ij}^- = \frac{s_{ij}}{n}
\]  

(11)

Another way to think of this is to imagine the endpoints of each event segment as having weights proportional to the average rate of attendance of the respective poles. If one distance (\( \delta_{ij}^+ \) or \( \delta_{ij}^- \)) is much larger than the other, the event is highly polarized and it is either rare or common.

In addition, the cosine of the angle between two event segments is the “correlation” between participation patterns in the events. Then if two event segments form a small angle and present the positive pole on the same side, the corresponding events will have similar participation patterns. If the two segments form a small angle but the positive poles are opposite, the events will have opposite participation patterns. These characteristics are preserved when events are plotted in the two-dimensional maps obtained using the first two principal coordinates \( \Psi \) (Eq. (2)).

For the southern women dataset, in Fig. 3 we note that the most polarized events are E1, E2 and E13, E14, which are the

---

4 Freeman (2003) provides a consensual description of the “true structure” of the southern women dataset by performing a meta-analysis of 21 analyses of these data. Two actor subgroups are found: “Group A” consisting of Evelyn, Laura, Theresa, Brenda, Charlotte, Frances, Eleanor, Pearl, Ruth, and “Group B” comprising Verne, Myra, Katherine, Sylvia, Nora, Helen, Dorothy, Olivia, Flora.
rarest events. In particular, we can say that participation in these polarized events highly characterizes actors. As for the correlation among participation patterns of events, we find that the most negative correlated events are E5 and E9, which lie on the same direction but their poles are opposite. From a relational point of view, this means that actors attending one of them are unlikely to attend the other. In the map we can imagine two orthogonal diagonal dimensions characterized by two groups of events: E13, E14, E10, E12 in the first quadrant, and E1, E2, E3, E4, and E6 in the second quadrant. Finally, in the actor space the variance of event attendance can be appreciated. For instance, the higher the polarization, the lower the event attendance variance. For example, E13 and E14 are the events with the lowest variances.

3.3. Jointly representing actors and events

In order to represent actors and events in a joint two-dimensional map we can use the biplot (Gabriel, 1995). In the literature, for the case of measurement data, several proposals have been out forward in order to obtain the most appropriate representation of rows and columns in a unique space. The easiest way, as in simple CA, would be to superimpose the actor and the event representation using the first two principal coordinates for both (Eqs. (1) and (2)) obtaining a symmetric map. However, given the asymmetry of actors and events, we propose to keep it in the representation by using the asymmetric biplot where the actors are represented in principal coordinates (Eq. (1)) and the events are in standard coordinates (Eq. (4)) (Greenacre, 2010). The direction vector defined by each event is the biplot axis. By projecting the points representing actors onto each biplot axis we can achieve an approximate value of the event participation profile and explain why actors are close or far apart in the event space.\footnote{The reverse representation, i.e. events in principal coordinates (Eq. (2)) and actors in standard coordinates (Eq. (3)), is less meaningful.} For instance, in Fig. 4 we note that Katherine, Sylvia and Nora have the highest coordinates on the diagonal direction corresponding to the events E13, E14, E10, E12. They are far from the origin because of their attendance at the rare and high polarized events E13 and E14. On the other side, for similar reasons we find the group of Laura, Theresa, Evelyn and Brenda lying on the other diagonal direction corresponding to events E1, E2, E3, E4 and E6. We notice also that Flora and Olivia have the highest coordinates on the biplot axis determined by event E11 as they are strongly characterized by their participation in this rare event (they also participate in a common event E9). The group of women in the center of the map is characterized by their participation in less polarized and more common events.

Fig. 4. MCA joint: asymmetric biplot of actors (dots) in principal coordinates and events (dashed lines) in standard coordinates for the southern women data. Triangles represent event positive poles, and squares represent event negative poles. The variance explained by the first axes is 40.2%, and by the second axis is 16.4%.
4. The use of supplementary points to fit external information in the MCA representation

In order to characterize actors and events by their attributes, the indicator matrix $Z$ is augmented by adding a set of columns (rows) corresponding to the complete disjunctive coding of the qualitative variables used as attributes.

The additional information has no influence on the MCA solution, because it does not play any role in determining the $\delta_0^{MCA}$ and the principal axes (Greenacre, 1984). More precisely, supplementary rows and columns can be viewed as additional points with zero masses in the row or column profile spaces. Even though they do not play any role in the analytical determination of the factorial axes, they can be projected onto the factorial plane as additional information. Then, their positions can be used to support the interpretation of the configuration (Blasius and Greenacre, 1994).

To exemplify this aspect of the analysis, we use a dataset collected in the framework of R&D policy evaluation. Our dataset refers to a network consisting of 140 organizations (actors), belonging to the IMAST technological district, that are involved in 24 granted joint R&D research projects (events) (http://www.imast.it). The available attributes of actors and of events are:

- for actors
  - organizations typology, i.e. firms, research centers, other organizations;
  - type of membership, i.e. stable member or occasional partner;

- for the events
  - type of grant: national (MIUR and FIRB), European;
  - field of research: aerospace, transportation, nano tech, electronics, and others (the last four being classified as no aerospace).

In Fig. 5 the IMAST data are represented through a bipartite graph. From the graph we note four projects that attract heavily more partners than members, and not well-characterized participation patterns for the rest.

The biplot in Fig. 6 portrays on the same map the actors, the events, and the actor and event attributes. We note that stable members are on the left side and are mainly firms and research centers participating to national project (PR-MIUR), while all the partners that occasionally participate to a specific project. The participation pattern of partners is characterized by being mainly involved in European Projects. Looking at the vertical axis it is possible to note an association between aerospace projects and firms, while reaches centers are mainly involved in project non-related to aerospace. Looking at the events we can spot four

---

6 In the MCA of the indicator matrix, by adopting the method of Nenadic and Greenacre (2007), supplementary variables can be represented as averages of respondent points in principal coordinates. For example, from the transition formula

$$
\Psi = D_{ni}^{-1} Z_{nm} \Phi \Sigma^{-1} = D_{ni}^{-1} Z_{nm} \Gamma.
$$

which express the event principal coordinates as the weighted average of the actor standard coordinates with weights equal to the even profiles, the principal coordinates of actor attributes are expressed as:

$$
\Psi = D_{ni}^{-1} Z_{nm} \Phi \Sigma^{-1} = D_{ni}^{-1} Z_{nm} \Gamma.
$$

where $\Psi$, $D_{ni}^{-1}$, $Z$ are, respectively, the principal coordinates, the matrix of weights and the complete disjunctive coding of the actor attributes that are added.

7 Technological districts are organizational networks established between firms, universities and research centers in Italy since 2002.
groups of projects (three on the left and one on the right); in each group the attendance pattern is very similar as indicated by the size of angles the vectors corresponding to the events.

Note that, the use of additional information may be almost mandatory in the analysis of large affiliation matrices, in which assessing the role of single actors or events is difficult and/or pointless. In this case, it may be more useful to analyze the relational patterns of subgroups of actors or events while preserving all the information contained in the original affiliation matrix.

5. Correspondence Analysis and the appraisal of relational similarities

In the analysis of affiliation networks, one point of interest is the assessment of the degree of similarity of actors/events network profiles, i.e. the structural equivalence among actors/events (Lorrain and White, 1971; Faust and Romney, 1985; Faust, 1988; Kovacs, 2010). In a two-mode network, two actors are structurally equivalent if they participate in exactly the same events and two events are structurally equivalent if they are attended by the same actors (Borgatti & Everett, 1992; Pizarro, 2007). Formally, given two actors \(a_i\) and \(a_k\), the structural equivalence property \(S\) states that: \(a_i S a_k\) if and only if \(f_{ij} = f_{kj} \forall j\). If two actors \(a_i\) and \(a_k\) are structurally equivalent they are indistinguishable, and one equivalent actor can substitute for the other because the two relational patterns are the same. Similar statements hold for the events.

To measure the degree of similarity among actors/events two main approaches have been proposed for one mode networks: the first based on the Euclidean distance (Burt, 1980), the other on correlation\(^8\) (Breiger et al., 1975).

For a two-mode network described by an affiliation matrix \(F\), in the commonly used conversion approach the above distance measures are applied to the elements of the matrices \(FF\) and \(FF\). In the direct approach, in line with Burt (Burt, 1980), similarities among actors/events could be measured by the Euclidean distance between the corresponding numerical vectors:

\[
\delta_E^2(f_i, f_j) = \sum_{j=1}^{m} (f_{ij} - f_{ij})^2, \\
\delta_R^2(f_i, f_j) = \sum_{i=1}^{n} (f_{ij} - f_{ij})^2, \tag{14}
\]

\(^8\) The relation between these two distances has been extensively discussed (Faust and Romney, 1985; Faust, 1988; Michaelson and Contractor, 1992).
It is easy to show that the distances in Eq. (14) are linearly related to the usual distances evaluated on the matrices $FF$ and $FF$.

If two actors/events are structurally equivalent, the distances in Eq. (14) are equal to zero and vice versa. Deviations from 0 can be used to interpret imperfect structural equivalence. An increase in the distance between actors/events corresponds to their decreased similarity.

When factorial methods as CA or MCA are used to graphically represent similarities of actors or events, the appraisal of $\chi^2$ distances is called for, instead of the usual Euclidean distances in Eq. (14). It is worth studying to what extent the $\chi^2$ distances portrayed on the factorial planes differ from the Euclidean ones, and the reasons for such a difference. The interest in the Euclidean distance is related to the visualization goals of the analysis. Euclidean distance, in fact, is the “ordinary” distance between two points on a plane that can be immediately evaluated by the human eye and brain.

For the MCA case the $\chi^2$ distances are the ones in Eqs. (8) and (9). A comparison of the two of Eq. (14) with Eqs. (8) and (9) shows that, with respect to the Euclidean distance $\delta_L^2$, the $\delta_C^2$ distances are based on a weighting system which involves the sizes of the events. However, the event size $s_j$ has a counterpart in its complement $n - s_j$, and consequently the column weights are somehow balanced.

For the sake of comparison, we also analyze the behavior of the $\chi^2$ distance in the case of simple CA, i.e. when the algorithm is directly applied to the $F$ matrix considered as a contingency table and the $f_{ij}$ assumed to be frequencies. In this case, the $\chi^2$ distance between two actor profiles $p_{a_i}, p_{a_k}$ is given by:

$$
\delta_C^2(p_{a_i}, p_{a_k}) = \sum_{i=1}^{m} \left( \frac{f_{ij}}{s_j} - \frac{f_{kj}}{s_k} \right)^2 \frac{L}{s_j},
$$

(15)

with $D_e = \text{diag}(s_1/L, \ldots, s_n/L)$, and $L$ the total number of ties, whereas, the $\chi^2$ distance between two event profiles, $p_{e_j}, p_{e_k}$ is:

$$
\delta_C^2(p_{e_j}, p_{e_k}) = (p_{e_j} - p_{e_k})^T D_e^{-1} (p_{e_j} - p_{e_k}) = \sum_{i=1}^{n} \left( \frac{f_{ij}}{s_j} - \frac{f_{kj}}{s_k} \right)^2 \frac{L}{s_j},
$$

(16)

with $D_e = \text{diag}(d_1/L, \ldots, d_m/L)$.

A comparison of the two Eq. (14) with Eqs. (15) and (16) shows that, with respect to the Euclidean distance $\delta_L^2$, the $\chi^2$ distances in CA $\delta_C^2$ are based on a complex weighting system. Indeed, the $\delta_C^2$ distances between actors or between events depends not only on the pattern of participation or of attendance encoded in $f_{ij}$, but also on the actor degree $d_i$ and the event size $s_j$. Specifically in the $\chi^2$ distance formulae (15) and (16), less active actors (i.e. actors with a small degree) and small size events (i.e. events with a low rate of attendance) are associated to larger weights. Thus, the same differences in $f$ values could have very different impacts on the $\delta_C^2$ value due to weights in denominators in Eqs. (15) and (16), i.e. a small Euclidean distance $\delta_L^2$ could correspond to a large $\delta_C^2$ distance.

In particular, actors (events) profiles can be pushed away from the origin in the event (actor) space because of their small degrees (sizes) and not only because of their degree of dissimilarity. This effect in MCA is less severe because the weighting system in MCA is simpler than the one involved in CA, the actor degree $d_i$ does not play any role in the row and column profiles, and the event size $s_j$ is balanced by its complement $n - s_j$.

The two quantities $d_i$ and $s_j$ depend on the network density $\theta$: the larger the number of ties (i.e. dense networks), the higher the rates of membership for actors ($d_i$) and the larger the sizes of the events ($s_j$). The opposite is true for sparse networks.

With no loss of generality, we focus on the distances among actors (Eqs. (8) and (15)). In the case of sparse networks, as the $d_i$s will generally be low in value, the corresponding weights in $\delta_C^2$ will be higher in value. In the case of denser networks, this effect changes since the average degree in the network increases and the relative weights become smaller in value, slightly affecting the CA distances. In MCA on the contrary, the density does not affect...
the weighting system involved in $\delta_{\text{MC}}^2$, as the actor degrees do not come into the computation of the distance. In fact, each profile is weighted by a constant term equal to $1/m$.

As for event size, CA weights are equal to $1/s_j$. In the case of sparse networks, $s_j$s tend on average to be small and the weights tend to be large in value. For increasing values of $\theta$, CA weights sharply decrease tending to $1/n$ (i.e. each actor attends all the events). Thus we observe two effects: (i) sparse and dense networks have, on average, two very different weighting systems and (ii) small differences in the event sizes could yield large differences in weights. On the contrary, the MCA weights are equal to $1/s_j(n - s_j)$ so that for each $s_j$ there is a counterpart term $(n - s_j)$. Therefore, sparse and dense networks are treated in the same way, by assigning equally large weights to both very rare and very common events (i.e. events having extreme sizes). Moreover, for a wide range of densities (specifically, from 0.2 to 0.8), MCA weights vary (as a quadratic form) in a small range of values and remain relatively stable. Hence, events with different, but not extreme, sizes have a similar role in the computation of $\delta_{\text{MC}}^2$.

In order to visually appreciate the CA and MCA weight behavior with respect to the density $\theta$, in Fig. 7 we plot the expected values\(^9\) of $d_i$ (left panel) and $s_j$ (right panel) terms involved in the $\delta_{\text{CA}}^2$ weights as a function of the density $\theta$ for both CA and MCA.

\(^9\) Under the hypothesis that the presence of an edge has a Bernoulli distribution with probability of success $\theta$, it is easy to prove that $E(d_i) = n\theta$ and $E(s_j) = n\theta$. 

---

**Fig. 8.** Scatter plots of CA and MCA distances vs the Euclidean distance for the case of small networks: (a) dense; (b) sparse; (c) clustered with large noise (i.e. dense); (d) block regular with small noise (i.e. sparse).
Fig. 9. Boxplots of the distributions of the $R^2$ index evaluated between the Euclidean distance, and CA or MCA chi-square distances over the simulated networks.
5.1. An experimental comparison of distances in two-mode network analysis

In explaining the differences between \( \chi^2_{Fe} \), \( \chi^2_{MCA} \) and \( \chi^2_{CA} \), the presence of structures and the network dimensions have to be taken into account along with density. A simulation study has therefore been undertaken to compare the behavior of the two \( \chi^2 \) distances with respect to the Euclidean one in a wide set of network configurations. We generated networks by using as design parameters of the simulation: (i) density (dense vs sparse networks), (ii) presence of structure (random networks vs networks clustered into structural equivalent blocks), and (iii) dimension (small – 20 actors and 10 events – vs large networks – 200 actors and 50 events). All together, we get 8 possible configurations.

For each fixed dimension (small or large) we started from a full affiliation network obtaining dense and sparse networks by randomly removing 25% and 75% of the ties, respectively. By using networks generated in this way we created a partition of each matrix into a given number of clusters \( h \). Each cluster on the main diagonal has a random size and is created with density equal to 1. A small amount of noise \( (p) \) was then added to each cluster, so the final block density is slightly less than 1. We performed 100 runs for each design by varying \( h \) and \( p \) for the clustered networks.

Fig. 8 reports the results obtained in one of the 100 runs for the case of small networks by varying the other two design parameters, i.e. we have four small networks: random sparse \((p=0.2)\), random dense \((p=0.6)\), clustered sparse \((h=3, p=0.2)\), and clustered dense \((h=3, p=0.6)\). Fig. 8 shows that MCA deviates less than CA from the Euclidean distance, with a smaller variability. Similar results hold for large networks.

For each run we compute the \( R^2 \) index to measure the closeness of \( \chi^2_{MCA} \) and \( \chi^2_{CA} \) to \( \chi^2_{Fe} \). The overall results for the simulation study are reported in Fig. 9 where boxplots portray the distributions of the resulting \( R^2 \) indices. In general we can see that MCA more than CA closely resembles the \( \chi^2_{Fe} \) for all the cases, given that the \( R^2 \) values are closer to one with a lower variability. In the case of sparse networks, this variability becomes negligible for MCA with respect to CA. This is due to the larger variability of actor degrees and event sizes which, in these configurations, produce a large variability in CA weights leaving MCA weights more stable.

However there are cases in which these two techniques can be interchangeably used and the advantages of MCA are less evident. In particular this happens in random setting when affiliation networks are (very) dense, whereas in non-random setting this occurs when the structural equivalent clusters are not well defined by the high presence of extra-cluster linkages (i.e. clustered dense design). The effect of “extreme” clusteringization is even stronger as the overall density is higher and the network size is small. In other words, when the effect of clusteringization is masked by a large amount of noise – i.e. a high presence of extra-cluster ties – and the number of clusters \( h \) is small, both techniques capture the same “amount” of relational similarities among nodes.

6. Concluding remarks

In this paper we have discussed the use of MCA to visually explore two-mode networks, to analyze the degree of similarity of actor/event network profiles, and to characterize actors and events on the basis of their affiliations and on their attributes. We have discussed how the actors-by-events structure of a two-mode network fits the case-by-variable structure that is assumed in MCA, rightly assigning a different role in the analysis to actors and to events. We have also shown that the interpretation of the analysis results is enhanced by adopting the doubling principle. In this respect, we note that in Greenacre’s original proposal, the application of the CA algorithm to a doubled matrix was conceived for the case of subjects that express rates or judgements on ordinal scales. This suggests to us that our approach could be extended to analysis of any two-mode networks in which \( f_{ij} \) expresses the degree of affiliation of the actor \( a_i \) to the event \( e_j \) measured on an ordinal scale. Further studies need to be undertaken in this direction.

We have also compared MCA, as in our proposal, with the more frequently used simple CA, from an analytic point of view and through a simulation study. The metric used in MCA, i.e. the weighting system involved in the \( \chi^2 \) distance, turns out to be stable with respect to the actor degrees and to the event sizes, and with respect to some network characteristics such as density and the presence of blocks. This weighting system, even though could have some undesirable effects on the factorial maps, makes it possible to preserve some valuable information about the network structure.

Finally we would like to highlight that in the CA family there are other methods that are worth investigating for the analysis of affiliation networks. For example Subset Correspondence Analysis (Greenacre and Pardo, 2006) could be useful as it is able to focus the analysis only on the participation pattern in specific events, while preserving the properties of the analysis of the complete indicator matrix.

Acknowledgments

We are grateful to the anonymous referees for their comments and suggestions. Special thanks are due to IMAST technological districts for making the data available and to Michael Greenacre and Vladimir Batagelj for their valuable comments. We acknowledge the financial contribution of the REPOS project (networks, public policies, and development) – FSE POR Campania 2007–2013.

References


