Applications of eigenvector centrality to small social networks

Anders K. H. Bengtsson* and Mary-Anne Holfve-Sabel†

January 29, 2013

Abstract

This article investigates conceptual and methodological questions that may arise in applying eigenvector centrality to small social networks such as school classes. The focus on small networks brings out surprising and subtle properties related to the interpretation of the measure. We investigate examples where the weighted adjacency matrix of the underlying social network quantifies inter-individual preferences of whom to work with or play with. We show that mathematical operations such as transposition and symmetrization of the weighted adjacency matrix enhances the power of the measure. It is demonstrated that it is not sufficient to work with the original weight matrix. By working with a symmetrized or a semi-symmetrized or a transposed weight matrix different characteristics of the social interaction are revealed. The method chosen depends on the purpose of the investigation. Identifying isolated or popular individuals in the network are also facilitated using these operations.

Keywords: eigenvector centrality, small social networks, peer relations, integration

*School of Engineering, University of Borås, Allégatan 1, SE-50190 Borås, Sweden. Email: anders.bengtsson@hb.se

†School of Education and Behavioural Sciences, University of Borås, Allégatan 1, SE-50190 Borås, Sweden.
Email: mary-anne.holfve-sabel@hb.se

1 Introduction

The present paper grew out of a concrete question on how to extract information from a huge amount of sociometric class-room data collected by the second author. During the PhD project *Classroom, Peers and Teacher*, student attitudes were investigated (Holfve-Sabel, 2006). The overall aim was to map quality dimensions within school classes without including achievement results. Instead the focus was on a number of complicated issues in educational settings such as affective variables and effects of socialization. The idea behind this was that each class mirrors a small society where both adults and learners interact and create norms for behavior and learning. This also means that classrooms may present substantial similarities but also unique differences. Furthermore, both questionnaires and sociometric data were hypothesized to offer methodological improvements in the analytical process. The attitudes were thus analyzed and presented using two-level confirmatory factor analysis in (Holfve-Sabel and Gustafsson, 2005). In the present paper we discuss a number of issues related to evaluation of sociometric data of the type that was collected during the above mentioned project.

Quantitative sociometric methods are useful to analyze links between individuals within a group or network (Cillessen, 2009). An interesting type of measure are the amount of integration or segregation of an individual within the network. There are many such measures considered in the literature (Kindermann and Gest, 2009). We will, however, exclusively consider eigenvector centrality with the aim to understand how it can be applied to small social networks.

Eigenvector centrality was introduced by Bonacich in 1972 (Bonacich, 1972, 1987) and has been explored in various directions. For reviews, see for example (Seary and Richards, 2003). For a textbook reference on networks, see (Newman, 2010). Apparently it was independently rediscovered by Echenique and Fryer (Echenique and Fryer Jr, 2007) in 2007. These authors called it Spectral Segregation Index and stressed that it provides both a measure for the whole group as well as for the individuals. They applied it to segregation (defined as weak contacts with people in other groups than ones own) in large populations. It was through Echenique's and Fryer's definition of segregated networks that the we initiated the present work.

We will keep the old terminology eigenvector centrality (EC). Our aim is to better understand its properties when applied to small networks such as school classes. We do this in a particular context provided by the data collected by the second author. Our discussion is nevertheless quite general. We will adopt the terms *individual index* for the eigenvector components, and *group index* for the eigenvalue. We show that it will be useful to consider various mathematical operations, such as transposition and symmetrizing, to the weighted adjacency matrix of the network in order to extract more information.

Let us describe the kind of sociometric data that we are interested in. As well as providing context for the theoretical and methodological discussions to follow, it allows the reader to gauge to what extent our discussion can be generalized to other similar contexts and types of basic data.

We consider well-defined groups of individuals, in our case school children, organized in classes in the usual way of elementary schools. We are interested in two kinds of inter-individual relations: preferences as to whom to work with in the classroom, and preferences as to whom to play with during breaks. The students were asked to name three other students that they preferred to work with or play with respectively. They were also asked to grade their choices into first, second and third. It was allowed not to make choices and to make choices in other classes. Among other data collected was information concerning gender and ethnicity.

Preferences as to whom to work with and play with are correlated in practice. That is indeed one interesting aspect to study, but as to the collection of the data, they are considered to be independent, i.e. the students are not supposed to make any correlations, but just asked to make these two kinds of choices as if they were independent.

This can be generalized to other kinds of groups of people and other kinds of relations between them. The number of choices required can vary as well as the recording of other characteristics of the individuals. Hence, the method is fairly general.

In order to make our discussion comprehensible we need to review some relevant mathematics in sections 2 and 3. In sections 4 and 5 we discuss and show by examples how eigenvalue centrality can be useful in analyzing small social networks. Our conclusions are found in section 6.

2 Mathematical description of social networks

A group of n people can be considered as a set of n named elements called *individuals*. There can be various types of relations between the individuals involving any number of them. Relations involving only one individual are called *properties* of the individuals. Relations involving exactly two individuals are particularly interesting. One example of such a relation is *friendship* which is clearly mutual, another is *dependency* which need not be mutual. The kinds of relations we are interested in (preferences as to whom to work with or play with) involve at least two individuals and need not be mutual. We will treat such relations as relations between pairs of people, and therefore any given individual might participate in several such relations.

2.1 Graphs

A social network can be described by a *directed graph*. Each individual corresponds to a *node* in the graph and these can be numbered by an index i and denoted by P_i. If there is a relation between individual P_i and individual P_j there is a *directed link* in the graph going from node P_i to node P_j. Between any two nodes there might therefore be zero, one or two directed links. The special case where there is at most one directed link between any two nodes is called a *uni-directed* graph. A path in a graph is a sequence of nodes where each consecutive node can be reached by following a link in the graph.

If we don't attach any importance to the direction of the relation between the individuals, or if we consider the relation to be mutual, then the network can be described by a *un-directed graph*. In such a graph, there is thus zero or one link between any two nodes. Alternatively, we can view an un-directed graph as a graph where there is either no link between a pair of nodes, or two links, one in each direction. Adopting this point of view, we don't have to introduce separate definitions for un-directed graphs. An undirected graph is simply a directed graph where to any directed link there is automatically a link in the other direction.

2.2 Adjacency and weight matrices

A directed graph can be described by an *adjacency matrix* A of zeros and ones. Consider a graph with n nodes. If there is a link from node i to node j then the corresponding matrix element a_{ij} of A is 1, otherwise it is zero. The adjacency matrix is thus an $n \times n$ square matrix.

If there is some way of quantifying the strength of the relation, we can consider instead a *weighted adjacency matrix*, or simply a *weight matrix*, W with elements w_{ij}. There is a considerable amount of arbitrariness in judging the strength of the relation. We will return to this issue in section 4.2. For now we assume the weights to be given or assigned based on some principle of judging strengths of relations.

No matter how that assignment is done, it seems reasonable to have some bound on the weights, for instance

$$\sum_{j=1}^{n} w_{ij} = d \tag{1}$$

Intuitively, this corresponds to a bound on the total strength of individual i's relations with other individuals in the network. The requirement (1) is called *row stochasticity.* For actual data this requirement is too strong. As we will see, data satisfying equation (1), leads to a homogeneous EC (all individual indices are equal) with no ability to detect differences in integration within the group. A much more practical, and commonly occurring, situation is to have

$$\sum_{j=1}^{n} w_{ij} \leq d \tag{2}$$

The importance of the weight matrix in our context comes from the fact that eigenvalue centrality is calculated directly from the weight matrix, given that the weight matrix satisfies an important mathematical restriction to which we turn now.

2.3 Connectivity and irreducibility

Consider a directed graph. Two nodes i and j in a graph is said to be *connected* if there is a path from i to j following directed links in the graph. If from every node in the graph it is possible to reach every other node, then the graph is said to be *strongly connected.* If the links in the graph represents some kind of inter-individual relationship, this means that every individual is connected to every other individual, directly or indirectly, through such relations.

In order to compute the EC for a network, its graph must be strongly connected. If it is not, then the network must be subdivided into strongly connected subgraphs. In terms of adjacency, or weight matrices, the requirement of strong connectedness for the graph Γ translates into *irreducibility* for the corresponding adjacency matrix A (or W).

A matrix is *irreducible* if it is not reducible. A matrix is *reducible* if there is some permutation of its rows and columns (such permutations correspond to a relabeling of the nodes of the underlying graph) such that it can be brought to the following block form

$$\begin{pmatrix} B & C \\ 0 & D \end{pmatrix} \tag{3}$$

where B, C and D are sub-matrices with B and D quadratic. Otherwise, there are no special requirements on the matrices B, C, D. What is important is the zero sub-matrix in the indicated position. An adjacency matrix with the block structure of (3) corresponds to a subset of nodes not being possible to reach from the rest. A theorem states that the adjacency matrix A is irreducible if and only if the corresponding graph Γ is strongly connected (Horn and Johnson, 1985).

2.4 Kinds of weight matrices

To every quadratic matrix W there is a corresponding *transposed* matrix W^T. A matrix W is *symmetric* if for every matrix element $w_{ij} = w_{ji}$. For a symmetric matrix, the transposed matrix is equal to the matrix itself. A matrix W is *semi-symmetric* if whenever $w_{ij} = 0$ we also have $w_{ji} = 0$. A *positive* matrix has only positive (non-zero) elements. For a *non-negative* matrix all matrix elements are either positive or zero. The kinds of weight matrices that we will be interested in are non-negative corresponding to all weights being non-zero numbers.

3 Eigenvector centrality

Assume that the network is strongly connected and thus that the corresponding weight matrix is irreducible. The basic idea behind eigenvector centrality is that the index s_i for each individual i should be a weighted mean value of the indices over all the individuals that are related to i

$$\mathbf{s}_i = \frac{1}{\sigma} \sum_{j=1}^{n} w_{ij} \mathbf{s}_j \tag{4}$$

This is an eigenvalue equation as is easily seen by writing the formula using matrix notation and multiplying with σ on both sides

$$\sigma \mathbf{s} = W \mathbf{s} \tag{5}$$

This is a self-consistency equation for the individual indices. The definition of the EC is based on the Perron-Frobenius theorem which say that there is a highest positive eigenvalue σ of the irreducible weight matrix W and a corresponding eigenvector $\mathbf{s}$ with elements that are either positive or zero. These are used to define the EC for the whole group as the highest eigenvalue σ of the weight matrix. The individual indices are measured by the elements s_i of the corresponding eigenvector $\mathbf{s}$.

It is practical to rescale the eigenvector so that the sum of its elements equals the eigenvalue σ. Thus, denoting the group EC with σ and the elements of the eigenvector $\mathbf{s}$ with s_i, we get the sum $|\mathbf{s}|$ of the elements

$$|\mathbf{s}| = \sum_{i=1}^{n} s_i \tag{6}$$

Denoting the elements of the normed eigenvector $\hat{\mathbf{s}}$ with $\hat{s}_i$ we get

$$\hat{s}_i = \frac{\sigma}{|\mathbf{s}|} s_i \tag{7}$$

This yields

$$\sum_{i=1}^{n} \hat{s}_i = \sigma \tag{8}$$

In performing this particular rescaling we are following (Echenique and Fryer Jr, 2007).

The pair of eigenvalue and normed eigenvector $(\sigma, \hat{\mathbf{s}})$ constitute eigenvector centrality for the network described by the weighted adjacency matrix W.

Having reserved the term EC for the pair $(\sigma, \hat{\mathbf{s}})$, we need some convenient terminology for the two parts of the pair. Let us simply say *group index* for the eigenvalue σ and *individual index* for any component $\hat{s}_i$ of the eigenvector $\hat{\mathbf{s}}$.

We will strengthen the intuition behind the interpretation of the index with some sample calculations. Let us first point out one counter-intuitive aspect of the index.

3.1 EC for row stochastic matrices

If all row sums in W have the same value d then the eigenvector $\hat{s}$ is constant in the sense that all its components are equal to d/n. Furthermore the highest eigenvalue σ is d. In this case we say that the individual indices are *homogeneous*. But this means that for such a weight matrix, EC cannot detect any differences in the individual integration indices.

This somewhat surprising result can be understood as follows. If we insert a constant vector $\mathbf{k}$

$$\mathbf{k} = \begin{pmatrix} k \\ k \\ \vdots \\ k \end{pmatrix} \quad \text{i.e.} \quad k_1 = k, k_2 = k, \ldots k_n = k \tag{9}$$

into equation (5), now with $\sigma = d$, we get

$$\begin{pmatrix} w_{11}k + w_{12}k + \ldots + w_{1n}k \\ w_{21}k + w_{22}k + \ldots + w_{2n}k \\ \vdots \\ w_{n1}k + w_{n2}k \ldots + w_{nn}k \end{pmatrix} = \begin{pmatrix} dk \\ dk \\ \vdots \\ dk \end{pmatrix} \tag{10}$$

Since all the row sums $w_{i1}+w_{i2}+\ldots+w_{in} = d$, we see that we have a solution. This solution $\mathbf{k}$ must then be the eigenvector associated with the eigenvalue d.

This is perhaps not what we would have expected to find. We can understand it by referring back to our context. In collected data there will in general be quite a few individuals making choices outside of the group, or not making the maximum allowed number of choices. It is therefore useful to have a kind of reference situation where this situation does not occur.

Let us introduce the concept of a *fully integrated group*. In such a group, all individuals make all their choices within the group. Furthermore, the group is strongly connected in the sense discussed in

2.3. It is easy to see that no matter how the weighting of the choices is done, as long as the same principle is applied to all the individuals, the weight matrix becomes row stochastic. Consequently, all individuals have the same index of integration. Considering the meaning we have given to a fully integrated group, this does indeed make sense, there is no difference between the individuals in such a group.

Detectable differences in individual indices therefore come from broken row stochasticity as expressed in equation (2).

4 Conceptual and methodological issues

In this section we discuss general conceptual issues pertaining to using EC for small social networks. Our discussion is illustrated by explicitly working through archetypical examples.

4.1 Context and questions

We have a set of n individuals making a certain number p of choices of preference among their peers. The choices are graded first, second, third and so on up to p:th. There are a set of phenomena that we wish the EC calculation to capture. Some of these are

(a) Individuals that consider themselves lacking relations, or only having weak relations and therefore do not make any choices or less than p choices.

(b) Individuals that make choices outside their own group.

(c) Individuals that tend not to be chosen by their peers.

(d) The choice relation may be un-symmetric.

Items (a) and (b) break the row stochasticity and will show up in the EC calculation as we will see below. Items (c) and (d) are more subtle and will be discussed below. It turns out that it is not sufficient just to work with the weight matrix as it comes out of the quantification stage. Instead by working with transposed, symmetrized and semi-symmetrized weight matrices, a much more thorough understanding of the network can be achieved.

4.2 Quantification

The basic data is ordinal. There is no way to ascertain that different individuals attach the same significance to first choice, second choice et cetera. Nevertheless, in order to calculate eigenvector centrality we need quantitative data in the form of a weight matrix. Therefore we assign weights $w_1, w_2, \ldots, w_p$ to the choices. This quantification is arbitrary apart from the natural assumption that $w_1 \geq w_2 \geq w_3 \geq \cdots \geq w_p$ and that the sum of the weights should be bounded by some number d as in equation (2).

An extreme weighting would be to downplay any significance of the order of the choices and give them all the same weight

$$w_1 = w_2 = w_3 = \ldots = w_p$$

The other extreme would be to downplay all but the first choice

$$w_1 = d, w_2 = 0, w_3 = 0, \ldots, w_p = 0$$

In between there is a lot of arbitrariness. This leads to the question of the sensitivity of the EC calculation towards variations of weightings. In practice, computation of EC is done by numerical algorithms. Thus, the best thing to do in actual computations is to vary the weights and record the variations in the computed indices. At least in sample calculations it seems that the arbitrariness in the weighting does not jeopardize the reliability of the EC computations. For actual data, numerical sensitivity analysis might be needed. This could be done by systematically varying the weights and studying the ensuing variations in the indices.

4.3 Mutuality

An important question is how to treat the case when a certain individual P_i have chosen another individual P_j, but not the other way around. We first assume that the network is strongly connected, if not, we have to work separately with the irreducible sub networks.

To begin with we have the elements $a_{ij} = 1$ and $a_{ji} = 0$ in the adjacency matrix, and after quantification, $w_{ij} \geq 0$ and $w_{ji} = 0$ in the weight matrix. We will now discuss various operations on the weight matrices.

Semi-symmetrization

Semi-symmetrization means setting $a_{ij} = 0$ and $a_{ji} = 0$ whenever either of them is zero. The resulting new adjacency matrix is called *semi-symmetric*. This procedure can be carried over to the weight matrix putting $w_{ij} = 0$ whenever $w_{ji} = 0$. We generalize this procedure to the operation of *semi-symmetrization* where all matrix elements in the original weight matrix are replaced by the geometric mean $\sqrt{w_{ij} \cdot w_{ji}}$.

There are circumstances where semi-symmetrization is reasonable. It tends to break row stochasticity and therefore it affects the indices. It is a rather brutal operation on the network weight matrix, but it can be useful as an instrument of analysis precisely because it breaks uni-directional links (Holfve-Sabel, 2010).

Symmetrization

Another way to treat the case $w_{ij} \geq 0$ and $w_{ji} = 0$ is to replace all matrix elements in W with the arithmetic mean $(w_{ij} + w_{ji})/2$. This makes the weight matrix symmetric, and we call this operation *symmetrization*. Symmetrization also tends to break row stochasticity and therefore affects the EC computation. Symmetrization is often well motivated.

Transposition

Eigenvector centrality computed for a fully integrated group does not detect any differences in the individual indices. However, fully integrated groups can look very different. An extreme example is when one individual P_a has been chosen as the first choice by all the other individuals, while another individual P_b has been chosen by only one other individual and in this case as the last choice. Intuitively we would consider P_a strongly integrated while P_b weakly integrated. Nonetheless, under the assumption that the network is strongly connected and fully integrated, all individuals have the same individual indices. The restriction "fully integrated" is weak, we only require that all individuals make all their allowed choices within the group. This inevitably makes the weight matrix row stochastic, and the homogeneity of EC follows.

This could be considered to be a weakness in EC but it can be understood. When computing the indices based on the original weight matrix, we must remember that what we have in the rows are weights coming out of each individuals own choices. One could say that the original weight matrix contains information about the network *as seen from the perspective of each and every individual.* When everyone make all their allowed number of choices p, it does indeed make sense that EC computes them to be equally integrated. Since the network is strongly connected and fully integrated, the mere fact that individual P_b belongs to it, is enough for the EC to compute P_b to be just as integrated as anyone else in the network.

Still, we would like to be able to see the difference between P_a and P_b. It turns out that a simple trick allows us to do precisely that. Transposing the weight matrix and computing EC based on W^T instead of W gives us the required information. The intuition behind this procedure comes precisely from the argument above where we understood that the original weight matrix gauges the standing of each individual in the network *from its own point of view.* The transposed matrix instead gauges the standing of each individual in the network *from the point of view of the whole network.*

One way of understanding this is to ask the question of how integrated any particular individual is from the point of view of the rest of the network. Then the focus is on the peer choices. The corresponding weights are found in columns of the W. The sum of these weights may very well not sum to d, i.e. W is not column stochastic, and therefore W^T is not row stochastic. An individual that is the first choice among all its peers will get a total weight (row sum in the transposed matrix) equal to $(p-1)w_1$ which in general is considerably higher than d. Thus, column sums may very well exceed the bound d on row sums.

In order to illustrate this we will consider a few archetypical models. We consider small networks comprising only six individuals numbered $1, 2, 3, 4, 5, 6$. When referring to a particular individual we write for instance no.3.

4.4 Archetypical examples

First we need to elaborate our base case, the fully integrated network.

A base case: Balanced fully integrated network In a fully integrated group, all individuals make all their choices within the group. If furthermore, every individual is chosen by exactly p other individuals, once as first choice, second choice, up to p:th choice, then we call the network *balanced fully integrated.* The weight matrix W

for such a network becomes both row and column stochastic and the same holds for the transpose W^T. All individuals get the same index whether we compute them based on W or W^T.

Example base case There are many balanced fully integrated choice structures. Table 1 shows a simple example.

Table 1: Example of choice structure in a balanced fully integrated network.

Individual	Choice		
	first	second	third
1	2	3	4
2	3	4	5
3	4	5	6
4	5	6	1
5	6	1	2
6	1	2	3

It is clear from table 1 that "everyone is equally chosen". It is a natural base case for EC calculations. There are no differences in integration to detect, neither from the individuals own point of view (W), nor from the point of view of the rest of the network (W^T).

Next translate this into a weight matrix. Suppose the first choice gets weight 0.6, second choice 0.3 and third choice 0.1. The weight matrix becomes

$$W = \begin{pmatrix} 0 & 0.6 & 0.3 & 0.1 & 0 & 0 \\ 0 & 0 & 0.6 & 0.3 & 0.1 & 0 \\ 0 & 0 & 0 & 0.6 & 0.3 & 0.1 \\ 0.1 & 0 & 0 & 0 & 0.6 & 0.3 \\ 0.3 & 0.1 & 0 & 0 & 0 & 0.6 \\ 0.6 & 0.3 & 0.1 & 0 & 0 & 0 \end{pmatrix} \quad (11)$$

This matrix is both row and column stochastic. But it is not symmetric as the choices need not be symmetric, only fully balanced according to our definition. The corresponding symmetrized weight matrix becomes

$$W_{\text{sym}} = \begin{pmatrix} 0 & 0.3 & 0.15 & 0.1 & 0.15 & 0.3 \\ 0.3 & 0 & 0.3 & 0.15 & 0.1 & 0.15 \\ 0.15 & 0.3 & 0 & 0.3 & 0.15 & 0.1 \\ 0.1 & 0.15 & 0.3 & 0 & 0.3 & 0.15 \\ 0.15 & 0.1 & 0.15 & 0.3 & 0 & 0.3 \\ 0.3 & 0.15 & 0.1 & 0.15 & 0.3 & 0 \end{pmatrix}. \tag{12}$$

This symmetrized weight matrix is of course also both row and column stochastic.

Let us also note the semi-symmetrized matrix

$$W_{\text{semisym}} = \begin{pmatrix} 0 & 0 & 0 & 0.1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0.1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0.1 \\ 0.1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0.1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0.1 & 0 & 0 & 0 \end{pmatrix} \tag{13}$$

This matrix is reducible. It corresponds to a small network consisting of three pairs: no.1 and no.4 choosing each other, and so on. It illustrates dramatically the tendency of semi-symmetrization to "break up" the network into smaller pieces. We now turn to a few interesting cases that contrast this base case.

A very popular individual Suppose no.1 is very popular among its peers. All the other individuals in the fully balanced network choose no.1 as their first choice. This have some further consequences for the choice structure as is seen in the table: no.4 effectively interchanges no.1 and no.5 while no.5 interchanges no.1 and no.6.

The weight matrix becomes

$$W = \begin{pmatrix} 0 & 0.6 & 0.3 & 0.1 & 0 & 0 \\ 0.6 & 0 & 0 & 0.3 & 0.1 & 0 \\ 0.6 & 0 & 0 & 0 & 0.3 & 0.1 \\ 0.6 & 0 & 0 & 0 & 0.1 & 0.3 \\ 0.6 & 0.1 & 0 & 0 & 0 & 0.3 \\ 0.6 & 0.3 & 0.1 & 0 & 0 & 0 \end{pmatrix} \tag{14}$$

The matrix is row stochastic and EC does not detect the conspicuous popularity of individual no.1. However, if we compute EC based on

Table 2: Example choice structure with a very popular individual.

Individual	Choice		
	first	second	third
1	2	3	4
2	1	4	5
3	1	5	6
4	1	6	5
5	1	6	2
6	1	2	3

the transpose W^T we get individual indices (which we show as a row vector)

$$\hat{\mathbf{s}}_{\text{tr}} = \big(0.38, 0.25, 0.12, 0.11, 0.07, 0.07\big) \qquad (15)$$

where we note the high index for no.1. The group EC is $\sigma_{\text{tr}} = 1$.

Can we understand the low indices for no.5 and no.6? Judging from the column sums (corresponding to the row sums in W_{tr}), where no.5 and no.6 gets larger values than no.3 and no.4, one could perhaps expect no.5 and no.6 to get higher indices than no.3 and no.4. However, no.5 and no.6 are not chosen by the very popular no. 1. The EC computation based on the transposed weight matrix manage to detect this.

Let us also study the symmetrized case. The individual indices for the symmetrized weight matrix become

$$\hat{\mathbf{s}}_{\text{sym}} = \big(0.30, 0.22, 0.15, 0.15, 0.15, 0.16\big) \qquad (16)$$

In this case the group EC is $\sigma_{\text{sym}} = 1.13$. If we rescale the sum of the individual indices to 1, we get

$$\frac{\hat{\mathbf{s}}_{\text{sym}}}{\sigma_{\text{sym}}} = \big(0.27, 0.19, 0.14, 0.13, 0.13, 0.14\big)\,. \qquad (17)$$

If we now compare "transposed" EC to "symmetrized" EC, we see that both methods manage to detect the popularity of individual no.1, but that the symmetrization gives a somewhat "smother" measure. This is to be expected. The symmetrized weight matrix is indeed just the mean of the original matrix and the transpose, i.e. $W_{\text{sym}} = (W + W_{\text{tr}})/2$ so when computing EC based on W_{sym} we are not just taking the point of view of the network but also of the individual.

A very unpopular individual Now consider no.6 to be very unpopular. Only no.3 have chosen no.6 as a third choice. The weight matrix is

$$W = \begin{pmatrix} 0 & 0.6 & 0.3 & 0.1 & 0 & 0 \\ 0 & 0 & 0.6 & 0.3 & 0.1 & 0 \\ 0 & 0 & 0 & 0.6 & 0.3 & 0.1 \\ 0.1 & 0.3 & 0 & 0 & 0.6 & 0 \\ 0.3 & 0.1 & 0.6 & 0 & 0 & 0 \\ 0.6 & 0.3 & 0.1 & 0 & 0 & 0 \end{pmatrix} \tag{18}$$

It is row stochastic and EC cannot detect the unpopularity of no.6. So we compute based on the transposed matrix instead. We get

$$\hat{\mathbf{s}}_{\text{tr}} = \big(0.10, 0.16, 0.26, 0.22, 0.23, 0.03\big) \tag{19}$$

The group EC is $\sigma_{\text{tr}} = 1$. The unpopularity of no.6 is evident, but perhaps it should be smoothed somewhat by its own choices within the group? Based on the symmetrized matrix we get

$$\hat{\mathbf{s}}_{\text{sym}} = \big(0.16, 0.19, 0.22, 0.19, 0.19, 0.09\big) \tag{20}$$

The group EC is now $\sigma_{\text{sym}} = 1.05$.

4.5 Concluding the examples

As these examples show, it is not enough to work only in terms of the original weight matrix. Not even in the case where we don't have row stochasticity. Then, as we have seen, EC do compute different individual indices, but still only from the point of view of each individual's standing in the network. By working with the transposed and the symmetrized matrices, we weigh in each individual's standing within the network from the point of view of the rest of the network.

Row stochasticity (and column stochasticity) is of great importance for the EC computation and in particular for its interpretation. With our choices of weighting principles, a fully integrated group will get a row stochastic weight matrix. For such a matrix, all individual indices evaluates to the same number. This follows immediately from the mathematics. However, as argued, even in a fully integrated group, different individual's can have different "popularity" depending on the extent to which they have been chosen by their peers. This should be reflected in an integration index.

As we have seen in the examples, we can extract these effects by transposing or symmetrizing the weight matrix. Transposing breaks the row stochasticity (unless the network is also balanced). Symmetrization also breaks row stochasticity in such a way that the weights of popular individuals gets enhanced.

So what should one choose, transposing or symmetrizing? If we argue based on the intuition that the original matrix shows each and every individuals level of integration in the network from its own perspective, while the transposed matrix does the same thing but now from the perspective of the rest of the network, then symmetrization can be viewed as a compromise between these two perspectives. Transposition, in this sense, is "tougher" in what the individual itself thinks of its standing does not matter. Again, semi-symmetrization is even tougher. If a relation is not mutual it does not count at all. It is however a method to spot isolated individuals (Holfve-Sabel, 2010).

Symmetrization of an originally reducible matrix may very well turn it irreducible. On the other hand, semi-symmetrization may very well (and often does) turn an irreducible matrix into a reducible. We can understand this in terms of links in the network. Symmetrization can be thought of as adding new directed links to the network (in cases when originally $w_{ij} = 0$ but $w_{ji} \neq 0$). Semi-symmetrization tends to remove links (again precisely in cases when originally $w_{ij} = 0$ but $w_{ji} \neq 0$).

Another effect that tends to break row stochasticity is non-choices or choices outside of the group. In cases when an fully integrated group is split according to for instance gender or ethnicity (or some other characteristic of the individuals) we cannot expect the subgroups to still be fully integrated. This shows up in the weight matrices not being row stochastic.

5 Network analysis

It remains to discuss what to do when the network is not strongly connected, which is the common situation encountered for real data. We look for a way to analyze such a network into strongly connected subnetworks.

First of all we want to find all strongly connected subgraphs of the network because these are what we can compute EC for directly. Such a strongly connected subgraph is characterized by the fact that from any one of the nodes inside the subgraph, it is possible to reach all other nodes in the subgraph. However, it may be that there are directed links out of the subgraph in question, leading into other subgraphs.

Following standard terminology (Newman, 2010), we call a strongly connected subgraph a *component*. Having found all the components (some of which may contain just a single node), the network can be pictured as a uni-directed graph with directed links between the components. We also introduce some new terms to capture more of the structure.

Some components have the property that no links lead out of them. We call these *final components*. In case the final component contains just one node, we call it simply a *final node*. There may be components with no links entering them from the outside. We call these *initial components*. In case the initial component contains just one node, we call it simply an *initial node*. Any component that is neither final nor initial, i.e. components having links going into it as well going out will be called *open components*. Components with no links in or out we call *closed components*. In case the closed component contains just one node, we call it a *closed node*.

Let us now illustrate all this with some pictorial examples. In order not to clutter the pictures we only draw at most one link between any two nodes in the original graph (corresponding to $p = 1$). Figure 1 shows an example of an open component with one incoming link and one outgoing.

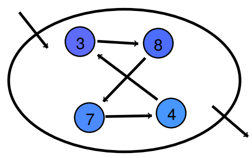

Figure 1: Example of an open component.

We chose to draw a line around the component and to color its nodes blue when the component is open. It is important to realize that the links going *into* and *out of* the open component *comes from* and *goes into* respectively *different* components in the network. If that is not the case, then we can find larger components by merging components. This merging of components is illustrated in Figure 2. Since there are links going both ways between these subgraphs, they should be merged into one component. Between any two components we draw at most only one representative link. The components form precisely an uni-directed graph.

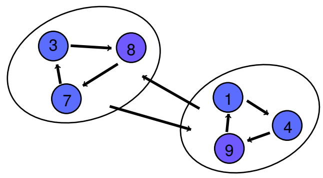

Figure 2: Two subgraphs that can be merged into a single component.

Next we illustrate in figure 3 an initial component, a final component and a closed component, using colors of green, red and grey respectively for the nodes inside.

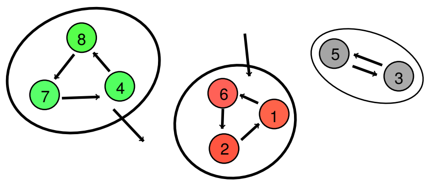

Figure 3: Example of initial, final and closed components.

5.1 Example of network analysis

Let us consider a network comprising of 26 individuals with the choice structure as in table 3 where the hyphens mean non-choices. It is constructed to bring out several aspects of the network analysis discussed above. First we analyze the network based on the original weight matrix and compute EC. We comment on the effect of transposing the weight matrix. Then we look at the symmetrized and semi-symmetrized cases and compute EC in these cases.

Figure 4 shows the result of analyzing the network in terms of components (for standard techniques, see (Newman, 2010)). The links inside the components are not shown and the links between components

Table 3: Choice structure in a network.

Individual	Choice			Individual	Choice		
	first	second	third		first	second	third
1	6	25	22	14	16	18	17
2	16	11	14	15	13	4	8
3	14	18	-	16	18	3	-
4	8	21	15	17	11	19	9
5	3	18	-	18	2	-	-
6	22	1	-	19	9	20	11
7	24	12	23	20	10	17	11
8	13	21	15	21	13	4	8
9	11	20	23	22	1	-	-
10	-	-	-	23	24	7	-
11	17	19	9	24	23	-	-
12	23	24	-	25	19	17	14
13	21	8	15	26	-	-	-

are representatives of what might be several links between particular nodes within the components. Note that this network is not strongly connected. EC can be computed for the components but not for the whole network (see however the next section).

EC may be computed for components with more than one member. We record the result of this computation in table 4 below.

Transposition

Transposing the weight matrix corresponds to reversing all the links in the network. As a consequence, the representative links between the components will be reversed. Open components stay open, closed stay closed, but finals and initials are interchanged. The result of computing EC is recorded in table 5. We remind the reader of the discussion in section 4.3. For instance comparing the indices for the closed component co.4 we see that it is fully integrated when computed from the original weight matrix but not when computed from the transposed weight matrix.

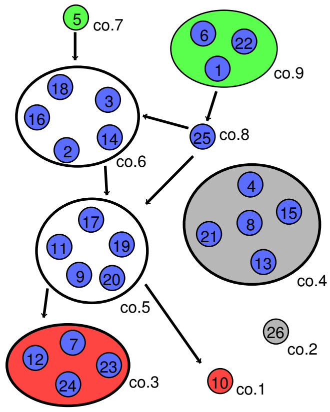

Figure 4: Analyzed network corresponding to table 3.

Symmetrization

Symmetrization consolidates the network, adding links where there are only links in one direction as well as recomputes the weight matrix as already discussed. As a consequence the representative links now go in both directions. Effectively, the components co.1, co.3, co.5, co.6, co.7, co.8 and co.9 are merged into one big component (which we now renumber as co.1) whereas components co.2 and co.4 stay closed. We record EC in table 6. The individual no.26 is not in the table.

These values are not directly comparable to the values for the original and transposed weight matrix, apart from the closed component number co.4 (which is disconnected from the rest of the network). Due to the large number of members in the merged component, the group index is shared among more individuals than in the smaller components of the original network. The same effect of number of individuals in a component is also seen when comparing the components within the symmetrized network. One could rescale the indices taking the effect of size into account.

Semi-symmetrization

Semi-symmetrization disaggregates the network, removing links where there are only links in one direction as well as recomputes the weight matrix as already discussed. As a consequence the network breaks up into closed components and closed individuals. In this case we get four closed components. EC is recorded in table 7. We keep the numbering from the original network. As an effect of the semi-symmetrization, co.3 has lost member 12, co.5 has lost member 20. The individuals not in the table are the ones that have got their links removed by semi-symmetrization.

Tables of EC for the example network of table 3 and figure 4.

Table 4: EC for the original network.

components and their index	Individuals and their indices				
co.3 0.788		7 0.223	12 0.211	23 0.201	24 0.153
co.4 1.000	4 0.200	8 0.200	13 0.200	15 0.200	21 0.200
co.5 0.914	9 0.181	11 0.228	17 0.228	19 0.177	20 0.100
co.6 0.767	2 0.147	3 0.177	14 0.169	16 0.159	18 0.115
co.9 0.732			1 0.252	6 0.273	22 0.207

Table 5: EC for the transposed network.

components and their index	Individuals and their indices				
co.3 0.788		7 0.114	12 0.043	23 0.300	24 0.331
co.4 1.000	4 0.107	8 0.201	13 0.335	15 0.064	21 0.293
co.5 0.914	9 0.159	11 0.275	17 0.215	19 0.161	20 0.105
co.6 0.767	2 0.175	3 0.080	14 0.085	16 0.204	18 0.224
co.9 0.732			1 0.278	6 0.228	22 0.225

Table 6: EC for the symmetrized network.

components and their index	Individuals and their indices				
co.4 1.063	4 0.161	8 0.226	13 0.283	15 0.145	21 0.248
co.1 0.993	1 0.017	2 0.053	3 0.032	5 0.016	6 0.010
	7 0.012	9 0.106	10 0.020	11 0.154	12 0.012
	14 0.043	16 0.047	17 0.137	18 0.044	19 0.112
	20 0.068	22 0.009	23 0.023	24 0.020	25 0.059

Table 7: EC for the semi-symmetrized network components.

components and their index	Individuals and their indices				
co.3 0.625			7 0.077	23 0.279	24 0.2
co.4 0.923	4 0.102	8 0.186	13 0.276	15 0.112	21 0.247
co.5 0.671		9 0.097	11 0.266	17 0.238	19 0.069
co.9 0.490			1 0.207	6 0.179	22 0.10

5.2 EC for non-strongly connected networks

EC cannot be directly computed for non-strongly connected graphs. However, given our context of a medium sized group of individuals making choices of preference among their peers, and our subsequent quantification of this data, at least some comparisons across components should be possible. Granted that the collection of the basic ordinal data and the quantification is consistently done for the whole group, the uni-directed graph of components retain some information about their common origin.

Consider a component co.M in the network containing m individuals. Together they have a total weight of md to distribute according to their choices. To the extent that the individuals make outside choices, some of this total weight, "leaks out" of the component. As a consequence, the group index σ_M computed from the original weight matrix will be less than d. Indeed, all links out off a component are weight drains and lower σ. What about links into the component? This corresponds to individuals in other components making choices in the component co.M. This, however, has no effect on σ_M. Component co.M gains nothing from incoming links. The reason is that the weight matrix for a component only contain information about links within the component. There is no way to maintain information about links from the outside. It does not help to transpose the weight matrix either. Even though transposition can be pictured as reversing the links in the graph, the original choice structure is not affected. The group index is the same for the transposed weight matrix, as is seen from tables 4 and 5. In conclusion then, a relatively low value for the group index indicates that the individuals in the component has relatively many outside choices and/or non-choices.

As regards individual indices, comparing them across components can be done to some extent, provided the differences in group indices and size of the components is accounted for. Consider two components co.K and co.L of sizes k and l, and with group indices σ_k and σ_l respectively. We have to rescale the eigenvectors $\hat{\mathbf{s}}_k$ and $\hat{\mathbf{s}}_l$ to make them comparable. A simple choice is

$$\hat{\mathbf{s}}'_k = \frac{k}{(k+l)}\hat{\mathbf{s}}_k \quad (21)$$

$$\hat{\mathbf{s}}'_l = \frac{l}{(k+l)}\hat{\mathbf{s}}_l \quad (22)$$

This compensates for the difference is size while retaining the information in the group indices

$$\sigma'_k = \frac{k}{(k+l)}\sigma_k \tag{23}$$

$$\sigma'_l = \frac{l}{(k+l)}\sigma_l \tag{24}$$

Perhaps the best way to take into account incoming links to components is to work with the symmetrized weight matrix. This has the effect of merging components. Consider components co.K and co.L with a link going from co.K to co.L, i.e. individuals in co.K have made choices in co.L. Symmetrizing the weight matrix will merge them, and the merged component will contain all the individuals and the individual indices for the members of co.L will be enhanced by the "outside" choices. Let us provide one final example to illustrate this very interesting point.

Example: Initial and final components

We design a small network with 7 individuals with the special choice structure of table 8.

Table 8: Example choice structure: an initial and a final component.

Individual	Choice		
	first	second	third
1	2	4	5
2	3	5	6
3	1	6	7
4	5	6	7
5	6	7	4
6	7	4	5
7	4	5	6

The individuals no.1 , no.2 and no.3 make their first choices among themselves, but their second and third choice among no.4, no.5, no.6 and no.7. The latter, make all their choices among themselves. We thus get one initial component (co.1) with members no.1, no.2 and no.3 and one final component (co.2) with members no.4, no.5, no.6 and no.7.

Computing EC from the original weight matrix yields homogeneous indices. We get $\sigma_{\text{co.1}} = 0.6$ shared equally among members of component co.1, and $\sigma_{\text{co.2}} = 1.0$ shared equally among members of component co.2. Computing EC from the transposed weight matrix gives the same result, since both components are balanced fully integrated. This illustrates the leak of weight from component co.1 lowering its index, while nothing is gained by component co.2.

Symmetrizing the weight matrix will merge the components. The group index become $\sigma = 1.0650$ and the individual indices

$$\hat{\mathbf{s}}_{\text{sym}} = \big(0.0870,\ 0.0873,\ 0.0870,\ 0.2008,\ 0.2043,\ 0.2040,\ 0.1947\big) \tag{25}$$

Now the influx of weight from component co.1 to component co.2 is captured in the individual indices.

We chose to quote four decimals here just to show how closely EC captures the choice structure. Note that individuals no.5 and no.6 get more choices than individuals no.4 and no.7 in co.2 from members in co.1. Furthermore, no.2 in co.1 gets a slightly higher index since it is related to the more popular no.5 and no.6. EC is indeed finely tuned to the underlying choice structure of the network.

6 Conclusions and outlook

Eigenvector centrality provides a subtle index of integration or segregation on the individual as well as the network level. This becomes clear when trying to apply it to small social networks where the computed indices can be compared with the corresponding intuitive concepts. In our context, EC computed from the original weight matrix measures integration from the point of view of the individual, whereas EC computed from the transposed weight matrix computes integration from the point of view of the peers. Computing EC with a symmetrized weight matrix moderates the individual and group indices between these extremes. Furthermore, the symmetrized weight matrix most naturally allows us to compare individual indices across strongly connected sub networks as we have argued in the previous section. The proposed method has specific strengths. An example is using the instrument in schools with segregated uptake areas to find whether the students actively make nominations regardless of ethnicity during school work and or breaks (Holfve-Sabel, 2012). Using a semi-symmetrized matrix, lonely students become visualized. These students make choices which are not bilateral. If the aim is to find leaderships, e.g. popular individuals among the students, the symmetrized or transposed matrices become useful. These methods also provide a measure of the class harmony by indicating either segregated smaller groups or large class networks.

Acknowledgements

The first author would like to thank Dr. Magnus Lundin for discussions. The second author would like to thank professor Jan-Eric Gustafsson for inspiration in pursuing new approaches to analyzing sociometric data.

References

Bonacich, P. (1972). Factoring and weighting approaches to clique identification. *Journal of Mathematical Sociology 2*, 113–120.

Bonacich, P. (1987). Power and centrality: A family of measures. *American Journal of Sociology 92*, 1170–1182.

Cillessen, A. H. N. (2009). Sociometric methods. See Rubin et al. (2009), pp. 82–99.

Echenique, F. and R. G. Fryer Jr (2007). A measure of segregation based on social interactions. *Quarterly Journal of Economics 122*, 441–485.

Holfve-Sabel, M.-A. (2006). *Attitudes towards Swedish comprehensive school. Comparisons over time and between classrooms in grade 6.* Ph. D. thesis, Gothenburg University. Göteborg Studies in Educational Sciences 242. http://gupea.ub.eu.se/dspace/handle/2077/10035.

Holfve-Sabel, M.-A. (2010). Characteristics of students without reciprocal friendship during school work. *Paper given at ECER, Helsinki*. http://hdl.handle.net/2320/6621.

Holfve-Sabel, M.-A. (2012). Segregation between schools does not mean the occurrence of segregation within classrooms. *Manuscript sent to Educational Enquiry*.

Holfve-Sabel, M.-A. and J. Gustafsson (2005). Attitudes towards school, teacher and classmates at classroom and individual levels: an application of two-level confirmatory factor analysis. *Scandinavian Journal of Educational Research 49(2)*, 187–202.

Horn, R. A. and C. R. Johnson (1985). *Matrix Analysis*. Cambridge University Press.

Kindermann, T. and S. Gest (2009). Assessment of the peer group: Identifying naturally occurring social networks and capturing their effects. See Rubin et al. (2009), pp. 100–117.

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

Rubin, K. H., W. M. Bukowski, and B. Laursen (Eds.) (2009). *Handbook of Peer Interactions, Relationships, and Groups*. Guilford Press, New York.

Seary, A. and W. Richards (2003). Spectral methods for analysing and visualizing networks: An introduction. In K. C. R. Beiger and P. Pattison (Eds.), *Dynamic Social Network Modelling and Analysis*, pp. 209–228. The National Academic Press.