Eigenvectors in Social Networks Analysis*

Omar Lizardo†[1]

[1]Department of Sociology, UCLA

November 25, 2025

*Reserved for acknowledgments
†olizardo@soc.ucla.edu

Abstract

I this chapter the author argues that the use of eigenvector-based ranking techniques in Social Network Analysis (SNA), particularly Eigenvector Centrality, should be viewed not as an "black box" mathematical fix to a circular definition problem, but as a justified "clear box" model based on an underlying generative process: the diffusion of status or prestige through a network. The author first reviews the historical context, starting with Bonacich's (1972) foundational work, and addresses the criticism that these spectral methods lack the transparent theoretical foundation of other centrality measures like closeness and betweenness. The core of the argument is demonstrated by showing that the canonical Eigenvector Centrality, PageRank scoring, and the HITS algorithm (Hubs and Authorities) can all be understood as variations of a simple "status distribution game" algorithm, which converges to the leading eigenvector of different proximity matrices (adjacency, Markov transition, and common-neighbors matrices, respectively). Furthermore, the paper generalizes this status distribution framework, demonstrating that running the algorithm on centered proximity matrices, such as those used in Principal Components Analysis (PCA) and the Modularity matrix, shifts the outcome from producing one-dimensional rank scores to revealing structural role sets and communities within the network.

1 Introduction

The use of linear algebra machinery, particularly the technique of eigenvector—also referred to as "spectral"—decomposition of square matrices, has a long history in the social and behavioral sciences, with some applications, such as Seeley's (1949), dating back to the late 1940s (Vigna 2016). Nevertheless, eigenvector decomposition techniques entered Social Network Analysis (hereafter SNA), and by implication sociology more generally, due to the pioneering work of Bonacich (1972) who unified earlier attempts to rank nodes based on the information contained in the network's adjacency matrix by showing that they were all equivalent to computing the leading (first) eigenvector of such a matrix and using the resulting scores to rank the nodes from highest to lowest.[1]

Since then, the Bonacich approach to spectral ranking of nodes in a network has entered the canon of core contributions in mathematical sociology and, more generally, in Network Science under the moniker of "Eigenvector Centrality." When considered under the family of centrality measures for social networks, the Bonacich (1972) eigenvector approach can be thought of as a generalization of degree centrality,[2] in that actors are ranked based not just on the number of direct (one-step) connections they have with other actors. Instead, each connection is weighted by the number of alters of the nodes at the other end, such that connections to better-connected actors count more in determining the focal node's ranking. In turn, alters who are themselves connected to better-connected others count more in determining a given actor's rank, as do connections to others who themselves have longer indirect connections to better-connected others (up to the diameter of the graph).

Thus, in contrast to degree, which is a purely "local" measure, the eigenvector ranking uses information on the graph's *global* connectivity structure, ordering actors based on all direct *indirect* connections to others, with shorter indirect connections to better-connected others contributing more to an actor's score than longer paths to the same others or shorter paths to others who were not so well-connected. In this way, the eigenvector ranking also carries with it an idea of "attenuation" based on the length of indirect paths connecting pairs of actors in the system, such that the contribution that each of the other actors make to the focal actors score is weighted not just by their own (direct and indirect) connectivity to others, but also by how "close" they are to the focal actor as given by the geodesic distance (Vigna 2016).

In this respect, the intellectual career of eigenvector decomposition in SNA resolves itself into a contribution to the broader—measurement-driven—topic of "centrality measures," with Bonacich (2007) himself contributing strong arguments for the conceptual and computational advantages of the eigenvector approach as compared to the other "big two" centralities in the Freeman (1978) "graph theoretic" tradition, such as closeness and betweenness. Yet, one contrast between eigenvector centrality and the "classical"

[1]The network adjacency matrix is a (typically) square matrix containing information on the pairwise proximities (e.g., friendship, advice, similarities, frequency of interaction) between actors in a network

[2]See Borgatti and Everett (2006: 469-471) for a typically lucid exposition of this argument.

Freeman centralities is that the latter's more transparent foundation in graph theory also links them more straightforwardly to underlying "mechanistic" models of network process, which can serve as their theoretical foundation and motivation (Friedkin 1991; Borgatti 2005).

For instance, closeness can be defined as the ranking of actors in a social network based on how *early* they would, on average, receive something diffusing through the network, across various runs of a diffusion process (which can be good or bad, depending on the nature of the diffusing thing), with actors who would receive the diffusing object the earliest ranked higher than those who would have to wait a long time. In a similar vein, betweenness has been classically defined as a measure of power through *gatekeeping* and control over the diffusion of information (or other goods) within the network, with actors whose structural position (e.g., being in the "middle" of many shortest paths between pairs of other actors) would allow them to control whether other actors receive something ranked highest and actors who lack such control— such as a pendant node—ranked lowest (Freeman 1980). Ironically, and perhaps because eigenvector decomposition is itself a somewhat "black box" technique, what is arguably the "best" approach to ranking nodes in social networks based on connectivity structure information (Bonacich 2007), also seems to have the shakiest theoretical foundations in terms of a transparent model of network process (Borgatti and Halgin 2011).

Given this backdrop, this chapter has two main goals. First, to provide a theoretical motivation for the use of eigenvector-based ranking approaches in social networks in terms of an underlying model of social process. This model is a version of a diffusion model, in which the status of every actor is the "thing" that diffuses through the network. Secondly, to revisit the issue of the relations between different variants of eigenvector-based ranking that have accumulated since Bonacich's foundational contribution—some of which have been proposed outside sociology—by showing that they all apply the same underlying status diffusion model, but differ in terms of the proximity matrix between actors that they see as serving as the substrate over which the diffusion process plays out. This exercise shows that the use of eigenvectors in SNA is much more general than the canonical Eigenvector Centrality scoring. The ultimate goal is to demystify the use of eigenvectors in SNA by showing them to be a well-justified "clear box" model, rather than a mechanical, obscure tool that produces rankings of actors with unclear or ambiguous connections to an underlying generative process model.

2 Using Eigenvectors to Rank Nodes

2.1 Networks as Prisms

As noted, the primary use of eigenvectors in SNA has been as a black-box technique to compute scores that can be used to rank nodes in a network— a now multidisciplinary practice Vigna (2016) has referred to as "spectral ranking." These scores are used as a measure of a structural index of node position called "Eigenvector Centrality," which analysts subsequently give various localized substantive interpretations (depending on the application) and may even include in linear regression models as either predictors or outcomes (e.g., Rossman et al. 2010).

Note the problem here. The idea of "Eigenvector Centrality" is a doubly obscure compound term with the first nominal referring to a linear algebra technique (itself obscure in terms of the mechanics of computation) and the second nominal referring to one of the most ambiguous and conceptually inflated ideas in the entire field of SNA. As Borgatti and Everett (2006) have noted, calling something a "centrality" measure does not clarify much beyond saying that you are trying to see whether a node is "well-placed" in the network, an idea that is subject to multiple (sometimes conflicting) interpretations. In this sense, both the various variations of the theme of betweenness or closeness centrality are in a better place, since, computation-wise, they reduce to sums or averages of relatively straightforward graph-theoretic quantities, and they can be tied to relatively transparent models of network processes, as intimated earlier.

Let us thus step back and see whether we can motivate a conception of "well-placedness" in a network that fits what the Eigenvector scores compute, but independently of the technical machinery underlying these scores. There is, indeed, one such well-motivated intuition as to what it means to be "well-placed" in a network. Here, and in contrast to the generative process models that characterize both closeness and betweenness (Borgatti 2005), the ties in the network are seen less as "pipes" that transmit stuff and more like "prisms" that reflect on each actor (Podolny 2001).

If you want to think about this version of "well-placedness" using the usual flow imagery that animates closeness and betweenness (Borgatti and Halgin 2011), then you can say that what is transmitted through the network is the *network itself*, or more accurately, the importance, status, and prestige of each actor, presumably flowing from those of high status to those with less. Under this interpretation, actors get status and prestige in the network from being connected to prestigious and high-status others (hence the "prismatic" nature of social ties under this model). Those others, in turn, get their status from being connected to high-status others, and so on ad infinitum.

Folowing Bonacich (1972), one way to quantify these ideas is as follows. If $\mathbf{s}$ is a vector containing the desired status scores, then the status of each actor i should be equal to:

$$s_i = \sum_j a_{ij} s_j \tag{1}$$

Where $a_{ij} = 1$ if i is adjacent to j in the graph that represents the network. Note that in this formula, the status of a given actor s_i is just the sum of the status scores of all the others each actor is connected to s_j, thus succinctly and elegantly capturing the sociological idea that anyone's status is just a function (in this case, the simplest additive function) of the status of the people they are connected to.

In matrix notation, if is $\mathbf{s}$ a column vector of status scores then:

$$\mathbf{s} = \mathbf{A}\mathbf{s} \tag{2}$$

Where $\mathbf{A}$ is the network's adjacency matrix; because $\mathbf{A}$ is an $N \times N$ matrix (where N is the number of actors in the network) and $\mathbf{s}$ is an $N \times 1$ column vector, the (matrix) multiplication of $\mathbf{A}$ times the vector will return another column vector of dimensions $N \times 1$, in this case $\mathbf{s}$ itself!

It is easy to see that this formulation of status in networks, despite its simplicity, elegance, and intuitive appeal, poses a problem because $\mathbf{s}$ appears on both sides of the equation, meaning that to know the status of any one node we would need to know the status of the others, but calculating the status of the others depends on knowing the status of the focal node, and so on. There's a chicken and the egg problem here.

Now it is usually at this step where the machinery of eigenvectors and eigenvalues is carted in as a purely technical solution to what began as a *conceptual* problem, for there is an obvious (to the math oriented) technical solution to the issue posed by equation 2, as there's a class of solvable (under some mild conditions imposed on the $\mathbf{A}$ matrix like being of full rank) linear matrix algebra problems that take the form:

$$\lambda\mathbf{s} = \mathbf{A}\mathbf{s} \tag{3}$$

Where λ is just a plain old number (a scalar). Once again, conditional on the aforementioned mild conditions being met, we can use various standing iterative algorithms that are guaranteed to find the values of λ and $\mathbf{s}$ (if they exist among the real numbers) that make the above equality true. When we do that successfully, we say that the value of λ we hit upon is an *eigenvalue* of the matrix and the values of the $\mathbf{s}$ vector we came up with are an *eigenvector* of the same matrix (technically, in the above equation, a right eigenvector). Eigenvalues and eigenvectors, like Don Quixote and Sancho Panza, come in pairs, because you need a unique combination of both to solve the equation. Typically, a given matrix (like the adjacency matrix $\mathbf{A}$) will have multiple pairs that will solve the equation (up to $N - 1$, or the rank of the matrix).

Typically, for most matrices with interesting structure (like a network adjacency matrix), we can order the eigenvalues in terms of magnitude from largest to lowest, meaning that there will be a largest eigenvalue (or a tie between multiple eigenvalues for the title of the largest). The eigenvector associated with the largest eigenvalue is called the leading (sometimes "dominant") eigenvector of the matrix. These are the usual scores taken as the Eigenvector Centralities in SNA.

Note that all this talk about eigenvalues and eigenvectors has already gotten pretty obscure; it is just matrix linear algebra gobbledygook. It has nothing to do with networks and social structure, or with the elegant definition of status and prestige we just outlined. In contrast, because the big two centrality measures have a direct foundation in graph theory, and graph theory is an *isomorphic* model of social structures (points map to actors/people and lines map to relations) the "math" we do with graph theory when computing centrality scores (which is typically simple arithmetic) is directly meaningful as a *model* of networks (Borgatti and Halgin 2011).

Eigenvalues and eigenvectors are *not*—at least as usually presented—a model of social structure in the way graph theory is (their first scientific application was in Chemistry and Physics). They are just a mechanical math fix to a circular equation problem. This is why it's a mistake to introduce network measures of status and prestige by jumping directly to the machinery of linear algebra.

A better approach is to see whether we can motivate the use of measures like those captured by Equation 3 using the simple model of the distribution of status and prestige we started with earlier. We will see that we can, and that doing that leads us back to solutions that are the mathematical equivalent of the eigenvector shenanigans.

2.2 Distributing Status to Others

Let us start with the simplest model of how people can get their status from others' status in a network. It is the simplest because it is based on degree.

Imagine everyone has the same "quantum" of status to distribute at the beginning (which can be stored in a vector of length N with constant values). Let us call these the "old" status scores. Then, at each step, people "send" their designated amount of status to all their contacts in the network and receive the same quantum from each of them. These are the "new" or updated scores, which we can calculate using Equation 2. Then we repeat the process, with everyone sending the amount of status they now have in their new scores and receiving the same from others. We repeat this many times and stop when the new scores computed at the current step are not much different from the old scores calculated in the previous step.

Words are nice, but it is better to formally write our proposed status distribution model as a little *program* that anyone can run on their computer if they wish (like a simple "agent-based model" of the status game). The pseudocode for such a program is shown in Algorithm 1.

Algorithm 1 Status distribution game algorithm

Ensure: $\delta > 1 \times 10^{-10}$

1: $\mathbf{s}^n \leftarrow \mathbf{1}$ ▷ Initializing status scores to the all-ones vector of length N
2: **while** $\delta > 1 \times 10^{-10}$ **do**
3: $\quad \mathbf{s}^o \leftarrow \mathbf{s}^n$ ▷ Old scores equal to previous new scores
4: $\quad \mathbf{s}^n \leftarrow \mathbf{A}\mathbf{s}^o$ ▷ Updating new scores based on old scores
5: $\quad \mathbf{s}^n \leftarrow \frac{\mathbf{s^n}}{||\mathbf{s}^n||_2}$ ▷ Normalizing new status vector using the L-2 (Euclidean) norm
6: $\quad \delta \leftarrow \sum_{k=1}^{N} |s_k^n| - |s_k^o|$ ▷ Computing difference between new and old score vectors
7: **end while**
8: **return** $\mathbf{s}^n$

Let us see what this Algorithm does. Line 1 initializes the initial status vector for each node to the all-ones vector $\mathbf{1}$ of length N. Then lines 2-7 implement a little looping routine of how status is distributed through the network, with the most important piece of pseudocode being line 4, where the current status scores for each node are updated to equal the sum of the status scores of its neighbors computed one iteration earlier, just like in equation 2. The loop stops when the difference (δ) between the old and the new scores is negligible, as checked in line 6. Note the normalization step on line 5, where we divide the new status score vector by a normalized version of the same vector. This is required to prevent the sum of status scores from growing without bound. While we could use any normalization (e.g., maximum or sum), algorithm 1 implements the Euclidean vector norm (also called the L-2 norm), which for any vector $\mathbf{s}$ is given by:

$$||\mathbf{s}||_2 = \sqrt{\sum_k (s_k)^2} \tag{4}$$

The perspicacious reader may already have suspected that I have pulled a fast one here, for algorithm 1 is nothing but an implementation of the so-called "power method", which is guaranteed to find the dominant eigenvector of matrix $\mathbf{A}$ (see Newman 2010: 346-348). I submit, however, that rather than a mechanical method for obtaining that leading eigenvalue of the adjacency matrix, the status game algorithm instantiates a generative process model of how status emerges in social networks. This model makes specific assumptions (e.g., the "quantum" of status each node gets to distribute is the same regardless of each node's structural characteristics; status is given by the *sum* of the status of alters) that other models (or "centrality measures") modify in systematic ways, as we will see.

2.3 PageRank Prestige Scoring

The model of status distribution implied by the "Eigenvector Centrality" approach we just considered holds that each actor distributes the same "quantum" of status to others in the network, independent of the number of connections they have. Status just replicates indefinitely. Thus, a node with 100 friends has 100 status units to distribute, and a node with 10 friends has 10 units. This is why the Bonacich (1972) eigenvector idea rewards nodes that are more connected to popular nodes. Even though everyone begins with a single unit of status, well-connected nodes by degree end up having more to distribute, and people connected to those nodes receive more status.

But what if status propagated in the network *proportionately* to the number of connections one had? For instance, if someone has 100 friends and only has so much time or energy, they would have only a fraction of status to distribute to others than a person with 10 friends. In that case, the node with 100 friends would have only 1/100 of the status units to distribute to each of its connections, while the node with 10 friends would have 1/10 units. Under this formulation, being connected to *discerning* others, that is, people who only connect to a few, is better than being connected to others who connect to everyone else indiscriminately. Note that this model of status, while being in many ways the opposite of that implied by the traditional "Eigenvector Centrality" is no less compelling, and in fact may be more appropriate in specific substantive settings.

How would we implement this model? First, let's create a variation of the undirected friendship nomination adjacency matrix called the $\mathbf{P}$ matrix. It is defined like this:

$$\mathbf{P} = \mathbf{D}^{-1}\mathbf{A} \tag{5}$$

Where $\mathbf{A}$ is our old friend, the adjacency matrix, and $\mathbf{D}^{-1}$ is a matrix containing the *inverse* of each node's degree along the diagonals and zeroes in every other cell. So the resulting matrix $\mathbf{P}$ is the original adjacency matrix, in which each non-zero entry is equal to one divided by the degree of the corresponding node in each row; each cell in the $\mathbf{P}$ matrix is thus equal to $1/k_i$ where k_i is the degree of the node in row i, meaning that each cell p_{ij} will contain a value between 0 and 1. Moreover, the $\mathbf{P}$ matrix is *asymmetric*; that is, p_{ij} is not necessarily equal to p_{ji}. In fact, p_{ij} will only be equal to p_{ji} when $k_i = k_j$ (nodes have the same degree). Finally, the $\mathbf{P}$ matrix is row stochastic. That is, the "outdegree" of each node in the matrix is forced to sum to 1.0. However, the "indegree" of each node in the same matrix is not so constrained. This means that inequalities in the system will be tied to the *incoming* weight of the ties of each node in the $\mathbf{P}$ matrix, which is given by either the column sums of the matrix or the row sums of the transpose of the same matrix $\mathbf{P}^T$.

Substantively, this means we are equalizing the total amount of prestige or status each node can distribute within the system to a fixed quantity. Nodes with many neighbors will dissipate this quantity by distributing it across a larger number of recipients; hence, their corresponding non-zero entries in the rows of $\mathbf{P}$ will be a small number. Meanwhile, nodes with few neighbors will have more to distribute and thus their neighbors will get a higher quantum of status from them. Hence, while the $\mathbf{P}$ matrix has many interpretations (e.g., as a Markov transition matrix for a random-walk diffusion process through the network), here, it just quantifies the idea that the amount of centrality each node can distribute is proportional to its degree, and that the larger the degree, the less there is to distribute.

Now, if we were to play the distribution game using the matrix $\mathbf{P}^T$ instead of $\mathbf{A}$ in algorithm 1 (e.g., substituting $\mathbf{s}^n \leftarrow \mathbf{P}^T\mathbf{s}^o$ in line 4), the resulting scores at convergence would be equivalent to the "PageRank" prestige score of Brin and Page (1998) (corresponding to the leading right eigenvector of the $\mathbf{P}^T$ matrix or the leading left eigenvector of the $\mathbf{P}$ matrix). Thus, the Bonacich (1972) "Eigenvector centralities" and the "PageRank" scores can be seen as alternative models of how status is distributed within a system that make distinct assumptions about the link between the "quantum" of status each actor has to distribute and that actor's degree, with Bonacich (1972) assuming this quantity is independent of degree and PageRank assuming that it declines proportionately to degree.[3]

[3] As Vigna (2016) notes, the PageRank scoring is a rediscovery of the original spectral ranking method proposed by Seeley (1949).

2.4 Hubs and Authorities

So far, we have considered two ways in which status can be distributed in undirected networks. What happens when we have a social network composed of directed relations? Well, consider two ways of showing your status in a system governed by directed relations (like advice). You can be highly sought after by others (be an "authority"), or you can be discerning in who you seek advice from, preferably seeking out people who are also sought after (e.g., be a "hub" pointing to high-quality others). These two forms of status are mutually defining (Bonacich and Lloyd 2001). The top authorities are those who are sought after by the top hubs, and the top hubs are the ones who seek the top authorities!

So this leads to a doubling of the Bonacich (1972) status accounting:

$$s_i^h = \sum_j a_{ij} s_j^a \tag{6a}$$

$$s_i^a = \sum_i a_{ij} s_i^h \tag{6b}$$

This says that the hub status score s^h of a node is the sum of the authority status scores s^a of the nodes they point to (sum over j; the outdegree), and the same authority status score of a node is the sum of the hub scores of the nodes that point to it (sum over i; the indegree). The same chicken-and-egg problem we encountered in the undirected case rears its ugly head here as well. To determine a node's hub score, we need the authority scores of the other nodes, but to figure out those scores, we need the hub scores of those nodes. Once again, it is likely that the machinery of eigenvectors will come to the rescue just like in the undirected case (Bonacich and Lloyd 2001). But, as in that case, it is more illuminating to view this as a variation of the status distribution model we have been considering.

To account for the duality expressed by equation 6, we need to make our status distribution game a bit more complicated (but not too much). A version of the more elaborate status distribution model for directed graphs is shown in algorithm 2. Note that everything is the status distribution game of algorithm 1, except that we now keep track of two mutually defining status scores, a and h. We first initialize the authority scores to the all ones vector $\mathbf{1}$ of length N in line 1. The while loop in lines 2-9 then updates the hub scores (to be the sum of the authority scores of each out-neighbor) in line 4, normalizes them in line 5, and updates the new authority scores to be the sum (across each in-neighbor) of these new normalized hub scores, which are then themselves normalized in line 7.

So at each step t, the authority and hub scores are calculated like this:

$$s_i^h(t) = \sum_j a_{ij} s_j^a(t-1) \tag{7a}$$

$$s_i^a(t) = \sum_j a_{ij}^T s_j^h(t) \tag{7b}$$

Algorithm 2 Status distribution game (directed) algorithm

Ensure: $\delta > 1 \times 10^{-10}$

$\mathbf{a}^n \leftarrow \mathbf{1}$ ▷ Initializing authority scores to the all-ones vector of length N

while $\delta > 1 \times 10^{-10}$ **do**

  $\mathbf{a}^o \leftarrow \mathbf{a}^n$ ▷ Old scores equal to previous new scores

  $\mathbf{h}^n \leftarrow \mathbf{A}\mathbf{a}^o$ ▷ Updating new hub scores based on old authority scores

  $\mathbf{h}^n \leftarrow \frac{\mathbf{h}^\mathbf{n}}{||\mathbf{h}^n||_2}$ ▷ Normalizing new hub score vector using the L-2 norm

  $\mathbf{a}^n \leftarrow \mathbf{A}^T\mathbf{h}^n$ ▷ Updating new authority scores based on current hub scores

  $\mathbf{a}^n \leftarrow \frac{\mathbf{a}^\mathbf{n}}{||\mathbf{a}^n||_2}$ ▷ Normalizing new authority score vector using the L-2 norm

  $\delta \leftarrow \sum_{k=1}^{N} |a_k^n| - |a_k^o|$ ▷ Computing difference between new and old authority score vectors

end while

return $\mathbf{h}^n$

return $\mathbf{a}^n$

Where a_{ij}^T is the corresponding entry in the transpose of the adjacency matrix. As you may have guessed, algorithm 2 is just an implementation of the "HITS" algorithm developed by (Kleinberg 1999).[4] More importantly, it is a model-based specification of a dual status distribution regime. Nodes can get status either by being "good pointers" to high-status others or by being pointed to by good pointers.

Of course, had we begun immediately with the eigenvector machinery (Bonacich and Lloyd 2001), we would have missed this connection to the status distribution game. Just like we saw with the traditional Bonacich (1972) Eigenvector centrality and the PageRank scoring, we can be sure that the hub and authority scores computed by algorithm 2 are the leading eigenvectors of *some* proximity matrix between nodes in the directed graph. Let us find out which ones.

Consider the matrices:

$$\mathbf{M}^h = \mathbf{A}\mathbf{A}^T \tag{8a}$$

$$\mathbf{M}^a = \mathbf{A}^T\mathbf{A} \tag{8b}$$

It is well known that the off-diagonal entries of $\mathbf{M}^h$ contain the common out-neighbors of each pair of nodes, and the off-diagonal entries of $\mathbf{M}^a$ contain their common in-neighbors. Meanwhile, the diagonal entries of $\mathbf{M}^h$ contain the outdegree vector of each node, while the diagonal entries of $\mathbf{M}^a$ contain the indegree vector.

In some substantive contexts, $\mathbf{M}^h$ and $\mathbf{M}^a$ matrices have special interpretations. For instance, in an information network, the papers that point to other papers are their shared references. Therefore, the number of common out-neighbors of two papers is

[4] HITS is an acronym for the unfortunate and impossible to remember name "Hypertext Induced Topic Selection" reflecting the origins of the approach in web-based information science.

called the *bibliographic coupling* score. The number of common in-neighbors between two papers, by way of contrast, is called their *co-citation* score. Both bibliographic coupling and co-citation scores capture two ways that nodes can be similar in a directed graph. In the social context of advice seeking, for instance, two people can be similar if they seek advice from the same others, or if they are sought after for advice by the same others.[5]

The $\mathbf{M}^h$ and $\mathbf{M}^a$ matrices, therefore, are two (unweighted) similarity matrices between the nodes in a directed graph. As you may also be suspecting, the hub and authorities scores are the leading eigenvectors of the $\mathbf{M}^h$ and $\mathbf{M}^a$ matrices (Kleinberg 1999; Bonacich and Lloyd 2001):

$$\lambda \mathbf{s}^h = (\mathbf{A}\mathbf{A}^T)\mathbf{s}^h \tag{9a}$$

$$\lambda \mathbf{s}^a = (\mathbf{A}^T\mathbf{A})\mathbf{s}^a \tag{9b}$$

Note that this also means we could have obtained the hub and authority scores using our initial status distribution game (algorithm 1), but we would have had to play the game twice, once for the matrix $\mathbf{M}^a$ and the other one for the matrix $\mathbf{M}^h$. Once again, this hints that the status distribution game is a *generalized* mechanism, and that the Bonacich, PageRank, and HITS scores are just special cases that make distinct assumptions about how status is distributed in the system. These assumptions are then encoded in the particular proximity matrix that is considered: Either the adjacency matrix $\mathbf{A}$, the Markov transition matrix $\mathbf{P}$, or the common-neighbors proximity matrix $\mathbf{M}$ (Taylor et al. 2021: 117). Other proximity matrices would thus yield different "status" rankings under their own proprietary assumptions.[6]

3 Using Eigenvectors to Group Nodes

3.1 PCA as HITS on a Centered Adjacency Matrix

As noted, eigenvector-based decomposition methods in SNA are mainly associated with the pioneering work of Bonacich (1972), which means that in the field, they are primarily thought of as a method to induce "centrality" scores (or prestige scores), and thus arrange nodes in an ordinal rank.

Yet, in computer and information science, eigenvector-based methods for relational data have a different intellectual history, being mainly viewed as methods for *clustering*

[5]Given the connection to the HITS dual status ranking, sometimes the $\mathbf{M}^h$ is called the *hub matrix* and the $\mathbf{M}^a$ is called the *authority matrix* (Ding et al. 2002).

[6]In other cases, such as the Katz (1953) proximity matrix, the result ends up being the substantively the same as using the original adjacency matrix, since the Katz centrality scores can be obtained by using the "perturbed" (see Vigna 2016) version of the adjacency matrix $\alpha\mathbf{A}$ in algorithm 1, which end up converging to the Bonacich scores as α approximates the leading eigenvalue of the matrix (Bonacich 1987).

nodes in a graph, as in the extensive literature on spectral clustering (Dam et al. 2021). They thus belong to the suite of strategies a sociologist would use to identify nodes occupying similar structural positions (Burt 1976) or belonging to cohesive subgroups (Moody and White 2003), which is usually considered a completely different task from centrality analysis.

We will see that these two tasks are actually more interrelated than the history of uptake of eigenvector-based techniques in SNA suggests. The basic idea goes like this: As noted, when played over a matrix indicating proximities or affinities between nodes (e.g., based on direct or indirect connectivity), the status game in algorithm 1 will yield a version of a "prestige" or "status" score, linearly ranking nodes in a network. However, when played over a matrix indicating *differences* in interaction propensities above or below some baseline, the same status diffusion game will reveal *partitions* in the social structure, either based on structural (dis)similarity or interaction propensities.

One such unheralded link in the SNA literature is that between Eigenvector scoring for networks represented as directed graphs (like the HITS approach reviewed earlier) and long-standing eigenvector-based techniques in the social and behavioral sciences, usually applied to cases by variables data, such as Principal Components Analysis (PCA) or Factor Analysis (FA). Using these methods, the spectral decomposition of a matrix summarizing the distance between variables (such as the correlation or covariance matrix) can reveal "axes" that summarize the most significant amount of variance among the cases. Is there a connection between these types of multivariate analysis techniques and Eigenvector-based approaches in social networks?

Saerens and Fouss (2005) argue that there is, indeed, an intimate relationship between the HITS dual ranking scores—and by implication the Bonacich (1972) Eigenvector Centrality—and one of the most widely used multivariate analysis techniques, Principal Components Analysis (PCA). They argue that HITS is just PCA applied to an *uncentered data matrix* which, in the case of relational data, is the network's adjacency matrix. If Saerens and Fouss (2005) are right then, we could also just argue that the PCA applied to one-mode network data is just HITS (or more accurately, the dual status distribution game described by algorithm 2) played over a *centered adjacency matrix*.

Let us see how this works. First, let us define the centered version of the network adjacency matrix. To center a typical data matrix (e.g., cases in rows, variables in columns), we subtract each case's score on a given variable (column) from the column mean (the average score across all cases on that variable). Similarly, to obtain the centered network adjacency matrix, we first need to compute the corresponding column means. Note that for the typical network data, this will be equivalent to a vector $\langle \mathbf{d}^{in} \rangle$ containing each node's in-degree (the sum of the columns) divided by the total number of nodes in the network, which is kind of a normalized degree centrality vector.[7] This means that each cell of the centered adjacency matrix $\langle \mathbf{A} \rangle$ is given by:

[7] Except the usual normalization is to divide by $n - 1$ (to have a maximum of 1.0).

$$\langle a_{ij} \rangle = a_{ij} - \langle d_j^{in} \rangle \tag{10a}$$

$$\langle \mathbf{d}^{in} \rangle = \frac{\mathbf{d}^{in}}{N} \tag{10b}$$

Where $\langle d_j^{in} \rangle$ is the j^{th} element of the $\langle \mathbf{d}^{in} \rangle$ vector of normalized in-degrees. Note that the centered adjacency matrix $\langle \mathbf{A} \rangle$ will contain entries equal to either $-\langle d_j^{in} \rangle$ when the two nodes are not connected or $1 - \langle d_j^{in} \rangle$ when the two nodes are adjacent, meaning that each of the columns sum to zero (e.g., $\sum_j \langle a_{ij} \rangle = 0$ for each node i). The key result is that playing the dual status game implied by algorithm 2 over the $\langle \mathbf{A} \rangle$ matrix produces equivalent scores as would be obtained by a standard PCA of the adjacency matrix, showing that indeed, PCA is a version of HITS played over a centered adjacency matrix.

The same result can be obtained by playing the status distribution game implied by algorithm 1 over the *covariance* matrix implied by the adjacency matrix of the network, as PCA resolves itself to the spectral decomposition of the covariance matrix between variables in the usual rectangular cases by variables data set. In the context of one-mode network data, the relevant covariance matrix $\mathbf{K}$ is the one between the column vectors recording each incoming tie for each node (the columns of the adjacency matrix). Each entry in this matrix records the covariance between pairs of nodes $\{i, j\}$, which is given by:

$$k_{i,j} = \frac{\sum_{q=1}^{N} [(\mathbf{a}_{i(q)} - \bar{\mathbf{a}}_i)(\mathbf{a}_{j(q)} - \bar{\mathbf{a}}_j)]_q}{N - 1} \tag{11}$$

Where $\mathbf{a}_i$ is the column vector of the adjacency matrix corresponding to the i^{th} node and $\bar{\mathbf{a}}_i$ is the mean of the entries of the vector. Running the status game described by algorithm 1 over the $\mathbf{K}$ matrix of pairwise covariances produces scores ($\mathbf{s}^{h(pca)}$) equivalent to authority scores produced by running algorithm 2 on the centered version of the adjacency matrix (returning the leading eigenvector of the node covariance matrix). The centered hub scores (corresponding to the PCA scores for the rows) can then be obtained from:

$$\mathbf{s}^{h(pca)} = \mathbf{A}\mathbf{s}^{h(pca)} \tag{12}$$

Which means that the $\mathbf{s}^{h(pca)}$ scores of each node are just the sum of the $\mathbf{s}^{a(pca)}$ scores of the nodes it points to.

The PCA version of the HITS scores assigns high positive values on each metric to nodes that are the "purest" types of hubs (or authorities). Accordingly, nodes that have high authority scores but low hub scores receive high values of $\mathbf{s}^{h(pca)}$, nodes with high hub scores and low authority scores receive high values of $\mathbf{s}^{a(pca)}$. In contrast, nodes with dual roles as both hubs and authorities (high scores on both) receive lower scores on the respective metrics. Thus, the PCA-like status distribution scores reveal clusters of nodes that play particular specialized "roles" in a network (Burt 1976). These should be evident from an embedding of each node in a two-dimensional space with axes defined by the $\mathbf{s}^{a(pca)}$ and $\mathbf{s}^{h(pca)}$ scores (the analogue of the usual joint display of individuals and variables in PCA).

3.2 Using Eigenvectors to Find Communities

Newman (2006) shows that we can use the spectral partitioning of a specialized centered matrix—the famous *modularity matrix*—to find clusters of nodes that interact selectively with one another (e.g., groups as opposed to roles). This approach can be seen as a variation and a generalization of the PCA for the the centered adjacency matrix we considered in the previous section, and thus a generalization of the status distribution game to find clusters of preferentially interacting nodes in networks.

To see this, recall that the modularity matrix $\mathbf{B}$ is defined as a variation or adjustment of the adjacency matrix $\mathbf{A}$, where each cell b_{ij} takes the value:

$$b_{ij} = a_{ij} - \frac{d_i^{out} d_j^{in}}{\sum_i \sum_j a_{ij}} \tag{13}$$

Where k_i^{out} is node i's outdegree and k_j^{in} is node j's indegree. Note that the entries in the modularity matrix have an "observed minus expected" structure.[8] In this case, we compare whether we see a link from i to j—as given by a_{ij} (or the strength of the link w_{ij} if the adjacency matrix is weighted)—against the chances of observing a link in a graph in which nodes connect at random with probability proportional to their degrees (or the out and in "node strength" $\sum_j w_{ij}$ and $\sum_i w_{ij}$ if the graph is weighted), as given by the right-hand side fraction. The denominator is total number of edges E (or the weighted graph volume $\sum_i \sum_j w_{ij}$ if the graph is weighted).[9]

The modularity matrix is distinctive in that, unlike the centered adjacency matrix in equation 10, which is only centered in one of the "ways" (the columns), the modularity matrix is *doubly centered*; that is for each node i, $\sum_j b_{ij} = 0$ *and* $\sum_i b_{ij} = 0$. Like the centered adjacency matrix, the modularity matrix will contain some negative entries, some positive entries, and some close to or actually zero. We interpret these as follows: If an entry is negative, it means that a link between the row and column nodes is much less likely to happen than expected given each node's degrees; a positive entry indicates the opposite; a larger than expected chance of a link forming, conditional on the degree of the nodes at the two ends of the edge. Entries close to zero indicate those nodes have close to a random chance of forming a tie.

[8] "Expected" given a suitable choice for a null model, which in equation 13 is the so-called *configuration model* given by the fraction on the right-hand side, which says that the connection between nodes is proportional to their degrees. Other null models are possible, yielding alternative versions of the modularity matrix (Newman 2006).

[9] Note that if the graph is undirected, the entries in the modularity matrix are just: $a_{ij} - \frac{d_i d_j}{\sum_i \sum_j a_{ij}}$.

As you may have guessed, playing the status distribution game of algorithm 1 on the modularity matrix **B** results in a vector of scores (the leading eigenvector of the modularity matrix). If the network has discernible community structure, some of these entries will be positive, some negative, and some near zero. Binarizing this vector such that entries above zero go in one bucket and entries below or equal to zero go in another bucket, results in a optimal bipartition of the nodes of the matrix into two communities, with maximal number of interactions within and minimal number of connections between (an approximation to the bipartition that maximizes a quality function based on the modularity).[10]

In this way, the same principle we saw illustrated above for the case of the PCA of the adjacency matrix applies: playing the status distribution game on a matrix that captures disaffinities between nodes (e.g., differences in the likelihood of interaction) yields distinct clusters of nodes rather than producing one-dimensional ranks. This approach can, of course, be generalized into a divisive clustering strategy (Newman 2006), similar to CONCOR (Breiger et al. 1975), where we first split the set of nodes into two maximally interacting communities, then we split those communities in two based on the same approach (running the status game in the modularity matrix subgraph) check to see if this increase results in an overall increase in the modularity, and stop when subsequent partitions do not improve the modularity score.

Concluding Remarks

The preceding analysis reframes the use of eigenvector-based methods in Social Network Analysis (SNA), moving them from an obscure, “black-box” mathematical technique to a transparent “clear-box” model grounded in an explicit generative social process. The chapter shows that Eigenvector Centrality, PageRank, and the HITS algorithm are all computationally equivalent to an iterative status-distribution game that converges to the leading eigenvector of a specific proximity matrix (e.g., an adjacency, Markov transition, or common-neighbors matrix). By establishing this common, intuitive framework, the paper argues that the differences between these measures are not arbitrary mathematical variations but rather systematic differences in the underlying assumptions about how status, prestige, or influence diffuses through a social network. Furthermore, this status distribution framework was shown to generalize beyond one-dimensional ranking; when applied to centered proximity matrices, such as those used in Principal Components Analysis (PCA) or the Modularity matrix, the same iterative process is shown to reveal structural groupings and communities rather than a simple ordinal rank, bridging the historical divide between centrality analysis and structural grouping in SNA.

Building on this demystification, future research should leverage this generalized status diffusion model to systematically explore a wider family of spectral measures. One

[10]For a suggested partition of a graph based on m groups, the modularity Q of the suggested partition is given by: $Q = \sum_{k=1}^{m} \left[\frac{\sum_{i \in k} \sum_{j \in k} a_{ij}}{\sum_i \sum_j a_{ij}} - \frac{(\sum_{i \in k} \sum_j a_{ij})(\sum_i \sum_{j \in k} a_{ij})}{(\sum_i \sum_j a_{ij})^2} \right]$.

clear path is to investigate the effects of playing the status game on other theoretically meaningful proximity matrices, such as those derived from blockmodeling or those capturing signed/multiplex relations, to see what new types of structural scores they reveal. Secondly, the model offers a clear-cut way to formally test and compare competing sociological theories of prestige and influence: Different theories can be codified as distinct proximity matrices or different rules for status distribution (e.g., non-linear functions of alter's status), allowing researchers to compare which model best predicts empirical outcomes. Finally, extending the framework to dynamic networks is a promising avenue (see e.g., Taylor et al. 2017) where the continuous iteration of the status game can model the real-time co-evolution of network structure and status/prestige, potentially offering new insight into mechanisms of stratification and social change.

References

Bonacich, Phillip. 1972. "Factoring and weighting approaches to status scores and clique identification." *The Journal of Mathematical Sociology* 2:113–120.

Bonacich, Phillip. 1987. "Power and Centrality: A Family of Measures." *American Journal of Sociology* 92:1170–1182.

Bonacich, Phillip. 2007. "Some unique properties of eigenvector centrality." *Social Networks* 29:555–564.

Bonacich, Phillip and Paulette Lloyd. 2001. "Eigenvector-like measures of centrality for asymmetric relations." *Social Networks* 23:191–201.

Borgatti, Stephen P. 2005. "Centrality and network flow." *Social Networks* 27:55–71.

Borgatti, Stephen P and Martin G Everett. 2006. "A graph-theoretic perspective on centrality." *Social networks* 28:466–484.

Borgatti, Stephen P and Daniel S Halgin. 2011. "On Network Theory." *Organization Science* 22:1168–1181.

Breiger, Ronald L, Scott A Boorman, and Phipps Arabie. 1975. "An algorithm for clustering relational data with applications to social network analysis and comparison with multidimensional scaling." *Journal of Mathematical Psychology* 12:328–383.

Brin, Sergey and Lawrence Page. 1998. "The anatomy of a large-scale hypertextual Web search engine." *Computer Networks and ISDN Systems* 30:107–117.

Burt, Ronald S. 1976. "Positions in Networks*." *Social Forces* 55:93–122.

Dam, Alje van, Mark Dekker, Ignacio Morales-Castilla, Miguel Á Rodríguez, David Wichmann, and Mara Baudena. 2021. "Correspondence analysis, spectral clustering and graph embedding: applications to ecology and economic complexity." *Scientific reports* 11:8926.

Ding, Chris, Xiaofeng He, Parry Husbands, Hongyuan Zha, and Horst D. Simon. 2002. "PageRank, HITS and a unified framework for link analysis." *Proceedings of the 25th annual international ACM SIGIR conference on Research and development in information retrieval* pp. 353–354.

Freeman, Linton C. 1978. "Centrality in Social Networks: Conceptual Clarification." *Social networks* 1:215–239.

Freeman, Linton C. 1980. "The gatekeeper, pair-dependency and structural centrality." *Quality and Quantity* 14:585–592.

Friedkin, Noah E. 1991. "Theoretical Foundations for Centrality Measures." *American Journal of Sociology* 96:1478–1504.

Katz, Leo. 1953. "A new status index derived from sociometric analysis." *Psychometrika* 18:39–43.

Kleinberg, Jon M. 1999. "Authoritative sources in a hyperlinked environment." *Journal of the ACM (JACM)* 46:604–632.

Moody, James and Douglas R White. 2003. "Structural Cohesion and Embeddedness: A Hierarchical Concept of Social Groups." *American sociological review* 68:103–127.

Newman, M. E. J. 2006. "Finding community structure in networks using the eigenvectors of matrices." *Physical Review E* 74:036104.

Newman, Mark E. J. 2010. *Networks: An introduction*. Oxford University Press.

Podolny, Joel M. 2001. "Networks as the pipes and prisms of the market." *The American journal of sociology* 107:33–60.

Rossman, Gabriel, Nicole Esparza, and Phillip Bonacich. 2010. "I'd Like to Thank the Academy, Team Spillovers, and Network Centrality." *American Sociological Review* 75:31–51.

Saerens, Marco and Francois Fouss. 2005. "HITS is Principal Components Analysis." *The 2005 IEEE/WIC/ACM International Conference on Web Intelligence (WI'05)* pp. 782–785.

Seeley, John R. 1949. "The net of reciprocal influence; a problem in treating sociometric data." *Canadian Journal of Psychology / Revue Canadienne de Psychologie* 3:234–240.

Taylor, Dane, Sean A. Myers, Aaron Clauset, Mason A. Porter, and Peter J. Mucha. 2017. "Eigenvector-Based Centrality Measures for Temporal Networks." *Multiscale Modeling & Simulation* 15:537–574.

Taylor, Dane, Mason A. Porter, and Peter J. Mucha. 2021. "Tunable Eigenvector-Based Centralities for Multiplex and Temporal Networks." *Multiscale Modeling & Simulation* 19:113–147.

Vigna, Sebastiano. 2016. "Spectral ranking." *Network Science* 4:433–445.