ELSEVIER

Social Networks 22 (2000) 357–365

SOCIAL NETWORKS

www.elsevier.com/locate/socnet

Eigenvector-centrality — a node-centrality?☆

Britta Ruhnau*

Department of Operations Research, Universität Dortmund, D-44221 Dortmund, Germany

Abstract

Networks of social relations can be represented by graphs and socio- or adjacency-matrices and their structure can be analyzed using different concepts, one of them called centrality. We will provide a new formalization of a "node-centrality" which leads to some properties a measure of centrality has to satisfy. These properties allow to test given measures, for example measures based on degree, closeness, betweenness or Bonacich's eigenvector-centrality. It turns out that it depends on normalization whether eigenvector-centrality does satisfy the expected properties or not. © 2000 Elsevier Science B.V. All rights reserved.

1. Introduction

The concept of centrality has been discussed for over 50 years now. An intensive research regarding appropriate measures originated from the analysis of communication patterns and their performance in small groups by Bavelas (1950) and Leavitt (1951). For example, measures based on the degree of a node, on its distance to other nodes, on the betweenness or on algebraic concepts like the eigenvector of a matrix have been suggested. For an overview see, e.g. Wasserman and Faust (1997). Though lot of information about different centrality-measures is disposal:

> "There is certainly no unanimity on exactly what centrality is or on its conceptual foundations, and there is very little agreement on the proper procedure for its measurement." (Freeman (1977), p. 217)

We think this statement is still valid, in particular if the measures should allow comparisons between different graphs. Here the influence of network size has to be removed, thus a kind of normalization is required. Our aim is to develop a formal definition of a "node-centrality" that matches the intuitive understanding of centrality and is related to a fixed scale. For interpretation of measurement results the scale a value is related to provides important

☆ This is a revised version of a paper presented at the Symposium on Operations Research (SOR) in Magdeburg in 1999.

* Tel.: +49-231-5407.
E-mail address: b.ruhnau@wiso.uni-dortmund.de (B. Ruhnau).

PII: S0378-8733(00)00031-9

information. The interpretation of a value of, say, 0.8 will differ essentially if the maximal value that can be reached is 1 or if it is 10. Another reason for a fixed scale is that comparison of nodes from different graphs independent of graph size becomes possible.

We then compare "node-centrality" to several well-known centrality concepts. In particular we will provide conditions under which eigenvector-centrality (see, e.g. Bonacich (1972b)) is a node-centrality. Here it will turn out that normalization of the eigenvector has an important influence.

The paper is organized as follows. First we recall some basic notation and Freemans general procedure for normalizing point-centralities and developing an appropriate graph centralization measure (Freeman, 1979). Based on the notion of the "star", our central definition is developed in Section 3 and applied to different centrality-measures.

In Section 4 we recall the definition of eigenvector-centrality and examine different ways of normalizing the eigenvector. Their influence on eigenvector-centrality as a node-centrality is discussed. Section 5 concludes.

2. Basic notation and point-centrality

A social network can formally be represented by a graph $G = (V, E)$, where V is the set of nodes, or points, respectively, and E the set of edges. We restrict our considerations to undirected graphs. The distance $\mathrm{dist}(v_i, v_j)$ is the length of a geodesic, i.e. the number of edges in a shortest path between v_i and v_j. $d(v_i)$ is the degree of node v_i, and $G_n = \{G||V| = n\}$ is the set of all undirected and connected graphs with n nodes.

The description of actors in social networks is often done in terms of some "structural features" like the degree, closeness or betweenness of an actor. These structural features have been used to create measures of centrality for single nodes in a graph. Here we recall briefly the definition of these measures. [1]

Degree-centrality $C_{\mathrm{D}}(v_i) = d(v_i)$, $v_i \in V$, is based on the idea that the number of "direct relations" an actor has is an important feature of his structural position. *Closeness-centrality* $C_{\mathrm{C}}(v_i) = 1/\sum_{j=1}^{n}\mathrm{dist}(v_i, v_j)$, $v_i \in V$, takes into account how far away an actor is located from all other actors, and *Betweenness-centrality* $C_{\mathrm{B}}(v_i) = \sum_{k=1}^{n}\sum_{j:j>k}^{n}[g_{kj}(v_i)/g_{kj}]$, $v_i \in V$, relies on the extent to which other nodes depend on v_i as transmitter of, say, communication. Here g_{kj} is the number of geodesics connecting v_k and v_j and $g_{kj}(v_i)$ the number of such geodesics containing v_i.

These three centrality-measures emphasize different structural aspects and represent different points of view. They have in common that they all depend on n, the size of the graph. This problem, "the effect of network size", was examined by Freeman (1979) for the above defined centrality-measures. Using

$$C'_X(v_i) := \frac{C_X(v_i)}{C^*_X} \tag{1}$$

with $C^*_X = \max_{G \in G_n} \max_i C_X(v_i)$, where the index X can either be D, C or B referring to degree, closeness or betweenness centrality, removes this effect. [2] $C'_X(v_i)$ is restricted to

[1] For a detailed description and further literature see Freeman (1979).
[2] See, e.g. Freeman (1979) and Wasserman and Faust (1997).

the interval [0,1], and using C^*_X for normalization guarantees that at least one node in some $G \in G_n$ reaches 1. Thus, $C'_X(v_i)$ is an absolute measure in the sense that we can interpret the values with respect to the scale from 0 to 1. In the following we call centrality-measures normalized in this way *point-centrality* (Freeman, 1979).

Based on a point-centrality $C'_X(v_i)$ a *graph-centralization index* is given by

$$C_X = \frac{\sum_i (C'_X(v^*) - C'_X(v_i))}{\max_{G \in G_n}(\sum_i (C'_X(v^*) - C'_X(v_i))} \qquad (2)$$

where the nominator is the difference in point-centralities to the maximal value occurring in the actual graph, $C'_X(v^*) = \max_i C'_X(v_i)$, and the denominator gives the maximal difference in point-centralities with respect to all $G \in G_n$. Obviously, $C_X \in [0, 1]$.

The graph determining the maximal value in the denominator of C_X is the star, $S_{1,n-1}$ with $E = \{(v_1, v_i)|i = 2, \dots, n\}$, for X equal to D, C or B. The center of the star, v_1, is, moreover, one node that determines C^*_X in these cases.

Calculability of the denominator in Eqs. (1) and (2) is a necessary condition not only for normalizing centrality-measures to point-centrality and graph-centralization but for meaningful interpretation. This should be kept in mind when the above described concept is applied to arbitrary centrality measures. We will come back to this subject later.

To avoid misunderstandings we call our concept that is developed in the next section "node-centrality", to distinguish from the above derived "point-centrality".

3. Node-centrality

The question "what centrality is" seems to be difficult to answer in general, so we have to rely on intuition.

> "A person located in the center of a star is universally assumed to be structurally more central than any other person in any other position in any other network of similar size." (Freeman (1979), p. 218)

We agree with Freeman when he states further that *"this intuition seems to be natural enough"*. We call the node that represents the above described position in a graph the *center of the star* and while most researchers approach centrality in terms of the "properties" of this node, i.e. maximal degree, minimal distance and maximal betweenness, we use this intuition as defining condition.

Definition 1. Let $G = (V, E)$ be an undirected and connected graph with $|V| = n$. Let *nc* be a function which assigns a real value to every node of G. $nc(v_i)$ is called a *node-centrality* of node v_i if

(I) $nc(v_i) \in [0, 1]$ for every $v_i \in V$, and
(II) $nc(v_i) = 1$ if and only if $G = S_{1,n-1}$ and $i = 1$.

Nonnegative values and a range from 0 to 1 seems to be a pure technical condition but, as we have seen before, it is helpful for interpretation. The significance of a value depends on minimal and maximal attainable values. Note that if they are known for some function *nc*, condition (I) can be satisfied simply by an affine transformation of *nc*. Moreover, condition (I) corresponds exactly to point-centrality.

Condition (II) is the formal description of Freemans intuitive statement.[3] The center of a star should reach the maximal value, that is 1, and no other node should reach this value. The latter constitutes the difference to point-centrality.

We now compare our notion of "node-centrality" to the centrality concepts introduced in Section 2. Note first that degree- or distance-based centrality-measures do not differentiate between the center of the star and, for example, nodes in the complete graph K_n. Thus,

Result 1. Degree-centrality $C'_{\mathrm{D}}(v_i)$ and closeness-centrality $C'_{\mathrm{C}}(v_i)$ are not node-centralities.

In contrast, betweenness-centrality fulfills the second condition (Freeman, 1977), and thus,

Result 2. Betweenness-centrality $C'_{\mathrm{B}}(v_i)$ is a node-centrality.

Another measure of centrality that is often used for the analysis of interlocking directorates (see, e.g. Bonacich (1972a,b) and Mizruchi and Bunting (1981)) is eigenvector-centrality (called "rank prestige" by Wasserman and Faust (1997)). In the next section we first recall the definition and then develop conditions which assure that this centrality-measure is a node-centrality.

4. Eigenvector-centrality

Based on the idea that an actor is more central if it is in relation with actors that are themselves central, we can argue that the centrality of some node does not only depend on the *number* of its adjacent nodes, but also on their value of centrality. For example, Bonacich (1972b) defines the centrality $c(v_i)$ of a node v_i as positive multiple of the sum of adjacent centralities, i.e.

$$\lambda c(v_i) = \sum_{j=1}^{n} a_{ij} c(v_j) \quad \forall i.$$

In matrix notation with $c = (c(v_1), \dots, c(v_n))$ this yields

$$Ac = \lambda c. \tag{3}$$

This type of equation is well known and solved by the eigenvalues and eigenvectors of A.

From the set of different eigenvectors only one seems to be an appropriate solution that can serve as a centrality measure (see, e.g. Bonacich (1972b)). As A is the adjacency-matrix of an undirected (connected) graph, A is nonnegative and due to the theorem of Perron–Frobenius, e.g. Cvetković et al. (1995), there exists an eigenvector of the maximal eigenvalue with only nonnegative (positive) entries.

[3] Note, that "node-centrality" combines this intuitive approach with the formal approach of a "point-centrality".

We call a nonnegative eigenvector $c \geq 0$ of the maximal eigenvalue *principal eigenvector* and an entry $c(v_i)$ *eigenvector-centrality* of node v_i.

Note that the principal eigenvector, and thus also a single entry, is not uniquely determined. Every multiple of the principal eigenvector satisfies Eq. (3), i.e. for $c \geq 0$, principal eigenvector of A, αc with $\alpha > 0$ is a principal eigenvector of A, too. These vectors can be transformed into each other by multiplication, but nevertheless they differ in interpretation and properties. To show this we pose some simple questions: Which is the maximal value a single entry in c can reach? Which node v_i in which graph $G \in G_n$ does the maximal entry correspond to?

The common concept for vector-normalization is that of a p-norm:

$$||y||_p := \begin{cases} (|y_1|^p + |y_2|^p + \cdots + |y_n|^p)^{1/p}, & 1 \leq p < \infty \\ \max_{i=1,\dots,n} |y_i|, & p = \infty \end{cases}$$

for any $y = (y_1, \dots, y_n)$.

Maximal entries in the principal eigenvector of graphs in dependence of p are investigated in Papendieck and Recht (2000). We refer to them for the proof of the statements below.

In the following sections we examine the effect that different normalization have on the interpretation of eigenvector-centrality.

4.1. Maximum norm

To normalize c with respect to the maximum norm $p = \infty$, we divide every entry $c(v_i)$ by the maximal entry $c(v^*) := \max_j c(v_j)$ occurring in the centrality-vector c

$$c_{\mathrm{m}}(v_i) := \frac{c(v_i)}{c(v^*)}.$$

In Bolland (1988, p. 236) c_{m} is called *continuous flow*.

Properties of c_{m}:

- $c_{\mathrm{m}}(v_i) \in [0, 1]$ for all i.
- Obviously, $\max_i c_{\mathrm{m}}(v_i) = 1$ for every graph.

Using continuous flow as a measure of centrality, the maximal value that can be reached is 1 and it is reached by some node in any graph, i.e. in an arbitrary social network of arbitrary size there is always an actor with maximal central position. Thus, $c_{\mathrm{m}}(v_i)$ gives rather a kind of "relative centrality" within the graph than an absolute value with respect to what is possible. To decide which actor in which of some different networks is "more central" is not possible using continuous flow.

This does not correspond to node-centrality where the maximum has to be reached only by the center of a star. Using continuous flow as a point-centrality one should be aware of this property while comparing nodes from different graphs, except regarding their relative centrality in the graph.

Moreover, defining a graph-centralization index based on continuous flow is problematic. As stated before, calculability of the denominator in Eq. (2) in Section 2, i.e. maximal difference in point-centrality for $G \in G_n$, is a necessary condition. However, in contrast to the assertion in Bolland (1988, p. 237), it is not true that the maximal centralization in

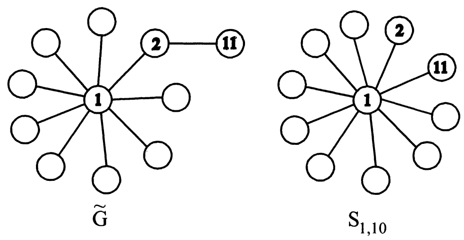

Fig. 1. Example graphs.

Eq. (2) for the continuous flow occurs in a star. That is the maximal value in the denominator is not

$$\max_{G \in G_n} \left(\sum_{i=1}^{n} (c_{\mathrm{m}}(v^*) - c_{\mathrm{m}}(v_i)) \right) = n - \sqrt{n-1} - 1.$$

As can be seen from the following counterexample there exist graphs reaching higher graph-centralization than the star if continuous flow is used as point-centrality.

Counterexample 2. Consider the graphs depicted in Fig. 1. For both graphs, $S_{1,n-1}$ and $\tilde{G}$, eigenvector-centrality is given in terms of the principal eigenvalue and the maximal entry $c(v_1)$ and $\tilde{c}(v_1)$, respectively. For the star $S_{1,n-1}$ we have

$$c = \left(c(v_1), \frac{c(v_1)}{\lambda}, \dots, \frac{c(v_1)}{\lambda} \right) \quad \text{with } \lambda = \sqrt{n-1}$$

and for $\tilde{G}$

$$\tilde{c} = \left(\tilde{c}(v_1), \frac{\mu \tilde{c}(v_1)}{\mu^2 - 1}, \frac{\tilde{c}(v_1)}{\mu}, \dots, \frac{\tilde{c}(v_1)}{\mu}, \frac{\tilde{c}(v_1)}{\mu^2 - 1} \right)$$

with principal eigenvalue

$$\mu = \sqrt{\frac{n-1}{2} + \sqrt{\left(\frac{n-1}{2}\right)^2 - (n-3)}}.$$

After normalization $c_{\mathrm{m}}(v_1) = \tilde{c}_{\mathrm{m}}(v_1) = 1$ and for $n = 11$ one easily calculates

$$\sum (c_{\mathrm{m}}(v^*) - c_{\mathrm{m}}(v_i)) = n - (n-1)^{0.5} - 1 = 6.83772$$

$$\begin{aligned}\sum (\tilde{c}_{\mathrm{m}}(v^*) - \tilde{c}_{\mathrm{m}}(v_i)) &= \left(1 - \frac{\mu}{\mu^2 - 1}\right) + 8\left(1 - \frac{1}{\mu}\right) + \left(1 - \frac{1}{\mu^2 - 1}\right) \\ &= 6.85645.\end{aligned}$$

For the creation of a graph-centralization it is necessary to determine the maximal value among all graphs. For the maximum norm, or continuous flow, this maximum is not known yet but it can be seen from the example that the maximum is not reached by the star.

Result 3. c_m can be used to range the nodes within a graph. c_m is not an appropriate measure for the comparison of nodes from different graphs or the evaluation with respect to a given standard. c_m is not a node-centrality.

4.2. Sum-norm

To normalize c with respect to $p = 1$, every entry $c(v_i)$ has to be divided by the sum of all values:

$$c_s(v_i) := \frac{c(v_i)}{\sum_{i=1}^{n} c(v_i)}.$$

$c_s(v_i)$ gives the proportion of centrality that v_i reaches in G. Again we are interested in the maximal value a single node can reach, and which node reaches this value. We have the following:

Properties of c_s:

- $c_s(v_i) \in \left[0, \frac{1}{(2\cos(\pi/(n+1)))^{-1}+1}\right]$ for all i
- $c_s(v_i) = \frac{1}{(2\cos(\pi/(n+1)))^{-1}+1} \Leftrightarrow G = K_2$.

Clearly, value 1 for $n > 1$ does not occur. But if we knew the maximal value reached for fixed n, we could multiply the vector by the inverse.

The upper bound for $c_s(v_i)$ is only reached by the nodes of K_2 (Papendieck and Recht, 2000).

For instances of n greater than 2 neither the value $\max_{G\in G_n} \max_{i=1,\dots,n} c_s(v_i)$ nor which node in which graph reaches this maximal value is a priori known. That is one has to enumerate the values for each G and each n. It is known, however, that the center of the star does not always attain the maximal value as can be seen from Fig. 1, where node 1 in $\tilde{G}$ reaches an entry greater than that of the center of the star.

What does that mean? By definition the interpretation of a node-centrality is based on the relation to the center of the star. Now, there is no such "general structure" as the star that determines the maximal value of c_s. Even worse, evaluation of some instances of n shows that the structure determining the maximal value changes with n. That is for $n = 4$ it is the center of the star, and for $n = 11$ it is definitely not the star, as calculation for $\tilde{G}$ in Fig. 1 shows.

Result 4. c_s can be used to range nodes according to their proportion of the centrality within a graph. Since the maximal value $\max_{G\in G_n} \max_i c_s(v_i)$ is in general not reached by the center of the star, c_s is not a node-centrality.

4.3. Euclidean-norm

Another often used normalization is the Euclidean norm, i.e. $p = 2$, that leads to

$$c_e(v_i) := \frac{c(v_i)}{(\sum_{i=1}^{n} c(v_i)^2)^{0.5}}.$$

Here the Euclidean length of the vector c_e is 1. This does not lead to an interpretation as easy and natural as for $p = 1$, but we have the following.

Properties of c_{e}:

- $c_{\mathrm{e}}(v_i) \in [0, \sqrt{1/2}]$ for all i
- $c_{\mathrm{e}}(v_i) = \sqrt{1/2} \Leftrightarrow G = S_{1,n-1}$ and $i = 1$ for all n.

These properties were already mentioned in Borgatti and Everett (1997). A proof can be found in Papendieck and Recht (2000).

The maximal value of $c_{\mathrm{e}}(v_i)$ is known and uniquely determined, it is $\sqrt{(1/2)}$ regardless of n. This maximum is reached only by the center of a star; in contrast to c_{m}, where the maximum is reached by some node in every graph, and c_{s}, where varying graphs determine the maximum.

Again, the value 1 cannot be reached for $n \geq 2$, but multiplying by the inverse of the maximal value $\sqrt{(1/2)}$, that is fix for all n, we reach

Result 5. For eigenvector-centrality c_{e} based on Euclidean normalization, the function

$$nc_{\mathrm{e}}(v_i) := \sqrt{2}\, c_{\mathrm{e}}(v_i)$$

is a node-centrality.

Now, after we have shown that only normalization based on the Euclidean norm leads to a node-centrality, the following question remains: How is it possible, that although the difference between normalizing with $p = 1$ and $p = 2$ is only to multiply with an appropriate factor, nodes from different graphs reach the maximal value?

The reason for this phenomenon is found in the relative values. To transform c_{s} into c_{e} we use

$$c_{\mathrm{e}}(v_i) = \frac{c_{\mathrm{s}}(v_i)}{||c_{\mathrm{s}}||_2}.$$

Let v_1 denote the center of the star and w_1 the center of $\tilde{G}$ in Fig. 1. With respect to $p = 1$, the center of the star reaches a value that is smaller than that of the center of $\tilde{G}$, $c_{\mathrm{s}}(v_1) < \tilde{c}_{\mathrm{s}}(w_1)$. But the Euclidean length of this vector $c_{\mathrm{s}} = (c_{\mathrm{s}}(v_1), \dots, c_{\mathrm{s}}(v_n))$ is smaller than that of $\tilde{c}_{\mathrm{s}} = (\tilde{c}_{\mathrm{s}}(w_1), \dots, \tilde{c}_{\mathrm{s}}(w_n))$, too, i.e. $||c_{\mathrm{s}}||_2 < ||\tilde{c}_{\mathrm{s}}||_2$ and this difference compensates the difference $c_{\mathrm{s}}(v_1) < \tilde{c}_{\mathrm{s}}(w_1)$.

For this reason it follows $c_{\mathrm{e}}(v_1) > \tilde{c}_{\mathrm{e}}(w_1)$, i.e. the relation between the eigenvector-centralities changes by normalization.

5. Conclusion

The concept of centrality is widely used in Social Network Analysis and has found different realizations regarding proper measures.

In this paper, the outstanding structural position of the center of a star was used as a defining property for "node-centrality": The maximal value of a centrality measure should be reached only by the center of the star. This concept reflects Freeman's intuitive definition of centrality, but it differs slightly from point-centrality.

We have discussed several centrality-concepts from the perspective of node-centrality. It turned out that betweenness-centrality is both a point- and a node-centrality, but degree and closeness based measures do not satisfy the properties of node-centrality. We have shown that eigenvector-centrality under Euclidean norm can be transformed into node-centrality, while other normalizations fail to satisfy the conditions. Moreover, they can lead to different "rankings" of nodes with respect to centrality, as we have seen from the example in Fig. 1.

We have seen that eigenvector-centrality with maximum norm, c_m, is only suitable for comparisons within a graph. For eigenvector-centrality with sum-norm, c_s, the top end of the scale is unknown for most n until now, thus complicating interpretation. The effect that normalization has on results and their interpretation should be noticed in choosing and applying eigenvector-centrality when investigating certain social networks.

Acknowledgements

I would like to thank the unknown reviewers for helpful comments.

References

Bavelas, A., 1950. Communication patterns in task-oriented groups. Journal of the Acoustical Society of America 22, 725–730.

Bolland, J.M., 1988. Sorting out centrality: an analysis of the performance of four centrality models in real and simulated networks. Social Networks 10, 233–253.

Bonacich, P., 1972a. Technique for analyzing overlapping memberships, In: Costner, H.L. (Ed.), Sociological Methodology, Jossey-Bass, San Francisco, pp. 176–185.

Bonacich, P., 1972b, Factoring and weighting approaches to status scores and clique identification. Journal of Mathematical Sociology 2, 113-120.

Borgatti, S.P., Everett, M.G., 1997. Network analysis of 2-mode data. Social Networks 19, 243–269.

Cvetković, D.M., Doob, M., Sachs, H.,1995. Spectra of Graphs, Barth, Heidelberg Leipzig, 3rd Edition.

Freeman, L.C., 1977. A set of measures of centrality based on betweenness. Sociometry 40, 35–41.

Freeman, L.C., 1979, Centrality in social networks: conceptual clarification. Social Networks 1, 215-239.

Leavitt, H.J., 1951. Some effects of certain communication patterns on group performance. Journal of Abnormal and Social Psychology 46, 38–50.

Mizruchi, M.S., Bunting, D., 1981. Influence in corporate networks: an examination of four measures. Administrative Science Quarterly 26, 475–489.

Papendieck, B., Recht, P., 2000. On maximal entries in the principal eigenvector of graphs. Linear Algebra and its Applications 310, 129–138.

Wasserman, S., Faust, K., 1997. Social network analysis methods and applications. Structural Analysis in the social Sciences, Vol. 8. Cambridge University Press, Cambridge (Corrected Reprint).