ON A MATRIX-BASED MEASURE OF CORENESS OF A NODE IN A NETWORK

V.M. Chelnokov, V.L. Zephyrova

Abstract

The paper focuses on the right and left eigenvectors of a network matrix that belong to the largest eigenvalue. It is shown that each of vector entries measures the walk centrality of the corresponding node's position in the network's link structure and of the positions of the node's adjacent nodes; as a result, it indicates to which degree the node can be associated with the structure's core - the structural coreness of the node. Both the entry-position relationship and the vectors' finding itself are based on an iterative computational scheme known as the power method. The paper studies the method's convergence for networks of different structure. Some possible applications are discussed. The paper includes also a numerical example dealing with a real network of 197 nodes and 780 links.

Introduction

In the paper we show that the right and left eigenvectors of a network matrix, corresponding to the largest eigenvalue, measure the degree of the *structural coreness* of the network nodes. Our coreness measure represented by the vectors' entries is a function on the network's node set. The nodes can be arranged in decreasing order of this function's value; as we show, such an ordering is also one of lowering the degree to which a node can be associated with the core of the network's link structure. Some nodes from the top of this ranking are taken as the network's core; besides, it's possible to compare any two nodes of the network with each other by this scale – an opportunity that's not less important than the core's revealing.

There are two types of (the degree of) coreness in any directed network: the ith entry of the right eigenvector measures the *out*-coreness of the position of the ith node in the link structure, while the ith entry of the left eigenvector gives the *in*-coreness of this node's position in the structure. The first value reflects the sum of weights of all walks starting at node i, while the second represents the total weight of the walks terminating there. The interconnection between walks and eigenvectors is explained by that the (i, j)th entry of the k-th power of the matrix of a network is equal to the total weight (number) of the network's k-step walks from node i to node j, on the one hand, and that these matrix powers are involved in the iterations of an eigenvector-computation scheme called power method, on the other. In an undirected network both types of coreness merge into one.

Network-node rankings by the *out*/*in*-coreness, under different names, are known in social networks, citation networks, hypertext networks, etc. The ranking by the *in*-coreness, in the case of a stochastic network matrix, turns out the final probability distribution for the Markov chain describing a random walk on the network. We give examples of using the coreness measure in models of text comprehension and nonclassical categorization.

We show that the convergence of the computational scheme has a solid basis. Well-known mathematical works on matrix analysis, upon which we base, do not include network-related interpretations relevant to our subject. On the other hand, the dispersed publications in different areas, in which indices computationally equivalent to our measure are described, do not inform readers on the subject's mathematical side and on that values they discuss represent the node coreness measure defined for networks of any nature. The purpose of our paper is to help to fill in this gap.

We implemented the paper's theoretical considerations as computer programs. They were used to obtain results in two numerical examples included in the paper; in one of them we deal with a real network containing 197 nodes and 780 links.

1 Notation

Consider a network with nodes labeled $1, 2, \ldots, n$ and links which, in general case, are directed. The network is represented by an $n \times n$ matrix of relationships A, in which the entry $A(i, j)$ is positive and equal to the weight of link (i, j), in case of the link's presence, and to zero otherwise $(i, j = 1, 2, \ldots, n)$.

When the network's links are all of the same weight, we have a standard adjacency matrix of a graph consisting of zeros and ones. Thus A is a real, non-negative, square and, in general case, nonsymmetric matrix.

As it's well known, for any real matrix A, the sets of eigenvalues (spectra) of A and of A', its transpose, are the same. If $\bar{x}$ and $\bar{y}$ are the column eigenvectors of A and A' with some one and the same eigenvalue λ, i.e. $A \cdot \bar{x} = \lambda \cdot \bar{x}$ and $A' \cdot \bar{y} = \lambda \cdot \bar{y}$, then $\bar{x}$ and $\bar{y}$ are called the *right* and *left* eigenvectors of A with eigenvalue λ, implying that $\bar{y}' \cdot A = \lambda \cdot \bar{y}'$, where $\bar{y}'$ is the row vector [1,2]. The eigenvalues of A and A' are numbered in order of nonincreasing absolute value: $|\lambda_1| \geqslant |\lambda_2| \geqslant |\lambda_3| \geqslant \ldots \geqslant |\lambda_n|$ (each λ is present here a number of times equal to its multiplicity); the eigenvectors, corresponding to λ_s, are denoted by $\bar{x}_s$ and $\bar{y}_s$ $(s = 1, 2, 3, \ldots, n)$; the real number $r(A) = |\lambda_1|$ is called the *spectral radius* of A (and A').

2 The power method

For any real square matrix, eigenvalue λ_1 and eigenvectors $\bar{x}_1$ and $\bar{y}_1$ can be found by means of the *power method*, one of the most popular in matrix computations [1, 2, 3, 4]. In our, "network", case, this method is not only a way to obtain the above results, but also an opportunity to see the relation of $x_1(i)$ and $y_1(i)$, the vectors' ith entries, to the position of node i within the network's link structure.

Suppose that the maximal eigenvalue λ_1 of matrix A has the algebraic multiplicity 1 and is unique, i.e. $|\lambda_1| > |\lambda_2|$; suppose also that it's real. Let $\bar{z}_0$ be an n-component vector not orthogonal to $\bar{x}_1$, and $\nu_k = \left\| A^k \bar{z}_0 \right\|$, where $\|.\|$ is a vector norm, e.g. Euclidean length. The method is based on that, under the given assumptions, vector $\bar{z}_k$ in the iterations

$$\bar{z}_k = \nu_k^{-1} A^k \bar{z}_0, k = 1, 2, \ldots, \tag{1}$$

tends to vector $\|\bar{x}_1\|^{-1} \bar{x}_1$ as $k \to \infty$, and the ratio $\nu_k / \nu_{k-1} \to \lambda_1$. In a form more convenient for computations,

$$\bar{z}_{k+1} = \alpha_k^{-1} A \bar{z}_k, k = 0, 1, 2, \ldots, \tag{2}$$

where $\alpha_k = \|A\bar{z}_k\| \to \lambda_1$ $(\alpha_0 \alpha_1 \ldots \alpha_{k-1} = \nu_k)$.

In practical computations, a small number ε, $0 < \varepsilon < 1$, can be specified in advance; then the iterations are performed, with computing

$e = 1 - \frac{1}{n}\sum_{i=1}^{n} \frac{z_k(i)}{z_{k-1}(i)}$ at each step, until $|e| > \varepsilon$. The more the ratio $\frac{|\lambda_1|}{|\lambda_2|}$, the less the required number of iterations [2].

The same computational scheme applied to matrix A' leads to vector $\bar{y}_1$.

The method's convergence is provided by the above assumptions concerning λ_1. Their plausibility will be discussed below. At this moment, we want only to note that in our case λ_1 is always positive, i.e. the spectral radius $r = \lambda_1$, and all entries of vectors $\bar{x}_1$ and $\bar{y}_1$ are positive ($\bar{x}_1 > 0$, $\bar{y}_1 > 0$). As a result, any non-negative, nonzero vector can be taken as the initial $\bar{z}_0$.

Note also that analogous computations with matrices, obtained by reducing the original matrix in a definite way, allow to find all next eigenvectors and eigenvalues (the method of deflation) [3].

3 The coreness measure

If some two nodes i and j occupy symmetric positions in the network's link structure, then $x_1(i) = x_1(j)$, $y_1(i) = y_1(j)$ (and the same is true for any other eigenvector). On the other hand, if entries relating to nodes i and j are different in some eigenvector, say, $x_1(i) \neq x_1(j)$, then i and j are dissimilar by their position in the structure. Formulas (1) and (2) help us to reveal two structural aspects of the $\bar{x}_1$'s and $\bar{y}_1$'s entries. In view of them, we treat each of the entries as expressing the degree to which the corresponding node can be associated with the structural core of the network – as measuring the node's *coreness*.

3.1 Coreness as centrality.

Let $A^{(k)}(i,q)$ denote the (i,q)th entry of matrix A^k, $k = 1, 2, \ldots$. This entry is equal to the total weight of different k-step walks that start at node i and terminate at node q (the length of each step is one link). This fact is well known in graph theory; it follows directly from the definition of matrix multiplication. Recall that a *walk* from i to q is a sequence of network links, in which the starting node of the first link is i, the starting node of each next link is the end node of the preceding link, and the end node of the last link is q; repetitions and loops are allowed. The *weight* of a k-step walk is defined as the product of the weights of the k links forming the walk sequence. When each link in the network is of weight 1, the entry indicates simply the number

of such walks.

Let $S_{out}^{(k)}(i)$ and $S_{in}^{(k)}(i)$ denote the total weights (numbers) of k-step walks from node i to all nodes of the network and from them to i, respectively: $S_{out}^{(k)}(i) = \sum_{q=1}^{n} A^{(k)}(i,q)$, $S_{in}^{(k)}(i) = \sum_{q=1}^{n} A^{(k)}(q,i)$; when $k = 1$, these values indicate $d_{out}(i)$ and $d_{in}(i)$, the out-degree and in-degree of node i, that is, weights (numbers) of outgoing and incoming links.

Theorem 1 *Let a network matrix A have the eigenvalue $\lambda_1 = r > |\lambda_2|$ and eigenvectors $\bar{x}_1 > 0$ and $\bar{y}_1 > 0$. Then, for any two nodes i, j of the network and any sufficiently large k: (a) if $x_1(i) > x_1(j)$, then $S_{out}^{(k)}(i) > S_{out}^{(k)}(j)$; (b) if $y_1(i) > y_1(j)$, then $S_{in}^{(k)}(i) > S_{in}^{(k)}(j)$.*

Proof. Consider the iterations according to formula (1), with the initial vector $\bar{z}_0$, whose entries are all equal to 1. We have $\nu_k z_k(i) = S_{out}^{(k)}(i)$ and $\lim_{k\to\infty} z_k(i) = \frac{x_1(i)}{\|\bar{x}_1\|}$, for $i = 1, 2, \ldots, n$. Therefore

$$\lim_{k\to\infty} \frac{S_{out}^{(k)}(i)}{S_{out}^{(k)}(j)} = \lim_{k\to\infty} \frac{z_k(i)}{z_k(j)} = \frac{x_1(i)}{x_1(j)},$$

whence the validity of (a) follows. Applying the iteration formula to matrix A', we have

$$\lim_{k\to\infty} \frac{S_{in}^{(k)}(i)}{S_{in}^{(k)}(j)} = \frac{y_1(i)}{y_1(j)},$$

whence it follows that (b) holds. This proves the theorem.

By this theorem, the greater the entry $x_1(i)$ the higher the rank of node i as the starting point of the network's walks, and the greater $y_1(i)$ the higher this node's rank as their terminating point. Ordering nodes by the values of $d_{out}(i)$ or $d_{in}(i)$ gives the ranks in the first approximation.

Thus one structural aspect of the vector entries is as follows: a node's x_1 and y_1 show themselves as values characterizing the structural centrality of the node, the importance of its role in the structure, the load it bears there. These are qualities commonly associated with the core of structure of any nature, material or mental.

3.2 Core as a compact zone.

Suppose that in a network the fraction of nodes with x_1 values greater than a given number is a rapidly decreasing function of this number, for example, a negative power of it (this power-law form of the function appears to be typical for large real networks). Then, to retain the basic relation $\lambda_1 x_1(i) = \sum_{q=1}^{n} A(i,q) x_1(q)$, $i = 1, 2, \ldots, n$, those nodes i, which are of high rank by x_1, should have $A(i,q) > 0$, i.e. be adjacent in the network, to those nodes q, which themselves are of high rank. Therefore these nodes form a dense, compact zone in the network; the corresponding subgraph is close to a complete graph to the extent to which specific structural conditions allow this. An analogous picture takes place for nodes of high rank by y_1.

Thus another structural aspect of the vector entries consists in that, for each vector, nodes of high rank by this vector are distinguished in the network by their compactness – a property which is also characteristic of core in structures of any nature.

Because of this and the preceding aspect of the value $x_1(i)$, we call it the ***out-coreness*** of node i; this value's denoted by $c_{out}(i)$. The value $y_1 = c_{in}(i)$ is called the ***in-coreness***. In case of an undirected network, the matrix is symmetric ($A' = A$), the right and left eigenvectors coincide with each other, and the coreness is denoted simply by $c(i)$.

Note that the second aspect isn't an attribute of other measures of node centrality, such as ordinary node degree and average or maximal distance between a specific node and the rest of nodes: there is no "interest" in mutual adjacency among nodes of high degree or of small distance.

3.3 Some applications.

The coreness measure has often appeared under other names. Its first applications are probably those related to networks of the kind of sociograms and tournaments ("the leader problem") [5, 6]. The value $c_{out}(i)$ was called there "the real strength" of player i. The measure c is known under the name "coreness" in social networks [7].

A significant role in applications belongs to the case of *stochastic* network matrix, when $\sum_{j=1}^{n} A(i,j) = 1$, $i = 1, 2, \ldots, n$. The matrix defines here a finite *Markov chain* with transition probabilities $A(i,j)$, for example, a

random walk on the network performed by a particle, which is at a node at each discrete moment k and moves to an adjacent node "between" moments k and $k+1$ ($k = 0, 1, 2, \ldots$). Consider iterations (2) with matrix A' and a nonnegative $\bar{z}_0$, such that $\sum_{i=1}^{n} z_0(i) = 1$ (the initial probability distribution). Let α be a norm defined for a nonnegative vector as the sum of its entries. Thus, for $k = 0, 1, 2, ...$, $\alpha_k = 1$ and (2) takes the form $z_{k+1}(i) = \sum_{j=1}^{n} A'(i,j)\, z_k(j) = \sum_{j=1}^{n} z_k(j)\, A(j,i)$, that is, $z_{k+1}(i)$ is equal to the absolute (unconditional) probability of that the particle is at node i at moment $k+1$. Under the conditions of convergence of (1) and (2), $\lim_{k\to\infty} \bar{z}_k = \bar{c}_{in} > 0$, that is, vector $\bar{c}_{in}$ gives the *final (stationary) probability distribution* for the Markov chain determined by matrix A; like other results of (2), this distribution is independent of $\bar{z}_0$ – in this case, the initial distribution. Thus the *in*-core nodes form the most probable place where the particle can be found after a sufficiently large time, in a stationary regime (c_{out} values are all equal to each other in this case).

The random-walk model proves effective when applying to the behavior of a user traveling by links in a hypertext network. It underlies the Web-page ranking algorithm used by popular search engine Google [8, 9]. Nodes (Web pages) of high c_{in} value act as 'authorities' to which the network eventually gets a user most often [8].

Node rankings by c are used in cognitive psychology. According to a text comprehension model, a text is represented in the reader's memory as a network of clauses (text units of meaning), interconnected by semantic links. Nodes with high c values represent *macrounits* of the text, i.e. units bearing the heaviest semantic load in it – forming its summary. Experiments confirm this, for they show that units with high c values are recalled faster than units with lower c values [10].

One more example relates to cognitive linguistics. *Nonclassical categories* (groupings of similar things), studied in this discipline, are distinct from classical categories in that they have core-periphery structure, caused by category items' varying degrees of membership, or of representativity, or of prototypicality. Items of a category can be weighted, for example, in the following way. Each item is characterized with a set of attributes taken from a common attribute list, which results in an item-by-attribute bipartite graph. Each attribute receives a weight equal to the number of items linked up

to it. Finally, a weight is calculated for each item by summing the weights of attributes associated with the item [11]. Clearly, this weighting scheme represents the first two iteration cycles of the coreness computing. An improvement would consist in the replacement of these weights with values of c.

4 The computational scheme convergence

The interpretation of vectors $\bar{x}_1$ and $\bar{y}_1$ as coreness measures is based on the supposition that iterations (1) and (2) converge, i.e. that the above-mentioned properties of the eigenvalue λ_1 take place, on the supposed positivity of vectors $\bar{x}_1$ and $\bar{y}_1$, and on some aspects of these vectors and their approximations, resulting from formulas (1) and (2). That the suppositions are justified follows from Perron and Frobenius classical results for non-negative matrices [1, 6, 12].

4.1 Perron-Frobenius theorem.

As it was established by Perron, if a real matrix A is positive, i.e. all entries in it are positive, then: (a) r, the spectral radius of A, is positive; (b) r is an eigenvalue of A, i.e. $r = \lambda_1$; (c) this eigenvalue belongs to a positive eigenvector; (d) this eigenvalue has algebraic multiplicity 1; whence the eigenvector is unique (up to scalar multiplication); and (e) $|\lambda| < r$, for every eigenvalue λ of A, such that $\lambda \neq r$.

Taken together, (a), (b), (d) and (e) mean that $r = \lambda_1 > |\lambda_2|$.

This theorem was extended to non-negative *irreducible* matrices – those whose networks are strongly connected (simply connected, if links are undirected). For such a matrix, as Frobenius proved, parts (a)-(d) of Perron's theorem hold; the truth of (e), however, is not guaranteed ([1], theorem 8.4.4).

Note that if matrix A is irreducible, then the transpose, A', is irreducible, too, since the only difference between their networks consists in that one of them has links going in the opposite direction by comparison with the other.

The eigenvalue r of an irreducible matrix A is called the *Perron root* [1] or the *dominant eigenvalue* [2] of matrix A (and A').

Our network matrices are merely non-negative, and it's very important for us to retain part (e), for if a network matrix has an eigenvalue λ, such that $\lambda \neq r$ and $|\lambda| = r$, then iterations (1) and (2) don't converge.

We focus therefore on a matrix class introduced by Frobenius: a non-negative, irreducible matrix is called *primitive*, if it has exactly one eigenvalue equal in modulus to r [1, 12].

4.2 The case of an undirected, connected network.

In this case, we have a symmetric, irreducible network matrix. Each network is either unipartite or bipartite; a bipartite network is characterized by that its node set can be divided into two parts in such a way, that each link connects nodes from different parts (a "multipartite" network remains bipartite). Bipartite networks are often used as data representation schemes.

Theorem 2 *The matrix of an undirected, connected network is not primitive, if and only if the network is bipartite.*

PROOF. A network matrix A is primitive, if and only if $A^k > 0$ for any, sufficiently large integer k ([1], theorem 8.5.2). Take a bipartite network. Let $1, 2, \ldots, np$ be nodes of some one of its parts ("p-nodes"), and $np+1$, $np+2$, $\ldots$, $np+nt = n$ be nodes of the other part ("t-nodes"). Then the structure of the network's matrix A is such that matrix A^{2k+1}, $k = 0, 1, 2, \ldots$, has positive entries only in the $np \times nt$ submatrix in the upper right corner and in the $nt \times np$ submatrix in the lower left corner, while matrix A^{2k}, $k = 1, 2, \ldots$, has them only in the square submatrix of order np in the upper left corner and in the square submatrix of order nt in the lower right corner. Thus all powers of A contain zero submatrices, whence A is not primitive.

Now suppose that a network matrix A is not primitive. Let M denote the adjacency matrix of the *network's graph*, which differs from the network associated with A only in that each link in it is of weight 1. Clearly, walks in the graph are the same as in the network, whence, for any given k, $A^k > 0$, if and only if $M^k > 0$, that is, either A and M both are primitive, or they both aren't primitive. Therefore M is not primitive, that is, in addition to eigenvalue $r(M)$, it has an eigenvalue not equal to $r(M)$, but equal to it in modulus. The only possibility for this is $-r(M)$, since M is symmetric and hence all eigenvalues of M are real. But that the adjacency matrix of a connected graph has an eigenvalue pair of this kind is equivalent to that the graph is bipartite ([6], theorem 3.4). Clearly, the "partiteness" of the graph is the same as of the network. Therefore the network is bipartite.

The theorem's proof is completed.

Note that a nonsymmetric, irreducible matrix isn't necessarily primitive, even if its network isn't bipartite. Take, for example, the adjacency matrix of a 3-node directed cycle. All three eigenvalues of it are of modulus 1 (two of them are complex numbers forming a conjugate pair).

Note also that, in the theorem's proof, the square submatrices in the corners of matrix A^{2k}, $k = 1, 2, \ldots$, are of the form P^k and T^k, where P and T are primitive matrices. The reasoning behind their primitiveness is as follows. Take matrix P. It's primitive, because the network it determines on the set of p-nodes is connected and non-bipartite (unipartite). The first follows from that the whole bipartite network is connected. The second is equivalent to that the graph of this "p-network" is unipartite, and that is so, because none of the eigenvalues of the graph's matrix can be negative (for they are squares of the eigenvalues of the matrix of the whole network's graph). When taking matrix T and its "t-network", we obtain the same situation. The eigenvalues of P and T are equal to the squares of the eigenvalues of A, and their eigenvectors' entries form a bipartite division of entries of each eigenvector of A.

4.3 The addition of loops.

Any non-negative, irreducible matrix, in which all entries on the main diagonal are positive, is primitive ([1], lemma 8.5.5). This fact is useful for practical computations. Thanks to it, when having a strongly connected network (simply connected, if links are undirected), we can be sure of the iteration convergence, using, for example, matrix $A + I$ instead of A, where I is a square matrix of order n, in which $I(i,i) = 1$ for $i = 1, 2, \ldots, n$, and $I(i,j) = 0$ for $i \neq j$. This would result in the increasing of each eigenvalue by 1 and remaining the eigenvectors unchanged. Formula (2) then takes the form:

$$z_{k+1}(i) = \alpha_k^{-1}\left[z_k(i) + \sum_{q=1}^{n} A(i,q)\, z_k(q)\right], \quad \alpha_k = \|\bar{z}_k + A\bar{z}_k\| \to r + 1. \quad (3)$$

The transition from A to $A + I$ is equivalent to the supplying each node with a loop (a link to itself), which makes the network certainly non-bipartite.

5 Two numerical examples

E x a m p l e 1. We took the network of cross-references of Linear Algebra Glossary from Google's Directory. The glossary's terms were taken as nodes, their complete number was 201. The cross-references were taken as links, and we made them, for simplicity, undirected. They all were of weight 1. The computer extracted a graph, forming the network's maximal connected component, and that's the graph, for which main results were obtained. It had the number of nodes $n = 197$, the number of links $m = 780$ and the average node degree $2m/n = 7.9$; its maximal and minimal node degrees were 54 and 1, the maximal distance between two nodes (diameter) 6, and the average distance between a node pair 2.9.

The computer program, used for obtaining the coreness vector $\bar{c}$ and the dominant eigenvalue $r = \lambda_1$, was based on formula (3). With $\varepsilon = 10^{-5}$ (see Section 2), it took 16 iterations to obtain the results. It was found that $r = 14.3196$ ($\lambda_2 = 8.43$). Table 1 gives the ranking of nodes by coreness value ($d\,(i)$ is the degree of node i).

Table 1:

Row no.	i	$c(i)/c_{max}$	$d(i)$	Term
1	46	1	54	*Eigenvalues*
2	185	0.76	40	*Symmetrix Matrix*
3	84	0.64	31	*Inverse Matrix*
4	47	0.56	26	*Eigenvectors*
5	146	0.55	28	*Positive Definite Matrix*
6	134	0.54	24	*Orthogonal Matrix*
7	190	0.52	20	*Transpose*
8	168	0.49	22	*Singular Matrix*
...	...	...	...	.....
197	3	0.003	1	*Adjoint Matrix*

It's noteworthy that the coreness measure demonstrates on this network the highest level of discriminating power: there are no equals among all the 197 values of c (while the node degree has only 26 distinct values)[1] .

[1]The complete data can be obtained from the authors.

Suppose that nodes with c values not less than 50% of the maximum are taken as the network's core. Then the subgraph at the core has the form shown in Fig.1.

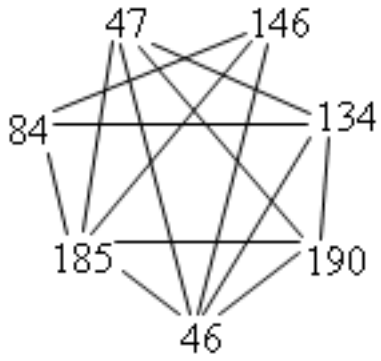

Figure 1: The core structure of the given 197-node graph – a connected, compact 7-node subgraph with 14 interlinks (there can be, at most, 21 links in a 7-node graph)

Figures 2 and 3 show, in log-log coordinates, the cumulative distributions of the node coreness and the node degree in the given 197-node graph. The two plots are similar to each other. Each has an approximately linear tail making up more than the curve's half – a part, that is well fit by a power law $J(x) \sim x^{-a}$, where $a = const > 1$. As a whole, the curves are well approximated by a function of the form $J(x) \sim (b+x)^{-a}$, where $b = const > 0$. Each curve reflects the presence of a small group of nodes being in sharp contrast to the rest of the graph's nodes. For example, a node subset of size more than 55% (99/197) of the graph's total number of nodes have coreness values not greater than 11% of the maximum, whereas a group of nodes, making up only 3.5% (7/197) of the total number, has values above 50% of the maximum. Similarly, a 3.5%-group of nodes has degrees greater than 44% (24/54) of the maximum, whereas a subset, containing 53% (105/197) of all nodes, has degrees below 11% (6/54) of the maximum.

E x a m p l e 2. In the first example, the core subgraph was characterized by an increased link density (closeness to a complete graph), in comparison

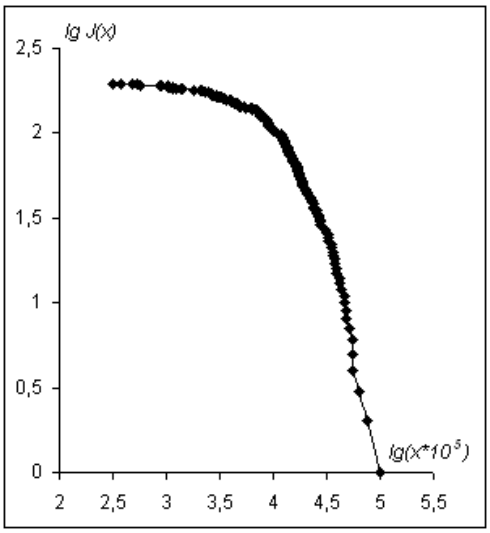

Figure 2: $J(x)$ is the number of nodes of coreness value c, such that $c/c_{max} \geqslant x$

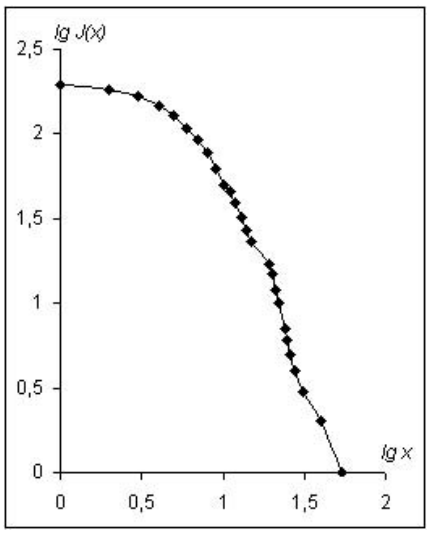

Figure 3: $J(x)$ is the number of nodes of degree $d \geqslant x$

to other parts of the graph. Such a case is frequent, but not necessary. It's possible that a network doesn't include any compact zone associated with its core, but nevertheless the core nodes are clearly indicated by the network's vector $\bar{c}$. They then exceed the rest of the network's nodes by their walk centrality (see Section 3.1). Note also that, in any case, the core subgraph "tends" to be connected (see Section 3.2).

As an illustration, take a rectangular lattice of size 10×7 – an undirected graph with $n = 70$ nodes and $m = 123$ links. Clearly, in this case, one cannot speak of any zone distinguished by its link density. As to coreness measure c, it assumes here 20 different values; equal c values correspond to symmetrical nodes. In Fig.4, the lattice nodes are labeled with ranks, from 1 to 20, indicating their positions in ordering by c value (nodes, having one and the same value, occupy the same position). The figure includes also a plot of the cumulative distribution of c.

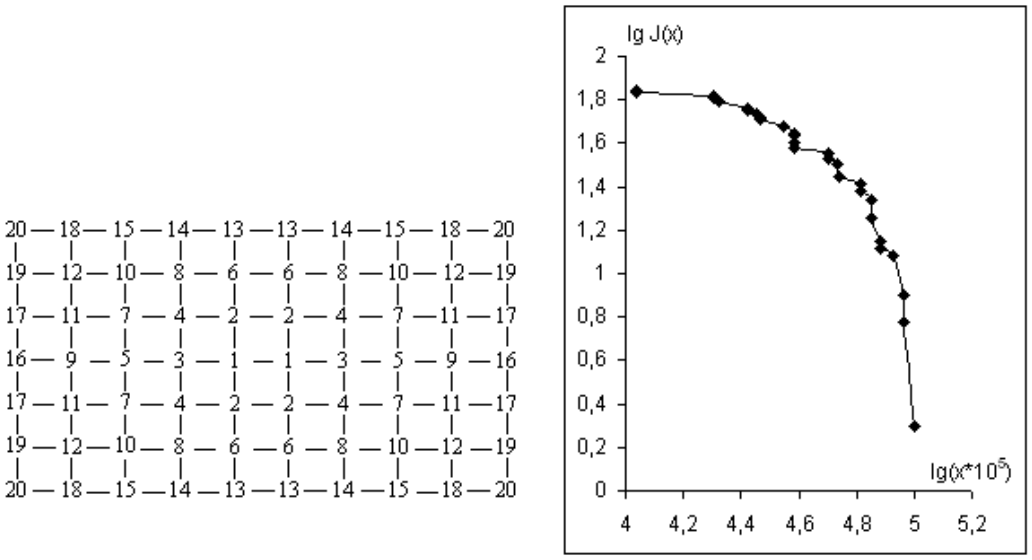

Figure 4: A rectangular lattice with nodes labeled according to their ranking by the coreness value. $J(x)$ is the number of nodes of coreness value c, such that $c/c_{max} \geqslant x$

Conclusion

The value, assumed by the coreness measure at a node, characterizes the node's position within the network's link structure. Nodes ranked highest by this measure, i.e. forming the core, are neighboring in the structure and make up a central part of it. Being adjacent or close to each other, the core nodes tend to form a dense nucleus, approaching to a complete subgraph as closely as it's possible with the specific link structure. The sparse part of a network is marked with decreased values of the measure and thus makes up the network's periphery. The core nodes are considered to form a central zone, for they surpass all other nodes of the network in the total weight of walks starting/terminating at them.

Walk-like pathways correspond to the kind of traveling characterized by a tendency to explore all associative possibilities, and not subordinated to the purpose of reaching a specific place by a route optimal in a formal sense. This kind of traveling happens, for example, when a user follows links of a hypertext network, in particular, WWW, searching for information on a

theme outlined only generally. The search engine *Google* ranks the Web nodes (pages), basing on the final probability distribution for a random walk, or Markov chain, on the Web. The final probabilities for a random walk coincide with the *in*-coreness values for the network, represented by the stochastic matrix of transition probabilities, determining this random walk.

In a network of information nodes and semantic links, the coreness value at a node correlates with the significance of the node as a macrounit of the semantic content of the network. In this case, coreness values, in combination with a depth-first search of the network graph, allow to compose, from the network nodes, coherent summaries on multiple themes [13, 14], and this may be used in an automatic summarization system.

Walks are also suitable to represent things different from chains of mental associations. For example, walks are plausible pathways for news, rumor and contagious disease propagation in social-contact networks [15]. Powerful computer models for simulating epidemics in communities of size up to million have been constructed [16, 17]. In these network models, a central role is attributed to hubs, nodes with extraordinary high degrees (numbers of links). It would be interesting to consider, instead of hubs, nodes ranked highest by the coreness value: when the degree distribution in a network follows a power law, these nodes most likely coincide with hubs, but cases are numerous when a network doesn't contain hubs at all, while these core nodes are well discernible (a simple example is rectangular lattice).

References

[1] Horn R.A. and Johnson C.R. Matrix analysis. Cambridge University Press, 1986.

[2] Golub G. and Van Loan C.F. Matrix computations. Baltimore: The John Hopkins University Press, 1989.

[3] Faddeyev D.K. and Faddeyeva V.N. Computational methods of linear algebra. Moscow: Fizmatgiz Press, 1963 (in Russian).

[4] Gelfand I.M. Lectures in linear algebra. Moscow: Nauka Press, 1966 (in Russian).

[5] Berge C. The theory of graphs and its applications. N.Y.: Wiley and Sons, 1962.

[6] Cvetković D.M., Doob M. and Sachs H. Spectra of graphs. N.Y.: Academic Press, 1980.

[7] Borgatti S.P. and Everett M.G. Models of core/periphery structures // Social Networks. 2000. V.21, No.4, P. 375-395.

[8] Kleinberg J.M. Authoritative sources in a hyperlinked environment // J. of the ACM. 1999. V.46, No.5, P. 604-632.

[9] Brin S. and Page L. The anatomy of a large-scale hypertextual Web-search engine // Proceedings of the 7th WWW Conference. Brisbane, Australia, 1998. P. 107-117.

[10] Kintsch W. and Welsch D.M. The construction-integration model: A framework for studying memory for text // Relating theory and data: Essays on human memory (W.Hockley, S.Lewandowsky, eds.). Hillsdale, NJ: Erlbaum, 1991. P. 367-386.

[11] Storms G., De Boeck P. and Ruts W. Prototype and exemplar-based information in natural language categories // J. of Memory and Language. 2000, V.42, No.1, P. 51-73.

[12] Voevodin V.V. and Kuznetzov Yu.A. Matrices and computations. Moscow: Nauka Press, 1984 (in Russian).

[13] Chelnokov V.M. and Zephyrova V.L. Hypertext macrodynamics // Lecture Notes in Computer Science. V. 1015. Berlin: Springer-Verlag, 1995. P. 105-120.

[14] Chelnokov V.M. and Zephyrova V.L. Collective phenomena in hypertext networks // Proceedings of the Hypertext'97 Conference (Southampton, UK). ACM, 1997. P. 220-221.

[15] Borgatti S.P. Centrality and AIDS // Connections. 1995. V.18, No.1, P. 112-115.

[16] Eubank S., Guclu H., Anil Kumar V.S., Marathe M.V., Srinivasan A., Toroczkai Z. and Wang N. Modelling disease outbreakes in realistic urban social networks // Nature. 13 May 2004. V.429, No.6988, P. 180-184.

[17] Barrett C.L., Eubank S.G. and Smith J.P. If smallpox strikes Portland... // Scientific American. 2005. V.292, No.3, P. 42-49.