

Boston University**OpenBU****<http://open.bu.edu>**

Physics

BU Open Access Articles

2016-01-01

The H-index of a network node and its relation to degree and coreness

This work was made openly accessible by BU Faculty. Please [share](#) how this access benefits you. Your story matters.

Version	Published version
Citation (published version):	Linyuan Lue, Tao Zhou, Qian-Ming Zhang, H Eugene Stanley. 2016. "The H-index of a network node and its relation to degree and coreness." NATURE COMMUNICATIONS, Volume 7, 7 pp. https://doi.org/10.1038/ncomms10168

<https://hdl.handle.net/2144/39935>

Boston University

ARTICLE

Received 3 Mar 2015 | Accepted 10 Nov 2015 | Published 12 Jan 2016

DOI: 10.1038/ncomms10168

OPEN

The H-index of a network node and its relation to degree and coreness

Linyuan Lü¹, Tao Zhou^{2,3}, Qian-Ming Zhang^{2,4} & H. Eugene Stanley^{1,4}

Identifying influential nodes in dynamical processes is crucial in understanding network structure and function. Degree, H-index and coreness are widely used metrics, but previously treated as unrelated. Here we show their relation by constructing an operator \mathcal{H} , in terms of which degree, H-index and coreness are the initial, intermediate and steady states of the sequences, respectively. We obtain a family of H-indices that can be used to measure a node's importance. We also prove that the convergence to coreness can be guaranteed even under an asynchronous updating process, allowing a decentralized local method of calculating a node's coreness in large-scale evolving networks. Numerical analyses of the susceptible-infected-removed spreading dynamics on disparate real networks suggest that the H-index is a good tradeoff that in many cases can better quantify node influence than either degree or coreness.

¹Alibaba Research Center for Complexity Sciences, Alibaba Business College, Hangzhou Normal University, Hangzhou 311121, China. ²Complex Lab, Web Sciences Center, University of Electronic Science and Technology of China, Chengdu 611731, China. ³Big Data Research Center, University of Electronic Science and Technology of China, Chengdu 611731, China. ⁴Department of Physics and Center for Polymer Studies, Boston University, Boston, Massachusetts 02215, USA. Correspondence and requests for materials should be addressed to L.L. (email: linyuan.lv@hznu.edu.cn) or to H.E.S. (email: hes@bu.edu).

The focus of network science research has been shifting from discovering macroscopic statistical regularities^{1–4} to uncovering the role played by such microscopic elements as nodes, links and motifs in the structure and dynamics of the system^{5–10}. Being able to effectively and efficiently identify the critical nodes associated with the specific dynamics of large-scale networks¹¹ will allow us to better control the outbreak of epidemics¹², conduct successful advertisements for e-commercial products¹³, prevent catastrophic outages in power grids or the Internet¹⁴, optimize the use of limited resources to facilitate information propagation¹⁵, discover drug target candidates and essential proteins¹⁶, and design strategies for communication breakdowns in human and telecommunication networks¹⁷.

The simplest way to measure the importance of a node is to determine its degree, that is, to count the number of its linked neighbours. Previous studies have shown that protecting, immunizing and regulating large-degree nodes can maintain network connectivity¹⁸, halt infectious disease propagation¹², enhance synchronizability¹⁹, improve transport capacity²⁰ and promote cooperation in evolutionary games²¹.

Recently, Kitsak *et al.*²² argued that the location of a node is more significant than the number of its linked neighbours, and they suggested that coreness is a better indicator of a node's influence on spreading dynamics than degree. The coreness of a node is measured by k -core decomposition²³, and a larger coreness value indicates that the node is more centrally located in the network. The k -core decomposition process is initiated by removing all nodes with degree $k = 1$. This causes new nodes with degree $k \leq 1$ to appear. These are also removed and the process is continued until the only nodes remaining are those of degree $k > 1$. The removed nodes and their associated links form the 1-shell. This pruning process is repeated for the nodes of degree $k = 2$ to extract the 2-shell, that is, in each stage the nodes with degree $k \leq 2$ are removed. The process is continued until all higher-layer shells have been identified and all network nodes have been removed. Then each node i is assigned a shell layer c_i , called the coreness of node i . Recent studies suggest that coreness is a good measure of a node's influence^{22,24}.

Calculating coreness requires global topological information, and its implementation is usually centralized, which can hinder its application to very large-scale dynamical networks. In contrast, degree is a simple local index, but of lower utility. As a tradeoff, Chen *et al.*^{25,26} proposed indices using the local neighbourhood information of individual nodes, which perform well but lack an underlying mathematical structure. This approach brings to mind the Hirsch index (also called the H-index)²⁷, which was originally used to measure the citation impact of a scholar or a journal^{27–29}. For a scholar or journal, the H-index is defined as the maximum value h such that there exists at least h papers, each with citation count $\geq h$.

Here we discuss the extension of the H-index concept to quantify how important a node is to its network³⁰. The H-index of a node is defined to be the maximum value h such that there exists at least h neighbours of degree no less than h . In the Supplementary Fig. 1, we compare calculating the H-index of a scholar and the H-index of a node. Degree, H-index and coreness seem to be independent but are actually interrelated. We construct an operator \mathcal{H} on a group of reals that returns a node's H-index when acting on its neighbours' degrees. By sequentially and synchronously applying the \mathcal{H} operator to each node, the returned value soon converges to coreness, that is, in terms of this operator, degree, H-index and coreness are its initial state, intermediate state and steady state, and all other intermediate states can also be treated as centrality measures. We further show that the convergence to coreness can be guaranteed even under an asynchronous updating process, thus

allowing a distributed computing algorithm to deal with large-scale dynamical networks. To see whether these centralities can measure a node's influence, we apply the standard susceptible-infected-removed (SIR) spreading model³¹ on eight real networks from disparate fields and calculate the correlation between each node's influence and its centrality values. Simulation results show that the H-index outperforms both degree and coreness in several cases, and thus can be considered a good tradeoff between degree and coreness.

Results

Mathematical relationship. We construct an operator, \mathcal{H} , which acts on a finite number of reals (x_1, x_2, \dots, x_n) and returns an integer $y = \mathcal{H}(x_1, x_2, \dots, x_n) > 0$, where y is the maximum integer such that there exist at least y elements in (x_1, x_2, \dots, x_n) , each of which is no less than y . For a scholar with n publications, x_1, x_2, \dots, x_n is the number of citations to these publications and $\mathcal{H}(x_1, x_2, \dots, x_n)$ is the scholar's H-index.

Denote $G(V, E)$ an undirected simple network, where V is the set of nodes and E is the set of links. The degree of an arbitrary node i is denoted by k_i and its neighbours' degrees are $k_{j_1}, k_{j_2}, \dots, k_{j_{k_i}}$. Then, we define the H-index of node i

$$h_i = \mathcal{H}(k_{j_1}, k_{j_2}, \dots, k_{j_{k_i}}). \quad (1)$$

We define $h_i^{(0)} = k_i$ to be the zero-order H-index of node i , and define the n -order H-index ($n > 0$) iteratively as

$$h_i^{(n)} = \mathcal{H}(h_{j_1}^{(n-1)}, h_{j_2}^{(n-1)}, \dots, h_{j_{k_i}}^{(n-1)}). \quad (2)$$

The H-index of node i is equal to the first-order H-index, namely $h_i^{(1)} = h_i$. A more detailed illustration can be found in Supplementary Note 1.

Theorem 1: for every node $i \in V$ of an undirected simple network $G(V, E)$, its H-index sequence $h_i^{(0)}, h_i^{(1)}, h_i^{(2)}, \dots$ will converge to the coreness of node i ,

$$c_i = \lim_{n \rightarrow \infty} h_i^{(n)}. \quad (3)$$

The proof is given in the Methods section. We give an example of iterative process from degree to coreness in the Supplementary Fig. 2.

This theorem shows that the degree, H-index and coreness are respectively the initial, intermediate and steady states under successive operations by \mathcal{H} . Given a network $G(V, E)$, the convergence time n_∞ is defined as the minimum number of iterations required to reach coreness from degree using the operator \mathcal{H} , that is, n_∞ is the minimum integer such that $h_i^{(n_\infty)} = h_i^\infty = c_i$ for $\forall i \in V$.

Using equation (2) and Theorem 1, we implement the synchronous updating in eight representative real networks drawn from disparate fields, including two social networks (Sex and Facebook), two collaboration networks (Jazz and NS), one communication network (Email), one information network (PB), one transportation network (USAir) and one technological network (Router). In brief, Jazz³² is a collaboration network of jazz musicians and consists of 198 nodes and 2,742 interactions, NS³³ is a co-authorship network of scientists working on network science, our Email³⁴ network is of e-mail interchanges between members of the Rovira i Virgili University (Tarragona), our Sex³⁵ network is of a bipartite sexual activity web community in which nodes are female (sex sellers) and male (sex buyers) and links between them are established when males write posts indicating sexual encounters with females, Facebook³⁶ is a sample of the friendship network of Facebook users, PB³⁷ is a network of US political blogs (the original links are directed, but here we treat them as undirected), USAir³⁸ is the US air transportation

network and Router³⁹ is a symmetrized snapshot of the structure of the Internet at the level of autonomous systems. Their topological features are shown in Table 1. Experiments show that the sequence of H-indices quickly converges to the coreness (n_∞ in Table 1). In addition to the degree, H-index and coreness, all other intermediate states $h^{(2)}, h^{(3)}, \dots$ can also be considered as centrality measures.

The resolution rate of the $h^{(n)}$ -index is the probability that two randomly chosen nodes will have different $h^{(n)}$. It is also a useful index for measuring the degree to which a network is coarse grained. Note that degree is the most distinguishable index and coreness is the least distinguishable, and that the resolution rate decreases as the index order increases (Supplementary Fig. 3; Supplementary Note 2). On the other hand, the calculation of degree requires less information, while coreness requires the most information. For a given node i , the information required to calculate $h_i^{(n)}$ can be measured by information coverage, which is defined as the ratio of the number of nodes with a distance no more than n from i to the network size. The coverage rate increases as the index order increases (Supplementary Fig. 4). Surprisingly, we find that in some networks, such as NS and Router, the coverage rate is <1 even for $h^{(n_\infty)}$, indicating that only partial information is required when calculating the coreness of a node in these networks (Supplementary Note 3).

Figure 1a shows the H-indices in different orders for a typical network Router. From left to right, we see the coarse graining process from degree to coreness. Figure 1b shows the probability distribution $p(h)$, defined as the probability that a randomly selected node's $h^{(n)}$ value is equal to h for the cases $n=0$, $n=1$ and $n=n_\infty$ of the network Router. Note that as the order n increases from 0 to 6 ($n_\infty=6$ for Router), the distribution narrows (see Supplementary Figs. 5–12 for the distributions of all H-family indices for the eight networks under study). Nevertheless, the distribution of values of high-order H-indices is still relatively broad, suggesting its potential as a good indicator of a node's importance²².

To show the different roles of different H-family indices, we iteratively construct a hierarchical tree (see ref. 40 for a similar method). As shown in Fig. 1c, the initialized network is of two levels, isomorphic to a star with L leaves. Here we set $L=4$. In each step, every leaf node becomes a star with L leaves, and the central node is connected with its original neighbours. After each step, the number of levels is increased by one, and the nodes in the more central positions (that is, of smaller levels) have a higher influence that is not reflected by their degree if the number of levels is high. Note that the three trees in Fig. 1c–e have 2, 3 and 4 levels, respectively, and in a hierarchical tree of D levels we need the index $h^{(0)}, h^{(1)}, \dots, h^{(D-2)}$ to quantify node influence. Such example clearly shows that a few low-order

H-indices are not always enough to distinguish different nodes' influences.

Asynchronous updating. The updating driven by \mathcal{H} uses only local information, and it rapidly converges to coreness. However, the updating from $h^{(n-1)}$ to $h^{(n)}$ is implemented synchronously according to equation (2), and thus in principle requires a centralized controller to set up a global clock that records the order n . In particular, if the target network is evolving, the addition of a single link will require the recalculation of the entire sequence of H-indices. This limits the application of H-indices to large-scale dynamical networks. Fortunately, the asynchronous updating can still guarantee a convergence to coreness, as shown in the following theorem.

Theorem 2: given an undirected simple network $G(V, E)$, for every node $j \in V$, we define $g_j = k_j$. In each iteration of the asynchronous updating process, a node i is randomly selected and its g value updated, that is,

$$\mathcal{H}(g_{j_1}, g_{j_2}, \dots, g_{j_{k_i}}) \rightarrow g_i, \quad (4)$$

where j_1, j_2, \dots, j_{k_i} are the neighbouring nodes of i . If $|V|$ is finite, this updating process will reach a steady state $(g_1^\infty, g_2^\infty, \dots, g_{|V|}^\infty)$ after a finite number of iterations such that the updating at any node will not change its g value, namely,

$$\forall i \in V, \quad g_i^\infty = \mathcal{H}(g_{j_1}^\infty, g_{j_2}^\infty, \dots, g_{j_{k_i}}^\infty). \quad (5)$$

In the steady state, for every node i we have $g_i^\infty = c_i$. The proof is given in the Methods section.

Note that in the updating process of equation (4) the g values are not associated with a temporal superscript. In fact, at a certain updating step the values $g_{j_1}, g_{j_2}, \dots, g_{j_{k_i}}$ could lie in different stages, some updated dozens of times and others never updated. Thus, for any node i , before it reaches the steady state $g_i^\infty = c_i$, all the intermediate values are not necessarily equal to any order of H-indices. Theorem 2 makes a significant step towards making feasible applications to large-scale dynamical networks possible: it guarantees that a decentralized and localized method can be used to calculate the coreness, and even if the network evolves in time, equation (4) can be used to calculate the coreness when new links and nodes are periodically added, and all the previously obtained g values are still usable.

Quantifying spreading influences. Epidemic spreading, one of the most significant dynamic behaviours in complex networks⁴¹, can be used to characterize such real processes as the spreading of infectious diseases⁴², the diffusion of microfinance⁴³ and the propagation of traffic congestion⁴⁴. To see whether the H-indices

Table 1 | The basic topological features and the convergence time of the eight real networks.

Networks	$ V $	$ E $	$\langle k \rangle$	$\langle d \rangle$	C	r	n_∞
Jazz	198	2,742	27.7	2.24	0.633	0.02	13
NS	379	914	4.82	6.04	0.798	−0.082	4
Email	1,133	5,451	9.62	3.61	0.254	0.078	16
Sex	15,810	38,540	4.88	5.79	0	−0.115	19
Facebook	63,731	817,090	25.64	8.09	0.253	0.177	63
PB	1,222	16,714	27.36	2.74	0.360	−0.221	18
USAir	332	2,126	12.81	2.74	0.749	−0.208	5
Router	5,022	6,258	2.49	6.45	0.033	−0.138	6

$|V|$ and $|E|$ are the number of nodes and links, respectively. $\langle k \rangle$ and $\langle d \rangle$ are the average degree and the average distance, respectively. C and r are the clustering coefficient¹ and assortative coefficient³, respectively. Nodes with degree 1 are excluded from the calculation of clustering coefficient. Sex is a bipartite network and thus is characterized by (or 'has') clustering coefficient zero. n_∞ is the convergence time to coreness, defined as the minimum steps required to reach coreness from degree by the operator \mathcal{H} .

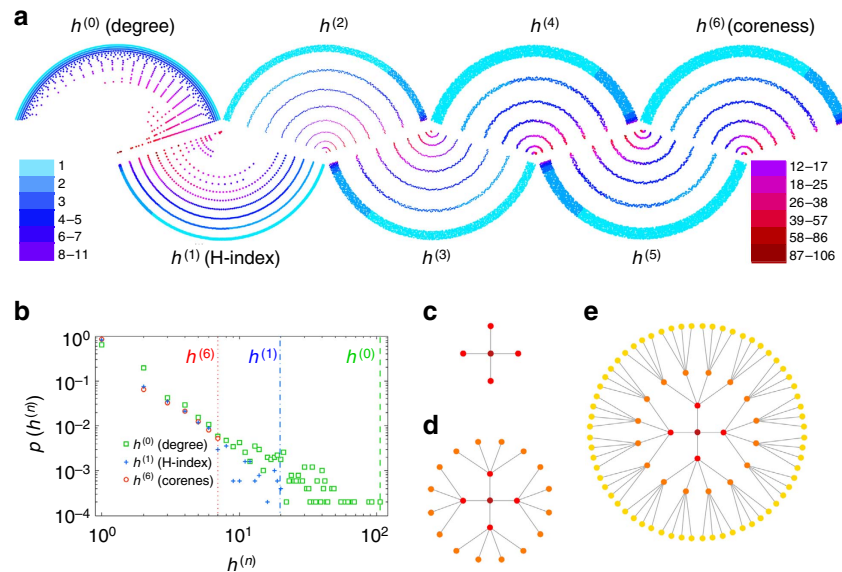


Figure 1 | Comparisons among H-indices in different orders for the network Router. The subplot **a** exhibits a visualized illustration, where the colour represents the node degree (from 1 to 106). The node location represents the $h^{(n)}$ -index. Nodes located at a ‘fan’ closer to the centre of the fan have higher $h^{(n)}$ values, and nodes located at the same layer of a fan have the same $h^{(n)}$ values. The subplot **b** shows the distributions of values of H-indices in different orders, where the green squares, blue crosses and red circles represent the cases for $n=0$, $n=1$ and $n=6$, respectively. The dash lines in different colours mark the largest values for the corresponding indices. The subplots **c**, **d** and **e** show an illustration of the hierarchical trees with 2, 3 and 4 levels, respectively.

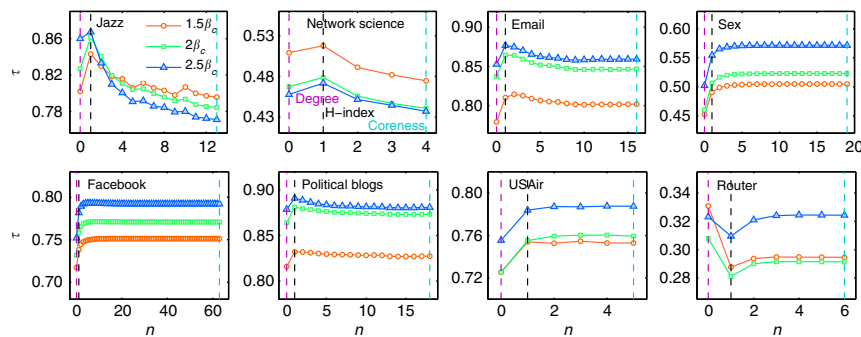


Figure 2 | The Kendall Tau between $h^{(n)}$ -index and the node influence index R for undirected networks. The value of n ranges from 0 to n_{∞} . The red circles, green squares and blue triangles represent the case of $\beta=1.5\beta_c$, $2\beta_c$ and $2.5\beta_c$, respectively. The dash lines in purple, black and cyan colours emphasize the τ values for degree (that is, $h^{(0)}$), H-index (that is, $h^{(1)}$) and coreness (that is, $h^{(\infty)}$), respectively. The influence R of a node is quantified using the average number of removed nodes after the dynamics over 1,000 independent runs.

can quantify the spreading of node influence, we study the standard SIR spreading model⁴¹ in which the influence R_i of node i is quantified using the average number of removed nodes after the dynamics over 1,000 independent runs, each of which begins with node i as the sole infected seed (see details in Methods).

Given the order n ($0 \leq n \leq n_{\infty}$), we have two sequences associated with the $|V|$ nodes: the $h^{(n)}$ -index $h_1^{(n)}, h_2^{(n)}, \dots, h_{|V|}^{(n)}$ and the influences $R_1, R_2, \dots, R_{|V|}$. To quantify to what extent the $h^{(n)}$ -index resembles node influence values, we apply the Kendall Tau (τ) coefficient⁴⁵, which lies in the range $-1 \leq \tau \leq 1$. The larger value of τ means a stronger correlation between the two sequences (see Methods for the definition of τ).

Figure 2 shows that the H-index of node i is highly correlated with the influence value R_i . In many cases (Jazz, NS, Email and PB), the H-index (that is, $h^{(1)}$ -index) outperforms both degree and coreness. In other cases (Sex, USAir and Facebook), coreness outperforms degree and the performance of low-order H-indices rapidly approaches that of coreness as the order increases from zero. Router is an exception, where degree performs the best and

the H-index the worst. Note that because Router is the most sparse network it may hinder spreading and make predicting influences more difficult (as indicated by the smallest τ value). Thus, the sequence of H-indices, starting from degree and driven by the operator \mathcal{H} , provides more alternative centralities in characterizing the importance of nodes, and the low-order H-indices are a good tradeoff between degree and coreness. We further compare the three best known H-indices (degree, H-index and coreness) with two well-known centrality indices (closeness and betweenness) for undirected networks. The definitions of closeness and betweenness are given in Methods. As shown in Table 2, all the H-family indices are competitive, and the H-index and closeness are the overall best performers, but the computational cost in calculating closeness is huge and thus the H-index is the better choice.

We further test two well-known dynamical processes: the susceptible-infected-susceptible (SIS) spreading model⁴⁶ and bond percolation⁴⁷. In the SIS model, the node influence index R is defined as the probability that this node will remain infected

Table 2 | The Kendall Tau between the node influence index R of SIR model and five centrality indices.

Networks	Degree	H-index	Coreness	Closeness	Betweenness
Jazz	0.8021	0.8431	0.7958	0.6961	0.4629
NS	0.5092	0.5178	0.4747	0.3510	0.3392
Email	0.7794	0.8103	0.8021	0.7747	0.6195
Sex	0.4525	0.4905	0.5049	0.7029	0.3834
Facebook	0.7173	0.7381	0.7513	0.6716	0.4851
PB	0.8159	0.8321	0.8274	0.7375	0.6589
USAir	0.7256	0.7540	0.7529	0.7453	0.5442
Router	0.3309	0.2877	0.2946	0.5975	0.3228

SIR, susceptible-infected-removed.

They are degree (that is, $h^{(0)}$), H-index (that is, $h^{(1)}$), coreness (that is, $h^{(\infty)}$), closeness and betweenness. The spreading rate β is set as $\beta = 1.5\beta_c$ and for other values of β , the results are very similar and can be found in the Supplementary Table 1 ($\beta = 2\beta_c$) and Supplementary Table 2 ($\beta = 2.5\beta_c$). In each row, the largest τ is highlighted in bold.

in the steady state. In bond percolation, the node influence index R is defined as either the probability that the target node belongs to the giant component or the size of the connected component that encompasses the target node. The results (Supplementary Tables 3–11) suggest that the H-family indices are competitive, especially for the H-index and coreness. Detailed information about the dynamical processes and the simulation results can be found in Supplementary Note 4.

Discussion

We discover an important relation among degree, H-index and coreness—centrality measures that have previously been treated as unrelated. We construct an operator \mathcal{H} that functions as a ‘necklace’ stringing together degree, H-index, coreness and other intermediate indices. All these indices are centralities that characterize each node’s importance. Using the operator \mathcal{H} to achieve the coreness looks like an inverse way to the iterative removal of nodes with degree less than k that is widely used to determine the k -core of a network. Indeed, they are different, as the iterative removal method cannot result in H-index, or any other H-family indices except for degree and coreness, and the steps required to achieve the final coreness for the two methods are also different.

Although the importance of a given node strongly depends on the type of dynamical processes under consideration and thus there is no single best centrality measure, we need effective and elegant centralities in practice. For example, although we know that degree is not an accurate centrality measure in quantifying node influence in many dynamical processes^{9,22}, it is still a useful estimation of node importance, even without the specification of dynamics. In despite of its bias and disadvantages^{48,49,50} the H-index is now becoming the most widely applied index for academic performances, ranging from individual scientists, scientific journals to universities and even countries. As indicated by the simulations on the SIR model, the SIS model and bond percolation, the H-family indices are effective in quantifying the spreading influences of nodes.

The asynchronous updating by \mathcal{H} can still guarantee the convergence to coreness, and thus one can use a decentralized local algorithm to calculate coreness, which is able to deal with evolving networks. However, randomly selecting nodes to update in each iteration may greatly extend the time required before arriving at the steady state, even in static networks. Thus, the process for selecting which node to update is a nontrivial issue. For example, we can shorten the convergence time by reducing the selection probability of nodes that have been updated many times but have g values seldom changed. In addition, the change of g value of a node will enhance the updating probabilities of its neighbouring nodes, making this issue more complicated and thus more interesting.

The methodologies and results presented here are also applicable to directed networks and weighted networks, in which degree is replaced by in-degree, out-degree or node strength. In this way we can also define the H_{in} -index, H_{out} -index and weighted H-index, as well as in-coreness, out-coreness and weighted coreness. An example of how to calculate the directed H-family indices is presented in Supplementary Fig. 13. To test the performance, we compare the directed H-family indices with PageRank⁵¹ and HITs⁵² on seven directed networks. The basic statistics of these seven directed networks are summarized in Supplementary Table 12 and a more detailed introduction given in Supplementary Note 5. The results (Supplementary Tables 13–24) suggest that the directed H-family indices are still very competitive (in-coreness performs overall best).

Methods

Proof of Theorem 1. From the definition of \mathcal{H} -function, for any node i and integer $n \geq 0$, we have $h_i^{(n)} \geq 0$ and $h_i^{(0)} \geq h_i^{(1)}$. Applying mathematical induction, we prove that $h_i^{(n)} \geq h_i^{(n+1)}$ for any node i and integer $n \geq 0$. If $h_i^{(n)} \geq h_i^{(n+1)}$ is valid for all $n \leq m$, we then prove it to also be valid for $n = m + 1$. From the definition, $h_i^{(m+2)} = \mathcal{H}(h_{j_1}^{(m+1)}, h_{j_2}^{(m+1)}, \dots, h_{j_{k_i}}^{(m+1)})$, according to the induction hypothesis, $h_j^{(m+1)} \leq h_j^{(m)}$ for any node j . Therefore, $h_i^{(m+2)} = \mathcal{H}(h_{j_1}^{(m+1)}, h_{j_2}^{(m+1)}, \dots, h_{j_{k_i}}^{(m+1)}) \leq \mathcal{H}(h_{j_1}^{(m)}, h_{j_2}^{(m)}, \dots, h_{j_{k_i}}^{(m)}) = h_i^{(m+1)}$, namely $h_i^{(m+1)} \geq h_i^{(m+2)}$. Since $h_i^{(0)}, h_i^{(1)}, h_i^{(2)}, \dots$ is a monotonously nonincreasing sequence, and each element is nonnegative, it has a nonnegative limitation. Thus, we can define h_i^∞ as the limitation of the sequence $h_i^{(0)}, h_i^{(1)}, h_i^{(2)}, \dots$.

We then introduce two simple relations. First, if $G'(V', E') \subseteq G(V, E)$, then from the definition of an \mathcal{H} -function it is obvious that for any node $i \in V'$ and any integer $n \geq 0$, $h_{i,G'}^{(n)} \leq h_{i,G}^{(n)}$, where the subscript G' indicates that the corresponding index is defined on the subgraph G' . Second, if we denote k_{\min} the minimal degree of G , then for any node i and any integer $n \geq 0$, $h_i^{(n)} \geq k_{\min}$. This second relation can also be proven by mathematical induction. It is clear that it is valid for $n = 0$. We next prove that if this relation is valid for all $n \leq m$, then it is also valid for $n = m + 1$. For any node i , $h_i^{(m+1)} = \mathcal{H}(h_{j_1}^{(m)}, h_{j_2}^{(m)}, \dots, h_{j_{k_i}}^{(m)})$, and according to the induction hypothesis, all elements in $(h_{j_1}^{(m)}, h_{j_2}^{(m)}, \dots, h_{j_{k_i}}^{(m)})$ are no less than k_{\min} and the number of elements in $(h_{j_1}^{(m)}, h_{j_2}^{(m)}, \dots, h_{j_{k_i}}^{(m)})$ is $k_i \geq k_{\min}$. Therefore, according to the definition of \mathcal{H} -function, $h_i^{(m+1)} \geq k_{\min}$.

If we denote G' the c_i -core of G , it is clear that $G' \subseteq G$ and in G' , $k_{\min} \geq c_i$. Therefore, $h_i^\infty \geq k_{\min} \geq c_i$. We denote $G''(V'', E'')$ the induced subgraph containing all nodes j such that $h_j^\infty \geq h_i^\infty$. Note that the node i itself also belongs to G'' . For any node $l \in V$, we find $h_l^\infty = \mathcal{H}(h_{j_1}^\infty, h_{j_2}^\infty, \dots, h_{j_{k_l}}^\infty)$, where j_1, j_2, \dots, j_{k_l} are the k_l neighbours of node l . For any node $j \in V''$, since in G , $h_j^\infty > h_i^\infty$, there are at least h_i^∞ neighbours of node j with h^∞ values no less than h_i^∞ . Thus, these neighbours also belong to V'' . Therefore, in G'' the degrees of all the nodes are no less than h_i^∞ , that is, G'' is a subgraph of G 's h_i^∞ -core. Because c_i is the coreness of i , $h_i^\infty \leq c_i$. Combining the two inequalities, we arrive at Theorem 1.

Proof of Theorem 2. For convenience, we introduce the systematic time step t . Initially we set $t = 0$, and for every node $j \in V$ we define $g_j^{(0)} = k_j$. Then, at each time step, we randomly select a node and perform the \mathcal{H} operator on it. If at time step $t > 0$ the node i is selected, then $g_i^{(t)} = \mathcal{H}(g_{j_1}, g_{j_2}, \dots, g_{j_{k_i}})$. The g value without

a temporal superscript indicates the most recently updated value, since only the current value is meaningful in the asynchronous updating procedure. Note that all neglected superscripts are smaller than t . Note also that for an arbitrary pair (t, j) , $g_j^{(t)}$ may not exist unless the node j is selected at time step t .

We first prove that if any node $j \in V$ has been selected at time steps t_1 and t_2 , and $t_2 > t_1 \geq 0$, then $g_j^{(t_1)} \geq g_j^{(t_2)}$. Note that for any node j selected at time step $t=1$ we have $g_j^{(0)} \geq g_j^{(1)}$. We apply mathematical induction and assume that the above inequality holds when $t_1 \leq n$ and $t_2 \leq n$, and we next prove this also holds for $t_1 \leq n+1$ and $t_2 \leq n+1$. If node i is the one selected at time step $t=n+1$, then the aforementioned inequality holds for all other nodes $j \neq i$. We denote t' ($0 \leq t' \leq n$) an arbitrary earlier updating time step of node i , and record two updates, that is, $g_i^{(t')} = \mathcal{H}(g_{i_1}^{(\phi_1)}, g_{i_2}^{(\phi_2)}, \dots, g_{i_{k_i}}^{(\phi_{k_i})})$ and $g_i^{(n+1)} = \mathcal{H}(g_{i_1}^{(\phi_1)}, g_{i_2}^{(\phi_2)}, \dots, g_{i_{k_i}}^{(\phi_{k_i})})$. Note that for any m ($1 \leq m \leq k_i$) we have $\phi_m \leq \phi_m \leq n$. According to the induction hypothesis, $g_m^{(\phi_m)} \geq g_m^{(n+1)}$, together with the definition of \mathcal{H} function, we have $g_i^{(t')} \geq g_i^{(n+1)}$. We denote the updating time steps of any node $i \in V$ to be $0 = t_0 < t_1 < t_2 < \dots$, then $g_i^{(t_0)}, g_i^{(t_1)}, g_i^{(t_2)}, \dots$ is a monotonously nonincreasing sequence and each element is nonnegative, and therefore it has a nonnegative limitation. At this point, we can define g_i^∞ as the limitation of the sequence $g_i^{(t_0)}, g_i^{(t_1)}, g_i^{(t_2)}, \dots$.

We first prove that for any node $j \in V$, $g_j^\infty \geq c_j$. Proving by contradiction, when this inequality does not hold we denote i to be the first node to reach a g value smaller than c_i , and the corresponding updating time step is t , that is, $g_i^{(t)} < c_i$ and before t for all nodes $j \in V$, $g_j \geq c_j$. Note that g_j without a superscript indicates the last updated value before t . Therefore, $g_i^{(t)} = \mathcal{H}(g_{i_1}, g_{i_2}, \dots, g_{i_{k_i}}) \geq \mathcal{H}(c_{j_1}, c_{j_2}, \dots, c_{j_{k_i}})$. According to Theorem 1, $\mathcal{H}(c_{j_1}, c_{j_2}, \dots, c_{j_{k_i}}) = c_i$, namely $g_i^{(t)} \geq c_i$. This leads to a contradiction and thus the inequality $g_j^\infty \geq c_j$ is validated.

Analogous to the proof of Theorem 1, after convergence, for any node $i \in V$, all nodes j such that $g_j^\infty \geq g_i^\infty$, including i itself, constitute an induced subgraph of G 's g_i^∞ -core. Since c_i is the coreness of i , $g_i^\infty \leq c_i$. Combining the two inequalities, we arrive at Theorem 2.

Spreading models. The standard SIR model, also referred to as the susceptible-infected-recovered model, is usually applied to analyse the propagation of opinions or news⁴¹. In the SIR model, there is a group of infected seed nodes and all other nodes are initially susceptible. At each time step, each infected node makes contact with its neighbours and each susceptible neighbour is infected with a probability β . Then, each infected node enters the removed state with a probability λ . For simplicity, we set $\lambda = 1$. To quantify the spreading influence of a target node i , we begin the spreading process with i being the sole infected seed. When there are no longer any infected nodes and the dynamic process ends, the number of removed nodes R_i is a measurement of the influence of node i . Because we use small β values in our simulations, the infected percentage of the population is also small. When β values are high the disease infects a large percentage of the population, irrespective of where it originated, and the influence of individual nodes cannot be measured. According to the heterogeneous mean-field theory^{47,53,54}, the epidemic threshold of SIR model is approximate to $\beta_c \approx \langle k \rangle / (\langle k^2 \rangle - \langle k \rangle)$. To be more precise, we determine the epidemic threshold β_c by simulation on real networks⁵⁵. We set $\beta = 1.5\beta_c$, $2\beta_c$ and $2.5\beta_c$ in this paper, and we have checked that the choice of theoretical or simulation threshold will not affect the conclusion. Because the fluctuation of R_i is large when β values are small, we use 1,000 independent implementations for averaging.

Benchmark centralities. We also compare two benchmark centrality indices (that is, closeness and betweenness) for undirected networks. Betweenness is one of the most popular geodesic-path-based ranking measures. It is defined as the fraction of shortest paths between all node pairs that pass through the node of interest. Betweenness is, in some sense, a measure of the influence of a node in terms of its role in spreading information^{20,56}. For a network $G = (V, E)$, the betweenness centrality of node i , denoted by $B(i)$, is defined as^{57,58}

$$B(i) = \sum_{s \neq i, s \neq t, i \neq t} \frac{g_{st}(i)}{g_{st}}, \quad (6)$$

where g_{st} is the number of shortest paths between nodes s and t , and $g_{st}(i)$ denotes the number of shortest paths between nodes s and t that pass through node i .

Closeness of node i is defined as the reciprocal of the sum of geodesic distances to all other nodes of V ^{59,60},

$$C(i) = \frac{1}{\sum_{j \in V \setminus i} d(i, j)}, \quad (7)$$

where $d(i, j)$ is the geodesic distance between i and j . Closeness can be used to measure of how far information will be able to spread from a given node to other reachable nodes in the network.

The Kendall Tau. We consider two sequences associated with $|V|$ nodes, $X = (x_1, x_2, \dots, x_{|V|})$ and $Y = (y_1, y_2, \dots, y_{|V|})$, as well as the $|V|$ two-tuples $(x_1, y_1), (x_2, y_2), \dots, (x_{|V|}, y_{|V|})$. Any pair of two-tuples (x_i, y_i) and (x_j, y_j) ($i \neq j$) are concordant if the ranks for both elements agree, that is, if both $x_i > x_j$ and $y_i > y_j$ or if both $x_i < x_j$ and $y_i < y_j$. They are discordant if $x_i > x_j$ and $y_i < y_j$ or if $x_i < x_j$ and $y_i > y_j$. If $x_i = x_j$ or $y_i = y_j$, the pair is neither concordant nor discordant. Comparing all $\frac{1}{2}|V|(|V|-1)$ pairs of two-tuples, the Kendall Tau is defined as $\tau = 2(n_+ - n_-) / [|V|(|V|-1)]$, where n_+ and n_- are the number of concordant and discordant pairs, respectively. The coefficient must be in the range $-1 \leq \tau \leq 1$, and if X and Y are independent, τ should be approximately zero and thus the extent to which τ exceeds zero indicates the strength of the correlation.

Notations. We summarize the notations used in this paper in Supplementary Table 25.

References

- Watts, D. J. & Strogatz, S. H. Collective dynamics of 'small-world' networks. *Nature* **393**, 440–442 (1998).
- Barabási, A.-L. & Albert, R. Emergence of scaling in random networks. *Science* **286**, 509–512 (1999).
- Newman, M. E. J. Assortative mixing in networks. *Phys. Rev. Lett.* **89**, 208701 (2002).
- Newman, M. E. J. & Girvan, M. Finding and evaluating community structure in networks. *Phys. Rev. E* **69**, 026113 (2004).
- Csermely, P. *Weak Links: Stabilizers of Complex Systems from Proteins to Social Networks* (Springer, 2006).
- Alon, U. Network motifs: theory and experimental approaches. *Nat. Rev. Genet.* **8**, 450–461 (2007).
- Cheng, X.-Q., Ren, F.-X., Shen, H.-W., Zhang, Z.-K. & Zhou, T. Bridginess: a local index on edge significance in maintaining global connectivity. *J. Stat. Mech.* **2010**, P10011 (2010).
- Lü, L., Zhang, Y.-C., Yeung, C. H. & Zhou, T. Leaders in social networks: the delicious case. *PLoS ONE* **6**, e21202 (2011).
- Klemm, K., Serrano, M. Á., Equiluz, V. M. & Miguel, M. S. A measure of individual role in collective dynamics. *Sci. Rep.* **2**, 292 (2012).
- Wang, P., Lü, J. & Yu, X. Identification of important nodes in directed biological networks: a network motif approach. *PLoS ONE* **9**, e106132 (2014).
- Pei, S. & Makse, H. A. Spreading dynamics in complex networks. *J. Stat. Mech.* **2013**, P12002 (2013).
- Pastor-Satorras, R. & Vespignani, A. Immunization of complex networks. *Phys. Rev. E* **65**, 036104 (2002).
- Leskovec, J., Adamic, L. A. & Huberman, B. A. The dynamics of viral marketing. *ACM Trans. Web* **1**, 5 (2007).
- Albert, R., Albert, I. & Nakarado, G. L. Structural vulnerability of the North American power grid. *Phys. Rev. E* **69**, 025103 (2004).
- Morone, F. & Makse, H. A. Influence maximization in complex networks through optimal percolation. *Nature* **524**, 65–68 (2015).
- Csermely, P., Korcsmros, T., Kiss, H. J., London, G. & Nussinov, R. Structure and dynamics of molecular networks: a novel paradigm of drug discovery: a comprehensive review. *Pharmacol. Ther.* **138**, 333–408 (2013).
- Resende, M. G. C. & Pardalos, P. M. *Handbook of Optimization in Telecommunications* (Springer, 2006).
- Albert, R., Jeong, H. & Barabási, A.-L. Error and attack tolerance of complex networks. *Nature* **406**, 378–382 (2000).
- Motter, A. E., Zhou, C. & Kurths, J. Network synchronization, diffusion, and the paradox of heterogeneity. *Phys. Rev. E* **71**, 016116 (2005).
- Yan, G., Zhou, T., Hu, B., Fu, Z.-Q. & Wang, B.-H. Efficient routing on complex networks. *Phys. Rev. E* **73**, 046108 (2006).
- Santos, F. C., Pacheco, J. M. & Lenaerts, T. Evolutionary dynamics of social dilemmas in structured heterogeneous populations. *Proc. Natl Acad. Sci. USA* **103**, 3490–3494 (2006).
- Kitsak, M. et al. Identification of influential spreaders in complex networks. *Nat. Phys.* **6**, 888–893 (2010).
- Dorogovtsev, S. N., Goltsev, A. V. & Mendes, J. F. F. K-core organization of complex networks. *Phys. Rev. Lett.* **96**, 040601 (2006).
- Pei, S., Muchnik, L., Andrade, J. S., Zheng, Z. M. & Makse, H. A. Searching for superspreaders of information in real-world social media. *Sci. Rep.* **4**, 5547 (2014).
- Chen, D.-B., Lü, L., Shang, M.-S., Zhang, Y.-C. & Zhou, T. Identifying influential nodes in complex networks. *Physica A* **391**, 1777–1787 (2012).
- Chen, D.-B., Gao, H., Lü, L. & Zhou, T. Identifying influential nodes in large-scale directed network: the role of clustering. *PLoS ONE* **8**, e77455 (2013).
- Hirsch, J. E. An index to quantify an individual's scientific research output. *Proc. Natl Acad. Sci. USA* **102**, 16569–16572 (2005).
- Braun, T., Glänzel, W. & Schubert, A. A Hirsch-type index for journals. *Scientometrics* **69**, 169–173 (2006).
- Hirsch, J. E. Does the h index have predictive power? *Proc. Natl Acad. Sci. USA* **104**, 19193–19198 (2007).

30. Korn, A., Schubert, A. & Telcs, A. Lobby index in networks. *Physica A* **388**, 2221–2226 (2009).
31. Hethcote, H. W. The mathematics of infectious diseases. *SIAM Rev.* **42**, 599–653 (2000).
32. Gleiser, P. & Danon, L. Community structure in Jazz. *Adv. Complex Syst.* **6**, 565 (2003).
33. Newman, M. E. J. Finding community structure in networks using the eigenvectors of matrices. *Phys. Rev. E* **74**, 036104 (2006).
34. Guimerà, R., Danon, L., Diaz-Guilera, A., Giral, F. & Arenas, A. Self-similar community structure in a network of human interactions. *Phys. Rev. E* **68**, 065103 (2003).
35. Rocha, L. E., Liljeros, F. & Holme, P. Simulated epidemics in an empirical spatiotemporal network of 50,185 sexual contacts. *PLoS Comput. Biol.* **7**, e1001109 (2011).
36. Viswanath, B., Mislove, A., Cha, M. & Gummadi, K. P. in: *Proceedings of the 2nd Workshop on Online Social Networks*. 37–42 (ACM, 2009).
37. Adamic, L. A. & Glance, N. in: *Proceedings of the 3rd International Workshop on Link Discovery*. 36–43 (ACM, 2004).
38. Batageli, V. & Mrvar, A. Pajek Datasets. Available at <http://vlado.fmf.uni-lj.si/pub/networks/data/>. 2007.
39. Spring, N., Mahajan, R., Wetherall, D. & Anderson, T. Measuring ISP topologies with Rocketfuel. *IEEE/ACM Trans. Networking* **12**, 2–16 (2004).
40. Ravasz, E. B. & Barabasi, A.-L. Hierarchical organization in complex networks. *Phys. Rev. E* **67**, 026112 (2003).
41. Pastor-Satorras, R., Castellano, C., Van Mieghem, P. & Vespignani, A. Epidemic processes in complex networks. *Rev. Mod. Phys.* **87**, 925 (2015).
42. Keeling, M. J. & Rohani, P. *Modeling Infectious Diseases in Humans and Animals* (Princeton Univ. Press, 2008).
43. Banerjee, A., Chandrasekhar, A. G., Duflo, E. & Jackson, M. O. The diffusion of microfinance. *Science* **341**, 1236498 (2013).
44. Li, D. *et al.* Percolation transition in dynamical traffic network with evolving critical bottlenecks. *Proc. Natl Acad. Sci. USA* **112**, 669–672 (2015).
45. Kendall, M. A new measure of rank correlation. *Biometrika* **30**, 81–89 (1938).
46. Pastor-Satorras, R. & Vespignani, A. Epidemic spreading in scale-free networks. *Phys. Rev. Lett.* **86**, 3200 (2001).
47. Newman, M. E. J. Spread of epidemic disease on networks. *Phys. Rev. E* **66**, 016128 (2002).
48. Egghe, L. Theory and practise of the g-index. *Scientometrics* **69**, 131–152 (2006).
49. Jin, B., Liang, L., Rousseau, R. & Egghe, L. The R-and AR-indices: complementing the h-index. *Chinese Sci. Bull.* **52**, 855–863 (2007).
50. Dorogovtsev, S. N. & Mendes, J. F. F. Ranking Scientists. *Nature Physics* **11**, 882–883 (2015).
51. Brin, S. & Page, L. The anatomy of a large-scale hypertextual Web search engine. *Comput. Networks ISDN Syst.* **30**, 107–117 (1998).
52. Kleinberg, J. Authoritative sources in a hyperlinked environment. *J. ACM* **46**, 604–632 (1999).
53. Cohen, R., Erez, K., Ben-Avraham, D. & Havlin, S. Resilience of the Internet to random breakdowns. *Phys. Rev. Lett.* **85**, 4626 (2000).
54. Castellano, C. & Pastor-Satorras, R. Thresholds for epidemic spreading in networks. *Phys. Rev. Lett.* **105**, 218701 (2010).
55. Shu, P., Wang, W., Tang, M. & Do, Y. Numerical identification of epidemic thresholds for susceptible-infected-recovered model on finite-size networks. *Chaos* **25**, 063104 (2015).
56. Guimerà, R., Das-Guilera, A., Vega-Redondo, F., Cabrales, A. & Arenas, A. Optimal network topologies for local search with congestion. *Phys. Rev. Lett.* **89**, 248701 (2002).
57. Freeman, L. C. A set of measures of centrality based on betweenness. *Sociometry* **40**, 35–41 (1977).
58. Freeman, L. C. Centrality in social networks conceptual clarification. *Soc. Networks* **1**, 215–239 (1979).
59. Bavelas, A. Communication patterns in task-oriented groups. *J. Acoust. Soc. Am.* **22**, 725–730 (1950).
60. Sabidussi, G. The centrality index of a graph. *Psychometrika* **31**, 581–603 (1966).

Acknowledgements

This work was partially supported by the National Natural Science Foundation of China (Grant Nos. 11205042, 11222543, 11075031, 61433014). L.L. acknowledges the research start-up fund of Hangzhou Normal University under Grant No. PE13002004039 and the EU FP7 Grant 611272 (project GROWTHCOM). The Boston University work was supported by NSF Grants CMMI 1125290, CHE 1213217 and PHY 1505000.

Author contributions

L.L. and T.Z. planned and performed this research, contributed equally to this work, and wrote this paper. Q.M.Z. performed research. H.E.S. planned the research and edited this paper. All authors discussed the results and reviewed the manuscript.

Additional information

Supplementary Information accompanies this paper at <http://www.nature.com/naturecommunications>

Competing financial interests: The authors declare no competing financial interests.

Reprints and permission information is available online at <http://npg.nature.com/reprintsandpermissions/>

How to cite this article: Lü, L. *et al.* The H-index of a network node and its relation to degree and coreness. *Nat. Commun.* **7**:10168 doi: 10.1038/ncomms10168 (2016).



This work is licensed under a Creative Commons Attribution 4.0 International License. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in the credit line; if the material is not included under the Creative Commons license, users will need to obtain permission from the license holder to reproduce the material. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>