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

Linyuan Lu¹, Tao Zhou²,³, Qian-Ming Zhang²,⁴ & H. Eugene Stanley¹,⁴

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.
The focus of network science research has been shifting from discovering macroscopic statistical regularities\textsuperscript{1–3} to uncovering the role played by such microscopic elements as nodes, links and motifs in the structure and dynamics of the system\textsuperscript{4–10}. Being able to effectively and efficiently identify the critical nodes associated with the specific dynamics of large-scale networks\textsuperscript{11} will allow us to better control the outbreak of epidemics\textsuperscript{12}, conduct successful advertisements for e-commercial products\textsuperscript{13}, prevent catastrophic outages in power grids or the Internet\textsuperscript{14}, optimize the use of limited resources to facilitate information propagation\textsuperscript{15}, discover drug target candidates and essential proteins\textsuperscript{16}, and design strategies for communication breakdowns in human and telecommunication networks\textsuperscript{17}.

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 connectivity\textsuperscript{18}, halt infectious disease propagation\textsuperscript{19}, enhance synchronizability\textsuperscript{19}, improve transport capacity\textsuperscript{20} and promote cooperation in evolutionary games\textsuperscript{21}.

Recently, Kitsak \textit{et al.}\textsuperscript{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 \textit{k}-core decomposition\textsuperscript{23}, and a larger coreness value indicates that the node is more centrally located in the network. The \textit{k}-core decomposition process is initiated by removing all nodes with degree \textit{k} = 1. This causes new nodes with degree \textit{k} ≤ 1 to appear. These are also removed and the process is continued until the only nodes remaining are those of degree \textit{k} > 1. The removed nodes and their associated links form the 1-shell. This pruning process is repeated for the nodes of degree \textit{k} = 2 to extract the 2-shell, that is, in each stage the nodes with degree \textit{k} ≤ 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 \textit{i} is assigned a shell layer \textit{c}_\textit{i}, called the coreness of node \textit{i}. Recent studies suggest that coreness is a good measure of a node’s influence\textsuperscript{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 \textit{et al.}\textsuperscript{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)\textsuperscript{27}, which was originally used to measure the citation impact of a scholar or a journal\textsuperscript{27–29}. For a scholar or journal, the H-index is defined as the maximum value \textit{h} such that there exists at least \textit{h} papers, each with citation count \textit{≥ h}.

Here we discuss the extension of the H-index concept to quantify how important a node is to its network\textsuperscript{30}. The H-index of a node is defined to be the maximum value \textit{h} such that there exists at least \textit{h} neighbours of degree no less than \textit{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 model\textsuperscript{31} 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 \((x_1, x_2, \ldots, x_n)\) and returns an integer \(y = \mathcal{H}(x_1, x_2, \ldots, x_n) \geq 0\), where \(y\) is the maximum integer such that there exist at least \(y\) elements in \((x_1, x_2, \ldots, x_n)\), each of which is no less than \(y\). For a scholar with \(n\) publications, \(x_1, x_2, \ldots, x_n\) is the number of citations to these publications and \(\mathcal{H}(x_1, x_2, \ldots, x_n)\) 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 \(k_{j_1}, k_{j_2}, \ldots, k_{j_{h_i}}\). Then, we define the H-index of node \(i\)

\[ h_i = \mathcal{H}\left(k_{j_1}, k_{j_2}, \ldots, k_{j_{h_i}}\right). \]  

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

\[ h_{i}^{(n)} = \mathcal{H}\left(h_{i}^{(n-1)}, h_{j_1}^{(n-1)}, \ldots, h_{j_{h_i}}^{(n-1)}\right). \]  

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 \(h_{i}^{(0)}, h_{i}^{(1)}, h_{i}^{(2)}, \ldots\) will converge to the coreness of node \(i\).

\[ c_i = \lim_{n \to \infty} h_{i}^{(n)}. \]  

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 \(n_\infty\) is defined as the minimum number of iterations required to reach coreness from degree using the operator \(\mathcal{H}\), that is, \(n_\infty\) 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, Jazz\textsuperscript{32} is a collaboration network of jazz musicians and consists of 198 nodes and 2,742 interactions, NS\textsuperscript{33} is a co-authorship network of scientists working on network science, our Email\textsuperscript{34} network is of e-mail interchanges between members of the Rovira i Virgili University (Tarragona), our Sex\textsuperscript{35} 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, Facebook\textsuperscript{36} is a sample of the friendship network of Facebook users, PB\textsuperscript{37} is a network of US political blogs (the original links are directed, but here we treat them as undirected), USAir\textsuperscript{38} is the US air transportation network.
network and Router$^{39}$ 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 ($n_\infty$ in Table 1). In addition to the degree, H-index and coreness, all other intermediate states $h^{(2)}, h^{(3)}, \ldots$ 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 $h^{(n)}$ 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).

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 importance.

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 $h^{(0)}, h^{(1)}, \ldots, h^{(D-2)}$ 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 $H$ 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 value updated, that is,

$$H(g_{j1}, g_{j2}, \ldots, g_{jh}) \rightarrow g_{i}.$$  \hspace{1cm} (4)

where $j_1, j_2, \ldots, j_h$ are the neighbouring nodes of $i$. If $|V|$ is finite, this updating process will reach a steady state $(g^{\infty}_{j1}, g^{\infty}_{j2}, \ldots, g^{\infty}_{j|V|})$ after a finite number of iterations such that the updating at any node will not change its $g$ value, namely,

$$\forall i \in V, \quad g_{i}^{\infty} = H(g_{j1}^{\infty}, g_{j2}^{\infty}, \ldots, g_{jh}^{\infty}).$$  \hspace{1cm} (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 $g_{j1}, g_{j2}, \ldots, g_{jh}$ 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 networks$^{41}$, can be used to characterize such real processes as the spreading of infectious diseases$^{42}$, the diffusion of microfinance$^{43}$ and the propagation of traffic congestion$^{44}$. To see whether the H-indices

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

| Networks | $|V|$ | $|E|$ | $(k)$ | $(d)$ | $C$ | $r$ | $n_\infty$ |
|----------|------|------|------|------|-----|-----|-----------|
| 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. $(k)$ and $(d)$ are the average degree and the average distance, respectively. $C$ and $r$ are the clustering coefficient$^1$ and assortative coefficient$^2$ 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. $n_\infty$ is the convergence time to coreness, defined as the minimum steps required to reach coreness from degree by the operator $H$.  

---

**NATURE COMMUNICATIONS** | DOI: 10.1038/ncomms10168 | www.nature.com/naturecommunications
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 $n_{\infty}$. The red circles, green squares and blue triangles represent the case of $\beta = 1.5\beta_i$, 2\$\beta_i$ and 2.5$\beta_i$, respectively. The dash lines in purple, black and cyan colours emphasize the $\tau$ values for degree (that is, $h^{(0)}$), H-index (that is, $h^{(1)}$) and coreness (that is, $h^{(6)}$), 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 model\(^4\) 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 $h_1^{(n)}, h_2^{(n)}, \ldots, h_{|V|}^{(n)}$ and the influences $R_1, R_2, \ldots, R_{|V|}$. To quantify to what extent the $h^{(n)}$-index resembles node influence values, we apply the Kendall Tau ($\tau$) coefficient\(^5\), which lies in the range $-1 \leq \tau \leq 1$. The larger value of $\tau$ means a stronger correlation between the two sequences (see Methods for the definition of $\tau$).

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 $\tau$ value). Thus, the sequence of H-indices, starting from degree and driven by the operator $H$, 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 model\(^6\) and bond percolation\(^7\). In the SIS model, the node influence index $R$ is defined as the probability that this node will remain infected...
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 centrality measures 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,\(^9\text{,}^{22}\) it is still a useful estimation of node importance, even without the specification of dynamics. In despite of its bias and disadvantages,\(^48\text{,}^{49}\text{,}^{50}\) 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 other words, for other values of $\beta$, the results are very similar and can be found in the Supplementary Table 1 ($\beta = 2b(t)$) and Supplementary Table 2 ($\beta = 2b(t)$). In each row, the largest $r$ is highlighted in bold.

### 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 | 0.8013  | 0.5049   | 0.7029    | 0.3834      |
| Sex      | 0.4525 | 0.4905  | 0.7513   | 0.6716    | 0.4851      |
| Facebook | 0.7173 | 0.7381  | 0.8274   | 0.7375    | 0.6859      |
| PB       | 0.8159 | 0.8221  | 0.7529   | 0.7453    | 0.5442      |
| USAir    | 0.7256 | 0.7540  | 0.2946   | 0.5975    | 0.3228      |
| Router   | 0.3309 | 0.2877  | 0.8021   | 0.7744    | 0.6195      |

SIR, susceptible-infected-removed.

They are degree (that is, $h_i^{(0)}$), H-index (that is, $h_i^{(1)}$), coreness (that is, $h_i^{(2)}$), closeness and betweenness. The spreading rate $\beta$ is set as $\beta = 1.5\beta_{\text{avg}}$ for the directed networks and $\beta = 1$ for the unweighted networks.

### 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 $n \leq m$, then we prove it to also be valid for $n = m + 1$. From the definition, $h_i^{(m+2)} = H(h_i^{(m+1)}, h_i^{(m+1)}, \ldots, h_i^{(m+1)})$, according to the induction hypothesis, $h_i^{(m+1)} \leq h_i^{(m)}$ for any node $j$. Therefore, $h_i^{(m+1)} = H(h_i^{(m+1)}, h_i^{(m+1)}, \ldots, h_i^{(m+1)}) \leq H(h_i^{(m)}, h_i^{(m)}, \ldots, h_i^{(m)}) = h_i^{(m)}$ for any node $j$. Namely, $h_i^{(m)} \leq h_i^{(m+1)}$. Therefore, $h_i^{(m+1)}$ is monotuously nonincreasing sequence, and each element is nonnegative, it has a nonnegative limit. Thus, we can define $h_i^{(\infty)}$ as the limit of the sequence $h_i^{(m)}, h_i^{(m+1)}, h_i^{(m+2)}, \ldots$.

We then introduce two simple relations. First, if $G'(V', E') \subseteq G(V, E)$, then from the definition of $\mathcal{H}$-function it is obvious that for any node $i \in V'$ and any integer $n \geq 0$, $h_i^{(n)}(G') \leq h_i^{(n)}(G)$, where the subscript $G'$ indicates that the corresponding index is defined on the subgraph $G'$. Second, if we denote $k_{\text{min}}$ the minimal degree of $G$, then for any node $i$ and any integer $n \geq 0$, $h_i^{(n)} \geq k_{\text{min}}$. This condition also can 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 $i$, $h_i^{(m+1)} = H(h_i^{(m)}, h_i^{(m)}, \ldots, h_i^{(m)})$, and according to the induction hypothesis, all elements in $(h_i^{(m)}, h_i^{(m)}, \ldots, h_i^{(m)})$ are no less than $k_{\text{min}}$ and the number of elements in $(h_i^{(m)}, h_i^{(m)}, \ldots, h_i^{(m)})$ is $k_i \geq k_{\text{min}}$. Therefore, according to the definition of $\mathcal{H}$-function, $h_i^{(m+1)} \geq k_{\text{min}}$.

If we denote $G'$ the $c$-core of $G$, it is clear that $G' \subseteq G$ and in $G'$, $k_{\text{min}} \geq c$. Therefore, $h_i^{(n)} \geq k_{\text{min}} \geq c$. We denote $G'(V', E')$ the induced subgraph containing all nodes $j$ such that $h_j^{(n)} \geq h_i^{(n)}$. Note that the node itself also belongs to $G'$. For any node $i \in V$, we find $h_i^{(n)} = H(h_i^{(n)}, h_i^{(n)}, \ldots, h_i^{(n)})$, where $j_1, j_2, \ldots, j_{k_i}$ are the $k_i$ neighbours of node $i$. For any node $j \in V'$, since in $G'$, $h_j^{(n)} \geq h_j^{(n)}$, there are at least $h_j^{(n)}$ neighbours of node $j$ with $h_j^{(n)}$ values no less than $h_j^{(n)}$. Thus, these neighbours also belong to $V'$. Therefore, in $G'$ the degrees of all the nodes are no less than $h_i^{(n)}$, that is, $G'$ is a subgraph of $G$'s $h_i^{(n)}$-core. Because $c$, the coreness of $i$, $h_i^{(n)} \leq c$. 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 $r > 0$ the node is selected, then $g_j^{(r)} = H(g_j^{(r-1)}, g_j^{(r-1)}, \ldots, g_j^{(r-1)})$. 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 \((i, j)\), \( g^{(1)}_{ij} \) 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 \( g^{(1)}_{ij} \geq g^{(1)}_{ij} \). Note that for any node \( j \) selected at time step \( t = 1 \) we have \( g^{(1)}_{ij} \geq g^{(1)}_{ij} \). We apply mathematical induction and assume that the above inequality holds when \( t_1 < t \) and \( t_2 < t \), and we next prove this also holds for \( t_2 < t + 1 \) and \( t_2 < t + n \). 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, \( g^{(1)}_{ij} = H(g^{(1)}_{ij}, \ldots, g^{(1)}_{ij}) \) and \( g^{(n+1)}_{ij} = H(g^{(n+1)}_{ij}, \ldots, g^{(n+1)}_{ij}) \). Note that for any \( m \) \((1 \leq m \leq n)\) we have \( \phi_{m,0} \leq \phi_{m,1} \leq n \). According to the induction hypothesis, \( g^{(1)}_{ij} \geq g^{(n+1)}_{ij} \), together with the definition of \( H \) function, we have \( g^{(1)}_{ij} \geq g^{(n+1)}_{ij} \). We denote the updating time steps of any node \( i \in V \) to be \( 0 = t_0 < t_1 < t_2 \ldots < t_n \). Then \( g^{(1)}_{ij}, g^{(n+1)}_{ij}, g^{(n+1)}_{ij} \) is a monotonously nonincreasing sequence and each element is nonnegative, and therefore it has a nonnegative limit. At this point, we can define \( g^{*}_{ij} \) as the limit of the sequence \( g^{(1)}_{ij}, g^{(n+1)}_{ij}, g^{(n+1)}_{ij} \).

We first prove that for any node \( j \in V \), \( g^{*}_{ij} \geq g^{*}_{ij} \). 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 \( t_0 \), that is, \( g^{(1)}_{ij} \geq c_i \) and before \( t_0 \) all nodes \( j \in V \) have \( g^{(1)}_{ij} \geq c_i \). Note that without a superscript indicates the last updated value before \( t_0 \). Therefore \( g^{(1)}_{ij} = H(g^{(1)}_{ij}, \ldots, g^{(1)}_{ij}) \geq H(g^{(n+1)}_{ij}, \ldots, g^{(n+1)}_{ij}) \). According to Theorem 1, \( H(c_i, c_i, \ldots, c_i) = \) const, namely \( g^{(1)}_{ij} \geq c_i \). This leads to a contradiction and thus the inequality \( g^{*}_{ij} \geq g^{*}_{ij} \) is validated.

Analogous to the proof of Theorem 1, after convergence, for any node \( i \in V \), all nodes \( j \) such that \( g^{*}_{ij} \geq g^{*}_{ij} \), including \( i \) itself, constitute an induced subgraph of \( G \)’s \( g^{*}_{ij} \)-core. Since \( c_i \) is the coreness of \( i \), \( g^{*}_{ij} \geq c_{ij} \). Combining the two inequalities, we arrive at Theorem 2.

### Spreading models

The standard SIR model, also referred to as the susceptible-infected-removed model, is usually applied to analyse the propagation of opinions or news.\(^{41}\) 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 \( \beta \). Then, each infected node enters the removed state with a probability \( \lambda \).

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 \( \beta \) values in our simulations, the infected percentage of the population is also small. When \( \beta \) 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, the epidemics threshold of SIR model is approximate to \( \beta_c \approx (k)/(\langle k^2 \rangle - \langle k \rangle) \). To be more precise, we determine the epidemic threshold \( \beta_c \) by simulation on real networks.\(^{25}\) 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 \( R_i \) when large \( \beta \) 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 the largest geodesic centrality of node \( i \), denoted by \( B(i) \), is defined as \( a^{27,35} \)

\[
B(i) = \sum_{j \neq i} \frac{g^{(1)}_{ij}}{g^{(1)}_{ij}} \leq \frac{1}{\sum_{j \neq i} d(i, j)},
\]

where \( g^{(1)}_{ij} \) is the number of shortest paths between nodes \( i \) and \( j \), and \( g^{(1)}_{ij} \) denotes the number of shortest paths between nodes \( i \) and \( j \) 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 \( \sum_{i} \).

\[
C(i) = \frac{1}{\sum_{j \neq i} d(i, j)},
\]

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.
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. Guimera, R., Danon, L., Diaz-Guilera, A., Giralt, 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. Batagelj, V. & Mrvar, A. Pajek Datasets. Available at http://vlado.fmf.uni-lj.si/pub/networks/datasets. 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. Egghé, L. Theory and practise of the g-index. Scientometrics 69, 131–152 (2006).
49. Jin, B., Liang, L., Rousseau, R. & Egghé, 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, 723–730 (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://www.npg.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).