

ARTICLE

Received 4 Feb 2011 | Accepted 16 Jun 2011 | Published 19 Jul 2011

DOI: 10.1038/ncomms1396

Ranking stability and super-stable nodes in complex networks

Gourab Ghoshal^{1,2,3} & Albert-László Barabási^{1,2}

Pagerank, a network-based diffusion algorithm, has emerged as the leading method to rank web content, ecological species and even scientists. Despite its wide use, it remains unknown how the structure of the network on which it operates affects its performance. Here we show that for random networks the ranking provided by pagerank is sensitive to perturbations in the network topology, making it unreliable for incomplete or noisy systems. In contrast, in scale-free networks we predict analytically the emergence of super-stable nodes whose ranking is exceptionally stable to perturbations. We calculate the dependence of the number of super-stable nodes on network characteristics and demonstrate their presence in real networks, in agreement with the analytical predictions. These results not only deepen our understanding of the interplay between network topology and dynamical processes but also have implications in all areas where ranking has a role, from science to marketing.

¹ Department of Physics, Biology and Computer Science, Center for Complex Network Research, Northeastern University, Boston, Massachusetts 02115, USA. ² Department of Medicine, Harvard Medical School, and Center for Cancer Systems Biology, Dana-Farber Cancer Institute, Boston, Massachusetts 02115, USA. ³ Media Laboratory, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139, USA. Correspondence and requests for materials should be addressed to A.-L.B. (email: alb@neu.edu or barabasi@gmail.com).

Originally introduced to rank web pages in the world wide web (www)¹, pagerank, a network-based diffusion algorithm, today is not only at the heart of Google and other search engines² but also the method of choice for ranking an extensive array of data in a wide range of network environments. It is used to rank physicists based on their citation patterns^{3,4}, disease-causing genes based on protein–protein interactions⁵, academic doctoral programs based on alumni placement⁶, roads or streets in terms of traffic⁷, ecological species based on their position in the food web⁸, highlight cancer genes in proteomic data⁹ and even to disambiguate words in lexical semantics¹⁰. The algorithm's popularity lies in both its perceived effectiveness and its easy to understand philosophy: rather than ranking objects based on difficult-to-measure intrinsic qualities, such as the utility of a webpage or the creativity of a researcher, it exploits the collective wisdom encoded in the network the object is part of, interpreting each link as an inherent vote.

Current advances in the statistical mechanics of complex networks^{11–18} have shown that the systems on which pagerank operates have significant differences in their network topology: some, such as the www, are scale-free^{19,20}, others, such as food webs, display a mixture of exponential and fat-tailed degree distributions²¹; the underlying networks have different sizes, average degree, path length, degree correlations^{22,23} and community decomposition^{24–26}. These topological differences are known to affect most network-based processes, from epidemic spreading^{27,28} to diffusion and network robustness^{29–34}. Yet the role of the underlying network structure in the effectiveness of pagerank remains unknown, prompting us to ask: could pagerank be inherently more accurate for some networks than for others? The key role ranking has from information retrieval to marketing makes this a question of major practical importance, affecting many aspects of our information society³⁵. Although the stability of pagerank to perturbations has been studied in computer science^{36–39}, we will show here that by focusing on the ranking stability of the top nodes we can obtain a series of fundamental results, that reshape our understanding of ranking stability. In particular, we find that, thanks to the fat-tailed nature of the degree distribution, a few super-stable nodes can emerge whose ranking becomes independent of what other nodes connect to them. We demonstrate the presence of such super-stable nodes in several real systems, from the www to citation networks.

Results

Pagerank algorithm and diffusion. The pagerank of a node in a network of N nodes with adjacency matrix A_{ij} can be calculated from

$$p_t(i) = (1 - \alpha) / N + \alpha \sum_j A_{ij} p_{t-1}(j) / k_{\text{out}}(j), \quad (1)$$

where $k_{\text{out}}(j)$ is the number of outgoing links from node j and α is a reset parameter¹. Equation (1) describes a diffusion process, where $p_t(i)$ is the frequency of visitation of node i by a particle at time step t that moves along the links of the network (encoded in the adjacency matrix A_{ij}) with probability α and jumps to a randomly chosen node with probability $1 - \alpha$. The stationary state of this diffusion process is the pagerank $p(i)$ of node i , determining its ranking relative to other nodes. In addition to its link to diffusion, mapping equation (1) to a Schrödinger-like wave equation helps elucidate the localization properties of the web graph^{40,41}.

The central hypothesis of the pagerank algorithm is that a link from a node i to node j serves as an ‘endorsement’ of node j by i . Moreover, the status of the recommending node is important—a letter of recommendation from a Nobel Laureate (that is, a node with high pagerank $p_{t-1}(j)$) carries far more weight than 10 letters from academics of lesser prominence. However, if the Laureate has

drafted a large number of recommendations for various candidates (has a high $k_{\text{out}}(j)$), then his (her) status as a recommender drops.

Pagerank typically operates on networks that are either mapped incompletely, such as the www⁴², or contain many false positives and negatives, such as protein interaction networks⁴³, raising a fundamental question: is the ranking of a node stable relative to other nodes in the face of such considerable network perturbations?

Ranking stability under degree-preserving perturbations. As random perturbations, from network incompleteness to noise, leave the relative degrees of the nodes largely unaltered, here we study the ranking stability under degree-preserving perturbations. This is achieved by randomly rewiring the network, while leaving the degree of each node (and hence the degree distribution $P(\mathbf{k})$) invariant. This approach is also motivated by the fact that the leading contribution to the pagerank of a node is its in-degree⁴⁴, therefore perturbations that randomly change a node's degree render the algorithm useless.

The ranking of a node with rank m is considered stable under network perturbations if changes in its pagerank p_m (where the subscript associates the pagerank to its rank, that is, the node with the highest pagerank has $m = 1$) leave the node's ranking m unchanged. Denoting with $\sigma(p_m)$ the fluctuations in p_m around its mean value $\langle p_m \rangle$ under different realizations of the degree-preserving perturbations, the m th ranked node has a stable rank if

$$\sigma(p_m) \leq \Delta(p_m), \quad (2)$$

where $\Delta(p_m) = p_m - p_{m+1}$. In other words, if the fluctuations in a node's pagerank p_m are small compared with the gap between its pagerank p_m and that of the node ranked below it p_{m+1} , the perturbation will not lower its ranking. Note, that $\Delta(p_m) / \sigma(p_m)$ is a monotonically decreasing function in m ; hence, if the gap exceeds the fluctuation for a specific rank m , then it will also exceed for all $m' < m$. To see whether the stability criteria equation (2) is ever satisfied, we calculated analytically the expected gap and fluctuations in the pagerank for networks with different degree distributions (Supplementary Methods, Supplementary Figs S1 and S2). We find that for a scale-free network ($P^{\text{SF}}(\mathbf{k}) \sim \mathbf{k}^{-\gamma}$)¹⁹, the gap follows $\Delta^{\text{SF}}(p_m) = \alpha N^{-(\gamma-2)/(\gamma-1)} \times f_{\Delta}^{\text{SF}}(m, \gamma)$, whereas for an exponential network ($P^{\text{EXP}}(\mathbf{k}) \sim e^{-\lambda \mathbf{k}}$)⁴⁵, we have $\Delta^{\text{EXP}}(p_m) = \alpha / N \times f_{\Delta}^{\text{EXP}}(m)$. The fluctuations $\sigma(p_m)$ in scale-free networks follow $\sigma^{\text{SF}}(p_m) = \alpha^2 N^{-(2\gamma-3)/2(\gamma-1)} \times g_{\sigma}^{\text{SF}}(m, \gamma)$, whereas in exponential networks $\sigma^{\text{EXP}}(p_m) = \alpha^2 / N \times g_{\sigma}^{\text{EXP}}(m, \lambda)$. Therefore, the stability ratio for the two networks is

$$\Delta^{\text{SF}}(p_m) / \sigma^{\text{SF}}(p_m) = N^{1/2(\gamma-1)} / \alpha \times F^{\text{SF}}(m, \gamma), \quad (3)$$

$$\Delta^{\text{EXP}}(p_m) / \sigma^{\text{EXP}}(p_m) = 1 / \alpha \times F^{\text{EXP}}(m, \lambda), \quad (4)$$

where the complete expressions for F^{SF} and F^{EXP} are provided in Supplementary Methods. Note that although the stability ratio equation (3) for scale-free networks depends on the system size, N , for exponential networks equation (4) is size invariant. The reason is that for an exponential distribution the top nodes have comparable degrees (Fig. 1a), whereas for a fat-tailed distribution (Fig. 1b) the degrees of the top nodes are well separated from each other. Indeed, the relative gap, $(k_{\text{max}} - k_{\text{max}-1}) / k_{\text{max}-1}$ between the two top-ranked nodes in exponential networks of size $N = 10^4$ is $\sim 10^{-2}$ (Fig. 1c), whereas for scale-free networks it is $10^0 - 10^1$, that is, two–three orders of magnitude larger (Fig. 1d). Consequently in an exponential network, the pagerank distribution of the top nodes are practically indistinguishable (Fig. 1e), indicating that the identity of the first-, second- or third-ranked node is different for each configuration. In contrast, the pagerank of the top node is well separated from the pagerank

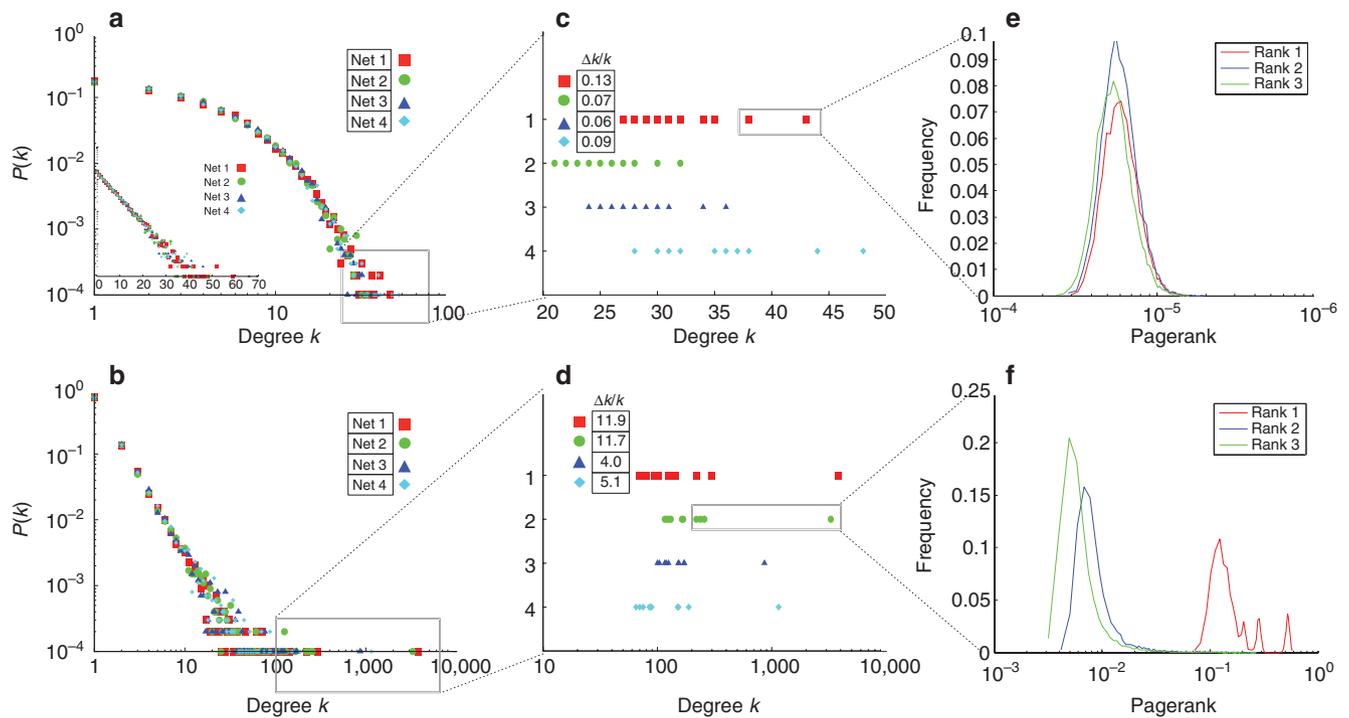


Figure 1 | The difference between exponential and scale-free distributions. The degree distributions of (a) four networks with exponential degree distribution $P(k) \sim e^{-\lambda k}$ with $\lambda = 0.2$ on a log-log scale (inset: log-linear) and (b) four networks with scale-free degree distribution $P(k) \sim k^{-\gamma}$ with $\gamma = 2.2$, average degree $\langle k \rangle = 5$ and size $N = 10^4$. (c, d) The magnified area of a and b showing the gap between the largest degrees. The difference between the largest degrees in a scale-free network is consistently large compared with the largest values in an exponential network. The legend shows the relative difference in the top two degrees' $\Delta k = (k_{\max} - k_{\max-1}) / k_{\max-1}$ for both degree distributions. (e, f) The pagerank distributions of the top three nodes selected from each network type. For the power law distribution, the peak of each curve can be clearly distinguished in contrast with the exponential network where the top nodes' pagerank are practically indistinguishable.

of the second- and third-ranked nodes in a scale-free network (Fig. 1f), indicating that the top-ranked node remains the same for each network configuration, being insensitive to perturbations.

Equation (4) predicts that for an exponential network the gap between consecutive pageranks never exceeds the fluctuations, making the ranking rather sensitive to perturbations. In contrast, according to equation (3), for certain (N , γ and α) combinations in a scale-free network the stability criteria equation (2) is satisfied, predicting the existence of a finite set of nodes whose ranking is stable to network perturbations. We will call these nodes super-stable, which means that by virtue of the many links they have, their ranking is independent of who points at them. Thus, degree-preserving perturbations do not alter their ranking, in contrast with the rest of the nodes in the network, whose ranking is sensitive to precisely which node points at them.

In Figure 2a,b, we show $\Delta(p_m)$ and $\sigma(p_m)$ for the top-ranked nodes in scale-free networks with sizes $N = 10^2$ and 10^4 . For $N = 10^2$, the fluctuations σ exceed the gap Δ for any rank m , indicating the absence of nodes whose rank is stable to perturbations. However, for $N = 10^4$ nodes with $m < m_c$, we have $\sigma(p_m) < \Delta(p_m)$, indicating that their rank is stable. In general, the stability ratio $\Delta(p_m) / \sigma(p_m)$ scales with system size as $N^{1/2(\gamma-1)}$ (a dependence absent in exponential networks), that is, the larger the scale-free network the more stable is the ranking of the top nodes. Therefore, it is easier to agree on the relative ranking of the top nodes in a large network than a small one, a rather counterintuitive result, given the cognitive limits that we face when we try to compare with each other a larger number of objects or services. The reason is that in larger systems the likelihood of the emergence of true outliers, whose pagerank is significantly greater than others, is greater.

Figure 2a,b suggests that the system size must exceed a critical size for super-stable nodes to emerge. Defining N_c as the minimum system size for which at least the top node's ranking is stable (that is, $\Delta(p_1) / \sigma(p_1) \geq 1$), we find that for scale-free networks with degree exponents in the range $2 \leq \gamma < 3$ we have $N_c = 0$, indicating that super-stable nodes emerge for any system size. For $\gamma \geq 3$, however, only networks whose size exceeds

$$N_c(\gamma) \approx (\alpha(\gamma-1) / (\gamma-3))^{1/2} 2^{(\gamma-1)} \quad (5)$$

can have super-stable nodes (Supplementary Methods). Therefore, $\gamma_c = 3$ represents a critical exponent for ranking stability, as illustrated by the (N , γ) phase diagram of Figure 2c: for $\gamma < \gamma_c = 3$, we always have at least one super-stable node, whereas for $\gamma > \gamma_c$ only for $N > N_c(\gamma)$ can super-stability emerge.

We also find that the number of super-stable nodes m_c scales as $m_c \sim N^{1/(2(\gamma-1))}$, which is a rather weak dependence—for $\gamma = 3$, to increase m_c by a factor of ten, one needs to increase the system size by five orders of magnitude. For large N and $2 < \gamma < 3$, the critical rank m_c depends on γ as $e^{\alpha(\gamma-2)/(2(\gamma-1))}$ and for $\gamma \gg 3$, it decays as $\gamma^{-\gamma}$. The resulting γ dependence is summarized in Figure 2d, indicating that the number of stable ranks is relatively small for all γ and that it peaks in the vicinity of $\gamma_c = 3$. The peak becomes increasingly pronounced for large N .

At the first glance, the peak at $\gamma_c = 3$ is unexpected: an increasing γ should decrease the gap as $\Delta(p_m) \sim N^{1/2(\gamma-1)}$. Note, however, that for $\gamma < 3$ the fluctuations $\sigma(p_m)$ diverge as $\sigma \sim N^{(3-\gamma)/(\gamma-1)}$, whereas $\sigma(p_m)$ is asymptotically size independent for $\gamma > 3$. Hence for $\gamma < 3$, the gap is large but so are the fluctuations, whereas for $\gamma > 3$ the gap decreases and the fluctuations are effectively constant. The best payoff between

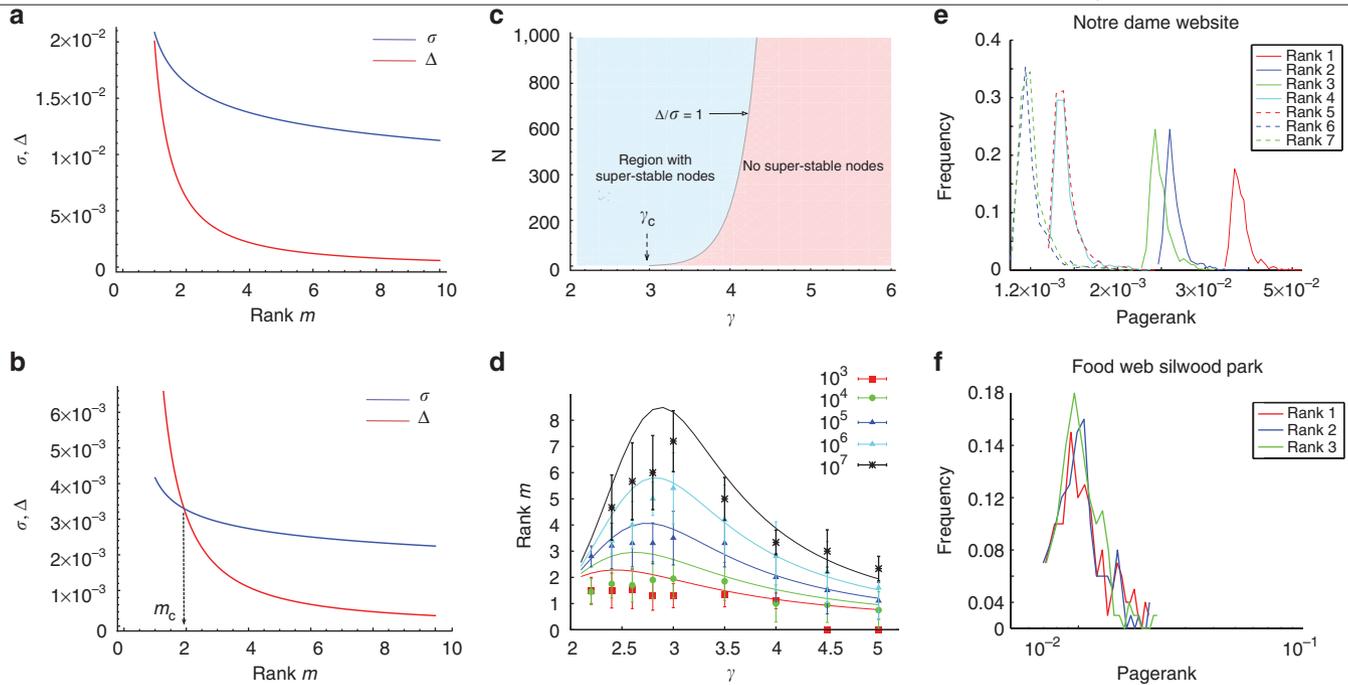


Figure 2 | Analytical and numerical evidence for super-stable nodes. (a, b) Pagerank gap $\Delta(p_m)$ and the fluctuation $\sigma(p_m)$ plotted as function of the rank m for two scale-free networks with the same exponent, $\gamma = 3.5$, and different system sizes, $N = 10^2$ and 10^4 , showing the emergence of super-stability for larger N . (c) The (N, γ) phase diagram showing the stable and unstable regimes that emerge for $\gamma > \gamma_c = 3$. (d) The critical rank m_c as a function of γ where the curves correspond to the calculated m_c and the points represent the measured m_c in numerical simulations on multiple networks. The peak near $\gamma_c = 3$ is more pronounced as N increases, with different scaling behaviors above and below γ_c . (e, f) Pagerank distributions of the top-ranked nodes in the www (scale-free) and the food web (exponential) shown in Table 1. The separated curves correspond to m_c (measured) values shown in the table. Note that in the www the top-ranked nodes are clearly identifiable, whereas in the food web there are no super-stable nodes as predicted by the calculation.

Table 1 | Super-stable nodes in real networks.

Network	N	γ	N_c	m_c predicted	m_c measured
<i>Scale-free</i>					
Google web sample	875,713	2.5	0	4	1
Notre Dame web sample	325,729	2.1	0	3	3
Stanford web sample	281,903	2.1	0	2	2
Berkeley–Stanford web sample	685,230	2.1	0	3	3
Hep-th citations	27,770	2.8	0	3	2
Wikipedia admin. voting	7,115	3.3	50	2	1
Amazon co-purchase	223,431	4.3	2,730	2	2
Mobile call graph	4,562,263	5.2	150,000	2	3
<i>Exponential</i>					
<i>C. elegans</i> neural net	307	—	∞	0	1
Food Web (Little Rock Lake)	183	—	∞	0	0
Food Web (Silwood Park)	154	—	∞	0	0

The second to fifth columns give the number of vertices in the network, N , the best estimate of the degree exponent of the degree distribution, γ , ‘—’ when the network has an exponential distribution, the predicted critical system size, N_c , the predicted critical rank, m_c , and the measured m_c , provided by the perturbation scheme described in the text. The networks are, in order, a snapshot of a portion of the Google web, the Notre Dame, Stanford and combined Berkeley–Stanford web sites, the citation network of papers found in the high-energy physics arXiv, the network of co-purchases from Amazon, the voting records for administrators in Wikipedia, the neural network of *C. elegans* and finally the Food Webs from Little Rock Lake and Silwood Park. The sources and reference of each data set are provided in Supplementary information.

these two regimes is in the vicinity of $\gamma_c = 3$, where $\sigma^2 \sim (1/2) \log(N\pi)$, resulting in the peak at γ_c (Supplementary Methods).

To test our analytical predictions, we generated networks with fixed $P(\mathbf{k})$ using the configuration model⁴⁶, and then ranked each node according to their pagerank. We perturbed the network by rewiring every edge (keeping the degree of each node unchanged) and determined the pagerank after each rewiring, helping us identify nodes whose ranking did not change as a result of the perturbation. We not only found that such nodes exist in the predicted topological regimes, but also plot the measured value of the critical rank m_c as a function of the exponent γ for various system sizes in Figure 2d, confirming the predicted trend: for large N , the m_c versus γ curves

develop a peak in the vicinity of γ_c . Most importantly, the analytical and the numerical results agree that the number of super-stable nodes is rather small—less than ten for a system with ten million nodes.

Evidence of super-stable nodes in real networks. To see whether super-stable nodes emerge in real systems, we collected data for a variety of real networks, ranging from samples of the www to citation networks and identified the nodes whose ranking do not change under rewiring perturbations. For each network with a fat-tailed degree distribution, we observed a few super-stable nodes whose number closely agrees with the analytical prediction (Table 1). For networks with an exponential degree distribution, the data support

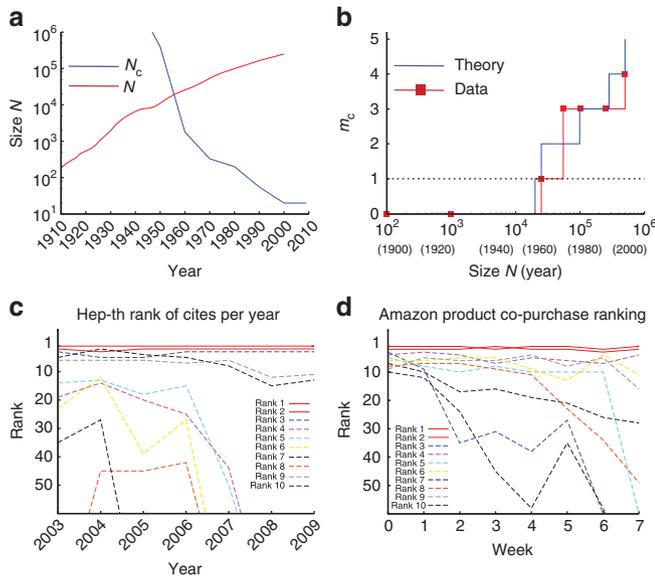


Figure 3 | Temporal stability of super-stable nodes. (a) The number of papers published in the Physical Review corpus with time and the predicted critical system size N_c . (b) The evolution of the number of super-stable nodes in the same network, m_c , as a function of size, N , and time (shown in parentheses). The squares correspond to the measured m_c from the Physical Review citation network, whereas the curve corresponds to the analytical prediction using the appropriate N and the best estimate of γ for each decade. (c) The number of citations per year for the top ten ranked papers, where ranking is provided by applying pagerank to all high-energy physics papers published by 2002. The two predicted super-stable nodes (solid red lines) continue to be the top cited papers for almost a decade. The rest of the papers display considerable fluctuations in rank, even dropping out of the top 50 lists. (d) The evolution of the ranks of the top ten most co-purchased books in Amazon. The curves represent the evolution of books, ranked in the top ten in our first sample, over the subsequent weeks. Although the ranking of most nodes fluctuate, the top two nodes—predicted to be super-stable (solid red lines)—maintain their top ranking.

our prediction that super-stable nodes are absent. The only exception is the neural network of *Caenorhabditis elegans*, which has one super-stable node, because of the fact that its in-degree is separated by an order of magnitude from the rest of the nodes. Note that the probability to have such a high in-degree in this network is $\sim 10^{-9}$, indicating that this node, a motor neuron responsible for locomotion, represents a clear deviation from the expected degree distribution. The different behavior of the pagerank for the two network classes is illustrated in Figure 2e,f, where we show the pagerank distributions for the top nodes in two networks, the www (scale-free) and the food web (exponential). In line with our predictions, for the www, the top nodes are clearly separated from the rest of the nodes, whereas for the food web the pagerank distributions of the top nodes are indistinguishable.

Probably, the strongest direct evidence supporting our predictions comes from the Physical Review citation network. We used the publication history of papers published in the Physical Review journals from 1893 to 2009, allowing us not only to identify the super-stable nodes from a static snapshot of the citation network, but also to track the emergence of super-stability in time. Two systematic changes in the network impact the number of super-stable nodes: (i) The network grows, increasing from $N=0$ in 1892 to $N=449,673$ in 2009 (Fig. 3a). (ii) The degree exponent decreases from $\gamma \approx 5$ in the 1950s to $\gamma \approx 3$ today (Supplementary Fig. S3). We therefore predicted the number of super-stable nodes in each decade between 1900 and 2000, by incorporating the changes in N and γ , and also identified m_c

directly from the real data. We find that super-stable papers do not emerge before the 1950s, as the combination of high-degree exponent and small N prohibits super-stability (Fig. 3b). However, as the degree exponent γ drops and N increases, between 1950 and 1960, N overcomes N_c , allowing for the emergence of the first super-stable paper ($m_c = 1$). In the subsequent decades, m_c gradually increases to four super-stable papers. As Figure 3b shows, the numerically identified m_c closely follows the analytical predictions, the difference being at most one super-stable paper in a decade. Hence, the data set not only confirms the existence of super-stable publications in the Physical Review corpus (for the list of super-stable papers, see Supplementary Table S1), but also shows that their emergence in time follows closely the analytical predictions.

Our ability to identify super-stable nodes from a single snapshot of the network raises an important question: how stable is the ranking with time? To answer this question, we collected time-resolved ranking data for the citation and co-purchasing networks (Supplementary Table S2), allowing us to quantify the temporal stability of the top nodes. In the high-energy citation network, the super-stable nodes were identified from a sample containing all papers published in 2002. We find, however, that in the subsequent 7 years these two papers maintain their top ranking, collecting the most citations each year. In contrast, the ranking of the rest of the papers, which do not demonstrate super-stability in the 2002 sample, fluctuates widely (Fig. 3c). Similarly, for the Amazon co-purchasing network, the two super-stable books continue to maintain their rank in samples collected on a weekly basis (Fig. 3d) and were the top ranked in a sample collected 6 years earlier (2005) as well. Additionally, in the Physical Review Corpus (Supplementary Fig. S4), most super-stable nodes maintain their status for a period of 6–10 years; some, such as the 1957 paper on the BCS theory of superconductivity, show super-stability for three decades. Taken together, we find that super-stable nodes, identified from a single snapshot of the network, show a remarkable temporal stability, a feature not shared by other nodes in the system.

Discussion

In summary, we find that real networks with heavy-tailed degree distributions naturally lead to a set of super-stable nodes that have such a high number of ‘recommendations’ (in-degree) that their ranking becomes independent of who recommends them. This is somewhat unexpected from the perspective of the network architecture: the scale-free nature of these networks normally implies a lack of objective criteria to distinguish hubs from non-hubs. The balance of rank stability and fluctuations do allow us, however, to identify a few hubs that respond in a distinct manner to perturbations.

Both our analytical predictions and numerical results indicate that the number of super-stable nodes is very small. As predicted by our scaling analysis, this number is largely unaffected by most network characteristics and only a significant increase in system size can increase their number. This suggests that across a large number of systems a small number of components (nodes) are bound to have a disproportionate role in the system. These nodes are often easy to identify: a simple link counting should place them at the top, limiting the usefulness of pagerank to rank nodes that are not super-stable.

It is often mentioned that the early success of Google compared with its competitors was not because of better coverage (which back then was inferior to that of the market leader Inktomi), but its pagerank algorithm, that offered a superior user experience through a better ranking of the relevant documents. Our results suggest that the success of pagerank was the inadvertent consequence of the scale-free nature of the web graph. Had the web been an exponential network, the ranking provided by pagerank would have been unreliable given the incompleteness of the web graph. Indeed, in 1999, Google indexed only 7.8% of the web⁴² and even today its coverage is less than half of the indexable web. Yet, the scale-free property of the web graph leads to the emergence of a small number of super-stable nodes, for which a simple count of the

in-degree offers the correct relative ranking. As the www grew, the ranking stability at the top increased, making the top-ranked nodes even easier to identify. Therefore, counterintuitively, we find that the growth of the web, instead of making search more difficult by offering more hits, helps select clear winners, offering better ranking clarity at the top.

Methods

Determining the gap and fluctuation in pagerank. Given a probability distribution $p(x)$, we can determine the expectation value of the largest x after we draw N numbers from $p(x)$. If we draw N numbers from a particular sample and one of them, x_j , lies between the interval $x + dx$, the probability that there are no other numbers with a greater value than x_j is given by $p(x)dx \times [1 - P(x)]^{N-1}$, where $P(x)$ is the cumulative distribution. As there are N ways of choosing x_j , the total probability is $\pi(x) = Np(x)[1 - P(x)]^{N-1}$. Similarly, we can determine the expectation value of the m th largest number $\langle x \rangle_m$. By definition, the m th ranked number has $m - 1$ numbers above it and $N - m$ below it, obtaining

$$\pi_m(x) = \frac{p(x)P(x)^{m-1}[1 - P(x)]^{N-m}}{B(N - m + 1, m)}, \quad (6)$$

where the denominator is the beta function. The expectation value is determined by,

$$\langle x \rangle_m = \int_0^\infty x \pi_m(x) dx. \quad (7)$$

Combining this with equation S1 (Supplementary Methods) gives the expectation value for the pagerank p_m of a node ranked m . The gap between the pagerank of a node ranked m and the node ranked one place below it p_{m+1} is $\Delta(p_m) = p_m - p_{m+1}$, whereas the fluctuation $\sigma(p_m)$ is determined by substituting p_m into equation S2. The details of the calculation are listed in Supplementary Methods.

Determination of critical values. According to the stability criteria equation (2), setting the ratio $\Delta(p_m)/\sigma(p_m)$ equal to one allows us to define a critical value for each relevant parameter, such that above that value we are in the stable regime and below in the unstable regime. We focus on two parameters: the critical system size $N = N_c$, which specifies the minimum system size for which any stable ranks exist, and $m = m_c$, which denotes the maximum rank in the network that is stable for system size $N > N_c$.

To find N_c in a scale-free network, we note that the maximum value of equation (3) as a function of m is at $m = 1$. Furthermore, the ratio is a monotonically decreasing function in m ; hence, if there is a critical value for the system size N_c , then it must at least hold for $m = 1$ and thus setting the ratio and m equal to one, we can derive an equation for N_c . The equation can only be solved numerically, but the scaling behavior of N_c can be extracted through a series of approximations (Supplementary Methods). Similarly, the critical rank m_c is derived by setting the ratio to one and N to a fixed value $N > N_c$.

References

- Brin, S. & Page, L. The anatomy of a large-scale hypertextual web search engine. *Comput. Netw. ISDN Syst.* **30**, 107–117 (1998).
- Langville, A. N. & Meyer, C. D. *Google's Pagerank and Beyond: THE Science of Search Engine Rankings* (Princeton University Press, 2006).
- Chen, P., Xie, H., Maslov, S. & Redner, S. Finding scientific gems with Google. *J. Informetrics* **1**, 8–15 (2007).
- Walker, D., Xie, H., Yan, K.-K. & Maslov, S. Ranking scientific publications using a simple model of network traffic. *J. Stat. Mech.* **6**, P06010 (2007).
- Chen, J., Aronow, B. J. & Jegga, A. G. Disease candidate gene identification and prioritization using protein interaction networks. *BMC Bioinformatics* **10**, 73 (2009).
- Schmidt, B. M. & Chingos, M. M. Ranking doctoral programs by placement: a new method. *Polit Sci Polit* **40**, 523–529 (2007).
- Jiang, B., Zhao, S. & Yin, J. Self-organized natural roads for predicting traffic flow: a sensitivity study. *J. Stat. Mech.* **7** (2008).
- Allesina, S. & Pascual, M. Googling food webs: can an eigenvector measure species' importance for coextinctions? *PLoS Comput. Biol.* **5**, e1000494 (2009).
- Ivan, G. & Grolmusz, V. When the web meets the cell: using pagerank for analyzing protein interaction networks. *Bioinformatics* **27**, 405–407 (2011).
- Navigli, R. & Lapata, M. An experimental study of graph connectivity for unsupervised word sense disambiguation. *IEEE Trans. Pattern Anal. Mach. Intell.* **32**, 4 (2010).
- Albert, R. & Barabási, A.-L. Statistical mechanics of complex networks. *Rev. Mod. Phys.* **74**, 47–97 (2002).
- Pastor-Satorras, R. & Vespignani, A. *Evolution and Structure of the Internet* (Cambridge University Press, 2004).
- Newman, M. E. J., Barabási, A.-L. & Watts, D. J. *The Structure and Dynamics of Networks* (Princeton University Press, 2006).
- Dorogovtsev, S. N., Goltsev, A. V. & Mendes, J. F. F. Critical phenomena in complex networks. *Rev. Mod. Phys.* **80**, 1275–1335 (2008).
- Cohen, R. & Havlin, S. *Complex Networks: Structure, Robustness and Function* (Cambridge University Press, 2010).
- Newman, M. E. J. *Networks: An Introduction* (Oxford University Press, 2010).
- Barrat, A., Boccaletti, S., Caldarelli, G., Chessa, A., Latora, V. & Motter, A. E. Complex networks: from biology to information technology. *J. Phys. A: Math. Theor.* **41**, 220301 (2008).
- Caldarelli, G. *Scale Free Networks* (Oxford University Press, 2007).
- Barabási, A.-L. & Albert, R. Emergence of scaling in random networks. *Science* **286**, 509–512 (1999).
- Albert, R., Jeong, H. & Barabási, A.-L. Diameter of the world wide web. *Nature* **401**, 130–131 (1999).
- Montoya, J. M. & Solé, R. V. Small world patterns in food webs. *J. Theor. Bio.* **214**, 405–412 (2002).
- Pastor-Satorras, A., Vázquez, A. & Vespignani, A. Dynamical and correlation properties of the internet. *Phys. Rev. Lett.* **87**, 258701 (2001).
- Newman, M. E. J. Assortative mixing in networks. *Phys. Rev. Lett.* **89**, 208701 (2002).
- Girvan, M. & Newman, M. E. J. Community structure in social and biological networks. *Proc. Natl Acad. Sci. USA* **99**, 7821–7826 (2002).
- Palla, G., Derényi, I., Farkas, I. & Vicsek, T. Uncovering the overlapping community structure of complex networks in nature and society. *Nature* **435**, 814–818 (2005).
- Ahn, Y.-Y., Bagrow, J. P. & Lehmann, S. Link communities reveal multiscale complexity in networks. *Nature* **466**, 761–764 (2010).
- Colizza, V. & Vespignani, A. Reaction-diffusion processes and metapopulation models in heterogeneous networks. *Phys. Rev. Lett.* **99**, 148701 (2007).
- Kitsak, M. *et al.* Identification of influential spreaders in complex networks. *Nat. Phys.* **6**, 888–893 (2010).
- Callaway, D. S., Newman, M. E. J., Strogatz, S. H. & Watts, D. J. Network robustness and fragility: percolation on random graphs. *Phys. Rev. Lett.* **85**, 5468 (2000).
- Cohen, R., Erez, K., ben-Avraham, D. & Havlin, S. Resilience of the internet to random breakdowns. *Phys. Rev. Lett.* **85**, 4626 (2000).
- Albert, R., Jeong, H. & Barabási, A.-L. Error and attack tolerance of complex networks. *Nature* **406**, 378–382 (2000).
- Cohen, R., Erez, K., ben-Avraham, D. & Havlin, S. Breakdown of the internet under intentional attack. *Phys. Rev. Lett.* **86**, 3682 (2001).
- Pastor-Satorras, R. & Vespignani, A. Epidemic spreading in scale-free networks. *Phys. Rev. Lett.* **86**, 3200 (2001).
- Motter, A. E. Cascade control and defense in complex networks. *Phys. Rev. Lett.* **93**, 098701 (2004).
- Colizza, V., Flammini, A., Serrano, M. A. & Vespignani, A. Detecting rich-club ordering in complex networks. *Nat. Phys.* **2**, 110–115 (2006).
- Ng, A. Y., Zheng, A. X. & Jordan, M. I. Stable algorithms for link analysis. *Proceedings of the 24th International Conference on Research and Development in Information Retrieval (SIGIR)* 258–266 (ACM, New York, NY, 2001).
- Chien, S., Dwork, C., Kumar, R., Simon, D. R. & Sivakumar, D. Link evolution: analysis and algorithms. *Internet. Math.* **1**, 277–304 (2004).
- Lempel, R. & Moran, S. Rank-stability and rank-similarity of link-based web ranking algorithms in authority-connected graphs. *J. Inf. Retr.* **8**, 245–264 (2005).
- de Kerchove, C., Ninove, L. & van Dooren, P. Maximizing pagerank via outlinks. *Linear Algebra Appl.* **429**, 1254–1276 (2008).
- Perra, N., Zlatić, V., Chessa, A., Conti, C., Donato, D. & Caldarelli, G. Pagerank equation and localization in the www. *Europhys. Lett.* **88**, 48002 (2009).
- Giraud, O., Georgeot, B. & Shepelyansky, D. L. Delocalization transition for the Google matrix. *Phys. Rev. E* **80**, 026107 (2009).
- Lawrence, S. & Giles, C. L. Accessibility of information on the web. *Nature* **400**, 107–109 (1999).
- Venkatesan, K. *et al.* An empirical framework for binary interactome mapping. *Nat. Methods* **6**, 83–90 (2009).
- Fortunato, S., Boguñá, M., Flammini, A. & Menczer, F. On local estimations of pagerank: a mean field approach. *Internet Math.* **4**, 245–266 (2009).
- Erdős, P. & Rényi, A. The evolution of random graphs. *Publ. Math. Inst. Hung. Acad. Sci.* **5**, 17–61 (1960).
- Newman, M. E. J., Strogatz, S. H. & Watts, D. J. Random graphs with arbitrary degree distributions and their applications. *Phys. Rev. E* **64**, 041902 (2001).

Acknowledgements

We thank G. Bianconi and J.P. Bagrow for useful discussions. This work was supported by the Network Science Collaborative Technology Alliance sponsored by the US Army Research Laboratory under Agreement Number W911NF-09-2-0053; the Office of Naval Research under Agreement Number N000141010968; the Defense Threat Reduction Agency awards WMD BRBAA07-J-2-0035 and BRBAA08-Per4-C-2-0033; and the James S. McDonnell Foundation 21st Century Initiative in Studying Complex Systems.

Author contributions

G.G. and A.-L.B. designed and performed the research and wrote 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: Ghoshal, G. & Barabási, A.-L. Ranking stability and super-stable nodes in complex networks. *Nat. Commun.* 2:394 doi: 10.1038/ncomms1396 (2011).