Ranking Stability and Super-Stable Nodes in Complex Networks

Gourab Ghoshal and Albert-László Barabási

Supplementary Information

Contents

Supplementary Figures **S-2**

Supplementary Figure S1 S-2

Supplementary Figure S2 S-2

Supplementary Figure S3 S-3

Supplementary Figure S4 S-4

Supplementary Figure S5 S-5

Supplementary Tables **S-6**

Supplementary Table S1 S-6

Supplementary Table S2 S-7

Supplementary Methods **S-8**

Pagerank and degree S-8

Analysis of the largest numbers and the gap distribution S-8

Stability condition S-12

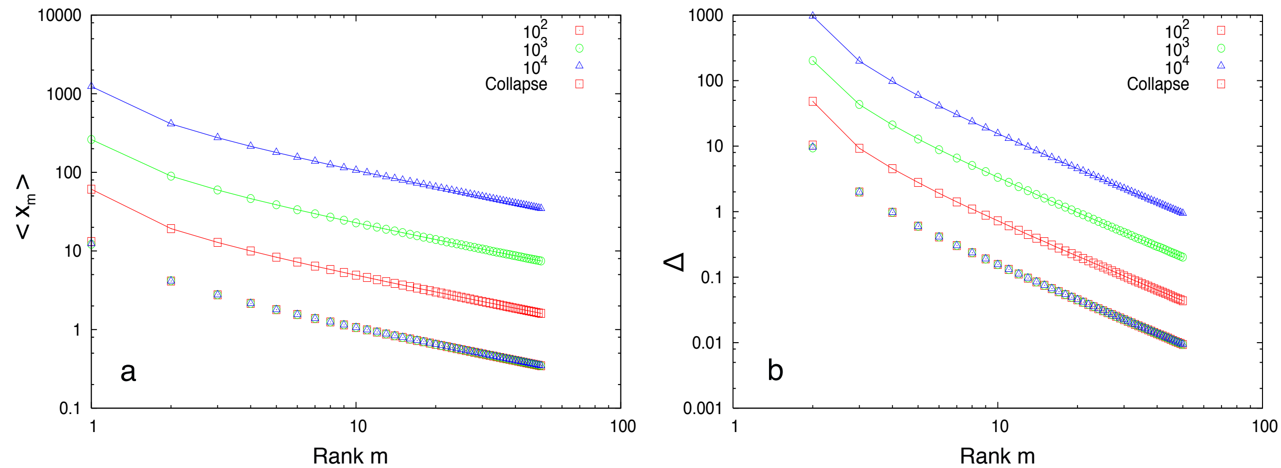
Critical values S-14

Data and empirical analysis S-16

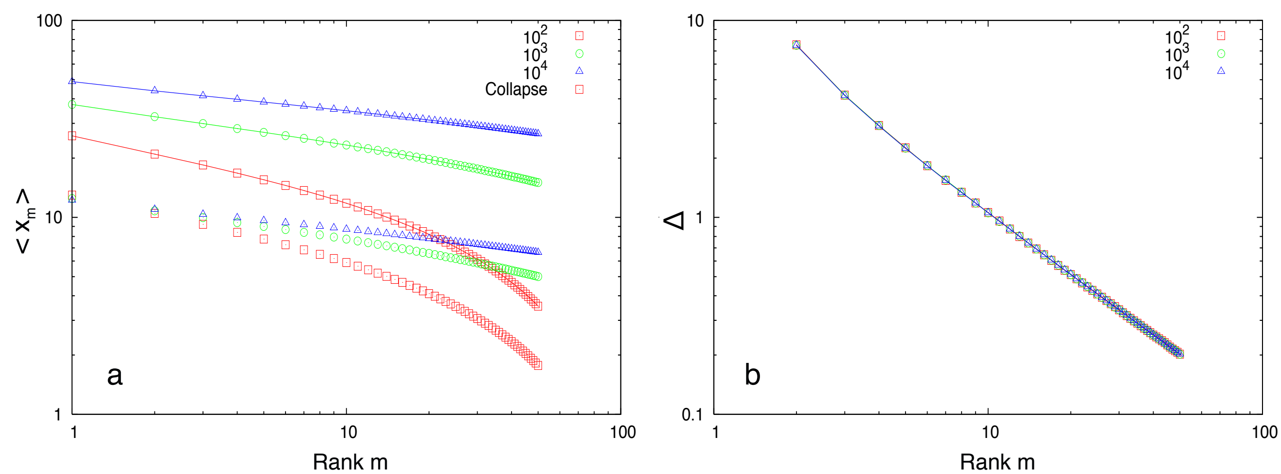
Effect of degree correlation on ranking S-18

Relation to previous results on pagerank stability S-19

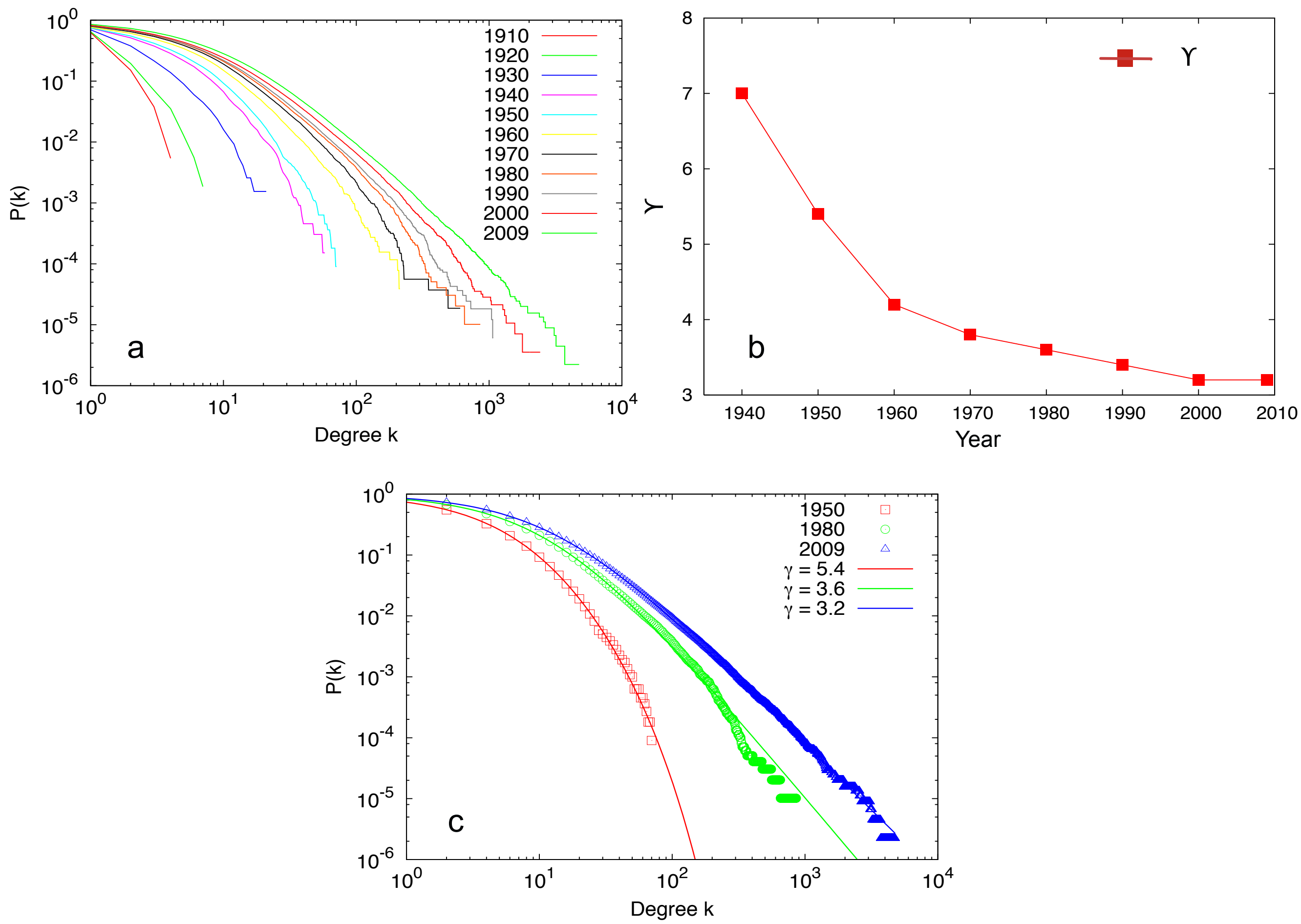
Supplementary References **S-22**



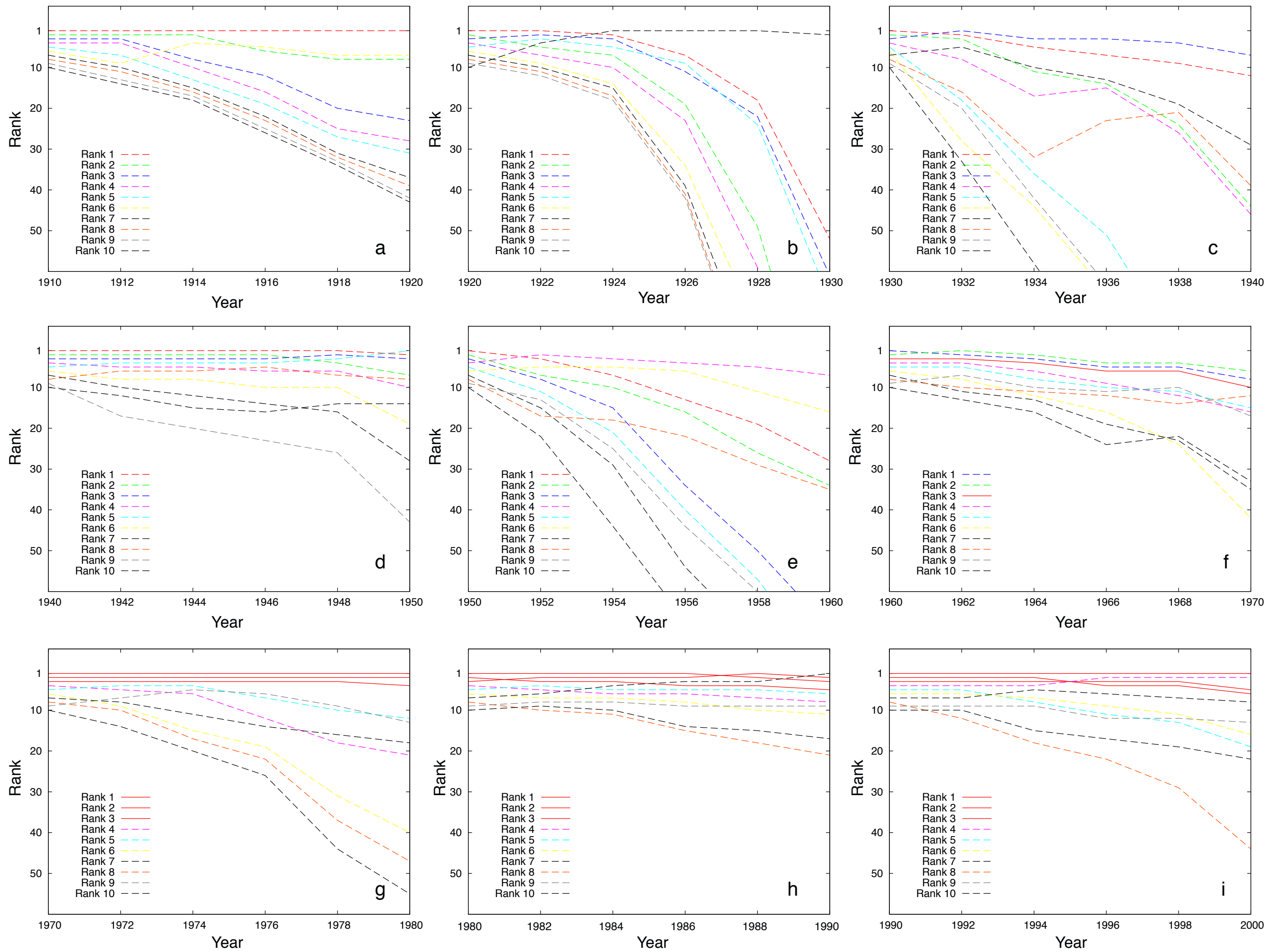
Supplementary Figure S1: **(a)** The expected values of the m largest numbers in a power-law distribution with $\gamma = 2.5$ (log-log scale) as per equation (S20). The three curves correspond to different sample sizes $N = 10^2, 10^3, 10^4$ and the top 50 values are shown. Points correspond to simulations, while the straight lines correspond to the analytical solution. The scaling with system size is clearly visible. The bottom curve corresponds to the collapse of the three curves after re-scaling, i.e after dividing out by the factor $N^{1/(\gamma-1)}$. **(b)** The same plot for the gap $\Delta = \langle x_m \rangle - \langle x_{m+1} \rangle$.



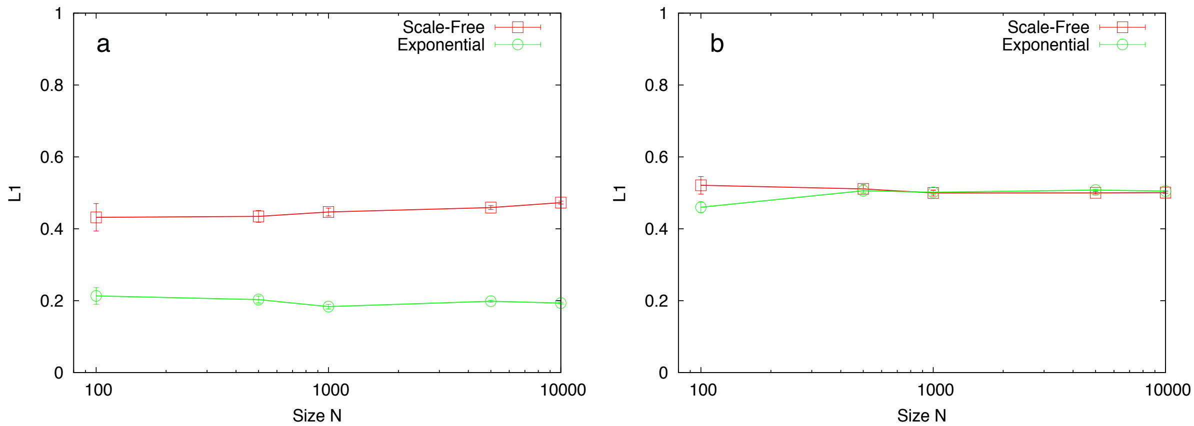
Supplementary Figure S2: **(a)**The expected values of the m largest numbers in an exponential distribution with $\lambda = 0.2$ (log-log scale) as per equation (S23). The three curves correspond to different sample sizes $N = 10^2, 10^3, 10^4$ and the top 50 values are shown. Points correspond to simulations, while the straight lines correspond to the analytical solution. The scaling with system size is visible, although there is a clear approach to a limiting value. The bottom curve corresponds to the collapse of the three curves after re-scaling. **(b)** The same plot for the gap $\Delta = \langle x_m \rangle - \langle x_{m+1} \rangle$. As is clearly seen the size of the gaps is completely independent of system size.



Supplementary Figure S3: **(a)** The evolution of the cumulative distribution function of the citation network from 1910-2010. The network develops a power-law distribution in the tail starting around 1940. **(b)** The change in the degree exponent γ with time. The figure shows the best estimate to γ after fitting the distribution to equation (S52). The exponent decreases in each decade until it stabilizes in 2000 to a value of $\gamma = 3.2$. Note that the degree exponent in each decade corresponds to the network containing all publications published from 1893 till that particular year. **(c)** The cumulative distribution function of the degree in different decades compared to the fit equation (S52).



Supplementary Figure S4: Temporal evolution of the top ten papers in each decade. **(a)** 1910-1920, **(b)** 1920-1930, **(c)** 1930-1940, **(d)** 1940-1950, **(e)** 1950-1960, **(f)** 1960-1970, **(g)** 1970-1980, **(h)** 1980-1990, **(i)** 1990-2000. The super-stable nodes are marked by solid red lines, whereas the rest are marked by dashed lines. Note the relative volatility in ranking before 1960 (when the first super-stable node emerges). Apart from the very top-ranked node in some of the decades preceding 1960, the rest of the ranking changes rapidly. After 1960, the super-stable nodes maintain their ranking, typically for 6-10 years.



Supplementary Figure S5: **(a)** The $L1$ distance between pagerank scores of the original and perturbed networks as a function of system size N. The red squares correspond to a scale-free network and the green circles correspond to an exponential network, both with average degree $\langle k \rangle = 1.5$. The points are averaged over N realizations for each N. There is no evident scaling with N in either network. **(b)** The $L1$ distance between the ranked lists in the same networks, where the ranks have been converted to a probability distribution using equation (S60). There is no scaling with size here as well, in addition the two networks are virtually indistinguishable.

Time	Size N cumulative	m_c predicted	m_c measured	Super-Stable Papers
1890-1910	185	0	0	---
1910-1920	539	0	0	---
1920-1930	1,941	0	0	---
1930-1940	6,591	0	0	---
1940-1950	11,066	1	0	---
1950-1960	25,677	1	1	1) M. Goldhaber and A.W. Sunyar, *Classification of Nuclear Isomers*, Phys. Rev. **83**, 906 (1951).
1960-1970	53,711	3	3	1) J. Bardeen, L.N. Cooper and J.R. Schrieffer, *Theory of Superconductivity*, Phys. Rev. **108**, 1175 (1957). 2) M. Gell-Mann, *Symmetries of Baryons and Mesons*, Phys. Rev **125**, 1067 (1962). 3) R.P. Feynman and M. Gell-Mann, *Theory of the Fermi Interaction*, Phys. Rev. **109**, 193 (1958).
1970-1980	98,622	3	3	1) J. Bardeen, L.N. Cooper and J.R. Schrieffer, *Theory of Superconductivity*, Phys. Rev. **108**, 1175 (1957). 2) M. Gell-Mann, *Symmetries of Baryons and Mesons*, Phys. Rev **125**, 1067 (1962). 3) S. Weinberg, *A model of Leptions*, Phys. Rev. Lett. **19**, 1264 (1967).
1980-1990	164,458	3	3	1) W. Kohn and L.J. Sham, *Self-Consistent Equations*, Phys. Rev. **136,** A1133 (1965). 2) S. Weinberg, *A model of Leptions*, Phys. Rev. Lett. **19**, 1264 (1967). 3) J. Bardeen, L.N. Cooper and J.R. Schrieffer, *Theory of Superconducitvity*, Phys. Rev. **108**, 1175 (1957).
1990-2000	282,384	4	3	1) W. Kohn and L.J. Sham, *Self-Consistent Equations*, Phys. Rev. **136,** A1133 (1965). 2) P. Hohenberg and W. Kohn, *Inhomogenous Electron Gas,* Phys. Rev. **136**, B864 (1964). 3) J.P. Perdew, *Self-interaction correction to density-functional approximation*, Phys. Rev. B **23**, 5048 (1981).
2000-2009	449,673	5	4	1) W. Kohn and L.J. Sham, *Self-Consistent Equations*, Phys. Rev. **136,** A1133 (1965). 2) P. Hohenberg and W. Kohn, *Inhomogenous Electron Gas,* Phys. Rev. **136**, B864 (1964). 3) J.P. Perdew, *Self-interaction correction to density-functional approximation*, Phys. Rev. B **23**, 5048 (1981). 4) J.P. Perdew, K. Burke and M. Erzerhof, *Generalized gradient approximation*, Phys. Rev. Lett. **77**, 3865 (1996).

Supplementary Table S1: Details of the Physical Review citation network as it evolves in time. In order from left to right, the columns show the decade the network corresponds to, the cumulative size of the network, the predicted number of super-stable nodes, the measured value of m_c and finally the references to the super-stable nodes (publications) when applicable.

	Date collected	Starting Book	Size N	γ	m_c
Sample 1	12/13/2010	Decision Points	$223,431$	4.3	2
Sample 2	12/20/2010	The Lost Symbol	$302,528$	4.3	2
Sample 3	12/27/2010	The Girl with the Dragon tattoo	$367,619$	4.2	3
Sample 4	1/3/2011	Mark Twain: An Autobiography	$327,615$	4.2	3
Sample 5	1/10/2011	Outliers: The Story of Success	$245,247$	4.3	2
Sample 6	1/17/2011	Decoded	$328,334$	4.2	3
Sample 7	1/24/2011	Light of the world	$322,953$	4.3	2

Supplementary Table S2: Details of a number of product co-purchase networks constructed by crawling the Amazon web site. The columns indicate the date when the crawl started, the starting book, the size of the sample after a week's worth of crawl, the best fit of the degree exponent γ to the tail and the measured number of super-stable nodes m_c.

Supplementary Methods

Pagerank and degree

The average pagerank of a node with degree $\mathbf{k} = (k_{in}, k_{out})$ is determined by its in-degree (44).

$$\overline{p}(\mathbf{k}) = \frac{(1-\alpha)}{N} + \frac{\alpha}{N} \times \frac{k_{in}}{\langle k_{in} \rangle}, \tag{S1}$$

where $\langle k_{in} \rangle$ is the average in-degree of the network, while the fluctuation in pagerank values for nodes with the same degree $\mathbf{k}$ is

$$\sigma(\mathbf{k}) \approx \frac{\alpha^2}{N} \left\langle \frac{k_{in}^2}{k_{out}} \right\rangle^{1/2} \langle {k_{in}}^{-3/2} \rangle \times k_{in}^{1/2}. \tag{S2}$$

Analysis of the largest numbers and the gap Distribution

Given a probability distribution $p(x)$ we can calculate the expectation value of the largest x if we draw N numbers from $p(x)$. The definition of the largest value is that there are no others greater than it. The probability for there to be any value larger than x is given by the cumulative distribution function defined as,

$$P(x) = \int_x^\infty p(x')dx'. \tag{S3}$$

Therefore the probability that there are none larger than it is given by $1 - P(x)$. Thus if we draw N numbers from a particular sample and one of them, say x_i, lies between the interval $x + dx$,the probability that there are no other numbers with a greater value than x_i is given by $p(x)dx \times [1 - P(x)]^{N-1}$. Since there are N ways of choosing i, the total probability is,

$$\pi(x) = Np(x)[1 - P(x)]^{N-1}, \tag{S4}$$

where the exponent reflect the fact that there are $N - 1$ values below the largest number. The expectation value of the largest number is,

$$\langle x_{max} \rangle = \int_0^\infty x\pi(x)dx, \tag{S5}$$

Using a similar approach we can calculate the expectation value of the m'th largest number. For this, there must be at least $m - 1$ values above it and $N - m$ values below it. Taking into account the appropriate combinatorial factor the total probability is,

$$\pi_m(x) = \frac{N!}{(N-m)!(m-1)!} p(x)P(x)^{m-1}[1 - P(x)]^{N-m}. \tag{S6}$$

The combinatorial factor can be written more compactly as,

$$\frac{N!}{(N-m)!(m-1)!} = \frac{\Gamma(N+1)}{\Gamma(N-m+1)\Gamma(m)} = \frac{1}{B(N-m+1, m)}, \tag{S7}$$

where $B(a,b)$ is the Legendre-Beta function. The expression is thus,

$$\pi_m(x) = \frac{p(x)P(x)^{m-1}[1-P(x)]^{N-m}}{B(N-m+1,m)}, \tag{S8}$$

and the s'th moment is given by,

$$\langle x^s \rangle_m = \int_0^\infty x^s \pi_m(x) dx. \tag{S9}$$

Choosing numbers from a power-law distribution (Scale-free networks)

Let us draw numbers from a power-law distribution of the form,

$$p(x) = (\gamma - 1)x^{-\gamma}. \tag{S10}$$

The cumulative distribution function is defined by,

$$P(x) = (\gamma - 1)\int_x^\infty x'^{-\gamma} dx' = x^{1-\gamma}. \tag{S11}$$

Substituting these into equation (S8) we get,

$$\frac{\gamma - 1}{B(n-m+1,m)} \times x^{-\gamma}(x^{1-\gamma})^{m-1}(1-x^{1-\gamma})^{N-m}. \tag{S12}$$

The expectation value for the m'th largest number is,

$$\langle x \rangle_m = \frac{\gamma - 1}{B(N-m+1,m)} \int_1^\infty x \times x^{-\gamma} \left[x^{1-\gamma}\right]^{m-1} \left[1 - x^{1-\gamma}\right]^{N-m} dx. \tag{S13}$$

Changing the integration variable to $y = 1 - x^{-(\gamma-1)}$, we get,

$$\langle x \rangle_m = \frac{1}{B(N-m+1,m)} \int_0^1 (1-y)^{m-1-1/(\gamma-1)} y^{N-m} dy = \frac{B(N-m+1, m - \frac{1}{\gamma-1})}{B(N-m+1,m)}. \tag{S14}$$

It is straightforward to show that the general expression for the s'th moment is,

$$\langle x^s \rangle_m = \frac{B(N-m+1, m - \frac{s}{(\gamma-1)})}{B(N-m+1,m)}, \tag{S15}$$

which in turn implies that for a power-law distribution with exponent γ, the s'th moment exists only if $\gamma - 1 > s/m$.

The expressions above are rather cumbersome, but we are interested in the highly ranked values, i.e small m for which $N >> m$, a regime where we can make a pretty strong approximation to the Beta function. By definition,

$$B(a,b) = \int_0^1 y^{a-1}(1-y)^{b-1}. \tag{S16}$$

Changing variables to $y = 1 - \frac{x}{a-1}$, we get,

$$B(a,b) = \frac{1}{(a-1)^b} \int_0^{(a-1)} \left[1 - \frac{x}{a-1}\right]^{a-1} x^{b-1} dx. \tag{S17}$$

In the limit of large a, (large N and small m as $a = N - m + 1$),

$$\left[1 - \frac{x}{a-1}\right]^{a-1} = e^{-x}, \tag{S18}$$

which implies,

$$B(a,b) \approx (a-1)^{-b} \int_0^{\infty} e^{-x} x^{b-1} dx = \Gamma(b)(a-1)^{-b}, \tag{S19}$$

where $\Gamma(x)$ is the Euler-Gamma function.

Using this approximation, we see that the expectation value of the m'th largest number scales as,

$$\langle x \rangle_m = \frac{B(N+1-m, m - \frac{1}{\gamma-1})}{B(N+1-m, m)} \approx (N-m)^{1/(\gamma-1)} \frac{\Gamma\left(m - \frac{1}{\gamma-1}\right)}{\Gamma(m)}. \tag{S20}$$

In Supplementary Figure S1a, we show the 50 largest values for a scale-free graph with the same exponent, for three different system sizes. The behavior described is clearly visible.

Once we know the expectation value of the largest m numbers, it is straightforward to calculate the difference between the m'th number and the $(m+l)$'th number where $l = (1, 2, \ldots, n-m)$. In the case of $l = 1$ we have, $\Delta = \langle x \rangle_m - \langle x \rangle_{m+1}$ and after a sequence of manipulations it can be shown that,

$$\Delta = N^{1/(\gamma-1)} \left(1 - \frac{m}{N}\right)^{1/(\gamma-1)} \frac{\Gamma\left(m - \frac{1}{\gamma-1}\right)}{\Gamma(m)} \times \frac{1}{m(\gamma-1)}. \tag{S21}$$

In Supplementary Figure S1b we plot the values of Δ for the same power-law as before for 3 different system sizes. As can be seen the agreement between theory and experiment is excellent. As per our analytical solution, the gap scales with system size, and after re-scaling with the appropriate factor, all three curves collapse into one.

Choosing numbers from an exponential distribution (Random networks)

Let us now draw numbers from an exponential distribution of the form,

$$p(x) = \lambda e^{-\lambda x}. \tag{S22}$$

Inserting this into equation (S8), we can show that the expectation value of the m'th largest number is,

$$\langle x \rangle_m = \frac{1}{\lambda} \times (H_N - H_{m-1}), \tag{S23}$$

where H_N is the N'th Harmonic number,

$$H_N = \sum_{k=1}^{N} \frac{1}{k}. \tag{S24}$$

For large N and the top ranks i.e. when $N >> m$, we can neglect the second term in equation (S23) and it can be shown that asymptotically we have,

$$\langle x \rangle_m \approx \frac{1}{\lambda} \times [\log(N) + \mathrm{O}(N^{-1})]. \tag{S25}$$

Here as well, the largest m values scale with system size, however the growth is only logarithmic and since the logarithm is a slowly varying function, i.e.,

$$\lim_{x \to \infty} \frac{\log(Ax)}{\log(x)} = 1, \tag{S26}$$

where A is some scalar, asymptotically $\langle x \rangle_m$ reaches a limiting value and does not change with large A. In Supplementary Figure S2a we plot the 50 largest values for an exponential distribution with $\lambda = 0.2$. The limiting behavior can be clearly seen. The bottom curve shows the collapse of the curves after division by the scale factor. Note that in this case the collapse only happens in the regime $N >> m$. For very large N however (on the order of 10^6), the logarithmic term is clearly dominant and the curves will collapse completely.

For the exponential distribution, the gap Δ is given by,

$$\Delta = \langle x \rangle_m - \langle x \rangle_{m+1} = \frac{1}{\lambda}[H_m - H_{m-1}] = \frac{1}{\lambda}\frac{1}{m}, \tag{S27}$$

In contrast with the power-law distribution Δ does not scale with N, but depends only on the parameter λ. Looking at equation (S27), the largest gap is for $m = 1$—the gap between the 1st and 2nd largest values—and is of the order $\lambda^{-1}m^{-1}$ and in general the gap is increasingly negligible for larger m. In Supplementary Figure S2b we show the size of the gaps as a function of system size. As the plot shows the gaps are indeed invariant with respect to N.

Stability condition

We claim that a vertex is stable if, given its pagerank, we perturb the network, and the resulting fluctuation in pagerank is less than the gap between its pagerank and that of the node ranked one position below it. Mathematically, if a vertex has pagerank p_m where m is its corresponding rank, then its stability is defined by the condition,

$$\sigma(p_m) \leq p_m - p_{m+1}. \tag{S28}$$

If we write $p_m - p_{m+1}$ as $\Delta(p_m)$, then depending on whether the gap is greater or lesser than the fluctuation in pagerank, we can define three different regimes:

$$\frac{\Delta(p_m)}{\sigma(p_m)} \begin{cases} > 1 & \text{Then there exists a finite } m \text{ that is stable.} \\ < 1 & \text{No such } m \text{ exists.} \\ = 1 & \text{Point marking the boundary between regimes.} \end{cases} \tag{S29}$$

Hence $\frac{\Delta(p_m)}{\sigma(p_m)} > 1$ represents out stability criteria for ranking.

Ranking stability in a Scale-free network

In order to determine the value of m at which the stability criteria is satisfied, we will need to combine the results of the previous sections. From Eqs. (S1) and (S20) the expected value of pagerank for a node of rank m in a scale-free network with exponent γ is,

$$p_m = \frac{(1-\alpha)}{N} + \frac{\alpha}{N\langle k_{in}\rangle} \times N^{1/(\gamma_{in}-1)} \frac{\Gamma\left(m - \frac{1}{\gamma_{in}-1}\right)}{\Gamma(m)}. \tag{S30}$$

We can write the gap $\Delta(p_m) = p_{m-1} - p_{m+1}$ as,

$$\Delta^{SF}(p_m) = \frac{\alpha}{N\langle k_{in}\rangle} \times N^{1/(\gamma_{in}-1)} \frac{\Gamma\left(m - \frac{1}{\gamma_{in}-1}\right)}{\Gamma(m)} \frac{1}{m(\gamma_{in}-1)}. \tag{S31}$$

The fluctuation $\sigma(p_m)$ is determined by combining Eqns. (S2) and (S20),

$$\sigma^{SF}(p_m) = \frac{\alpha^2}{N} \left\langle \frac{k_{in}^2}{k_{out}} \right\rangle^{1/2} \langle k_{in}\rangle^{-3/2} \times N^{1/2(\gamma_{in}-1)} \left[\frac{\Gamma(m - (\gamma_{in}-1)^{-1})}{\Gamma(m)}\right]^{1/2}. \tag{S32}$$

We divide the two quantities to get,

$$\frac{\Delta^{SF}(p_m)}{\sigma^{SF}(p_m)} = N^{1/2(\gamma_{in}-1)} \frac{\langle k_{in}\rangle^{1/2}}{\alpha} \left\langle \frac{k_{in}^2}{k_{out}} \right\rangle^{-1/2} \left[\frac{\Gamma(m - (\gamma_{in}-1)^{-1})}{\Gamma(m)}\right]^{1/2} \frac{1}{m(\gamma_{in}-1)}. \tag{S33}$$

From the equation above, we see that as N increases, the gap increases faster than the fluctuation by a factor of $N^{1/2(\gamma_{in}-1)}$ and thus the top ranked nodes are more stable in *larger* networks than in smaller ones. In general, however, for finite sizes the relation between the gap and the fluctuation will be sensitive to a combination of the value of the exponent γ_{in} and the reset parameter α.

A discussion is warranted, on the terms in the brackets $\langle \ldots \rangle$ in Eqn. (S33). Of particular interest is the second moment of the in-degree distribution k_{in}. Depending on the value of the exponent γ this term can be either finite or infinite. To see why, we write it explicitly,

$$\left\langle \frac{k_{in}^2}{k_{out}} \right\rangle = (\gamma_{in}-1)(\gamma_{out}-1)\int_1^\infty \int_1^\infty k_{in}^{2-\gamma_{in}} k_{out}^{-(1-\gamma_{out})} dk_{in} dk_{out} = \frac{\gamma_{in}-1}{\gamma_{in}-3} \times \frac{\gamma_{out}-1}{\gamma_{out}}, \tag{S34}$$

which diverges as $\gamma_{in} \to 3$. Note however, that even though there is no formal definition of the fluctuation in an infinite system due to the absence of any length scale, this is certainly not true in a finite system. In particular we know that in a finite system of size N, there is a well defined notion of a maximum expected degree and hence pagerank. On average, then, the fluctuations in the system cannot exceed the limit set by this maximum. We expect that the fluctuations scale with system size, such that in the thermodynamic limit $N \to \infty$, the fluctuations asymptotically diverge. Keeping this in mind, for a given size N, we can integrate the equation above by specifying an upper cutoff at a max value $k = K$—determined by setting $m = 1$ in equation (S14),

$$\begin{aligned} \left\langle \frac{k_{in}^2}{k_{out}} \right\rangle &= (\gamma_{in}-1)(\gamma_{out}-1)\int_1^{K_{in}} \int_1^{K_{out}} k_{in}^{2-\gamma_{in}} k_{out}^{-(1-\gamma_{out})} dk_{in} dk_{out} \\ &= \frac{\gamma_{in}-1}{3-\gamma_{in}}\left(K_{in}^{3-\gamma_{in}} - 1\right) \times \frac{\gamma_{out}-1}{\gamma_{out}}\left(1 - K_{out}^{-\gamma_{out}}\right). \end{aligned} \tag{S35}$$

One can see that when $\gamma_{in} > 3$, the term $K_{in}^{3-\gamma_{in}} \to 0$ (as K scales with N) and there is sign change such that $(3-\gamma_{in}) \to (\gamma_{in}-3)$ and we recover the original expression Eqn. (S34). For sake of completeness we will retain the term involving K_{in} for all regimes as we will see shortly that it is important to do so. Note however, that $K_{out}^{-\gamma_{out}} \to 0$ for any positive value for γ_{out} so we can simply drop it without loss of generality.

Finally, substituting the expression for K_{in} into equation (S35) we get,

$$\left\langle \frac{k_{in}^2}{k_{out}} \right\rangle = \pm C_\pm \left(\frac{\gamma_{in}-1}{\gamma_{in}-3}\right)\left(\frac{\gamma_{out}-1}{\gamma_{out}}\right), \tag{S36}$$

where $C_\pm$ is a scale factor taking on the following values,

$$\begin{aligned} C_+(\gamma_{in} > 3) &= 1 - N^{(3-\gamma_{in})/(\gamma_{in}-1)}\Gamma(1-(\gamma_{in}-1)^{-1})^{3-\gamma_{in}}, \\ C_-(\gamma_{in} < 3) &= N^{(3-\gamma_{in})/(\gamma_{in}-1)}\Gamma(1-(\gamma_{in}-1)^{-1})^{3-\gamma_{in}} - 1. \end{aligned} \tag{S37}$$

For $\gamma_{in} = 3$, using L'Hôpital's rule we note that the ratio,

$$\lim_{\gamma_{in} \to 3} \frac{C_\pm}{\pm(\gamma_{in}-3)} = \frac{1}{2}\log(N\pi). \tag{S38}$$

There are three distinct scaling regimes in the $N \to \infty$ limit. For $\gamma_{in} < 3$ the fluctuations diverge as some power of system size N, for $\gamma_{in} > 3$, the fluctuations go to a finite value Eqn. (S34), whereas at $\gamma_{in} = 3$, the fluctuations are logarithmically divergent. However for a finite system the fluctuations are well defined for all regimes by the expression above, with $\gamma_{in} = 3$ marking the boundary between two regimes.

Ranking stability in an exponential network

From Eqns. (S1) and (S20) the expected value of the pagerank of a node of rank m in an exponential network with exponent λ is,

$$p_m = \frac{(1-\alpha)}{N} + \frac{\alpha}{N\langle k_{in}\rangle}\frac{1}{\lambda}[H_N - H_{m-1}] \tag{S39}$$

The gap $\Delta(p_m) = p_{m-1} - p_{m+1}$ is thus,

$$\Delta^{exp}(p_m) = \frac{\alpha}{N} \times \frac{1}{m}. \tag{S40}$$

Note that the gap is independent of the parameter λ or in other words does not depend on the average degree $\langle k\rangle = \lambda^{-1}$. The fluctuation $\sigma(p_m)$ is given by,

$$\sigma^{exp}(p_m) = \frac{\alpha^2}{N} \times (2\lambda \mathrm{e}^{-\lambda})^{1/2} \times \left[\Gamma(0,\lambda)(H_N - H_{m-1})\right]^{1/2}, \tag{S41}$$

where $\Gamma(0,\lambda)$ is the Incomplete Gamma Function. We divide the two quantities to get,

$$\frac{\Delta^{exp}(p_m)}{\sigma^{exp}(p_m)} = \frac{1}{\alpha}\frac{1}{m}(2\lambda \mathrm{e}^{-\lambda})^{-1/2} \times \left[\Gamma(0,\lambda)(H_N - H_{m-1})\right]^{-1/2}. \tag{S42}$$

Note that in this case as the system size increases, the ratio *decreases* (as the gap is size independent, while the fluctuation grows logarithmically with N). Also it can be shown that the ratio is *always* less than one except in the limits $\alpha \to 0$ and $\lambda \to 0$. Thus in the exponential case there are no stable ranks for any realistic parameter choice.

Critical values

In the previous section, we saw that we can define three different regimes—equation (S29)—as per the ratio between the gap and the fluctuation. The ratio crucially depends on the specific values of (N, α, γ, m). Consequently setting the ratio equal to one, then allows us to define a critical value for each of the parameters, such that above that value we are in the stable regime and below in the unstable regime. We will focus particularly on two of them; one for the system size N and the other for the rank m. The first critical value $N = N_c$ specifies a minimum system size for which any stable ranks exist. The second critical value $m = m_c$ will tell us for a fixed system size $N > N_c$ the maximum rank in the network that is stable.

Critical system size N_c

Substituting the expression for the moments in Eqn. (S36) into Eqn. (S33) we find that,

$$\begin{aligned}\frac{\Delta(p_m)}{\sigma(p_m)} &= \frac{N^{1/2(\gamma_{in}-1)}}{\alpha}\left[\frac{\Gamma(m-(\gamma_{in}-1)^{-1})}{\Gamma(m)}\right]^{1/2}\frac{1}{m(\gamma_{in}-1)} \\ &\times \left(\frac{\gamma_{out}}{(\gamma_{out}-1)(\gamma_{in}-2)}\right)^{1/2}\left(\pm\frac{\gamma_{in}-3}{\pm 1 - N^{(3-\gamma_{in})/(\gamma_{in}-1)}\Gamma(1-(\gamma-1)^{-1})^{3-\gamma_{in}}}\right)^{1/2}\end{aligned} \tag{S43}$$

where as before $\pm$ depends on the regime we are in. In order to find a critical value for the system size N_c we note that the maximum value of Eqn. (S33) as a function of m for any choice of parameter values of (N, α, γ) is at $m = 1$. Furthermore, Eqn. (S33) is a monotonically decreasing function in m, so if there is a critical value for the system size, N_c, then it must at *least* hold for $m = 1$. Setting the ratio equal to one, $\gamma_{in} = \gamma_{out}$ (our results are not sensitive to γ_{out}) and $m = 1$, after re-arranging terms we find that N_c is a solution to the equation,

$$\begin{aligned} \pm\frac{N^{1/(\gamma-1)}}{(1 - N^{(3-\gamma)/(\gamma-1)}\Gamma(1-(\gamma-1)^{-1})^{3-\gamma})} &= \alpha^2(\gamma-1)^2\left[\frac{(\gamma-1)(\gamma-2)}{\pm\gamma(\gamma-3)}\right] \\ &\times \ (\Gamma(1-(\gamma-1)^{-1})^{-1}. \end{aligned} \tag{S44}$$

The equation above is transcendental in nature, and thus there is no known algebraic solution to it. Instead we resort to numerical methods, and it can be shown that in the limit of $\alpha \to 1$ (i.e the point at which we sample the topology exclusively and forbid random jumps), the only real solution to N_c is for $\gamma \geq 3$. The physical interpretation of this, is that below this limit there is no meaningful minimum critical value for which a stable rank exists. A stable rank *always* exists irrespective of system size. This is related to the fact that below $\gamma = 3$ we have hubs, and owing to their highly connected nature, hubs are always stable. For $\gamma > 3$ there is indeed a minimum critical system size for a stable rank to exist. In the limit $\gamma >> 3$, we can simplify Eqn. (S44) considerably and it is straightforward to show that,

$$N_c \approx \left[\frac{\alpha(\gamma-1)}{(\gamma-3)^{1/2}}\right]^{2(\gamma-1)}. \tag{S45}$$

This equation defines the minimum system size for which there exists a stable rank, as a function of the reset parameter α and the exponent γ. Note that since $0 < \alpha < 1$, N_c is very weakly dependent on α and is mostly dominated by γ. The curve describing N_c then divides the space into two phases (See Figure 2c in the manuscript). Above the curve is the stable regime (i.e the point at which at least one stable rank exists) and below it is the unstable regime (where no stable rank exists), while the curve equation (S45) represents the boundary.

Critical rank m_c

In a similar fashion as above, we can derive an expression for a critical value m_c for a fixed $N > N_c$. Once again setting Eqn. (S43) to one and re-arranging terms, we have,

$$\begin{aligned} \frac{1}{m}\left[\frac{\Gamma(m-(\gamma_{in}-1)^{-1})}{\Gamma(m)}\right]^{1/2} &= \alpha\ N^{-1/2(\gamma-1)}(\gamma-1)\left(\frac{(\gamma-1)(\gamma-2)}{\gamma}\right)^{1/2} \\ &\times \ \left(\pm\frac{1-N^{(3-\gamma)/(\gamma-1)}\Gamma(1-(\gamma-1)^{-1})}{\gamma-3}\right)^{1/2}. \end{aligned} \tag{S46}$$

To get an explicit expression for m_c, we will need to use numerical methods, however we can get a good idea of the qualitative behavior by making a few careful approximations. The most challenging part is the ratio of Gamma functions. However note, that in the limit $m >> 1$ the ratio follows,

$$\frac{\Gamma(m-a)}{\Gamma(m)} \approx m^{-a}, \tag{S47}$$

Applying this approximation we have,

$$\left(\frac{\Gamma(m-(\gamma-1)^{-1})}{\Gamma(m)}\right)^{-1/2} \approx m^{-1/2(\gamma-1)}. \tag{S48}$$

We can therefore write,

$$\left[\frac{\Gamma(m-(\gamma-1)^{-1})}{\Gamma(m)}\right]^{1/2}\frac{1}{m} \approx m^{-(2\gamma-1)/2(\gamma-1)}. \tag{S49}$$

The expression for the critical value of the rank m_c is then given by,

$$\begin{aligned} m_{c\pm} = N^{1/(2\gamma-1)}\Bigg[\quad & \alpha(\gamma-1) \quad \left(\frac{(\gamma-1)(\gamma-2)}{\gamma}\right)^{1/2} \\ \times \quad & \left(\pm\frac{1-N^{(3-\gamma)/(\gamma-1)}\Gamma(1-(\gamma-1)^{-1})}{\gamma-3}\right)^{1/2}\Bigg]^{-(\gamma-1)/(\gamma-1/2)}. \end{aligned} \tag{S50}$$

Thus, the number of stable ranks increases with the system size (for $N > N_c$) while the dependence on the exponent γ is highly nontrivial. However in the large N limit, the different scaling behaviors in the tail on either side of $\gamma = 3$ are,

$$m_c \sim \begin{cases} \gamma^{-\gamma} & \text{For } \gamma >> 3. \\ \mathrm{e}^{A(\gamma-2)/(2\gamma-1)} & \text{For } \gamma \text{ near } 2, \end{cases} \tag{S51}$$

where A is a constant.

Data and empirical analysis

Data Sources

The datasets for the scale-free networks listed in Table 1 in the manuscript (with the exception of the Amazon co-purchasing data and the Mobile Call graph) are available to download from the *Stanford Large Network Database* maintained by Jure Leskovec at `http://snap.stanford.edu/data`. The dataset for the Food Webs is available at the homepage of the EU project COSIN *COevolution and Self-organization In dynamical Networks* from `http://www.cosinproject.org`. The Mobile Call Graph data is taken from (47) while the neural network of *C. Elegans* is from (48). In order to check the temporal evolution of super-stable nodes we collected time resolved data for citations of High Energy Physics papers, samples of the Amazon co-purchasing product network and the entire publication history of Physical Review papers spanning over 100 years namely between 1893 and 2009.

High energy physics theory citation network

Our starting point for the citation network is a sample (49) of High Energy Physics Theory papers published in the period from 1993 to 2002 consisting of 27,770 papers with 352,807 edges (citations). Our analysis predicts three super-stable nodes ($m_c = 3$), close to the measured value of $m_c = 2$. The top cited paper in our data (2414 cites) was authored by Juan Maldacena, *The Large N limit of super conformal field theories and super gravity*, where he proposed the AdS/CFT correspondence—showing that String theory and gravity are equivalent in some limit. The second and third most cited papers at the time (1775, 1641 cites) authored by Ed Witten and S. Gubser respectively, represent follow up work on the same topic. In order to check their ranking stability over time, we collected data using the TopCites resource form the SPIRES database at `http://www.slac.stanford.edu/spires/topcites`. This lists the top 50 most cited High Energy Physics papers (ranked by the citations that they collect each year) found in the arXiv, on an annual basis starting from 1991. We find that the top two papers from our initial sample, continue to be the top-cited papers (in terms of cites per year) from 2003 to 2009 over a period of seven years (Figure 3c), while the third paper lost its prominence for a while but made a recovery around 2005 and continues to maintain its third rank.

Product co-purchasing network

Our product co-purchasing network is constructed from crawls of the Amazon website. We start from a pre-selected book (e.g. *The Lost Symbol* by Dan Brown) and conduct a Breadth First Search (BFS) by examining the products that were co-purchased with this item. The number of total unique items returned is fixed by the depth of the BFS tree (in practice a crawl of a week returned on average 300,000 items). A product co-purchasing network is thus constructed with item A "citing" item B, if B was co-purchased with A. In Table 1 in the manuscript we presented the first of a number of samples that we got from a crawl of the Amazon database, and in roughly good agreement with theory we found 2 super-stable nodes. The super-stable nodes were identified to be *Publication Manual of the American Psychological Association* and *Diagnostic and statistical manual of mental orders DSM-IV*. These two books continued to maintain their stability in terms of their ranking in all subsequent samples gathered on a weekly basis starting from different books (Figure 3d). We list the details of these samples in Supplementary Table S2. Furthermore, an examination of a much older dataset (50) of the Amazon co-purchased products dating from 2005 also revealed these to be the 2 most co-purchased books at the time.

Physical Review citation network

We collected the publication history of papers published in the Physical Review journals from 1893-2009, available upon request from the American Physical Society—`data-requests@ridge.aps.org`. From the data we constructed temporal citation networks with the resolution of a decade (1900, 1910, 1920 ...), containing all publications including and up to that decade. In Supplementary Figure S3a we plot the cumulative distribution function of the degrees in the citation network from 1910 to 2010. As the figure shows, the topology of the network changes considerably in time. In the earlier years it is difficult to fit the degree distribution owing to the small size of the network. Around the 1940's, the distribution starts to develop a power-law tail but the exponent γ continues to vary till 2009. Using a fit to the forms,

$$\begin{aligned} p(k) &= (\gamma-1)k_0^{1-\gamma}\,(k+k_0)^{-\gamma}\,, \\ P(k) &= \left(1+\frac{k}{k_0}\right)^{-(\gamma-1)}, \end{aligned} \tag{S52}$$

where the second expression is the cumulative distribution function and k_0 corresponds to the minimum value for which the power-law holds, in Supplementary Figure S3b we plot the change in γ with time and find that it decreases from $\gamma = 5$ in the 1950's to $\gamma = 3.2$ in the past decade. In Supplementary Figure S3c we plot the distributions for the years 1950, 1980 and 2009, as well as the corresponding fits as per equation (S52). (Note that adding an exponential cutoff e^{-k/k_c} to equation (S52) gives a value of $k_c = 10^6$.)

To check whether our prediction for a critical system size N_c manifests itself in the data, in Figure 3b in the manuscript, we plot the measured value of m_c as a function of size N as the citation network grows in time. We find that for a long period (1900-1950) there are no super-stable nodes. Only in 1960 as the number of papers published reach the critical size $N_c = 25000$, does the first super-stable node emerge. Subsequently, as predicted by our analytical results, the number of super-stable nodes increases slowly with size. To compare the results with theoretical predictions, on the same figure we plot the analytical curve (equation (S50)) where we used the appropriate estimated γ and adjusted size N for each decade. Overall we find a good qualitative agreement between the two. In Figure 3a in the manuscript, we plot the growth in papers as a function of time as well as the analytical expression for N_c (equation S44). The theory predicts the emergence of super-stability around the mid 1950's ($N_c^{th} = 20,000$), whereas in the data we see the first super-stable node around 1960 at $N_c^{data} = 25,000$. Thus, the predicted N_c^{th} and the measured N_c^{data} are in fairly close agreement.

In order to check the temporal stability of the super-stable nodes, we tracked the top nodes in each decade. In Supplementary Figure S4 we show the evolution of the ranking of the top ten nodes in each decade ranging from 1910 to 2000. The super-stable papers—solid lines in the plots—maintain their ranking over a remarkably long period of time, while the rest of the nodes lose their ranking fairly rapidly. Overall, it appears that super-stable nodes maintain their rank for six to ten years. However, even while they may lose their super-stability they continue to occupy a high rank in the network after extremely long time periods. In particular the top paper in 1970 (BCS theory of Superconductivity) fell only four places in rank after thirty years. In Supplementary Table S1 we present a number of statistics of the citation network as it evolves in time from 1893 to 2009.

Effect of degree-correlations on ranking

In our mean-field analysis of pagerank, we ignored the effect of degree correlations. A complete calculation is rather involved, but the qualitative effect of the correlations on our results can be uncovered by a simple scaling analysis. The effect of degree-correlations in scale-free networks is a well-studied problem (51–54). In particular the main result of interest is that the presence of correlations affects the scaling of the maximal degree. In scale-free networks, it can be shown that the corresponding correction to the scaling—see (53)—is of the form:

$$k_{max} \sim \min\left(N^{\frac{1}{\gamma-1}},\ N^{\frac{1-\beta}{5-\gamma}},\ N^{\frac{1-\beta}{\gamma-1}}\right), \tag{S53}$$

where $-1 \leq \beta \leq 1$ is the measure of degree-degree correlations. A positive value implies assortativity, while a negative value implies disassortativity. The exponent β is linear in Newman's assortativity coefficient r (55). The first term in the parentheses corresponds to the *natural cutoff* in the absence of correlations. The second term is the scaling in the region $2 \leq \gamma < 3$ and the last term is the scaling for $\gamma > 3$. Thus the corresponding corrections to the gap-fluctuation ratio equation (S33) are:

$$\frac{\Delta^{SF}(p_m)}{\sigma^{SF}(p_m)} \sim \begin{cases} N^{\frac{1}{2(\gamma-1)}} & \text{If } \beta \leq 0 \\ N^{\frac{1-\beta}{2(\gamma-1)}} & \text{If } \beta > 0 \text{ and } \gamma \geq 3 \\ N^{\frac{1-\beta}{2(5-\gamma)}} & \text{If } \beta > 0 \text{ and } 2 \leq \gamma < 3. \end{cases} \tag{S54}$$

In the case of disassortative networks, the scaling of the ratio is exactly the same as used in the manuscript, so the results presented are unaffected by the presence of these correlations. Now consider the networks presented in Table 1 in the manuscript. It is known that the world wide web, protein interaction networks, the food web and neural networks are disassortative—see (55). Thus for these the scaling analysis holds without any changes. The citation network and the product co-purchasing network are known to be weakly assortative.

As an illustrative example, let us assume a worst-case scenario and take a high assortativity value $\beta = 0.3$. The difference in scaling for the co-purchasing network is,

$$\frac{\Delta^{SF}(p_m)}{\sigma^{SF}(p_m)} \sim \begin{cases} N^{\frac{1}{2(4.3-1)}} = N^{0.15} & \text{for } \beta = 0 \\ N^{\frac{(1-0.3)}{2(4.3-1)}} = N^{0.1} & \text{for } \beta = 0.3, \end{cases} \tag{S55}$$

while for the high energy physics citation network it is,

$$\frac{\Delta^{SF}(p_m)}{\sigma^{SF}(p_m)} \sim \begin{cases} N^{\frac{1}{2(2.8-1)}} = N^{0.28} & \text{for } \beta = 0 \\ N^{\frac{(1-0.3)}{2(5-2.8)}} = N^{0.16} & \text{for } \beta = 0.3. \end{cases} \tag{S56}$$

Thus it can be seen that the change in scaling even in the presence of strong positive correlations is negligible. The precise number of super-stable nodes and the critical system size may vary (and as can be seen it will be even smaller for positively correlated networks) but the qualitative behavior remains unaffected, except in the rather extreme and unrealistic case of a perfectly correlated network, where $\beta = 1$. Note, that the actual citation and co-purchasing networks have negligible assortativity: $\beta = 0.002$ and $\beta = 0.015$ respectively.

Relation to previous results on pagerank stability

The efficiency of search algorithms (of which pagerank is only one example) is a well studied problem in computer science (36–39). While most theoretical analysis has focused on Kleinberg's HITS algorithm (56), there are a number of key features known about pagerank. Next we summarize the most pertinent results, and their relation to our work.

Pagerank stability to the reset parameter α.
It has been shown by (36) that should the network be perturbed in any fashion (specifically if the pagerank of l vertices are changed), then in the infinite graph limit, the difference between the pagerank vector of the original network p, and of the perturbed network $\tilde{p}$ has an upper-bound:

$$||\tilde{p} - p||_1 \leq \frac{2\sum_{j=1}^{l} p_j}{1-\alpha}, \tag{S57}$$

where $||\ldots||_1$ is the L1 or Manhattan distance, which for two N dimensional vectors p, q is defined by,

$$\sum_{i=1}^{N} |p_i - q_i| \, . \tag{S58}$$

This result implies that provided $1-\alpha$ (the probability to jump to a randomly chosen node in the graph) is not close to zero and the perturbed pages have small pagerank, then the perturbed vector is close to the original pagerank vector. Under these conditions pagerank is considered to be *score-stable*.

Pagerank stability to link addition.

In (37) the monotonicity of pagerank is proven. The authors consider a transition matrix $\mathbf{P}$ of a regular Markov chain and the pagerank vector p to be the stationary distribution of the Markov chain. Then a perturbation to $\mathbf{P}$ in the form of an error matrix $\mathbf{E}$ with all entries 0 except in one element E_{ij} is considered. This represents the addition of a *single* edge from node i to node j. The question addressed by the paper is this: does the addition of a single edge increase or decrease the transition rate from i to j? In other words does the addition of an inlink increase a node's pagerank? The authors rigorously prove that single-edge addition in fact increases j's pagerank and thus pagerank satisfies the property of monotonicity. Note, that this result is captured in equation (S1). Then the authors demonstrate empirically that the $L1$ distance between the perturbed vector $\tilde{p}$ and the original vector p is sufficiently small (within a tolerable error bound) for large scale edge addition. The case of edge deletion is not considered, nor the addition of multiple edges.

Pagerank stability in authority-connected graphs

In (57) it was noted that score-stability does not imply rank-stability. In other words while the pagerank of a page might not deviate much from its original value under perturbations, that does not imply that the absolute rank of the node derived from pagerank will be the same. This was touched upon in detail by (37). In particular the authors focus on a special class of networks called authority-connected graphs. These are graphs where for every pair of nodes i, j there exists a co-citation path connecting the two. Two vertices are said to be co-cited if there exists a third node k that links to both nodes i and j implying a topical connection between them. For these graphs the authors show that the ranking is extremely sensitive to change in the link structure. They thus conclude that pagerank is rank-unstable on authority connected graphs.

Differences between our work and previous publications.

(i) The results quoted above provide bounds on the efficiency of pagerank in the infinite graph limit (or the thermodynamic limit in physics). Real world networks are finite systems and consequently it is important to know what happens when we cannot take the infinite limit. In our approach we use the statistics of rare events for finite N which then allows us to perform a scaling analysis for precisely how the system diverges as it approaches infinity (even in the case of infinite moments where we renormalize by the maximal degree).
(ii) None of the previous work consider the effects of network topology on the efficiency of pagerank. As a result one of the holes in the analysis is the distinction between score and rank stability, a distinction that can only be bridged explicitly by defining the mapping between the score and rank via the network topology as we did in this paper. We combine a mean field analysis of the pagerank

(which isolates the indegree as the main contribution) and combine it with order statistics to make the connection between topology and ranking explicit.
(iii) The metrics used in the work discussed above focus on the efficiency of the global output of the algorithm. That is, pagerank is considered score (rank) stable only if the entire list is stable. In this approach the phenomena of super-stability (which is a property limited to the top ranks only) cannot be uncovered as no distinction is made between how the nodes with high pagerank respond to perturbations as opposed to nodes with low pagerank. Our definition of local stability overcomes this problem by calculating the difference in response to perturbations between the top (bottom) ranked nodes in terms of the gap to fluctuation ratio.

Let us demonstrate the difference by applying the distance metric used in (36,37) to our perturbation scheme. In Supplementary Figure S5a we plot the $L1$ distance between the original and perturbed pagerank vectors of two networks (averaged over N realizations for each size N), one exponential and one scale-free with the same average degree $\langle k \rangle = 1.5$ as a function of size N. As is clearly seen while the value of $L1$ is different for the two networks, there seems to be no clear scaling with size N. In addition, it is unclear what the physical meaning of a value of $L1 = 0.3$ is. Thus at least when it comes to the extensive property of scale-free networks the $L1$ distance fails to distinguish between scale-free and exponential networks.

What about rank stability? There are a number of ways to measure the distance between two ranked lists, such as the Kendall Tau (58)or the Spearman ρ (59). However in the spirit of measuring the distance between two probability distributions we use the approach of (60). Over the top c documents, the distribution weight associated with each rank $1 \leq r \leq c$ is of the form:

$$w(r) = \frac{1}{2c}\left(1 + \frac{1}{r} + \frac{1}{r+1} + \ldots + \frac{1}{c}\right), \tag{S59}$$

which can be written in closed form:

$$w(r) = \frac{1 + \psi(1+c) - \psi(r)}{2c}, \tag{S60}$$

where $\psi(r) = \dfrac{\Gamma'(r)}{\Gamma(r)}$ is the Digamma function. Note, in particular that the weights capture the distinction between the top ranked and lower ranked items in the list, in that it rewards the algorithm for getting the top queries right while not attaching much importance to the bottom of the list. In Supplementary Figure S5b we plot the $L1$ distance between the ranked lists of the same networks and find once again that the two networks are indistinguishable; there is an absence of scaling with N and moreover the distance measure is exactly the same for both networks. This despite the fact that we deliberately favored the algorithm towards matching the top queries correctly. Thus the $L1$ metric fails to capture the scaling properties, shows a difference between the values for the pagerank scores of an exponential and scale-free network but completely fails to distinguish between them when it comes to the difference between ranked lists. In addition there is no non-trivial way to connect the distance in scores to the distance in ranks.

Additionally one can use different distance measures, such as the Euclidean distance or L_2 norm, the Kullback-Leibler Divergence (61), the Jensen-Shannon Divergence (62), the Mutual Information (63) and many others. Applying each distance metric leads to qualitatively different results. While in the thermodynamic limit they might have the same asymptotic properties and thus when it comes to estimating bounds they are all appropriate, for finite systems it is always debatable as to which metric is appropriate for a given problem, and what the numerical value of this measure *physically* implies. (Apart of course from the complete dissimilarity/similarity limits 0,1.)

In summary the essential difference between previous approaches and our results, is that we consider the local stability of pagerank to perturbations. In particular we study the effect of specific real-world topologies on the quality of ranking in pagerank and this allows us to directly connect the score-stability to the rank stability via a direct mapping between the two measures. We have defined a principled perturbation to the system based on the fact that to first order a node's pagerank is linear in its indegree. This allows us to investigate the scaling properties in finite systems and to distinguish between the efficiency of pagerank on homogenous and heterogenous topologies, a novel result to the best of our knowledge. In addition we uncover a rich behavior within the class of heterogenous topologies in terms of the moments of the degree distribution.

Supplementary References

47. Onnela, J.-P. *et al* Structure and tie strengths in mobile communication networks. *Proc. Natl. Acad. Sci. USA* **104**, 7332-7336 (2007).

48. White, J. G., Southgate, E., Thomson J. N. and Brenner, S. The structure of the nervous system of the nematode caenorhabditis elegans. *Phil. Trans. R. Soc. London B* **314**, 1-340 (1986).

49. Leskovec, L., Kleinberg, J. and Faloutsos, C. Graphs over time: densification laws, shrinking diameters and possible explanations. *ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD)*, (2005).

50. Leskovec, L., Adamic, L. and Adamic, B. The dynamics of viral marketing. *ACM Transactions on the Web* **1**, 1 (2007).

51. Vazquez, A., Boguñá, M., Moreno, Y., Pastor-Satorras, R. and Vespignani, A. Topology and correlations in structured scale-free networks. *Phys. Rev. E* **67**, 046111 (2003).

52. Dorogovtsev, S., Mendes, J., Povolotsky, A., and Samukhin, A. Organization of complex networks without multiple connections *Phys. Rev. Lett.* **95**, 195701 (2005).

53. Weber, S. and Porto, M. Generation of arbitrarily two-point-correlated random networks. *Phys. Rev. E* **76**, 046111 (2007).

54. Xulvi-Brunet, R. and Sokolov, I.M. Reshuffling scale-free networks: From random to assortative. *Phys. Rev. E* **70**, 066102 (2004).

55. Newman, M. E. J. Assortative mixing in networks. *Phys. Rev. Lett.* **89**, 208701 (2002).

56. Kleinberg, J. M. Authoritative sources in a hyperlinked environment. ACM **46**, 604 (1999).

57. Borodin, A., Roberts, G. O., Rosenthal, J. S. and Tsaparas, P. Finding authorities and hubs from link structures on the world wide web. *Proceedings of 10'th International World Wide Web Conference* p. 415 (2001).

58. Kendall, M. A new measure of rank correlation. *Biometrika* **30**, 81-93 (1938).

59. Spearman, C. The proof and measurement of association between two things. *Amer. J. Psychol.* **15** 72-101 (1904).

60. Aslam, J. A. and Pavlu, V. Query hardness estimation using Jensen-Shannon divergence among multiple scoring functions. *Proceedings of the 29'th European conference on IR research*, (Springer-Verlag, Berlin, 2007), p. 198.

61. Kullback, S. and Leibler, R. A. On information and sufficiency. *Annal. Math. Stat.* **22**, 79-86 (1951).

62. Lin, J. Divergence measures based on the shannon entropy. *IEEE Trans. Infor. Theory* **37**, 145-151 (1991).

63. Guiasu, S. *Information theory with applications*, McGraw-Hill, New York, (1977).