Asymmetric network connectivity using weighted harmonic averages

2011 EPL 93 40002

(http://iopscience.iop.org/0295-5075/93/4/40002)

A LETTERS JOURNAL EXPLORING THE FRONTIERS OF PHYSICS

February 2011

EPL, **93** (2011) 40002 www.epljournal.org
doi: 10.1209/0295-5075/93/40002

Asymmetric network connectivity using weighted harmonic averages

GREG MORRISON[(a)] and L. MAHADEVAN[(b)]

School of Engineering and Applied Sciences, Harvard University - Cambridge, MA 02138, USA

received 22 November 2010; accepted in final form 21 January 2011
published online 24 February 2011

PACS 02.10.Ox – Combinatorics; graph theory

Abstract – We propose a non-metric measure of the "closeness" felt between two nodes in an undirected, weighted graph using a simple weighted harmonic average of connectivity, that is a real-valued Generalized Erdös Number (GEN). While our measure is developed with a collaborative network in mind, the approach can be of use in a variety of artificial and real-world networks. We are able to distinguish between network topologies that standard distance metrics view as identical, and use our measure to study some simple analytically tractable networks. We show how this might be used to look at asymmetry in authorship networks such as those that inspired the integer Erdös numbers in mathematical coauthorships. We also show the utility of our approach to devise a ratings scheme that we apply to the data from the NetFlix prize, and find a significant improvement using our method over a baseline.

A variety of complex natural and artificial systems can be viewed as a network [1], with a set of nodes representing objects and a set of edges connecting these nodes representing interactions between objects. Such systems include protein [2] or metabolic [3,4] networks, computer networks and the world wide web [5,6], disease propagation in populations [7,8], and networks of human [7,9] or other animal [10,11] interactions. While much of the study of networks generally involves characterizing both its internal structure [3,8,9] and the propagation of dynamical processes in it [1,12], a basic question that continues to be of interest is that of characterizing "closeness" or "similarity" between nodes in such networks. Various metrics of the distance between nodes in an undirected network have been developed including the integer distance [13] (identical to the classic Erdös numbers [14] which measure the authorship distance to the famous Hungarian mathematician), resistance distance [10,15,16], or a variety of measures determined from mutual nearest neighbors of nodes [17]. These approaches often have both a topological (incorporating the nearest-neighbor or global topology of the network) *and* geometric (satisfying the definition of a distance metric) character to them, although non-metric measures on undirected networks can be determined as well [17–19], such as the transition probabilities or hitting times for a random walk. In this paper, we develop a framework for determining the "closeness" between nodes in a weighted network by developing a real-valued Generalized Erdös Numbers (GENs) in the context of a social network. The GENs are inherently topological entities that incorporates nonlocal information about connectivity and are of use beyond the collaborative networks in which they were developed. The GENs are asymmetric (*i.e.* $E_{ij} \neq E_{ji}$) even if the underlying adjacency matrix is symmetric, thus failing to satisfy the requirements of a metric. Using analytically tractable symmetric networks, we show that the GENs can distinguish between topologies that are identical when viewed through the lens of common distance metrics [13,15,17]. We develop a basic predictor for a small subset of the NetFlix data [20] to demonstrate the utility of the GENs in real-world networks, and find significant improvement over a baseline prediction.

In order to develop a natural measure of the "closeness" between two nodes in a collaborative network, we consider one of the simplest possible connected graphs: a linear network of exactly three nodes (diagrammed in fig. 1(a)). With Erdös indexed as 0, we define his closeness to himself as $E_{00} = 0$, as is the case in all distance metrics [13,15]. For a node B (Bob, say) directly connected to exactly one other node (Alice in this case), we define the closeness felt by Bob towards Erdös as $E_{0B} = E_{0A} + w_{AB}^{-1}$, with w_{AB} the weight of the edge joining Alice and Bob. The

(a)E-mail: gmorriso@seas.harvard.edu
(b)E-mail: lm@seas.harvard.edu

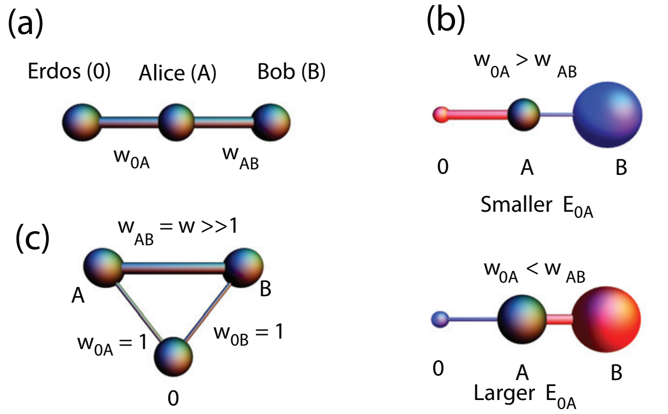

Fig. 1: (Colour on-line) (a) A simple linear author network. Bob has published only with Alice, so $E_{0B}=E_{0A}+w_{0A}^{-1}$. The competition between the connections with Erdös and leads to E_{0A} given in (1). (b) Schematic of the relationship between weights and GENs. Smaller sizes denote closeness to Erdös (large spheres implying large E_{0i}). Red coloring denotes strong interactions, while blue denotes weak interactions. E_{0A} increases as w_{AB} increases. (c) A simple cycle for two authors weakly connected to Erdös but strongly connected to one another.

determination of E_{0A} is more ambiguous, since Alice is connected to two nodes. If we were to use a distance metric [13,15–17] to determine the closeness between Erdös and Alice, we would find $E_{0A}=w_{0A}^{-1}$, independent of the fact that Alice is connected to Bob. This is a sensible result when considering distances, where closeness is determined from the (fixed) locations of nodes, and is thus independent of their connectivity. However, there are a variety of settings where Alice's interactions with one neighbor depends on her connectivity to the rest of the network, with collaborative networks being one example. In most sociological settings, Alice has a finite amount of time with which she can spend with her friends (Erdös and Bob), which must be shared between the two [19]. In a random walk the transition probability [18] from node i to node j depends on the degree of node i, and thus i's total connectivity. In cases where a finite resource must be allocated between neighbors, we expect that the closeness felt from node i towards node j will depend on the connectivity of i, and this closeness can therefore not be captured by a distance metric.

To incorporate the effect of multiple connections between Alice and the other nodes, we assume that the closeness Alice feels to Erdös is a function of the closeness felt by all other nodes connected to Alice towards Erdös. In particular, we expect $E_{0A}=f(\{E_{0i}+w_{Ai}^{-1}\})$, where $E_{0i}+w_{Ai}^{-1}$ would be the closeness node i feels to Erdös in the absence of all other edges. We expect I) that the unknown functional form of f should penalize large values of $E_{0i}+w_{Ai}^{-1}$ (*i.e.* that nodes that feel close to Erdös contribute more than nodes that feel far from Erdös when computing E_{0A}), and II) that connections with high weight have a higher contribution than those of low weights. These expectations are diagrammed schematically in fig. 1(b), and suggest the use of a weighted harmonic mean of the form

$$\frac{1}{E_{0A}}=\frac{1}{w_{0A}+w_{AB}}\left(\frac{w_{0A}}{E_{00}+w_{0A}^{-1}}+\frac{w_{AB}}{E_{0B}+w_{AB}^{-1}}\right), \qquad (1)$$

where the necessity of using the scaled weight $w_{Ai}/(w_{A0}+w_{AB})$ will be addressed below. We note that although (1) is the simplest and most natural functional form that satisfies the constraints above, other forms are certainly possible. Furthermore, the centrality of Erdös in a network may be replaced by that of any node i, so that we can generalize (1) to define the closeness felt by node j towards node i as

$$\frac{1}{E_{ij}}=\frac{1}{W_j}\sum_{l\in C_j}\frac{w_{jl}}{E_{il}+w_{jl}^{-1}}, \qquad (2)$$

where w_{jl} is the weight of the edge between j and l, C_j is the set of nodes directly connected to node j, and $W_j=\sum_l w_{jl}$ is the strength of node j. The reason for scaled weights w_{jl}/W_j becomes clearer in (2): using unscaled weights w_{jl} would imply that node j would have a low Generalized Erdös Number E_{ij} (*i.e.* feel very close to node i) by having many connections (large W_j), even if these connections led to nodes with high Erdös numbers E_{il}. We note that if $w_{jl}=\epsilon\delta_{jl_0}$ for some l_0, then $E_{jl_0}\sim\epsilon^{-1}\to\infty$ as $\epsilon\to 0$, so as a node with vanishing weight for its only connection to the network will have a diverging GEN as it becomes "disconnected".

To illuminate some aspects of the GENs we first consider a simple, three-node cycle shown in fig. 1(c), where two strongly connected nodes with weight $w_{AB}=w\gg 1$ are weakly connected to a third node (indexed 0). Solving (2) for the GENs E_{ij} for this simple network yields $E_{0A}=E_{0B}\sim(1+\sqrt{5})/2+O(w^{-1})\approx 1.62$ for large w, showing that the two nodes move away from the third (E_{0A} and E_{0B} increase) as their connection strengthens, with $E_{0A}=E_{0B}=1$ for $w=0$. Nodes A and B move towards each other as w increases, as can be seen by computing $E_{AB}=E_{BA}\sim w^{-1}+O(w^{-2})$. The third node has a low strength and is closer to the other nodes than they are to it ($E_{A0}=E_{B0}\sim\sqrt{2}+O(w^{-1})\approx 1.41$). Figure 1(c) thus displays the inherent asymmetry in the Erdös numbers ($E_{A0}\neq E_{0A}$), indicating that E_{ij} is not a distance metric.

The GENs differ from common distance metrics in a number of ways, as can be seen by examining some more complex networks. In a fully connected network of $N+1$ nodes of constant interaction strength $w_{ij}=w(1-\delta_{ij})$, each node has an identical Erdös number ($E_{0i}=E$ for $i\neq 0$), with the equivalent of (2) given by

$$\frac{wN}{E}=w^2+\frac{w^2(N-1)}{wE+1}, \qquad (3)$$

yielding $E=\sqrt{N}/w$. As the strength w of each connection between nodes increases, the Erdös number of all

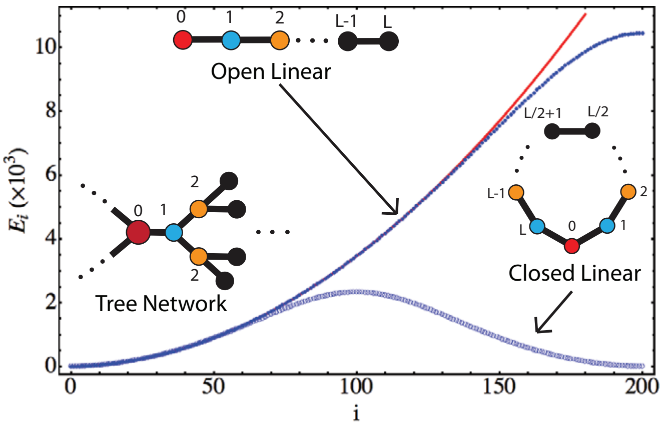

Fig. 2: (Colour on-line) The GENs computed for the open (filled symbols) and closed (open symbols) linear networks, along with the theoretical scaling of $E_{0i} = i(i+4)/3$ (solid red line). Insets schematically diagram the open and closed linear networks, as well as a tree network with $m = 3$ connections per node and length $L = 3$ (discussed in the text).

nodes decrease, since all nodes become closer to each other as well as to Erdös. However, as the number of nodes increases, with edges added to keep the network fully connected, the importance of an individual edge is lessened and all nodes will feel less close to one another. This is in contrast with other measures such as the resistance distance, which decreases as new nodes are added or integer distance, which remains constant independent of N.

We next consider generalizations of the simple networks (fig. 1) to extended linear networks and a cycle-free tree (fig. 2) where each node is connected to exactly m nodes, except for the endpoints. For the open networks with $m \geqslant 2$, the resistance distance between any two points is given by the integer distance, $R_{ij} = D_{ij}/w = |i-j|/w$, since there are no cycles. The GEN between a node i and the base of a branch (indexed 0) satisfies

$$\frac{wm}{E_{0i}} = \frac{w^2}{wE_{0,i-1}+1} + \frac{w^2(m-1)}{wE_{0,i+1}+1} \qquad (4)$$

with the boundary conditions $E_{00} = 0$ and $E_{0L} = E_{0,L-1} + w^{-1}$. The closed linear network can be studied using the same difference equation, with the boundary conditions $E_{00} = E_{0,L+1} = 0$ after insertion of a virtual node. $E_{0i}(w) = E_{0i}(w=1)/w$ for constant interaction strength w, so the weights can be factored out, and are ignored below. While the difference equations are not exactly solvable, if $m = 2$ we can see that $E_{0i} = i(i+4)/3$ is a solution that satisfies (4) and the boundary condition $E_{00} = 0$. For $i \approx L$, deviations from this predicted scaling are expected to occur due to the boundary condition at the distant ends. Interestingly, the quadratic scaling $E_{0i} \sim i^2$ for distant nodes matches the time for particle diffusion from node 0 to i, taking time $\tau_{0i} \sim i^2$. For tree networks with large m (inset of fig. 2), we find that $E_{0i} = E_{0,i-1} + (m-1)^i$ asymptotically satisfies the difference equation with the boundary condition $E_{00} = 0$. The tree network produces an exponential growth with i for large m, rather than the quadratic growth seen for $m = 2$, clearly showing that the GENs are able to distinguish between the global topology of these very different networks more accurately than a resistance distance approach.

In order to determine the numerical values of the GENs for this linear network (with $w = 1$), we determine an iterative solution for E_{0i}, with $2/E_{0i}^{(t)} = (E_{0,i-1}^{(t-1)}+1)^{-1} + (E_{0,i+1}^{(t-1)}+1)^{-1}$ and $E_{0i}^{(0)} = i(i+4)/3$. $E_{0i}^{(t)}$ is computed until $\epsilon(t) = \max_j |E_{0i}^{(t)} - E_{0,i}^{(t-1)}| < 0.01$. The resulting numerical solutions to the GENs are shown in fig. 2, with the solid red line denoting the predicted quadratic growth, $E_{0i} = i(i+4)/3$. The predicted scaling agrees well with the numerical results[1], with deviations occurring near the $i = N$ endpoint for the open network and near the $i = N/2$ midpoint for the closed network.

Moving away from simple, symmetric networks, we now examine the structure of more complex networks using the coauthorship network [21] of Paul Erdös to see the differences between GENs and integer distance between nodes. The 17,724 edges in the (unweighted, with $w_{ij} = 0$ or 1) network incorporate all coauthorships with Erdös as well as coauthorships between authors with integer Erdös number 1 and those with integer Erdös number $\leqslant 2$, with a total of $N = 9778$ authors. Coauthorships between a pair of mathematicians both of whom have integer Erdös number 2 (*i.e.* edges between non-neighbors of Erdös) are omitted in this network. In fig. 3, we show the GENs only for coauthors direct coauthors of Erdös who all have integer Erdös number 1, and see both the closeness felt by Erdös towards his coauthors (E_{i0}, red squares) and the closeness felt by his coauthors towards him (E_{0i}, blue circles). Mathematicians who have very low degree ($W_i \sim 1$) feel very close to Erdös, since they are connected directly to him and few other nodes. In particular, a coauthor of Erdös with $W_i = 1$ will necessarily have $E_{0i} = 1$. Nodes with high degree ($W_i \gg 1$) feel less close to Erdös, because their attention is divided between a large number of nodes, as was the case in our artificial networks. On the other hand, the closeness felt by Erdös towards his coathors has the opposite behavior: Erdös is less close to nodes with low degree, because his attention is divided between many coauthors, but is closer to nodes with high degree because of the many paths between him and these coauthors. The asymmetry in the GENs, with $E_{0i} \neq E_{i0}$, would not be discernible using a distance metric such as the integer distance between nodes, which is necessarily symmetric. Quantitatively, we see that the closeness felt towards Erdös scales as $E_{0i} \sim W_i^{0.34}$ and the closeness felt by Erdös scales as $E_{i0} \sim W_i^{-0.45}$; the particular exponents in the growth or decay of the GENs depends strongly on the global topology of the network,

[1]The numerical results for the tree network with $m \geqslant 5$ also agree with the theoretically predicted exponential growth.

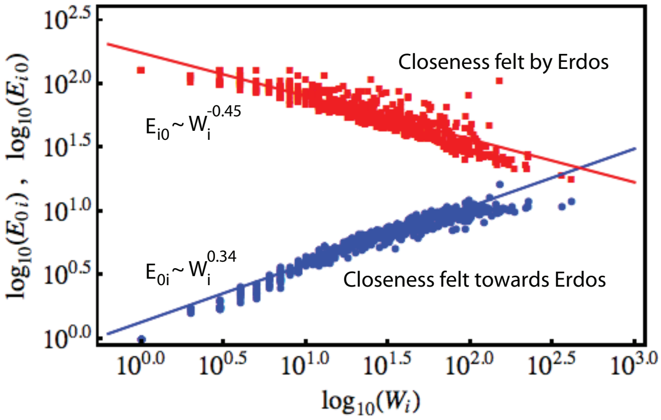

Fig. 3: (Colour on-line) The GENs for the network of mathematical coauthors of Paul Erdös. Blue circles show the closeness felt by coauthors of Paul Erdös towards him, and red squares show the closeness felt by Erdös towards his coauthors. Authors who have low degree (who have published few papers) feel much closer to Erdös than those with hig h degree (see the text). Alternatively, Erdös feels closer to those coauthors with high degree, who have worked not only with Erdös, but also with other authors to which Erdös feels close. The best-fit nonlinear scalings of $E_{0i} \sim W_i^{0.34}$ and $E_{i0} \sim W_i^{-0.45}$ have a power law growth or decay whose exponent will depend on the global topology of the network, and the variations of the GENs for authors with the same degree is due to local variations in the network connectivity.

and are not expected to be universal. There is a crossover when $E_{0i} \sim E_{i0}$, and is expected since a virtual author connected to all nodes in the network (with $W_i = N$) will not feel close to any node, including Erdös, while Erdös will feel very close to this virtual node due to the large number of paths. We would expect for a tree-like network (where there is a single path between Erdös and any other author) that the transition with $E_{0i} = E_{i0}$ would occur for nodes having the same degree as Erdös: $W_i = W_0 = 511$ in contrast with the observed cross-over for the real network that occurs for $W_i = 465$.

Having analyzed both idealized and real networks using GENs, we now show how they can be used as a basis for deducing higher-order information about the structure of networks. As an example, we consider the data provided for the NetFlix prize [20], a competition to improve algorithms for the prediction of future movie ratings based on previous ratings. Here, we use the GENs as a means to characterize an interaction "energy" between nodes when predicting the rating user i gives to movie l, $p_i^{(l)}$, using the Boltzmann weighted average taken from statistical mechanics,

$$p_i^{(l)} = \sum_{j \in S_l} r_j^{(l)}\ e^{-\bar{w}\beta E_{ij}} \Big/ \sum_{j \in S_l} e^{-\bar{w}\beta E_{ij}}, \qquad (5)$$

with $\bar{w}$ the average weight of each edge (see below). β is a free parameter (an inverse temperature), describing how important distant nodes are in determining the predicted rating. $\bar{w}\beta E_{ij}$ is the 'interaction energy' between users i and j, determining which nodes are important to the average and which are not and assigning a lower weight to the latter. In order to compute the Erdös numbers in (5), we need to generate a weighted graph use the NetFlix data.

While we could represent the NetFlix data as a bipartite network [19,22], where the users and movies form sets of disjoint nodes, we instead use the movie ratings (an integer between 1 and 5) to determine a weight between two users, using the simple power law of the form

$$w_{ij} = \sum_{l \in M_{ij}} (5 - |\Delta r_{ij}^{(l)}|)^{\alpha} \qquad (6)$$

with $\Delta r_{ij}^{(l)} = r_i^{(l)} - r_j^{(l)}$, $r_i^{(l)}$ the rating user i gave to movie l ($0 \leqslant |\Delta r_{ij}^{(l)}| \leqslant 4$), and M_{ij} is the set of movies that both user i and j have rated ($w_{ij} = 0$ if i and j have rated no movies in common). If users i and j disagree on all movies (*i.e.* one rates a 5 while the other rates a 1), the weight between them is $w_{ij} = |M_{ij}|$, while perfect agreement (both rating 1's or 5's) gives a weight $w_{ij} = 5^{\alpha} \times |M_{ij}|$. Implicit in this definition is that users who seek out the same movies have more similar tastes than those who do not (even if they do not agree), and that users who agree on movies are more likely to have similar tastes than those who disagree. The free parameter α determines the importance of agreement, with $\alpha = 0$ implying that disagreement in the ratings are irrelevant, while ratings that agree become dominant as $\alpha \to \infty$.

To test our prediction scheme, we select a subset of the full NetFlix dataset comprised of N users and 6000 movies (the parameters are listed in table 1). For varying values of N and α, we choose k users from the data set in order to test the efficacy of our approach (k is shown in the third column in table 1). For each node i selected, we iteratively perform the followings steps for each movie l user i has seen: I) remove the rating user i gave to movie l from the network, II) compute the GENs for this modified network using (2), and III) compute the predicted rating user i gives to movie l using (5) as a function of β, $p_i^{(l)}(\beta)$. The average improvement as a function of β is determined from the RMSD $\rho_i^2(\beta) = \sum_l [r_i^{(l)} - p_i^{(l)}]^2/n_i$, where n_i is the number of movies that user i has seen.

Table 1: Parameters used in the NetFlix analysis. N is the number of users in the dataset, and k is the number of users for whom predictions were made. The number of nodes with $n_i \geqslant n_{min} = 30$ out of the k considered are shown, as well as the average percent improvement for all nodes with at least 30 movies seen.

Case	N	k	α	No. with $n > 30$	$\kappa_{\beta=2}(n_{min} = 30)$
1	3000	553	2	304 (55%)	3.56%
2	3000	557	4	302 (54%)	4.71%
3	3000	1297	8	782 (60%)	3.35%
4	6000	368	8	188 (51%)	4.53%

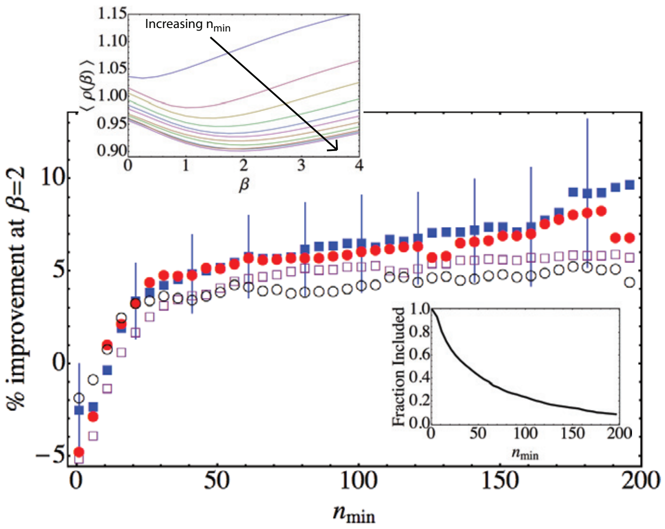

Fig. 4: (Colour on-line) Percent improvement at $\beta = 2$ compared to $\beta = 0$ ($\kappa_{\beta=2}$) as a function of n_{min}. See table 1 for parameters. Case 1 is shown as open circles, case 2 as filled circles, case 3 as open squares, and case 4 as filled squares. Error bars (using the standard deviation of the mean) are shown only for case 4, with the errors for the other cases being smaller. Upper inset shows $\langle\rho(\beta)\rangle$ as a function of β for varying n_{min} for case 3 (higher curves correspond to smaller n_{min}). The lower inset shows the fraction of users satisfying $n \geqslant n_{min}$ for case 3.

The RMSD ρ_i depends strongly on the number of movies (n_i) that the user has seen, as is shown by computing the average RMSD restricted to users with $n_i \geqslant n_{min}$. In the upper inset of fig. 4, a pronounced minimum in $\langle\rho(\beta)\rangle$ occurs for increasing n_{min}. The relative improvement of (5) over an unweighted average ($\kappa_\beta = 1 - \langle\rho(\beta)\rangle/\langle\rho(0)\rangle$) is significant for $n_{min} \geqslant 30$ as seen in the main panel of fig. 4. Restricting ourselves to users with $n_i \geqslant 30$ ratings gives an improvement of at least 3–5% at $\beta = 2$ for all values of α and k examined (over 50% of the nodes included in the average in all cases, see table 1). For very well connected nodes (with $n_{min} = 200$ or about 8% of the nodes in each case, see the lower inset of fig. 4) the average improvement is quite significant, ranging from 4.5–9.5%. The dependence of the improvement on n_{min} is somewhat unsurprising, as the preferences of users who have seen very few movies will be much more difficult to predict. We also note that the negative improvement for small n_{min} is due to the fact that the positions of the minimum in $\langle\rho(\beta)\rangle$ saturate at $\beta = 2$ for large n_{min}, but are far from this value for small n_{min} (upper inset of fig. 4).

In conclusion, our minimal measure of connectivity in networks uses a simple weighted harmonic average and leads to Generalized Erdös Numbers (GENs) which are real-valued, asymmetric and take the global topology of the network into account. They can be used to characterize connectivity in a range of networks, as is demonstrated by examining a variety of simple artificial networks and the Erdös coauthorship network, and further can form the basis for detailed probes of the structure of complex networks. We show the latter by using GENs to construct a ranking scheme for data sets from the Netflix prize, where it outperforms a baseline predictor. The weighted average in (5) can be implemented in other prediction schemes, and a more complex form for the weighting between nodes (incorporating temporal information, for example) may give further improvements in predictions. A natural next step of any measure of connectedness is to to use it in additional applications: problems associated with community detection in graphs, as well as the dynamics of diffusion, epidemics and the behavior of dynamic networks with time-dependent edge weights beckon.

* * *

The authors thank M. Venkadesan for numerous useful discussions and the suggestion of examining the NetFlix data, J. Grossman for the publicly available Erdös Number Project data, and the comments of anonymous reviewers for improvements in the text.

REFERENCES

[1] Newman M., Barabasi A. L. and Watts D. J., *The Structure and Dynamics of Networks* (Princeton, Princeton, NJ) 2006.

[2] King A. D., Pržulj N. and Jurisica I, *Bioinformatics*, **20** (2004) 3013.

[3] Jeong H., Tombor B., Albert R., Oltval Z. N. and Barabási A. L., *Nature*, **407** (2000) 651.

[4] Patil K. R. and Nielsen J., *Proc. Natl. Acad. Sci. U.S.A.*, **102** (2005) 2686.

[5] Page L., Brin S., Motwani R. and Winograd T., Technical Report 1999-66, 1999.

[6] Adamic L. A., Huberman B. A., Barabási A-L., Albert R., Jeong H. and Biaconi G., *Science*, **287** (2000) 2115.

[7] Keeling M. J., *Proc. R. Soc. London, Ser. B*, **266** (1999) 859.

[8] Pastor-Satorras R. and Vespignani A., *Phys. Rev. Lett.*, **86** (2001) 3200.

[9] Travers J. and Milgram S., *Sociometry*, **32** (1969) 425.

[10] McRae B. H., *Evolution*, **60** (2006) 1551.

[11] Proulx S. R., Promislow D. E. L. and Phillips P. C., *Trends Ecol. Evol.*, **20** (2006) 345.

[12] Albert R., Jeong H. and Barabási A. L., *Nature*, **406** (2000) 378.

[13] Buckley F. and Harary F., *Distance in Graphs* (Addison-Wesley, Redwood City, Cal.) 1990.

[14] de Castro R. and Grossman J. W., *Math. Intell.*, **21** (1999) 51.

[15] Klein D. J. and Randic M., *J. Math. Chem.*, **12** (1993) 81.

[16] Balaban A. T. and Klein D. J., *Scientometrics*, **55** (2002) 59.

[17] Limben-Nowell D. and Kleinberg J., *J. Am. Soc. Inf. Sci.*, **58** (2007) 1019.

[18] Bilke S. and Peterson C., *Phys. Rev. E*, **64** (2001) 036106.
[19] Zhou T., Ren J., Medo M. and Zhang Y.-C., *Phys. Rev. E*, **76** (2007) 046115.
[20] Koren Y., http://www.netflixprize.com/assets/GrandPrize2009_BPC_BellKor.pdf.
[21] The network data was downloaded from the Erdös Number Project website, maintained by J. Grossman, and available at http://www.oakland.edu/enp/.
[22] Zhou T., Kuscsik Z., Liu J-G, Medo M., Wakeling J. R. and Zhang Y-C., *Proc. Natl. Acad. Sci. U.S.A.*, **107** (2010) 4511.