

Generalized Erdős Numbers for network analysis

Greg Morrison^{1,2}, Levi Dudte¹, and L. Mahadevan¹

1: School of Engineering and Applied Sciences,
Harvard University, Cambridge MA 02138

2: Laboratory for the Analysis of Complex Economic Systems,
IMT Institute for Advanced Studies, Lucca Italy 55100

July 17, 2015

Abstract

In this paper we consider the concept of ‘closeness’ between nodes in a weighted network that can be defined topologically even in the absence of a metric. The Generalized Erdős Numbers (GENs) [1] satisfy a number of desirable properties as a measure of topological closeness when nodes share a finite resource between nodes as they are real-valued and non-local, and can be used to create an asymmetric matrix of connectivities. We show that they can be used to define a personalized measure of the importance of nodes in a network with a natural interpretation that leads to a new global measure of centrality and is highly correlated with Page Rank. The relative asymmetry of the GENs (due to their non-metric definition) is linked also to the asymmetry in the mean first passage time between nodes in a random walk, and we use a linearized form of the GENs to develop a continuum model for ‘closeness’ in spatial networks. As an example of their practicality, we deploy them to characterize the structure of static networks and show how it relates to dynamics on networks in such situations as the spread of an epidemic.

The study of complex networks has increased enormously in recent years due to their applicability to a wide range of physical[2, 3], biological[4], epidemiological[5, 6], and sociological[7] systems. Two basic goals in this regard are to understand and quantify the structure of the network to better characterize the relationship between the interacting members of the network (the nodes), while also characterizing the dynamical processes on the network[7] that may shed light on the processes by which they form [8].

Understanding the topological properties of the network on both a global and local level can be useful in approaching both of these goals. Global properties of interest may include simple measures of the distribution of node properties, such as the degree distribution, $P_d(k)$, with k the number of edges incident upon a node; strength distribution, $P_s(s)$, with s the total weight incident upon a node; or distribution of clustering coefficients,

$P_c(c)$, with c the fraction of triplets of connected nodes that are closed[9, 10]. Community structure in the network[11, 12, 13], which partitions the network into densely connected sub-networks with more links within communities than between communities, has been extensively studied and may provide more detailed information about the relationship between nodes than simple distributions. Community structure in networks can indicate the existence of underlying similarities between nodes in the network, and may have a great impact on dynamical processes occurring on the network (such as a random walk[14, 15, 16] or epidemic spreading[5, 17, 18]), and can influence the material properties of granular systems[2].

While global properties of networks can be used to assess the attributes of the nodes on an aggregate level, it is also of great interest to understand the topological properties of nodes on an individual, local level as well. Node centrality is the classic example of a topological measure of an individual node, which assesses the ‘importance’ of a node in a variety of contexts. The most basic measure of a node’s centrality is simply related to its degree, a property of the node that is based solely on the local topology of its connectivity. The centrality of individual nodes can also be measured incorporating the global topology of the network in a variety of ways, including PageRank[19], betweenness[16], or random walk[14] centralities. Each of these measures reduces the global properties of the network into a individualized local measure of importance, permitting a rank-ordering of their importance in the network[20, 21].

In many contexts[22, 23], not all members of the network will necessarily agree on the importance of the same node: nodes that have a direct connection between them will be more important to each other than distant nodes in the network. Nodes that are central to the network as a whole may have very low importance from the perspective of sub-networks. The universality of importance is further complicated by the fact that we may expect the influence between a pair of nodes to be asymmetric even if they are directly connected[23] (the importance assigned by an important node towards an unimportant one is not necessarily the same as the importance assigned in the opposite direction), which may have important consequences in real-world systems[4]. The determination of a personalized measure of node importance that incorporates the global topology in an asymmetric measure is therefore an important but non-trivial problem.

In this paper, we use the recently developed Generalized Erdős Numbers[1, 12] (GENs), which provide a nonlinear asymmetric measure of the pairwise ‘closeness’ between nodes based on the network’s global topology, as a proxy for the importance one node feels towards another. We show that this measure of closeness can be linked to the dynamics of random walks and spreading of epidemics on networks. Using the GENs, we develop a linearized importance that provides an intuitive measure of the (asymmetric) importance one node feels towards another based on the global network topology. This linearized importance is considered in a continuum approximation for a spatial network with homogeneous

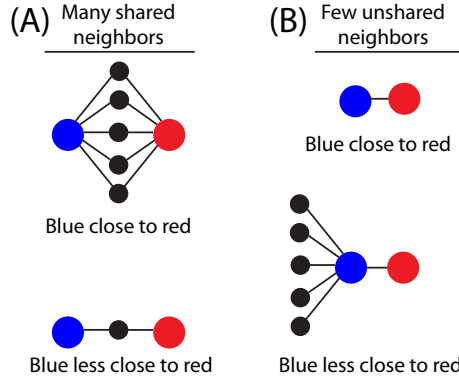


Figure 1: Two competing requirements for global ‘closeness’ in a network with shared resources. In (A), many short paths between nodes increase the closeness between them. This is similar to the resistance distance between nodes: additional parallel paths between them reduce their resistance distance. In (B), the finite resources of the high-degree blue node suggest it should be less close to the red node than for the lower-degree blue node above, as resources are shared with the other neighbors as well. This is similar to the transition probability from the blue node in a random walk: the more connections the blue node has, the lower probability of visiting the red node.

distance-dependent links between material points on a sphere which is coupled to long-range distance-independent connection strengths, and is shown to reduce to an inhomogeneous Fredholm equation with a kernel depending on the topology of the spatially-independent portion of the network. Finally, we show that a global measure of centrality defined using the GENs is consistent with other centrality measures, and that the asymmetry in the GENs is consistent with the asymmetry of the mean first passage time (MFPT) between nodes in a random walk. The flexibility of the GENs in understanding the topology of complex networks shows their utility in a wide range of problems.

1 The Generalized Erdős Numbers

When nodes represent objects in a physical space[24, 25, 26, 3, 27], the distance between nodes, D_{ij} , is a naturally defined (metric) measure of closeness between the objects. Due to the generality of networks (where nodes and edges abstractly represent ‘objects’ and ‘interactions,’ respectively), there can be no guarantee of a naturally defined distance metric[28, 3], and, in some cases, the network topology itself must define a measure of closeness (Δ_{ij}) based solely on the matrix of weights between nodes i and j , w_{ij} (with an undirected network where $w_{ij} = w_{ji}$ is assumed throughout this paper). The closeness, Δ_{ij} , will be small for nodes that are close to one another and large for distant nodes, with

a simple and common choice being $\Delta_{ij} = w_{ij}^{-1}$ (so strongly connected nodes are ‘close’, and disconnected nodes are ‘far’). Alternatively, in an unweighted network, the length of the shortest path between a pair of nodes is a natural definition[28, 29] and is the basis for the classic Erdős numbers in the context of an unweighted collaboration network[30]. Improvements on this measure which incorporate the effect of multiple paths between nodes (see Fig. 1(A) for a schematic diagram) include the resistance distance[15, 31], self consistent similarity measures[32], and communicability [33], to name only a few. An additional approach to defining similarity between nodes is found by positing a multidimensional ‘latent space’ of node properties[34], with the assumption that nodes that are close in the latent space are likely to be connected in the network and each node’s position in the space inferred from the observed connectivity. Each of these methods incorporates the global topology of the network into a symmetric measure of closeness between pairs of nodes ($\Delta_{ij} = \Delta_{ji}$).

Finite resources are shared in some networks, with examples including collaboration on networks (where time with one collaborator reduces the available time for others), multi-core processor components[35] (where finite memory or other hardware must be shared), and random walks (where the walker can only move to a single neighbor at a time with a transition probability $P_{i \rightarrow j} = w_{ij}/W_i$ with $W_i = \sum_k w_{ik}$ the total strength of the node i). In the context of these networks of limited resources, closeness measures such as resistance distance may be undesirable[23], because the addition of a new edge in the network should be detrimental to some nodes (those who receive less of the finite resource due to the new edge) and beneficial to others (those who receive more due to the edge). For closeness measures based on the direct weight between nodes (where the ‘closeness’ between i and j is often taken to be w_{ij}^{-1}) or resistance distance between nodes, it is straightforward to see that the newly measured closeness between nodes i and j $\Delta_{ij}^{(new)} \leq \Delta_{ij}^{(old)}$ for all pairs, i.e. the addition of an edge can never cause nodes to feel less close to one another. This is not sensible in the context of nodes that share a finite resource with its neighbors, as shown in Fig. 1(B): if a node i has many neighbors, each receives less of the resource than if i had few neighbors. A quantity such as the transition probability in a random walk, $P_{i \rightarrow j}$, is asymmetric and ensures that nodes are closer if they have few neighbors, pictured in Fig. 1(B) (so a walker is more likely to pass between them than if they had many connections). However, it is not a global measure of closeness because the transition probability incorporates only the nearest neighbor connections between nodes. It is useful to develop a measure of closeness that incorporates these two (sometimes contradictory) aspects depicted in Fig. 1: nodes feel close if there are many paths between them, but popular nodes are less close to their neighbors than unpopular nodes.

We have recently shown[1] that the Generalized Erdős Numbers (E_{ij} , or GENs), describing the closeness node j feels towards node i , satisfy the expected properties for the sharing of

finite resources described in Fig. 1. The GENs are defined as

$$\frac{W_j}{E_{ij}} = w_{ij}^2 + \sum_{\substack{l \in C_j \\ l \neq i}} \frac{w_{jl}}{E_{il} + w_{jl}^{-1}} \quad E_{ii} \equiv 0, \quad (1)$$

where C_j is the set of nodes directly connected to j . This form is chosen such that the node i is as close as possible to itself and that if j is connected to only one node k , j 's closeness to i satisfies $E_{ij} \equiv E_{ik} + w_{jk}^{-1}$. If there are multiple paths between nodes, the closeness j feels to i is strengthened if there is a direct connection between them but also includes a contribution from all other neighbors of j weighted by their connection strength. By choosing a harmonic mean for the form of the contribution, we bias our measure of closeness towards neighbors that themselves feel close to i . The GENs are defined using the global topology of the network, and E_{ij} is finite even if i and j share no neighbors (as may not be the case for more local measures of closeness [23]).

This definition of the GENs in Eq. 1 is nonlinear, and the exact values of E_{ij} for complex networks are not easily determined analytically. E_{ij} can be computed numerically in an iterative fashion[1], with $E_{ij} \equiv E_{ij}^{(\infty)}$ and the recursive definition $W_j/E_{ij}^{(t+1)} = \sum_l w_{jl}/[E_{il}^{(t)} + w_{jl}^{-1}]$ (with the constraint that $E_{ii}^{(t)} = 0$ continually enforced). In this paper, the iteration is halted when $\max_{ij} |E_{ij}^{(t+1)} - E_{ij}^{(t)}| \leq \epsilon = 0.005$. The method also requires an initial guess, $E_{ij}^{(0)}$, with the simple initial guess throughout this paper that $E_{ij}^{(0)} = 1$ (the method is robust to variations in this initial value).

Eq. 1 is only one way of satisfying the expectations shown in Fig. 1, and there is a great deal of functional freedom in satisfying these constraints. For example, any measure $E_{ij}^{(g)}$ of the form $W_j g(E_{ij}^{(g)}) = \sum_k w_{kj} g(E_{ik}^{(g)} + w_{kj}^{-1})$ will satisfy the desired behavior depicted in Fig. 1 for a monotonically decreasing $g(x)$, with $g(x) = x^{-1}$ in the definition of Eq. 1. Another alternate definition will satisfy a triangle inequality at the cost of additional computational complexity by replacing the direct connection strength, w_{kj} , with the closeness, \tilde{E}_{kj} , in the denominator: $W_j/\tilde{E}_{ij} = w_{ij}^2 + \sum_{k \neq i} w_{ij}/(\tilde{E}_{ik} + \tilde{E}_{kj})$. While these alternate definitions may be of interest in certain contexts, we continue to use Eq. 1 throughout this paper, due to its simplicity and previously demonstrated successes in prediction algorithms[1] and community detection methods[12]. Variations in the definition of E_{ij} will certainly change the numerical values of the closeness, but the qualitative behavior of the closeness between nodes is expected to be robust to perturbations of the definition of the GENs.

2 The GENs in homogeneous networks

While Eq. 1 is not exactly solvable for all but the simplest of network topologies, the general properties of the GENs can be explored for sufficiently homogeneous networks. The unweighted Erdős-Rényi (ER) networks have a degree distribution sharply peaked about the average ($k_i \approx \langle k \rangle$, where k_i is the degree of the node i in an unweighted network), and we expect the closeness between nodes will still be broadly distributed due to the complex network topology. The average closeness between nodes can be derived by assuming that $E_{ij} = E_c$ (the ‘typical connected’ closeness) if i and j are connected, and the ‘typical disconnected’ closeness, $E_{ij} = E_d$, if they are not directly connected. In a unweighted regular network, with all nodes having the same degree $k_i = k$, it is possible to examine the average closeness between connected and disconnected nodes using the GENs. For homogeneous degree distributions such as the ER networks, we expect an approximation $k_i \approx \langle k \rangle$ to be reasonable, with fluctuations in the degree expected to have a relatively minor impact, particularly for high mean degree. For these homogeneous networks, we assume that nodes that are directly connected feel a typical closeness E_c between each other, and another closeness $E_d \geq E_c$ to nodes they are not. If i and j are directly connected they have on average $(k-1)^2/(N-2)$ neighbors in common (since both have exactly k edges, one of which connects to the other), and they have $k^2/(N-2)$ neighbors in common if they are not connected. We must split the sum in Eq. 1 into two parts: a sum over nodes neighboring both i and j , and a sum over nodes only connected to j . This gives the approximate equations for an unweighted network of constant degree

$$\frac{k}{E_c} \approx 1 + \frac{(k-1)^2}{N-2} \frac{1}{E_c+1} + \left(k-1 - \frac{(k-1)^2}{N-2} \right) \frac{1}{E_d+1} \quad (2)$$

and

$$\frac{k}{E_d} \approx \frac{k^2}{N-2} \frac{1}{E_c+1} + \left(k - \frac{k^2}{N-2} \right) \frac{1}{E_d+1}. \quad (3)$$

It is possible to solve E_c exactly in terms of k , N , and the unknown E_d , with

$$E_c = \frac{2 + kE_d^2 - N}{-2 + kE_d + N}. \quad (4)$$

An exact solution for E_d is unwieldy, but it is possible to find a solution for E_d asymptotically for large N and we find that

$$E_d \sim \left(\frac{(k+1)N}{k} \right)^{\frac{1}{2}} + \frac{k}{k+1} + O(N^{-1/2}). \quad (5)$$

Comparing this expression to the numerical solution of the equation shows less than 1% deviation for $N \sim 1000$ and $k \lesssim 300$, suggesting the truncation to terms of order $O(N^0)$ is

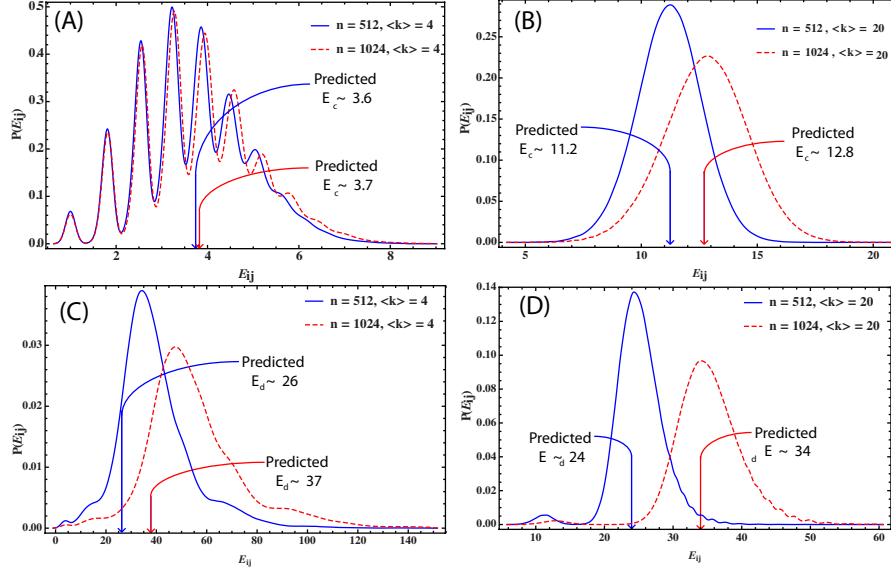


Figure 2: The distribution of E_{ij} split into cases where i and j are directly connected in (A,B) and not directly connected in (C,D) for Erdős-Rényi networks with $N = 512$ or 1024 and $\langle k \rangle = 4$ (A,C) or $\langle k \rangle = 20$ (B,D). Note the changing axes in all figures. The predicted average of E_c and E_d are marked using the same color schemes as in the figures (where the predictions include the higher-order terms found in Appendix A). For $\langle k \rangle = 20$, there is excellent agreement between the theoretical and simulated averages. For $\langle k \rangle = 4$, the GENs are far more heterogeneous than can be captured using the simple model in Eqs. 5 and 6, and the theoretical predictions do not agree well with the observed behavior for both connected and disconnected nodes.

sufficient for large N over a wide range of k . A good approximation for E_c can be found by setting $k = \kappa N$ and taking the limit of $\kappa \rightarrow 0$. We find

$$E_c \approx \frac{k + 2\sqrt{\frac{k^3}{N(k+1)}}}{1 + \sqrt{\frac{k(k+1)}{N}}} \sim k + O(k^2 N^{-\frac{1}{2}}). \quad (6)$$

where latter is the scaling for sufficiently large N . Even for $N \sim 1000$, higher order terms can contribute in the series for only moderate values of k , and the full expression is required to obtain an accurate estimate.

In Fig. 2, we see the distribution of the GENs for Erdős-Rényi networks with varying N and $\langle k \rangle$. In (A-B) we see that changing $\langle k \rangle$ radically alters the mean values of E_{ij} as well as the shape of the distributions, while changing N only marginally affects the distribution of the

connected GENs. For $\langle k \rangle = 4$ the distribution of E_{ij} exhibits multiple peaks in Fig. 2(A), with each local maximum corresponding to a different degree of the node j and with the width of the distribution about the peak coming from differing degrees of the node i . Such heterogeneity is less apparent for high-degree nodes (Fig. 2(B)), where fluctuations in the degree of i or j have less of an impact on the GENs, and the distributions are unimodal. For disconnected nodes, the distributions have a single dominant peak (Fig. 2(C-D)), and the location of the peaks are well predicted for $\langle k \rangle = 20$. Due to the heterogeneity and significance of degree fluctuations for the smaller $\langle k \rangle = 4$, there are large differences between the predicted and observed averages.

In contrast to the homogeneous degree distribution of the Erdős-Rényi random network model, Barabasi-Albert (BA) networks[8] have a scale-free, heterogeneous degree distribution, and Fig. 3 shows that the distribution for the GENs for BA networks are likewise heterogeneous for directly connected nodes. The distribution for the GENs between nodes that share an edge (shown in Fig. 3A-B) appear to have a relatively fat tail and approximately satisfy $Pr(E_{ij} = E) \sim E^{-\lambda}$ for nodes that share a direct connection, with an empirically determined scaling exponent near 1.5 for $\langle k \rangle = 4$ and around 2.1-2.2 for $\langle k \rangle = 20$ (shown in Fig. 4). This is in comparison to the fat-tailed degree distribution with the $P(k) \sim k^{-3}$ scaling of the BA networks for both values of $\langle k \rangle$. Interestingly, the distribution of the GENs for disconnected nodes does not depend as strongly on the scale-free nature of the degree distribution, with similar qualitative features found in both Fig. 2C-D for the ER networks and Fig. 3C-D for the BA networks. While the existence of hubs in the BA networks tends to give a higher probability of finding smaller values of E_{ij} for disconnected nodes in comparison to ER networks, the most likely values of E_{ij} are similar for disconnected nodes in either network topology (in contrast to the radically different distributions in for connected nodes). We have considered only unweighted networks in this analysis, and allowing weighted edges further complicates the analysis of the ‘typical’ GEN between nodes unless a homogeneity assumption on the distribution of weights is likewise made.

3 Random Walks and the GENs

The mean first passage time (MFPT) from node i to node j (τ_{ij}) are of interest in many contexts, and because of the limited resource represented by the random walker, it is worthwhile to see the relationship between the rate of travel between nodes and how ‘close’ they are as measured by the GENs. Tetali[36] has shown that the MFPT in an unbiased random walk can be computed directly from the resistance distance[15, 16] R_{ij} with $\tau_{ij} = \frac{1}{2} \sum_l k_l (R_{ij} + R_{jl} - R_{il})$. It is easily seen that the MFPTs are asymmetric ($\tau_{ij} \neq \tau_{ji}$), in general, as it is easier to reach a high-degree node than a low degree node (much like the asymmetry in the GENs with $E_{ij} \neq E_{ji}$). We intuitively expect that if node j feels

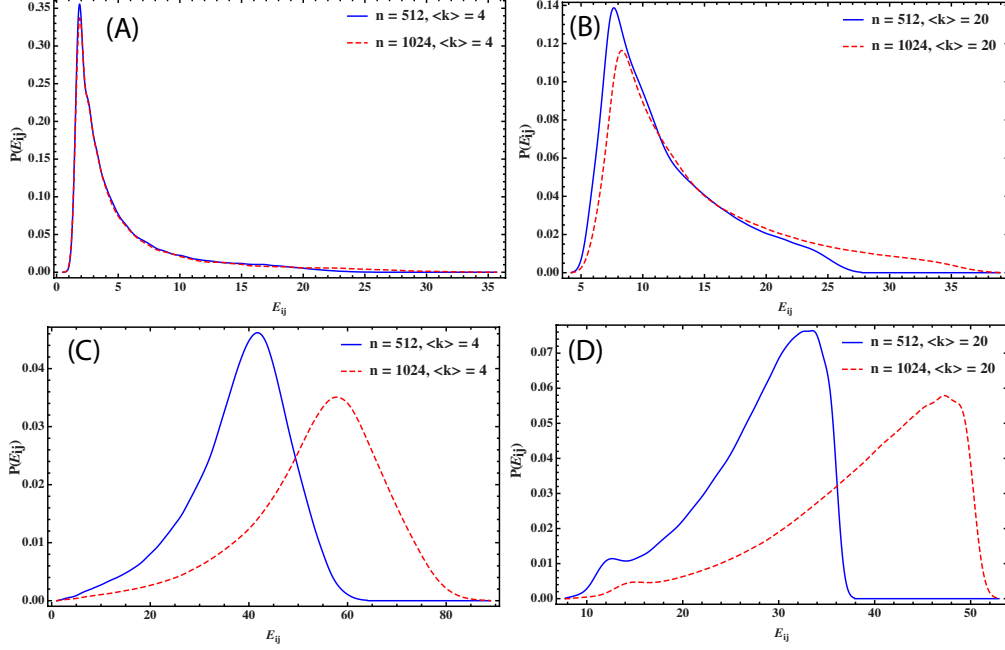


Figure 3: Distributions of the GENs for the Barabasi-Albert networks for nodes that do share a direct connection (A,B) and do not share a connection (C,D) for $\langle k \rangle = 4$ (A,C) and $\langle k \rangle = 20$ (B,D) and with $N = 512$ (blue solid lines) or 1024 (red dashed lines). The distributions are far smoother than those seen in Fig. 2 for small $\langle k \rangle$ due to the more heterogeneous degree distribution of the BA networks. The behavior is consistent with the ER networks: the mean values depend strongly on $\langle k \rangle$ and weakly on N if the nodes share a direct connection, while the opposite is true if the nodes are not neighbors. Because the fat tail of the degree distribution (with $P(k) \sim k^{-3}$) provides a broader distribution to the connected GENs distribution than in the ER case (Fig. 2), the mean values are further from the peaks of the distribution and are not indicated in the figure.

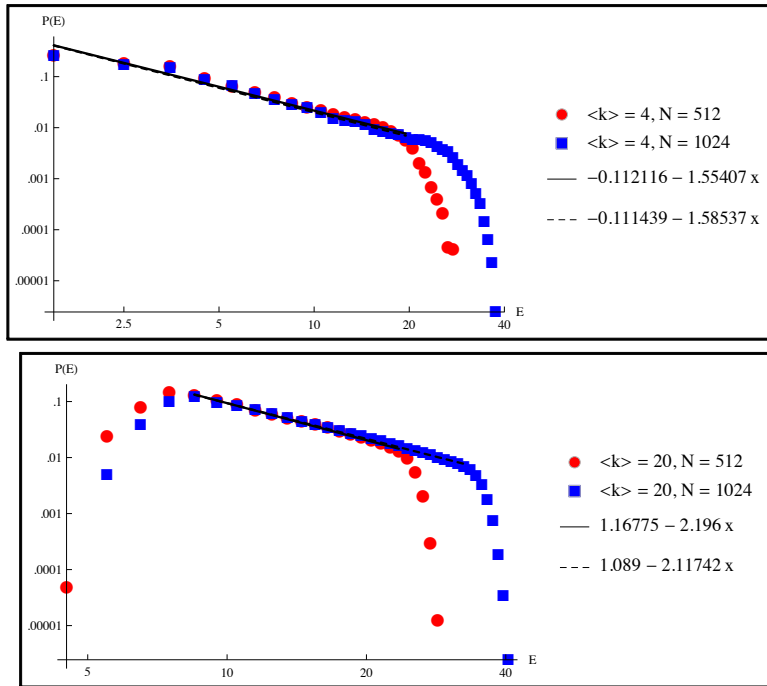


Figure 4: Fitting the heavy tail of the distribution of the nearest-neighbor GEMs for the Barabasi-Albert networks in Fig. 3 (A-B). Over a wide range of values, there is an approximate power law decay which is very weakly dependent on N ($N = 512$ in the red circles and $N = 1024$ in the blue squares) but does depend on the average connectivity ($\langle k \rangle = 4$ shown above and $\langle k \rangle = 20$ below). This is consistent with the weak N dependence seen for the Erdős-Rényi networks in Eqs. 5 and 6.

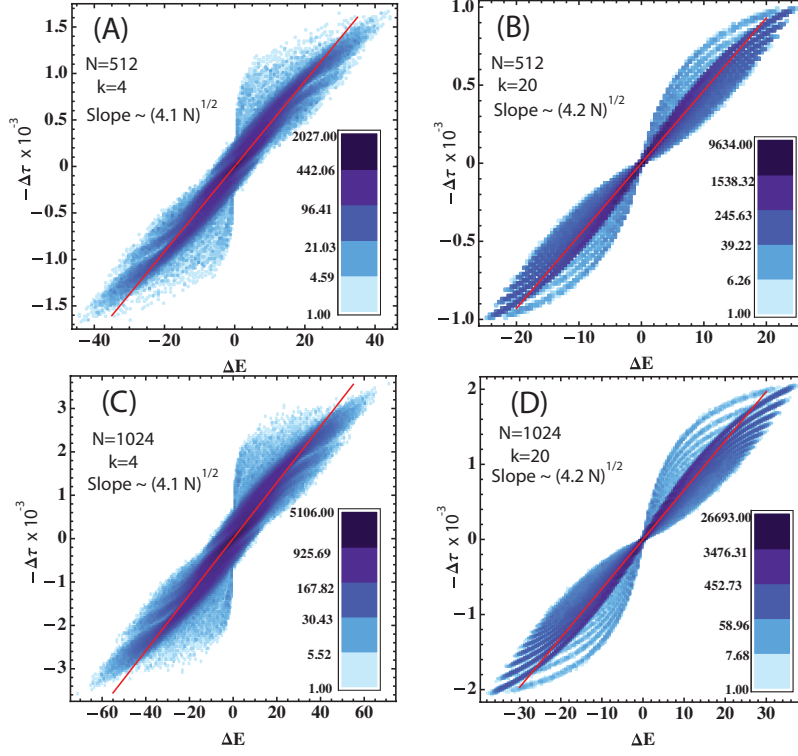


Figure 5: Asymmetry in the Barabási-Albert GENs $\delta E_{ij} = E_{ij} - E_{ji}$ compared with the asymmetry in the MFPTs for those networks, $\delta\tau_{ij} = \tau_{ij} - \tau_{ji}$. In these density plots, darker colors correspond to a greater observed frequency of the same $(\delta E, \delta\tau)$ pair. Shown are two values of $N = 512$ and 1024 , as well as two values of $\langle k \rangle = 4$ and 20 . It is visually clear that the asymmetry of the GENs are highly correlated to the asymmetry in the MFPT (with the slope of the best fit line indicated).

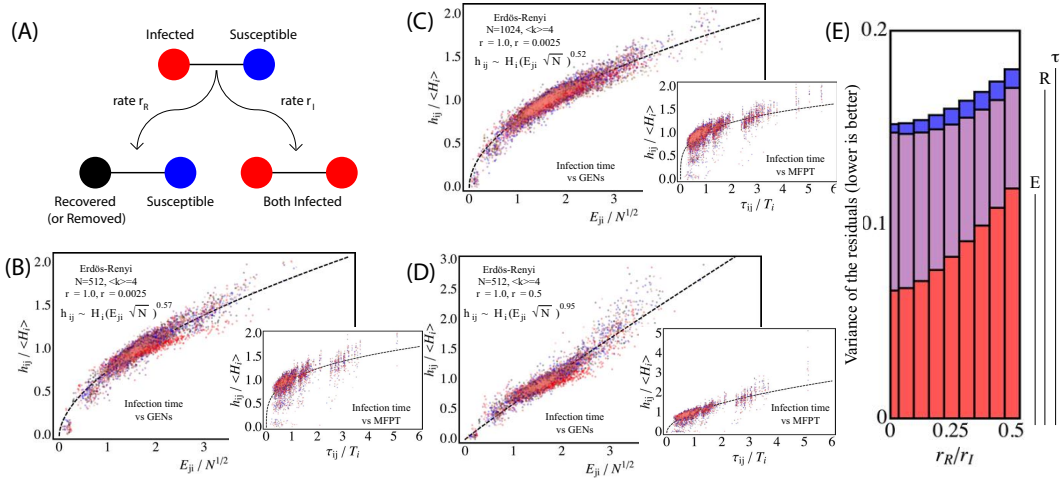


Figure 6: The harmonic mean of the infection time of node j with a single initially infected node i , h_{ij} in an Erdős-Rényi network. The SIR model is diagrammed in (A). The main panels of (B-D) compared h_{ij} with the GENS E_{ji} while a comparison with the MFPT in a random walk τ_{ij} is shown in the inset, with different colors denoting a different initially infected node i . In all, h_{ij} is scaled by the average infection time of the nodes, H_i , and likewise the MFPT is scaled in terms of $T_i = (N-1)^{-1} \sum_{i \neq k} \tau_{ik}$. (B) shows $N = 512$ nodes with $r_R = 0.0025R_I$. (C) $N = 512$ and $r_R = 0.5r_I$. (D) $N = 1024$ and $r_R = 0.0025r_I$, all with $\langle k \rangle = 4$. A functional relationship between h_{ij}/H_i and E_{ij} is immediately apparent in (B-D), which corresponds well to a simple power law for $r_R \ll r_I$, while the insets show that the MFPT is a poor predictor of average infection time. The dashed lines show the best power law fit for the data, with the scaling indicated in the figure. The functional behavior is not universal, depending on the size of the network and the SIR model parameters. (E) shows the standard deviation of the residuals as a function of r_R/r_I for a power law fit $h_{ij}/H_i \sim f(x_{ij})$ for $x_{ij} = E_{ij}$ (red), R_{ij} (purple), and τ_{ij}/T_i (blue).

‘close’ to node i (small E_{ij}) but node j does not (large E_{ji}), a random walker will pass more readily from j to i than from i to j . We therefore compare $\delta E_{ij} = E_{ij} - E_{ji}$ to the difference in the MFPT between nodes $\delta\tau_{ij} = \tau_{ij} - \tau_{ji}$ in Fig. 5 and empirically find the asymmetry in the MFPT is highly correlated with the asymmetry in the GENs, with an empirical scaling of $\delta\tau_{ij} \sim -\delta E_{ji}\sqrt{\alpha N}$ with N the number of nodes in the network and $\alpha \sim 4$ a topology-dependent constant. The fact that $\delta\tau_{ij} \propto \delta E_{ji}$, even when there are no direct connections between i and j , indicates that the GENs are able to capture the importance of the global network topology even for distant nodes. It is interesting to note that while the relative proportionality appears to be strongly dependent on the size of the network and only weakly dependent on the average degree, it is visually apparent that $\langle k \rangle$ sets the scale of the fluctuation statistics from the best fit line. For $\langle k \rangle = 4$, the asymmetry density is relatively disperse (Fig. 5(A) and (C)) with no obvious structure, while for $\langle k \rangle = 20$ there is clear clustering of the density about curved bands (Fig. 5(B) and (D)).

A similar result, using the same values of N and $\langle k \rangle$ for a Erdős-Rényi network, is shown in Fig. 7 and displays the same qualitative features: the asymmetry in the GENs are highly correlated with the asymmetry in the mean first passage times, and there is an apparent scaling of $\delta E_{ij} \sim \delta\tau_{ij}/\sqrt{\alpha_{ER}N}$ with $\alpha \sim 4 - 7$ depending on the average degree. This coefficient is on the order of the scaling coefficient for the BA networks in Fig. 5, with $\alpha_{BA} \sim 4$. While the structure of the deviation from the best fit line is interesting, note that the logarithmic scale of the coloring means that these extreme outliers are still relatively rare compared to the much higher density near the best fit line. The overall correlation between the asymmetries of the GENs and the MFPTs indicates that δE is a meaningful quantity, capturing the topological details of the network. We have found that our iterative method for computing the GENs (implemented in C++) converges far more rapidly than directly computing the MFPT via a matrix pseudo-inversion[15] for the networks we have considered (using the ‘pinv’ function in Matlab).

4 An application to epidemic spreading

The spreading of an epidemic has been studied by many authors and in a wide range of contexts[37, 18, 38, 39, 17], with the SIR model being one of the simplest and most commonly used. The susceptible-infected-recovered (SIR) model assumes that a population of susceptible individuals becomes infected due to interactions with previously infected individuals, and infected individuals may recover and become non-infectious. A simple schematic of the SIR model is shown in Fig. 6(A), with infections occurring at a constant rate, r_I , when individuals interact and the recovery at constant rate, r_R . A number of more complex models have been considered extensively for a homogeneously mixed population of individuals [39], but non-uniform interactions between individuals, represented by

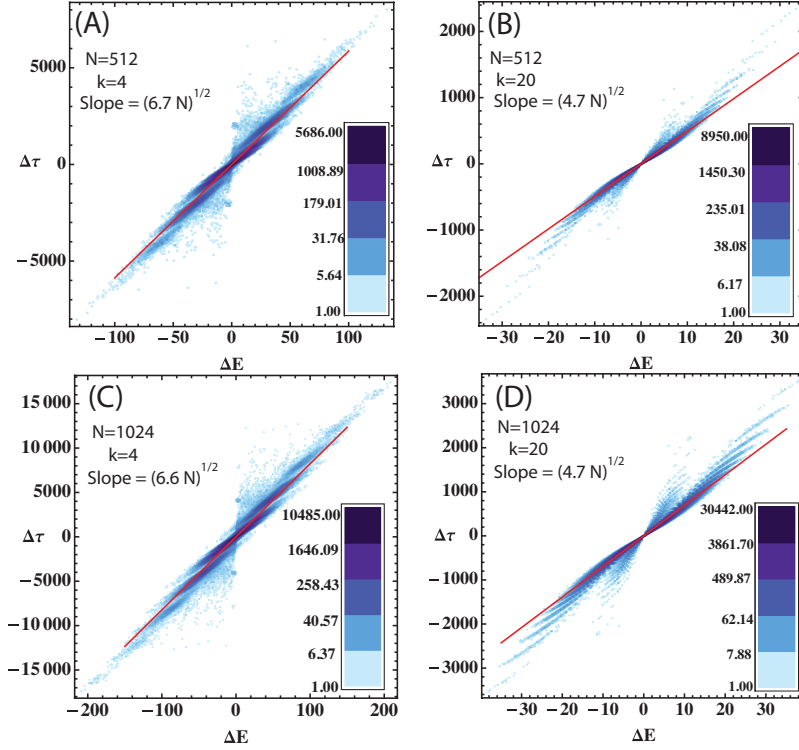


Figure 7: Asymmetry in the Erdős-Rényi GENs $\Delta E_{ij} = E_{ij} - E_{ji}$ compared with the asymmetry in the MFPTs for those networks, $\Delta\tau_{ij} = \tau_{ji} - \tau_{ij}$. The figure labels are identical to those in Fig. 5, with a similar behavior in scaling and variation from the best fit line.

networks, can have a profound impact on the dynamics of epidemic spreading in the SIR model[5, 17, 18]. The existence of epidemic thresholds[5, 40] for homogeneous networks (or the lack thereof for scale free networks[17]) are well-studied global quantities of interest, while more local quantities such as the probability of a particular node i becoming infected, sparking an epidemic[41], and quarantine or immunization strategies[38, 42] have also been examined.

While it is clearly useful to understand the global properties of the epidemic (such as the expected number of infected individuals), a particular individual j may also be interested in its own probability of becoming infected given the current state of the disease and may reasonably be less concerned if no neighbors are infected than if many neighbors are infected. However, it is not straightforward to analytically calculate how long the disease will take to reach j from any point in the network, and it would be useful to have a measure for how ‘close’ the epidemic is from an individual node. If the infection begins with a single node i , we expect that the disease will more rapidly propagate to nodes for which i feels topologically close, and it is therefore worthwhile to compare the infection times in a SIR epidemic with the GENs as a proxy for closeness. To see the relationship between infection time and closeness, we simulate an SIR epidemic, using Gillespie dynamics[43] on an Erdős-Rényi graph (with a uniform probability of connection and each node having $\langle k \rangle = 4$ average degree). This allows us to determine the harmonic mean of the infection time of a node j given an initial infection at i over the K simulations, $h_{ij}^{-1} = \sum_{k=1}^K [t_{i \rightarrow j}^{(k)}]^{-1}$ (the harmonic mean is chosen to avoid diverging infection times in simulation k , $t_{i \rightarrow j}^{(k)}$, where the disease dies out before j is infected). Because the GENs do not naturally include the timescales of the system dynamics (r_I and r_R) we normalize the harmonic mean by $H_i = \frac{1}{N-1} \sum_{j \neq i} h_{ij}$, which is the ‘typical’ time at which a node becomes infected if the disease originates with i .

In Fig. 6(B-D) we compare h_{ij}/H_i to E_{ji} (how close node i feels towards node j) for a network of 512 nodes and $r_R = 0.0025r_I$ in (B); $N = 1024$ and $r_R = 0.0025r_I$ in (C); and $N = 512$ and $r_R = 0.5r_I$ in (D). In all cases, h_{ij}/H_i tends to increase with E_{ji} (as is expected: nodes i feels close to are infected faster), although the functional dependence of the infection time on the topological closeness depends on the parameters in a non-universal way. We also show in the insets that the scaled mean first passage time between nodes in a random walk (τ_{ij}/T_i) does a surprisingly poor job of predicting the infection time in comparison with the GENs. This is likely due to the significant differences between the SIR model and a random walk: if a walker departs from node i at timestep t , he cannot depart from i again at time $t + 1$. In effect, there is only one ‘infectious’ node at a time in a random walk which makes the MFPT a poor estimator of the infection time in a SIR simulation. The resistance distance between nodes R_{ij} provides a fit quality similar to that of the scaled MFPT (data not shown). In Fig. 6(E), the quality of the fit is quantified using the standard deviation of the residuals of the fit. In all cases, the GENs are much

more highly correlated with the average time of infection between pairs of nodes in the network. Fig. 6 indicates that the GENs E_{ij} can more meaningfully capture the impact of network topology on the dynamics of epidemic spreading than other global measures of pairwise ‘closeness’ between nodes.

5 Measuring a personalized importance

The GENs incorporate a simple idea of what is meant by the ‘closeness’ between nodes in a network where limited resources are shared, and we expect that a node j that feels ‘close’ to node i (having small E_{ij}) considers node i to be ‘important’ in some sense. We may therefore regard the inverse of the closeness between nodes ($\psi_{ij} = E_{ij}^{-1}$) as an un-normalized personalized measure of importance, allowing a ranking of all nodes in the network from the perspective of the node j . Because ψ_{ij} measures the importance of i from a particular node j (rather than the network at large), it is not equivalent to a centrality measure.

To gain insight into the meaning of the personalized importance, Eq. 1 can be expanded in the limit of large E_{ij} for $i \neq j$ (a valid expansion for $k_i = k$ shown in the SI, and is observed to be reasonable for more complex networks), yielding $E_{ij}^{-1} \approx W_j^{-1}[w_{ij}^2 + \sum_{i \neq l \in C_j} w_{il} E_{ij}^{-1}] + O(E_{ij}^{-2})$. We can use this lowest-order expansion to define the linearized importance assigned by node j towards node i (φ_{ij}) as

$$\varphi_{ij} \equiv \frac{w_{ij}^2}{W_j} + \frac{1}{W_j} \sum_{i \neq l \in C_j} w_{jl} \varphi_{il}, \quad (7)$$

where φ_{ii} is undefined (since $E_{ii} = 0$). Eq. 7 provides a natural interpretation of the meaning of personalized importance. The importance j assigns to i is a combination of two terms: a weighted average of the importance his neighbors assign to i , and an importance of w_{ij} assigned in the case of a direct connection between i and j . Defining $(\boldsymbol{\varphi})_{ij} = \varphi_{ij}$ and $(\mathbf{L})_{ij} = W_j \delta_{ij} - w_{ij}$ the graph Laplacian[31, 44, 15], Eq. 7 can be written $(\boldsymbol{\varphi} \mathbf{L})_{ij} = w_{ij}^2 - w_{ij} \varphi_{ij}$, relating the linearized importance directly to the graph Laplacian. Despite the simplicity of this interpretation of the linearized importance, φ_{ij} is of limited use in unweighted networks (diagrammed in Fig. 8: if $w_{ij} = \{0, 1\}$, Eq. 7 reduces to $k_j \times \varphi_{ij} = \sum_{l \in C_j} \varphi_{il}$ with the solution $\varphi_{ij} = 1$ for all connected nodes. Only in the context of weighted networks is a non-trivial measure of importance obtained. The full nonlinear expression for the GENs (which incorporates higher order terms than the linear approximation) does not suffer from this difficulty and can still determine meaningful measures of importance between nodes.

The usefulness of the nonlinear importance $\psi_{ij} = E_{ij}^{-1}$ on a network can rapidly determine meaningful relationships between nodes in complex networks. To illustrate this,

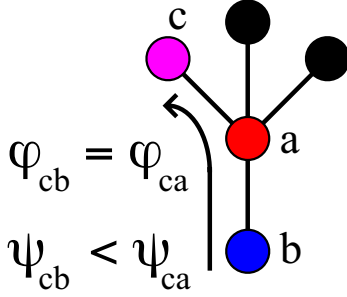


Figure 8: In this simple network, $\varphi_{cb} = \varphi_{ca}$, meaning the importance node b assigns to c (with no direct connection between them) is identical to the importance a assigns to c (which share a direct connection). Intuitively, we would expect the lack of direct connection to imply b feels c is less important (due to the lack of direct connection $w_{bc} = 0$). This condition is satisfied using the nonlinear importance, with $\psi_{cb} < \psi_{ca}$.

we consider the benchmark of Liancichinetti, Fortunato, and Radicchi (LFR)[11], which constructs a network of communities of variable sizes n (distributed as $P(n) \propto n^{-\beta}$), a scale-free distribution of the nodes (with $P(k) \propto k^{-\gamma}$), and which is characterized by the mixing parameter, μ , as the fraction of inter-community edges. We have previously shown[12] that the GENs can be used to detect the community structure underlying this benchmark. When measuring the importance of a node, a global measure of centrality will generally focus on nodes with high degree, but due to the heterogeneous density of edges between communities, we expect a meaningful definition of the importance j assigns to i to differ significantly depending on if i and j are in the same community. Note that the determination of the GENs does not require *a priori* knowledge of the community structure. In Fig. 9(A), we determine the distribution of importance ψ_{ij} between nodes i and j that do not share a direct connection ($w_{ij} = 0$) for nodes within i 's community (red) and outside of i 's community (blue). There is an immediately apparent difference in the distributions, with a greater probability of a high importance if i and j are in the same community due to the increased number of shared neighbors (even in the absence of a direct connection). Increasing the LFR parameter μ (which increases the number of edges between communities) reduces the difference in the distributions, but varying the other system parameters has only a minor impact on the clear distinction between the two distributions (data not shown). A large overlap between neighbors is not required to accurately detect meaningful differences in the importance between pairs of nodes, though: in Fig. 9(B) we show cliques of n nodes with A , B , and C having no neighbors in common, but due to the network topology, A is closer to B than to C (depicted in the inset of Fig. 9(B)). The GENs determine that B feels A is more important than C does. ψ_{ij} incorporates direct connections between nodes, proximity in the network, and shared neighbors into a meaningful measure of personalized importance.

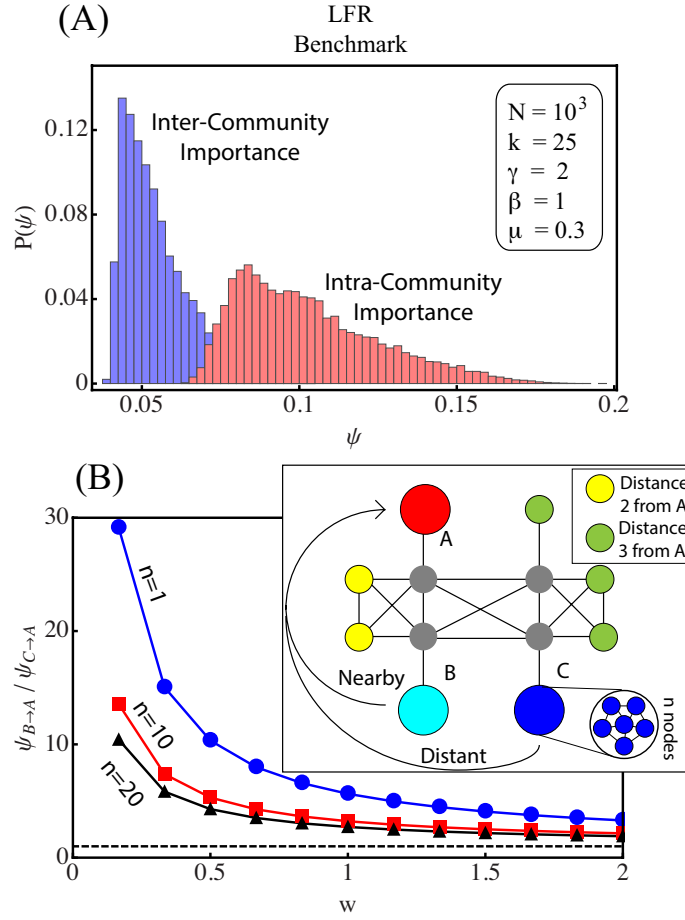


Figure 9: (A) The GENs applied to the LFR benchmark[11]. The red shows the distribution of importance for nodes i and j that are in the same community but do not share a direct connection. The blue shows the distribution for those in different communities (and still sharing no direct connection). Due to the high density of links inside of the communities, the GENs accurately indicate that ψ_{ij} is likely to be larger if i and j are in the same community. (B) In the inset, we diagram a simple network topology where cliques are in greater proximity to one another but still have no neighbors in common. Each circle represents a clique of n nodes (fully connected with each edge having unit strength) and each edge an all-to-all connection between cliques with each edge having weight w . Clique B is in close proximity to auxiliary nodes with integer distance 2 from A (highlighted in yellow) while clique C is in closer proximity to auxiliary nodes of integer distance 3 (highlighted in green). In the main figure, ψ_{ij} indicates that A is more important from the perspective of B than from C for varying values of n and w due to the closer proximity of B's neighbors to A than C's neighbors, with $\psi_{AB}/\psi_{AC} > 1$.

k_i	Degree[45]	Node Degree
Pr_i	PageRank[19]	i^{th} component of first eigenvector of \mathbf{B}
b_i	Betweenness[16]	times i is visited in a random walk
	Centrality	$\propto \sum_{j \in C_i} \sum_{l, m \neq i} R_{il} + R_{jm} - R_{im} - R_{jl} $.
c_i	Random Walk [14]	The rate information flows from other
	Centrality	nodes to i : $c_i^{-1} - c_j^{-1} = \tau_{ji} - \tau_{ij}$

Table 1: Some common methods of measuring the importance of a node i . These are compared directly to the Erdős Centrality $\Psi_i = \sum_{l \in C_i} E_{il}^{-1}$ in Fig. 4 of the main text. This list is not exhaustive, and other measures of centrality are possible[45, 46]. R_{ij} denotes the resistance distance[15, 16] between nodes i and j , τ_{ij} the mean first passage time between i and j in a random walk, and $\mathbf{B}_{ij} = (1 - d)w_{ij}/W_j + d/N$ is the matrix of transition probabilities for PageRank (where d is the teleport probability of visiting a non-neighbor in PageRank). Throughout the paper, each measure of centrality is normalized such that the sum of all centralities in the network is 1.

6 Global importance and Erdős centrality

Having defined a pairwise measure of the importance a node j assigns to i using ψ_{ij} (or the linearized φ_{ij}), we naturally expect that we can leverage this definition into a global measure of the importance of node i . There already exist a wide variety of methods for measuring centrality from a global perspective, including degree[45, 16, 46], PageRank[46, 19], random walk[14], and betweenness centrality[14, 16] (briefly described in Table 1). Each measure ranks high-degree nodes above low-degree nodes but take the global network topology into account in different ways. These methods produce qualitatively similar but quantitatively different node rankings, as reflected by the fractional intersection between the top- n orderings[47], $\sigma_{XY}(n) = \frac{1}{n} \sum_k |\mathbf{o}_X(k) \cap \mathbf{o}_Y(k)|/k$, with $\mathbf{o}_X(k)$ the top- k ordering using method X . $\sigma_{XY}(n)$ is shown comparing the ordering due to PageRank with that found using the betweenness and random walk centrality measures for a Barabási-Albert network topology in Fig. 10(A). It is clear that there is an immediate drop to $\sigma \gtrsim 0.95$ for small n for these well known centralities (i.e. good but not perfect agreement on the top few nodes), after which the methods tend to vary slowly or not at all for larger top- n lists.

To convert our personalized importance measures into a single global measure, we define $\Psi_i = \sum_{l \in C_i} \psi_{il}$ as the sum of the importance the neighbors of i assign to it (akin to the approach of Ref. [32]). For an unweighted network (with $\varphi_{ij} = 1$ for all i and j), the linearized $\Phi_i = \sum_{j \in C_i} \varphi_{ij} = k_i$ is equivalent to the degree centrality. In Fig. 10(B), we compare Ψ_i to the other measures of centrality in Table 1 for a single realization of a Barabási Albert network with $N = 512$ and $\langle k \rangle = 20$. In all cases, there is an obvious correlation between these measures of centrality but with significant difference between

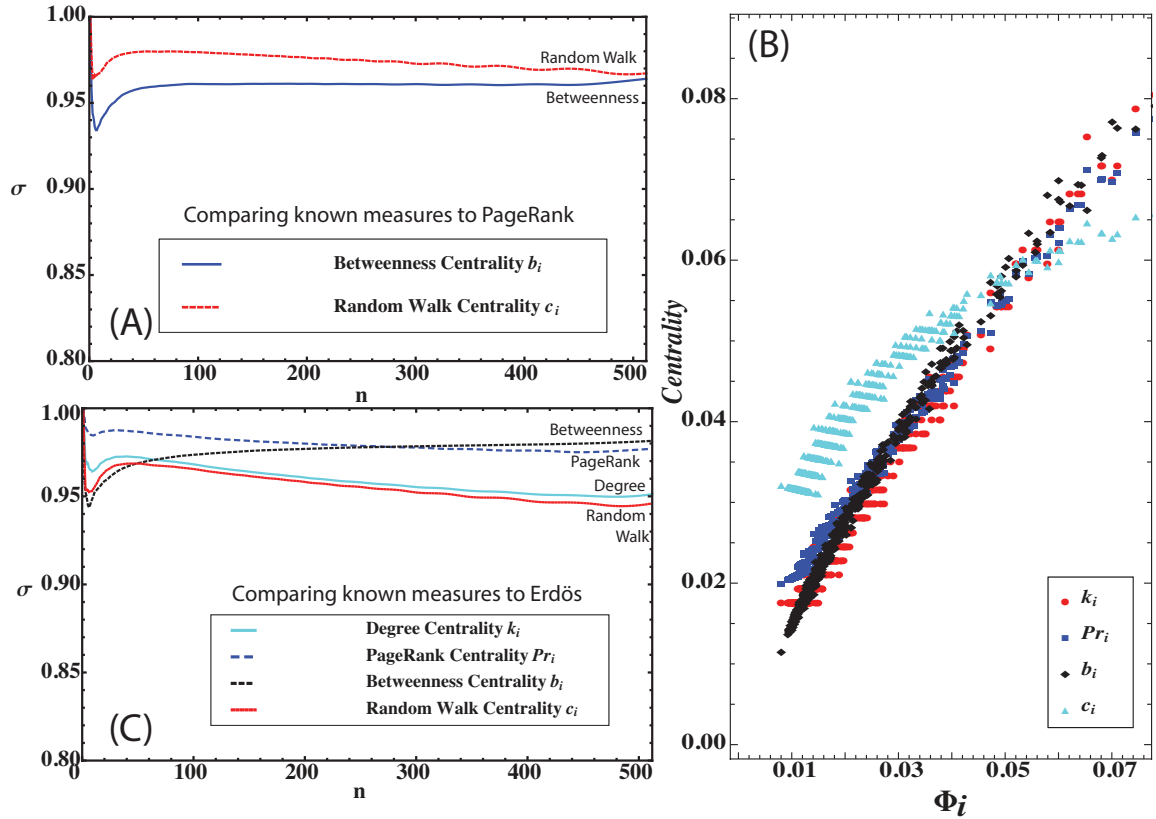


Figure 10: Centrality for a Barabási-Albert network with $\langle k \rangle = 20$. (A) uses the intersection metric $\sigma_{XY}(n)$ to estimate the similarity in top- n lists between PageRank and two centrality measures listed in Table 1 in the Supplementary information. While there is good correspondence between these measures, each measure gives a slightly different ordering for small n . (B) The Erdős centrality (x-axis) compared to the 4 common centrality measures (y-axis) shows an obvious positive correlation overall. Circles shows degree centrality, squares PageRank, diamonds betweenness centrality and triangles random walk centrality. The clustering of some data near discrete values is due to the heterogeneity of the Erdős centrality for nodes of equal degree (this effect is more pronounced in Fig. 11 with $\langle k \rangle = 4$). (C) shows the intersection metric between the top n elements of the Erdős centrality ($\sigma_E(n)$) and the top n elements of the other centrality measures for varying n . The GENs are about as consistent with known measures of centrality as these measures are amongst themselves (shown in (A)).

centrality measures in some cases for both central and non-central nodes alike. Degree and random walk centrality both tend to be strongly clustered around discrete values, leading to the banded structures seen in Fig. 10(B), and the difference between these and PageRank seen in Fig. 10(A). The Erdős centrality takes the global topology into account differently than these measures of centrality (similar to the betweenness and PageRank centrality). When comparing the Erdős centrality to those in Table. 1, the similarity between the orderings from most- to least-important[48, 49], using different measures is of particular interest[20, 21]. We compare the Erdős centrality ordering to the other measures of centrality using $\sigma_{XY}(n)$ [47] in Fig. 10(C) and see that the Erdős centrality is consistent with other measures of centrality: a sharp drop initially with slow variation for larger n and $\sigma(n) \gtrsim 0.95$ throughout. Despite the different formulations between the Erdős centrality and PageRank, the top- n list for Ψ_i compares best to the list from Pr_i (dashed purple line) for high degree nodes but begins to agree better with betweenness centrality when less central nodes are also included.

We have also computed the Erdős centrality and other measures of centrality for a BA network with $N = 512$ and $\langle k \rangle = 4$. While the discreteness of the degree distribution is still apparent in Fig. 10B (with the tight clustering of points about discrete values), it has a greater impact in Fig. 11, where k attains fewer values due to the significantly smaller average degree. The general trends are qualitatively similar to that in Fig. 10, though: the betweenness centrality (Fig. 11(C)) depends on the degree in a qualitatively different manner than any other measure (and is more clearly correlated to the Erdős centrality for low-ranked nodes). The Erdős centrality does not appear to be linearly related to the random walk centrality (although they are still correlated), and clear bands are observed due to nodes of equal degree for both the degree and PageRank measures of centrality.

It is interesting to note that ψ_{ij} can define a modified connectivity matrix, incorporating both direct connections in the adjacency matrix and non-local paths between nodes. A random walk performed with a transition probability $\mathbf{B}'_{ji} = \psi_{ij} / \sum_{l \neq i} \psi_{lj}$ (with the convention $\mathbf{B}'_{ii} = 0$, meaning the walker never remains at i) has a similar interpretation: a walker at node j has a relatively high probability of moving to node i if they are directly connected in the original network (if $w_{ij} > 0$), but has a non-zero probability of jumping to a disconnected node. The matrix \mathbf{B}' can be compared to the PageRank transition probability matrix \mathbf{B} , which has a uniform probability of teleporting to any node in the network (regardless of the network topology). \mathbf{B}' has a jump probability that is not purely random but rather preferentially lands on nodes that j finds important, in contrast to PageRank's uniform teleport probability. An alternate Erdős measure of centrality can be defined as the largest eigenvector of the matrix \mathbf{B}' (akin to PageRank, the steady state probability of being found at a node i under the transition probability \mathbf{B}). However, the computational efficiency of the simple sum $\Psi_i = \sum_{j \in C_i} \psi_{ij}$ and its clear correlation with known measures of centrality (seen in Figs. 10 and 11) makes our current definition of the Erdős centrality

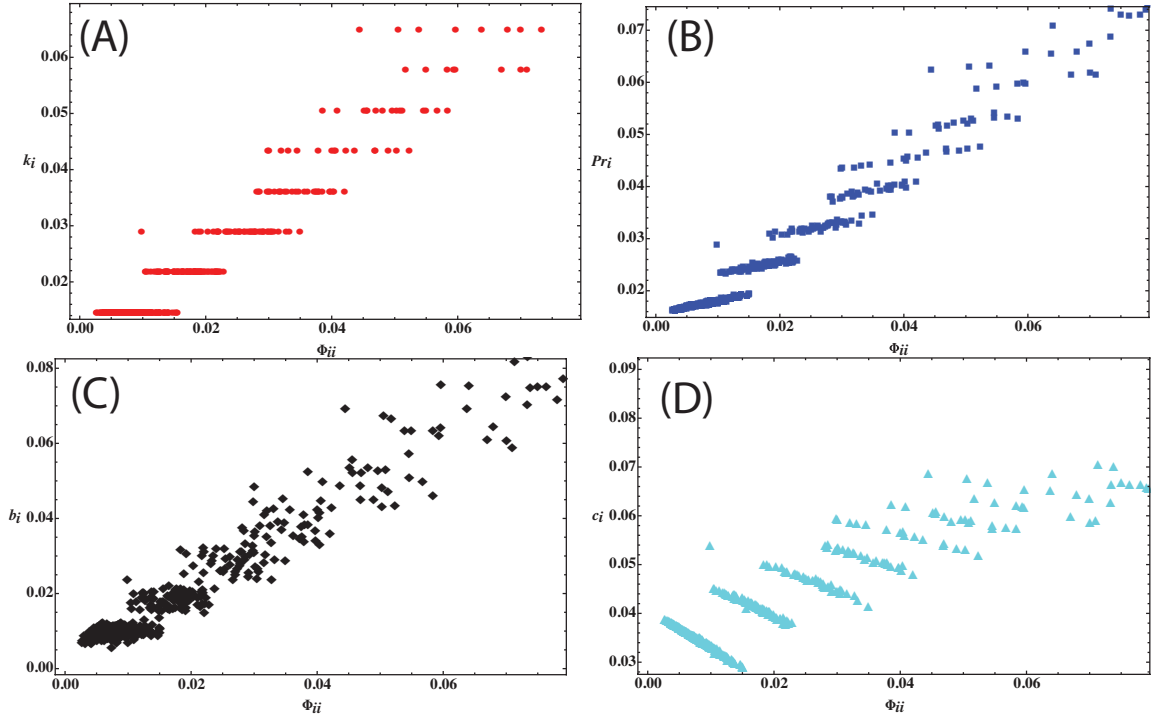


Figure 11: Comparing the Erdős centrality with other measures of centrality for Barabási-Albert networks with a lower average degree of $\langle k \rangle = 4$. The symbol color and shape are the same as in Fig. 10. There is pronounced clustering in the data due to the strong dependence the degree and PageRank measures of centrality have on the node degree, and the Erdős centrality will sometimes label a node as more important than another node with a somewhat higher degree. The overall rankings are still consistent, with high degree nodes being typically more important.

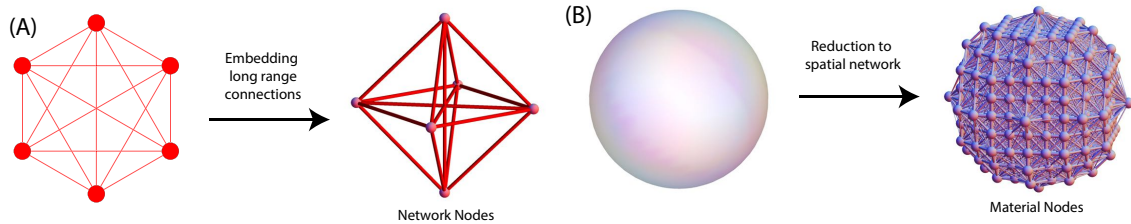


Figure 12: Schematic diagram of the continuum model with long range connections. (A) shows the reduction of the continuum to a network of material nodes, with the weight between them dependent only on their distance, $w(\mathbf{x}, \mathbf{x}') = e^{-\lambda|\mathbf{x}-\mathbf{x}'|/a}$. (B) shows the embedding of long range connections between network nodes, with weight specified by the network rather than depending on the distance between the nodes.

preferable.

Because the GENs can be used to describe random walks efficiently while prescribing a topology-dependent teleportation via the matrix \mathbf{B}' (with $\mathbf{B}'_{ij} = \psi_{ij} / \sum_{l \neq i} \psi_{lj}$ discussed in Sec. 6), they may be of use in future studies as well. Walkers on \mathbf{B}' will tend to remain trapped in regions of locally dense edges, and implementing such a walk in community detection methods may increase the accuracy of the determined partition[50]. Similarly, the modified walk may be of use in modeling systems where jumps to non-neighboring (but topologically close) nodes is desirable, such as epidemic spreading in a social network. Epidemics can be modeled[18] as spreading with some probability to direct contacts (due to the assumed high probability they interact directly) and with a lower probability ‘jumping’ to nodes without a direct contact (with non-neighbor transmission possible due to rare transient interactions). We could expect that non-neighbor interactions will not have a uniform probability, but rather that each node is more likely to transiently interact with friends-of-friends (or individuals s/he feels topologically close to). Epidemic spreading on \mathbf{B}' will capture our expectation that transmission to friends with whom we have regular contact is most likely, but transmission to friends of friends is more likely than transmission to very distant people in the social network.

7 GENs in the continuum limit

We showed in the previous sections that the Generalized Erdős Numbers can be useful in describing complex networks a variety of contexts by incorporating the global network topology into a measure of closeness between nodes. Many physical networks exist in a metric space[3], such as wireless sensor networks[24, 25] or transportation[26, 27] networks,

where the strength of the direct interaction between two nodes depends on their locations. In many contexts, such as transportation networks[6], distance-independent connection strengths may coexist with geometrically defined links (for example, surface mobility versus air travel). It is natural to wonder what impact this may have on the GENs, and how topological closeness is influenced by physical proximity. We assume that a network of long-range connections has been externally defined, and that each node in that network can be assigned to a location in some d dimensional space (a schematic is shown in Fig. 12A). Each node i in the original, distance-independent network (of N nodes) occupies a location \mathbf{y}_i , with an interaction strength w_{ij} . We refer to these nodes as ‘network nodes’, each of which has at least one direct connection due to the connectivity of the original network (referred to as a long-range connection). To introduce geometry into the problem, we also define a network of N^* ‘material nodes’ into the network at locations $\{\mathbf{x}_j\}$, whose interactions are purely geometric (schematically diagrammed in Fig. 12B for a spherical geometry). The distance-dependent interaction between the nodes of any type (both network and material nodes) at locations \mathbf{z} and \mathbf{z}' is given by $u(\mathbf{z}, \mathbf{z}')$, with the constraint that $u(\mathbf{z}, \mathbf{z}') = u(\mathbf{z}', \mathbf{z})$ to ensure the network remains undirected. Note that distance-dependent connections are drawn between network nodes as well, so the total weight between \mathbf{y}_i and \mathbf{y}_j is $w_{ij} + u(\mathbf{y}_i, \mathbf{y}_j)$. If $N^* = 0$ or $u(\mathbf{x}, \mathbf{x}') = 0$ the physical geometry is irrelevant and the original network is recovered, while if $N = 0$ or $w_{ij} = 0$, then the weights between nodes are determined only by node locations.

Replacing discrete equations with continuum models has been of use in a wide variety of systems[51], and for networks with large N^* , we can establish a continuum limit of the definition of the GENs in Eq. 1 of the main text. In SI sec. A, we develop a linearized continuum model for the approximate pairwise importance $\varphi(\mathbf{x}, \mathbf{x}')$ and describe a more complete (and complex) nonlinear continuum model in SI sec. D. An inhomogeneous Fredholm equation is found for the purely geometric network (where the number of long-range network nodes $N = 0$ or equivalently $w_{ij} = 0$), with

$$\varphi(\mathbf{x}, \mathbf{x}') = \frac{u^2(\mathbf{x}, \mathbf{x}')}{U(\mathbf{x}')} + \frac{1}{U(\mathbf{x}')} \int d\mathbf{z} u(\mathbf{x}', \mathbf{z}) \varphi(\mathbf{x}, \mathbf{z}) \left(\rho(\mathbf{z}) - \delta(\mathbf{x} - \mathbf{z}) \right), \quad (8)$$

where $\rho(\mathbf{z})$ is the local density of material nodes at \mathbf{z} and $U(\mathbf{x}) = \int d\mathbf{z} \rho(\mathbf{z}) u(\mathbf{x}, \mathbf{z})$ is the strength the material node at \mathbf{x} . The delta function removes the (undefined) factor of $\varphi(\mathbf{x}, \mathbf{x})$ (necessary because of the discrete requirement that $E_{ii} = 0$), and is equivalent to the removal of a self-energy. Note that if $u(\mathbf{x}, \mathbf{x}') \neq \text{constant}$, the linearized importance will yield a non-trivial result ($\varphi(\mathbf{x}, \mathbf{x}') \neq \text{constant}$). It is straightforward to show that the solution to the Eq. 8 is $\varphi(\mathbf{x}, \mathbf{x}') = \bar{\varphi}(\mathbf{x}, \mathbf{x}') - \bar{\varphi}(\mathbf{x}, \mathbf{x})$, where $U(\mathbf{x}') \bar{\varphi}(\mathbf{x}, \mathbf{x}') = u^2(\mathbf{x}, \mathbf{x}') + \int d\mathbf{z} \rho(\mathbf{z}) u(\mathbf{x}', \mathbf{z}) \bar{\varphi}(\mathbf{x}, \mathbf{z})$ is the importance without removal of the self-energy term. There are a wide range of methods to determine $\bar{\varphi}(\mathbf{x}, \mathbf{x}')$ analytically[52, 53], but determining an exact expression is a non-trivial task even for simple geometries (see SI Sec C).

If we combine short ranged geometrical interactions with long range, geometry-independent

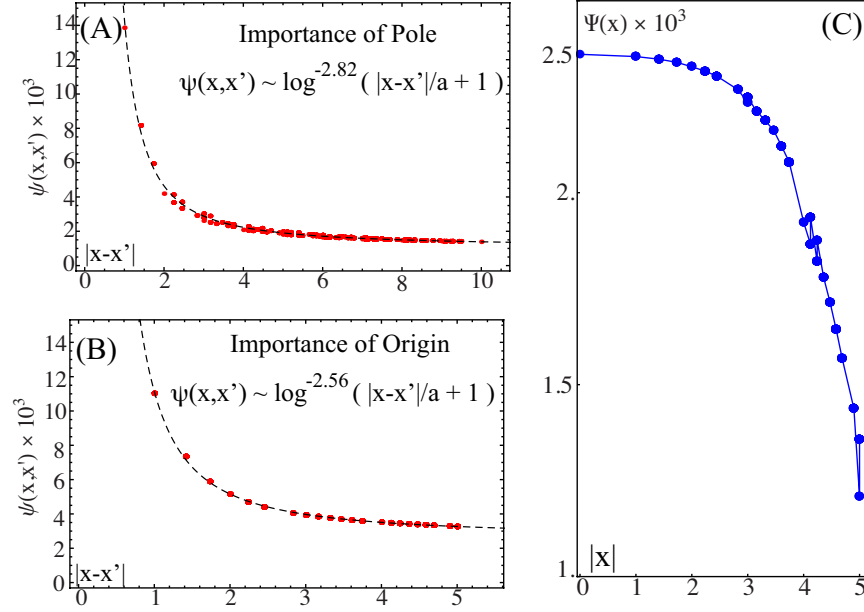


Figure 13: (A) shows the importance that a node at \mathbf{x} assigns to a node at one of the poles of the sphere $\mathbf{x} = (R, 0, 0)$ for all nodes in a sphere of radius $R = 5a$ with a weight $u(\mathbf{x}, \mathbf{x}') = e^{-|\mathbf{x}-\mathbf{x}'|/a}$ and no long-range connections. The importance assigned by a node at \mathbf{x}' towards the reference point at \mathbf{x} decreases with an empirically determined scaling which is not expected to be universal. (B) shows the importance of the origin ($\mathbf{x} = (0, 0, 0)$) as a function of the distance $|\mathbf{x}'|$ and decreases more slowly than the importance assigned to a node at the pole due to \mathbf{x} 's central location in the network. (C) shows the global centrality of each node $\Psi(\mathbf{x})$ as a function of its distance from the origin. Sharp variations for particular values of $\Psi(\mathbf{x})$ are due to nodes with a differing local connectivity but an identical distance from the origin (e.g. $|\mathbf{x}| = 5a$ for $\mathbf{x} = (0, 0, 5)$ and for $\mathbf{x} = (0, 4, 3)$). The overall decrease in $\Psi(\mathbf{x})$ as a function of $|\mathbf{x}|$ is readily apparent, with the origin being most important as expected.

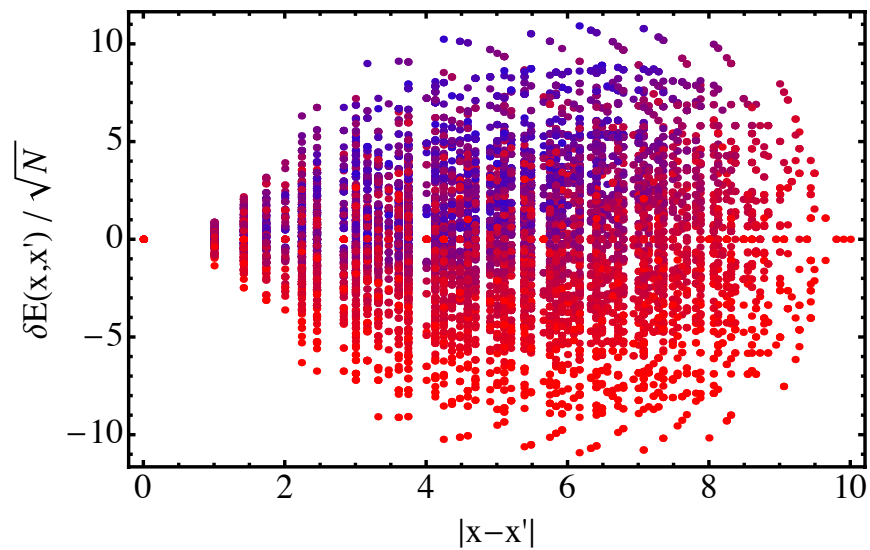


Figure 14: The asymmetry $\delta E(\mathbf{x}, \mathbf{x}') = E(\mathbf{x}, \mathbf{x}') - E(\mathbf{x}', \mathbf{x})$ for a sphere of radius $R = 5a$ (with no long-range interactions) as a function of the distance between the nodes in the absence of any long range connections. Deeper red points correspond to \mathbf{x}' near the origin and blue points correspond to \mathbf{x}' near the boundary of the sphere, so nodes on the periphery feel the central nodes are more important than vice versa. The relationship between δE_{ij} and $\delta \tau_{ij}$ in Fig. 2 of the main text suggests that random walks reach the center of the sphere more readily than a specified point on the surface.

edges, the expression for the GENs and the linearized importance φ becomes more complex. Eq. 8 is valid only in the absence of long range connections, but we show in SI Sec. B that the linearized can still be written in the simple form

$$\varphi(\mathbf{x}, \mathbf{x}') = \varphi_0(\mathbf{x}, \mathbf{x}') + \int_{\Omega} d\mathbf{z} K_1(\mathbf{x}, \mathbf{x}', \mathbf{z}) \varphi(\mathbf{x}, \mathbf{z}), \quad (9)$$

with the complicated functional forms of $\varphi_0(\mathbf{x}, \mathbf{x}')$ the kernel $K_1(\mathbf{x}, \mathbf{x}', \mathbf{z})$ explicitly given in SI Sec B. While Eq. 9 is linear (and therefore analytically tractable), all methods for determining $\varphi(\mathbf{x}, \mathbf{x}')$ will require integration of functions involving $\frac{1}{U(\mathbf{z})}$, which we show in SI Sec. C is difficult even in simple domains and for simple interactions $u(\mathbf{x}, \mathbf{x}')$. Eq. 8 and 9 can still be solved numerically over finite domains using well known methods[52, 53] in cases where exact results can not be obtained. The global importance, $\Phi(\mathbf{x})$, is straightforwardly computed from Eq. 9 using $\Phi(\mathbf{x}) = \int d\mathbf{z} \rho(\mathbf{z}) \varphi(\mathbf{x}, \mathbf{z})$ (assuming $u(\mathbf{x}, \mathbf{x}') > 0$ for all $|\mathbf{x} - \mathbf{x}'| < \infty$).

While Eq. 9, describing the continuum linearized importance, is analytically approachable, the full nonlinear theory for the continuum system is entirely intractable. Because the undesirable aspects of the linearized importance shown in Fig. 8 are expected to persist even in the continuum limit, we expect the nonlinear theory to be required to yield meaningful estimates of personalized closeness or global centrality. To understand the effect of the material interactions, we determine the GENs for a sphere of radius R formed from a lattice of equally spaced material nodes with an inter-node spacing a (i.e. with a constant node density ρ). It is straightforward to simply solve for the GENs in Eq. 1 of the main text numerically for any geometry, which equivalent to solving the integral equation via quadrature[52]. For $R = 5a$, the $N + N^* = 515$ nodes in the network are connected with a strength $u(\mathbf{x}, \mathbf{x}') = e^{-\lambda|\mathbf{x} - \mathbf{x}'|/a}$ (with $\lambda = 1$ chosen here). For a purely material sphere (where the interaction strength is determined entirely by $u(\mathbf{x}, \mathbf{x}')$ and the number of network nodes is $N = 0$), the closeness between two points is determined primarily by their relative distance, as shown in Fig. 13(A-B). While the expected decrease in the importance of \mathbf{x} from the perspective of another node at \mathbf{x}' for increasing $|\mathbf{x} - \mathbf{x}'|$ is observed, the qualitative behavior of the importance depends strongly on the locations of both points, with the empirically derived scaling of $\psi(\mathbf{x}, \mathbf{x}') \sim \log^{\beta(\mathbf{x})}(|\mathbf{x} - \mathbf{x}'|/a + 1) + \text{const}$. We also find that nodes towards the center of the sphere have a higher global importance $\Phi(\mathbf{x})$ than those towards the boundary (Fig. 13(C)), due to the greater number of paths towards the center than along the surface. We likewise expect the asymmetry in the closeness, $\delta E(\mathbf{x}, \mathbf{x}') = E(\mathbf{x}, \mathbf{x}') - E(\mathbf{x}', \mathbf{x})$, is skewed such that more external nodes feel closer to internal nodes than vice versa, due to the greater overall importance of the central nodes. This is confirmed in Fig. 14, where $\delta E(\mathbf{x}, \mathbf{x}') < 0$ is far more likely to be seen at the boundary of the sphere, regardless of the distance between the nodes. Coupled with Fig. 2 of the main text we expect that $\delta\tau(\mathbf{x}, \mathbf{x}') \sim -\delta E(\mathbf{x}, \mathbf{x}') \times \sqrt{\alpha N}$, for some topology-dependent α , will accurately predict the asymmetry in the mean first passage time of a random walk

between nodes in the sphere.

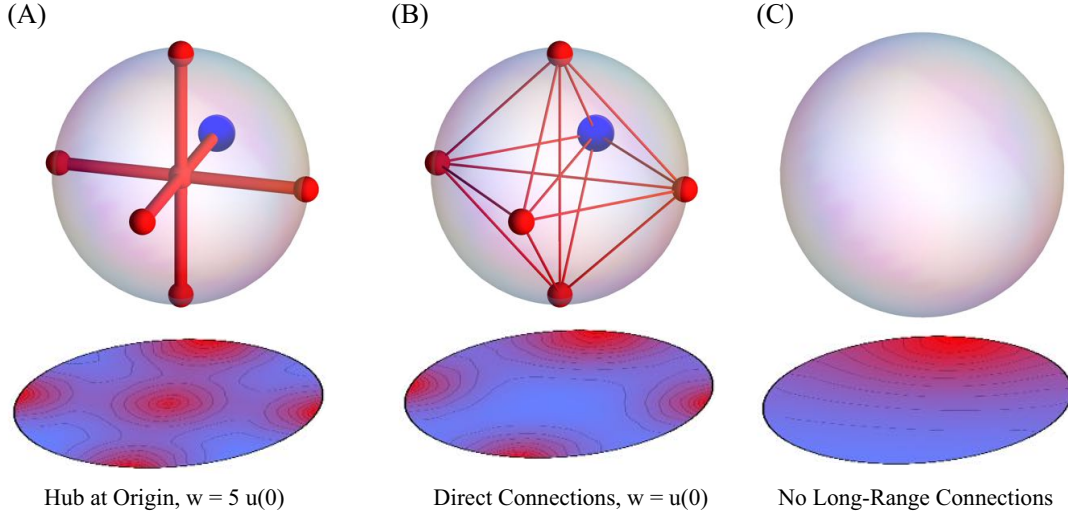


Figure 15: The importance from a node located in the $x - y$ plane felt towards the node at $\mathbf{x} = (R, 0, 0)$. A schematic of the network topology is shown on top, and the importance $\psi(\mathbf{x}, \mathbf{x}')$ for all points in the $x - y$ plane are shown below, with red corresponding to a higher importance. (A) shows the effect of long-range connections between the poles and the center of the sphere (with each long-range edge having weight $w_{ij} = 5$) for all edges relative to $\mathbf{x} = (R, 0, 0)$ (marked with the blue sphere). While each pole feels somewhat closer to nearby nodes than it does to distant nodes, it is still clear that the poles are strongly connected. (B) shows a network with the same strength of all nodes but the origin and a differing topology of direct connections between poles with weight $w_{ij} = 1$. Surprisingly, the importance assigned between poles is lower in (B) than (A) (i.e. stronger indirect connections produce a greater pole-to-pole importance than weaker direct connections), and the range of importance due to the distance-dependence connections is increased. (C) shows the material-only case with $N = 0$ (shown in Fig. 13 (A) as well).

The influence of long-range network topology greatly increases the complexity of the problem and obscures our understanding of the closeness between nodes in the sphere. To study the influence of the topology of the long range connections, we consider two very simple networks embedded in our sphere: long range connections between the poles of the sphere passing through a central hub, and direct connections between the poles (pictured in Fig. 15). In Fig. 15(A), we show the network nodes at the poles are indirectly linked to each other through strong connections to the center of the sphere of strength $w_{ij} = 5$ (so $W = \frac{1}{2} \sum_{ij} w_{ij} = 30$). Also shown is a cross-section of the sphere, showing the closeness between points to one of the poles (marked with the blue dot in the sphere above). The

total strength of the distance-dependent interactions is $U(\mathbf{x}) = \sum_l u(\mathbf{x}, \mathbf{x}_l) \approx 7.6$ at the poles, so the contribution of the material nodes is comparable to that of the long-range connections. Unsurprisingly, the center of the sphere feels each of the poles are important (the central red region in the cross-section in Fig. 15(A)), as do each of the other network-node poles. We compare the central-hub topology in Fig. 15(A) to that of direct connections between poles, depicted in Fig. 15(B). To keep the strength of the poles constant, the long range connections all have weight $w_{ij} = 1$ (but $W = 15$). Despite the constant total strength of the poles, the increased number of edges per pole decreases the importance each pole assigns to the other, reflected in the fact that the pole-to-pole importance, $\psi(\mathbf{x}, \mathbf{x}')$, in (B) is reduced by about 5% relative to (A). Indirect connections between the poles can therefore lead to a greater importance between them than direct connections due to the effect of short-ranged interactions with material nodes. The decay length of the importance assigned by material nodes to the network nodes is also increased for the direct connections in (B) relative to the range in (A) (not shown), suggesting that the reduced importance between network nodes is due to an increased importance of the material nodes. In situations where finite resources are shared (e.g. a random walk or diffusion of information), direct connections to a central hub will more easily allow the resource to be targeted to specific sites in comparison to multiple direct connections.

8 Conclusions

In this paper, we have shown the utility of the Generalized Erdős Numbers in a variety of contexts on both a local and a global level. The asymmetry of the GENs is highly correlated to the asymmetry in the MFPT, another asymmetric quantity of use in network science. The ability of the GENs to generate a global measure of closeness between nodes in a network has been shown in the context of defining both a global and personalized measure of the importance or centrality. Our global Erdős centrality is consistent with other well known centrality measures with the added benefit of a greater heterogeneity for nodes of equal degree in many cases (see Fig. ErdosCentrality.fig). The personalized importance ψ_{ij} from which the Erdős centrality is derived is shown to be useful in other contexts, with a specific example studying the influence specific locations of a sphere can have on other locations. Having demonstrated the usefulness of the GENs in many contexts here, future work will apply these methods to concrete problems involving network growth or dynamics on networks in more concrete cases of physical or biological importance.

Acknowledgements: We thank O. Peleg and G. Strang for useful comments on the manuscript.
grant info

References

- [1] G. Morrison and L. Mahadevan. Asymmetric network connectivity using weighted harmonic averages. *Europhys. Lett.*, 93:40002, 2011.
- [2] D. S. Bassett, E. T. Owens, K. E. Daniels, and M. A. Porter. Influence of network topology on sound propagation in granular materials. *Phys. Rev. E*, 86:041306, 2012.
- [3] M. Barthelemy. Spatial networks. *Phys. Rep.*, 499:1, 2011.
- [4] J. Bacompte, P. Jordano, and J. M. Olesen. Asymmetric coevolutionary networks facilitate biodiversity maintenance. *Science*, 312:431, 2006.
- [5] M. E. J. Newman. Spread of epidemic disease on networks. *Phys. Rev. E*, 66:016128, 2002.
- [6] D. Balcan, V. Colizza, B. Goncalves, H. Hu, J. J. Ramasco, and A. Vespignani. Multiscale mobility networks and the spatial spreading of infectious diseases. *Proc. Natl. Acad. Sci.*, 106:21484, 2009.
- [7] M. Keeling. The implications of network structure for epidemic dynamics. *Theor. Pop. Biol.*, 67:1, 2005.
- [8] A-L. Barabási and R. Albert. Emergence of scaling in random networks. *Science*, 286:509, 1999.
- [9] A. Barrat, M. Barthelemy, R. Pastor-Satorras, and A. Vespignani. The architecture of complex weighted networks. *Proc. Natl. Acad. Sci.*, 101:3747, 2004.
- [10] M. E. J. Newman. *Networks: An Introduction*. Oxford University Press, 2010.
- [11] A. Liancichinetti, S. Fortunato, and F. Radicchi. Benchmark graphs for testing community detection algorithms. *Phys. Rev. E*, 78:46110, 2008.
- [12] G. Morrison and L. Mahadevan. Discovering communities through friendship. *PLoS One*, 7:e38704, 2012.
- [13] M. Girvan and MEJ Newman. Community structure in social and biological networks. *Proc. Natl. Acad. Sci.*, 99:7821, 2002.
- [14] J. D. Noh and H. Rieger. Random walks on complex networks. *Phys. Rev. Lett.*, 92:118701, 2004.
- [15] D. J. Klein and M. Randic. Resistance distance. *J. Math. Chem.*, 12:81, 1993.
- [16] M. E. J. Newman. A measure of betweenness centrality based on random walks. *Social Networks*, 27:39, 2005.

- [17] R. Pastor-Satorras and A. Vespignani. Epidemic spreading in scale-free networks. *Phys. Rev. Lett.*, 86:3200, 2001.
- [18] F. Ball and P. Neal. Network epidemic models with two levels of mixing. *Math. Biosci.*, 212, 2008.
- [19] Massimo Franceschet. Pagerank: Standing on the shoulders of giants. *Communications of the ACM*, 54:92, 2011.
- [20] G. Ghoshal and A-L Barabási. Ranking stability and super-stable nodes in complex networks. *Nature Comm.*, 2:1, 2011.
- [21] N. Blumm et. al. Dynamics of ranking processes in complex systems. *Phys. Rev. Lett.*, 109:128701, 2012.
- [22] M. E. J. Newman. Scientific collaboration networks. ii. shortest paths, weighted networks, and centrality. *Phys. Rev. E.*, 64:016132, 2001.
- [23] T. Zhou, J. Ren, M. Medo, and Y. Zhang. Bipartite network projection and personal recommendation. *Phys. Rev. E*, 76:046115, 2007.
- [24] Y. Wang. *Topology Control for Wireless Sensor Networks*. Springer, 2008.
- [25] N. Masuda, H. Miwa, and N. Konno. Geographical threshold graphs with small-world and scale-free properties. *Phys. Rev. E*, 71:036108, 2005.
- [26] M. Kurant and P. Thiran. Layered complex networks. *Phys. Rev. Lett.*, 96:138701, 2006.
- [27] A. De Montis, M. Barthelemy, A. Chessa, and A. Vespignani. The structure of interurban traffic: a weighted network analysis. *Envir. and Planning B*, 34:905, 2007.
- [28] S. Boccaletti, V. Latora, Y. Moreno, M. Chavez, and DU Hwang. Complex networks: Structure and dynamics. *Phys. Rep.*, 424:175, 2006.
- [29] A. W. Rives and T. Galitski. Modular organization of cellular networks. *Proc. Natl. Acad. Sci.*, 100:1128, 2003.
- [30] R. DeCastro and J. Grossman. Famous trails to paul erdős. *Math. Intel.*, 21:51, 1999.
- [31] D. Babic, DJ Klein, I. Lukovits, S. Nikolic, and N. Trrinajstic. Resistance distance matrix: A computational algorithm and its application. *Int. J. Qantum Chem.*, 90:166, 2002.
- [32] E. A. Leicht, P. Holme, and MEJ Newman. Vertex similarity in networks. *Phys. Rev. E*, 73:026120, 2006.
- [33] E. Estrada and N. Hatano. Communicability in complex networks. *Phys. Rev. E*, 77:036111, 2008.

- [34] P. D. Hoff, A. E. Raftery, and M. S. Handcock. Latent space approaches to social network analysis. *J. Am. Stat. Assoc.*, 97:1090, 2002.
- [35] A. Fedorova, S. Blagodurov, and S. Zhuravlev. Managing contention for shared resources on multicore processors. *Comm. ACM*, 53:49, 2010.
- [36] P. Tetali. Random walks and the effective resistance of networks. *J. Theoretical Prob.*, 4:101, 1991.
- [37] R. M. Anderson and R. M. May. *Infectious Diseases of Humans*. Oxford University Press, 1992.
- [38] C. Lagorio et. al. Quarantine-generated phase transition in epidemic spreading. *Phys. Rev. E*, 83:026102, 2011.
- [39] H. W. Hethcote. The mathematics of infectious diseases. *SIAM REV.*, 42:599, 2000.
- [40] J. C. Miller. A note on a paper by Erik Volz: SIR dynamics in random networks. *J. Math. Biol.*, 62:349, 2011.
- [41] L. A. Meyers. Contact network epidemiology: Bond percolation applied to infectious disease prediction and control. *Bull. Amer. Math. Soc.*, 44:63, 2007.
- [42] R. Pastor-Satorras and A. Vespignani. Immunization of complex networks. *Phys. Rev. E*, 65:036104, 2002.
- [43] D. T. Gillespie. A general method for numerically simulating the stochastic time evolution of coupled chemical reactions. *J. Comp. Phys.*, 22:403, 1976.
- [44] M. Hein, J-Y Audibert, and U. von Luxburg. Graph laplacians and their convergence on random neighborhood graphs. *J. Mach. Learn. Res.*, 8:1325, 2007.
- [45] T. Opsahl, F. Agneessens, and J. Skvortez. Node centrality in weighted networks: Generalizing degree and shortest paths. *Social Networks*, 32:245, 2010.
- [46] S. P. Borgatti. Centrality and network flow. *Social Networks*, 27:55, 2005.
- [47] R. Fagin, R. Kumar, and D. Sivakumar. Comparing top k lists. *SIAM J. Disc. Math.*, 17:134, 2003.
- [48] T. H. Haveliwala. Topic-sensitive pagerank: a context-sensitive ranking algorithm for web search. *IEEE Trans. Knowl. Data Eng.*, 15:784, 2003.
- [49] N. Blumm et. al. Dynamics of ranking processes in complex systems. *Phys. Rev. Lett.*, 109:128701, 2012.
- [50] P. Pons and M. Latapy. Computing communities in large networks using random walks. *Comp. and Info. Sci. ISCIS*, 2005:284, 2005.

- [51] M. E. J. Newman and D. J. Watts. Mean-field solution of the small-world network model. *Phys. Rev. Lett.*, 84:3201, 2000.
- [52] A. D. Polyanin and A. V. Manzhirov. *Handbook of Integral Equations*. CRC Press, 1998.
- [53] F. G. Tricomi. *Integral Equations*. Interscience Publishers, 1957.

A Formulation of the linearized continuum model

In order to illustrate the emergence of both the node density and the delta function constraints in the continuum limit (Eq. 8 of the main text), it is useful to begin by calculating $U(\mathbf{x})$ in the case where network nodes are not included (or only between material nodes). We presume that there are a sufficiently large number of material nodes N^* so that we can replace the sums over all nodes (such as in Eq. 1 of the main text) with an integral. The interaction between material the nodes i and j is $u(\mathbf{x}_i, \mathbf{x}_j)$. For the discrete system, the total strength of the material node j is given by

$$U_j = \sum_{i \neq j} u(\mathbf{x}_i, \mathbf{x}_j) = \sum_i u(\mathbf{x}_i, \mathbf{x}_j) - u(\mathbf{x}_j, \mathbf{x}_j), \quad (10)$$

where the second term imposes the constraint of no self-interactions (no loops) in the network. This constraint can be avoided by allowing self-loops and will have no impact on the continuum model for the linearized importance, but such a change would have an minor impact on the non-linear theory. Because the GENs were originally developed for networks without self-interactions, we retain that assumption in this paper. If each material node were on a lattice with spacing a , the volume excluded by each node would be a^d . We could then write

$$U_i = \sum_j \frac{a^d u(\mathbf{x}_i, \mathbf{x}_j)}{a^d} - u(\mathbf{x}_i, \mathbf{x}_i) \rightarrow \int_{\Omega} d\mathbf{z} \rho u(\mathbf{z}, \mathbf{x}_i) \left[1 - \delta(\mathbf{x} - \mathbf{z}) \right] \equiv U(\mathbf{x}_j) - u(\mathbf{x}_j, \mathbf{x}_j) \quad (11)$$

in the limit of $N^* \rightarrow \infty$ and $a \rightarrow 0$, where $\rho = 1/a^d$ is the constant number density of the nodes at the location \mathbf{z} and with the constraint $\rho V = N^*$, where V is the volume of the domain and $U(\mathbf{x})$ is the total strength with the integral over all space (i.e. without neglecting the self-loop). We have kept the continuum version of $U(\mathbf{x})$ to be the integral over all space for convenience, but must remember to remove the contribution from the self loop as in Eq. 11. For material nodes that are not on a lattice, we can replace the effective volume a^d with the volume excluded by the node j $v(\mathbf{x}_i)$, and defining the variable density, $\rho(\mathbf{x}) = 1/v(\mathbf{x})$, produces the strength

$$U(\mathbf{x}) = \sum_j \frac{v(\mathbf{x}_j) u(\mathbf{x}_i, \mathbf{x}_j)}{v(\mathbf{x}_j)} \rightarrow \int d\mathbf{z} \rho(\mathbf{z}) u(\mathbf{z}, \mathbf{x}) \quad (12)$$

where again $U_j \rightarrow U(\mathbf{x}_j) - u(\mathbf{x}_j, \mathbf{x}_j)$. A similar limit will be found for any other sum that must avoid self-loops in the continuum limit, such as in Eq. 1 of the main text.

B Appendix: Including long-range connections in the continuum model

Labeling $\varphi^{(n \rightarrow m)}(\mathbf{x}, \mathbf{y})$ as the importance a network node at \mathbf{y} (connected to a long range edge) assigns to a material node at \mathbf{x} (with no long range connection), the linearized importance between material nodes reduces to the coupled equations

$$\begin{aligned}
 U(\mathbf{x}')\varphi(\mathbf{x}, \mathbf{x}') &= u^2(\mathbf{x}, \mathbf{x}') + \int d\mathbf{z} u(\mathbf{x}', \mathbf{z})\varphi(\mathbf{x}, \mathbf{z}) \left[\rho(\mathbf{z}) - \delta(\mathbf{x} - \mathbf{z}) \right] \\
 &\quad + \sum_k u(\mathbf{x}', \mathbf{y}_k) \left(\varphi^{(n \rightarrow m)}(\mathbf{x}, \mathbf{y}_k) - \varphi(\mathbf{x}, \mathbf{y}_k) \right)
 \end{aligned} \tag{13}$$

and

$$\begin{aligned}
 \left(U(\mathbf{y}_l) + W_k \right) \varphi^{(n \rightarrow m)}(\mathbf{x}, \mathbf{y}_k) &= u^2(\mathbf{x}, \mathbf{y}_k) + \int d\mathbf{z} u(\mathbf{y}_k, \mathbf{z})\varphi(\mathbf{x}, \mathbf{z}) \left[\rho(\mathbf{z}) - \delta(\mathbf{x} - \mathbf{z}) \right] \\
 &\quad + \sum_l [w_{kl} + u(\mathbf{y}_k, \mathbf{y}_l)] \varphi^{(n \rightarrow m)}(\mathbf{x}, \mathbf{y}_l) - u(\mathbf{y}_k, \mathbf{y}_l) \varphi(\mathbf{x}, \mathbf{y}_l),
 \end{aligned} \tag{14}$$

where Eq. 13 describes the importance between material nodes, and Eq. 15 describes the importance from the N network nodes to the material nodes. Eq. 13 is identical to Eq. 8 of the main text with a set of additional δ -function constraints at the location of the network nodes (where long-range connections contribute). Note that if $w_{ij} = 0$, the substitution of $\varphi^{(n \rightarrow m)}(\mathbf{x}, \mathbf{y}_l) = \varphi(\mathbf{x}, \mathbf{y}_l)$ into Eq. 15 recovers Eq. 13, which in turn reduces to Eq. 8 of the main text as expected. The discrete nature of the network nodes allows a direct solution in terms of $\varphi(\mathbf{x}, \mathbf{x}')$, and it is convenient to define $\mathbf{M} = (\mathbf{L}_u + \mathbf{L}_0)^{-1}$, with $(\mathbf{L}_0)_{ij} = W_i \delta_{ij} - w_{ij}$ the un-normalized graph Laplacian of the network nodes (in the absence of any material nodes) and $(\mathbf{L}_d)_{ij} = U(\mathbf{y}_i) \delta_{ij} - u(\mathbf{y}_i, \mathbf{y}_j)$ incorporates the effects of the material nodes on the strength of the network nodes. It is important to note that $\mathbf{L}_d + \mathbf{L}_0$ is *not* the graph Laplacian of the entire network, and can be inverted so long as $u(\mathbf{z}, \mathbf{z}') > 0$. It is straightforward to derive the solution for $\varphi^{(n \rightarrow m)}$ in terms of φ , with

$$\begin{aligned}
 [(\mathbf{L}_u + \mathbf{L}_0)\varphi^{(n \rightarrow m)}(\mathbf{x})]_k &= u^2(\mathbf{x}, \mathbf{y}_l) + \int d\mathbf{z} \varphi(\mathbf{x}, \mathbf{z}) u(\mathbf{y}_l, \mathbf{z}) \left[\rho(\mathbf{z}) - \delta(\mathbf{x} - \mathbf{z}) \right] \\
 &\quad - \sum_m \varphi(\mathbf{x}, \mathbf{y}_m) u(\mathbf{y}_l, \mathbf{y}_m),
 \end{aligned} \tag{15}$$

where $[\varphi^{(n \rightarrow m)}(\mathbf{x})]_k = \varphi^{(n \rightarrow m)}(\mathbf{x}, \mathbf{x}_k)$. Eq. 15 can be solved directly, and we find and substitution into Eq. 13 yields

$$\begin{aligned}
U(\mathbf{x}')\varphi(\mathbf{x}, \mathbf{x}') &= u^2(\mathbf{x}, \mathbf{x}') + \sum_{kl} u(\mathbf{x}', \mathbf{y}_k)\mathbf{M}_{kl}u^2(\mathbf{x}, \mathbf{y}_l) - \sum_l u(\mathbf{x}', \mathbf{y}_l)\varphi(\mathbf{x}, \mathbf{y}_l) \\
&+ \int d\mathbf{z}\varphi(\mathbf{x}, \mathbf{z})[\rho(\mathbf{z}) - \delta(\mathbf{x} - \mathbf{z})] \left[u(\mathbf{x}', \mathbf{z}) + \sum_{kl} u(\mathbf{x}', \mathbf{y}_k)\mathbf{M}_{kl}u(\mathbf{y}_l, \mathbf{z}) \right] \\
&- \sum_{kl} u(\mathbf{x}', \mathbf{y}_k)\mathbf{M}_{kl} \sum_m \varphi(\mathbf{x}, \mathbf{y}_m)u(\mathbf{y}_l, \mathbf{y}_m)
\end{aligned} \tag{16}$$

where $\mathbf{M} = (\mathbf{L}_u + \mathbf{L}_0)^{-1}$. Eq. 16 is linear in $\varphi(\mathbf{x}, \mathbf{x}')$, meaning that it can be solved exactly knowing the locations of and interactions between network nodes. In particular, we can write

$$\varphi(\mathbf{x}, \mathbf{x}') = \varphi_0(\mathbf{x}, \mathbf{x}') + \int_{\Omega} d\mathbf{z}K_1(\mathbf{x}, \mathbf{x}', \mathbf{z})\varphi(\mathbf{x}, \mathbf{z}) \tag{17}$$

with

$$\varphi_0(\mathbf{x}, \mathbf{x}') = \frac{u^2(\mathbf{x}, \mathbf{x}')}{U(\mathbf{x}')} + \frac{1}{U(\mathbf{x}')} \sum_{kl} u(\mathbf{x}', \mathbf{y}_k)\mathbf{M}_{kl}u^2(\mathbf{x}, \mathbf{y}_l) \tag{18}$$

is the direct connection term, renormalized to include the effects of the long-distance connection, and the kernel becomes

$$K_1(\mathbf{x}, \mathbf{x}', \mathbf{z}) = \left[1 - \frac{\delta(\mathbf{x} - \mathbf{z})}{\rho(\mathbf{x})} - \sum_m \frac{\delta(\mathbf{y}_m - \mathbf{z})}{\rho(\mathbf{y}_m)} \right] K_1^{(b)}(\mathbf{x}', \mathbf{z}) \tag{19}$$

with

$$K_1^{(b)}(\mathbf{x}', \mathbf{z}) = \frac{\rho(\mathbf{z})}{U(\mathbf{x}')} \left[u(\mathbf{x}', \mathbf{z}) + \sum_{kl} u(\mathbf{x}', \mathbf{y}_k)\mathbf{M}_{kl}u(\mathbf{y}_l, \mathbf{z}) \right] \tag{20}$$

is a ‘bare’ kernel (in the absence of any delta function constraints). It is worthwhile to note that the the dominant contribution to $\varphi(\mathbf{x}, \mathbf{x}')$ comes from $\varphi_0(\mathbf{x}, \mathbf{x}')$ in Eq. 17 (with higher order contributions coming from integrals weighted by u). If long-range interactions are expected to be perturbative and local interactions are very short-ranged, $\varphi_0(\mathbf{x}, \mathbf{x}')$ may be a sufficient approximation to avoid the complexity of an exact evaluation of $\varphi(\mathbf{x}, \mathbf{x}')$. However, the undesirable behavior of the linearized importance, pictured in Fig. 8 of the main text, suggests that this further-approximated value of $\varphi(\mathbf{x}, \mathbf{x}')$ should be used with caution.

The most straightforward approach to solving Eq. 9 of the main text is the development of a Neumann series, and it is simple to show that defining $\varphi(\mathbf{x}, \mathbf{x}') = \sum_{n=0}^{\infty} \varphi_n(\mathbf{x}, \mathbf{x}')$ with

$\varphi_n(\mathbf{x}, \mathbf{x}') = \int d\mathbf{z} K_n(\mathbf{x}, \mathbf{x}', \mathbf{z}) \varphi_0(\mathbf{x}, \mathbf{z})$ for $n \geq 1$ and $K_n(\mathbf{x}, \mathbf{x}', \mathbf{z}) = \int d\mathbf{z}' K_1(\mathbf{x}, \mathbf{x}', \mathbf{z}') K_{n-1}(\mathbf{x}, \mathbf{z}', \mathbf{z})$. The delta function constraints make this form of the kernel difficult to work with, but it is possible to reduce $K_n(\mathbf{x}, \mathbf{x}', \mathbf{z})$ in terms of the more easily computed $K_n^{(b)}(\mathbf{x}', \mathbf{z}) = \int d\mathbf{z} K_1^{(b)}(\mathbf{x}', \mathbf{z}') K_{n-1}^{(b)}(\mathbf{z}', \mathbf{z})$. It is not difficult to show that for a given location \mathbf{x} and known $\{\mathbf{y}_l\}$ that the full kernel has the form

$$K_n(\mathbf{x}, \mathbf{x}', \mathbf{z}) = \left(1 - \frac{\delta(\mathbf{z} - \mathbf{x})}{\rho(\mathbf{x})} - \sum_l \frac{\delta(\mathbf{z} - \mathbf{y}_l)}{\rho(\mathbf{y}_l)} \right) \left(K_n^{(b)}(\mathbf{x}', \mathbf{z}) \right. \quad (21)$$

$$\left. + \sum_{m=1}^{n-1} \left[b_m K_m^{(b)}(\mathbf{x}', \mathbf{z}) + c_m K_m^{(b)}(\mathbf{x}', \mathbf{z}) + \sum_l d_{ml} K_m^{(b)}(\mathbf{x}', \mathbf{y}_l) \right] \right),$$

with large and unwieldy recursive relations defining the coefficients b_m , c_m and d_{ml} , each of which depend on \mathbf{x} and $\{\mathbf{y}_l\}$. Importantly, though, the coefficients do not depend on \mathbf{z} , the variable of integration in the definition of $\varphi_n(\mathbf{x}, \mathbf{x}')$. It is therefore in principle possible to determine $\varphi(\mathbf{x}, \mathbf{x}')$ exactly with knowledge of $K_n^{(b)}(\mathbf{x}, \mathbf{x}')$, which can be determined using standard methods. However, as noted in SI Sec. C, integrals over $1/U(\mathbf{x}')$ are expected to be very unwieldy, and numerical work is likely required to determine the propagator.

C Determining $U(x)$ for a simple interaction

As long as our domain is finite, material nodes on the boundary of the domain will be less connected than material points in the interior. This will be reflected in a non-constant $U(\mathbf{x})$, which gives rise to a variable importance of the material nodes. Computing $U(\mathbf{x})$ is not necessarily trivial. If we suppose that $u(\mathbf{x}, \mathbf{x}') = u(|\mathbf{x} - \mathbf{x}'|)$ and that the domain is a sphere of radius R with a constant node density $\rho(\mathbf{x}) = V^{-1}$, we find

$$U(\mathbf{x}) = \frac{1}{V} \int_{\Omega} d^3 \mathbf{x}' u(|\mathbf{x} - \mathbf{x}'|) = \frac{1}{V} \int \frac{d^3 \mathbf{k}}{(2\pi)^3} e^{i\mathbf{k} \cdot \mathbf{x}} \hat{u}(|\mathbf{k}|) \int_{\Omega} d^3 \mathbf{y} e^{-i\mathbf{k} \cdot \mathbf{y}} \quad (22)$$

$$= \frac{1}{V} \int_0^{\infty} \frac{dk}{2\pi} \frac{4\hat{u}(k)}{k^2 x} \left(\sin(kR) - kR \cos(kR) \right) \sin(kx), \quad (23)$$

where $\hat{u}(\mathbf{k})$ is the Fourier Transform of $u(\mathbf{x})$. As a specific example, if $u(|\mathbf{x} - \mathbf{y}|) = e^{-l|\mathbf{x} - \mathbf{y}|}$, with $\hat{u}(k) = 8\pi l / (k^2 + l^2)^2$ in three dimensions, we find

$$U(x) = \frac{6}{R^3 l^3} + \frac{3e^{-lR}}{R^3 l^4 x} \left(\cosh(lx)[lx + l^2 x R] - \sinh(lr)[l^2 R^2 - 3lR - 3] \right), \quad (24)$$

yielding an expression for the total strength at the material node x expressed in terms of elementary functions only. Other forms of the interaction strength (for example, a gaussian) tend to produce more complex expressions, and we choose to use only an exponential decay

for the interaction strength between material nodes. Regardless of the (relative) simplicity of Eq. 24, terms involving integrals of $1/U(\mathbf{x}')$ will be difficult to evaluate exactly (as the reader may readily verify), meaning that the Neumann series[52, 53] described in appendix B is not easily computed analytically.

D The Nonlinear continuum model

The development of the continuum model for the nonlinear GENs follows the construction in Appendix A in a straightforward fashion. The discrete case of the GENs constrain $E_{ii} = 0$, but in the continuum limit we will have a self-energy contribution (similar to that of the self-interaction due to $u(\mathbf{x}_i, \mathbf{x}_i) \neq 0$ above). The δ -function constraints for the nonlinear GENs is most easily developed by writing

$$\frac{U_j}{E(\mathbf{x}_i, \mathbf{x}_j)} = u^2(\mathbf{x}_i, \mathbf{x}_j) + \sum_l \frac{u^2(\mathbf{x}_j, \mathbf{x}_l)}{E(\mathbf{x}_i, \mathbf{x}_l)u(\mathbf{x}_j, \mathbf{x}_l) + 1} - \frac{u^2(\mathbf{x}_i, \mathbf{x}_j)}{E(\mathbf{x}_i, \mathbf{x}_i)u(\mathbf{x}_i, \mathbf{x}_j) + 1} - \frac{u^2(\mathbf{x}_j, \mathbf{x}_j)}{E(\mathbf{x}_i, \mathbf{x}_j)u(\mathbf{x}_j, \mathbf{x}_j) + 1}, \quad (25)$$

where the first term on the right hand side correctly handles the direct connection between nodes i and j ; the second term is the sum over all nodes (including i and j); the third term removes the contribution due to a non-zero $E(\mathbf{x}_i, \mathbf{x}_i)$; and the fourth term removes the contribution due to the direct connection between the nodes at \mathbf{x}_i and xv_j . It is not difficult to see that in the continuum limit, this expression becomes

$$\frac{U(\mathbf{x}') - u(\mathbf{x}, \mathbf{x}')}{E(\mathbf{x}, \mathbf{x}')} + \frac{u(\mathbf{x}', \mathbf{x}')}{E(\mathbf{x}, \mathbf{x}')u(\mathbf{x}, \mathbf{x}') + 1} = u^2(\mathbf{x}, \mathbf{x}') + \int d\mathbf{z} \rho(\mathbf{z}) \frac{u^2(\mathbf{x}', \mathbf{z})}{E(\mathbf{x}, \mathbf{z})u(\mathbf{x}', \mathbf{z}) + 1} - \frac{u^2(\mathbf{x}, \mathbf{x}')}{E(\mathbf{x}, \mathbf{x})u(\mathbf{x}, \mathbf{x}') + 1} \quad (26)$$

after recalling that $U_j = U(\mathbf{x}_j) - u(\mathbf{x}_j, \mathbf{x}_j)$. The second term on the left hand side of the equation is due to the removal of the self-loop. It is clear that Eq. 26 can not be expressed in terms of any standard integral equation, and the nonlinearity guarantees that analytic work is unlikely to be fruitful without further approximation. If we assume that $E(\mathbf{x}, \mathbf{x}') \gg u^{-1}(\mathbf{x}, \mathbf{x}')$ for all \mathbf{x} and \mathbf{x}' and define $\varphi(\mathbf{x}, \mathbf{x}') \sim E^{-1}(\mathbf{x}, \mathbf{x}')$, it is straightforward to see that

$$U(\mathbf{x}')\varphi(\mathbf{x}, \mathbf{x}') = u^2(\mathbf{x}, \mathbf{x}') + \int d\mathbf{z} \varphi(\mathbf{x}, \mathbf{z})u(\mathbf{x}', \mathbf{z}) \left(\rho(\mathbf{z}) - \delta(\mathbf{x} - \mathbf{z}) \right), \quad (27)$$

which is Eq. 8 of the main text. The fact that Eq. 8 of the main text is the well known inhomogeneous Fredholm Equation[52, 53] means that analytic work is possible using the

linearized importance. The linearized importance was discussed in more detail in Appendix B.

In the continuum model, the long range connections between nodes at the specified points $\{\mathbf{y}_l\}$ must be enforced by delta-function constraints (since the material nodes an infinitesimal distance away from \mathbf{y}_j will not have a long-range connection), much like the self-loops are removed. The closeness between material nodes, $E(\mathbf{x}_i, \mathbf{x}_j)$, is still expected to be a continuous function, but we must also account for the long-range connections between network nodes. Because Eq. 1 of the main text determines the closeness felt by a network node \mathbf{y}_j towards a material node \mathbf{x}_i in terms of the direct connection to other material nodes $\{\mathbf{y}_l\}$ (with strength w_{jl}) and the closeness felt by \mathbf{y}_l towards \mathbf{x}_i . When computing the closeness felt by any node towards a material node, we do *not* need to compute the closeness felt by any node towards a network node. It suffices to compute the material-material closeness $E(\mathbf{x}_i, \mathbf{x}_j)$ as well as the network-material closeness $E^{(n \rightarrow m)}(\mathbf{x}_i, \mathbf{y}_j)$. In the discrete case, this labeling is irrelevant, but it is essential in the continuous case. We then write

$$\begin{aligned} \frac{U(\mathbf{x}_j) - u(\mathbf{x}_j, \mathbf{x}_j)}{E(\mathbf{x}_i, \mathbf{x}_j)} &= u^2(\mathbf{x}_i, \mathbf{x}_j) + \sum_{\substack{l \neq i, j \\ \mathbf{x}_l \notin \{\mathbf{y}_l\}}} \frac{u^2(\mathbf{x}_j, \mathbf{x}_l)}{E(\mathbf{x}_i, \mathbf{x}_l)u(\mathbf{x}_j, \mathbf{x}_l) + 1} \\ &\quad + \sum_{\mathbf{y}_l} \frac{u^2(\mathbf{x}_j, \mathbf{y}_l)}{E^{(n \rightarrow m)}(\mathbf{x}_i, \mathbf{x}_l)u(\mathbf{x}_j, \mathbf{x}_l) + 1} \end{aligned} \quad (28)$$

and

$$\begin{aligned} \frac{U(\mathbf{y}_i) + W_j - u(\mathbf{y}_j, \mathbf{y}_j)}{E^{(n \rightarrow m)}(\mathbf{x}_i, \mathbf{y}_j)} &= u^2(\mathbf{x}_i, \mathbf{y}_j) + \sum_{\substack{l \neq i, j \\ \mathbf{x}_l \notin \{\mathbf{y}_l\}}} \frac{u^2(\mathbf{y}_j, \mathbf{x}_l)}{E(\mathbf{x}_i, \mathbf{x}_l)u(\mathbf{y}_j, \mathbf{x}_l) + 1} \\ &\quad + \sum_{\mathbf{y}_l \neq \mathbf{y}_j} \frac{u^2(\mathbf{y}_j, \mathbf{y}_l)}{E^{(n \rightarrow m)}(\mathbf{x}_i, \mathbf{y}_l)u(\mathbf{y}_j, \mathbf{y}_l) + 1} \end{aligned} \quad (29)$$

While we have clearly delineated the distinction between $E(\mathbf{x}, \mathbf{x}')$ and $E^{(n \rightarrow m)}(\mathbf{x}, \mathbf{y}')$, there is no approximation here and Eqs. 28 and 29 are identical to Eq. 1 of the main text. In the continuum limit, only the sum over the material nodes can be converted to an integral, as the number of network nodes N remain fixed. We can then write

$$\begin{aligned} \frac{U(\mathbf{x}') - u(\mathbf{x}', \mathbf{x}')}{E(\mathbf{x}, \mathbf{x}')} + \frac{u^2(\mathbf{x}', \mathbf{x}')}{E(\mathbf{x}, \mathbf{x}')u(\mathbf{x}', \mathbf{x}') + 1} &= u^2(\mathbf{x}, \mathbf{x}') + \sum_l \frac{u^2(\mathbf{x}', \mathbf{y}_l)}{E^{n \rightarrow m}(\mathbf{x}, \mathbf{y}_l)u(\mathbf{x}', \mathbf{y}_l) + 1} \quad (30) \\ &\quad + \int d\mathbf{z} \frac{u^2(\mathbf{x}', \mathbf{z})}{E(\mathbf{x}, \mathbf{z})u(\mathbf{x}', \mathbf{z}) + 1} \left(\rho(\mathbf{z}) - \delta(\mathbf{x} - \mathbf{z}) - \sum_l \delta(\mathbf{y}_l - \mathbf{z}) \right) \end{aligned}$$

and

$$\begin{aligned}
\frac{U(\mathbf{y}_j) - u(\mathbf{y}_j, \mathbf{y}_j) + W_j}{E^{(n \rightarrow m)}(\mathbf{x}, \mathbf{y}_j)} + \frac{u^2(\mathbf{y}_j, \mathbf{y}_j)}{E^{(n \rightarrow m)}(\mathbf{x}, \mathbf{y}_j)u(\mathbf{y}_j, \mathbf{y}_j) + 1} = u^2(\mathbf{x}, \mathbf{y}_j) \quad (31) \\
+ \sum_l \frac{[u(\mathbf{y}_j, \mathbf{y}_l) + w_{jl}]^2}{E^{n \rightarrow m}(\mathbf{x}, \mathbf{y}_l)[u(\mathbf{y}_j, \mathbf{y}_l) + w_{jl}] + 1} \\
+ \int d\mathbf{z} \frac{u^2(\mathbf{y}_j, \mathbf{z})}{E(\mathbf{x}, \mathbf{z})u(\mathbf{y}_j, \mathbf{z}) + 1} \left(\rho(\mathbf{z}) - \delta(\mathbf{x} - \mathbf{z}) - \sum_l \delta(\mathbf{y}_l - \mathbf{z}) \right)
\end{aligned}$$

The full nonlinear version of the continuum limit is again clearly not analytically approachable, but the linearized version can be reduced to a single Fredholm equation, as is discussed further in Appendix B. If desired, the closeness felt by a material node towards a network node can be similarly computed (after defining a material-to-network closeness $E^{(m \rightarrow n)}(\mathbf{y}_i, \mathbf{x})$), which we do not explicitly compute here.

E Convergence of the Neumann Series

It is important to note that if $u(\mathbf{x}, \mathbf{x}') = u(|\mathbf{x} - \mathbf{x}'|)$, which is a natural choice for the form of the location dependence of the short range interactions, it is unlikely that the Neumann series will converge in an unbounded domain. The convergence of the series $\sum_n \varphi_n^{(b)}(\mathbf{x}, \mathbf{x}')$ is dependent on the fact that the l_2 norm of the kernel is bounded over the domain. For constant ρ and in an unbounded domain $U(\mathbf{x}') = U$ is a constant for all \mathbf{x} (since every node is connected in the same way to the other nodes in the infinite space) and it is straightforward to show that the l_2 norm is

$$\int d\mathbf{x}' d\mathbf{z} \rho(\mathbf{x}') \rho(\mathbf{z}) \left[K_1^{(b)}(\mathbf{x}', \mathbf{z}) \right]^2 = \frac{\rho}{U^2} \int d\mathbf{z} d\mathbf{x}' u(|\mathbf{x}' - \mathbf{z}|) \rightarrow \infty \quad (32)$$

after a change of variables to $\mathbf{z}' = \mathbf{z} - \mathbf{x}'$ (valid due to the domain's infinite extent). For any infinite domain, then, $\varphi^{(b)}(\mathbf{x}, \mathbf{x}')$ is undefined. With the δ -function constraints included, there will not be a divergence in φ , but the sum $\varphi(\mathbf{x}, \mathbf{x}') = \sum_n \varphi_n(\mathbf{x}, \mathbf{x}')$ will be weakly converging (e.g. a divergent sequence of alternating sign), and while theoretically exact will likely not be useful if numerical work is required. These issues may all be avoided if the domain is finite with volume V , at the cost of a more complex form for $U(\mathbf{x})$ (described in SI Sec. C).