communications physics

ARTICLE

https://doi.org/10.1038/s42005-022-00949-5 OPEN

Linking the network centrality measures closeness and degree

Tim S. Evans [1]✉ & Bingsheng Chen [1]

Measuring the importance of nodes in a network with a centrality measure is an core task in any network application. There many measures available and it is speculated that many encode similar information. We give an explicit non-linear relationship between two of the most popular measures of node centrality: degree and closeness. Based on a shortest-path tree approximation, we give an analytic derivation that shows the inverse of closeness is linearly dependent on the logarithm of degree. We show that our hypothesis works well for a range of networks produced from stochastic network models and for networks derived from 130 real-world data sets. We connect our results with previous results for other network distance scales such as average distance. Our results imply that measuring closeness is broadly redundant unless our relationship is used to remove the dependence on degree from closeness. The success of our relationship suggests that most networks can be approximated by shortest-path spanning trees which are all statistically similar two or more steps away from their root nodes.

[1] Centre for Complexity Science, and Theoretical Physics Group, Imperial College London, London SW7 2AZ, UK. ✉email: T.Evans@imperial.ac.uk

Network science has proved to be an exceptionally useful tool, especially with the large-scale data sets now available in every discipline. Many systems are driven by pairwise interactions at the microscopic level and the network formalism is ideal as it represents these bilateral relationships as edges between nodes. The real power of network analysis is that it gives us tools to look at the emergent behaviour of such systems even when driven by interaction's on meso- and macroscopic scales[1].

A good example is the key task of finding the most important nodes in the system. In network science, this is done by using a centrality measure which assigns nodes a centrality value with larger values indicating greater importance. Centrality has been developed over the last 70 years[2–15] and is a core part of many introductory texts on Network Science[16–19]. The simplest centrality measure is a degree, the number of neighbours each node has, probably the first node property examined in any study. However, the degree is insensitive to the wider network structure which is the primary goal of network analysis.

In order to probe the broader network structure, many centrality measures are based on the distance between nodes defined as the length of the shortest path between nodes. For instance, our focus is on one of the oldest and most widely used centrality measures, closeness[2], the inverse of the average distance from a node to all others. Closeness is an important indicator in many different contexts. It has been used to investigate the role of different academic disciplines when academics choose their journals and bibliographies[12]. Closeness centrality measures the impact of an author on a field and their social capital[10]. When used to select potential leads in customer data, closeness led to a significant gain in the success rate[9]. In an air transport network, it has been shown that the closeness of a city is highly correlated with socio-economic indicators such as gross regional domestic product[13]. Closeness has been applied to biological networks[6] and closeness measures were able to identify more than 50% of the global regulators within the top 2% of the ranked genes[8]. Essential genes have been found to have higher closeness than nonessential genes in protein-interaction networks[7]. In a metabolic network, the closeness of nodes can identify the most important metabolites[20]. Centrality measures such as closeness are an important tool to analyse many types of data.

However, there are a vast number of centrality indices available, as visualised nicely by Schoch[21,22], suggesting that there is a lot of redundancy. Many different centrality measures encode similar information as seen in the strong correlations between centrality indices[19,23–33]. In particular, Pearson correlation coefficients are invariably used which are most sensitive to linear correlations between centrality measures. That is we measure $\rho_{cd} = \langle (c_u - \bar{c})(d_u - \bar{d})\rangle/(\sigma_c\sigma_d)$ where c_u and d_u are two centrality measures for node u, the average is over all nodes with the means $(\bar{c}, \bar{d})$ and standard deviations (σ_c, σ_d) found using the same ensemble average. There seems to be no clear consensus from these studies other than there are often strong relationships between centrality measures but these vary from network to network. For instance, in the introduction of Schoch et al.[30] it is suggested that "Reported results, however, are often inconsistent with regard to the similarity of centrality indices". Of particular interest here is the conjecture made in the Introduction section of Valente et al.[27] where it is stated that "We expect that measures of degree and closeness centrality will be more highly correlated with each other than with other measures because they are both based on direct ties." Later, in the discussion of results by Valente et al.[27], the authors conclude that "The amount of correlation between degree, betweenness, closeness, and eigenvector indicates that these measures are distinct, yet conceptually related" and the closeness-degree pair is only the third most correlated pair of centrality indices in their study.

In this paper, we will focus on the relationship between closeness centrality and degree. Our result is that closeness centrality and degree have a non-linear relationship, namely that the inverse of closeness ('farness'[34], a normalised version of 'status'[5,6,35]) is linearly dependent on the logarithm of the degree. This explains why linear correlation measures often link degree and closeness centrality but at the same time, no general pattern has been seen before. Equally our results suggest that studies based on linear correlation measures may well miss important other features in the landscape of centrality measures.

This non-linear relationship means that measuring closeness gives the same broad information as closeness making the computationally expensive calculation of the latter redundant to a first approximation. To extract useful information from closeness the dependence on degree must first be removed and only then will closeness reveal important and distinct information on individual nodes.

Our starting point is that the shortest paths from any one node to all other nodes can be arranged as a spanning tree. We then conjecture that the branches of this tree are statistically similar, implying that the closeness of a node can only depend on the number of such branches, i.e., the degree of the node. If we also assume that the number of nodes in each branch grows exponentially, we find that the inverse of closeness is linearly dependent on the logarithm of degree centrality. The success of this non-linear relationship in a wide range of real-world networks suggests that both these conjectures are broadly true, an insight that can be exploited in a much wide range of network studies.

Results and discussion

Theory

General definitions. For simplicity, we will assume throughout this paper that we are analysing a simple graph $\mathcal{G}$ with just one component. We will denote the degree of each node v as k_v. A path in a network of length ℓ is a sequence of $(\ell+1)$ distinct nodes such that each consecutive pair of nodes in the path is connected by an edge. We will define the distance between two nodes u and v in a network to be the length of the shortest path between two nodes, denoted here as d_{uv}.

The closeness c_v[2,16–19] of a vertex v is then defined to be the inverse of the average distance from v to every other vertex in the graph, so

$$\frac{1}{c_v} = \frac{1}{(N-1)}\sum_{u\in\mathcal{V}\setminus v} d_{uv} \tag{1}$$

where $\mathcal{V}$ is the set of nodes and $N = |\mathcal{V}|$ is the number of nodes. Clearly the closer a vertex v is to other vertices in the network, the larger the closeness. Thus closeness mimics the properties of points in a geometric shape where those points closest to the geometric centre will have highest closeness.

Trees[16–19] are connected networks with no loops so the number of edges is always one less than the number of nodes. Here we use a spanning tree[36] which is a connected subgraph of the original graph $\mathcal{G}$ containing all the original vertices $\mathcal{V}$ but a subset of $(N-1)$ edges that are just sufficient to keep every node connected to all others. We are also going to work with rooted trees $\mathcal{T}(r)$ in which we have singled out one special node, the root r of the tree.

Estimate of closeness. We start from the idea that some of the statistical properties of real-world networks may be captured by spanning trees[36]. Here we are interested in closeness which uses the lengths of shortest paths between nodes so the most useful trees for this work are the shortest-path trees, $\mathcal{T}(r)$, that contains one shortest path from a root node r to each remaining node in

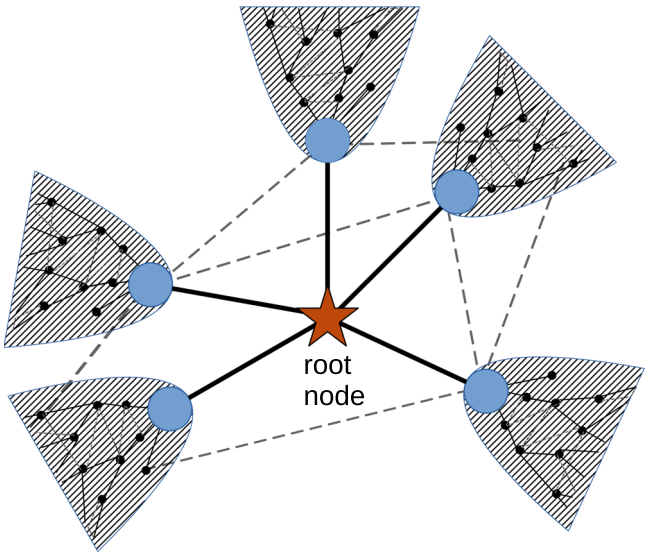


Fig. 1 An illustration of the shortest-path tree approximation. Every node r, here the red star, is considered to be the root node of a shortest-path tree $\mathcal{T}(r)$. This has k_r nearest neighbours, as indicated by the five solid black lines. Each of the neighbouring nodes, here the blue circles, is treated as the root of a branch of the shortest-path tree. These branches are treated as being statistically identical with a branching number $(1+\bar{z})$ as indicated here through the use of the same shaded shape rooted on each neighbouring node. The grey dashed lines represent some of the many edges in the graph $\mathcal{G}$ that are not included in the shortest-path tree.

the network. As our networks are unweighted, the shortest-path trees always exist and are easily defined as part of a breadth-first search algorithm, see the Supplementary Note 2 on the Shortest-Path Tree Algorithm for more details. Every node can act as a root node so there is at least one shortest-path tree, $\mathcal{T}(r)$, for every node r. These trees are not unique as there can be many shortest paths between a pair of nodes.

Our picture for these shortest-path trees is shown in Fig. 1. We start with the observation that close to the root node the structure of these shortest-path trees will vary and in particular, the number of nearest neighbours k_r of the root vertex r will vary. However, as we move further away from the root node, the number of nodes $n_r(\ell)$ at some distance ℓ from root node r grows exponentially with each step in most networks. This is the origin of the small-world effect seen in many networks, where the distance between nodes is typically much smaller than is found in similar size networks that are constrained by Euclidean geometry, such as a regular square grid of streets or a random geometric graph. Regardless of the local context of a root node, the shortest-path trees quickly access a similar set of nodes in the main bulk of the network provided there is no large-scale inhomogeneity in the network. Thus we conjecture that the structure and statistical properties of these trees away from the root node are similar for all possible root nodes. The contribution to the closeness of each node in the bulk is bigger as they are further from the root and more numerous. So we expect that the largest contributions to closeness always come from the same bulk regions where we can expect statistical similarity.

The most important difference when comparing different root nodes, therefore, comes from the local structure of each root node. In particular, the initial value for the exponential growth in the number of nodes at a distance ℓ from the root will have a large effect. This will depend on the local structure and the simplest effect comes from the number of immediate neighbours the root node has, i.e. the degree of the root node k_r. The simplest approximation for the growth of these shortest-path trees is therefore $n_r(\ell) = k_r\bar{z}^{\ell}$, where $\bar{z}$ is some measure of the rate of exponential growth of the shortest-path tree. Note that our assumption of statistical similarity suggests that the branching factor of these trees is, on average, the same so we use a single parameter $\bar{z}$ to represent the exponential growth from any root node r.

Our crude approximation is clearly going wrong when we look at large distances from the root as eventually any real network will run out of vertices so $n(\ell) = 0$ for large ℓ. One can model the end of the exponential growth in different ways but we will use the simplest. Namely, we will define a sharp upper cutoff L_r and assume that $n_r(\ell) = 0$ for $\ell > L_r$ which gives

$$N = 1 + \sum_{\ell=1}^{L_r} k_r\,\bar{z}^{\ell-1} = 1 + k_r\frac{(\bar{z}^{L_r}-1)}{(\bar{z}-1)}\,. \tag{2}$$

We can invert this expression to express this cutoff L_r in terms of parameters N and k_r, giving us $L_r = L(N, k_r)$ where for large N

$$L(N,k) \approx \frac{\ln\big(N(\bar{z}-1)/k\big)}{\ln(\bar{z})}\,. \tag{3}$$

Individual distances are integers but it is clear from the form in Eq. (3) that we need L to be a real number; in some sense L is an average over the actual distances from the root to the leaves (nodes with degree one) of the tree. This also tells us that the distance scale associated with a node depends on the logarithm of that node's degree. Since the inverse closeness is a sum of distances, this distance scale L controls the result and that explains why $\ln(k)$ appears in our expression for inverse closeness. As an aside, the $\ln(N)$ dependence of this distance scale L in Eq. (3) reflects the small-world effect seen in most networks.

We now have that $n_r(\ell) = k_r\bar{z}^{\ell-1}$ for $\ell < L(N, k_r)$ and zero for larger ℓ which depends on local parameters k_r, the degree of each root node, and two global parameters, the total number of nodes N and some measure of the growth rate of the shortest-path trees $\bar{z}$. We can now rewrite the closeness c_r Eq. (1) of a vertex r in terms of $n_r(\ell)$ as $1/c_r = (N-1)^{-1}\sum_{\ell=1}^{L_r}\ell n_\ell$ to find that

$$\frac{1}{c_r} = \frac{1}{(N-1)}\sum_{\ell=1}^{L_r}\ell k_r\bar{z}^{\ell-1} = \frac{k}{(N-1)}\left(\frac{(L_r+1)\bar{z}^{L_r}}{\bar{z}-1} - \frac{(\bar{z}^{L_r+1}-1)}{(\bar{z}-1)^2}\right) \tag{4}$$

By using Eqs. (2) and (3) we can eliminate L_r to find that

$$\frac{1}{c_r} = -\frac{1}{\ln(\bar{z})}\ln(k_r) + \beta\,. \tag{5}$$

Our calculation shows the parameter β is also independent of the root vertex r chosen but it is a function of the global network parameters N and $\bar{z}$ so that $\beta = \beta(\bar{z}, N)$ where

$$\beta(\bar{z},N) = \left(\frac{1}{(\bar{z}-1)} + \frac{\ln(\bar{z}-1)}{\ln(\bar{z})}\right) + \frac{1}{\ln(\bar{z})}\ln(N)\,. \tag{6}$$

Our prediction is that the inverse of closeness c_r of any node r should show a linear dependence on the logarithm of the degree k_r of that node with a slope that is the inverse of the log of the growth parameter $\bar{z}$.

In our analysis, we will not assume the parameter β is given by Eq. (6). By adding one additional fitted parameter we lose a little predictive power since many parameters are needed to characterise a network. This leaves us with a conjecture based on the number of nodes N and the degree of each node k_r which are usually known. Then in principle, we have two unknown global parameter values which we find from a linear fit to our data for c_r and k_r giving $\bar{z}^{(\mathrm{fit})}$ and $\beta^{(\mathrm{fit})}$.

Table 1 Results for artificial networks.

Network type	N	$1/\ln(\bar{z}^{(\mathrm{fit})})$	$\beta^{(\mathrm{fit})}$	$\bar{z}^{(\mathrm{fit})}$	$\beta(\bar{z}^{(\mathrm{fit})}, N)$	$\rho(c,k)$	χ_r^2
ER	1000	0.46 ± 0.01	4.29 ± 0.01	8.87 ± 0.20	4.23 ± 0.03	0.94	1.02
	2000	0.42 ± 0.01	4.52 ± 0.01	10.64 ± 0.18	4.28 ± 0.03	0.93	1.02
	4000	0.43 ± 0.01	4.82 ± 0.01	9.99 ± 0.12	4.67 ± 0.02	0.93	1.03
BA	1000	0.30 ± 0.01	3.59 ± 0.02	28.03 ± 3.11	3.09 ± 0.13	0.75	1.16
	2000	0.32 ± 0.01	3.86 ± 0.01	22.76 ± 2.22	3.46 ± 0.07	0.70	1.29
	4000	0.31 ± 0.01	4.03 ± 0.01	25.17 ± 2.61	3.60 ± 0.08	0.65	1.16
Config-BA	1000	0.35 ± 0.01	3.76 ± 0.02	17.41 ± 1.42	3.46 ± 0.14	0.75	1.19
	2000	0.36 ± 0.01	4.01 ± 0.02	16.08 ± 1.24	3.78 ± 0.15	0.70	1.28
	4000	0.35 ± 0.01	4.19 ± 0.01	17.41 ± 1.41	3.94 ± 0.08	0.66	1.19

Table of results for one example of a simple graph with average degree 10.0 produced using one of three artificial models with the same average degree $\langle k\rangle = 10.0$ but with a different number of nodes, N. Each 'ER' network is a standard Erdős-Réyni network, a 'BA' network is produced using pure preferential attachment in the Barabási-Albert model, and the 'Config-BA' network is a configuration model version of a Barabási-Albert model network. The results for $1/\ln(\bar{z}^{(\mathrm{fit})})$ and $\beta^{(\mathrm{fit})}$ come from linear fits of inverse closeness, $1/c_v$ to the logarithm of degree, $\ln(k_v)$ for each vertex v, i.e., $1/c_v = (\ln(\bar{z}))^{-1}\ln(k_v) + \beta$ (Eq. (5)). The value of β derived from $\bar{z}^{(\mathrm{fit})}$ and N using Eq. (6) is also shown for comparison. The fits are very good as indicated by the column for the reduced chi-square χ_r^2. The Pearson correlation measure between closeness and degree $\rho(c,k)$ is given for comparison with earlier work.

Numerical results for theoretical models. We looked at the relationship between closeness and degree using simple networks produced from three different theoretical models[17–19]: the Erdős-Réyni (ER) model[37] the Barabási-Albert model with pure preferential attachment[38] and the configuration model[39] network starting from a network generated with the same Barabási-Albert model. In the first and third model, the edges are completely randomised so there are no vertex-vertex correlations. The last two models both have fat-tailed degree distributions. Our networks built from artificial models were created using standard methods in the `networkx` package[40].

For a single network, we get several nodes with the same degree and we use this variation to find a mean and standard error in the mean shown. The fit is done using Eq. (5) with two free parameters $\bar{z}^{(\mathrm{fit})}$ and $\beta^{(\mathrm{fit})}$ and the goodness of fit measures in Table 1 shows this is a good fit, confirmed visually by the plots in Fig. 2a. Note that a good linear fit of status $((N-1)/c)$ to the logarithm of degree for one example of each of the Erdős-Réyni and Barabási-Albert models was found in Fig. 1 of Wuchty & Stadler[6].

Roughly speaking we find that mean inverse closeness values for any one degree are typically within 2% of the prediction made from the best fit as we can see in "Results" are shown in Fig. 2b. The small deviations seen in Fig. 2c, especially for the Config-BA model, are for higher degree values where the data are sparse and uncertainties are large. This means no firm conclusions can be drawn from Fig. 2c about the presence of higher-order corrections to our form Eq. (5).

We now turn to look at the actual values obtained from these fits of data on closeness and degree from the artificial networks to Eq. (5). As Table 1 shows there is a small amount of variation in value of $\bar{z}^{(\mathrm{fit})}$, the fit for the shortest-path tree growth factor, with the size of the network. What is of more interest are the differences in values between these three types of artificial networks. All these networks had an average degree of about 10.0 and an infinite tree with a constant degree 10 (a Bethe lattice) would have a growth factor $\bar{z} = 9$, one less than the average degree. So the best fit values for the growth factor $\bar{z}$ in the Erdős-Réyni networks are a little higher than this while the Barabási-Albert network and its randomised version are a lot bigger.

Another possible reference value for the shortest-path tree growth factor $\bar{z}$ is the average degree of a neighbour in a random graph with the same degree distribution which is $\langle k\rangle_{\mathrm{nn}} = \langle k^2\rangle/\langle k\rangle$. This is the relevant value for diffusive processes on a random graph. For our finite Erdős-Réyni networks we have that $\langle k\rangle_{\mathrm{nn}} \approx \langle k\rangle$ so again the growth factor found to give the best fit, $\bar{z}^{(\mathrm{fit})}$, in actual Erdős-Réyni networks is still a bit higher than this estimate. For the Barabási-Albert networks and their randomised versions, the $\langle k\rangle_{\mathrm{nn}}$ is around twenty-two to twenty-five for the networks in Table 1. This value is much closer but still not in complete agreement. This suggests our shortest-path trees are sampling nodes in different ways from diffusion but still with a bias to higher degree nodes.

Since spanning trees have many fewer edges than the original graphs, it is perhaps somewhat surprising that we find that the growth factors are comparable with any measures of the average degree in the original network. So the high values of $\bar{z}$ are telling us that the shortest-path trees are sampling the nodes of their networks with a large bias towards high-degree nodes in the parts of the tree close to the root node and that is why we need such a high growth rate $\bar{z}^{(\mathrm{fit})}$ when we fit our data for closeness. That way when we prune the edges to produce a tree we will still have high degrees in the tree close to the root node. The corollary is that the outer parts of shortest-path trees are dominated by leaves (degree one nodes) and other low-degree nodes, and these also correspond to low-degree nodes in the original network.

It is also clear that node correlations play an important role as these are present in the Barabási-Albert model but absent in the randomised version. The large difference in $\bar{z}$ values for these two cases show such node correlations are important and yet, the non-linear relationship in Eq. (6) still holds well in these artificial networks, with or without these correlations.

The β parameter in Eq. (5) is harder to interpret but Table 1 shows a comparison between the two values of β. The first is $\beta^{(\mathrm{fit})}$ derived from a two-parameter fit of the data to Eq. (5). The second value is $\beta(\bar{z}^{(\mathrm{fit})}, N)$ the value predicted using Eq. (6) where we use the $\bar{z}$ value obtained from the same two-parameter fit and the number of nodes N. What we can see is that the values derived using Eq. (6), $\beta(\bar{z}^{(\mathrm{fit})}, N)$, are consistently poorer than the values $\beta^{(\mathrm{fit})}$ derived from a two-parameter fit. It highlights that the details of our theoretical form, such as the precise formula for β, here Eq. (6), can be improved. However, our simple calculation has captured the important features of the problem so that the form Eq. (5) does work in these theoretical models provided we treat both $\bar{z}$ and β in Eq. (5) as free parameters to be determined.

Numerical results for real data. For networks representing real-world data, we used data that is open access and easily obtained[41–44]. We aimed for a wide range of networks both in terms of size and in terms of the type of interaction encoded in these real-world networks.

The first set of eighteen data sets we refer to as the Konect-SNAP networks. These were derived from real-world data and were chosen to reflect five broad categories of networks: social networks (`social-...`), communication networks (`commun-...`),

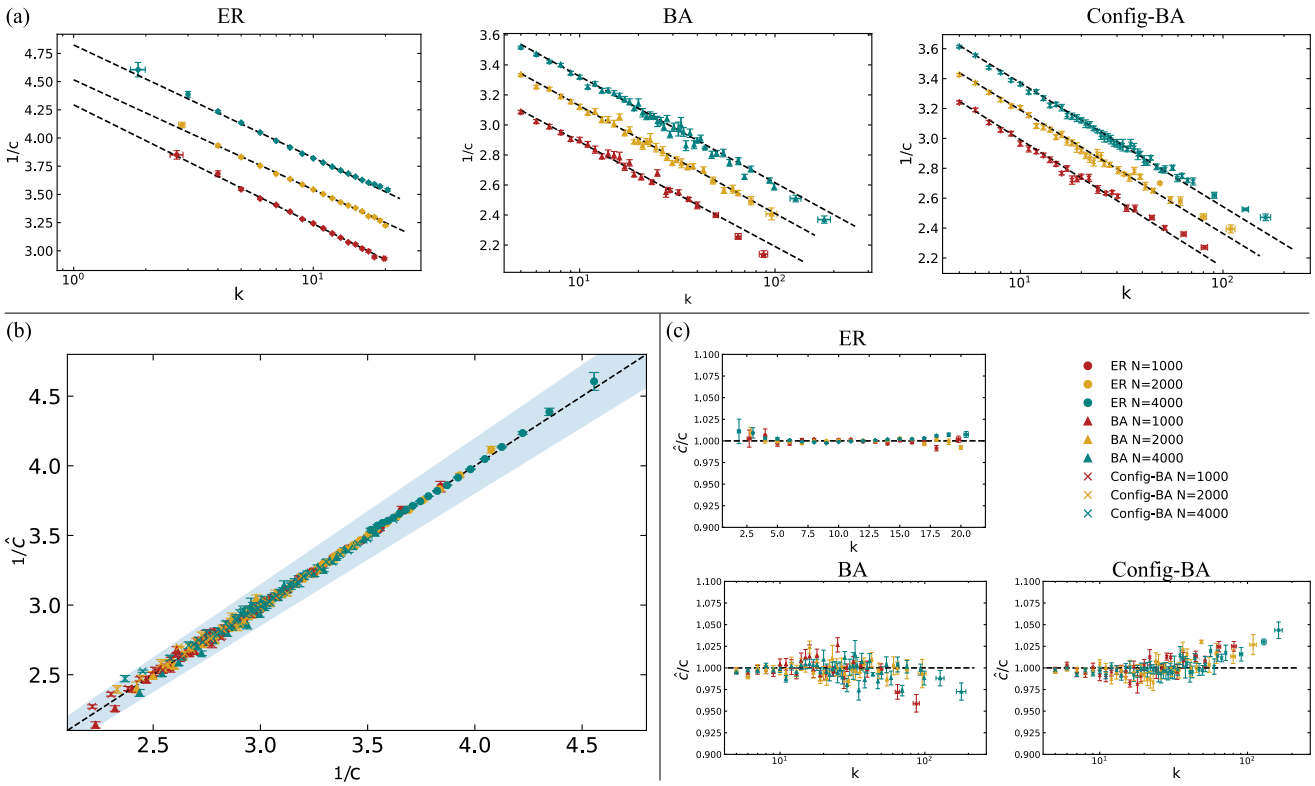


Fig. 2 Closeness and degree for artificial networks. In (**a**), each plot shows results for networks formed from one artificial model: the Erdős-Réyni (ER) model, the Barabási-Albert (BA) model, and the configuration model network starting from a Barabási-Albert model (Config-BA). The dashed lines shows the best linear fit of $1/c$ to $\ln(k)$ using Eq. (5). The same data from all nine artificial networks is shown in the scatter plot in (**b**) with data $1/\hat{c}$ against predicted value c obtained from the best fit Eq. (5) and the shaded region corresponds to a 5% deviation from the theoretical prediction. Panel **c** shows the fractional error, the fitted value of closeness divided by data value. The results are for three different sized networks: $N = 1000$ (red points) $N = 2000$ (blue points) and $N = 4000$ (yellow points) where N is the number of nodes. All networks have average degree 10.0 and 100 realisations were taken for each case. The values of closeness for each value of degree are binned, the mean is shown as the data point with error bars the standard error of the mean. The results show that the non-linear correlation of closeness and degree predicted in Eq. (5) works most of the time within a 2% variation. There are some hints of small but systematic at higher degree value but the data are sparse and less reliable here.

citation networks (`citation-...`), co-author networks (`coauth-...`), and hyperlink networks (`hyperlink-...`). A more detailed description of these Konect-SNAP networks is given in Supplementary Note 3 on data sets.

Summary statistics are given in Table 2. The reduced chi-square χ_r^2 measure is between 1.05 and 1.61 for ten, more than half, of our examples and another four networks have values between 2.09 and 2.86. Given the wide range of both size and nature of these networks and the simplicity of our theoretical derivation, this level of agreement may not have been expected. We also give the Pearson correlation measure between closeness and degree, $\rho(c, k)$, and this is generally high as has been noted before[23–32]. The success of our non-linear relationship between closeness and degree is not incompatible with high $\rho(c, k)$ values.

The data for each network is shown in more detail in Fig. 3. Again, we can see that within the error bars the average closeness at each degree generally follows the form we predict within 5% when the best fit parameters are used. Further, the uncertainties estimated for these data points suggest that the vast majority of average closeness values are statistically consistent with the predicted value for that degree, something already captured by the reduced chi-square values in Table 2.

Finally, we looked at our relationship between degree and closeness in 112 networks taken from the Netzschleuder archive of 276 network data sets[44]. Our only selection criterion was that we could automatically download a network and that it could be analysed successfully by our standard code without further work. This excluded several examples in this archive such as those with multiple networks (e.g. amazon_copurchases), or some which were too large for our code (e.g. academia_edu). Supplementary Note 3 contains further information on these Netzschleuder networks. As a result, our sample contains many for which we would not expect much success: some are very small, some are very dense, and some have added additional known structures, e.g., a bipartite network. At the same time there are many for which we would expect to be successful, typically anything sparse and large. We represented each as a simple graph and analysed the largest connected component which had up to 40,000 nodes and an average degree of <300.

The results are shown in Fig. 4 with additional information and a table of results on the Netzschleuder networks provided in Supplementary Note 4. We found 50 networks had an excellent fit with a reduced chi-square χ_r^2 close to 1.0 and always <2.0. Another 13 networks gave a reasonable fit $2.0 \le \chi_r^2 < 3.0$ and 9 networks had $3.0 \le \chi_r^2 < 4.0$. For the remaining networks, our code did not find a χ_r^2 for 13 of these as there was at most one node for each degree value and our code could not estimate the uncertainty in the measurement of closeness. Overall, of the 99 networks where we had a χ_r^2 result, our degree-closeness relationship was very successful ($\chi_r^2 < 2.0$) in these arbitrary networks 50% of the time.

Table 2 Results for the largest components the Konect-SNAP networks.

Network	N	$1/\ln(\bar{z}^{(\mathrm{fit})})$	$\beta^{(\mathrm{fit})}$	$\bar{z}^{(\mathrm{fit})}$	$\beta(\bar{z}^{(\mathrm{fit})}, N)$	$\rho(c,k)$	χ_r^2
social-karate-club	34	0.460 ± 0.066	2.997 ± 0.095	8.81 ± 2.76	2.69 ± 0.25	0.77	1.23
social-jazz	198	0.367 ± 0.015	3.349 ± 0.048	15.28 ± 1.72	2.98 ± 0.08	0.86	13.15
social-hamster	1788	0.353 ± 0.009	4.129 ± 0.020	17.05 ± 1.22	3.68 ± 0.07	0.68	1.20
social-oz	217	0.403 ± 0.010	3.492 ± 0.038	11.96 ± 1.02	3.22 ± 0.08	0.89	2.99
social-highschool	70	0.561 ± 0.039	3.734 ± 0.079	5.95 ± 0.74	3.48 ± 0.17	0.87	1.25
social-health	2539	0.537 ± 0.008	5.605 ± 0.016	6.43 ± 0.17	5.31 ± 0.06	0.75	1.07
commun-email	1133	0.394 ± 0.007	4.309 ± 0.014	12.64 ± 0.54	3.83 ± 0.05	0.84	1.06
commun-UC-message	1893	0.264 ± 0.003	3.526 ± 0.008	43.92 ± 2.16	3.01 ± 0.03	0.72	2.28
commun-EU(core)-email	986	0.259 ± 0.004	3.324 ± 0.012	47.63 ± 2.67	2.80 ± 0.03	0.84	2.48
commun-DNC-email	1833	0.222 ± 0.010	3.499 ± 0.012	91.16 ± 18.84	2.67 ± 0.08	0.41	1.38
commun-DIGG-reply	29,652	0.388 ± 0.002	5.078 ± 0.003	13.12 ± 0.18	5.05 ± 0.02	0.61	1.60
citation-DBLP	12,494	0.361 ± 0.003	4.856 ± 0.004	15.98 ± 0.35	4.45 ± 0.03	0.54	1.17
citation-Cora	23,166	0.503 ± 0.004	6.639 ± 0.008	7.31 ± 0.13	6.14 ± 0.04	0.48	1.10
coauthor-astro-ph	14,845	0.441 ± 0.004	5.735 ± 0.010	9.67 ± 0.21	5.30 ± 0.04	0.61	1.07
coauthor-netsci	379	0.382 ± 0.080	6.553 ± 0.119	13.74 ± 7.51	3.31 ± 0.48	0.35	1.27
coauthor-pajek	6927	0.259 ± 0.002	3.894 ± 0.002	47.45 ± 1.33	3.30 ± 0.02	0.64	12.56
hyperlink-polblog	1222	0.240 ± 0.004	3.316 ± 0.011	64.84 ± 4.44	2.72 ± 0.03	0.72	2.15
hyperlink-blogs	1222	0.239 ± 0.004	3.316 ± 0.011	65.07 ± 4.48	2.71 ± 0.03	0.72	2.18

These networks are derived from a variety of derived from real-world data, see the description of Konect-SNAP networks in Supplementary Note 3 on data sets for more details. The results for $1/\ln(\bar{z}^{(\mathrm{fit})})$ and $\beta^{(\mathrm{fit})}$ come from linear fits of inverse closeness, $1/c_v$ to the logarithm of degree, $\ln(k_v)$ for each vertex v, i.e., $1/c_v = (\ln(\bar{z}))^{-1}\ln(k_v) + \beta$ Eq. (5). The value of β derived from $\bar{z}^{(\mathrm{fit})}$ and N using Eq. (6) is also shown as $\beta(\bar{z}^{(\mathrm{fit})}, N)$ for comparison. The fits are very good as reduced chi-square χ_r^2 values show. The Pearson correlation measure between closeness and degree $\rho(c,k)$ is given for comparison with earlier work.

However, from Fig. 4 we can see that the density of a network, the number of edges divided by the number of node pairs, has an impact on the success rate. Our sample has included several very dense networks where there was a lot of additional information on the edges so the existence of an edge was much less important. Indeed from Table 3 we see that for Netzschleuder networks with densities below 0.04 we have around 70% fitting with a reduced chi-square of <2.0, rising to roughly 80% for $\chi_r^2 < 3.0$.

Using closeness. Our numerical results confirm our analytical work that the inverse of closeness depends linearly on the logarithm of degree Eq. (5) for most networks. This is a correlation, true on average but not an exact result for every node. It has long been known that nodes with larger degrees tend to have smaller closeness which leads to significant correlation measures[19,23–33] but this is often discussed in terms of the Pearson correlation coefficient which is most sensitive to linear correlations. A non-linear relationship between degree and closeness is not discussed in previous studies and our relationship puts this intuitive understanding of a degree-closeness relationship on a firm footing.

Our work suggests that in the majority of networks, to a first approximation, closeness of individual nodes captures little more information on average than is contained in the degree. There is no point spending time calculating closeness if you only want a rough measure of centrality, you may as well just use degree. However, closeness measurements can give useful information on a network if used correctly.

First, closeness and degree measurements yield our fitted parameters, $\beta^{(\mathrm{fit})}$ and $\bar{z}^{(\mathrm{fit})}$. These characterise each network and can be used to compare different types of networks and even networks of different sizes. In particular, $\bar{z}^{(\mathrm{fit})}$ characterises the exponential growth that is a feature of most complex networks and which is behind many well-known phenomena as we shall note later in this discussion.

There are a number of ways to determine $\bar{z}$ and β from our relation Eq. (5). We have found the most effective approach is the simpler method where we determined $\bar{z}$ and β from a linear fit of our data to Eq. (5). There are alternatives but these throw light on various possible approximations rather than being of practical use, see the sections on determining $\bar{z}$ and β in Supplementary Note 1 and results for higher-order polynomial fits in Supplementary Note 4.

For networks where degree and closeness are linked by our relationship Eq. (5), our work shows that useful information from closeness centrality for individual nodes can only come by comparing values for individual nodes against the expected value derived using our relationship Eq. (5) using the logarithm of degree. The only useful information in closeness values is the deviation from their expected value. In these situations, we could start by examining the degree centrality of every node. This would be the primary measure of centrality. We then fit our closeness values using Eq. (5) to produce an expected value of closeness $c_v^{(\mathrm{fit})}$ for each node. Finally, we use this fit to find nodes which are noticeably more (or less) central than expected. One way to do this is to look at the normalised closeness

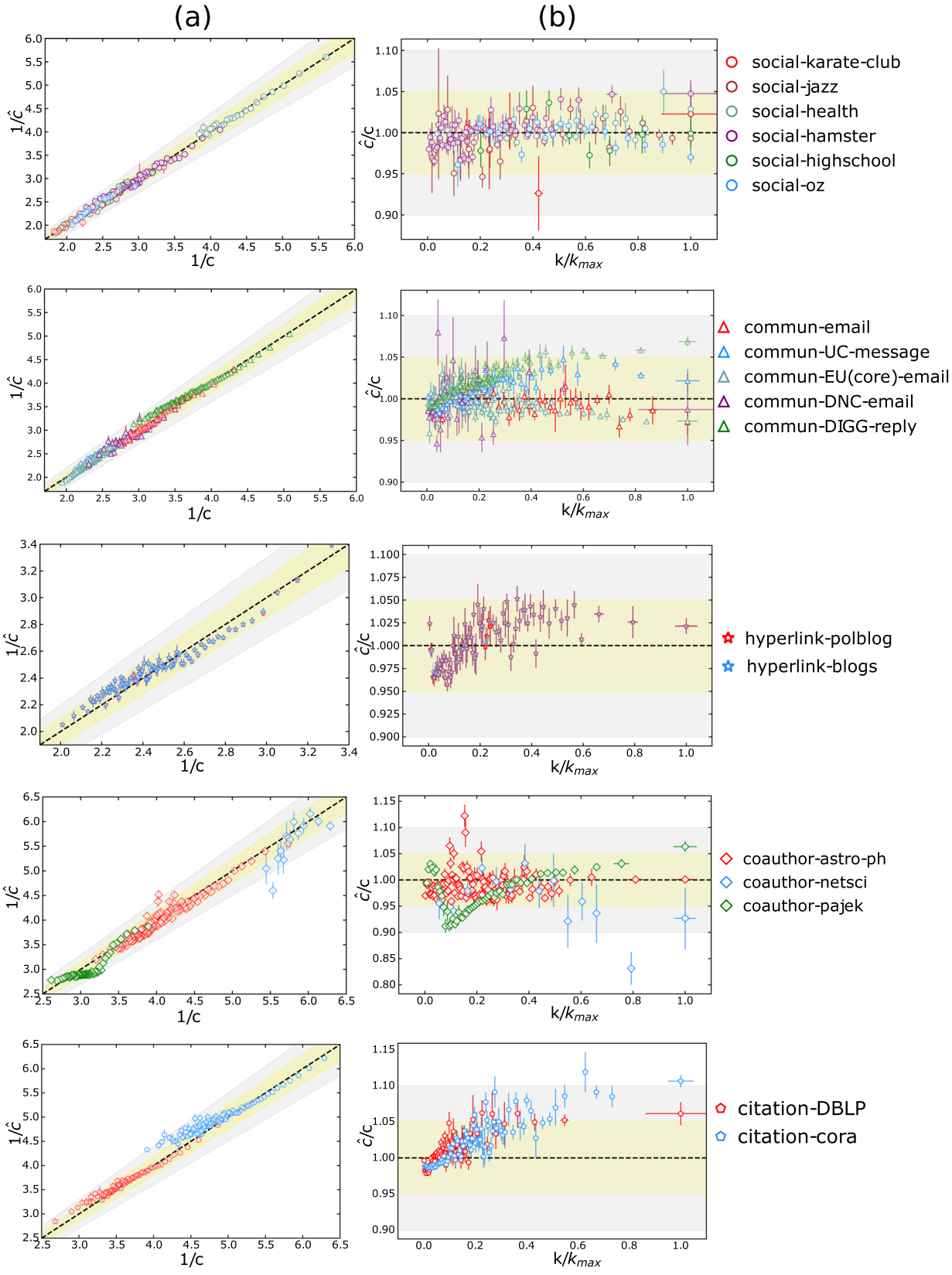
$$c_v^{(\mathrm{norm})} = \frac{c_v}{c_v^{(\mathrm{fit})}} \,. \tag{7}$$

Our normalised closeness measure $c_v^{(\mathrm{norm})}$ will highlight the outliers which would then be of most interest. See Supplementary Note 4 to see some examples of the fluctuations in closeness values around our predicted value.

One could also compare closeness to degree by running the configuration model, which keeps node degree constant, and measuring the closeness of each node in such a null-model network. However our method, requiring a simple linear fit to data already acquired, will be much faster than running the configuration model.

General network insights. The success of our conjecture also suggests that most networks satisfy two key assumptions built into our derivation.

First we assumed that the number of nodes a distance ℓ from any node grows exponentially. Such exponential growth is common in networks as it is the mechanism behind the concepts of the "six degrees of separation" and the "small-world" effect[45] often reported in networks. More formally, this is linked to length



scales in networks models with N vertices which grow as $O(\ln(N))$. This is to be contrasted with a network controlled by the geometry of a d-dimensional Euclidean space (such as Random Geometric Graphs) where the number of nodes at distance ℓ from a node grows as a power law ℓ^{d-1} and length scales in such networks grow as $O(N^{1/d})$. We will consider the length scales in more detail in the Network Length Scales subsection below.

This non-Euclidean behaviour of most networks highlights one situation where closeness is a useful centrality measure independent of and uncorrelated with degree, so a situation where our relationship Eq. (5) will fail. That is for graphs embedded in

Fig. 3 Closeness and degree for Konect-SNAP networks. Results for the largest components of eighteen Konect-SNAP networks derived from real-world data, see Table 2 for the statistics of each dataset. The yellow shaded region corresponds to 5% deviation and grey region corresponds to a 10% deviation. Plots in column (**a**) show the inverse of the predicted result $1/\hat{c}$ from the best fit against the inverse of the mean measured value $1/c = 1/\langle c \rangle_k$ averaged over nodes with the same degree k. Both axes are essentially $\ln(k)$. Column (**b**) shows the ratio of best fit value $\hat{c}$ over measured value c as a function of degree k rescaled by the largest degree in each network k_{max}. If the prediction matched data perfectly, points will lie on the dashed lines. The error bars represent from standard error of mean of the inverse closeness. For majority of points, we can see our prediction Eq. (5) captures the relation between closeness and degree, usually with within a 5% margin.

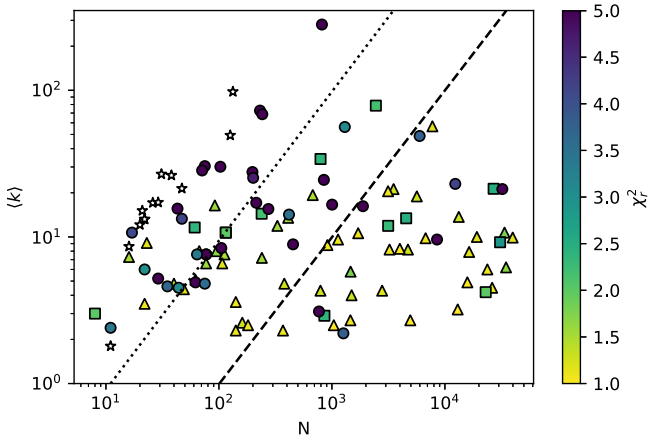


Fig. 4 Closeness and degree for Netzschleuder real-world networks. Results for the largest components of an additional 112 networks taken from the Netzschleuder archive[44]. Each point represents one network, plotted using the number of nodes N and the average degree $\langle k \rangle$ as coordinates. The symbol colour indicates reduced chi-square (χ_r^2) value with yellow being the best results (closest to 1.0). The symbol shape also indicates χ_r^2 values with triangles representing best results where $\chi_r^2 < 2.0$, squares are for $2.0 \leq \chi_r^2 < 3.0$ and circles represent poor performance with $3.0 \leq \chi_r^2$. The stars (white fill) are networks where our χ_r^2 calculation failed due to a lack of an error estimate on each point which occurs when we have one node per k value. To the right (left) of the dashed (dotted) line, networks have a density $\langle k \rangle/(N-1)$ of <0.01 (>0.1).

Table 3 The success rate for sparse Netzschleuder networks.

Upper Density	Number Networks	% Networks $\chi_r^2 < 2.0$	% Networks $\chi_r^2 < 3.0$
0.01	32	66%	84%
0.02	44	70%	84%
0.03	48	69%	81%
0.04	55	67%	80%

These results are for sparse networks taken from the 112 Netzschleuder networks analysed in this paper. Each row gives the statistics for all networks whose largest connected component has a density less than or equal to the density given. The percentage of these networks whose reduced chi-square is <2 or <3 is given in the last two columns.

Euclidean space. This is somewhat ironic as this type of planar network, such as a network representing the connections between intersections of streets in a city, is the prototypical example used to motivate the idea that the network measure closeness is related to our intuition about the concept of centrality in a network. Indeed, Bavelas[2] only used planar graph examples to develop closeness so a link between closeness and degree in most networks was never an issue in the original motivation for the closeness measure.

The second assumption that our work supports is that the branches of the shortest-path trees are statistically similar as illustrated in Fig. 1. The success of our analysis suggests this assumption works well whenever we are looking at measurements that depend on the bulk of the network. This simple approximation may therefore help analyse other network measurements, and we consider some of these in the Network Length Scales subsection below.

Closeness and real-world networks. Given the simple assumptions and approximations used in deriving our relationship Eq. (5), it is perhaps surprising to find that this closeness-degree relation works well for so many networks based on real-world data. So how often is our relationship a success? Also, can we understand when our closeness-degree relationship may succeed and when it is likely to fail?

For the first set of eighteen Konect-SNAP networks shown in Fig. 3 and Table 2, we had eleven (61%) with a reduced chi-square of <2.0 and all but two (89%) have a reduced chi-square with <3.0. The success rate for the 112 Netzschleuder networks, as measured by reduced chi-square, is lower. Of the ninety-nine Netzschleuder networks where we had a reduced chi-square measurement, fifty (51%) had a reduced chi-square of <2.0 while sixty-three (64%) had a a reduced chi-square of below 3.0. The lower success rate for the Netzschleuder networks can be understood as we are sampling without bias a varied collection of network data sets and, in many cases, we have information on the networks which suggests we might never have expected success.

The simplest issue is network density, the fraction of possible edges that are actually present in a network. If the density is very high, the distances between nodes will be low, perhaps just one or two. This leads to very small variations in the distances encoded in the closeness of each node making it harder to distinguish meaningful patterns in closeness measurements so our relationship is more likely to struggle with high-density networks. This trend is clear in Fig. 4. Of the two Konect-SNAP networks with poor chi-square, one, the `social-jazz` network, has a high density of 0.13. Conversely, we find that the success rate is much higher in our Netzschleuder networks for networks with density below 0.04: roughly two-thirds of these low-density networks fit our degree-closeness relationship well with $\chi_r^2 < 2.0$ and this rises to 80% if we accept anything with $\chi_r^2 < 3.0$, see Table 3.

More generally, a failure of our relationship is a powerful tool to highlight where there is an additional structure in the network. Our theoretical analysis assumes a generic network where every node sees the same exponential growth in the number of nodes at distance ℓ away as discussed in the General Network Insights subsection. If this generic structure is not present, we have no reason to expect our relationship to work well.

To see this structural issue, consider some of the notable exceptions to our low-density criterion. Looking at Fig. 4, the low-density networks with N between 700 and 2000 and $\langle k \rangle$ between 2.0 and 3.5, you can see two good networks with good fits (triangles) and three with poor fits (two circles and a square). One circle comes from the crime network with $N = 1263$, $\langle k \rangle = 2.2$ for which we find a reduced chi-square of $\chi_r^2 = 3.8$. The description of the crime network mentions that the data was

obtained using "snowball sampling from five initial homicides" so we are looking at a biassed sample of a much larger network. We suspect this sampling produces a different structure from a typical complex network leading to a failure of our relationship in this case. The other circle is the plant_pol_kato network with $N = 768$, $\langle k \rangle = 3.1$ and $\chi_r^2 = 34.1$ which is a "bipartite network of plants and pollinators" from a forest. Our analysis does not allow for a bipartite structure and in this case the difference between the two types of nodes is large with 91 plant species nodes but 715 nodes representing species of insects. The square represents the unicodelang network with $N = 858$, $\langle k \rangle = 2.9$ and $\chi_r^2 = 2.4$. This is still a good fit but it is also a "bipartite network of languages and the countries" so again, a better result might be achieved if we adapted our approach for two-mode networks.

In some cases, the meta-data we have for a network may already tell us about large-scale structure, such as a bipartite network, that will invalidate our relationship Eq. (5). In simple cases, it may be possible to adjust our derivation to find a more appropriate relation that matches the known structure. For instance, for bipartite networks we could represent our shortest-path trees using two growth rate parameters, $\bar{z}_a$ and $\bar{z}_b$, for odd and even distances from the root node.

In other cases, failure may indicate the presence of a structure that is not already known. For example, the network may have strong inhomogeneities such as high-degree nodes clustering together in a dense core. That would invalidate our assumption that all the branches of the shortest-path trees look statistically similar. We can see this type of problem in discussions of the average path length in random graph models[46,47]. There it was noted that the typical $\ln(N)$ dependence of length scales in complex networks is not seen when these models have a degree distribution of the form $p(k) \sim k^{-\gamma}$ with $2 < \gamma < 3$. In these cases this failure is linked to the network taking on a rather different structure, perhaps an "octopus"[46], a star-like structure with one dense core connected to many legs. This type of network invalidates our assumptions and we would not expect to see our closeness-degree relationship.

Network length scales. We can compare our analytical approach to that of other theoretical work on length scales in networks. The vast majority of the analytical results on distance in networks is for well-defined simple theoretical models, typically the Erdős-Réyni graph[17,46,48,49], the Barabási-Albert model[50,51], or scale-free random graphs[52] (our BA-Configuration model). The focus in the literature tends to be on characteristic length scales for a graph, such as the average path length $\langle \ell \rangle$ or the diameter, and not on length scales associated with each vertex, such as the closeness that we consider.

What is distinctive about our approach is that we focus on generic network properties that appear to hold in many networks. So our approach can be used on a much wider range of networks and in particular work for real-world networks not just for one simple model. The downside is that we do not have the mathematical rigour of those working with simple models. So, let us consider how our approach can be related to other length scale calculations in the literature.

We have already noted that the concepts of the "six degrees of separation" and the "small-world" effect[45] are linked to a network models where the length scale grows as $O(\ln(N))$, much slower that for $O(N^{1/d})$ behaviour expected for networks embedded in d-dimensional Euclidean space. One common way to study this more precisely is to consider the average distance between nodes $\langle \ell \rangle$, where each node pair contributes equally to the average. This is what we obtain if we take half the average of the inverse closeness Eq. (1) over all vertices, $\langle \ell \rangle = (2N)^{-1} \sum_r (c_r)^{-1}$. From our results for closeness, we can see that the N dependence in $\langle \ell \rangle$ comes from the $\ln(N)$ term in the expression for β of Eq. (6), see the section on average shortest-path length in Supplementary Note 1 for more details. So our result is consistent with the behaviour found more rigorously for the average path length in random graph models[46,47]. The success of our method in terms of this average path length result shows that small-world behaviour can be linked to the simple network features built into our method, those discussed in the General Network Insights subsection.

Our theoretical cutoff $L(N, k)$ of Eq. (3) gives another length scale for each node and again this has the same small-world $\ln(N)$ behaviour. One might conjecture that this cutoff length scale $L(N, k)$ could be linked to the eccentricity length scale. The eccentricity[6,16] e_r of a node r is the largest distance from r to any other node, $e_r = \max\{d_{rv} | v \in \mathcal{V}\}$. In that case we are making a new conjecture that eccentricity should also be linearly dependent of $\ln(k)$ with a gradient of $-1/\ln(\bar{z}^{(\mathrm{fit})})$ to match our theoretical cutoff $L(N, k)$. As seen in Fig. 1 of Wuchty & Stadler[6] and confirmed in our own analysis, eccentricity does depend linearly on $\ln(k)$ but the gradient does not seem to match the prediction from our theoretical cutoff $L(N, k)$. Roughly speaking our cutoff $L(N, k)$ represents a typical average large scale, not an extremal value of a distribution, so we should never expect a close link with eccentricity. For more details see the section on eccentricity and $L(N, k)$ in Supplementary Note 1.

As the largest eccentricity is the diameter of a graph, we can see the $O(\ln(N))$ behaviour of our $L(N, k)$ expression as matching this behaviour seen analytically in the diameter in simple models. However, just as for eccentricity and $L(N, k)$, we don't expect to be able to get a precise handle on diameter in our approach.

We also see strong similarities between our approach and that used in several papers[48,52] where the network is reduced to a set of rings of nodes, each ring containing all the nodes at the same distance ℓ from a root node. For instance, the mean first-passage time of random walkers on a network to a given vertex v is just the inverse of closeness where now closeness is defined in terms of a new distance function d_{uv} in Eq. (1), where d_{uv} is the average first-passage time for a random walker to move from vertex u to vertex v (this is also known as Markov centrality[53] or Random walk closeness centrality). Mean first-passage time has been observed to be proportional to degree[48] and applying our simple approximations to the "ring" method of Baronchelli & Loreto[48] quickly reveals this feature, see the section on the ring calculations of first passage times in Supplementary Note 1 for more details.

Overall, our approach can give insights into the behaviour of many network length scales, sometimes only very roughly, sometimes with more precision. It can never match the precision of the analytic calculations done for the simplest models but our approach can be used in a much broader range of networks.

Improvements. We have already noted one simple improvement when working with bipartite networks. That is to use two growth rate parameters, $\bar{z}_a$ and $\bar{z}_b$ for odd and even distances from the root node in the shortest-path tree of bipartite networks.

Other extensions are suggested by probing our numerical results in more detail. For our smaller set of 18 Konect-SNAP networks, beyond the two poorly fitting cases, our relationship Eq. (5) is very successful for most individual nodes within a 5% margin, a success which may not be expected given the simple analytical derivation. However, we can see some clear if small trends in the deviations in Fig. 3. We suggest these trends highlight the limitations of our analytical approach but it is possible to improve our theoretical methods.

At the simplest level, we could replace the sharp cutoff used for $n_\ell(r)$ where $n_\ell(r) = 0$ for $\ell > L$. There are examples of these distributions for some simple models in Baronchelli & Loreto[48]. A better cutoff may well lead to better predictions for β allowing one to fit a function with one independent parameter rather than two that we used by keeping β as a fitted parameter. However, while fitting one rather than two parameters may be theoretically satisfying, it does not seem much of a gain for the analysis of real-world networks.

Another option might be to calculate a different network parameter, namely the second degree $k_r^{(2)} = n_{\ell=2}(r)$[54] for each node r. By finding the number of nodes two steps away from every node, we can make a better approximation for $n_\ell(r)$, that is $n_0(r) = 1$, $n_1(r) = k_r$, and $n_\ell(r) = k_r^{(2)}\bar{z}^{\ell-2}$ for $2 \le \ell \le L_r$ and $n_\ell(r) = 0$ for $\ell > L_r$. This approach cannot be worse than the method used here as the latter is included as a special case where the second degree $k_r^{(2)} = \bar{z}k_r$ for all nodes r. To leading order we get the same type of result, namely that $1/c_r = (\bar{z})^{-1}\ln(k_r^{(2)}) + \beta$ since the degree k_r now only contributes a small number of terms to closeness. So in this approach using second degree we need to measure a different set of N parameters, the second degree of each node. Finding second degree is slower numerically than degree but both scale in the same way with increasing network size. The success of our simpler method here points to the idea that second degree and degree may often be correlated so it is likely that using second degree may only enhance results in a few cases.

More serious changes will be needed to the calculation if other effects neglected here, such as community structure or degree assortativity, are to be included.

Distance and logarithm of degree. The logarithm of degree $\ln(k)$ has been found to play an important role in network analysis before. A large fraction of papers on networks will show degree distributions where the horizontal axis is the dependent variable $\ln(k)$ and not simply the degree k. A more specific example comes from Zhou et al.[55] where the ratio of the degrees of nodes at the two ends of each edge (largest value in the numerator) is used to assign a 'distance' $\eta(u, v)$ to each edge (u, v). This is equivalent to defining $\lambda(u, v) = \ln(\eta(u, v)) = |\ln(k_u) - \ln(k_v)|$. In fact one can quickly see that while both η and λ are semi-distances on the set of edges in the formal mathematical sense, only λ is also a semi-metric and so λ is in some sense the more natural 'distance' measure in a qualitative sense. Our work suggests that an alternative view is to replace the logarithm of degree by the inverse of closeness. Since $(c_u)^{-1}$ is the actual average of the shortest-path distances from u to all other nodes, we can immediately see it is natural to work with inverse closeness when considering distances. For instance Zhou et al.[55] we could look at a different edge measure $\tilde{\lambda}(u, v) = |(c_u)^{-1} - (c_v)^{-1}|$. While the inverse closeness is a more natural distance, the degree is much easier to calculate in practice. Our work allows researchers to move between these two pictures.

Closeness and Gromov centrality. It has been noted[56] that the inverse of closeness is related to another centrality measure based on the average of the Gromov product. This centrality measure captures the extent to which the triangle inequality is not saturated between three nodes in a network, and so this is deeply connected to the geometry of a network. Our result leads to a natural prediction for this Gromov Centrality G_r^D of node r defined on the scale of the network's diameter D[56]. This Gromov Centrality is defined on other network length scales, G_v^ℓ, and Babul et al.[56] suggest there are useful generalisation of closeness. Our approximations will prove just as effective for such generalisations of closeness.

Conclusions

We have derived a non-linear relationship between degree and closeness Eq. (5), two of the most important centrality measures in Network Science. This was achieved by assuming that every node can be seen as the root of a shortest-path spanning tree where for all nodes these trees are statistically similar two or more steps away from the root. We have shown that this degree-closeness relationship works on a large number of artificial and real-world networks, particularly low-density networks where there is no special macroscopic structure.

Our results mean that measuring closeness is of little use unless our relationship is used to remove the dependence on degree from closeness. Further, most networks can be seen as shortest-path spanning trees which are statistically similar two or more steps away from their root nodes. Equally, if our relationship fails, it is an indication that our basic assumptions about network structure are wrong and so there are strong inhomogeneities and macroscopic structure in the data.

Data availability

All the data used in this paper is publicly available from several repositories. Our copies were downloaded from one of the following: KONECT[42], SNAP[43], and the Netzschleuder[44] repositories. More detailed information on some of the networks used along with additional references are also given in Supplementary Note 3 on data sets. The networks used in this study along with tables of many of the results are available in ref. [57].

Received: 23 September 2021; Accepted: 17 June 2022;
Published online: 02 July 2022

References

1. Brandes, U., Robins, G., McCranie, A. & Wasserman, S. What is network science? *Netw. Sci.* **1**, 1–15 (2013).
2. Bavelas, A. Communication patterns in task-oriented groups. *J. Acoust. Soc. Am.* **22**, 725–730 (1950).
3. Sabidussi, G. The centrality index of a graph. *Psychometrika* **31**, 581–603 (1966).
4. Freeman, L. C. Centrality in social networks conceptual clarification. *Soc. Netw.* **1**, 215–239 (1978).
5. Hage, P. & Harary, F. Eccentricity and centrality in networks. *Soc. Netw.* **17**, 57–63 (1995).
6. Wuchty, S. & Stadler, P. F. Centers of complex networks. *J. Theor. Biol.* **223**, 45–53 (2003).
7. Hahn, M. W. & Kern, A. D. Comparative genomics of centrality and essentiality in three eukaryotic protein-interaction networks. *Mol. Biol. Evol.* **22**, 803–806 (2004).
8. Koschützki, D. & Schreiber, F. Centrality analysis methods for biological networks and their application to gene regulatory networks. *Gene Regul. Syst. Biol.* **2**, GRSB.S702 (2008).
9. Kiss, C. & Bichler, M. Identification of influencers — measuring influence in customer networks. *Decis. Support Syst.* **46**, 233–253 (2008).
10. Yan, E. & Ding, Y. Applying centrality measures to impact analysis: A coauthorship network analysis. *J. Am. Soc. Inf. Sci. Technol.* **60**, 2107–2118 (2009).
11. Landherr, A., Friedl, B. & Heidemann, J. A critical review of centrality measures in social networks. *Bus. Inf. Syst. Eng.* **2**, 371–385 (2010).
12. Ni, C., Sugimoto, C. & Jiang, J. Degree, closeness, and betweenness: application of group centrality measurements to explore macro-disciplinary evolution diachronically. In Noyons, E., Ngulube, P., & Leta, J. (eds) *Proceedings of ISSI 2011*, (International Society for Scientometrics and Informetrics) 605 (2011).
13. Wang, J., Mo, H., Wang, F. & Jin, F. Exploring the network structure and nodal centrality of China's air transport network: a complex network approach. *J. Transport Geogr.* **19**, 712–721 (2011).
14. Brandes, U. & Hildenbrand, J. Smallest graphs with distinct singleton centers. *Netw. Sci.* **2**, 416–418 (2014).
15. Das, K., Samanta, S. & Pal, M. Study on centrality measures in social networks: a survey. *Soc. Netw. Anal. Mining* **8**, 13 (2018).
16. Wasserman, S. & Faust, K. *Social Network Analysis: Methods and Applications (Structural Analysis in the Social Sciences)* (Cambridge University Press, 1994).

17. Newman, M. *Networks: An Introduction* (Oxford University Press, 2010).
18. Coscia, M. *The Atlas for the Aspiring Network Scientist* (Michele Coscia, 2021).
19. Latora, V., Nicosia, V. & Russo, G. *Complex Networks: Principles, Methods and Applications* (Cambridge University Press, 2017).
20. Ma, H.-W. & Zeng, A.-P. The connectivity structure, giant strong component and centrality of metabolic networks. *Bioinformatics* **19**, 1423–1430 (2003).
21. Schoch, D. *A Positional Approach for Network Centrality*. Ph.D. thesis, Universität Konstanz (2015).
22. Schoch, D. Periodic table of network centrality. https://schochastics.net/sna/periodic.html (2016).
23. Bolland, J. M. Sorting out centrality: an analysis of the performance of four centrality models in real and simulated networks. *Soc. Netw.* **10**, 233–253 (1988).
24. Rothenberg, R. B. et al. Choosing a centrality measure: epidemiologic correlates in the colorado springs study of social networks. *Soc. Netw.* **17**, 273–297 (1995).
25. Faust, K. Centrality in affiliation networks. *Soc. Netw.* **19**, 157–191 (1997).
26. Lee, C.-Y. Correlations among centrality measures in complex networks. Preprint at https://arxiv.org/abs/physics/0605220 (2006).
27. Valente, T. W., Coronges, K., Lakon, C. & Costenbader, E. How correlated are network centrality measures? *Connections* **28**, 16 (2008).
28. Batool, K. & Niazi, M. A. Towards a methodology for validation of centrality measures in complex networks. *PLoS ONE* **9**, e90283 (2014).
29. Lozares, C., López-Roldán, P., Bolibar, M. & Muntanyola, D. The structure of global centrality measures. *Int. J. Soc. Res. Methodol.* **18**, 209–226 (2015).
30. Schoch, D., Valente, T. W. & Brandes, U. Correlations among centrality indices and a class of uniquely ranked graphs. *Soc. Netw.* **50**, 46–54 (2017).
31. Oldham, S. et al. Consistency and differences between centrality measures across distinct classes of networks. *PLoS ONE* **14**, e0220061 (2019).
32. Bringmann, L. F. et al. What do centrality measures measure in psychological networks? *J. Abnormal Psychol.* **128**, 892–903 (2019).
33. Arnaudon, A., Peach, R. L. & Barahona, M. Scale-dependent measure of network centrality from diffusion dynamics. *Phys. Rev. Res.* **2**, 033104 (2020).
34. Opsahl, T., Agneessens, F. & Skvoretz, J. Node centrality in weighted networks: generalizing degree and shortest paths. *Soc. Netw.* **32**, 245–251 (2010).
35. Harary, F. Status and contrastatus. *Sociometry* **22**, 23 (1959).
36. Šubelj, L. Algorithms for spanning trees of unweighted networks. Preprint at https://arxiv.org/abs/2205.06628 (2021).
37. Erdős, P. & Réyni, A. On random graphs. i. *Publ. Mathematicae* **6**, 290–297 (1959).
38. Barabási, A.-L. & Albert, R. Emergence of scaling in random networks. *Science* **286**, 173 (1999).
39. Molloy, M. & Reed, B. A critical point for random graphs with a given degree sequence. *Random Struct. Algorithms* **6**, 161–180 (1995).
40. Hagberg, A. A., Schult, D. A. & Swart, P. J. Exploring network structure, dynamics, and function using networkx. In Varoquaux, G., Vaught, T. & Millman, J. (eds.) *Proc. 7th Python in Science Conference (SciPy2008)*, 11–15 (2008).
41. Batagelj, V. Pajek datasets. http://vlado.fmf.uni-lj.si/pub/networks/data/ (2017).
42. Kunegis, J. The KONECT project. http://konect.cc/ (2013).
43. Leskovec, J. & Krevl, A. SNAP Datasets: Stanford large network dataset collection. http://snap.stanford.edu/data (2014).
44. Peixoto, T. P. The Netzschleuder network catalogue and repository. https://networks.skewed.de/ (2020).
45. Watts, D. J. & Strogatz, S. H. Collective dynamics of 'small-world' networks. *Nature* **393**, 440–442 (1998).
46. Chung, F. & Lu, L. The average distances in random graphs with given expected degrees. *Proc. Natl Acad. Sci. USA* **99**, 15879–15882 (2002).
47. Dorogovtsev, S. N., Mendes, J. F. F. & Samukhin, A. N. Metric structure of random networks. *Nucl. Phys. B* **653**, 307–338 (2003).
48. Baronchelli, A. & Loreto, V. Ring structures and mean first passage time in networks. *Phys. Rev. E* **73**, 026103 (2006).
49. Blondel, V. D., Guillaume, J.-L., Hendrickx, J. M. & Jungers, R. M. Distance distribution in random graphs and application to network exploration. *Phys. Rev. E* **76**, 066101 (2007).
50. Bollobás, B. & Riordan, O. The diameter of a scale-free random graph. *Combinatorica* **24**, 5–34 (2004).
51. Evans, T., Calmon, L. & Vasiliauskaite, V. The longest path in the Price model. *Sci. Rep.* **10**, 10503 (2020).
52. Cohen, R. & Havlin, S. Scale-free networks are ultrasmall. *Phys. Rev. Lett.* **90**, 058701 (2003).
53. White, S. & Smyth, P. Algorithms for estimating relative importance in networks. In *Proc. Ninth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining - KDD '03* (ACM Press), 266–275 (2003).
54. Falkenberg, M. et al. Identifying time dependence in network growth. *Phys. Rev. Res.* **2**, 023352 (2020).
55. Zhou, B., Meng, X. & Stanley, H. E. Power-law distribution of degree-degree distance: a better representation of the scale-free property of complex networks. *Proc. Natl Acad. Sci. USA* **117**, 14812–14818 (2020).
56. Babul, S'A., Devriendt, K. & Lambiotte, R. Gromov Centrality: A Multi-Scale Measure of Network Centrality Using Triangle Inequality Excess. Tech. Rep., Oxford University (2022).
57. Evans, T. S. & Chen, B. Linking the network centrality measures closeness and degree: additional data. figshare. https://doi.org/10.6084/m9.figshare.19216812 (2022).

Acknowledgements

T.S.E. would like to thank Max Hart, Oskar Hogburg and Luke Melville for initial investigations on this topic.

Author contributions

T.S.E. performed the analytical research and B.C. performed the numerical research. T.S.E. and B.C. designed the research, analysed data and wrote the paper.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s42005-022-00949-5.

Correspondence and requests for materials should be addressed to Tim S. Evans.

Peer review information *Communications Physics* thanks Filipi N. Silva and the other, anonymous, reviewer(s) for their contribution to the peer review of this work. Peer reviewer reports are available.

Reprints and permission information is available at http://www.nature.com/reprints

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit http://creativecommons.org/licenses/by/4.0/.

© The Author(s) 2022