DOI 10.1093/comnet/cnad016
ISSN 2051-1310
ESSN 2051-1329

Publisher: Oxford University Press

Journal of Complex Networks (2023) **3**, Advance Access Publication on 13 June 2023
https://doi.org/10.1093/comnet/cnad016

Correlation distances in social networks

PÁDRAIG MACCARRON AND SHANE MANNION
MACSI, Department of Mathematics and Statistics, University of Limerick, Limerick V94 T9PX, Ireland
*Corresponding author. Email: padraig.maccarron@ul.ie

AND

THIERRY PLATINI
Applied Mathematics Research Centre, Coventry University, Coventry CV1 5FB, UK

[Received on 21 December 2022; editorial decision on 30 March 2023; accepted on 5 May 2023]

In this work, we explore degree assortativity in complex networks, and extend its usual definition beyond that of nearest neighbours. We apply this definition to model networks, and describe a rewiring algorithm that induces assortativity. We compare these results to real networks. Social networks in particular tend to be assortatively mixed by degree in contrast to many other types of complex networks. However, we show here that these positive correlations diminish after one step and in most of the empirical networks analysed. Properties besides degree support this, such as the number of papers in scientific coauthorship networks, with no correlations beyond nearest neighbours. Beyond next-nearest neighbours we also observe a disassortative tendency for nodes three steps away indicating that nodes at that distance are more likely different than similar.

Keywords: social networks; assortativity; correlation distances.

1. Introduction

In recent decades, many quantities have been introduced to shed light on the structural properties of complex networks. Of particular importance is degree assortativity, which provides a measure of the correlations in the degree of neighbouring nodes. Its importance was underlined due to it distinguishing the structure of social networks, and other types of complex networks [1] (though this is found to not always be the case for online social networks [2]). Studies of assortativity have focused on nearest neighbour correlations, since neighbours interact directly only with their nearest neighbours, by construction.

By drawing parallels with spin systems, such as that described by the Ising model, one can argue that system components may exert influence indirectly, beyond their immediate neighbourhood. This is captured in the idea of *correlation length*. In the Ising model, while an individual spin has no direct interaction with more distant spins, a re-orientation of a spin can influence its neighbours causing a disturbance to propagate over a large area of the lattice, modelling the observed behaviour of ferromagnets. Regions separated by more than this characteristic distance can be thought of as essentially independent [3]. Here, we examine whether degree correlations in complex networks exist beyond those found between nearest neighbours.

In this work, we study correlations on undirected, unweighted networks as a function of network distance. The most elementary feature of a network node is its *degree*, defined as the number of edges adjacent to that node. The network-wide correlation between degrees of neighbouring nodes is commonly known as *assortativity* [4]. A network with strong degree correlations between nodes is called assortative, and with anti-correlations, disassortative. We use the shortest path length between two nodes to represent the distance, when this is greater than one we are moving beyond the traditional nearest neighbours.

In social networks, where node degree represents the number of friends that an individual has, assortativity indicates that popular individuals tend to have popular friends, and unpopular individuals unpopular friends. This notion is related to that of *homophily*, whereby individuals associate with people similar to themselves. This leads to a tendency of individuals to associate mostly with others of a similar race or ethnicity, age, religion or interest [5].

Previous studies have suggested that nodes' influence may extend beyond the immediate neighbourhood in social networks (e.g. Ref. [6]). It has even been claimed that individuals can have up to 'three degrees of influence' on other nodes in a network [7]; though serious questions have been raised about the methods of these studies [8]. Here, we use the assortativity at different distances to test this. We study path lengths up to the diameter of our graphs for 16 empirical social networks. We then use one co-authorship network in three time points, if nodes are expected to influence each other at a distance, naïvely we might expect the beyond nearest neighbour correlations to increase over time.

The article is structured as follows: we outline the similar research in this area, then we describe the theoretical framework for our correlation measure and introduce an expression to quantify assortativity as a function of distance. We then apply this to 16 social networks to observe correlation behaviour beyond nearest neighbours. We then perform simulations on random graphs, configuration model variants of each empirical network, as well as generated assortative networks, to attempt to generalize the results we observe.

2. Background

We are not the first to consider beyond nearest neighbours in complex networks. This is of course not surprising with similar ideas existing in spin systems in statistical physics. Related works have examined how correlations vary with distance in a number of ways, typically using the length of the shortest path between connected vertices, though not always.

References [9] and [10] approach the problem in similar ways. Both use probability distributions to calculate the long-range degree correlations within the network. The former exclusively uses Erdős-Rényi graphs to demonstrate their findings while the latter also shows some results for some empirical networks. In this article, we do not use these probability distributions. Instead, we look at the correlations between the degrees of nodes at different distances as an extension to the calculation of assortativity, in a similar fashion to what is done in Ref. [11]. This allows us to express the correlation between the degrees of nodes as a function of the distance between them.

Our work expands on what is done by Mayo *et al.* [11] in two ways. First, we apply our analysis to many more datasets (in particular social networks). Second, we include several types of random graphs in our analysis not covered by Mayo *et al.* [11]. Namely we study configuration model graphs based on some of the networks studied, Erdős-Rényi graphs, and Erdős-Rényi graphs which we rewire to be assortative. Although Mayo *et al.* [11] are more interested in average degrees at distances, they do look at correlations on one particular social network which is disassortative. Therefore, they do not find global results for social networks with correlations at a distance. We find that social networks become disassortative by $d = 3$.

Another approach is taken by Rybski *et al.* [12]. They are primarily interested in random network models such as the Barabási-Albert model, Cayley trees and fractal models, in addition to just a couple of real-world networks. The approach in this article is inspired by time-series analysis techniques. They look at how the mean degree along shortest paths of length differs with . Due to the differences between this approach and our own, results between the two are not directly comparable.

Arcagni *et al.* [13] discuss several measures of assortativity. One is based upon degree-based paths similar to that used by Rybski *et al.* [12]. Another is based on random walks in a network and finally a measure based on shortest paths as we have here, and as was described by Mayo *et al.* [11].

A measure of assortativity is introduced by Allen-Perkins *et al.* [14] which they call two-walk assortativity. This measure, in contrast to the one discussed in this article, looks at the correlation between the number of second neighbours of nodes connected by walks of length 2. One difference between the results we obtain here and those obtained by Allen-Perkins *et al.* [14] is that they find that no graphs which are degree assortative and two-walk disassortative, and even conjecture that such graphs cannot exist. With the measure we study here however, we find graphs that are assortative at $d = 1$ and disassortative at $d = 2$ (e.g. the Pretty Good Privacy (PGP) network shown in Fig. 2).

As we can see from the approaches taken by various authors in this section, there are many ways in which assortativity can be extended beyond nearest neighbors. We argue, however, that the derivation we use here is simplest to follow and most easily extended to all possible distances between connected nodes in a network. Furthermore, we apply our measure to a large number of datasets focusing on social networks. To our knowledge, this is the largest such analysis of long-range correlations in social networks to date.

3. Theory

In this section, we define a number of quantities centred around the concept of assortativity. We start by defining a graph or network as an ordered pair $G = (\mathscr{V}, \mathscr{E})$, where $\mathscr{V}$ is the set of nodes and $\mathscr{E}$ is the set of edges. The number of nodes and edges in a network are written as $N = |\mathscr{V}|$ and $L = |\mathscr{E}|$, respectively. Denoting the adjacency matrix by A, with elements A_{ij}, the degree k_i of node i is given by $k_i = \sum_j A_{ij}$. The mean degree of the network is given by $k\rangle = \sum_i k_i/N$. Although our results may be extended to directed graphs, for the purposes of this study, we focus solely on undirected networks, meaning that $A_{ij} = A_{ji}$. It follows that $2L = \sum_{i,j} A_{ij} = \sum_i k_i = \sum_k kp_k$, where p_k is the degree distribution so the final term is just N times the mean degree of the network.

The shortest path between nodes i and j is the path from i to j that traverses the minimal number of edges. We define the length of this path by λ_{ij}, noting that the path itself is often not unique. It is common for networks to consist of disconnected components or sets of nodes that are mutually reachable by traversing edges. When this is the case, there exists no path between nodes i and j belonging to distinct components and, by convention, we then set $\lambda_{ij} = \infty$. To define the average path length , it is convenient to introduce , the number of components of a graph and to define C_m, with $m \in \{1, 2, \ldots, \}$), as the mth component, which contains n_m nodes. Summing over all pairs of a connected component, the average path length is given by

$$= \sum_{m=1} \frac{1}{n_m(n_m - 1)} \sum_{ij \in \mathscr{C}_m} \lambda_{ij}. \tag{3.1}$$

The greatest shortest path length $_{\max}$ is known as the diameter of the network.

3.1 *Assortativity*

Consider the two nodes at the extremities of a randomly chosen edge. The probability of one node having degree k and the other degree q is

$$\Phi(k,q) = \frac{1}{2L}\sum_{i,j} A_{ij}\delta_{k_i k}\delta_{k_j q}, \tag{3.2}$$

where $\delta_{k_i k}$ is the Kronecker delta which is one when $k_i = k$ and zero otherwise.

We denote the average degree of the node at the end of an edge by $E[k]$. We stress that an average over edges is not the same as an average over nodes and use the notation $E(k)$ to represent the average summing over the set of edges. These quantities are, however, closely related. Further, note that since we have undirected graphs, $\mathrm{E}[q] = \mathrm{E}[k]$. Summing over q, the marginal probability is given by

$$\phi(k) = \sum_q \Phi(k,q). \tag{3.3}$$

Using this expression, it is possible to show that $\mathrm{E}(k)$ and the mean degree over the nodes of the network are related by

$$\mathrm{E}(k) = \sum_k k\phi(k) = \frac{\sum_k k^2 p_k}{\sum_k k p_k}, \tag{3.4}$$

which is the degree variance divided by the mean (i.e. the second moment of the degree distribution divided by the first moment). Note that any network besides those whose nodes have identical degree has positive degree variance. This implies that, except in the degree-regular case, $\mathrm{E}(k)$ is greater than the mean degree of the nodes of the network. This means that on average, individuals are connected to those more popular than themselves. This observation is commonly known as the friendship paradox.

Our starting point towards understanding degree correlations and assortativity is the expected value of the product of the degrees of the nodes at the end of an edge chosen uniformly at random, $\mathrm{E}(kq)$, defined by

$$\mathrm{E}[kq] = \sum_{k,q} kq\Phi(k,q). \tag{3.5}$$

For uncorrelated, undirected graphs, the probability distribution $\Phi(k,q)$ can be factorized as $\Phi(k,q) = \phi(k)\phi(q)$. The latter relation immediately implies

$$\mathrm{E}[kq] = \mathrm{E}[k]\mathrm{E}[q] = \mathrm{E}[k]^2 = \mathrm{E}[q]^2. \tag{3.6}$$

It follows that a network is said to have positive correlations when $\mathrm{E}[kq] > \mathrm{E}[k]\mathrm{E}[q]$ and negative correlations when $\mathrm{E}[kq] < \mathrm{E}[k]\mathrm{E}[q]$. We can now introduce the degree assortativity r as

$$r = \frac{\mathrm{E}[kq] - \mathrm{E}[k]\mathrm{E}[q]}{\sigma_k^2}, \tag{3.7}$$

which is in a form of the Pearson correlation coefficient [4], with σ_k^2 the variance of k,

$$\sigma_k^2 = \mathrm{E}[k^2] - \mathrm{E}[k]^2. \tag{3.8}$$

Normalizing by the variance ensures that $-1 \leq r \leq 1$. If $r > 0$, the network is said to be assortative, while if $r < 0$, it is said to be disassortative. If $r = 0$, the network is uncorrelated. Using Equations (3.5) and (3.2), the assortativity in Equation (3.7) can be written explicitly as

$$r = \frac{1}{2L} \sum_{i,j} \frac{A_{ij}(k_i - \mathrm{E}[k])(k_j - \mathrm{E}[k])}{\sigma_k^2}. \tag{3.9}$$

It is helpful to understand the assortativity in this way as we determine whether each edge contributes positively or negatively to the assortativity by knowing if the nodes are at different or the same sides as the mean. This expression also highlights that assortativity is the measure of the correlations between degree fluctuations (around the mean value) on two neighbouring nodes. In particular, if $A_{ij} = 1$, the product,

$$(k_i - \mathrm{E}[k])(k_j - \mathrm{E}[k]), \tag{3.10}$$

shows that edges for which, at both extremities $k_i > \mathrm{E}[k]$ and $k_j > \mathrm{E}[k]$ or, $k_i < \mathrm{E}[k]$ and $k_j < \mathrm{E}[k]$ contribute positively to the assortativity. However, edges for which $k_i < \mathrm{E}[k] < k_j$ or $k_j < \mathrm{E}[k] < k_i$ will contribute negatively. Therefore, the assortativity expresses the weighted balance between the number of edges for which both extremities have a degree under or over the mean and the number of edges for which the mean sits between the degree of each extremities.

We propose an alternative way to approach the assortativity and define the probability $\Psi(K, \Delta)$, where, if $A_{ij} = 1$, $K = k_i + k_j$ and $\quad = k_i - k_j$. This is the probability that the degrees of the nodes at the end of a randomly chosen edge sum to K, and differ by . This is given by

$$\Psi(K, \Delta) = \frac{1}{2L} \sum_{ij} A_{ij} \delta_{k_i+k_j,K} \delta_{k_i-k_j,\ } . \tag{3.11}$$

Naturally, the averages of any observable $\mathscr{O}(K, \Delta)$ are defined as $E[\mathscr{O}(K, \Delta)] = \sum_{K,\ } \Psi(K, \Delta)\mathscr{O}(K, \Delta)$. We can show that the assortativity becomes

$$r = \frac{\sigma_K^2 - \sigma^2}{\sigma_K^2 + \sigma^2}, \tag{3.12}$$

where σ_K^2 and σ^2 are defined as

$$\sigma_K^2 = E[K^2] - E[K]^2, \tag{3.13}$$

$$\sigma^2 = E[\ ^2]. \tag{3.14}$$

With this picture we now see that the assortativity compares the variance of the marginal distribution $\psi_+(K)$ and $\psi_-(\Delta)$,

$$\psi_+(K) = \frac{1}{2L} \sum_{ij} A_{ij} \delta_{k_i+k_j,K}. \tag{3.15}$$

$$\psi_{-}(\Delta) = \frac{1}{2L}\sum_{ij} A_{ij}\delta_{k_i - k_j}, \quad . \tag{3.16}$$

Plotting this distribution allows one to visualize the behaviour of the assortativity. Finally, note that the expected statistical error for the assortativity can be calculated using either the jackknife method [15] or the bootstrap method [16].

3.2 *Generalization*

In this section, we provide a definition of assortativity beyond nearest neighbours. We consider the simplest possible approach, by defining the matrix $A(d)$ with elements being one (respectively, zero) whenever there exist at least one (respectively, no) shortest path, of length d, between nodes i and j. In other words, $A_{ij}(d) = 1$ if there is a path of length d between the two nodes and if $A_{ij}(l) = 0$ for all $l < d$. This allows us to define $\Phi(\kappa, \eta, d)$ as

$$\Phi(\kappa, \eta, d) = \frac{1}{2L(d)}\sum_{ij} A_{i,j}(d)\delta_{k_i\kappa}\delta_{k_j\eta}, \tag{3.17}$$

with $2L(d) = \sum_{i,j} A_{ij}(d)$.

The assortativity at a distance d is

$$r(d) = \sum_{ij} \frac{A_{ij}(d)}{2L(d)} \frac{(k_i - \mathrm{E}[k](d))(k_j - \mathrm{E}[k](d))}{\sigma_k^2(d)}, \tag{3.18}$$

with $\sigma_k^2(d) = \mathrm{E}[k^2](d) - \mathrm{E}[k](d)^2$. This can be extended to other properties of nodes other than degree as we will demonstrate later (see Fig. 3).

4. Application

In this section, we analyse the assortativity of 16 different social networks of varying size. We focus on Equation (3.16) for these empirical networks, and then calculate the assortativity at a distance for each network.

The datasets we use are:

(i) the students at Faux Mesa high school friendship network [17, 18];
(ii) the American jazz musicians from 1912 to 1940 [19];
(iii) the face-to-face contact of participants at an infectious diseases exhibition in Dublin's Science Gallery [20];
(iv) a coauthorship network of scientists working on network theory and experiment [21];
(v) a Friendship network at Moreno high school [22];
(vi) a Friendship network between users of the petster website [23];
(vii) a key-sharing network from the PGP web of trust [24];
(viii) collaboration networks of authors on the condensed matter *arXiv* from 1995 to 2005 [25];

TABLE 1 *The number of nodes is given by N, L is the number of edges, k is the mean degree, k_{max} is the largest degree in the network, k^2 is the mean of the square of the degree (related to the variance). The average path length is given by and the diameter by $_{max}$. The assortativity is denoted by r and the symbol * signals that the statistical error is larger than this value indicating there are no degree–degree correlations*

Network	N	L	k	$k_{\max}$	k^2		$_{\max}$	r
Faux Mesa high school	147	202	2.75	13	11.70	6.81	16	0.12
Jazz Musicians	198	2,742	27.70	100	1,070.24	2.24	6	0.02*
Infectious diseases	410	2,765	13.49	50	252.43	3.63	9	0.23
Network Science	1,461	2,742	3.75	34	26.05	5.82	17	0.46
Moreno-Health	2,539	10,455	8.24	27	86.41	4.56	10	0.25
Petster	2,426	16,631	13.71	273	582.93	3.59	10	0.05
PGP web of trust	10,680	24,316	4.55	205	85.98	7.49	24	0.24
cond-mat arXiv	16,726	47,594	5.69	107	73.57	6.63	18	0.19
Astrophysics	18,771	198,050	21.10	504	1,379.51	4.19	14	0.21
Twitter	23,370	32,831	2.81	238	108.17	6.30	15	−0.48
Gplus	23,628	39,194	3.32	2,761	1,250.88	4.03	8	−0.39
Munmun	30,398	86,312	5.68	285	159.74	4.67	12	0.01
Facebook Wall	46,952	193,494	8.24	223	202.87	5.60	18	0.25
Facebook	63,731	817,035	25.64	1,098	2,256.80	4.32	15	0.18
Slashdot	79,116	467,731	11.82	2534	1,729.86	4.04	12	−0.07
Enron email	87,273	299,220	6.86	1,728	1,147.24	4.89	13	−0.17

(ix) the Astrophysics collaboration network [26];

(x) a Twitter user network [27];

(xi) a Google+ user network [27];

(xii) a reply network for news site digg [28, 29];

(xiii) a Facebook wall post network [30];

(xiv) a sample of Facebook users in the New Orleans region [31];

(xv) a network of users of the technology news site slashdot [32] and

(xvi) the network of emails sent at Enron [33, 34].

All datasets bars (ii), (iii) and (viii) can be found on Konect.cc [35]. Konect is an open-source library of network datasets taken from a wide range of scientific areas.

Some basic properties of these networks are listed in Table 1. In Refs [1, 4], it is observed that social networks tend to be assortative while non-social networks are not. Here, four of these social networks are found to be disassortative, and one is almost zero (with the error being larger than the value). The disassortativity in the network of Marvel Universe characters is described as being one of the reasons it is dissimilar to real social networks in Ref. [36]. The Twitter and Google+ networks show the same disassortative behaviour of other large online social networks [2]. The slashdot network also contains

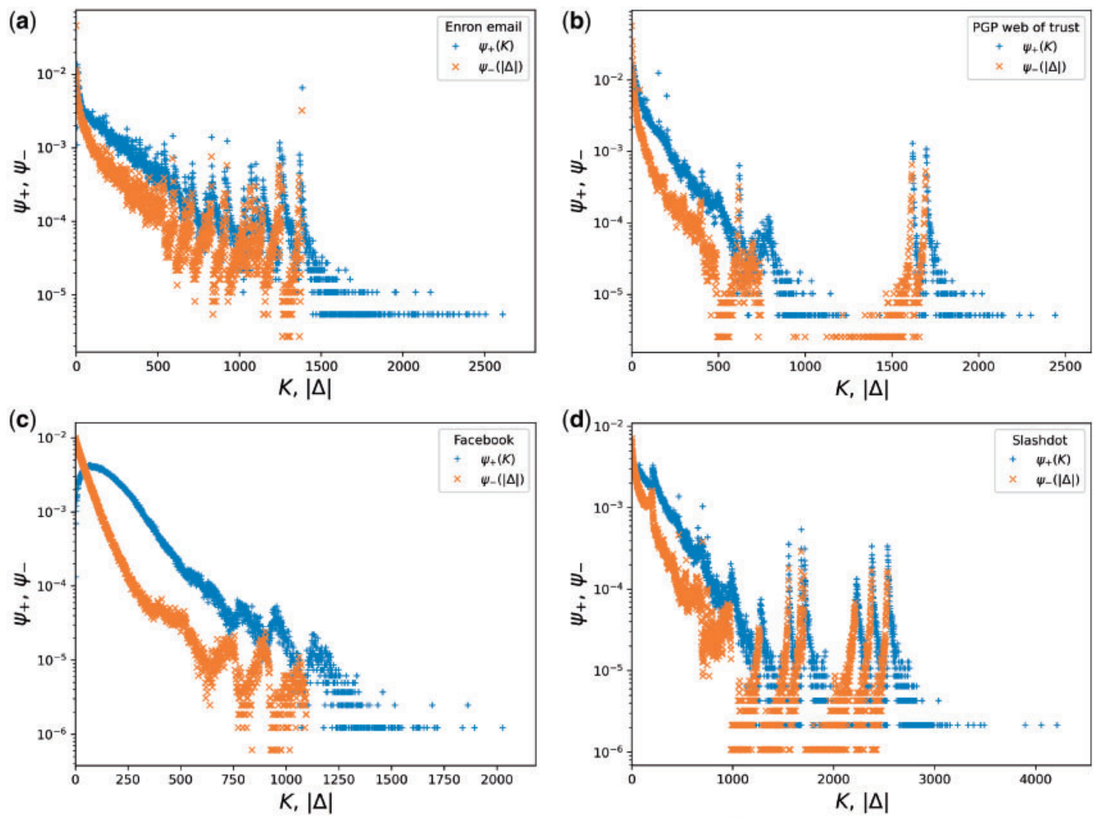

FIG. 1. Distributions $\psi_+(K)$ and $\psi_-(\Delta)$, as indicated, of the sum and difference in degrees of nodes at either extremity of an edge, respectively. Panel (a) shows the Enron email network. This network is disassortative and there are numerous edges with a difference of degrees of $|\quad| = 1,382$. In contrast, the PGP web of trust $\psi_-(\Delta)$ distribution is shown in Panel (b). This is assortative and has less noise in the tail of the $\psi_-(\Delta)$ distribution. Panel (c) displays the Facebook users network which is also assortative. Panel (d) however is for the online social network of slashdot users. This network is disassortative and contains a large number of fluctuations as $|\quad|$ increases.

'hostile' or 'negative' edges. As shown in Refs. [37, 38], hostile edges contribute towards disassortativity, removing the negative edges however only raises the assortativity from -0.07 to -0.06 and does not have much effect on the other network properties. To examine this further, for the slashdot an Enron email network, we plot the $\psi_+(K)$ and $\psi_-(\Delta)$ distributions from Equations (3.15) and (3.16) in Fig. 1, and compare them to two assortative networks.

In Fig. 1(a), the distributions $\psi_+(K)$ and $\psi_-(\Delta)$ are shown for the Enron email network. In the tail of the $\psi_-(\Delta)$ distribution in particular, we observe a high fraction of edges with $|\quad| = 1,382$. Hence, the node with the highest degree interacts with multiple nodes with a degree of one which interact with no other nodes in the network. This strongly contributes towards the disassortativity. The PGP web of trust network on the other hand has a relatively low maximum degree and a low degree variance. It is assortative, and we observe in Fig. 1(b) that there are comparatively few fluctuations in . However, note one large peak around 1,700, as this peak for $\psi_-(\Delta)$ is very close to the peak for $\psi_+(K)$, this shows the presence of a hub interacting with low degree vertices (i.e. for the sum of the degrees K to be close

TABLE 2 *The number of pairs $n_{\text{pairs}}(d)$ and their corresponding assortativity values $r(d)$ up to $d = 3$ for each of the 12 networks. The value in parentheses after the assortativity is the error in the last digit calculated using the bootstrap method. For most networks, the assortativity goes to zero (or fluctuates around it) after a small number of steps. A notable exception is that of jazz musicians which becomes more disassortative as d increases*

Network	$n_{\text{pairs}}(1)$	$r(1)$	$n_{\text{pairs}}(2)$	$r(2)$	$n_{\text{pairs}}(3)$	$r(3)$
Faux Mesa high school	202	0.12(8)	410	−0.14(5)	561	−0.05(5)
Jazz Musicians	2,742	0.02(2)	10,652	−0.13(1)	5,067	−0.31(2)
Infectious diseases	2,765	0.23(2)	13,150	−0.00(1)	24,631	−0.06(1)
network science	2,742	0.46(3)	3,980	−0.03(2)	6,365	−0.03(2)
Moreno-Health	10,455	0.25	71,624	0.09	365,757	−0.00
Petster	16,631	0.05	222,693	−0.14	757,751	−0.17
PGP web of trust	24,316	0.24(1)	188,183	0.01(1)	932,993	−0.09(1)
cond-mat arXiv	47,594	0.19(1)	275,120	0.07(1)	1,439,255	−0.02(1)
Astrophysics	198,050	0.21	4,441,041	−0.08	35,788,064	−0.19
Twitter	32,831	−0.48	1,110,048	0.12	2,534,785	−0.11
Gplus	39,194	−0.39	12,928,409	0.00	4,494,516	−0.01
Munmun-Digg	86,312	0.01	2,169,143	−0.06	33,925,631	−0.14
Facebook Wall	193,494	0.25	3,149,984	0.06	32,678,260	−0.10
Facebook	817,035	0.18	31,073,418	−0.04	357,008,740	−0.14
Slashdot	467,731	−0.07	47,324,811	−0.08	696,650,899	−0.11
Enron email	299,220	−0.17	27,781,955	−0.05	282,174,082	−0.07

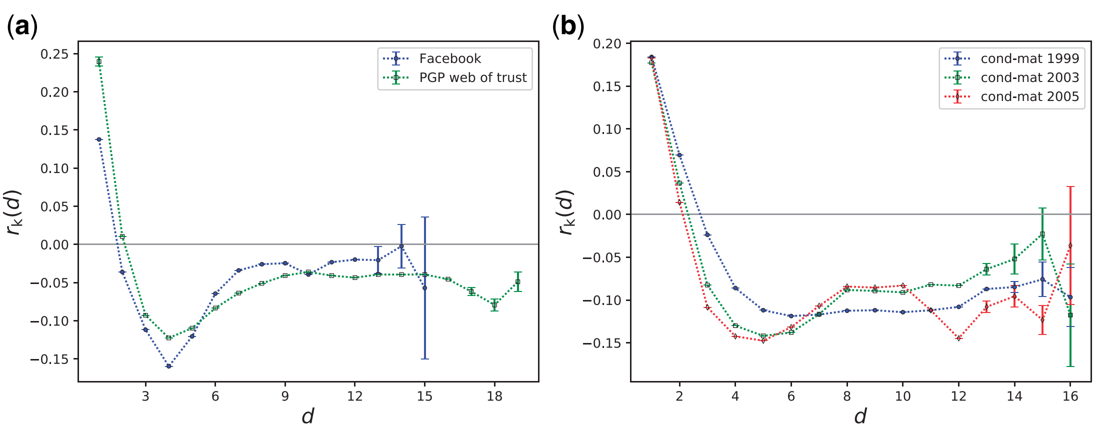

FIG. 2. The degree–degree correlations $r(d)$ plotted versus the distance d. Panel (a) shows that as the distance d increases for the PGP and Facebook networks, they lose their correlations. Panel (b) shows the correlations for three time periods on the condensed matter arXiv, which after $d = 2$ become anti-correlated. Error bars are calculated using the bootstrap method (they become more prominent at the end due to a significantly smaller number of pairs separated for large d).

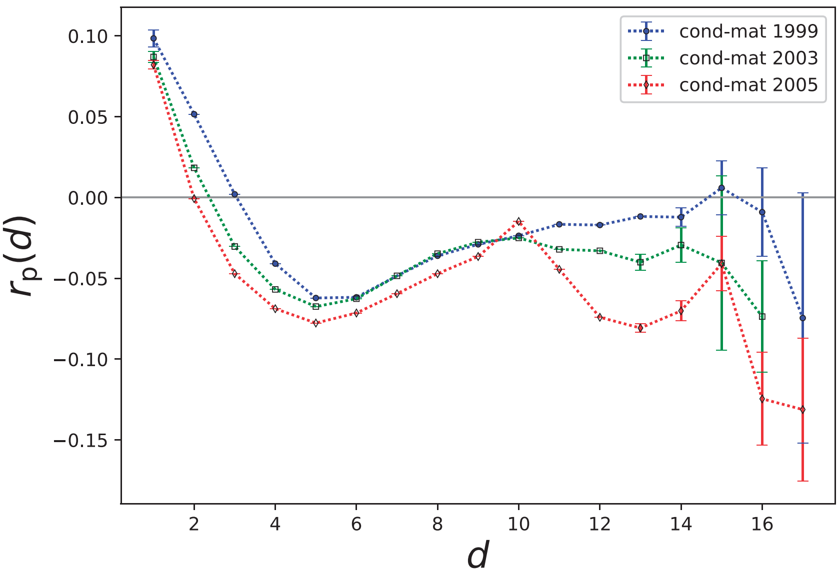

FIG. 3. Correlations in the number of papers p per author in the scientific coauthorship network derived from the condensed matter arXiv. Networks are the result of aggregation from the year 1995 to that indicated.

to difference). There are many more of these in the Slashdot network and each will contribute to disassortativity.

In Fig. 1(c) and (d), the $\psi_+(K)$ and $\psi_-(\Delta)$ distributions are shown for the Facebook and slashdot networks, respectively. The Facebook network is assortatively mixed by degree while the slashdot one is not. There are large fluctuations around $| \quad | \approx 1,500$ and $| \quad | \approx 2,500$ for the slashdot users which drives the disassortativity. However, the distribution for Facebook users decays with fewer fluctuations.

4.1 *Influence*

We turn our attention to correlations at a distance for these 16 social networks. In Table 2, the correlations for pairs of nodes separated up to an edge distance of $d = 3$ are displayed. For $d = 1$, the number of pairs of nodes $n_{\text{pairs}}(d)$, and $r(d)$, simply correspond to the number of edges and the traditional assortativity of the network, respectively. As d increases, the number of pairs increases rapidly. For most networks, $r(d)$ decreases beyond a distance of $d = 1$, and tends to fluctuate about $r(d) = 0$ for values of d near the average path length . None of the networks have positive correlations at $d = 3$, and only two networks have $r(2) > 0$.

In Fig. 2(a), $r(d)$ is plotted as a function of d for two of the larger assortative networks, namely the PGP web of trust and Facebook networks. In each case, the networks are assortative at $d = 1$, with r_k decreasing until $d = 4$, by which point they are disassortative. Beyond $d = 4$, values of r_k plateau below zero. In these networks, users distance $d = 3$ apart are unlikely to have similar degrees. This is not an unexpected result; as shown in Ref. [39], an individual's friends are not representative of a social network, instead they are a biased population. Similarly, that reference shows that high-degree nodes have a disproportionately large effect on networks at a distance of $d = 2$. This effect seems to increase with d, implying that as distances increase, people become more dissimilar.

This method is quite different to the degrees of influence observed by Christakis and Fowler [6, 7]. There, the authors study social networks evolving in time, and focus on correlations between node attributes besides degree, such as obesity and happiness. In contrast, we focus on degree correlations, on networks that have either been aggregated in time or extracted at a snapshot in time. One exception, however, is the condensed matter *arXiv* network, which we examine over multiple time periods; 1995–1999, 1995–2003 and 1995–2005 [25]. Figure 2(b) shows the degree correlations r as a function of distance d for these time periods. They each have positive values of r for $d = 1$, which become negative as d increases. Note that we have adopted the subscript k to distinguish degree correlations from correlations in the number of scientific papers, which we study in the next figure.

Although our focus is on degree correlations, one may study correlations as a function of distance for arbitrary node attributes. The scientific coauthorship network, for instance, details the number of papers p_i that author i has published. Studying correlations in this quantity provides a measure of the output assortativity of the collaboration network. These results are shown in Fig. 3, where we plot output correlation r_p as a function of distance d. These are correlated initially among nearest neighbours, implying productive authors have productive coauthors, and unproductive authors have unproductive coauthors. Small but non-zero correlations exist at $d = 2$ in the first two time-periods but are decorrelated for the final time point. For $d > 2$, there are no positive correlations present. Therefore, on average, two authors three steps away from one another in this coauthorship network are unlikely to have a similar number of publications and this seems to become less similar as time goes on. At higher d values then, authors are more likely to have a different number of publications to their distant neighbours.

5. Simulations

Having studied the behaviour of the assortativity for increasing d in empirical networks, we turn our attention to simulated networks. Here, we examine three types of simulated network; Erdős-Rényi graphs, Erdős-Rényi graphs that we have made assortative by rewiring, and configuration model graphs corresponding to each of the empirical networks in Table 1.

5.1 *Erdős-Rényi graphs*

Here, we generate Erdős-Rényi graphs and calculate $r(d)$ for all values of d such that $1 \leq d \leq {}_{\max}$, where ${}_{\max}$ is the diameter of the graph. This is repeated 1,000 times each for Erdős-Rényi graphs of size $N = 1,000$, with average degree $k\rangle = 5$, 10 and 15. We then repeat this 1,000 times each for graphs of size $N = 10,000$, again with $k\rangle = 5$, 10 and 15. The results of these calculations are shown below in Fig. 4.

As we can see, there are no positive correlations between the degrees of nodes at any value of d. Negative correlations begin to emerge in all cases at or beyond the average path lengths of the graphs, indicated as dashed vertical lines. This finding is in line with the work of Mayo *et al.* [11], who find negative correlations emerge when the average degree is greater than one. While high-degree nodes are exponentially suppressed in Erdős-Rényi graphs, negative correlations still emerge due to the finite size of the simulations. That is, paths of higher values of d still connect central nodes to nodes at the periphery of the graph. It is reasonable to expect that the degrees of such pairs of nodes would be dissimilar, thus reducing the assortativity at these d values.

5.2 *Configuration model networks*

We now examine the assortativity as a function of d for configuration model variants of the networks: Infectious disease through Munmun-Digg in Table 1. For each network, 100 instances of the

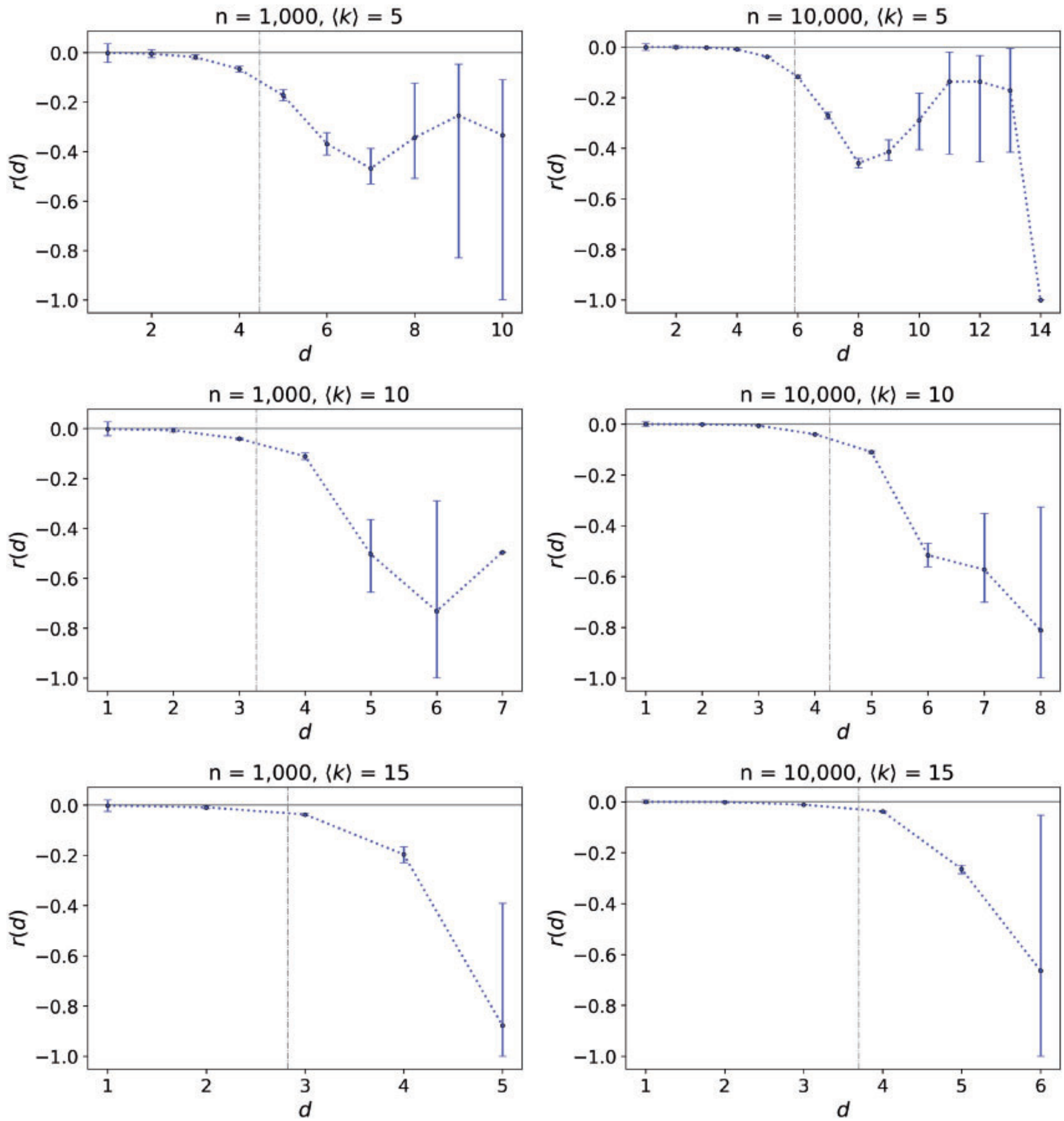

FIG. 4. Results for Erdős-Rényi graphs with 1,000 nodes (left) and 10,000 nodes (right). Points are average values for 1,000 instances of each size and average degree, error bars are 95% coverage bars for the assortativity values observed. Vertical dashed lines represent the mean average path length of each set of 1,000 graphs. As we can see in all cases, the networks became anti-correlated as d increases.

corresponding configuration model network were created and the assortativity calculated for all d values up to the diameter of each graph. Results for the network science coauthorship, PGP web of trust and Munmun-Digg networks can be seen in Fig. 5. These results are representative of the remaining networks not shown.

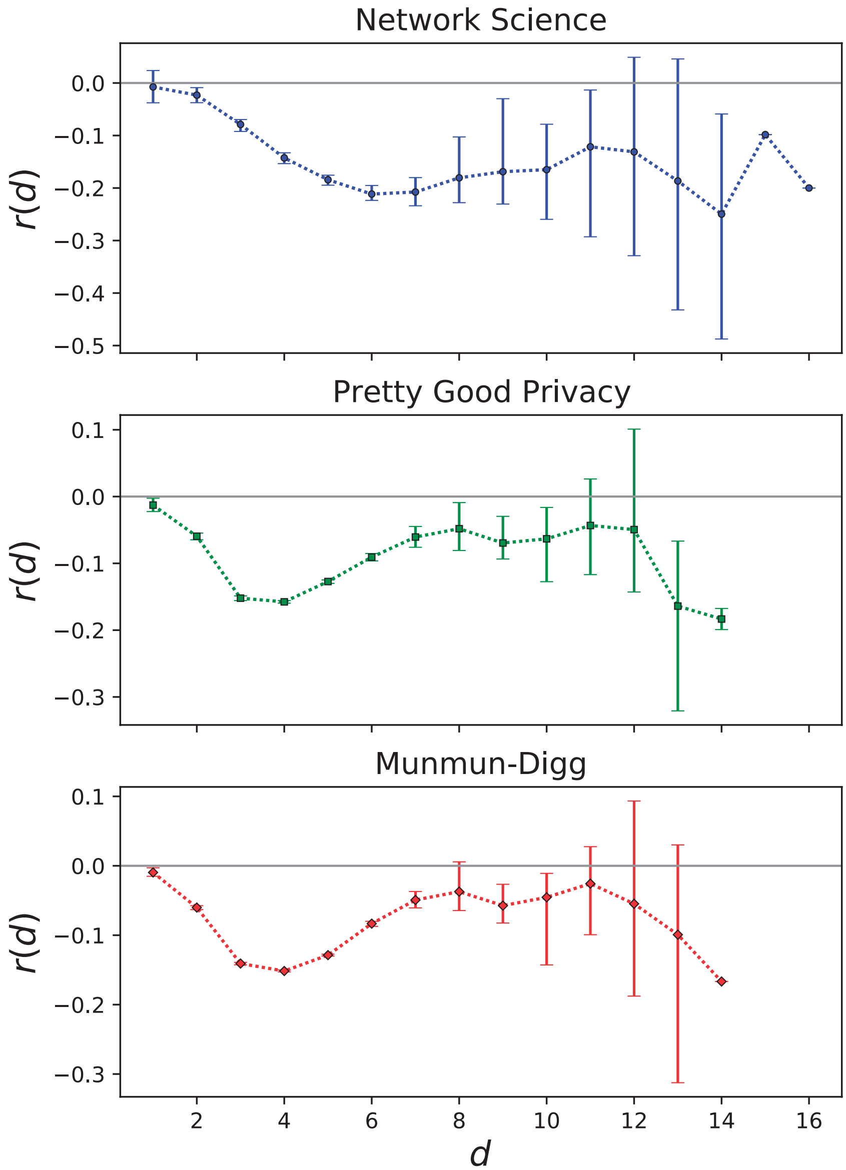

FIG. 5. Degree assortativity r as a function of shortest paths length d, for configuration model variants of the network science coauthorship network, the PGP security network and the Munmun-Digg network. Error bars are 95% coverage bars for the assortativity values.

In all cases, we find that the assortativity rapidly decreases before beginning to increase and approach zero for higher values of d. In the three cases shown here we also see large fluctuations at the largest value of d for each network. These fluctuations are due to the fact that at these large d values, there are very few pairs separated by a shortest path of that length d.

One possible explanation for the decrease in assortativity is that the quantity reaches its minimum when the influence of hubs in the network is at its maximum. For example, high-degree nodes will be connected to many other nodes by shortest paths of lengths 2 and 3, and these nodes will have degrees that are dissimilar to that of the hub node, thus reducing the assortativity. As d increases further, the number of pairs of nodes separated by shortest paths of length d increases. Many of these pairs then will be positively assortative, which counteracts the effect of the hub nodes.

To illustrate this further, we plot the number of pairs at each distance d in Fig. 6. Here, we see after the average path length (dashed vertical line), these is a sharp drop in the number of pairs. Beyond this point, the hubs have probably reached every periphery node, but nodes at different ends of the network still have not been used. As the hub is likely to be at a different side of $E(k)$ to a periphery node in Equation (3.9), their influence will contribute negatively to the assortativity. This difference in degree for neighbours can be seen in Fig. 1, these spikes will only increase as the number of paths increase.

5.3 *Modified Erdős-Rényi graphs*

Here, we start with Erdős-Rényi graphs of size $N = 1,000$ and average degree $\langle k\rangle = 5$, 10 and 15. We run a rewiring algorithm on 1,000 instances of each graph and calculate the assortativity for all values of d up to the diameter of the rewired graph. The algorithm works by randomly removing edges that decrease the assortativity, and replacing them with edges that increase the assortativity, as per Equation (3.9). This process is repeated until the desired value of first-neighbour assortativity is reached, choosing $r = 0.4$ for graphs size $N = 1,000$ and $r = 0.25$ for graphs of size $N = 10,000$. The results for this process are shown in Fig. 7. This approach is then repeated 500 times for Erdős-Rényi graphs of size $N = 10,000$ and average degrees $\langle k\rangle = 5$ and 10, and 250 times for Erdős-Rényi graphs of size $N = 10,000$ and average degree $\langle k\rangle = 15$. The reduced number of iterations here is due to the increased computational cost of rewiring larger graphs. As the number of nodes and edges increases, the contribution that each edge makes to the overall assortativity decreases and so many more iterations of the rewiring algorithm are required to reach the desired assortativity value.

As we can see, all of the graphs started out assortative at $d = 1$ due to the rewiring algorithm, but these positive correlations quickly disappear as d increases, and beyond the average path lengths, the results are very similar to those seen for the Erdős-Rényi graphs in Fig. 4. This is not a great surprise, but serves to highlight that the degree distribution plays a role in the behaviour of the decay of the assortativity as the value of d increases.

6. Conclusions

Social networks are generally found to be assortatively mixed by degree, a property separating them from other complex networks. Here, we extend this to go beyond nearest neighbours in order to test if they remain assortative or if this changes as the distance increases. We find that for all cases, as the edge distance between nodes increases, degree–degree correlations vanish. In some cases, they become anti-correlated before becoming decorrelated. In all 16 social networks, nodes three steps away from each other are more likely to have dissimilar degrees. That is, nodes three steps away from each other are more likely to have degrees on either side of the average value $\mathrm{E}(k)$. The same is true of number of papers in the condensed matter arXiv coauthorship network, and in that dataset, over time, both quantities become less correlated as time increases.

We performed simulations on three different types of network to test this. With Erdős-Rényi random graphs, correlations remain close to zero as expected from mean-field arguments for distances d up to the

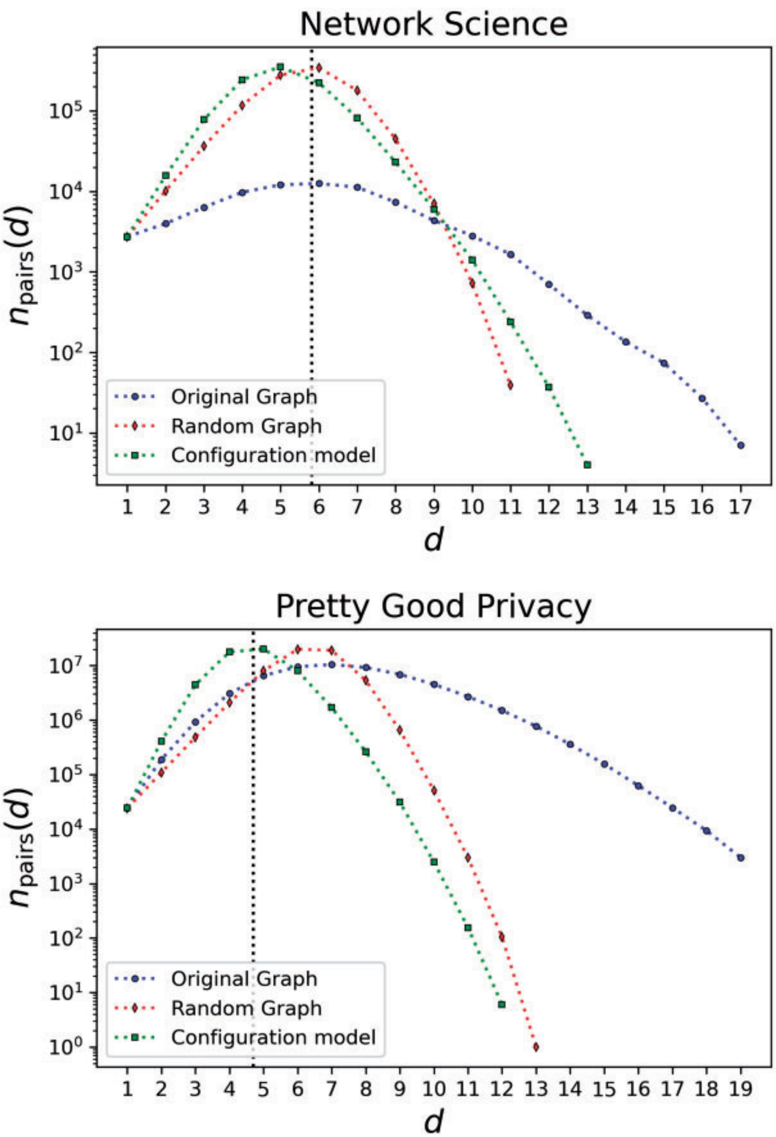

FIG. 6. The number of pairs for each shortest path length d for the Network Science coauthors (top) and PGP web of trust (bottom). The dashed vertical line represents the average path length of the original network.

average path length. Beyond this, while correlations become negative, there are fewer pairs corresponding to these distances, and fluctuations are large as a result. Further, this seems to be a reflection of the finite size of the simulations, with pairs of distant nodes tending to consist of one high-degree node, and one peripheral node at the end of a chain. In contrast, configuration model variants of real-world networks, correlations quickly become negative as a function of d. Finally, in Erdős-Rényi graphs rewired for assortativity, correlations become negative for distance beyond the average path length, as before. This is due to their similarity to Erdős-Rényi random graphs, despite rewiring.

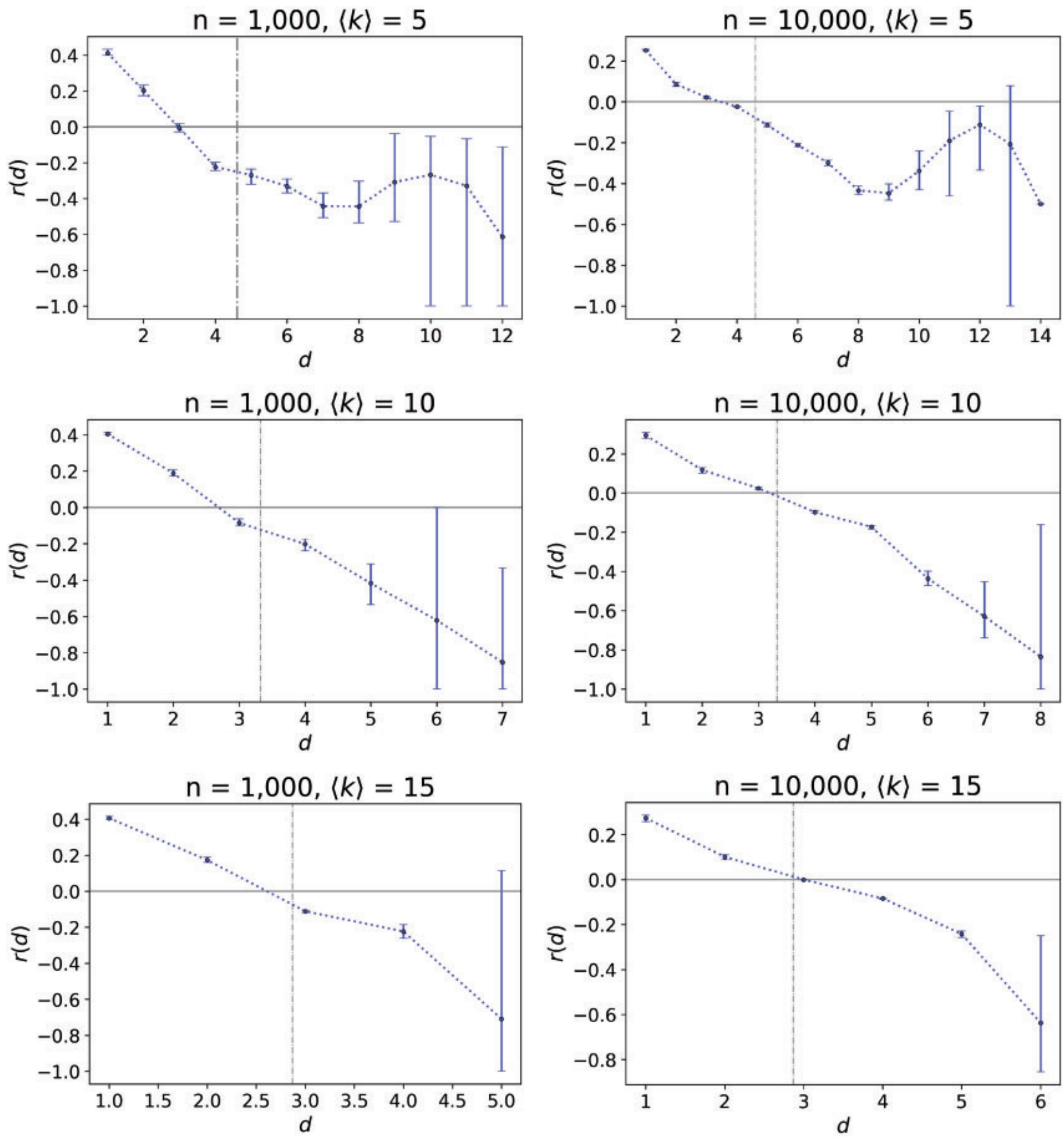

FIG. 7. Correlations between the degrees of nodes separated by shortest paths of length d for Erdős-Rényi graphs of size 1,000 nodes (left, 1,000 instances for each value of k) and 10,000 nodes (right, 500 instances each for $k = 5$ and $k = 10$,250 instances for $k = 15$) that have been rewired to be assortative. Vertical dashed lines represent the mean average path length of each set of 1,000 graphs. The error bars are 95% coverage bars for the assortativity values.

We attribute the decreasing degree correlations to the right-skewed distributions of the empirical datasets and, hence, the influence of hubs in these networks. Hubs are nodes with large degrees, and due to the initial assortativity in social networks, they are in general more likely to connect to nodes with a similar degree. However, they will have quite different degrees to others as the distance increases.

With the calculation for the assortativity, if the degrees of the two nodes of an edge are on either side of the average degree at the end of an edge then this contributes negatively towards the assortativity value. Therefore, as the distance increases, a pair is more likely to contain a well-connected node and one at the periphery. As the number of pairs increases significantly at each step, these high-degree nodes interact more and more with low-degree nodes. However, by a distance around the average path length, a hub may be connected to every node in the graph. Beyond this distance then, the influence of hubs on the assortativity value is diminished.

To investigate this further, we next aim to explore the role of the degree distribution with beyond nearest-neighbour correlations and identify how much influence hubs play on this. We also aim to explore more of the behaviour of disassortative networks, both in empirical data and simulations. We are also interested in the idea of categorical assortativity and whether this will have similar effects going deeper into the network. Finally, the assortativity at a distance here is calculated from the point of view of one shortest path between a pair, this could be weighted by the number of shortest paths, this is another extension we intend to explore.

In Ref. [11], using an average degree approach for three social networks, it was also shown that the assortativity decreases as the distance increases. Here, we show that for more social networks, as well as simulated networks, this is always the case. Similarly on a network at different time points, both assortativity and a further property become negative at distances above 2. These anti-correlations get stronger as time increases. This result is the opposite of what the ‘three degrees of influence’ work would suggest. Our work instead implies, that in a social network, the more distant you are from someone, the more likely you are different to them. This is related to the idea of the ‘friendship paradox’—that your friends, on average, have more friends than you—and is also likely because social networks are right-skewed.

This again implies that our observations of the correlations between nodes separated by distances greater than one here are tied to the degree distributions of the graphs, and the future work outlined above will help to explain this connection.

Acknowledgements

The authors would like to thank Ralph Kenna for ideas on assortativity and conceptual discussions, as well as Samuel Unicomb for working through the mathematics and making clarifications.

Funding

S.M. is funded by the Science Foundation Ireland (Grant number 18/CRT/6049).

References

1. Newman, M. E. & Park, J. (2003) Why social networks are different from other types of networks. *Phys. Rev. E*, **68**, 036122.
2. Hu, H.-B. & Wang, X.-F. (2009) Disassortative mixing in online social networks. *Europhys. Lett.*, **86**, 18003.
3. Wilson, K. G. (1979) Problems in physics with many scales of length. *Sci. Am.*, **241**, 158–179.
4. Newman, M. E. (2002) Assortative mixing in networks. *Phys. Rev. Lett.*, **89**, 208701.
5. Mcpherson, M., Smith-Lovin, L. & Cook, J. M. (2001) Birds of a feather: Homophily in social networks. *Annu. Rev. Sociol.*, 415–444.
6. Christakis, N. A. & Fowler, J. H. (2007) The spread of obesity in a large social network over 32 years. *N. Eng. J. Med.*, **357**, 370–379.

7. Christakis, N. A. & Fowler, J. H. (2009) *Connected: The Surprising Power of Our Social Networks and How They Shape Our Lives*. Boston, Massachusetts, USA: Little, Brown Spark.
8. Lyons, R. (2011) The spread of evidence-poor medicine via flawed social-network analysis. *Stat. Polit. Policy*, **2**, Article 2. http://www.bepress.com/spp/vol2/iss1/2.
9. Mizutaka, S. & Hasegawa, T. (2020) Emergence of long-range correlations in random networks. *J. Phys. Complex.*, **1**, 035007.
10. Fujiki, Y., Takaguchi, T. & Yakubo, K. (2018) General formulation of long-range degree correlations in complex networks. *Phys. Rev. E*, **97**, 062308.
11. Mayo, M., Abdelzaher, A. & Ghosh, P. (2015) Long-range degree correlations in complex networks. *Comput. Soc. Netw.*, **2**, 1–13.
12. Rybski, D., Rozenfeld, H. D. & Kropp, J. P. (2010) Quantifying long-range correlations in complex networks beyond nearest neighbors. *Europhys. Lett.*, **90**, 28002.
13. Arcagni, A., Grassi, R., Stefani, S. & Torriero, A. (2017) Higher order assortativity in complex networks. *Eur. J. Oper. Res.*, **262**, 708–719.
14. Allen-Perkins, A., Pastor, J. M. & Estrada, E. (2017) Two-walks degree assortativity in graphs and networks. *Appl. Math. Comput.*, **311**, 262–271.
15. Efron, B. (1979) Computers and the theory of statistics: Thinking the unthinkable. *SIAM Rev.*, **21**, 460–480.
16. Efron, B. & Tibshirani, R. (1986) Bootstrap methods for standard errors, confidence intervals, and other measures of statistical accuracy. *Stat. Sci.*, pp. 54–75.
17. Hunter, D. R., Goodreau, S. M. & Handcock, M. S. (2008) Goodness of fit of social network models. *J. Am. Stat. Assoc.*, **103**, 248–258.
18. Resnick, M. D., Bearman, P. S., Blum, R. W., Bauman, K. E., Harris, K. M., Jones, J., Tabor, J., Beuhring, T., Sieving, R. E., Shew, M. & Others (1997) Protecting adolescents from harm: Findings from the National Longitudinal Study on Adolescent Health. *JAMA*, **278**, 823–832.
19. Gleiser, P. M. & Danon, L. (2003) Community structure in jazz. *Adv. Complex Syst.*, **6**, 565–573.
20. Isella, L., Stehlé, J., Barrat, A., Cattuto, C., Pinton, J.-F. & Van Den Broeck, W. (2011) What's in a crowd? Analysis of face-to-face behavioral networks. *J. Theor. Biol.*, **271**, 166–180. https://www.sciencedirect.com/science/article/pii/S0022519310006284.
21. Newman, M. E. (2006) Finding community structure in networks using the eigenvectors of matrices. *Phys. Rev. E*, **74**, 036104.
22. Moody, J. (2001) Peer influence groups: Identifying dense clusters in large networks. *Soc. Netw.*, **23**, 261–283.
23. Kunegis, J. (2013a) Konect. http://konect.cc/networks/petster-hamster/. Accessed 1 November 2022.
24. Boguñá, M., Pastor-Satorras, R., Díaz-Guilera, A. & Arenas, A. (2004) Models of social networks based on social distance attachment. *Phys. Rev. E*, **70**, 056122.
25. Newman, M. E. (2001) The structure of scientific collaboration networks. *Proc. Natl Acad. Sci. USA.*, **98**, 404–409.
26. Leskovec, J., Kleinberg, J. & Faloutsos, C. (2007) Graph evolution: Densification and shrinking diameters. *ACM Trans. Knowl. Discov. Data*, **1**, 2–es. doi: https://doi.org/10.1145.
27. Leskovec, J. & Mcauley, J. (2012) Learning to discover social circles in ego networks. *Adv. Neural Inf. Process. Syst.*, **25**.
28. De Choudhury, M., Sundaram, H., John, A. & Seligmann, D. D. (2009) Social synchrony: Predicting mimicry of user actions in online social media. *2009 International Conference on Computational Science and Engineering*, vol. 4. Vancouver, BC, Canada: IEEE, pp. 151–158.
29. Kunegis, J. (2013b) Konect. http://konect.cc/networks/munmun_digg_reply/. Accessed 1 November 2022.
30. Viswanath, B., Mislove, A., Cha, M. & Gummadi, K. P. (2009a) On the evolution of user interaction in facebook. *Proceedings of the 2nd ACM Workshop on Online Social Networks*. Barcelona, Spain, pp. 37–42.
31. Viswanath, B., Mislove, A., Cha, M. & Gummadi, K. P. (2009b) On the evolution of user interaction in facebook. *Proceedings of the 2nd ACM Workshop on Online Social Networks*. Barcelona, Spain, pp. 37–42.

32. Kunegis, J., Lommatzsch, A. & Bauckhage, C. (2009) The slashdot zoo: Mining a social network with negative edges. *Proceedings of the 18th International Conference on World Wide Web*. Madrid, Spain, pp. 741–750.

33. Klimt, B. & Yang, Y. (2004) The enron corpus: A new dataset for email classification research. *European Conference on Machine Learning*. Springer, pp. 217–226.

34. Leskovec, J., Lang, K. J., Dasgupta, A. & Mahoney, M. W. (2009) Community structure in large networks: Natural cluster sizes and the absence of large well-defined clusters. *Internet Math.*, **6**, 29–123.

35. Kunegis, J. (2013c) Konect: The koblenz network collection. *Proceedings of the 22nd International Conference on World Wide Web*. pp. 1343–1350. http://konect.cc/. Accessed 1 November 2022.

36. Gleiser, P. M. (2007) How to become a superhero. *J. Stat. Mech. Theory Exp.*, **2007**, P09020.

37. Mac Carron, P. & Kenna, R. (2012) Universal properties of mythological networks. *Europhys. Lett.)*, **99**, 28002.

38. Szell, M. & Thurner, S. (2010) Measuring social dynamics in a massive multiplayer online game. *Soc. Netw.*, **32**, 313–329.

39. Newman, M. E. (2003) Ego-centered networks and the ripple effect. *Soc. Netw.*, **25**, 83–95.