

ARTICLE

Received 19 May 2015 | Accepted 13 Sep 2015 | Published 20 Oct 2015

DOI: 10.1038/ncomms9627

OPEN

# Quantifying randomness in real networks

Chiara Orsini<sup>1,2</sup>, Marija M. Dankulov<sup>3,4</sup>, Pol Colomer-de-Simón<sup>5</sup>, Almerima Jamakovic<sup>6</sup>, Priya Mahadevan<sup>7</sup>, Amin Vahdat<sup>8</sup>, Kevin E. Bassler<sup>9,10</sup>, Zoltán Toroczkai<sup>11</sup>, Marián Boguñá<sup>5</sup>, Guido Caldarelli<sup>12</sup>, Santo Fortunato<sup>13</sup> & Dmitri Krioukov<sup>1,14</sup>

Represented as graphs, real networks are intricate combinations of order and disorder. Fixing some of the structural properties of network models to their values observed in real networks, many other properties appear as statistical consequences of these fixed observables, plus randomness in other respects. Here we employ the *dk*-series, a complete set of basic characteristics of the network structure, to study the statistical dependencies between different network properties. We consider six real networks—the Internet, US airport network, human protein interactions, technosocial web of trust, English word network, and an fMRI map of the human brain—and find that many important local and global structural properties of these networks are closely reproduced by *dk*-random graphs whose degree distributions, degree correlations and clustering are as in the corresponding real network. We discuss important conceptual, methodological, and practical implications of this evaluation of network randomness, and release software to generate *dk*-random graphs.

<sup>1</sup>CAIDA, University of California San Diego, San Diego, California 92093, USA. <sup>2</sup>Information Engineering Department, University of Pisa, Pisa 56122, Italy. <sup>3</sup>Scientific Computing Laboratory, Institute of Physics Belgrade, University of Belgrade, Belgrade 11080, Serbia. <sup>4</sup>Department of Biomedical Engineering and Computational Science, Aalto University School of Science, Helsinki 00076, Finland. <sup>5</sup>Departament de Física Fonamental, Universitat de Barcelona, Barcelona 08028, Spain. <sup>6</sup>Communication and Distributed Systems group, Institute of Computer Science and Applied Mathematics, University of Bern, Bern 3012, Switzerland. <sup>7</sup>Palo Alto Research Center, Palo Alto, California 94304, USA. <sup>8</sup>Google, Mountain View, California 94043, USA. <sup>9</sup>Department of Physics and Texas Center for Superconductivity, University of Houston, Houston, Texas 77204, USA. <sup>10</sup>Max Planck Institut für Physik komplexer Systeme, Dresden 01187, Germany. <sup>11</sup>Department of Physics and Interdisciplinary Center for Network Science and Applications, University of Notre Dame, Notre Dame, IN 46556, USA. <sup>12</sup>IMT Alti Studi, Lucca 55100, Italy. <sup>13</sup>Department of Computer Science, Aalto University School of Science, Helsinki 00076, Finland. <sup>14</sup>Department of Physics, Department of Mathematics, Department of Electrical and Computer Engineering, Northeastern University, Boston, Massachusetts 02115, USA. Correspondence and requests for materials should be addressed to C.O. (email: chiara@caida.org) or to D.K. (email: dima@neu.edu).

Network science studies complex systems by representing them as networks<sup>1</sup>. This approach has proven quite fruitful because in many cases the network representation achieves a practically useful balance between simplicity and realism: while always grand simplifications of real systems, networks often encode some crucial information about the system. Represented as a network, the system structure is fully specified by the network adjacency matrix, or the list of connections, perhaps enriched with some additional attributes. This (possibly weighted) matrix is then a starting point of research in network science.

One significant line of this research studies various (statistical) properties of adjacency matrices of real networks. The focus is often on properties that convey useful information about the global network structure that affects the dynamical processes in the system that this network represents<sup>2</sup>. A common belief is that a self-organizing system should evolve to a network structure that makes these dynamical processes, or network functions, efficient<sup>3–5</sup>. If this is the case, then given a real network, we may ‘reverse engineer’ it by showing that its structure optimizes its function. In that respect the problem of interdependency between different network properties becomes particularly important<sup>6–10</sup>.

Indeed, suppose that the structure of some real network has property  $X$ —some statistically over- or under-represented subgraph, or motif<sup>11</sup>, for example—that we believe is related to a particular network function. Suppose also that the same network has in addition property  $Y$ —some specific degree distribution or clustering, for example—and that all networks that have property  $Y$  necessarily have property  $X$  as a consequence. Property  $Y$  thus enforces or ‘explains’ property  $X$ , and attempts to ‘explain’  $X$  by itself, ignoring  $Y$ , are misguided. For example, if a network has high density (property  $Y$ ), such as the interarial cortical network in the primate brain where 66% of edges that could exist do exist<sup>12</sup>, then it will necessarily have short path lengths and high clustering, meaning it is a small-world network (property  $X$ ). However, unlike social networks where the small-world property is an independent feature of the network, in the brain this property is a simple consequence of high density.

The problem of interdependencies among network properties has been long understood<sup>13,14</sup>. The standard way to address it, is to generate many graphs that have property  $Y$  and that are random in all other respects—let us call them  $Y$ -random graphs—and then to check if property  $X$  is a typical property of these  $Y$ -random graphs. In other words, this procedure checks if graphs that are sampled uniformly at random from the set of all graphs that have property  $Y$ , also have property  $X$  with high probability. For example, if graphs are sampled from the set of graphs with high enough edge density, then all sampled graphs will be small worlds. If this is the case, then  $X$  is not an interesting property of the considered network, because the fact that the network has property  $X$  is a statistical consequence of that it also has property  $Y$ . In this case we should attempt to explain  $Y$  rather than  $X$ . In case  $X$  is not a typical property of  $Y$ -random graphs, one cannot really conclude that property  $X$  is interesting or important (for some network functions). The only conclusion one can make is that  $Y$  cannot explain  $X$ , which does not mean however that there is no other property  $Z$  from which  $X$  follows.

In view of this inherent and unavoidable relativism with respect to a null model, the problem of structure–function relationship requires an answer to the following question in the first place: what is the right base property or properties  $Y$  in the null model ( $Y$ -random graphs) that we should choose to study the (statistical) significance of a given property  $X$  in a given network<sup>15</sup>? For most properties  $X$  including motifs<sup>11</sup>, the choice of  $Y$  is often just the degree distribution. That is, one usually checks if  $X$  is present in random graphs with the same degree

distribution as in the real network. Given that scale-free degree distributions are indeed the striking and important features of many real networks<sup>1</sup>, this null model choice seems natural, but there are no rigorous and successful attempts to justify it. The reason is simple: the choice cannot be rigorously justified because there is nothing special about the degree distribution—it is one of infinitely many ways to specify a null model.

Since there exists no unique preferred null model, we have to consider a series of null models satisfying certain requirements. Here we consider a particular realization of such series—the  $dk$ -series<sup>16</sup>, which provides a complete systematic basis for network structure analysis, bearing some conceptual similarities with a Fourier or Taylor series in mathematical analysis. The  $dk$ -series is a converging series of basic interdependent degree- and subgraph-based properties that characterize the local network structure at an increasing level of detail, and define a corresponding series of null models or random graph ensembles. These random graphs have the same distribution of differently sized subgraphs as in a given real network. Importantly, the nodes in these subgraphs are labelled by node degrees in the real network. Therefore, this random graph series is a natural generalization of random graphs with fixed average degree, degree distribution, degree correlations, clustering and so on. Using  $dk$ -series we analyse six real networks, and find that they are essentially random as soon as we constrain their degree distributions, correlations, and clustering to the values observed in the real network ( $Y = \text{degrees} + \text{correlations} + \text{clustering}$ ). In other words, these basic local structural characteristics almost fully define not only local but also global organization of the considered networks. These findings have important implications on research dealing with network structure–function interplay in different disciplines where networks are used to represent complex natural or designed systems. We also find that some properties of some networks cannot be explained by just degrees, correlations, and clustering. The  $dk$ -series methodology thus allows one to detect which particular property in which particular network is non-trivial, cannot be reduced to basic local degree- or subgraph-based characteristics, and may thus be potentially related to some network function.

## Results

**General requirements to a systematic series of properties.** The introductory remarks above instruct one to look not for a single base property  $Y$ , which cannot be unique or universal, but for a systematic series of base properties  $Y_0, Y_1, \dots$ . By ‘systematic’ we mean the following conditions: (1) inclusiveness, that is, the properties in the series should provide strictly more detailed information about the network structure, which is equivalent to requiring that networks that have property  $Y_d$  ( $Y_d$ -random graphs),  $d > 0$ , should also have properties  $Y_{d'}$  for all  $d' = 0, 1, \dots, d - 1$ ; and (2) convergence, that is, there should exist property  $Y_D$  in the series that fully characterizes the adjacency matrix of any given network, which is equivalent to requiring that  $Y_D$ -random graphs is only one graph—the given network itself. If these  $Y$ -series satisfy the conditions above, then whatever property  $X$  is deemed important now or later in whatever real network, we can always standardize the problem of explanation of  $X$  by reformulating it as the following question: what is the minimal value of  $d$  in the above  $Y$ -series such that property  $Y_d$  explains  $X$ ? By convergence, such  $d$  should exist; and by inclusiveness, networks that have property  $Y_{d'}$  with any  $d' = d, d + 1, \dots, D$ , also have property  $X$ . Assuming that properties  $Y_d$  are once explained, the described procedure provides an explanation of any other property of interest  $X$ .

The general philosophy outlined above is applicable to undirected and directed networks, and it is shared by different

approaches, including motifs<sup>11</sup>, graphlets<sup>17</sup> and similar constructions<sup>18</sup>, albeit they violate the inclusiveness condition as we show below. Yet one can still define many different  $Y$ -series satisfying both conditions above. Some further criteria are needed to focus on a particular one. One approach is to use degree-based tailored random graphs as null models for both undirected<sup>19–21</sup> and directed<sup>22,23</sup> networks. The criteria that we use to select a particular  $Y$ -series in this study are simplicity and the importance of subgraph- and degree-based statistics in networks. Indeed, in the network representation of a system, subgraphs, their frequency and convergence are the most natural and basic building blocks of the system, among other things forming the basis of the rigorous theory of graph family limits known as graphons<sup>24</sup>, while the degree is the most natural and basic property of individual nodes in the network. Combining the subgraph- and degree-based characteristics leads to  $dk$ -series<sup>16</sup>.

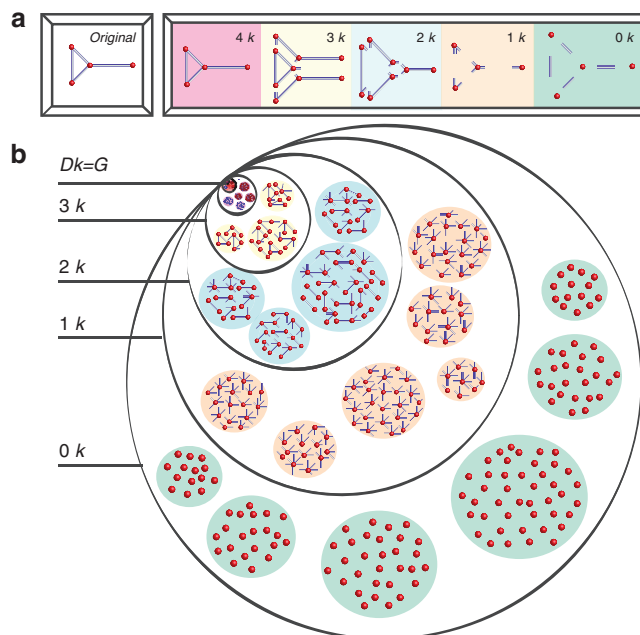
**$dk$ -series.** In  $dk$ -series, properties  $Y_d$  are  $dk$ -distributions. For any given network  $G$  of size  $N$ , its  $dk$ -distribution is defined as a collection of distributions of  $G$ 's subgraphs of size  $d = 0, 1, \dots, N$  in which nodes are labelled by their degrees in  $G$ . That is, two isomorphic subgraphs of  $G$  involving nodes of different degrees—for instance, edges ( $d = 2$ ) connecting nodes of degrees 1, 2 and 2, 2—are counted separately. The  $0k$ -distribution is defined as the average degree of  $G$ . Figure 1 illustrates the  $dk$ -distributions of a graph of size 4.

Thus defined the  $dk$ -series subsumes all the basic degree-based characteristics of networks of increasing detail. The zeroth element in the series, the  $0k$ -distribution, is the coarsest characteristic, the average degree. The next element, the  $1k$ -distribution, is the standard degree distribution, which is the number of nodes—subgraphs of size 1—of degree  $k$  in the network. The second element, the  $2k$ -distribution, is the joint degree distribution, the number of subgraphs of size 2—edges—between nodes of degrees  $k_1$  and  $k_2$ . The  $2k$ -distribution thus defines 2-node degree correlations and network's assortativity. For  $d = 3$ , the two non-isomorphic subgraphs are triangles and wedges, composed of nodes of degrees  $k_1, k_2$  and  $k_3$ , which defines clustering, and so on. For arbitrary  $d$  the  $dk$ -distribution characterizes the  $d$ -degree  $k$ -relations in  $d$ -sized subgraphs, thus including, on the one hand, the correlations of degrees of nodes located at hop distances below  $d$ , and, on the other hand, the statistics of  $d$ -cliques in  $G$ . We will also consider  $dk$ -distributions with fractional  $d \in (2, 3)$  which in addition to specifying two-node degree correlations ( $d = 2$ ), fix some  $d = 3$  substatistics related to clustering.

The  $dk$ -series is inclusive because the  $(d + 1)k$ -distribution contains the same information about the network as the  $dk$ -distribution, plus some additional information. In the simplest  $d = 0$  case for example, the degree distribution  $P(k)$  ( $1k$ -distribution) defines the average  $\bar{k}$  ( $0k$ -distribution) via  $\bar{k} = \sum_k kP(k)$ . The analogous expression for  $d = 1, 2$  are derived in Supplementary Note 1.

It is important to note that if we omit the degree information, and just count the number of  $d$ -sized subgraphs in a given network regardless their node degrees, as in motifs<sup>11</sup>, graphlets<sup>17</sup> or similar constructions<sup>18</sup>, then such degree- $k$ -agnostic  $d$ -series (versus  $dk$ -series) would not be inclusive (Supplementary Discussion). Therefore, preserving the node degree ( $k$ ) information is necessary to make a subgraph-based ( $d$ ) series inclusive. The  $dk$ -series is clearly convergent because at  $d = N$ , where  $N$  is the network size, the  $Nk$ -distribution fully specifies the network adjacency matrix.

A sequence of  $dk$ -distributions then defines a sequence of random graph ensembles (null models). The  $dk$ -graphs are a set



**Figure 1 | The  $dk$ -series illustrated.** (a) shows the  $dk$ -distributions for a graph of size 4. The  $4k$ -distribution is the graph itself. The  $3k$ -distribution consists of its three subgraphs of size 3: one triangle connecting nodes of degrees 2, 2 and 3, and two wedges connecting nodes of degrees 2, 3 and 1. The  $2k$ -distribution is the joint degree distribution in the graph. It specifies the number of links (subgraphs of size 2) connecting nodes of different degrees: one link connects nodes of degrees 2 and 2, two links connect nodes of degrees 2 and 3, and one link connects nodes of degree 3 and 1. The  $1k$ -distribution is the degree distribution in the graph. It lists the number of nodes (subgraphs of size 1) of different degree: one node of degree 1, two nodes of degree 2, and one node of degree 3. The  $0k$ -distribution is just the average degree in the graph, which is 2. (b) illustrates the inclusiveness and convergence of  $dk$ -series by showing the hierarchy of  $dk$ -graphs, which are graphs that have the same  $dk$ -distribution as a given graph  $G$  of size  $D$ . The black circles schematically show the sets of  $dk$ -graphs. The set of  $0k$ -graphs, that is, graphs that have the same average degree as  $G$ , is largest. Graphs in this set may have a structure drastically different from  $G$ 's. The set of  $1k$ -graphs is a subset of  $0k$ -graphs, because each graph with the same degree distribution as in  $G$  has also the same average degree as  $G$ , but not vice versa. As a consequence, typical  $1k$ -graphs, that is,  $1k$ -random graphs, are more similar to  $G$  than  $0k$ -graphs. The set of  $2k$ -graphs is a subset of  $1k$ -graphs, also containing  $G$ . As  $d$  increases, the circles become smaller because the number of different  $dk$ -graphs decreases. Since all the  $dk$ -graph sets contain  $G$ , the circles 'zoom-in' on it, and while their number decreases,  $dk$ -graphs become increasingly more similar to  $G$ . In the  $d = D$  limit, the set of  $Dk$ -graphs consists of only one element,  $G$  itself.

of all graphs with a given  $dk$ -distribution, for example, with the  $dk$ -distribution in a given real network. The  $dk$ -random graphs are a maximum-entropy ensemble of these graphs<sup>16</sup>. This ensemble consists of all  $dk$ -graphs, and the probability measure is uniform (unbiased): each graph  $G$  in the ensemble is assigned the same probability  $P(G) = 1/\mathcal{N}_d$ , where  $\mathcal{N}_d$  is the number of  $dk$ -graphs. For  $d = 0, 1, 2$  these are well studied classical random graphs  $\mathcal{G}_{N,M}$  (ref. 25), configuration model<sup>26–28</sup> and random graphs with a given joint degree distribution<sup>29</sup>, respectively. Since a sequence of  $dk$ -distributions is increasingly more informative and thus constraining, the corresponding sequence of the sizes of  $dk$ -random graph ensembles is non-increasing and shrinking to 1,  $\mathcal{N}_0 \geq \mathcal{N}_1 \geq \dots \geq \mathcal{N}_N = 1$ , Fig. 1. At low  $d = 0, 1, 2$  these numbers  $\mathcal{N}$  can be calculated either exactly or approximately<sup>30,31</sup>.

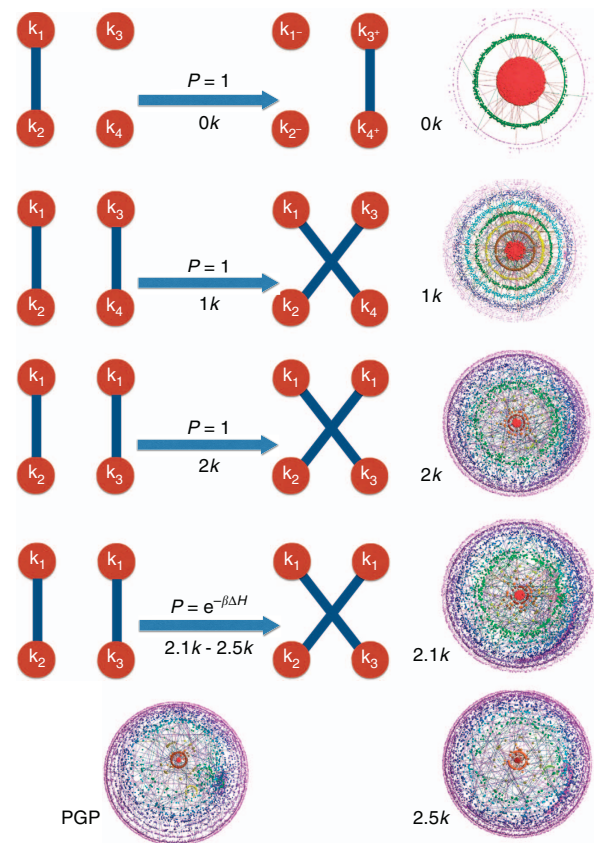
We emphasize that in  $dk$ -graphs the  $dk$ -distribution constraints are sharp, that is, they hold exactly—all  $dk$ -graphs have exactly the same  $dk$ -distribution. An alternative description uses soft maximum-entropy ensembles belonging to the general class of exponential random graph models<sup>32–35</sup> in which these constraints hold only on average over the ensemble—the expected  $dk$ -distribution in the ensemble (not in any individual graph) is fixed to a given distribution. This ensemble consists of all possible graphs  $G$  of size  $N$ , and the probability measure  $P(G)$  is the one maximizing the ensemble entropy  $S = -\sum_G P(G) \ln P(G)$  under the  $dk$ -distribution constraints. Using analogy with statistical mechanics, sharp and soft ensemble are often called microcanonical and canonical, respectively.

As a consequence of the convergence and inclusiveness properties of  $dk$ -series, any network property  $X$  of any given network  $G$  is guaranteed to be reproduced with any desired accuracy by high enough  $d$ . At  $d=N$  all possible properties are reproduced exactly, but the  $Nk$ -graph ensemble trivially consists of only one graph,  $G_{\text{self}}$ , and has zero entropy. In the sense that the entropy of  $dk$ -ensembles  $S_d = \ln \mathcal{N}_d$  is a non-increasing function of  $d$ , the smaller the  $d$ , the more random the  $dk$ -random graphs, which also agrees with the intuition that  $dk$ -random graphs are ‘the less random and the more structured’, the higher the  $d$ . Therefore, the general problem of explaining a given property  $X$  reduces to the general problem of how random a graph ensemble must be so that  $X$  is statistically significant. In the  $dk$ -series context, this question becomes: how much local degree information, that is, information about concentrations of degree-labelled subgraphs of what minimal size  $d$ , is needed to reproduce a possibly global property  $X$  with a desired accuracy?

Below we answer this question for a set of popular and commonly used structural properties of some paradigmatic real networks. But to answer this question for any property in any network, we have to be able to sample graphs uniformly at random from the sets of  $dk$ -graphs—the problem that we discuss next.

**$dk$ -random graph sampling.** Soft  $dk$ -ensembles tend to be more amenable for analytic treatment, compared with sharp ensembles, but even in soft ensembles the exact analytic expressions for expected values are known only for simplest network properties in simplest ensembles<sup>36,37</sup>. Therefore, we retreat to numeric experiments here. Given a real network  $G$ , there exist two ways to sample  $dk$ -random graphs in such experiments:  $dk$ -randomize  $G$  generalizing the randomization algorithms in refs 38,39, or construct random graphs with  $G$ 's  $dk$ -sequence from scratch<sup>16,40</sup>, also called direct construction<sup>41–44</sup>.

The first option,  $dk$ -randomization, is easier. It accounts for swapping random (pairs of) edges, starting from  $G$ , such that the  $dk$ -distribution is preserved at each swap, Fig. 2. There are many concerns with this prescription<sup>45</sup>, two of which are particularly important. The first concern is if this process ‘ergodic’, meaning that if any two  $dk$ -graphs are connected by a chain of  $dk$ -swaps. For  $d=1$  the two-edge swap is ergodic<sup>38,39</sup>, while for  $d=2$  it is not ergodic. However, the so-called restricted two-edge swap, when at least one node attached to each edge has the same degree, Fig. 2, was proven to be ergodic<sup>46</sup>. It is now commonly believed that there is no edge-swapping operation, of this or other type, that is ergodic for the  $3k$ -distribution, although a definite proof is lacking at the moment. If there exists no ergodic  $3k$ -swapping, then we cannot really rely on it in sampling  $dk$ -random graphs because our real network  $G$  can be trapped on a small island of atypical  $dk$ -graphs, which is not connected by any  $dk$ -swap chain to the main land of many typical  $dk$ -graphs. Yet we note that in



**Figure 2 | The  $dk$ -sampling and convergence of  $dk$ -series illustrated.** The left column shows the elementary swaps of  $dk$ -randomizing (for  $d=0, 1, 2$ ) and  $dk$ -targeting (for  $d=2.1, 2.5$ ) rewiring. The nodes are labelled by their degrees, and the arrows are labelled by the rewiring acceptance probability. In  $dk$ -randomizing rewiring, random (pairs of) edges are rewired preserving the graph's  $dk$ -distribution (and consequently its  $d'k$ -distributions for all  $d' < d$ ). In  $2.1k$ - and  $2.5k$ -targeting rewiring, the moves preserve the  $2k$ -distribution, but each move is accepted with probability  $p$  designed to drive the graph closer to a target value of average clustering  $\bar{c}$  ( $2.1k$ ) or degree-dependent clustering  $\bar{c}(k)$  ( $2.5k$ ):  $p = \min(1, e^{-\beta\Delta H})$ , where  $\beta$  the inverse temperature of this simulated annealing process,  $\Delta H = H_a - H_b$ , and  $H_{a,b}$  are the distances, after and before the move, between the current and target values of clustering:  $H_{2.1k} = |\bar{c}_{\text{current}} - \bar{c}_{\text{target}}|$  and  $H_{2.5k} = \sum_i |\bar{c}_{\text{current}}[k_i] - \bar{c}_{\text{target}}[k_i]|$ . The right column shows LaNet-vi (ref. 65) visualizations of the results of these  $dk$ -rewiring processes (Supplementary Methods), applied to the PGP network, visualized at the bottom of the left column. The node sizes are proportional to the logarithm of their degrees, while the colour reflects node cohesiveness<sup>65</sup>. As  $d$  grows, the shown  $dk$ -random graphs quickly become more similar to the real PGP network.

an unpublished work<sup>47</sup> we showed that five out of six considered real networks were virtually indistinguishable from their  $3k$ -randomizations across all the considered network properties, although one network (power grid) was very different from its  $3k$ -random counterparts.

The second concern with  $dk$ -randomization is about how close to uniform sampling the  $dk$ -swap Markov chain is after its mixing time is reached—its mixing time is yet another concern that we do not discuss here, but according to many numerical experiments and some analytic estimates, it is  $O(M)$ <sup>16,29,38–40,46</sup>. Even for  $d=1$  the swap chain does not sample  $1k$ -graphs uniformly at random, yet if the edge-swap process is done correctly, then the sampling is uniform<sup>20,21</sup>.



A simple algorithm for the second  $dk$ -sampling option, constructing  $dk$ -graphs from scratch, is widely known for  $d=1$ : given  $G$ 's degree sequence  $\{k_i\}$ , build a  $1k$ -random graph by attaching  $k_i$  half-edges ('stubs') to node  $i$ , and then connect random pairs of stubs to form edges<sup>27</sup>. If during this process a self-loop (both stubs are connected to the same node) or double-edge (two edges between the same pair of nodes) is formed, one has to restart the process from scratch since otherwise the graph sampling is not uniform<sup>48</sup>. If the degree sequence is power-law distributed with exponent close to  $-2$  as in many real networks, then the probability that the process must be restarted approaches 1 for large graphs<sup>49</sup>, so that this construction process never succeeds. An alternative greedy algorithm is described in ref. 42, which always quickly succeeds and gives an efficient way of testing whether a given sequence of integers is graphical, that is, whether it can be realized as a degree sequence of a graph. The base sampling procedure does not sample graphs uniformly, but then an importance sampling procedure is used to account for the bias, which results in uniform sampling. Yet again, if the degree distribution is a power law, then one can show that even without importance sampling, the base sampling procedure is uniform, since the distribution of sampling weights that one can compute for this greedy algorithm approaches a delta function. Extensions of the naive  $1k$ -construction above to  $2k$  are less known, but they exist<sup>16,29,44,50</sup>. Most of these  $2k$ -constructions do not sample  $2k$ -graphs exactly uniformly either<sup>46</sup>, but importance sampling<sup>20,44</sup> can correct for the sampling biases.

Unfortunately, to the best of our knowledge, there currently exists no  $3k$ -construction algorithm that can be successfully used in practice to generate large  $3k$ -graphs with  $3k$ -distributions of real networks. The  $3k$ -distribution is quite constraining and non-local, so that the  $dk$ -construction methods described above for  $d=1, 2$  cannot be readily extended to  $d=3$  (ref. 16). There is yet another option— $3k$ -targeting rewiring, Fig. 2. It is  $2k$ -preserving rewiring in which each  $2k$ -swap is accepted not with probability 1, but with probability equal to  $\min(1, \exp(-\beta\Delta H))$ , where  $\beta$  is the inverse temperature of this simulated-annealing-like process, and  $\Delta H$  is the change in the  $L^1$  distance between the  $3k$ -distribution in the current graph and the target  $3k$ -distribution before and after the swap. This probability favours and, respectively, suppresses  $2k$ -swaps that move the graph closer or farther from the target  $3k$ -distribution. Unfortunately, we report that in agreement with<sup>40</sup> this  $2k$ -preserving  $3k$ -targeting process never converged for any considered real network—regardless of how long we let the rewiring code run, after the initial rapid decrease, the  $3k$ -distance, while continuing to slowly decrease, remained substantially large. The reason why this process never converges is that the  $3k$ -distribution is extremely constraining, so that the number of  $3k$ -graphs  $\mathcal{N}_3$  is infinitesimally small compared with the number of  $2k$ -graphs  $\mathcal{N}_2$ ,  $\mathcal{N}_3/\mathcal{N}_2 \ll 1$  (refs 16,30). Therefore, it is extremely difficult for the  $3k$ -targeting Markov chain to find a rare path to the target  $3k$ -distribution, and the process gets hopelessly trapped in abundant local minima in distance  $H$ .

Therefore, on one hand, even though  $3k$ -randomized versions of many real networks are indistinguishable from the original networks across many metrics<sup>47</sup>, we cannot use this fact to claim that at  $d=3$  these metrics are not statistically significant in those networks, because the  $3k$ -randomization Markov chain may be non-ergodic. On the other hand, we cannot generate the corresponding  $3k$ -random graphs from scratch in a feasible amount of compute time. The  $3k$ -random graph ensemble is not analytically tractable either. Given that  $d=2$  is not enough to guarantee the statistical insignificance of some important properties of some real networks, see ref. 47 and below, we, as in ref. 40, retreat to numeric investigations of  $2k$ -random graphs

in which in addition to the  $2k$ -distribution, some substatistics of the  $3k$ -distribution is fixed. Since strong clustering is a ubiquitous feature of many real networks<sup>1</sup>, one of the most interesting such substatistics is clustering.

Specifically we study  $2.1k$ -random graphs, defined as  $2k$ -random graphs with a given value of average clustering  $\bar{c}$ , and  $2.5k$ -random graphs, defined as  $2k$ -random graphs with given values of average clustering  $\bar{c}(k)$  of nodes of degree  $k$  (ref. 40). The  $3k$ -distribution fully defines both  $2.1k$ - and  $2.5k$ -statistics, while  $2.5k$  defines  $2.1k$ . Therefore,  $2k$ -graphs are a superset of  $2.1k$ -graphs, which are a superset of  $2.5k$ -graphs, which in turn contain all the  $3k$ -graphs,  $\mathcal{N}_2 > \mathcal{N}_{2.1} > \mathcal{N}_{2.5} > \mathcal{N}_3$ . Therefore if a particular property is not statistically significant in  $2.5k$ -random graphs, for example, then it is not statistically significant in  $3k$ -random graphs either, while the converse is not generally true.

We thus generate 20  $dk$ -random graphs with  $d=0, 1, 2, 2.1, 2.5$  for each considered real network. For  $d=0,1,2$  we use the standard  $dk$ -randomizing swapping, Fig. 2. We do not use its modifications to guarantee exactly uniform sampling<sup>20,21</sup>, because: (1) even without these modifications the swapping is close to uniform in power-law graphs, (2) these modifications are non-trivial to efficiently implement, and (3) we could not extend these modifications to the  $2.1k$  and  $2.5k$  cases. As a consequence, our sampling is not exactly uniform, but we believe it is close to uniform for the reasons discussed above. To generate  $dk$ -random graphs with  $d=2.1, 2.5$ , we start with a  $2k$ -random graph, and apply to it the standard  $2k$ -preserving  $2.xk$ -targeting ( $x=1, 5$ ) rewiring process, Fig. 2. The algorithms that do that, as described in ref. 40, did not converge on some networks, so that we modified the algorithm in ref. 10 to ensure the convergence in all cases. The details of these modifications are in Supplementary Methods (the parameters used are listed in Supplementary Table 4), along with the details of the software package implementing these algorithms that we release to public<sup>51</sup>.

**Real versus  $dk$ -random networks.** We performed an extensive set of numeric experiments with six real networks—the US air transportation network, an fMRI map of the human brain, the Internet at the level of autonomous systems, a technosocial web of trust among users of the distributed Pretty Good Privacy (PGP) cryptosystem, a human protein interaction map, and an English word adjacency network (Supplementary Note 2 and Supplementary Table 3 present the analysed networks). For each network we compute its average degree, degree distribution, degree correlations, average clustering, averaging clustering of nodes of degree  $k$  and based on these  $dk$ -statistics generate a number of  $dk$ -random graphs as described above for each  $d=0, 1, 2, 2.1, 2.5$ . Then for each sample we compute a variety of network properties, and report their means and deviations for each combination of the real network,  $d$ , and the property. Figures 3–6 present the results for the PGP network; Supplementary Note 3, Supplementary Figs 1–10, and Supplementary Tables 1–2 provide the complete set of results for all the considered real networks. The reason why we choose the PGP network as our main example is that this network appears to be 'least random' among the considered real networks, in the sense that the PGP network requires higher values of  $d$  to reproduce its considered properties. The only exception is the brain network. Some of its properties are not reproduced even by  $d=2.5$ .

Figure 2 visualizes the PGP network and its  $dk$ -randomizations. The figure illustrates the convergence of  $dk$ -series applied to this network. While the  $0k$ -random graph has very little in common with the real network, the  $1k$ -random one is somewhat more similar, even more so for  $2k$ , and there is very little visual

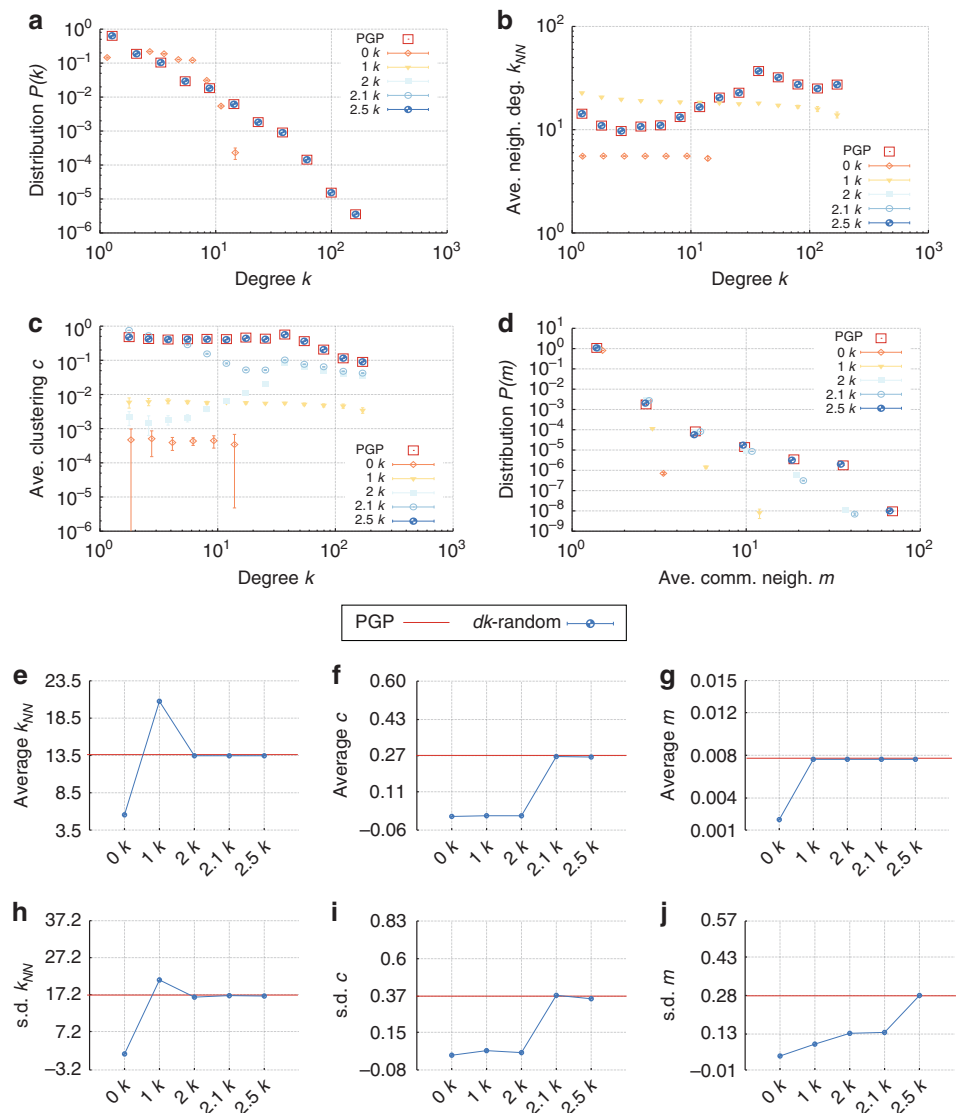
difference between the real PGP network and its 2.5k-random counterpart. This figure is only an illustration though, and to have a better understanding of how similar the network is to its randomization, we compare their properties.

We split the properties that we compare into the following categories. The microscopic properties are local properties of individual nodes and subgraphs of small size. These properties can be further subdivided into those that are defined by the  $dk$ -distributions—the degree distribution, average neighbour degree, clustering, Fig. 3—and those that are not fixed by the  $dk$ -distributions—the concentrations of subgraphs of size 3 and 4, Fig. 4. The mesoscopic properties— $k$ -coreness and  $k$ -density (the latter is also known as  $m$ -coreness or edge multiplicity, Supplementary Note 1), Fig. 5—depend both on local and global aspects of network organization. Finally, the macroscopic properties are truly global ones—betweenness, the distribution of hop lengths of shortest paths, and spectral properties, Fig. 6. In Supplementary Note 3 we also report some extremal properties, such as the graph diameter (the length of the longest shortest

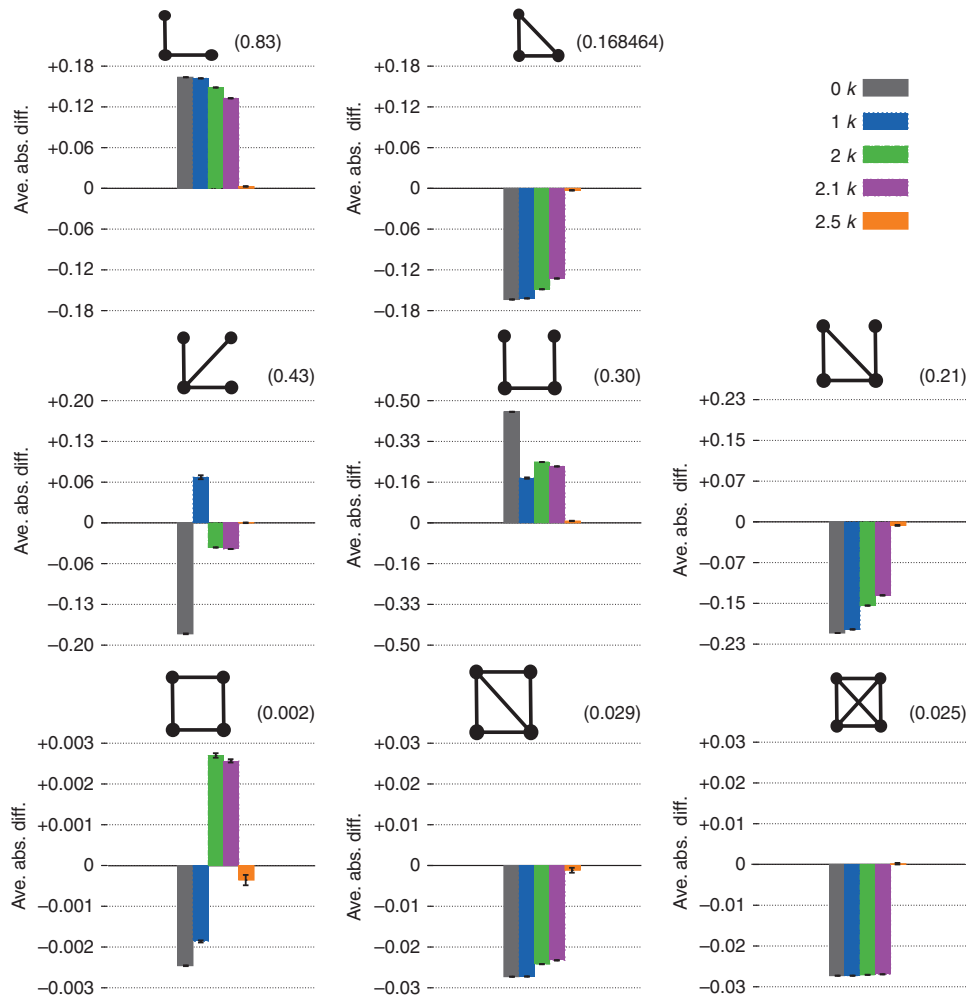
path), and Kolmogorov–Smirnov distances between the distributions of all the considered properties in real networks and their corresponding  $dk$ -random graphs. The detailed definitions of all the properties that we consider can be found in Supplementary Note 1.

In most cases—henceforth by ‘case’ we mean a combination of a real network and one of its considered property—we observe a nice convergence of properties as  $d$  increases. In many cases there is no statistically significant difference between the property in the real network and in its 2.5k-random graphs. In that sense these graphs, that is, random graphs whose degree distribution, degree correlations, and degree-dependent clustering  $\bar{c}(k)$  are as in the original network, capture many other important properties of the real network.

Some properties always converge. This is certainly true for the microscopic properties in Fig. 3, simply confirming that our  $dk$ -sampling algorithm operates correctly. But many properties that are not fixed by the  $dk$ -distributions converge as well. Neither the concentration of subgraphs of size 3 nor the distribution of



**Figure 3 | Microscopic properties of the PGP network and its  $dk$ -random graphs.** (a) The degree distribution  $P(k)$ , (b) average degree  $\bar{k}_{nn}(k)$  of nearest neighbours of nodes of degree  $k$ , (c) average clustering  $\bar{c}(k)$  of nodes of degree  $k$ , (d) the distribution  $P(m)$  of the number  $m$  of common neighbours between all connected pairs of nodes, and (e–g) the means and (h–j) s.d. of the corresponding distributions. The error bars indicate the s.d. of the metrics across different graph realizations.



**Figure 4 | The densities of subgraphs of size 3 and 4 in the PGP network and its  $dk$ -random graphs.** The two different graphs of size 3 and six different graphs of size 4 are shown on each panel. The numbers on top of panels are the concentrations of the corresponding subgraph in the PGP network, while the histogram heights indicate the average absolute difference between the subgraph concentration in the  $dk$ -random graphs and its concentration in the PGP network. The subgraph concentration is the number of given subgraphs divided by the total number of subgraphs of the same size. The error bars are the s.d. across different graph realizations.

the number of neighbours common to a pair of nodes are fully fixed by  $dk$ -distributions with any  $d < 3$  by definition, yet  $2.5k$ -random graphs reproduce them well in all the considered networks. Most subgraphs of size 4 are also captured at  $d = 2.5$  in most networks, even though  $d = 3$  would not be enough to exactly reproduce the statistics of these subgraphs. We note that the improvement in subgraph concentrations at  $d = 2.5$  compared with  $d = 2.1$  is particularly striking, Fig. 4. The mesoscopic and especially macroscopic properties converge more slowly as expected. Nevertheless, quite surprisingly, both mesoscopic properties ( $k$ -coreness and  $k$ -density) and some macroscopic properties converge nicely in most cases. The  $k$ -coreness,  $k$ -density, and the spectral properties, for instance, converge at  $d = 2.5$  in all the considered cases other than Internet’s Fiedler value. In some cases a property, even global one, converges for  $d < 2.5$ . Betweenness, for example, a global property, converges at  $d = 1$  for the Internet and English word network.

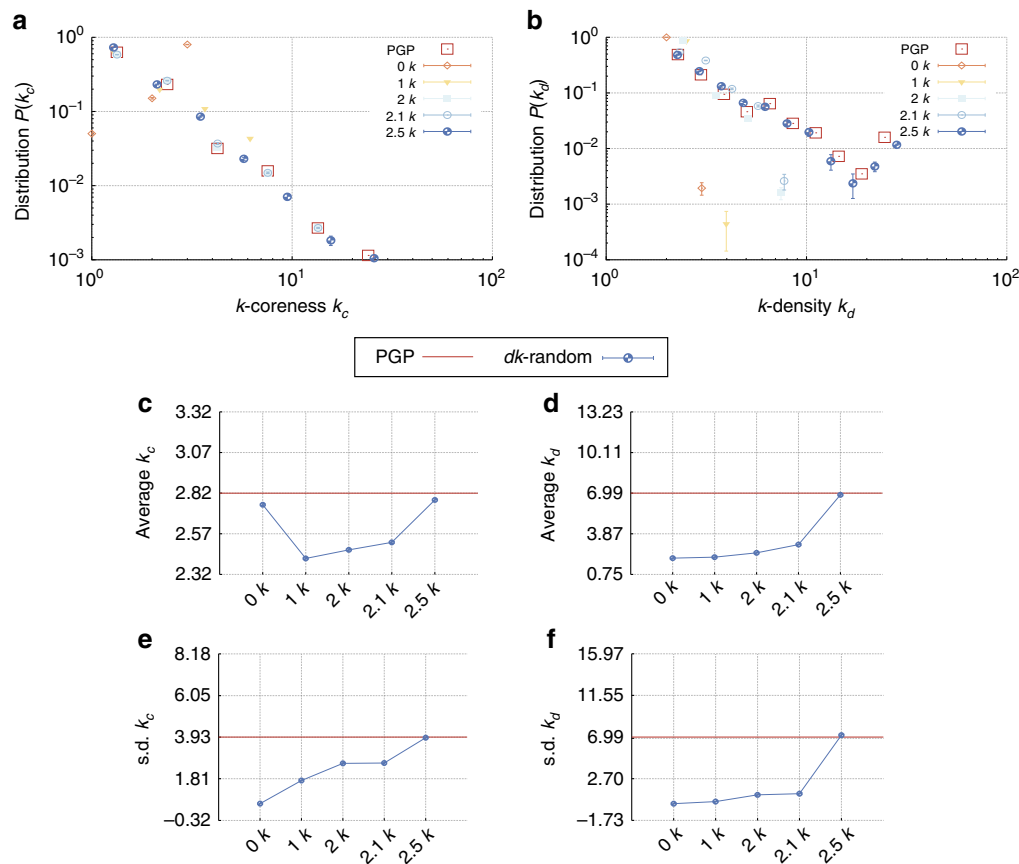
Finally, there are ‘outlier’ networks and properties of poor or no  $dk$ -convergence. Many properties of the brain network, for example, exhibit slow or no convergence. We have also experimented with community structure inferred by different algorithms, and in most cases the convergence is either slow or non-existent as one could expect.

### Discussion

In general, we should not expect non-local properties of networks to be exactly or even closely reproduced by random graphs with local constraints. The considered brain network is a good example of that this expectation is quite reasonable. The human brain consists of two relatively weakly connected parts, and no  $dk$ -randomization with low  $d$  is expected to reproduce this peculiar global feature, which likely has an impact on other global properties. And indeed we observe in Supplementary Note 3 that its two global properties, the shortest path distance and betweenness distributions, differ drastically between the brain and its  $dk$ -randomizations.

Another good example is community structure, which is not robust with respect to  $dk$ -randomizations in all the considered networks. In other words,  $dk$ -randomizations destroy the original peculiar cluster organization in real networks, which is not surprising, as clusters have too many complex non-local features such as variable densities of internal links, boundaries and so on, which  $dk$ -randomizations, even with high  $d$ , are expected to affect considerably.

Surprisingly, what happens for the brain and community structure does not appear representative for many other considered combinations of real networks and their properties.



**Figure 5 | Mesoscopic properties, the  $k$ -coreness and  $k$ -density distributions, in the PGP network and its  $dk$ -random graphs.** The figure shows the distributions  $P(k_{c,d})$  of (a) node  $k$ -coreness  $k_c$  and (b) edge  $k$ -density  $k_d$ , and their (c,d) means and (e,f) s.d. The  $k_c$ -core of a graph is its maximal subgraph in which all nodes have degree at least  $k_c$ . The  $k_d$ -core of a graph is its maximal subgraph in which all edges have multiplicity at least  $k_d - 2$ ; the multiplicity of an edge is the number of common neighbours between the nodes that this edge connects, or equivalently the number of triangles that this edge belongs to. A node has  $k$ -coreness  $k_c$  if it belongs to the  $k_c$ -core but not to the  $k_c + 1$ -core. An edge has  $k$ -density  $k_d$  if it belongs to the  $k_d$ -core but not to the  $k_d + 1$ -core. The error bars indicate the s.d. of the metrics across different graph realizations.

As a possible explanation, one can think of constraint-based modelling as a satisfiability (SAT) problem: find the elements of the adjacency matrix (1/0, True/False) such that all the given constraints in terms of the functions of the marginals (degrees) of this matrix are obeyed. We then expect that the  $3k$ -constraints already correspond to an NP-hard SAT problem, such as 3-SAT, with hardness coming from the global nature of the constraints in the problem. However, many real-world networks evolve based mostly on local dynamical rules and thus we would expect them to contain correlations with  $d < 3$ , that is, below the NP-hard barrier. The primate brain, however, has likely evolved through global constraints, as indicated by the dense connectivity across all functional areas and the existence of a strong core-periphery structure in which the core heavily concentrates on areas within the associative cortex, with connections to and from all the primary input and subcortical areas<sup>12</sup>.

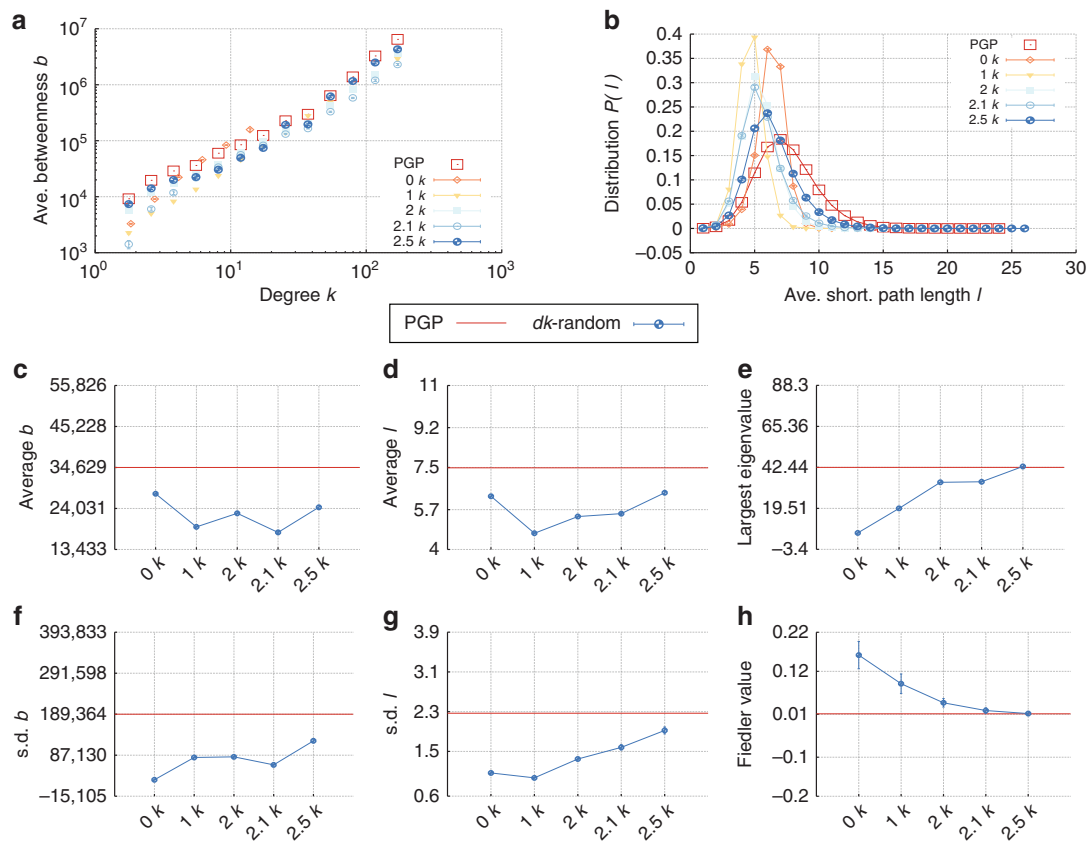
However, in most cases, the considered networks are  $dk$ -random with  $d \leq 2.5$ , that is,  $d \leq 2.5$  is enough to reproduce not only basic microscopic (local) properties but also mesoscopic and even macroscopic (global) network properties<sup>6–10</sup>. This finding means that these more sophisticated properties are effectively random in the considered networks, or more precisely, that the observed values of these properties are effective consequences of particular degree distributions and, optionally, degree correlations and clustering that the networks have. This further implies that attempts to find explanations for these

complex but effectively random properties should probably be abandoned, and redirected to explanations of why and how degree distributions, correlations and clustering emerge in real networks, for which there already exists a multitude of approaches<sup>52–57</sup>. On the other hand, the features that we found non-random do require separate explanations, or perhaps a different system of null models.

We reiterate that the  $dk$ -randomization system makes it clear that there is no *a priori* preferred null model for network randomization. To tell how statistically significant a particular feature is, it is necessary to compare this feature in the real network against the same feature in an ensemble of random graphs, a null model. But one is free to choose any random graph model. In particular, any  $d$  defines a random graph ensemble, and we find that many properties, most notably the frequencies of small subgraphs that define motifs<sup>11</sup>, strongly depend on  $d$  for many considered networks. Therefore, choosing any specific value of  $d$ , or more generally, any specific null model to study the statistical significance of a particular structural network feature, requires some non-trivial justification before this feature can be claimed important for any network function.

Yet another implication of our results is that if one looks for network topology generators that would veraciously reproduce certain properties of a given real network—a task that often comes up in as diverse disciplines as biology<sup>58</sup> and computer science<sup>59</sup>—one should first check how  $dk$ -random these





**Figure 6 | Macroscopic properties of the PGP network and its  $dk$ -random graphs.** (a) The average betweenness  $\bar{b}(k)$  of nodes of degree  $k$ , (b) the distribution  $P(l)$  of hop lengths  $l$  of the shortest paths between all pairs of nodes, the (c,d) means and (f,g) s.d. of the corresponding distributions, (e) the largest eigenvalues of the adjacency matrix  $A$ , and (h) the Fiedler value, which is the spectral gap (the second largest eigenvalue) of the graph's Laplacian matrix  $L = D - A$ , where  $D$  is the degree matrix,  $D_{ij} = \delta_{ij}k_i$ ,  $\delta_{ij}$  the Kronecker delta, and  $k_i$  the degree of node  $i$ . The error bars indicate the s.d. of the metrics across different graph realizations.

properties are. If they are  $dk$ -random with low  $d$ , then one may not need any sophisticated mission-specific topology generators. The  $dk$ -random graph-generation algorithms discussed here can be used for that purpose in this case. We note that there exists an extension of a subset of these algorithm for networks with arbitrary annotations of links and nodes<sup>60</sup>—directed or coloured (multilayer) networks, for instance.

The main caveat of our approach is that we have no proof that our  $dk$ -random graph generation algorithms for  $d=2.1$  and  $d=2.5$  sample graphs uniformly at random from the ensemble. The random-graph ensembles and edge-rewiring processes employed here are known to suffer from problems such as degeneracy and hysteresis<sup>35,61,62</sup>. Ideally, we would wish to calculate analytically the exact expected value of a given property in an ensemble. This is currently possible only for very simple properties in soft ensembles with  $d=0, 1, 2$  (refs 36,37). Some mathematically rigorous results are available for  $d=0, 1$  and for some exponential random-graph models<sup>28,34</sup>. Many of these results rely on graphons<sup>24</sup> that are applicable to dense graphs only, while virtually all real networks are sparse<sup>49</sup>. Some rigorous approaches to sparse networks are beginning to emerge<sup>63,64</sup>, but the rigorous treatment of global properties, which tend to be highly non-trivial functions of adjacency matrices, in random graph ensembles with  $d > 2$  constraints, appear to be well beyond the reach in the near future. Yet if we ever want to fully understand the relationship between the structure, function and dynamics of real networks, this future research direction appears to be of a paramount importance.

**References**

- Newman, M. E. J. *Networks: An Introduction* (Oxford Univ. Press, 2010).
- Barrat, A. *Dynamical Processes on Complex Networks* (Cambridge Univ. Press, 2008).
- Newman, M. E. J. The structure and function of complex networks. *SIAM Rev.* **45**, 167–256 (2003).
- Adilson, E. Motter and Zoltán Toroczkai. Introduction: optimization in networks. *Chaos* **17**, 026101 (2007).
- Burda, Z., Krzywicki, A., Martin, O. C. & Zagorski, M. Motifs emerge from function in model gene regulatory networks. *Proc. Natl Acad. Sci. USA* **108**, 17263–17268 (2011).
- Vázquez, A. et al. The topological relationship between the large-scale attributes and local interaction patterns of complex networks. *Proc. Natl Acad. Sci. USA* **101**, 17940–17945 (2004).
- Guimerá, R., Sales-Pardo, M. & Luis, A. N. A. Classes of complex networks defined by role-to-role connectivity profiles. *Nat. Phys.* **3**, 63–69 (2007).
- Takemoto, K., Oosawa, C. & Akutsu, T. Structure of n-clique networks embedded in a complex network. *Phys. A* **380**, 665–672 (2007).
- Foster, D. V., Foster, J. G., Grassberger, P. & Paczuski, M. Clustering drives assortativity and community structure in ensembles of networks. *Phys. Rev. E* **84**, 066117 (2011).
- Colomer-de-Simón, P., Serrano, M. Á., Beiró, M. G., Alvarez-Hamelin, J. I. & Boguñá, M. Deciphering the global organization of clustering in real complex networks. *Sci. Rep.* **3**, 2517 (2013).
- Milo, R. et al. Network motifs: simple building blocks of complex networks. *Science* **298**, 824–827 (2002).
- Markov, N. T. et al. Cortical high-density counterstream architectures. *Science* **342**, 1238406 (2013).
- Luis, A. Amaral and Roger Guimera. Complex networks: Lies, damned lies and statistics. *Nat. Phys.* **2**, 75–76 (2006).
- Colizza, V., Flammini, A., Ángeles Serrano, M. & Vespignani, A. Detecting rich-club ordering in complex networks. *Nat. Phys.* **2**, 110–115 (2006).

15. Trevino, III J., Nyberg, A., Del Genio, C. I. & Bassler, K. E. Fast and accurate determination of modularity and its effect size. *J. Stat. Mech.* **15**, P02003 (2015).
16. Mahadevan, P., Krioukov, D., Fall, K. & Vahdat, A. Systematic topology analysis and generation using degree correlations. *Comput. Commun. Rev.* **36**, 135–146 (2006).
17. Yaveroglu, Ö. N. *et al.* Revealing the hidden language of complex networks. *Sci. Rep.* **4**, 4547 (2014).
18. Karrer, B. & Newman, M. E. J. Random graphs containing arbitrary distributions of subgraphs. *Phys. Rev. E* **82**, 066118 (2010).
19. Coolen, A. C. C., Fraternali, F., Annibale, A., Fernandes, L. & Kleinjung, J. *Modelling Biological Networks via Tailored Random Graphs* 309–329 (John Wiley & Sons, Ltd, 2011).
20. Coolen, A. C. C., Martino, A. & Annibale, A. Constrained Markovian Dynamics of Random Graphs. *J. Stat. Phys.* **136**, 1035–1067 (2009).
21. Annibale, A., Coolen, A. C. C., Fernandes, L., Fraternali, F. & Kleinjung, J. Tailored graph ensembles as proxies or null models for real networks I: tools for quantifying structure. *J. Phys. A Math. Gen.* **42**, 485001 (2009).
22. Roberts, E. S., Schlitt, T. & Coolen, A. C. C. Tailored graph ensembles as proxies or null models for real networks II: results on directed graphs. *J. Phys. A Math. Theor.* **44**, 275002 (2011).
23. Roberts, E. S. & Coolen, A. C. C. Unbiased degree-preserving randomization of directed binary networks. *Phys. Rev. E* **85**, 046103 (2012).
24. Lovász, L. *Large Networks and Graph Limits* (American Mathematical Society, 2012).
25. Erdős, P. & Rényi, A. On Random Graphs. *Publ. Math.* **6**, 290–297 (1959).
26. Bender, E. & Canfield, E. The asymptotic number of labelled graphs with given degree distribution. *J. Comb. Theor. A* **24**, 296–307 (1978).
27. Newman, M. E. J., Strogatz, S. H. & Watts, D. J. Random graphs with arbitrary degree distributions and their applications. *Phys. Rev. E* **64**, 26118 (2001).
28. Chatterjee, S., Diaconis, P. & Sly, A. Random graphs with a given degree sequence. *Ann. Appl. Probab.* **21**, 1400–1435 (2011).
29. Stanton, I. & Pinar, A. Constructing and sampling graphs with a prescribed joint degree distribution. *J. Exp. Algorithm.* **17**, 3.1 (2012).
30. Bianconi, G. The entropy of randomized network ensembles. *Eur. Lett.* **81**, 28005 (2008).
31. Barvinok, A. & Hartigan, J. A. The number of graphs and a random graph with a given degree sequence. *Random Struct. Algorithm.* **42**, 301–348 (2013).
32. Holland, P. W. & Leinhardt, S. An exponential family of probability distributions for directed graphs. *J. Am. Stat. Assoc.* **76**, 33–50 (1981).
33. Park, J. & Newman, M. E. J. Statistical mechanics of networks. *Phys. Rev. E* **70**, 66117 (2004).
34. Chatterjee, S. & Diaconis, P. Estimating and understanding exponential random graph models. *Ann. Stat.* **41**, 2428–2461 (2013).
35. Horvát, S., Czabarka, É. & Toroczkai, Z. Reducing degeneracy in maximum entropy models of networks. *Phys. Rev. Lett.* **114**, 158701 (2015).
36. Squartini, T. & Garlaschelli, D. Analytical maximum-likelihood method to detect patterns in real networks. *N. J. Phys.* **13**, 083001 (2011).
37. Squartini, T., Mastrandrea, R. & Garlaschelli, D. Unbiased sampling of network ensembles. *N. J. Phys.* **17**, 023052 (2015).
38. Maslov, S., Sneppen, K. & Alon, U. *Handbook of Graphs and Networks* chapter 8 (Wiley-VCH, 2003).
39. Maslov, S., Sneppen, K. & Zaliznyak, A. Detection of topological patterns in complex networks: Correlation profile of the Internet. *Phys. A* **333**, 529–540 (2004).
40. Gjoka, M., Kurant, M. & Markopoulou, A. In *2013 Proceedings of IEEE INFOCOM 1968–1976* (IEEE, 2013).
41. Kim, H., Toroczkai, Z., Erdős, P. L., Miklós, I. & Székely, L. A. Degree-based graph construction. *J. Phys. A Math. Theor.* **42**, 392001 (2009).
42. Del Genio, C. I., Kim, H., Toroczkai, Z. & Bassler, K. E. Efficient and exact sampling of simple graphs with given arbitrary degree sequence. *PLoS ONE* **5**, e10012 (2010).
43. Kim, H., Del Genio, C. I., Bassler, K. E. & Toroczkai, Z. Constructing and sampling directed graphs with given degree sequences. *N. J. Phys.* **14**, 023012 (2012).
44. Bassler, K. E., Del Genio, C. I., Erdős, P. L., Miklós, I. & Toroczkai, Z. Exact sampling of graphs with prescribed degree correlations. *N. J. Phys.* **17**, 083052 (2015).
45. Zlatic, V. *et al.* On the rich-club effect in dense and weighted networks. *Eur. Phys. J. B* **67**, 271–275 (2009).
46. Czabarka, É., Dutle, A., Erdős, P. L. & Miklós, I. On realizations of a joint degree matrix. *Discret. Appl. Math.* **181**, 283–288 (2015).
47. Jamakovic, A., Mahadevan, P., Vahdat, A., Boguñá, M. & Krioukov, D. How small are building blocks of complex networks. Preprint at <http://arxiv.org/abs/0908.1143> (2009).
48. Milo, R., Kashtan, N., Itzkovitz, S., Newman, M. E. J. & Alon, U. On the uniform generation of random graphs with prescribed degree sequences. Preprint at <http://arxiv.org/abs/cond-mat/0312028> (2003).
49. Del Genio, C., Gross, T. & Bassler, K. E. All scale-free networks are sparse. *Phys. Rev. Lett.* **107**, 1–4 (2011).
50. Gjoka, M., Tillman, B. & Markopoulou, A. In *2015 Proceedings of IEEE INFOCOM*. pages 1553–1561 (IEEE, 2015).
51. de Simon, P. C. RandNetGen: a Random Network Generator. URL: <http://polcolomer.github.io/RandNetGen/>.
52. Dorogovtsev, S. N., Mendes, J. & Samukhin, A. Size-dependent degree distribution of a scale-free growing network. *Phys. Rev. E* **63**, 062101 (2001).
53. Klemm, K. & Eguluz, V. Highly clustered scale-free networks. *Phys. Rev. E* **65**, 036123 (2002).
54. Vázquez, A. Growing network with local rules: Preferential attachment, clustering hierarchy, and degree correlations. *Phys. Rev. E* **67**, 056104 (2003).
55. Serrano, M. Á. & Baguna, M. Tuning clustering in random networks with arbitrary degree distributions. *Phys. Rev. E* **72**, 036133 (2005).
56. Papadopoulos, F., Kitsak, M., Serrano, M. Á., Boguñá, M. & Krioukov, D. Popularity versus similarity in growing networks. *Nature* **489**, 537–540 (2012).
57. Bianconi, G., Darst, R. K., Iacovacci, J. & Fortunato, S. Triadic closure as a basic generating mechanism of communities in complex networks. *Phys. Rev. E* **90**, 042806 (2014).
58. Kuo, P., Banzhaf, W. & Leier, A. Network topology and the evolution of dynamics in an artificial genetic regulatory network model created by whole genome duplication and divergence. *Biosystems* **85**, 177–200 (2006).
59. Medina, A., Lakhina, A., Matta, I. & Byers, J. In *MASCOTS 2001, Proceedings of the Ninth International Symposium in Modeling, Analysis and Simulation of Computer and Telecommunication Systems* 346–353 (Washington, DC, USA, 2001).
60. Dimitropoulos, X., Krioukov, D., Riley, G. & Vahdat, A. Graph annotations in modeling complex network topologies. *ACM T Model. Comput. S* **19**, 17 (2009).
61. Foster, D., Foster, J., Paczowski, M. & Grassberger, P. Communities, clustering phase transitions, and hysteresis: Pitfalls in constructing network ensembles. *Phys. Rev. E* **81**, 046115 (2010).
62. Roberts, E. S. & Coolen, A. C. C. Random graph ensembles with many short loops. In *ESAIM Proc. Surv.* **47**, 97–115 (2014).
63. Bollobás, B. & Riordan, O. Sparse graphs: metrics and random models. *Random Struct. Algorithm.* **39**, 1–38 (2011).
64. Borgs, C., Chayes, J. T., Cohn, H. & Zhao, Y. An  $L^p$  theory of sparse graph convergence I: Limits, sparse random graph models, and power law distributions. Preprint at <http://arxiv.org/abs/1401.2906> (2014).
65. Beiró, M. G., Alvarez-Hamelin, J. I. & Busch, J. R. A low complexity visualization tool that helps to perform complex systems analysis. *N. J. Phys.* **10**, 125003 (2008).

## Acknowledgements

We acknowledge financial support by NSF Grants No. CNS-1039646, CNS-1345286, CNS-0722070, CNS-0964236, CNS-1441828, CNS-1344289, CNS-1442999, CCF-1212778, and DMR-1206839; by AFOSR and DARPA Grants No. HR0011-12-1-0012 and FA9550-12-1-0405; by DTRA Grant No. HDTRA-1-09-1-0039; by Cisco Systems; by the Ministry of Education, Science, and Technological Development of the Republic of Serbia under Project No. ON171017; by the ICREA Academia Prize, funded by the *Generalitat de Catalunya*; by the Spanish MINECO Project No. FIS2013-47282-C2-1-P; by the *Generalitat de Catalunya* Grant No. 2014SGR608; and by European Commission Multiplex FP7 Project No. 317532.

## Author contributions

All authors contributed to the development and/or implementation of the concept, discussed and analysed the results. C.O., M.M.D., and P.C.S. implemented the software for generating  $dk$ -graphs and analysed their properties. D.K. wrote the manuscript, incorporating comments and contributions from all authors.

## Additional information

**Supplementary Information** accompanies this paper at <http://www.nature.com/naturecommunications>

**Competing financial interests:** The authors declare no competing financial interests.

**Reprints and permission** information is available online at <http://npg.nature.com/reprintsandpermissions>.

**How to cite this article:** Orsini, C. *et al.* Quantifying randomness in real networks. *Nat. Commun.* 6:8627 doi: 10.1038/ncomms9627 (2015).



This work is licensed under a Creative Commons Attribution 4.0 International License. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in the credit line; if the material is not included under the Creative Commons license, users will need to obtain permission from the license holder to reproduce the material. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>