ARTICLE

Check for updates

Growing scale-free simplices

Kiriil Kovalenko [1,14], Irene Sendiña-Nadal [2,3,14 ✉], Nagi Khalil [4,14], Alex Dainiak[5], Daniil Musatov[1,5,6], Andrei M. Raigorodskii[1,6,7,8], Karin Alfaro-Bittner [9,10], Baruch Barzel[11,12,14] & Stefano Boccaletti[1,2,9,13,14]

The past two decades have seen significant successes in our understanding of networked systems, from the mapping of real-world networks to the establishment of generative models recovering their observed macroscopic patterns. These advances, however, are restricted to pairwise interactions and provide limited insight into higher-order structures. Such multi-component interactions can only be grasped through simplicial complexes, which have recently found applications in social, technological, and biological contexts. Here we introduce a model to grow simplicial complexes of order two, i.e., nodes, links, and triangles, that can be straightforwardly extended to structures containing hyperedges of larger order. Specifically, through a combination of preferential and/or nonpreferential attachment mechanisms, the model constructs networks with a scale-free degree distribution and an either bounded or scale-free generalized degree distribution. We arrive at a highly general scheme with analytical control of the scaling exponents to construct ensembles of synthetic complexes displaying desired statistical properties.

[1] Moscow Institute of Physics and Technology, National Research University, Dolgoprudny, Moscow, Russia. [2] Complex Systems Group & GISC, Universidad Rey Juan Carlos, Móstoles, Madrid, Spain. [3] Center for Biomedical Technology, Universidad Politécnica de Madrid, Madrid, Spain. [4] ESCET & GISC, Universidad Rey Juan Carlos, Madrid, Móstoles, Spain. [5] Russian Academy of National Economy and Public Administration, Moscow, Russia. [6] Caucasus Mathematical Center, Adyghe State University, Maykop, Russia. [7] Mechanics and Mathematics Faculty, Moscow State University, Moscow, Russia. [8] Institute of Mathematics and Computer Science, Buryat State University, Ulan-Ude, Russia. [9] Unmanned Systems Research Institute, Northwestern Polytechnical University, Xi'an, China. [10] Departamento de Física, Universidad Técnica Federico Santa María, Valparaíso, Chile. [11] Department of Mathematics, Bar-Ilan University, Ramat Gan, Israel. [12] Gonda Multidisciplinary Brain Research Center, Bar-Ilan University, Ramat Gan, Israel. [13] CNR - Institute of Complex Systems, Sesto Fiorentino, Italy. [14] These authors contributed equally: Kiriil Kovalenko, Irene Sendiña-Nadal, Nagi Khalil, Baruch Barzel, Stefano Boccaletti. ✉email: irene.sendina@urjc.es

All beauty, richness, and harmony in the emergent dynamics of a complex system largely depend on the specific way in which its elementary components interact. The last 20 years have seen the birth and development of the multidisciplinary field of network science, wherein a variety of systems in physics, biology, social sciences, and engineering has been modeled by networks of coupled units, in the attempt to unveil the mechanisms underneath the observed systems' functionality[1–6].

But the fundamental limit of such a representation is that networks capture only pairwise interactions, whereas the function of many real-world systems not only involves dyadic connections, but rather is the outcome of collective actions at the level of groups of nodes[7]. For instance, in ecological systems, three or more species compete for food or territory[8]. Similar multi-component interactions appear in functional[9–13] and structural[14] brain networks, protein interaction networks[15], semantic networks[16], multiauthors scientific collaborations[17], offline and online social networks[18,19], trigenic interactions in gene regulatory networks[20,21], and spreading of social behavior due to multiple, simultaneous, social interactions[22].

Simplicial complexes (SCs), being structures formed by simplices of different dimensions (nodes, links, triangles, tetrahedra, etc.), can effectively map the relationships between any number of components. Originally introduced over two decades ago[23], SCs are becoming increasingly relevant, thanks to the enhanced resolution of current datasets and the recent advances in data analysis techniques[24,25]. As real data are being accumulated, we encounter a theoretical challenge: how to synthesize SCs that faithfully reproduce the observed structural features. Significant progresses were made in extending to SCs static graph models, such as random graphs[26–28], configuration models[29–31], or activity-driven models[32]. The fact is however that in many circumstances (such as in the case of scientists collaborating with and citing each other), the network is the result of a growing process, and therefore another approach to model SCs is that of out-of-equilibrium models. Most of the models falling within this category have a geometrical interpretation[33] and are variations of the so-called "network geometry with flavor" (NGF) model[34–38], which aims at providing a theoretical basis to characterize the underlying geometry of complex networks[39] or to explore hidden geometries in complex materials by aggregating SCs as fundamental building blocks[40]. While the original NGF model adds one[34] or more[37] maximal simplices of fixed dimension at each time step of the growth process, Fountoulakis et al.[41] introduced even the option to remove an existing simplex when a new one is added.

In this paper, we discuss the NGF model proposed by Courtney and Bianconi[37] that is able to grow SCs of order two, i.e., structures made of nodes, links and triangles, making use of preferential and nonpreferential rules, and we generalize this model by proposing a model that makes a combination of the two mechanisms. The resulting SCs are characterized by two distributions, the classic degree distribution $P(k)$, capturing the fraction of nodes with degree k, and the generalized degree distribution $P(k_\ell)$, where k_ℓ characterizes the number of triangles supported by each link $l = (i, j)$. We show that our generative model always yields a power-law scaling in $P(k)$, recovering the ubiquitously observed scale-free property[42,43] and, at the same time, it allows full control over $P(k_\ell)$, i.e., bounded or scale-free with any desired scaling exponent. Indeed, $P(k_\ell)$ has been shown to play a crucial role in the emergence of collective behavior, such as synchronization[44]. We will furthermore show that our methods can be straightforwardly extended to grow structures containing hyperedges of size larger than two.

Results and discussion

SCs growth models. Let us, for the time being and unless otherwise specified, concentrate on NGFs models of SCs of order two proposed by Courtney and Bianconi[37]. The purpose is to grow a network of N nodes featuring a transitivity coefficient $T = 1$[45,46], thus implying that each link is part of a connected triplet of nodes (a triad) which forms a triangle. This is done by starting at $t = 0$ with an elementary network seed, consisting in an initial clique (an all-to-all connected network) of size $N_0 \ll N$. Then, at each successive times $t = 1, 2, 3, ..., N - N_0$, a new node is added to the graph. The added node selects m_{tri} already existing links, and forms connections with the $2m_{tri}$ nodes located at the ends of such links, thus generating m_{tri} new triangles (when $m_{tri} > 1$, a further condition is enforced that the m_{tri} selected links are not pairwise adjacent, to avoid multiple links from the added node to single existing nodes in the graph).

Such a procedure can be conducted with or without adopting a preferential attachment rule. In the latter case, the probability that the node added at time t selects the specific link (i,j) to form its connections is taken to be $P_{ij}(t) = \frac{1}{N_0(N_0-1)/2 + 2m_{tri}(t-1)}$, i.e., the m_{tri} links are randomly selected among all those which already exist in the graph at time $t - 1$ with equal probability (and therefore without applying a preferential rule). In the former case, instead, one has that $P_{ij}(t) = \frac{k_{ij}(t-1)}{\sum_{i,j} k_{ij}(t-1)}$, which implies that the larger is the number of triangles a given edge (i,j) is part of at time $t - 1$ the larger the probability for that edge of being selected to form a new triangle with the added node. These models are the NGF models proposed by Courtney and Bianconi[37] for $d = 2, s = 0$ (uniform attachment) and $d = 2, s = 1$ (preferential attachment).

Figure 1a, c reports the degree distribution $P(k)$ of the networks (of size $N = 10,000$ nodes) generated by the two methods. One immediately sees that in both cases the graphs feature a clear power-law scaling, i.e., the scale-freeness which is indeed characterizing the vast majority of real-world networks[1–6,42,43]. In Fig. 1b, d, we report, instead, a visualization of a typical synthesized network with $N = 200$ and $m_{tri} = 1$.

We also measure the generalized degree k_ℓ. Given a graph of N nodes and its adjacency matrix A (the $N \times N$ matrix with entries $a_{ij} = 1$ if nodes i and j are connected by a link, and $a_{ij} = 0$ otherwise), the generalized degree k_{ij} of the link between node i and j is the (i, j)-entry of the matrix $A \oslash A^2$ (with the symbol $\oslash$ standing for the Hadamard product), that is, $k_{ij} = (A)_{ij} \oslash (A^2)_{ij}$, which accounts for the number of triangles in which the link ij participates. Looking at Fig. 1, one can immediately see that while the two generated networks are essentially heterogeneous in the node degree, the properties of k_ℓ appear instead to be very different. In particular, the nonpreferential case leads to a much more restricted range of k_ℓ values as compared to that generated by the preferential rule (see the two color bars at the right of the panels). Moreover, the structure obtained with nonpreferential attachment is homogeneous in terms of k_ℓ (almost all the six colors, each one representing a given value of k_ℓ, are visible). On the opposite, the preferential rule generates an SC with a high heterogeneity in k_ℓ: almost all links are red (they have the lowest as possible value of k_ℓ), and only one link (the black one) displays a value of k_ℓ equal to the maximum in the distribution.

These latter features are more quantitatively visible in Fig. 2, which reports the distribution $P(k_\ell)$ of the generalized degree k_ℓ for the two cases. Comparing Fig. 2a, b, one immediately realizes that while $P(k_\ell)$ is exponentially decaying when the growth is realized with nonpreferential attachment (Fig. 2a), its scaling is a clear power law in the presence of preferential attachment.
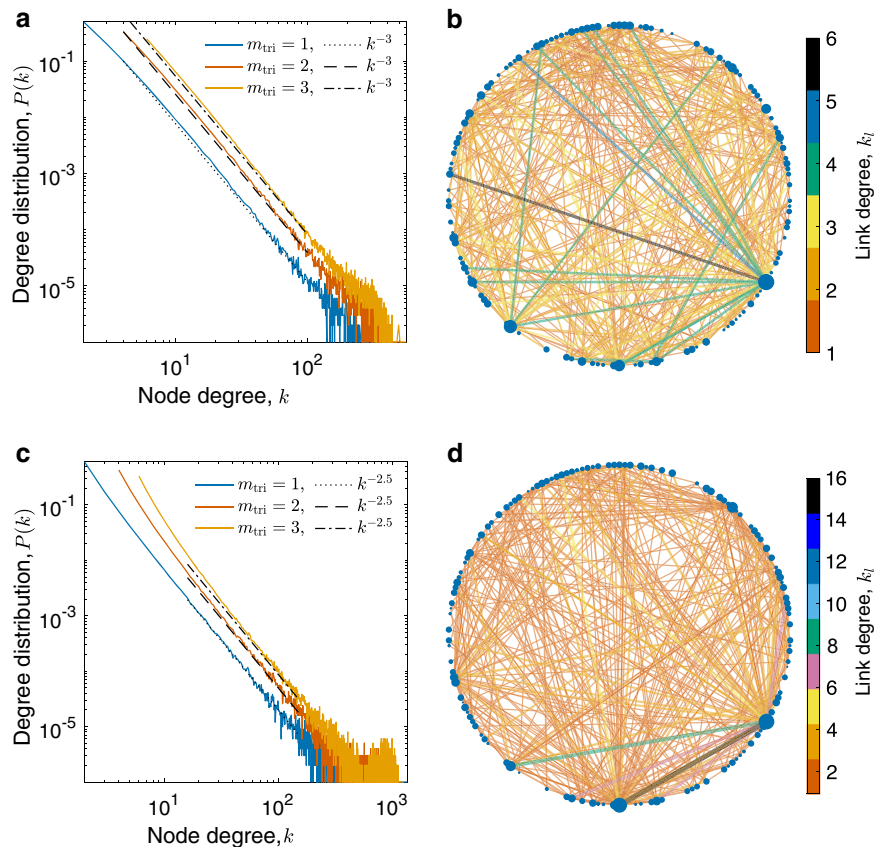
Fig. 1 The structures generated in the preferential and the nonpreferential cases. a, b The model without preferential attachment and **c, d** the model with preferential attachment. **a, c** Log–log plot of the node degree distribution $P(k)$ of the resulting networks for different values of the number of triangles m_{tri} added at each step of the growth process (see color code in the legend). Data are obtained as an ensemble average over 100 different realizations of a network with size $N = 10,000$ nodes. Dotted, dashed, and dash-dotted lines in **a** and **c** correspond to the analytical predictions given by Eqs. (3) and (9). **b, d** Schematic visualization of generated networks with $N = 200$ and $m_{tri} = 1$. The size of the nodes correlates with their influence in the network in terms of their eigenvector centrality[55], such that a node with large size implies a high eigenvector centrality, and therefore the node is connected to many nodes also of large size with high eigencentrality. The width of each link $\ell = (i,j)$ is proportional to the square root of the link degree k_ℓ, the number of triangles the link is adjacent to, and the color of the links encodes the supported number of triangles k_ℓ as reported in the bars at the right of both panels.
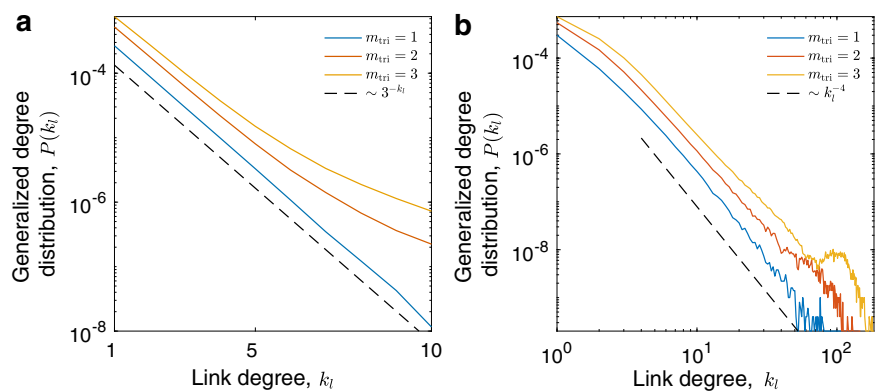

Fig. 2 The distribution of the generalized degree k_ℓ. The distribution $P(k_\ell)$ vs. the generalized—link—degree k_ℓ (number of triangles a link ℓ is adjacent to) obtained by growing a network of size $N = 10,000$ with the nonpreferential (**a**) and with the preferential (**b**) attachment models. The data refer to an ensemble average over 100 different realizations of the growth process. Notice that **a** is in log–linear scale, whereas **b** is in log–log scale. Legends in both panels report the color code for the number of triangles m_{tri} added at each step of the growth process. The dashed line in **a** is used for exponential solution given by Eq. (6), while the dashed line in **b** is used for the power-law solution given by Eq. (12).

A relevant conclusion at this stage is that the two cases are imprinting a completely different topology in the triangular structures (reflected by completely different scaling properties in the distribution of the generalized degree k_ℓ).

Nonpreferential model (NGF[37] with $d = 2$, $s = 0$). We now furnish a full analytic treatment, and provide the rigorous expressions for the distributions $P(k)$ and $P(k_\ell)$. We start with the case of nonpreferential attachment, and we call $N(k,t)$ the number

of nodes with degree k at time t. Its rate equation reads

$$\frac{dN(k,t)}{dt} = \frac{2m_{tri}}{\sum_k kN(k,t)}[-kN(k,t) + (k-1)N(k-1,t)] + \delta_{k,2m_{tri}}, \quad (1)$$

where $\frac{dN(k,t)}{dt} \equiv N(k,t+1) - N(k,t)$, and δ is the Kronecker delta function.

For $N(t) \equiv \sum_k N(k,t) \simeq t$, one seeks a solution of the form $N(k,t) = tP(k)$, where $P(k)$ is assumed to be time independent. Since the total number of edges is $\sim 2m_{tri}t$, one has $\sum_k kN(k,t) \simeq 4m_{tri}t$, and the rate equation for $P(k)$ becomes

$$P(k) = \frac{k-1}{k+2}P(k-1) + \frac{2}{k+2}\delta_{k,2m_{tri}}. \quad (2)$$

Such a latter equation is the same as Eq. (6) introduced by Boccaletti et al.[47], and its solution (for $k \geq 2m_{tri}$) is

$$P(k) = \frac{4m_{tri}(2m_{tri}+1)}{k(k+1)(k+2)} \sim k^{-3}, \quad (3)$$

which perfectly fits the data (see the dotted, dashed-dotted, and dashed black lines in Fig. 1a). A similar, exponential, degree distribution has also been reported in refs. [36,39,48].

Then, one can consider $N_e(k_\ell,t)$ as the number of edges participating in k_ℓ triangles at time t. Its rate equation is

$$\frac{dN_e(k_\ell,t)}{dt} = -m_{tri}\frac{N_e(k_\ell,t)}{N_e(t)} + m_{tri}\frac{N_e(k_\ell-1,t)}{N_e(t)} + 2m_{tri}\delta_{k_\ell,1} \pm \dots, \quad (4)$$

where $N_e(t) \simeq 2m_{tri}t$ and the unwritten terms account for the formation of triangles from two or more linked edges. For the distribution of the generalized degree k_ℓ, one has that $P(k_\ell) = N_e(k_\ell,t)/N_e(t)$, and the recursive equation is

$$P(k_\ell) = \frac{1}{3}P(k_\ell-1) + \frac{2}{3}\delta_{k_\ell,1}, \quad (5)$$

admitting the following solution:

$$P(k_\ell) = \frac{2}{3^{k_\ell}}, \quad k_\ell \geq 1. \quad (6)$$

The exponential function (6) is reported as a dashed line in Fig. 2a, and one can see that the fit with numerical simulations is rather good, especially for the case $m_{tri} = 1$.

Preferential model (NGF[37] with $d = 2$, $s = 1$). The analytic treatment of the preferential attachment case is far more complicated, as it implies the demonstration of a couple of theorems and the extensive use of a known lemma. We limit here ourselves to furnish the main results without reporting all the (sometimes cumbersome) formal mathematical steps, whereas the interested reader can find the full details within the "Methods" section. Calling again $N(k,t)$ the number of vertices with degree k at time t, its recurrence relation (see details in "Methods") can be written as

$$N(k,t+1) = N(k,t)\left(1 - \frac{k}{3t}\right) + N(k-2,t)\frac{k-2}{3t} + O\left(\frac{1}{t^2}\right) \quad (7)$$

for $k > 2m_{tri}$, and

$$N(2m_{tri}, t+1) = N(2m_{tri}, t)\left(1 - \frac{2m_{tri}}{3t}\right) + 1 + O\left(\frac{1}{t^2}\right) \quad (8)$$

for $k = 2m_{tri}$.

In order to obtain an expression for $P(k)$, one supposes that $\frac{N(k,t+1)}{t+1} = \frac{N(k,t)}{t}$ for large t. Therefore, one gets $P(2m_{tri}) = \frac{3}{3+2m_{tri}}$ and $P(k) = P(k-2)\frac{k/2+1}{k/2+1.5}$, which ultimately gives $P(k) =$

$\frac{3}{3+2m_{tri}}\prod_{l=m_{tri}+1}^{k/2}\frac{l-1}{l+1.5}$ or alternatively

$$P(k) = \frac{3}{3+2m_{tri}}\frac{\Gamma(k/2)\Gamma(m_{tri}+2.5)}{\Gamma(m_{tri})\Gamma(k/2+2.5)} \sim k^{-2.5}, \quad (9)$$

where Γ is here the gamma function. The power-law scaling predicted by Eq. (9) fits remarkably well the numerical data (see the dotted, dashed-dotted, and dashed black lines in Fig. 1c) and was also measured in refs. [36,39].

As for $P(k_\ell)$, one calls again $N_e(k_\ell,t)$ the number of edges participating in k_ℓ triangles at time t. The recurrence relation for $N_e(k_\ell,t)$ (see details in "Methods") is

$$N_e(k_\ell, t+1) = N_e(k_\ell,t)\left(1 - \frac{k_\ell}{3t}\right) + N_e(k_\ell-1,t)\frac{(k_\ell-1)}{3t} + O\left(\frac{1}{t^2}\right) \quad (10)$$

for $k_\ell > 1$, and

$$N_e(1, t+1) = N_e(1,t)\left(1 - \frac{m_{tri}}{3m_{tri}t}\right) + 2m_{tri} + O\left(\frac{1}{t^2}\right) \quad (11)$$

for $k_\ell = 1$. Imposing that $\frac{N_e(k_\ell,t+1)}{t+1} = \frac{N_e(k_\ell,t)}{t}$ for large t, one gets an equation for $P(k_\ell)$ which reads as $P(k_\ell) = P(k_\ell-1)\frac{k_\ell-1}{k_\ell+3}$, with $P(1) = \frac{3}{4}$. The solution is

$$P(k_\ell) = \frac{3}{4}\prod_{l=2}^{k_\ell}\frac{l-1}{l+3} = \frac{3}{4}\frac{4!}{k_\ell(k_\ell+1)(k_\ell+2)(k_\ell+3)} \sim k_\ell^{-4}. \quad (12)$$

The power-law function (12) is reported as a dashed line in Fig. 2b, and one can observe that the fit is, once again, extremely good.

Mixed model. Finally, our study can be extended and generalized to a mixed model, through which it is possible to effectively imprint any desired power-law scaling in the triangular structure of the network. Contrary to the previous two cases, our mixed model cannot be encompassed within the framework of the NGF[36]. In particular, we consider the case in which the probability that the node added at time t selects the specific link (i,j) to form its connections is

$$P_{ij}(t) = A\frac{2}{N_\ell(t-1)} + B\frac{3k_{ij}(t-1)}{2\sum_{ij}k_{ij}(t-1)}, \quad (13)$$

for some constants A and B. Here, B is nonnegative, $N_\ell(t) \sim 2m_{tri}t$ is the number of links at moment t and $\sum_{ij}k_{ij}(t-1) = 3m_{tri}t$ is the sum of the generalized degrees of all edges. Notice that $A = 0$ and $B = \frac{2}{3}$ ($A = \frac{1}{2}$ and $B = 0$) recovers the preferential (non-preferential) case discussed above.

From the constraint that the sum of all probabilities must be equal to 1, it follows that A and B must obey $2A + \frac{3}{2}B = 1$, so that $A = 1/2 - 3/4B$. Furthermore, $P_{ij}(t)$ must be nonnegative for all (i,j) (also those for which $k_{ij} = 1$), and this gives the following bounds for A and B: $-1 \leq A \leq \frac{1}{2}$ and $0 \leq B \leq 2$. Under these conditions, one can analytically demonstrate (see "Methods" for details) that for strictly positive B values, the resulting networks display scale-free distributions $P(k) \sim k^{-\gamma_\ell}$ and $P(k_\ell) \sim k_\ell^{-\gamma_\ell}$ with exponents given by

$$\gamma = 1 + \frac{1}{A+B} = 1 + \frac{4}{2+B}, \quad (14)$$

and

$$\gamma_\ell = 1 + \frac{2}{B}. \quad (15)$$

For $B = 0$, one has instead (see "Methods" for full details) $P(k_\ell) = \frac{2}{3^{k_\ell}}$. Equation (14) implies that γ values are between 2
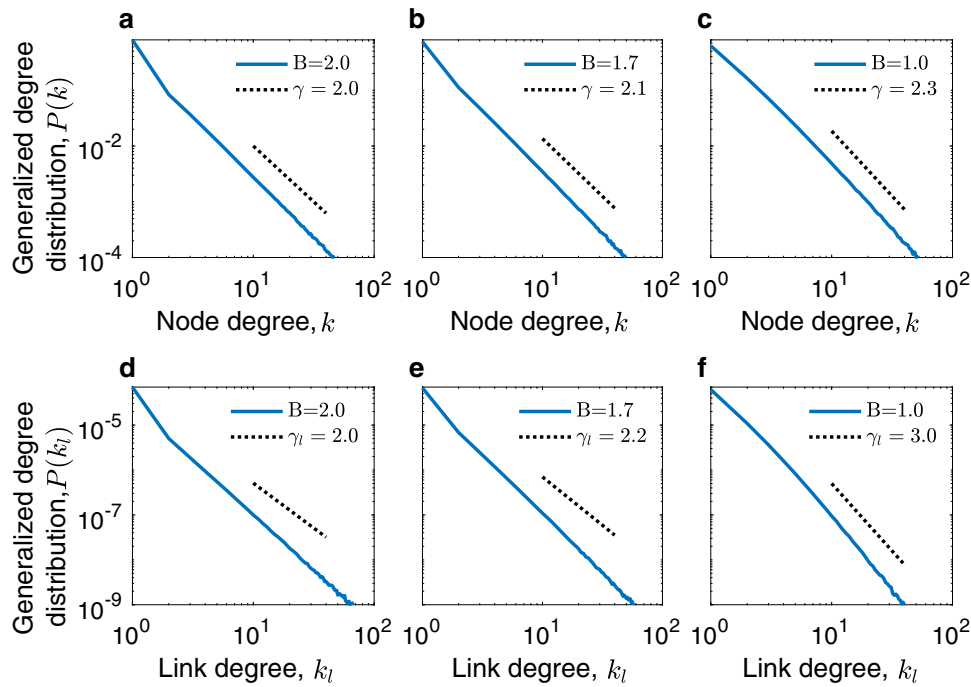
Fig. 3 Topological properties of the mixed model. Node degree probability distributions $P(k)$ (**a–c**) and link degree probability distributions $P(k_\ell)$ (**d–f**) obtained by growing networks of size $N = 50{,}000$ with the mixed model, at three different values of the parameter B introduced in Eq. (13). The data (blue lines) refer to an average over 100 different realizations of the growth process. Dotted lines report, for comparison, the scaling exponents γ and γ_ℓ predicted by Eqs. (14) and (15), respectively.

($B = 2$) and 3 ($B = 0$), whereas γ_ℓ is equal to 2 for $B = 2$, and tends to infinity as B tends to 0 according to Eq. (15). On its turn, this means that choosing B between 1 and 2, γ and γ_ℓ can be preselected *ad libitum* between 2 and 3, i.e., the imprinted structures of links and triangles feature well defined mean values of the degrees, but unbounded fluctuations as the system grows in size.

In Fig. 3, we report $P(k)$ and $P(k_\ell)$ for three distinct values of B. It is seen that the fit between analytic predictions and numerically generated data is always remarkably good. Moreover, the figure demonstrates that our method constitutes a highly general scheme by means of which one can construct, in a fully flexible way, ensembles of synthetic complexes displaying any desired statistical properties [from the condition of Fig. 3a, d featuring a super scale-freeness—where even the mean degrees diverge in the thermodynamic limit—to any milder condition which characterizes in fact many networks from the real world].

Extension to d uniform hypergraphs. The mixed model for SCs can be straightforwardly extended to grow uniform d hypergraphs. For simplicity, let us start by setting $N_0 = d + 1$ nodes in a ring, such that, for $d = 2$ we have a triangle (2-hyperlink), for $d = 3$ a square (3-hyperlink), for $d = 4$ a pentagon (4-hyperlink), and so on. At each time step, a structure of $d − 1$ new nodes forming an open ring with $d − 2$ edges (see the sketches in Fig. 4a–c) are added to the network, in order to close a d hyperedge with an existing link. Once again, the probability that a specific link (i,j) is chosen to close the hyperedge is

$$P_{ij}(t) = A \frac{2}{N_\ell(t-1)} + B \frac{3 k_{ij,d}(t-1)}{2 \sum_{ij} k_{ij,d}(t-1)}, \quad (16)$$

$N_\ell(t)$ being the total number of links at time t, and $k_{ij,d}(t)$ the generalized degree of the (i,j) link, that is, the total number

(at time t) of d uniform hyperedges which are incorporating the link ij. Figure 4 reports the node degree distributions $P(k)$ (d–f) and the generalized degree distribution $P(k_{\ell,d})$ (g–i), where $k_{\ell,d}$ now characterizes the number of triangles (for $d = 2$), squares (for $d = 3$), or pentagons (for $d = 4$) supported by each link $l = (i, j)$. The reported curves are obtained at different values of B, and the color code is visible in the legend of (a). It is possible to see that, in all cases, the grown structures display the same scaling behavior predicted by our mixed model for the case of SCs of order 2. The extension of our techniques to the growth of SC's of higher order will be reported elsewhere.

In summary, complex networks encode the basic architecture of social, biological, and technological networks, touching upon the most crucial challenges of modern science, from the spread of epidemics in social networks[20,49] to the resilience of our eco-systems and critical infrastructure[50].

In the context of pairwise interactions, the most natural measure of centrality is a node's individual degree, capturing its potential dynamic impact on the system[51]. The discovery that most real-world networks exhibit extreme levels of degree heterogeneity was disruptive—indicating that networks are highly centralized, with a potentially disproportionate role played by a small fraction of their components[42,43].

As we deepen our investigation into the interaction patterns of complex systems, it becomes increasingly clear, however, that higher order structures, beyond pairwise interactions, underlie much of the observed richness of real-world networks. Hence, we sought the fundamental rules that prescribe centrality, and govern its distribution, in an SC environment. In addition to this, the proposed models are expected to exhibit the small-world property, simultaneously featuring short average distances which increase proportionally to the logarithm of the network size (see Fig. 5) and large modularity and clustering coefficients—as new nodes enter the network closing triads. A natural measure is the
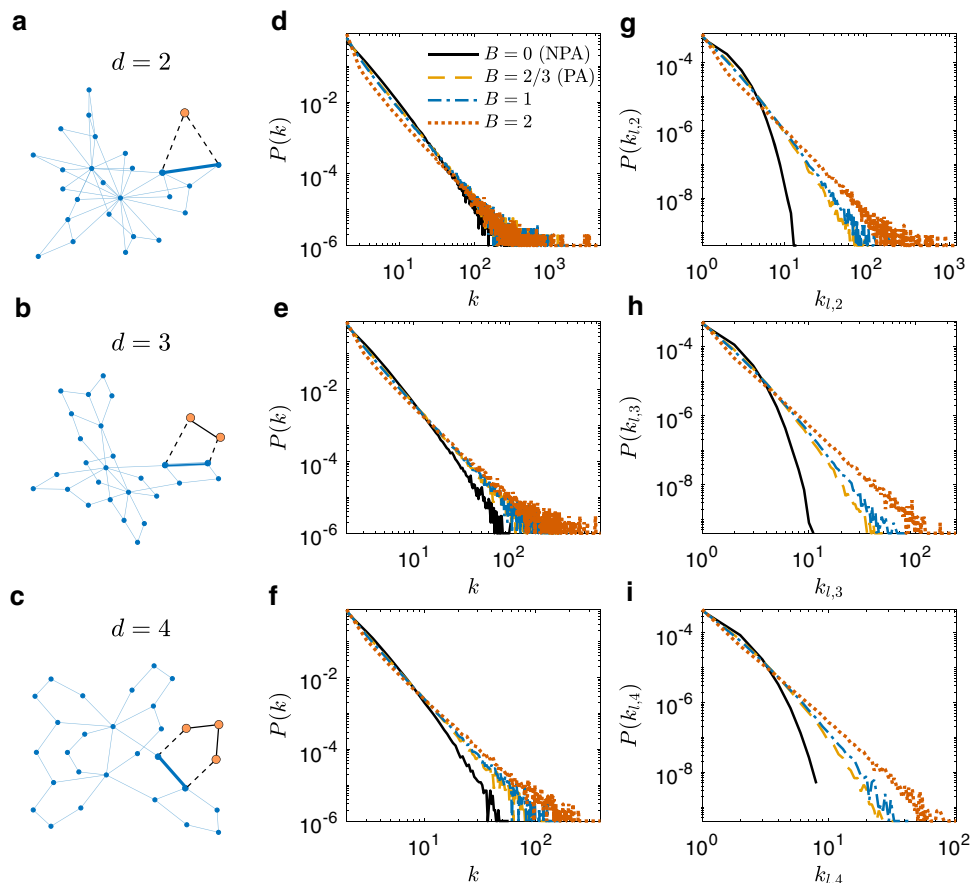
Fig. 4 Extending the mixed model to uniform *d* hypergraphs. A *d* hypergraph is a graph in which all hyperlinks contain $d + 1$ nodes, being *d* the order of the interaction. **a–c** Sketches of the processes through which uniform *d* hypergraphs are grown for $d = 2$ (a structure formed by 2-hyperlinks with triads of nodes forming triangles), $d = 3$ (a graph formed by 3-uniform hyperlinks, with all four nodes forming squares), and $d = 4$ (a uniform 4-hypergraph with all groups of five nodes forming pentagons). The procedure is such that at each time step, a chain of $d - 1$ new nodes (orange circles connected with solid lines) is added to the network to form a uniform *d* hyperedge by connecting the ends of the chain (dashed links) to an existing link (blue thick link). **d–f** The node degree *k* probability distributions $P(k)$ and **g–i** the probability distribution $P(k_{\ell,d})$ of the generalized link degree $k_{\ell,d}$, that is, the number of triangles ($d = 2$), squares ($d = 3$), and pentagons ($d = 4$) each link participates in, obtained by growing hypergraphs of size $N = 10,000$ at four different values of the parameter *B* introduced in Eq. (16) (reported in the legend of **d** where NPA and PA stand for nonpreferential and preferential attachment, respectively). Data refer to an average over 100 different realizations of the growth process.
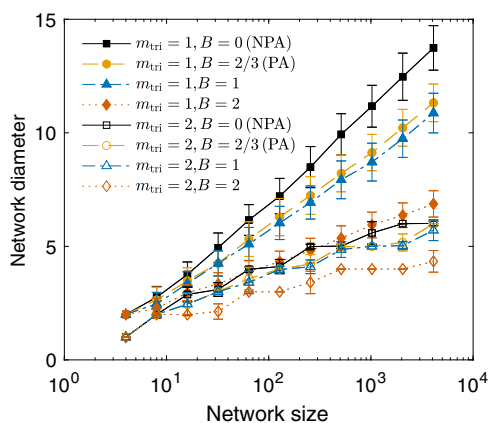

Fig. 5 Network diameter as a function of the network size *N*. Networks are grown with the mixed model for different values of the parameter *B* introduced in Eq. (13) and for the number of triangles added at each step of the growth process $m_{\mathrm{tri}} = 1$ (full symbols) and $m_{\mathrm{tri}} = 2$ (void symbols). Each point is an ensemble average over 100 network realizations and error bars represent the standard deviation. Notice the linear–log scale and that the network diameter is proportional to $\log N$. NPA and PA stand for nonpreferential and preferential attachment, respectively.

number of complexes, here triangles, that a link participates in. Indeed, an SC represents a potentially functional unit, such as a collaboration of a social team[17], or a trio of interacting biochemical agents[20,21]. A component that is part of many such complexes is, therefore, likely central in the functionality of the system.

Several growing network models have been introduced and studied, which help exposing the roots of SC heterogeneity, shedding light on the emergence of centrality beyond the degree distribution. As we seek to understand the behavior of complex systems, their resilience and dynamic functionality, we hope that our comprehension into the microscopic processes of their formation can provide meaningful macroscopic insights.

Methods

In this final section, we furnish all the details of the analytical results regarding our models. The starting point is that, as the synthesized networks have transitivity coefficient $T = 1$ (i.e., no links exist which do not form part of at least a triangle), the degree of each vertex is the number of triangles containing that vertex multiplied by 2, and one has that $N_{\mathrm{v}}(k_{\mathrm{v}}, t) = N(2k, t)$, where $N_{\mathrm{v}}(k_{\mathrm{v}}, t)$ is the number of nodes participating in k_{v} triangles and $N(2k, t)$ is the number of nodes having degree $2k$ at time *t*. As a consequence one has that $P(k_{\mathrm{v}}) = P(k)$, and is therefore entitled to concentrate on either one of such distributions, depending on which one finds the simpler analytical treatment.

The nonpreferential attachment case. Let $N_v(k, k_v, t)$ be the number of nodes with k neighbors participating in k_v triangles at time t. The rate equation is

$$\frac{dN_v(k, k_v, t)}{dt} = -\frac{2m_{tri}}{\sum_k kN(k, t)} kN_v(k, k_v, t)$$
$$+ \frac{2m_{tri}}{\sum_k kN(k, t)} (k-1)N_v(k-1, k_v-1, t) \quad (17)$$
$$+ \delta_{k, 2m_{tri}} \delta_{k_v, m_{tri}}$$
$$\pm \dots ,$$

where the unwritten terms account for the formation of triangles from two or more linked edges. In the sequel, we assume that the chosen edges are not linked (none of the nodes of a selected edge is linked to any of the nodes of another selected edge, which is always the case for $m_{tri} = 1$). Hence, the number of edges and triangles related to a given node increases one by one. Only the new nodes entering the system have the number of edges ($2m_{tri}$) double than the number of triangles in which they are participating (m_{tri}). This way, for k big enough, one has

$$kN_v(k, k_v, t) \simeq k_v N_v(k, k_v, t). \quad (18)$$

After summing over all values of k, one obtains an approximate equation for $N_v(k_v, t)$, the number of nodes participating in k_v triangles

$$\frac{dN_v(k_v, t)}{dt} \simeq -\frac{2m_{tri}}{\sum_k kN(k, t)} k_v N_v(k_v, t)$$
$$+ \frac{2m_{tri}}{\sum_k kN(k, t)} (k_v - 1)N_v(k_v - 1, t) \quad (19)$$
$$+ \delta_{k_v, m_{tri}}.$$

One then can proceed as in the case of the degree distribution $P(k)$ (see the main text), and seek a solution of the form $N_v(k_v, t) = tP(k_v)$:

$$P(k_v) = \frac{k_v - 1}{k_v + 2} P(k_v - 1) + \frac{2}{2 + k_v} \delta_{k_v, m_{tri}}. \quad (20)$$

The solution for $k_v \geq m_{tri}$ is

$$P(k_v) = \frac{2m_{tri}(m_{tri} + 1)}{k_v(k_v + 1)(k_v + 2)} \sim k_v^{-3}. \quad (21)$$

Therefore, one has also that $P(k) \sim k^{-3}$ which coincides with Eq. (3) in the main text.

The preferential attachment case. In order to obtain $P(k)$ and $P(k_\ell)$, one needs here to make use of the following lemma reported by Chung and Lu[52]:

Lemma 1 *Suppose that a sequence $\{a_t\}$ satisfies a recurrence relation*

$$a_{t+1} = \left(1 - \frac{b_t}{t + t_1}\right) a_t + c_t, \quad (22)$$

where t_0 and t_1 are arbitrary, positive, fixed, values. Furthermore, suppose that $\lim_{t\to\infty} b_t = b > 0$ and $\lim_{t\to\infty} c_t = c > 0$.
Then, $\lim_{t\to\infty} \frac{a_t}{t}$ exists and one has

$$\lim_{t\to\infty} \frac{a_t}{t} = \frac{c}{1 + b}. \quad (23)$$

For instance (and indicating by $\mathbb{E}(\cdot)$ the expectation value), $P(k)$ for $k = 2m_{tri}$ can be obtained by setting $b_t = b = \frac{2m_{tri}}{3}$ and $c_t = c = 1$. Application of Lemma 1 ensures that the limit $\lim_{t\to\infty} \frac{\mathbb{E}(N(2m_{tri}, t))}{t}$ exists and is equal to $\frac{3}{3 + 2m_{tri}}$, leading to $P(1) = \frac{3}{3 + 2m_{tri}}$.

On the other hand, $P(k)$ for $k > 2m_{tri}$ is obtained assuming that $P(k-2)$ exists, and applying Lemma 1 again with $b_t = b = \frac{k}{3}$ and $c_t = \mathbb{E}(N(k-2, t))\frac{k-2}{3t}$ (i.e., taking $c = P(k-2)\frac{k-2}{3}$). Such a choice, indeed, entitles one to write a recurrence relation, and to obtain an explicit formula for $P(k)$

$$P(k) = \frac{3}{3 + 2m_{tri}} \frac{\Gamma(k/2)\Gamma(m_{tri} + 2.5)}{\Gamma(m_{tri})\Gamma(k/2 + 2.5)} \sim k^{-2.5}, \quad (24)$$

which coincides with Eq. (9) of the main text. Here, Γ is the gamma function.

Furthermore, one can demonstrate that $P(k)$ is sharp, by the use of the following theorem:

Theorem 1 *For any fixed $\varepsilon > 0$ and $\delta > 0$ and for any large enough t, the difference between the number of vertices with degree k at time t and $P(k)t$ is smaller than εt with probability larger than $1 - \delta$.*

The theorem can be proved by considering the martingale $X_l = \mathbb{E}(N(k, t)|\mathcal{F}_l)$ (where $\mathcal{F}_l$ is the σ-algebra generated by the probability space at time l). It is rather easy to show that $|X_{l+1} - X_l|$ is bounded by 4, and therefore the theorem follows from the Azuma–Hoeffding inequality[53,54].

When trying to obtain $P(k_\ell)$, one can encounter a problem in the case in which the triangles added at step t have one or more common edges, so that the number of different edges added at time t might not be equal to $2m_{tri}$.

Denoting by $e(t, k_\ell)$ the number of added edges of degree k_ℓ at time t, the more accurate recurrence relation for $k_\ell > 1$ is the following:

$$\mathbb{E}(N_e(k_\ell, t+1)) = \mathbb{E}(N_e(k_\ell, t))\left(1 - \frac{k_\ell}{3t}\right)$$
$$+ \mathbb{E}(N_e(k_\ell - 1, t))\frac{(k_\ell - 1)}{3t} + O\left(\frac{1}{t^2}\right) + \mathbb{E}(e(t, k_\ell)), \quad (25)$$

and for $k_\ell = 1$ one has

$$\mathbb{E}(N_e(1, t+1)) = \mathbb{E}(N_e(1, t))\left(1 - \frac{1}{3t}\right) + O\left(\frac{1}{t^2}\right) + \mathbb{E}(e(t, 1)). \quad (26)$$

Using Theorem 1 (and some nontrivial math, of which we omit the technical details), one can prove that $\lim_{t\to\infty} e(t, 1) = 2m_{tri}$ and $\lim_{t\to\infty} e(t, k_\ell) = 0$ for $k_\ell > 1$, so that one has $\lim_{t\to\infty} \mathbb{E}(e(t, 1)) = 2m_{tri}$ and $\lim_{t\to\infty} \mathbb{E}(e(t, k_\ell)) = 0$ for $k_\ell > 1$.

Furthermore, one can prove another theorem:

Theorem 2 *The number of different edges divided by t converges always to $2m_{tri}$.*

It follows that $P(k_\ell) = \lim_{t\to\infty} \frac{\mathbb{E}(N_e(k_\ell, t))}{2m_{tri} t}$.

Finally, one can apply Lemma 1 with $b_t = b = \frac{1}{3}$ and $c = \lim 1 + O(\frac{1}{t^2}) = 1$ to get that $P(1) = \lim_{t\to\infty} \frac{\mathbb{E}(N_e(1, t))}{2m_{tri} t}$ exists and is equal to $\frac{3}{4}$. For $k_\ell > 1$, one assumes that $P(k_\ell - 1) = \lim_{t\to\infty} \frac{\mathbb{E}(N_e(k_\ell - 1, t))}{2m_{tri} t}$ exists, and applies again Lemma 1 with $b_t = b = \frac{k_\ell}{3}$ and $c = \lim c_t = P(k_\ell - 1)\frac{(k_\ell - 1)}{3t}$. Hence, $P(k_\ell) = \lim_{t\to\infty} \frac{\mathbb{E}(N_e(k_\ell, t))}{2m_{tri} t}$ exists and is equal to $P(k_\ell - 1)\frac{k_\ell - 1}{k_\ell + 3}$. From such a recurrence relation, one finally obtains the explicit formula for $P(k_\ell)$

$$P(k_\ell) = \frac{18}{k_\ell(k_\ell + 1)(k_\ell + 2)(k_\ell + 3)} \sim k_\ell^{-4}, \quad (27)$$

which is identical to Eq. (4) of the main text.

Finally, since the degree of the vertex is the number of triangles containing this vertex multiplied by 2, one has $N_v(k_v, t) = N(2k, t)$ and as a consequence one obtains

$$P(k_v) = \frac{3}{3 + 2m_{tri}} \frac{\Gamma(k_v)\Gamma(m_{tri} + 2.5)}{\Gamma(m_{tri})\Gamma(k_v + 2.5)} \sim k_v^{-2.5}. \quad (28)$$

The mixed case. The recurrence relation for $N_v(k_v, t)$ (the number of vertices participating in k_v triangles at time t) is

$$\mathbb{E}(N_v(k_v, t+1)) = \mathbb{E}(N_v(k_v, t))\left(1 - A\frac{m_{tri}}{t} - (A + B)\frac{k_v}{t}\right)$$
$$+ \mathbb{E}(N_v(k_v - 1, t))\left(A\frac{m_{tri}}{t} + (A + B)\frac{(k_v - 1)}{t}\right) + O\left(\frac{1}{t^2}\right). \quad (29)$$

For $t > 0$ and $k_v = m_{tri}$ one has

$$\mathbb{E}(N_{t+1}^v(m_{tri})) = \mathbb{E}(N_v(m_{tri}, t))\left(1 - A\frac{m_{tri}}{t} - (A + B)\frac{m_{tri}}{t}\right) + 1. \quad (30)$$

Furthermore, the use of Lemma 1 gives

$$P(m_{tri}) = \frac{1}{1 + 2Am_{tri} + Bm_{tri}} \quad (31)$$

and

$$P(k_v) = P(k_v - 1)\frac{Am_{tri} + (A + B)(k_v - 1)}{1 + Am_{tri} + (A + B)k_v}. \quad (32)$$

Therefore, one has

$$P(k_v) = \frac{1}{1 + 2Am_{tri} + Bm_{tri}} \prod_{l=m_{tri}+1}^{k_v} \frac{Am_{tri} + (A + B)(l - 1)}{1 + Am_{tri} + (A + B)l}. \quad (33)$$

or

$$P(k_v) = \frac{1}{1 + 2Am_{tri} + Bm_{tri}} \cdot \prod_{l=m_{tri}+1}^{k_v} \frac{\frac{Am_{tri}}{A+B} + l - 1}{\frac{1+Am_{tri}}{A+B} + l}$$
$$= \frac{1}{1 + 2Am_{tri} + Bm_{tri}} \cdot \frac{\Gamma(k_v + \frac{Am_{tri}}{A+B})\Gamma(m_{tri} + \frac{1+Am_{tri}}{A+B} + 1)}{\Gamma(m_{tri} + \frac{Am_{tri}}{A+B})\Gamma(k_v + \frac{1+Am_{tri}}{A+B} + 1)} \sim k_v^{-\gamma_v}, \quad (34)$$

where $\gamma_v = 1 + \frac{1}{A+B}$.

In addition, one can consider $N_\ell(k_\ell, t)$, the number of links with degree k_ℓ at moment t. Its recurrence equation is

$$\mathbb{E}(N_\ell(k_\ell, t+1)) = \mathbb{E}(N_\ell(k_\ell, t))\left(1 - A\frac{1}{t} - B\frac{k_\ell}{2t}\right)$$
$$+ \mathbb{E}(N_\ell(k_\ell - 1, t))\left(A\frac{1}{t} + B\frac{k_\ell - 1}{2t}\right) + O\left(\frac{1}{t^2}\right). \quad (35)$$

For $k_\ell = 1$ one has

$$\mathbb{E}(N_\ell(1, t+1)) = \mathbb{E}(N_\ell(1,t))\left(1 - A\frac{1}{t} - B\frac{1}{2t}\right) + 2m_{\text{tri}} + O\left(\frac{1}{t^2}\right). \quad (36)$$

One can apply again Lemma 1, and obtain

$$P(1) = \frac{1}{1 + A + \frac{1}{2}B}, \quad P(k_\ell) = P(k_\ell - 1)\frac{2A + B(k_\ell - 1)}{2 + 2A + Bk_\ell}. \quad (37)$$

Therefore, one gets

$$P(k_\ell) = \frac{1}{1 + A + \frac{1}{2}B}\prod_{l=2}^{k_\ell}\frac{2A + B(l-1)}{2 + 2A + Bl}. \quad (38)$$

Let us consider separately the cases $B = 0$ (the fully nonpreferential case) and $B > 0$.

For $B = 0$, one has

$$P(k_\ell) = \frac{1}{1+A}\prod_{l=2}^{k_\ell}\frac{2A}{2 + 2A}. \quad (39)$$

Since $A = \frac{1}{2}$, one obtains

$$P(k_\ell) = \frac{2}{3}\prod_{l=2}^{k_\ell}\frac{1}{3} = \frac{2}{3}\frac{1}{3^{k_\ell-1}}, \quad (40)$$

which fully coincide with the exponential scaling derived in the main text for the nonpreferential case.

On the other hand, when instead $B > 0$, one has

$$P(k_\ell) = \frac{1}{1 + A + \frac{1}{2}B}\prod_{l=2}^{k_\ell}\frac{\frac{2A}{B} + l - 1}{\frac{2}{B} + \frac{2A}{B} + l} =$$
$$= \frac{1}{1 + A + \frac{1}{2}B}\frac{\Gamma(\frac{2+2A}{B} + 2)\Gamma(\frac{2A}{B} + k_\ell)}{\Gamma(\frac{2A}{B} + 1)\Gamma(\frac{2+2A}{B} + k_\ell + 1)} \sim k_\ell^{-\gamma_\ell}, \quad (41)$$

where $\gamma_\ell = 1 + \frac{2}{B}$.

Data availability

Data sharing not applicable to this article as no datasets were generated or analysed during the current study.

Code availability

All the codes used in this study can be found in Supplementary Software file.

References

1. Dorogovtsev, S. N. & Mendes, J. F. F. Evolution of networks. *Adv. Phys.* **51**, 1079–1187 (2002).
2. Newman, M. The structure and function of complex networks. *SIAM Rev.* **45**, 167–256 (2003).
3. Boccaletti, S., Latora, V., Moreno, Y., Chavez, M. & Hwang, D.-U. Complex networks: structure and dynamics. *Phys. Rep.* **424**, 175–308 (2006).
4. Estrada, E. *The Structure of Complex Networks: Theory and Applications* (Oxford University Press, Inc., 2011).
5. Boccaletti, S. et al. The structure and dynamics of multilayer networks. *Phys. Rep.* **544**, 1–122 (2014).
6. Latora, V., Nicosia, V. & Russo, G. *Complex Networks: Principles, Methods and Applications* (Cambridge University Press, 2017).
7. Battiston, F. et al. Networks beyond pairwise interactions: structure and dynamics. *Phys. Rep.* **874**, 1–92 (2020).
8. Levine, J., Bascompte, J., Adler, P. & Allesina, S. Beyond pairwise mechanisms of species coexistence in complex communities. *Nature* **546**, 56–64 (2017).
9. Petri, G., Scolamiero, M., Donato, I. & Vaccarino, F. Topological strata of weighted complex networks. *PLoS ONE* **8**, e66506 (2013).
10. Sizemore, A., Giusti, C. & Bassett, D. S. Classification of weighted networks through mesoscale homological features. *J. Complex Netw.* **5**, 245–273 (2016).
11. Lee, H., Kang, H., Chung, M. K., Kim, B.-N. & Lee, D. S. Persistent brain network homology from the perspective of dendrogram. *IEEE Trans. Med. Imaging* **31**, 2267–2277 (2012).
12. Petri, G. et al. Homological scaffolds of brain functional networks. *J. R. Soc. Interface* **11**, 20140873 (2014).
13. Lord, L.-D. et al. Insights into brain architectures from the homological scaffolds of functional connectivity networks. *Front. Syst. Neurosci.* **10**, 85 (2016).
14. Sizemore, A. E. et al. Cliques and cavities in the human connectome. *J. Comput. Neurosci.* **44**, 115–145 (2018).
15. Estrada, E. & Ross, G. J. Centralities in simplicial complexes. applications to protein interaction networks. *J. Theor. Biol.* **438**, 46–60 (2018).
16. Sizemore, A. E., Karuza, E. A., Giusti, C. & Bassett, D. S. Knowledge gaps in the early growth of semantic feature networks. *Nat. Hum. Behav.* **2**, 682–692 (2018).
17. Patania, A., Petri, G. & Vaccarino, F. The shape of collaborations. *EPJ Data Sci.* **6**, 18 (2017).
18. Freeman, L. C. Q-analysis and the structure of friendship networks. *Int. J. Man-Mach. Stud.* **12**, 367–378 (1980).
19. Andjelković, M., Tadić, B., Maletić, S. & Rajković, M. Hierarchical sequencing of online social graphs. *Phys. A: Stat. Mech. Appl.* **436**, 582–595 (2015).
20. Harush, U. & Barzel, B. Dynamic patterns of information flow in complex networks. *Nat. Commun.* **8**, 2181 (2017).
21. Kuzmin, E. et al. Systematic analysis of complex genetic interactions. *Science* **360**, 1–9 (2018).
22. Centola, D. The spread of behavior in an online social network experiment. *Science* **329**, 1194–1197 (2010).
23. Aleksandrov, P. *Combinatorial Topology*, Vol. 1 (Dover Publications, Inc, 1998).
24. Salnikov, V., Cassese, D. & Lambiotte, R. Simplicial complexes and complex systems. *Eur. J. Phys.* **40**, 014001 (2018).
25. Sizemore, A. E., Phillips-Cremins, J. E., Ghrist, R. & Bassett, D. S. The importance of the whole: topological data analysis for the network neuroscientist. *Netw. Neurosci.* **3**, 656–673 (2019).
26. Kahle, M. *Topology of Random Simplicial Complexes: a Survey. AMS Contemp. Math* **620**, 201–222 (2014).
27. Costa, A. & Farber, M. Random simplicial complexes. In *Configuration Spaces: Geometry, Topology and Representation Theory* (eds Callegaro, F. et al.), 129–153 (Springer International Publishing, 2016).
28. Iacopini, I., Petri, G., Barrat, A. & Latora, V. Simplicial models of social contagion. *Nat. Commun.* **10**, 2485 (2019).
29. Courtney, O. T. & Bianconi, G. Generalized network structures: the configuration model and the canonical ensemble of simplicial complexes. *Phys. Rev. E* **93**, 062311 (2016).
30. Young, J.-G., Petri, G., Vaccarino, F. & Patania, A. Construction of and efficient sampling from the simplicial configuration model. *Phys. Rev. E* **96**, 032312 (2017).
31. Bianconi, G., Kryven, I. & Ziff, R. M. Percolation on branching simplicial and cell complexes, and its relation to interdependent percolation. *Phys. Rev. E* **100**, 062311 (2019).
32. Petri, G. & Barrat, A. Simplicial activity driven model. *Phys. Rev. Lett.* **121**, 228301 (2018).
33. Boguñá, M. et al. Network geometry. *Nat. Rev. Phys.* **3**, 114–135 (2021).
34. Bianconi, G. & Rahmede, C. Complex quantum network manifolds in dimension d > 2 are scale-free. *Sci. Rep.* **5**, 13979 (2015).
35. Bianconi, G., Rahmede, C. & Wu, Z. Complex quantum network geometries: evolution and phase transitions. *Phys. Rev. E* **92**, 022815 (2015).
36. Bianconi, G. & Rahmede, C. Network geometry with flavor: from complexity to quantum geometry. *Phys. Rev. E* **93**, 032315 (2016).
37. Courtney, O. T. & Bianconi, G. Weighted growing simplicial complexes. *Phys. Rev. E* **95**, 062301 (2017).
38. Mulder, D. & Bianconi, G. Network geometry and complexity. *J. Stat. Phys.* **173**, 783–805 (2018).
39. Bianconi, G. & Rahmede, C. Emergent hyperbolic network geometry. *Sci. Rep.* **7**, 41974 (2017).
40. Šuvakov, M., Andjelković, M. & Tadić, B. Hidden geometries in networks arising from cooperative self-assembly. *Sci. Rep.* **8**, 1987 (2018).
41. Fountoulakis, N., Iyer, T., Mailler, C. & Sulzbach, H. Dynamical models for random simplicial complexes. *arXiv*: 1910.12715 (2019).
42. Barabási, A.-L. & Albert, R. Emergence of scaling in random networks. *Science* **286**, 509–512 (1999).
43. Albert, R. & Barabási, A.-L. Statistical mechanics of complex networks. *Rev. Mod. Phys.* **74**, 47–97 (2002).
44. Gambuzza, L. V. et al. The master stability function for synchronization in simplicial complexes. *Nat. Commun.* **12**, 1255 (2021).
45. Barrat, A. & Weigt, M. On the properties of small-world network models. *Eur. Phys. J. B* **13**, 547–560 (2000).
46. Newman, M. E. J. Scientific collaboration networks. i. network construction and fundamental results. *Phys. Rev. E* **64**, 016131 (2001).
47. Boccaletti, S., Hwang, D.-U. & Latora, V. Growing hierarchical scale-free networks by means of nonhierarchical processes. *Int. J. Bifurc. Chaos* **17**, 2447–2452 (2007).
48. Dorogovtsev, S., Mendes, J. & Samukhin, A. Size-dependent degree distribution of a scale-free growing network. *Phys. Rev. E* **63**, 062101 (2001).

49. Vespignani, A. Modelling dynamical processes in complex socio-technical systems. *Nat. Phys.* **8**, 32–39 (2012).
50. Hacohen, A., Cohen, R., Efroni, S., Barzel, B. & Bachelet, I. Digitizable therapeutics for decentralized mitigation of global pandemics. *Sci. Rep.* **9**, 14345 (2019).
51. Barzel, B. & Barabási, A.-L. Universality in network dynamics. *Nat. Phys.* **9**, 673–681 (2013).
52. Chung, F. & Lu, L. *Complex Graphs and Networks* (American Mathematical Society, 2004).
53. Hoeffding, W. Probability inequalities for sums of bounded random variables. *J. Am. Stat. Assoc.* **58**, 13–30 (1963).
54. Azuma, K. Weighted sums of certain dependent random variables. *Tohoku Math. J.* **19**, 357–367 (1967).
55. Bonacich, P. Factoring and weighting approaches to status scores and clique identification. *J. Math. Sociol.* **2**, 113–120 (1972).

Acknowledgements

Author contributions

K.K., I.S.-N., N.K., S.B., and B.B. equally contributed to the manuscript. S.B., B.B., I.S.-N., and D.M. conceived the research study. K.K. and N.K. carried out all the analytical treatments. I.S.-N. wrote the codes and conducted the simulations and the analysis of the data. A.M.R. proved Theorem 2. K.K., I.S.-N., N.K., A.D., D.M., A.M.R., K.A.-B., B.B., and S.B. discussed the results, drew the conclusions, and jointly wrote the manuscript.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s42005-021-00538-y.

Correspondence and requests for materials should be addressed to I.S.-N.

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

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