

Scaling laws for properties of random graphs that grow via successive combination

PETER GRINDROD[†] 

Mathematical Institute, University of Oxford, Oxford OX2 6GG, UK

[†]Corresponding author. Email: grindrod@maths.ox.ac.uk

[Received on 11 March 2022; editorial decision on 28 April 2022; accepted on 16 May 2022]

We consider undirected graphs that grow through the successive combination of component sub-graphs. For any *well-behaved* functions defined for such graphs, taking values in a Banach space, we show that there must exist a scaling law applicable when successive copies of the same component graph are combined. Crucially, we extend the approach introduced in previous work to the successive combination of component random sub-graphs. We illustrate this by generalizing the preferential attachment operation for the combination of stochastic block models. We discuss a further wide range of random graph combination operators to which this theory now applies, indicating the ubiquity of growth scaling laws (and asymptotic decay scaling laws) within applications, where the modules are quite distinct, yet may be considered as instances drawn from the same random graph. This is a type of statistically self-similar growth process, as opposed to a deterministic growth process incorporating exact copies of the same motif, and it represents a natural, partially random, growth processes for graphs observed in the analysis of social and technology contexts.

Keywords: random graphs; successive growth; scaling laws

1. Introduction

We focus on graphs which grow by successively combining various smaller graphs. In concept, this subsumes the reverse of the popular community detection process of partitioning larger networks into small densely connected sub-communities, called modules; while any edges connecting up the distinct modules are relatively sparse. This decomposition is normally achieved by maximizing some convenient property (usually called ‘modularity’) over the set of all possible partitions [1, 2].

Instead, here we will successively combine graphs, potentially creating composite high-modularity graphs, and consider the behaviour of various functions that may be defined over them. While including a variety of graph combination mechanisms, we will consider only well-behaved functions that are insensitive to a small number of extra edges to be added-in to the combination. In many applications to social, organizational, biological or urban networks, we observe scaling laws that govern the behaviour of functional properties of graphs, or properties of dynamical systems that are defined over the graphs, as the underlying graphs grow [3–6].

In that context, it is important to differentiate between two distinct types of scaling law behaviour, which should not be conflated (yet may be related). There is the scaling behaviour exhibited within a single large graph: by the decay of its degree distribution, for example, in scale-free graphs, which are often formed by mechanisms such as preferential attachment [7, 8]. There is also the scaling law exhibited by the evolution of graph properties as the graphs themselves grow, which is observed when we examine and compare such objects of very different sizes (and thus different stages of growth), such as cities, organizations and organisms [3–6]. We are primarily focused on the latter here, though

we will allow graphs to grow by the successive combination of components, including via preferential attachment-type mechanisms.

In previous work [9], these functions of graphs were assumed to be real valued and the graph combination mechanism was limited to the disjoint union of components (together with the possible addition of sparse connecting edges). Importantly, that analysis [9] also explained why power-law scaling is consequently ubiquitous, as observed [3], in both asymptotic growth and asymptotic decay to some constant. In this article, in Section 2, we shall show that for *reasonable* functions of graphs taking values in a Banach space, that satisfy a type of Lipschitz condition, and for a very wide variety of graph–graph combination mechanisms, there will exist scaling laws. Most typically these will describe power-law growth, or inverse power-law asymptotic decay. As in [9], the analysis Section 2 is limited initially to the case where we combine multiple isomorphisms of some single graph.

In Section 3, we extend those considerations to deal with the situation where we combine multiple copies of a single random graph. In that case, within any realization, each addition might be distinct from all others, yet all conforming to the same random (sub)graph model. In one step, this creates large graphs whose modules are themselves generated as random graphs, and which necessarily must possess scaling laws (for growth and asymptotic decay).

A straightforward example of this process is a component random graph that conforms to a stochastic block model [10], with relatively edge-dense sub-blocks being added successively to the whole. For example in Fig. 1, we show the result of a stochastic block model with $q = 5$ blocks each of 10 vertices, where the probability of an edge between vertices within the same block is 0.5, whilst the probability of an edge between two vertices within different blocks is 0.005. We also show the eigenvectors of the associated graph (combinatorial) Laplacian together with the corresponding eigenvectors: clearly, the first few eigenvectors indicate the component five sub-blocks. This model is equivalent to a disjoint union of q separate stochastic blocks, with each being the Gilbert random graph [11], $\mathcal{G}(K = 10, p = 0.5)$ on K vertices with each possible edge present with independent probability p , amended to include a relatively small number of edges connecting pairs of vertices within distinct blocks (in this case there are 1000 possible inter-block edges, each present independently with probability 0.005, so we expect to observe around five such edges, and we actually observe six in Fig. 1).

Suppose that we increase the number of blocks, q . Then, the average observed vertex degree, z , grows linearly, ($z = 4.5 + 0.05(q - 1)$), due to the increasing number of possible inter-block edges, and the total number of observed edges grows quadratically: both functionals exhibit power-law growth.

A more intricate function of interest, for such an evolving stochastic block model growing as q increases, is explored by choosing a random starting vertex and measuring the fraction, ξ , of all vertices that may be reached within relatively short walks of length 3 or less. Clearly $\xi \leq 1$.

In Fig. 2, we consider such graphs generated by the stochastic block model introduced above, as q , increases. In all cases each block is the random Gilbert graph [11], $\mathcal{G}(K = 10, p = 0.5)$, while all possible inter-block edges are present with independent probability of 0.005. For each graph, we selected an initial vertex at random and calculated the fraction, ξ , of the full vertex set, V , that was reachable by walks of length 3 or less. In fact, for each value of q , and the consequent observed graph, in Fig. 2, we show the result for 20 alternative random starting vertices, also the mean reachable fraction of the vertex set. Even though the walks are relatively short, the small-world effect of the inter-block edges, and the linearly increasing average vertex degree mean that a large fraction of the vertex set may be reached. ξ is capped, of course, and we observe inverse power law convergence of its expectation to unity, $1 - \langle \xi \rangle \sim q^{-5/2}$.

This example is rather simplistic, deploying a disjoint union combination of blocks, together with the imposition of some extra, more rare, inter-block edges. A more realistic model might include a more

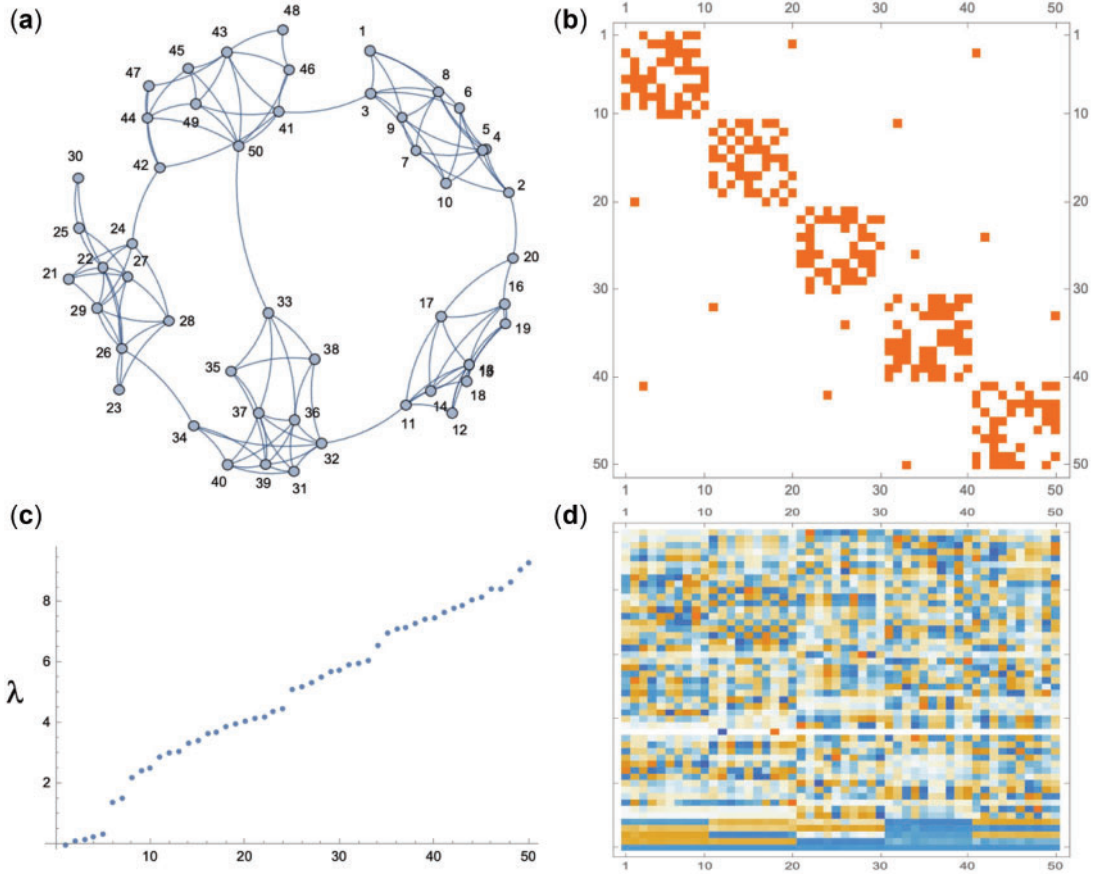


FIG. 1. A single realization of an undirected stochastic block random graph with $q = 5$ blocks each of 10 vertices where the probability on an edge e_{ij} is 0.5 if i and j belong to the same block and 0.005 otherwise: (a) the graph; (b) the adjacency matrix; (c) the ordered eigenvalues of the combinatorial Laplacian; (d) the associated eigenvectors.

sophisticated combination mechanism that would allow for both the additional blocks to be incorporated and also the addition of associated inter-block attachments. To achieve this, the binary operation employed to combine any two random graphs may itself include some biased random elements. That would still result in a new, combined, random graph. In Section 4, we introduce a such method to combine successive Gilbert random graphs, $\mathcal{G}(K, p)$ (or any other component random graphs) via preferential attachment. This results in a ‘Preferential Attachment Stochastic Block Model’ (PASBM) graph, generalizing the usual preferential attachment model, which is introduced in Section 4. This combination operation is more flexible and practical than the above example. It extends previous work on the growth of graphs via preferential attachment (since if $K = 1$ blocks are just single vertices which are preferentially attached) [7, 8].

In Section 5, we interpret Twitter reciprocated mentions networks for British cities, illustrating that they are composed of attached sub-graphs (modules) each of which is structurally similar (at least as represented to their associated spectra).

In Section 6, we set out a rather wide range of random graph combination (growth) operators to which the theory presented in Section 3 applies, including both additive (attachment) and hierarchical processes,

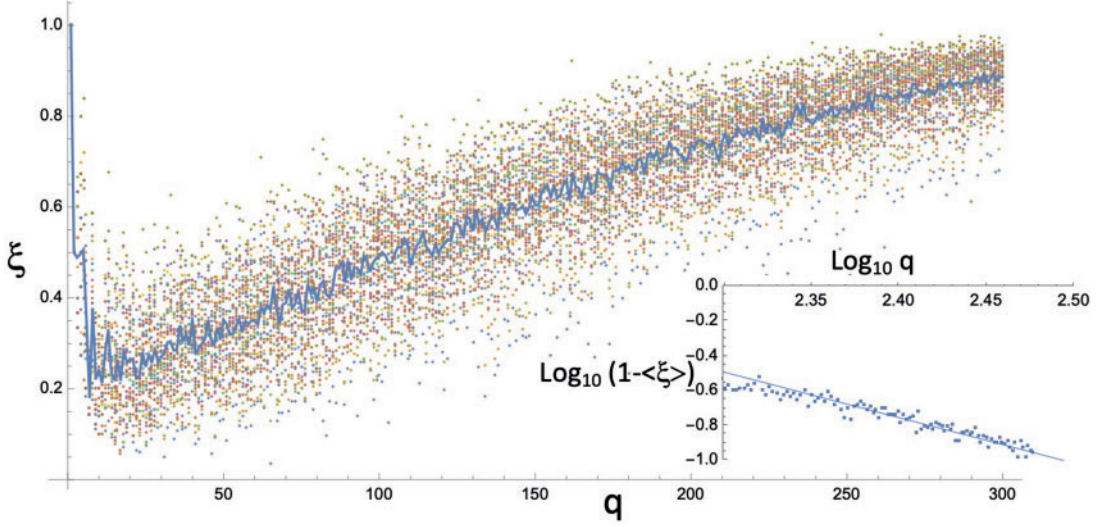


FIG. 2. For random stochastic block model graphs as q increases (see text), we show the fraction, ξ , of the vertex set, V , that is, reachable from a single random starting vertex by walks of length 3 or less versus q . For each value of q (and a consequent realized graph), we plot 20 results for different random starting vertices, and the expected reachable fraction, $\langle \xi \rangle$, of the vertex set (solid curve). Note $|V| = 10q$. We observe $1 - \langle \xi \rangle \sim q^{-5/2}$ (see inset).

indicating the ubiquity of growth scaling laws (and asymptotic decay scaling laws) within applications where these models are suitable.

2. Scaling laws for successively combined graphs

Let \mathcal{S} denote the set of all undirected graphs on a finite number of vertices. For each graph, $S \in \mathcal{S}$, we let $V(S)$ denote its vertex set and $E(S)$ denote its edge set. Let \diamond denote a commutative binary operation over \mathcal{S} .

For example, \diamond might denote the disjoint union of two graphs, \sqcup , as in [9].

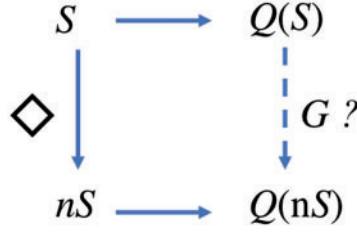
Alternatively, we might adjust the definition of \mathcal{S} so that each graph in \mathcal{S} also has a unique distinguished vertex, as a *pointed space*; then \diamond might denote the disjoint union, except for the two distinguished nodes, which are to be identified together as the one new distinguished vertex, inheriting neighbours from both original graphs (this is a kind of *star* combination of the two graphs).

For any positive integer $n \in \mathbb{Z}^+$, let nS denote the \diamond operation on n successive copies of S :

$$nS = S \diamond S \diamond \dots \diamond S = S \diamond (n-1)S, \quad n = 2, 3, \dots$$

Let $Q : \mathcal{S} \rightarrow \mathcal{B}$ be a function over \mathcal{S} taking values in a Banach space, \mathcal{B} . We are interested in functions which are relatively *well behaved* with respect to the edge set of S : in particular functions where a small relative change in the edge set may only result in a small change in Q . Since \mathcal{S} contains discrete objects, we will assume the following condition: there exists $C > 0$ such that for any graph $S_1 \in \mathcal{S}$, if we add some new edges to produce a new graph, S_2 , then

$$\|Q(S_2)\|_{\mathcal{B}} - \|Q(S_1)\|_{\mathcal{B}} \leq C \frac{|E(S_2/S_1)|}{|E(S_1)|}.$$

FIG. 3. Under what circumstances does a scaling law G exist?

Here, $|E(S_2/S_1)| = |E(S_2)| - |E(S_1)|$ is the change in the number of edges. As S_1 and $E(S_1)$ get larger, the addition of a fixed number of edges become a smaller relative change, and thus the change in Q tends to zero in the limit.

This rules out many functions, Q , which count graphs properties. For example, if Q counts the connected components of a graph, then even for the very largest disconnected graph, the addition of a single edge can reduce Q by one. However, many functions are included such as the spectral radius of the adjacency matrix, the maximal modularity achievable via any partition [12], the average degree of vertices, the reachable fraction of a connected the network by walks of length k or less starting from a single random vertex (as in Fig. 2). This restriction is important within applications, since the data specifying graphs may contain some errors in either the observation or the interpretation of edges: with both false positives and false negatives. Many properties of graphs may not be robust to such errors.

We define a *scaling function*, $G : \mathcal{B} \times \mathbb{Z}^+ \rightarrow \mathcal{B}$, for Q and \diamond , if it exists, to be a mapping that describes how the function Q behaves under the \diamond -combination of multiple isomorphisms of any graph in $S \in \mathcal{S}$, that satisfies

$$Q(nS) = G(Q(S), n), \quad (1)$$

for all $S \in \mathcal{S}$ and all $n \in \mathbb{Z}^+$.

We are interested in setting out a sufficient condition on the pair (\diamond, Q) for the existence of a scaling function, see Fig. 3.

The following result is a generalization of that given in [9] (where $\mathcal{B} = \mathbb{R}$ and \diamond represented the disjoint union of two graphs).

THEOREM Suppose that for all $S_0 \in \mathcal{S}$ there exists $C(S_0) > 0$ such that Q satisfies

$$\|Q(S_1 \diamond S_0) - Q(S_2 \diamond S_0)\|_{\mathcal{B}} \leq C(S_0) \|Q(S_1) - Q(S_2)\|_{\mathcal{B}} \text{ for all } S_1, S_2 \in \mathcal{S}. \quad (2)$$

for all S_1 and $S_2 \in \mathcal{S}$. Then, there exists a scaling function, G , satisfying (1).

Proof. Firstly, suppose that for any S_1 and S_2 in \mathcal{S} such that $Q(S_1) = Q(S_2)$ we also have $Q(nS_1) = Q(nS_2)$ for all $n \geq 1$. Then a scaling function, G , satisfying (1) may be constructed in a pointwise fashion, without any ill-definition: for any pair (\hat{Q}, n) , such that \hat{Q} is in the range of Q we may select any S for which $Q(S) = \hat{Q}$ and assign the corresponding value $Q(nS)$ to $G(\hat{Q}, n)$, without any inconsistency.

So, if (1) fails, then we there must exist at least one pair of graphs S_1 and S_2 in \mathcal{S} and a smallest integer $n^* > 1$, such that $Q(nS_1) = Q(nS_2)$ for $n = 1, \dots, n^* - 1$: that is, Q cannot distinguish between

nS_1 and nS_2 for $n < n^*$; and yet $Q(n^*S_1) \neq Q(n^*S_2)$, so that Q does distinguish between n^*S_1 and n^*S_2 . Such a situation clearly would make any desired scaling function, G , ill-defined.

Secondly, we will assume the inequality (2) together with the converse of (1) and establish a contradiction.

Assuming the converse of (1), then there exist S_1 and S_2 in \mathcal{S} such that $Q(nS_1) = Q(nS_2)$ for $n = 1, \dots, n^* - 1$ and yet $Q(n^*S_1) \neq Q(n^*S_2)$. Then, by using the commutativity of \diamond , we have

$$\begin{aligned} \|Q(n^*S_1) - Q(n^*S_2)\|_{\mathcal{B}} &= \|Q(S_1 \diamond (n^* - 1)S_1) - Q(S_2 \diamond (n^* - 1)S_2)\|_{\mathcal{B}} \\ &= \|Q(S_1 \diamond (n^* - 1)S_1) - Q(S_2 \diamond (n^* - 1)S_1) + Q(S_2 \diamond (n^* - 1)S_1) - Q(S_2 \diamond (n^* - 1)S_2)\|_{\mathcal{B}} \\ &\leq \|Q(S_1 \diamond (n^* - 1)S_1) - Q(S_2 \diamond (n^* - 1)S_1)\|_{\mathcal{B}} + \|Q(S_2 \diamond (n^* - 1)S_1) - Q(S_2 \diamond (n^* - 1)S_2)\|_{\mathcal{B}}. \end{aligned}$$

Employing (2) twice we have

$$\|Q(n^*S_1) - Q(n^*S_2)\|_{\mathcal{B}} \leq C((n^* - 1)S_1)\|Q(S_1) - Q(S_2)\|_{\mathcal{B}} + C(S_2)\|Q((n^* - 1)S_1) - Q((n^* - 1)S_2)\|_{\mathcal{B}}.$$

But both of these terms vanish, so $Q(n^*S_1) = Q(n^*S_2)$, which is the required contradiction. Hence if (2) holds, then (1) must hold. \square

REMARK The expression (2) is a Lipschitz-type condition.

REMARK The argument in [9] may be extended to show that the functional equation (1), together with the initial condition $G(Q, 1) = Q$, has solutions in the form of both power law growth and power law decay to a constant, $G(Q, n) = n^{-\alpha} + \beta(1 - n^{-\alpha})$, for constants $\alpha \in \mathbb{R}$, $\beta \in \mathcal{B}$. This is achieved by applying (1) twice over, so that $G(Q, nm) = G(G(Q, m), n)$, and deploying an ansatz.

If the binary combination operation is not commutative, then we will require Q to be Lipschitz in both arguments of \diamond , since the proof above theorem relies on this, and we replace it with the following.

COROLLARY Suppose that for all $S_0 \in \mathcal{S}$ there exists $C(S_0) > 0$ such that Q satisfies

$$\max\{\|Q(S_1 \diamond S_0) - Q(S_2 \diamond S_0)\|_{\mathcal{B}}, \|Q(S_0 \diamond S_1) - Q(S_0 \diamond S_2)\|_{\mathcal{B}}\} \leq C(S_0)\|Q(S_1) - Q(S_2)\|_{\mathcal{B}}, \quad (3)$$

for all $S_1, S_2 \in \mathcal{S}$. Then, there exists a scaling function, G , satisfying (1).

3. Scaling laws for random graphs

In this section, we extend the above considerations to random graphs, as some natural graph-building operations have a stochastic elements: for example, *preferential attachment* constructions [7].

A random graph W is a probability distribution, $P_W(S|X)$ defined over \mathcal{S} , conditional on any known prior information, X . Let \mathcal{W} denote the set of all random graphs. \mathcal{W} is a complete metric space under the L^1 metric: $d(W_1, W_2) = \sum_{S \in \mathcal{S}} |P_{W_1}(S|X) - P_{W_2}(S|X)|$. If fact \mathcal{W} is isomorphic to the non-negative elements on the 1-sphere of l_1 , the space of real-valued semi-infinite sequences, since \mathcal{S} is countable (and thus may be ordered: for each value of n , the number of vertices within any given graph, there are exactly $2^{n(n-1)/2}$ graphs in \mathcal{S}).

A random graph may be fully described by its probability distribution, or else, more succinctly, by the random process which generates elements drawn from it [13], for example, by defining an independent probability for every possible edge. Most commonly studied is the random graph proposed by Gilbert, denoted $\mathcal{G}(K, p)$, in which every possible edge occurs independently with probability, $p \in (0, 1]$. It is clear how one may draw an element, S , from such a graph by testing every possible edge over the K vertices and including each with independent probability p . Alternatively, if S is any given graph over K vertices, with edge set E , then we have the more cumbersome distribution

$$P_{G(K,p)}(S|X) = p^{|E|} (1-p)^{\frac{K(K-1)}{2} - |E|}.$$

All graphs defined on K vertices with exactly $|E|$ edges, including those arrived at via any vertex reordering, are thus equally probable

Alternatively, a circular ‘range-dependent’ random lattice on K vertices, arranges those vertices in a circle, like the hours of the clock, with the *longer range*, more distant, vertices being successively less likely to be connected [14], similar to the well-known ‘small world’ networks [15].

Next, we use \square to denote a binary operation over \mathcal{W} , so that $W_1 \square W_2$ is composite random graph, formed according to some mechanism, that is conditional on both W_1 and W_2 (as well as X).

As before, we let $nW = W \square \dots \square W$ denote the corresponding combination of n copies of W .

Now let $Q : \mathcal{W} \rightarrow \mathcal{B}$ be a mapping taking values in a Banach space, \mathcal{B} . For example, $Q(W)$ might denote the expected value of some given function over \mathcal{S} under the distribution $P_W(S|X)$. Then, we may extend the arguments given in the above to yield the following result.

THEOREM If Q and \square are such that for all $W_0 \in \mathcal{W}$, there is a constant $C(W_0) > 0$, and for all W_1 and $W_2 \in \mathcal{W}$, we have

$$\max\{\|Q(W_1 \square W_0) - Q(W_2 \square W_0)\|_{\mathcal{B}}, \|Q(W_0 \square W_1) - Q(W_0 \square W_2)\|_{\mathcal{B}}\} \leq C(S_0) \|Q(W_1) - Q(W_2)\|_{\mathcal{B}}, \quad (4)$$

then a scaling function $G : \mathcal{B} \times \mathbb{Z}^+ \rightarrow \mathbb{R}$ exists, satisfying

$$Q(nW) = G(Q(W), n),$$

for all $W \in \mathcal{W}$ and all $n \in \mathbb{Z}^+$.

Note this scaling law refers to the Q -valued property (in \mathcal{B}) of the random graph nW as it grows with n .

4. A PASBM

Using the results from Section 3, we introduce a novel form of a random graph, with familiar parentage, constructed as follows.

We generate successive sub-component graphs, all from the same random Gilbert graph, $\mathcal{G}(K, p)$, with parameters K and p fixed. At each successive step, we have an ‘existing combination graph’ (ECG) and a new Gilbert graph. At the first step, the ECG is just a single graph drawn from $\mathcal{G}(K, p)$. At each step, we combine the two graphs by introducing an edge between a vertex within the ECG selected with probability proportional to the present ECG’s vertex degree distribution, and a vertex within the new Gilbert graph, again selected with probability proportional to the observed vertex degree distribution.

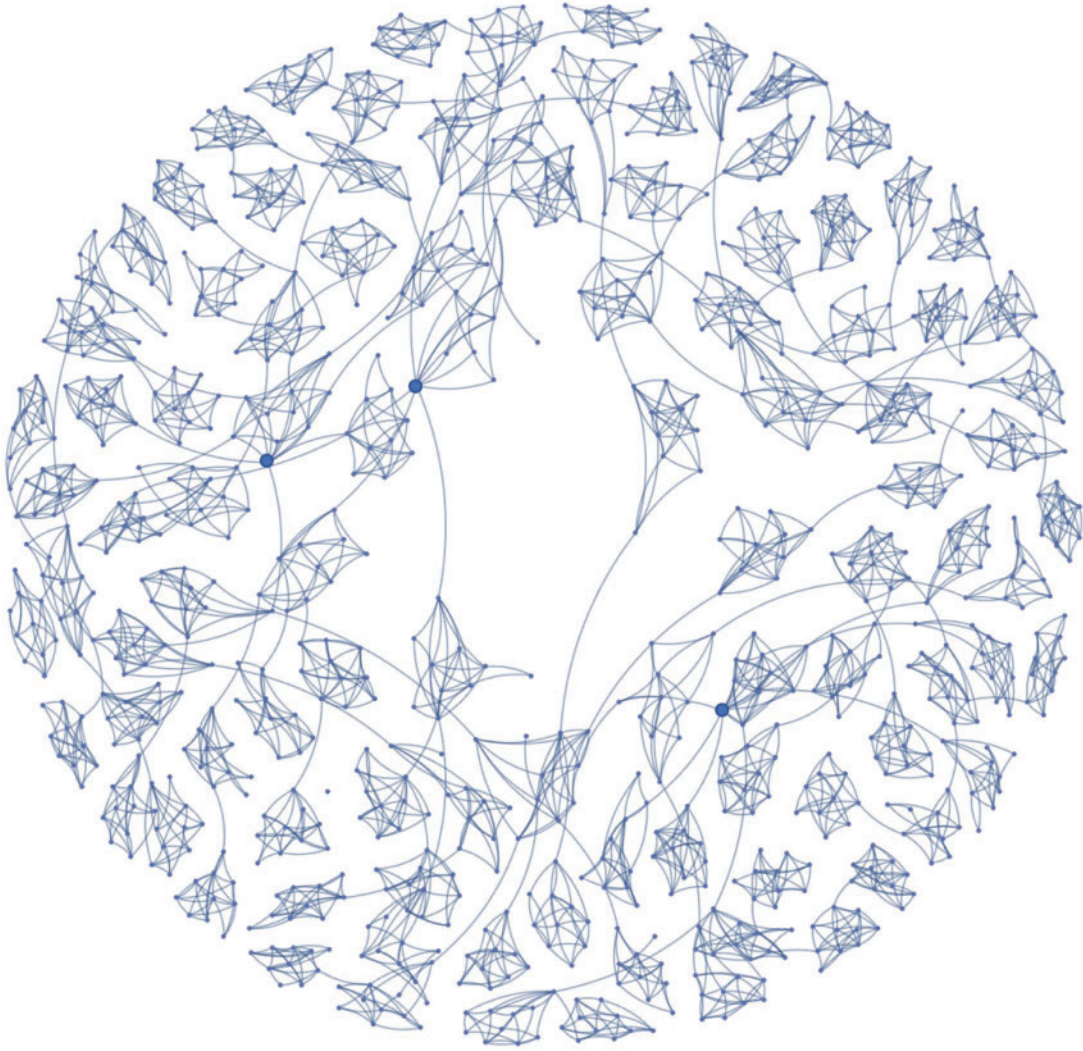


FIG. 4. An instance of a preferential attachment stochastic block model (PASBM) random graph with $(K, p, n) = (10, 0.6, 100)$. The three vertices with equal highest degree (equal to 11) are highlighted.

After the n steps, we have a random graph over Kn vertices, where the successive n Gilbert graphs have been preferentially attached. In Fig. 4, we show such a graph, where $(K, p, n) = (10, 0.6, 100)$. So, for the PASBM graph, the combination operation, \square , involves (degree-biased) random preferential attachment. Note if $K = 1$, then we preferentially attach a single new vertex to the ECG at each step, and thus we reduce to the usual preferential attachment model [7, 8].

As the PASBM graph grows with n , by attaching new SBM graphs (while the total number of vertices is nK), there are certain properties that appear to remain stationary. Two such are the cumulative distribution for the (real) eigenvalues of the self-adjoint adjacency matrix, A , and the associated combinatorial

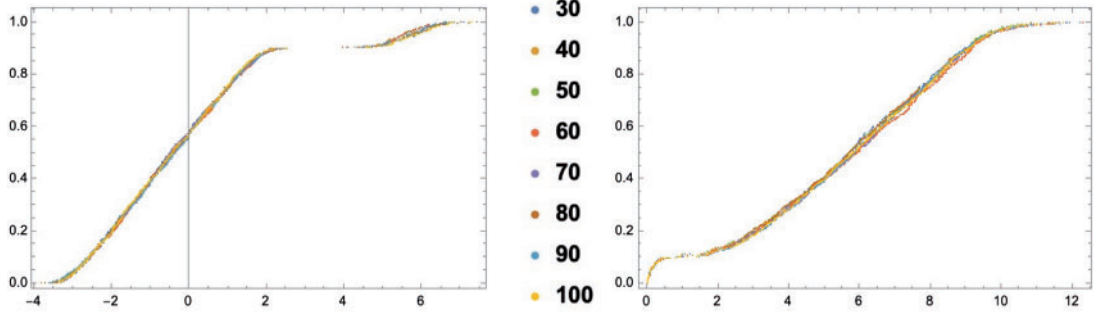


FIG. 5. The cumulative distribution of eigenvalues for the adjacency matrix (left) and the combinatorial Laplacian (right) associated with PASBM random graphs with $(K, p) = (10, 0.6)$, and $n = 30, 40, \dots, 100$.

Laplacian, $\Delta = D - A$ (where D is the diagonal vertex degree matrix). In Fig. 5, we show the cumulative distributions of the eigenvalues for $(K, p) = (10, 0.6)$, and $n = 30, 40, \dots, 100$. Of course, the density of eigenvalues grows linearly with n (there are nK eigenvalues, as n increases). The convergence occurs because at each step the ECG contains more and more vertices with relatively high degree, and thus the new edges to be added to the ECG at each step are shared out across more and more competing possibilities. This convergence of the cumulative spectral distribution is a very attractive feature in itself and is will be the subject of future investigation.

Now consider a function Q , of these graphs, which is itself the sum of a function g taken over the eigenvalues, λ_i , of either the adjacency matrix, or the Laplacian. Then, we have

$$Q(n\mathcal{G}(K, p)) = \sum_{i=1}^{nK} g(\lambda_i) = nK \left(\frac{1}{nK} \sum_{i=1}^{nK} g(\lambda_i) \right).$$

The adjacency matrix and Laplacian are both properties of a graph, and their spectra are invariant under any reordering of the vertices. So, any function of either spectra is a function of the graph's structure (regardless of the vertex ordering). In contrast, matrix-valued functions must necessarily include the eigenvector information (within the Jordan normal form of both the matrix and the consequent matrix-valued function). Here though, we are just considering a function Q in the form of a sum of some given function g taken over the eigenvalues (of either the adjacency matrix or the Laplacian). This is well-defined and is not dependent upon the vertex ordering. (For undirected graphs, these eigenvalues are all real, of course.)

Assuming, from Fig. 5, that the distribution of the eigenvalues converges as $n \rightarrow \infty$ to a density function, say $f(\lambda)$, then we must have asymptotic linear growth,

$$\lim_{n \rightarrow \infty} Q(n\mathcal{G}(K, p)) = nK \int f(\lambda) g(\lambda) d\lambda.$$

Similarly suppose Q , is the sum of a given monomial in M variables, as each variable is drawn from the set of eigenvalues of either the adjacency matrix or the Laplacian (in Fig. 5). We have

$$Q(n\mathcal{G}(K, p)) = \sum_{i_1=1}^{nK} \sum_{i_2=1}^{nK} \dots \sum_{i_M=1}^{nK} \lambda_{i_1}^{\alpha_1} \lambda_{i_2}^{\alpha_2} \dots \lambda_{i_M}^{\alpha_M} = \prod_{k=1}^M \sum_{i_k=1}^{nK} \lambda_{i_k}^{\alpha_k} = (nK)^M \prod_{k=1}^M \left(\frac{1}{nK} \sum_{i=1}^{nK} \lambda_i^{\alpha_k} \right),$$

where we have $\alpha_k > 0$ ($k = 1, \dots, M$), and $\sum_{k=1}^M \alpha_k = d > 1$, the monomial degree. Then, under the same convergence assumption, we must have asymptotic power-law growth,

$$\lim_{n \rightarrow \infty} Q(n\mathcal{G}(K, p)) = (nK)^M \left\{ \prod_{k=1}^M \int f(\lambda) \lambda^{\alpha_k} d\lambda \right\}.$$

REMARK There is nothing special about the Gilbert graph, $\mathcal{G}(K, p)$; we might successively combine graphs drawn from any suitable random graph, as stochastic blocks, by employing the preferential attachment method implicit in the above definition of \square .

5. An application to UK cities

Since the early work on growth scaling laws, reviewed in [3], it has become common to compare various performance measures across a wide range of cities (modular urban conurbations) of very different sizes, and to extract empirical power-law growth laws. This is based on a plausible assumption that the cities are themselves composed of a class of similar basic components (modules) but differ in size. While this may be true of the underlying infrastructure and public services, when it comes to city-based social networks this may not be so. Cities may fall into fundamentally distinct classes.

Consider the highly modular UK city-based social media ‘reciprocated mentions networks’ [12] in Fig. 6 (data available at [16]), which represents a standardized (Twitter-based) section of the social network within each city.

In Fig. 6, we also show the Bristol graph partitioned so as to maximize the modularity function [1, 2], indicating the high modularity of these social networks. In [12], fictitious cities were constructed, of any size, from existing, real, cites, by re-sampling the separate modules independently (with repetition). These modules were successively combined and sparsely connected. In [12], this was achieved in an ad hoc manner. Now, using the PASBM approach, we could attach new modules on to existing ones. That analysis fabricates an ensemble of new cities, each of the size of Leeds, say, yet made from modular components from Bristol, say. By contrasting the real Leeds graph with an ensemble of city graphs of a similar size, fabricated from Bristol’s modules, we may decide whether it is reasonable or not to consider Leeds as an up-scaled version of Bristol (in [12], it is shown to be not reasonable, by observing the modularity function, which itself scales with an inverse power-law asymptote).

Though high modularity implies the existence of scaling laws for any reasonable functions (from the results of Sections 2 and 3, and earlier in [9]), it is simply not a given that all UK cities of rather distinct sizes are composed of the same class of basic modules.

In Fig. 7, we have decomposed Bristol into a 10-module partition (by maximizing the modularity function). For each module taken in isolation (in order of decreasing size), we depict the cumulative spectral distributions for the corresponding adjacency matrix. As in Fig. 5, we see that these distributions are similar while the density of the eigenvalues decreases. We also show the same procedure carried out on a Gilbert random graph of the same size and density as Bristol. In the latter case, each module is really a Gilbert graph itself (on the requisite number of vertices). In both cases, the similarity of the distributions suggests that all modules are drawn from a random graph, with size being the only real difference. Of course, the Bristol modules (like the whole Bristol graph) possess a plethora of eigenvalue sitting arbitrarily close to zero. However, we may characterize such graphs, a Gilbert graph is clearly not a good candidate.

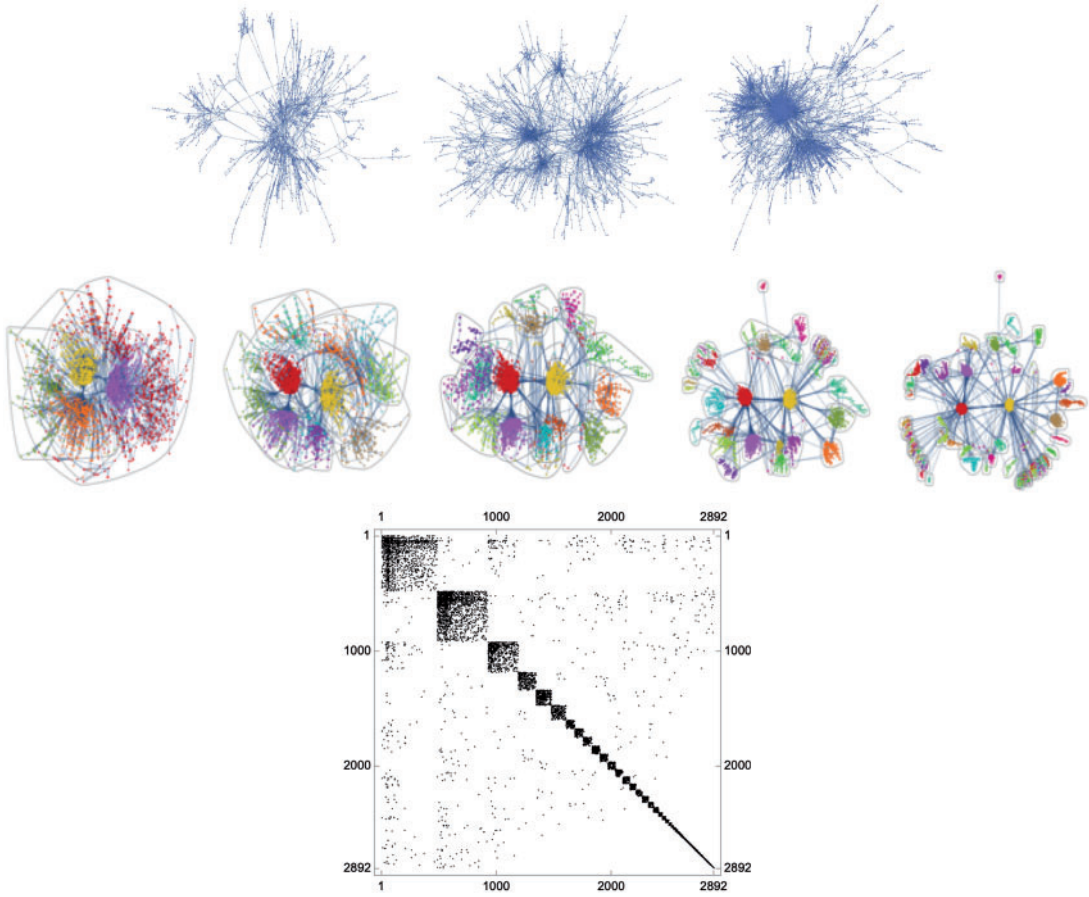


FIG. 6. Above: reciprocated mentions networks from three UK cities: Edinburgh, Bristol and Leeds, where $(|V|, |E|) = (1645, 2146), (2892, 4538)$ and $(5263, 9319)$, respectively. Middle: the Bristol graph successively partitioned into 5, 10, 20, 40, 74 (the optimum) modules, from [12]. Below: Bristol's adjacency matrix reordered as a SBM with 74 blocks/modules of decreasing size.

Given the stochastic block nature of these modularity decompositions, the spectra of the whole graphs are close to the union of those for the component blocks. This is in line with the convergence behaviour, such as observed in Fig. 5, as more blocks are added during growth. The main point is that Bristol's modules have similar spectra and they could all be drawn separately from some fundamental model and then attached together via a PASBM.

6. Alternative random graph combination operations

Our characterization of \square as a binary operation over \mathcal{W} , leaves much room for further thought. Beyond disjoint union and preferential attachment, which are both types of aggregation, we now discuss some alternatives well worth considering as prospective models for system growth, since all will yield scaling laws, with respect to growth, whenever suitably behaved functions, Q , are observed.

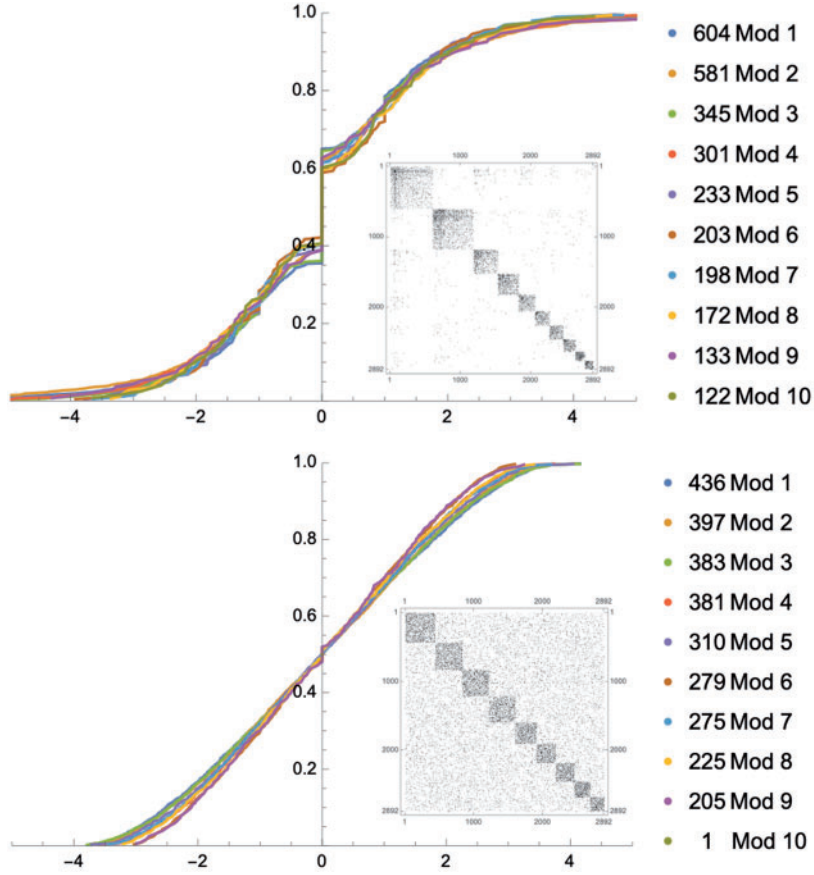


FIG. 7. Above: cumulative spectral distributions for each module from the 10-module partition of Bristol (re-ordered adjacency plot included). Below: the same for a single Gilbert graph of the same size and similar density as Bristol, partitioned into 10 modules. In the keys ‘N Mod J’ means module number J contains exactly N vertices in the partitions.

6.1 Mechanistic aggregative combination

So far we have combined two random graphs via (biased and random) *preferential attachment*. In that case, we first took the disjoint union and then selected a single vertex from each component sub-graph, according to a probability distribution proportional to the (sub-graph) vertex degrees, and then joined them with a single new edge.

Similarly, we might have first taken the disjoint union and then selected a single vertex from each component sub-graph, according to a probability distribution proportional to the r power of the vertex (sub-graph) degrees, and joined them with a single new edge. As $r \rightarrow 0$, this is equivalent to randomly choosing a vertex within each graph, with equal probabilities, and joining them: this might be termed *non-preferential attachment*. As $r \rightarrow \infty$ this is equivalent to picking a vertex with the maximum degree within each graph and then joining them: this might be termed *most popular vertex attachment* (a random element is necessary only if there exists more than one vertex with maximum degree in either graph).

We may generalize these mechanisms to include up to $J > 1$ connected edges between the random sub-graphs, each drawn independently with repetition: this might be termed called *multiple-edge preferential attachment/non-preferential attachment/most popular vertex attachment*.

These are all *aggregative* operations, since they embellish the disjoint union of the sub-components, maintaining the component vertices.

We have already mentioned *star attachment*. This applies where each of the graphs drawn from a random graph additionally have a unique *distinguished* vertex. Two such graphs are combined by first taking their disjoint union, and then followed by identifying both components' distinguished vertices together as one single distinguished vertex in the combined graph, retaining all of the edges from both of the component graphs.

6.2 Product combination

Hierarchical network models [17] have been proposed two decades ago, where a deterministic self-similar network is constructed iteratively, by successively combining copies of the same graph. This was motivated by the failure of the other simple scale-free models to produce and maintain high clustering coefficients [18]. The combination operator can be extended to produce a hierarchical model combining successive random graphs. The random graph must have a distinguished, 'core', vertex (as with star attachment), the none core vertices are termed 'peripheral'. At each new iteration/level multiple copies of (in our case) a random graph are attached, with edges connected all peripheral nodes to the core vertex from an instance at the previous iteration/level.

In this construction, the iterated operator, \square , is not commutative. However the scaling theorem from Section 3 still applies, under the extended condition given in Section 4, guaranteeing a scaling law for certain functions, under the corollary given in Section 3.

Hierarchical models combining random graphs are analogous to statistical self-similar fractals, rather than deterministically self-similar fractals.

A hierarchical approach also underpins the Kronecker graphs [19], introduced over a decade ago, with graphs combining via a product; and that non-commutative combination operation calls for the the corollary from Section 3 in order to guarantee scaling laws as graphs grow iteratively.

6.3 Directly combining probability distributions

Let $h : [0, 1]^2 \rightarrow \mathbb{R}^+$ be any mapping of two variables. Then, for any two random graphs W_1 and W_2 in \mathcal{W} , and all graphs $S \in \mathcal{S}$, we define \square by

$$P_{W_1 \square W_2}(S|X, W_1, W_2) = \frac{h(P_{W_1}(S|X), P_{W_2}(S|X))}{\sum_{S' \in \mathcal{S}} h(P_{W_1}(S'|X), P_{W_2}(S'|X))}.$$

Restricting our interest to graphs on K vertices, by direct observation will have $\mathcal{G}(K, p_1) \square \mathcal{G}(K, p_2) = \mathcal{G}(K, h(p_1, p_2))$.

If, in addition, $h(x, x) = x$ for $x \in [0, 1]$ then $W \square W = W$.

This type of definition may actually be impractical in many ways, as it seems non-constructive. Yet one can imagine sampling (using MCMC e.g. [20]) graphs from such a combined probability distribution, albeit relatively inefficiently. This will be the subject of future work.

7. Conclusions

We have shown that there are a wide number of methods by which one might successively combine random graphs so as to create a growing random graph. Regardless of that chosen combination operator, which may itself include some random elements, there is a strong possibility that observable (summary) performance measures (taking values in some Banach space) will possess a scaling law, as the underlying graph grows. Such scaling might control unbounded growth or the decay to some asymptotic constant. The deciding factor will be whether the function of interest, denoted by Q above, is *well behaved* meaning that it satisfies a Lipschitz-type condition under alternative combinations.

By making these, quite general, considerations we have created a framework that subsumes *aggregative combination* processes and also *product combination* processes, including Hierarchical and Kronecker models introduced elsewhere. We also introduced the particular class of PASBM graphs. Of course, within any particular instance with the general framework developed here one would need to check that the function of interest, Q , satisfies the required conditions. In a large part though this demonstrates the ubiquity of scaling laws for functions of underlying growing networks, as sequential operations are made to incorporate an increasing number of random graphs as sub-components.

Power laws are an important sub-class of scaling laws, for both growth and decay. Their relative preponderance within analyses of aggregative processes is well known. They are of course solutions of the general scaling law functional equation, though exponential scaling laws can also exist.

The importance of this research is in extending the previous work to consider successive combinations of random graphs and thus allowing the combination operator to be very general. It is thus applicable far beyond the standard applications to the aggregative growth of highly modular graphs in modelling cities, organization and organisms.

Acknowledgements

P.G. is pleased to acknowledge the support and advice given by Renaud Lambiotte and Desmond J. Higham.

REFERENCES

1. NEWMAN, M. E. J. (2010) *Networks: An Introduction*. Oxford, New York: Oxford University Press.
2. BLONDEL, V. D., GUILLAUME, J.-L., LAMBIOTTE, R. & LEFEBVRE, E. (2008) Fast unfolding of communities in large networks. *J. Stat. Mech. Theory Exp.*, **2008**, p10008.
3. WEST, G. (2017) *Scale: The Universal Laws of Growth, Innovation, Sustainability, and the Pace of Life in Organisms, Cities, Economies, and Companies*. London, UK: Penguin Press.
4. BETTENCOURT, L. M., LOBO, J., HELBING, D., KÜHNERT, C. & WEST, G. B. (2007) Growth, innovation, scaling, and the pace of life in cities. *Proc. Natl. Acad. Sci. USA*, **104**, 7301–7306.
5. HIGHAM, D. J., BATTY, M., BETTENCOURT, L. M. A., GREETHAM, D. V. & GRINDROD, P. (2017) An overview of city analytics. *R. Soc. Open Sci.*, **4**. <https://doi.org/10.1098/rsos.161063>.
6. WEST, G., BROWN, J. & ENQUIST, B. (1999) The fourth dimension of life: fractal geometry and allometric scaling of organisms. *Science*, **284**, 1677–1679.
7. BARABASI, A.-L. & ALBERT, R. (1999) Emergence of scaling in random graphs, *Science*, **286**, 509–512.
8. BOLLOBAS, B., RIORDAN, O., SPENCER, J. & TUSNADY, G. (2001) The degree sequence of a scale-free random graph process. *Random Struct. Algorithms*, **18**, 279290.
9. GRINDROD, P. & HIGHAM, D. J. (2018) High modularity creates scaling laws. *Sci. Rep.*, **8**, 9737.
10. LEE, C. & WILKINSON, D. J. (2019) A review of stochastic block models and extensions for graph clustering. *Appl. Netw. Sci.*, **4**, 122.

11. GILBERT, E. N. (1959) Random graphs. *Ann. Math. Stat.*, **30**, 1141–1144.
12. GRINDROD, P. & LEE, T. E. (2016) Comparison of social structures within cities of very different sizes. *R. Soc. Open Sci.*, **3**, 150526150526.
13. BOLLOBÁS, B. (2001) *Random Graphs*, 2nd edn. Cambridge University Press.
14. GRINDROD, P. (2014) *Mathematical Underpinnings of Analytics*. Oxford: Oxford University Press.
15. WATTS, D. J. & STROGATZ, S. H. (1998) Collective dynamics of ‘small-world’ networks. *Nature*, **393**, 440–442.
16. GRINDROD, P. & LEE, T. E. (2016) Data from: Comparison of social structures within cities of very different sizes. Dryad: <https://datadryad.org/stash/dataset/doi:10.5061/dryad.2gf23>.
17. RAVASZ, E. B. & BARABÁSI, A. L. (2003) Hierarchical organization in complex networks. *Phys. Rev. E*, **67**, p026112.
18. NOH, J. (2003) Exact scaling properties of a hierarchical network model. *Phys. Rev. E*, **67**, p045103.
19. LESKOVEC, J., CHAKRABARTI, D., KLEINBERG, J., FALOUTSOS, C. & GHAHRAMANI, Z. (2010) Kronecker graphs: an approach to modelling networks. *J. Mach. Learn. Res.*, **11**, 985–1042.
20. TAO, T. (2016) An improved MCMC algorithm for generating random graphs from constrained distributions. *Netw. Sci.*, **4**, 117–139.