Supporting Information

Palla et al. 10.1073/pnas.0912983107

SI Text

Thermodynamic Limit of Graph Sequences. The infinitely large system size limit is very important in statistical physics, as the properties of the studied phenomena are manifested in the most pure way in this limit. Similarly, in complex network theory the infinitely large limit graph of a converging graph sequence can be considered as a "platonic" network exhibiting the fundamental common properties of the graphs in the sequence in the most pure way. But under what conditions can we say that a given sequence of graphs is converging to something nontrivial and yet sufficiently universal to be conceptually meaningful? A simple intuitive condition is that the statistical features commonly used to characterize a network (e.g., degree distribution, clustering coefficient, etc.) should converge. The actual definition is based on homomorphisms (adjacency preserving maps) as follows. For two simple graphs F and G, let $\hom(F,G)$ denote the number of homomorphisms from $V(F)$ (the nodes in F) to $V(G)$ (the nodes in G). The *homomorphism density* $t(F,G)$ is defined as the probability that a random map from $V(F)$ to $V(G)$ is a homomorphism, given by

$$t(F,G)=\frac{\hom(F,G)}{|V(G)|^{|V(F)|}}. \qquad [S1]$$

A sequence (G_n) of graphs is *convergent*, if the sequence $t(F,G_n)$ has a limit for every simple graph F. Loosely speaking, this condition can be interpreted as the convergence of the probability of finding any given finite subgraph in the sequence of networks.

We note that when G has bounded average degree and F is connected, it is more natural to normalize the homomorphism density by dividing with $|V(G)|$ instead of $|V(G)|^{|V(F)|}$. Therefore, we let $\mathrm{inj}(F,G)$ denote the number of injective homomorphisms from graph F to graph G and define

$$s(F,G)\equiv\frac{\mathrm{inj}(F,G)}{|V(G)|}. \qquad [S2]$$

[Thus, $s(F,G)$ is the average number of labeled copies of F, such that a specified node of F goes on a specified node of G.]

Convergent graph sequences are related to 2D functions in a nontrivial way (1, 2). First of all, we can construct a convergent graph sequence using a symmetric measurable function, $0\le W(x,y)\le 1$, defined on the unit square as follows. For a given network size N, we distribute N points independently, uniformly at random on the $[0,1]$ interval. These points correspond to the nodes in the network, and each pair of nodes is linked with the probability given by $W(x,y)$ at the coordinate of the according points. In the $N\to\infty$ limit the obtained graph sequence is converging. What is even more surprising, it can be proven that we can represent any convergent graph sequence by a 2D function, since for any convergent graph sequence one can find a $W(x,y)$ providing the same limiting subgraph densities.

The average degree of nodes in a random graph generated from a given $W(x,y)$ using the construction above can be given simply as

$$\langle d\rangle = N\iint W(x,y)dxdy. \qquad [S3]$$

Thus, in the $N\to\infty$ limit the obtained network becomes dense. In contrast, real networks are usually sparse in the sense that their average degree is not expected to grow with increasing size. A solution to this problem was proposed by Bollobás et al., by redefining the linking probabilities as $W(x,y)/N$, resulting in a network with an average degree independent of N. They showed that, depending on the choice of $W(x,y)$, a wide range of sparse networks can be generated.

Our approach is different from this method in that instead of using a construction into which we build in the level of complexity from the beginning, we generate complexity by using tensorial products of increasing power as $N\to\infty$. This is a qualitatively new picture, corresponding to reality to a higher degree (larger graphs are more complex/inhomogeneous/structured in nature than smaller graphs). In addition, we achieve this using a relatively simple construction.

The Multifractal Link Probability Measure. As explained in the main text, our link probability measure corresponds to a multifractal defined on the unit square. This can be best envisaged as a surface getting rougher and rougher at each iteration (as shown in Fig. 1 of the main text), where the height of the surface signals the measure associated to the rectangle beneath. The "roughening" is governed by Eqs. **1–2**) in the main text: When iterated, each rectangle is divided to smaller rectangles following the pattern given by the generating measure, and the measure associated to these new rectangles is given by the measure of the starting rectangle, multiplied by the corresponding element of p_{ij} defining the generating measure. (In parallel, the ratio between the area of a new rectangle divided by the area of starting rectangle is equal to the area of the corresponding rectangle in the generating measure.) For clarity, in Fig. S1 we show the expressions for the elements of the link probability measure displayed in Fig. 1 of the main text.

Entropy of the Generated Graph Ensemble. The concept of entropy in the context of graph ensembles was introduced by Bianconi in ref. 3 to measure the complexity or the level of order in networks. The graph ensembles considered there are analogous to the microcanonical ensemble in statistical physics: All configurations fulfilling a chosen criterion (e.g., a fixed average degree or a fixed degree sequence) are equally probable. Thus, the corresponding entropy is equal simply to the logarithm of the number of all possible graphs:

$$S=\ln\mathcal{N}. \qquad [S4]$$

In our case the probability to obtain a given network configuration from a fixed link probability measure is strongly dependent on the configuration. Therefore, our graph ensemble is more close to the canonical ensemble, where the entropy can be defined as

$$S=-\sum_G p_G\ln p_G, \qquad [S5]$$

with the sum running over all possible graphs and p_G denoting the probability of a given configuration G. In the special case of equally probable configurations ($p_G\equiv p$ for all G), due to the normalization criterion $\sum_G p_G=1$ the above formula for S simplifies to expression [**4**] given by Bianconi. However, in general for a graph G generated by the multifractal graph generator the p_G can be given as

$$p_G = \prod_{(IJ)\in G} p_{ij}(k) \prod_{(IJ)\notin G} (1 - p_{ij}(k)), \qquad \text{[S6]}$$

where the first product is over all links, the second product is over all missing links, and $p_{ij}(k)$ denotes the link probability measure at the coordinates corresponding to the node pair (I,J).

Although Eqs. **5–6** give a simple formulation of the entropy, its evaluation is far from trivial due to the very large number of possible graphs. In fact, any configuration with N nodes is a possible outcome of our method from the empty graph with absolutely no edges to the completely filled graph with $N(N-1)/2$ connections. Thus, the total number of graphs is equal to $\mathcal{N} = 2^{N(N-1)/2}$, which means that the exact evaluation of S is only plausible for graphs with a very few nodes only. For larger number of nodes one can try to reveal the frequency of graphs, $\rho(p_G)$ in function of p_G with sampling of the possible graphs. With the help of $\rho(p_G)$ the number of graphs with $p_G \in [p'_G, p''_G]$ can be expressed as $\mathcal{N}\int_{p'_G}^{p''_G}\rho(p_G)dp_G$, and the entropy defined in Eq. **5** can be formulated as

$$S = -\mathcal{N}\int p_G \ln(p_G)\rho(p_G)dp_G. \qquad \text{[S7]}$$

However, our preliminary experiments based on sampling showed that $\rho(p_G)$ shows a hugely varying nature; thus, it must be explored with a number of samples of similar order to the actual number of graphs to get a reasonable estimate of S. Nevertheless, these experiments indicated that the entropy is increasing faster than linear, but slower than quadratic in function of the number of nodes. This means that similarly to the case of, e.g., graph ensembles with fixed degree sequence, the entropy per node is increasing with the system size; however, its increase is slower than in a graph ensemble without any restriction on the structure of graphs.

Limiting Cases of the Multifractal Model. A shortcoming of our model is that it can lead to a network in which the majority of nodes are isolated in the $N \to \infty$ limit. However, as we shall see, this effect is negligible for graphs in the size range of real networks.

Analytical results. In general, if $W_1(x,y), W_2(x,y), \ldots, W_k(x,y)$ is a sequence of symmetric measurable functions on the unit square [with $0 \le W_k(x,y) \le 1$ for any k], let us define $w_k(x)$ as the average linking probability for a node at position x given by

$$w_k(x) = \int_0^1 W_k(x,y)dy. \qquad \text{[S8]}$$

Similarly, let ω_k denote the average link probability for the whole network, which can be expressed as

$$\omega_k = \int_0^1 w_k(x)dx. \qquad \text{[S9]}$$

Let us choose the number of nodes N_k associated to $W_k(x,y)$ in such a way that the average degree of nodes converges to a constant (nonzero) $\langle d\rangle$ for $k \to \infty$, and thus

$$N_k\omega_k \to \langle d\rangle. \qquad \text{[S10]}$$

(This means that the number of links is around $\langle d\rangle/2$.) The degree distribution of a node at position x can be given by a binomial distribution as

$$\rho(d,x) = \binom{N_k}{d} w_k(x)^d [1 - w_k(x)]^{N_k - 1 - d}. \qquad \text{[S11]}$$

In the thermodynamic limit this can be approximated by a Poisson distribution written as

$$\rho(d,x) \simeq \frac{[N_k w_k(x)]^d}{d!} e^{-N_k w_k(x)}. \qquad \text{[S12]}$$

The degree distribution of the whole network is obtained by integrating $\rho(d,x)$, resulting in

$$\rho(d) = \frac{1}{d!}\int_0^1 [N_k w_k(x)]^d e^{-N_k w_k(x)}dx. \qquad \text{[S13]}$$

In particular, the probability that a randomly chosen node will be isolated (having degree zero) is

$$\rho(d=0) = \int_0^1 e^{-N_k w_k(x)}dx. \qquad \text{[S14]}$$

From **[9]** and **[10]** it follows that the average value of $w_k(x)$ is around $\langle d\rangle/N_k$. In case $w_k(x)$ is actually independent of x, then $w_k(x) = \langle d\rangle/N_k$, and

$$\rho(d=0) \simeq e^{-\langle d\rangle}. \qquad \text{[S15]}$$

However, if $w_k(x)$ is such that its typical value is much smaller than its average, then typically $e^{-N_k w_k(x)} \simeq 1$ resulting in

$$\rho(d=0) \simeq 1, \qquad \text{[S16]}$$

which means that the majority of nodes becomes isolated. The condition for avoiding this degeneracy can be formulated as

$$\int_0^1 e^{-N_k w_k(x)}dx < c, \qquad \text{[S17]}$$

where $c < 1$ is a constant.

In case of the multifractal graph generator (or a more general "tensoring" construction), the above condition is not fulfilled, unless $w_k(x)$ is independent of x. As mentioned in the main text, by using a measure preserving bijection between $[0,1]$ and $[0,1]^k$, our model can be formulated in a more general form using the tensorial product $W_k \equiv W^{\otimes k} = W \otimes \cdots \otimes W$ defined as

$$W(x_1,\ldots,x_k,y_1,\ldots,y_k) = W(x_1,y_1)\ldots W(x_k,y_k). \qquad \text{[S18]}$$

The marginals **[8]** in this representation are given by

$$\begin{aligned} w_k(x_1,\ldots,x_k) &= \int_{[0,1]^k} W(x_1,y_1)\ldots W(x_k,y_k)dy_1\ldots dy_k \\ &= w(x_1)\ldots w(x_k), \end{aligned} \qquad \text{[S19]}$$

where $w(x) = \int_0^1 W(x,y)dy$. Similarly, **[9]** is transformed into

$$\omega_k = \int_{[0,1]^k} w_k(x_1,\ldots,x_k)dx_1\ldots dx_k = \omega^k, \qquad \text{[S20]}$$

where $\omega = \int_0^1 w(x)dx$. Thus, according to **[10]**, we should choose $N_k \simeq \langle d\rangle/\omega^k$.

Unfortunately, these functions do not satisfy condition **[17]** unless $w(x)$ is constant. Indeed, if $(x_1,\ldots,x_k,y_1,\ldots,y_k)$ is a random point in $[0,1]^k$, then

$$\ln w_k(x_1,\ldots,x_k) = \ln w(x_1) + \cdots + \ln w(x_k) \sim k\int_0^1 \ln w(x)dx \qquad \text{[S21]}$$

almost surely by the law of large numbers. Let $\alpha = \exp(\int \ln w(x)dx)$, then $\alpha < \omega$ by the Jensen inequality (except if w is constant), and the value of w_k is almost always close to α^k, while its average is ω^k. Since $(\alpha/\omega)^k \to 0$ if $k \to \infty$, this shows that if [**17**] holds, then $w(x)$ is constant.

On the other hand, if $w(x) = \omega$ for any x, then the expected degree of the nodes becomes independent from their position and the degree distribution converges to a Poisson distribution, just like in case of an Erdős–Rényi graph. In this case it is also easy to calculate the number of copies of any connected graph F with l nodes in a graph G_k obtained from W_k. There are $N_k(N_k-1)\ldots(N_k-l+1) \sim N_k^l$ ways to map $V(F)$ into $[n_k]$ injectively, and for each map, the probability that it is a homomorphism is $t(F,W_k) = t(F,W)^k$. Hence

$$\mathsf{E}(\mathrm{inj}(F,G_k)) \sim N_k^l t(F,W)^k = \langle d\rangle^l \left(\frac{t(F,W)}{\omega^l}\right)^k. \qquad \text{[S22]}$$

Since we have a sparse graph, we want to normalize this by N_k; so the normalized number of copies of F is

$$\mathsf{E}(s(F,G_k)) = \frac{\hom(F,G_k)}{N_k} \sim N_k^{l-1} t(F,W)^k = \langle d\rangle^{l-1}\left(\frac{t(F,W)}{\omega^{l-1}}\right)^k. \qquad \text{[S23]}$$

For example, the normalized number of triangles is

$$\frac{1}{6}s(K_3,G_k) \sim \langle d\rangle^2\left(\frac{t(K_3,W)}{\omega^2}\right)^k. \qquad \text{[S24]}$$

It is easy to see that if $w(x) = \omega$ is constant, then

$$t(F,W) \le \omega^{l-1}, \qquad \text{[S25]}$$

where equality holds if and only if F is a tree or W is an equivalence relation such that there is a partition $S_1 \cup \cdots \cup S_m = [0,1]$ and

$$W(x,y) = \begin{cases} 1, & \text{if } x,y \in S_i \quad \text{for some } i, \\ 0, & \text{otherwise.} \end{cases} \qquad \text{[S26]}$$

From [**23**] and [**25**] we gain

$$\mathsf{E}(s(F,G_k)) \to \begin{cases} \langle d\rangle^{l-1}, & \text{if } F \text{ is a tree or } W \text{ is an equivalence relation,} \\ 0, & \text{otherwise.} \end{cases} \qquad \text{[S27]}$$

Using high concentration inequalities one can prove that this convergence happens almost surely (not just in expectation). We see from [**27**] that the sequence G_k is convergent with probability 1 in the Benjamini–Schramm sense.

There are a number of possibilities which one could try to cure the degeneracy of the thermodynamic limit of our model shown here; however, these are out of the scope of the present study. We could modify the tensoring construction by adding to $W^{\oplus k}$ a constant c_k tending to 0 reasonably slowly. Another possibility is to modify $W^{\oplus k}$ to $(W^{\oplus k})^{a_k}$, where $a_k \to 0$.

Numerical Studies. ***Fraction of isolated nodes.*** Next, let us investigate the magnitude of the above effect for graphs in the size range of real networks. For this purpose we generated networks from randomly chosen generating measures (with $m = 4$, equal-sized boxes) iterated from $k = 1$ to $k = 11$. The number of nodes at $k = 1$ was set to 1,000 and to 5,000, respectively, and for $k > 1$ it was adjusted using Eq. **3** in the main text. (Thus the average degree of the graphs remained the same during the iterations.) In Fig. 2. we show the results obtained by averaging over 1,000 samples for both settings by plotting $\rho(d = 0)$ as a function of N. In spite of the increasing tendency of the curves, at the last iteration with network sizes above 10^9, the fraction of isolated nodes is still very low. Thus, the effect of isolated nodes becoming dominant is negligible on the scale of real world applications.

Absence of convergence in the realistic size range. In spite of the analytical results for the convergence in the thermodynamic limit, the degree distribution often shows an oscillatory behavior in the size range of real networks. This is shown in Fig. 3 for an $m = 3$ generating measure iterated from $k = 1$, $N = 100$ to $k = 11$, $N = 4.24 \cdot 10^8$. As k is getting larger, the more oscillations can be observed in $\rho(d)$ towards the large degrees.

Calculating Statistical Distributions. A serious advantage of our model is that the statistical properties characterizing the network topology can be calculated analytically. In the *Appendix* of the main text we give a derivation for the degree distribution, based on the generating-function formalism. The definition and the most important properties of the generating functions can be summarized as follows.

The generating functions. If a random variable ξ can take nonnegative integer values according to some probability distribution $\mathcal{P}(\xi = n) \equiv \rho(n)$, then the corresponding generating function is given by

$$G_\rho(x) \equiv \langle x^\xi\rangle = \sum_{n=0}^{\infty}\rho(n)x^n. \qquad \text{[S28]}$$

The generating function of a properly normalized distribution is absolute convergent for all $|x| \le 1$ and hence has no singularities in this region. For $x = 1$ it is simply

$$G_\rho(1) = \sum_{n=0}^{\infty}\rho(n) = 1. \qquad \text{[S29]}$$

The original probability distribution and its moments can be obtained from the generating function as

$$\rho(n) = \frac{1}{n!}\frac{d^n G_\rho(x)}{dx^n}\bigg|_{x=0}, \qquad \text{[S30]}$$

$$\langle \xi^l\rangle = \sum_{n=0}^{\infty} n^l\rho(n) = \left[\left(x\frac{d}{dx}\right)^l G_\rho(x)\right]_{x=1}. \qquad \text{[S31]}$$

Finally, if $\eta = \xi_1 + \xi_2 + \cdots + \xi_I$, where $\xi_1, \xi_2, \ldots, \xi_I$ are independent random variables (with nonnegative integer values), then the generating function corresponding to $\mathcal{P}(\eta = n) \equiv \sigma(n)$ is given by

$$G_\sigma(x) = \langle x^\eta\rangle = \left\langle \prod_{i=1}^{I} x^{\xi_i}\right\rangle = \prod_{i=1}^{I}\langle x^{\xi_i}\rangle = G_{\rho_1}(x)G_{\rho_2}(x)\ldots G_{\rho_I}(x). \qquad \text{[S32]}$$

The clustering coefficient. Similarly to the degree distribution, the clustering coefficient of nodes falling into the same row of the link probability measure is expected to be the same. The clustering coefficient of a node in row i can be obtained by calculating the number of triangles containing the node, divided by the number of link pairs originating from the node. Since the triangles are equivalent to link pairs originating from the node having their other end connected by a third link, the expected clustering coefficient of a node in row i can be given as

$$\langle C_i(k)\rangle = \frac{\frac{1}{2}\sum_{j=1}^{m^k}[l_j(k)]^2[p_{ij}(k)]^2 p_{jj}(k) + \sum_{j=1}^{m^k}\sum_{q=j+1}^{m^k} l_j(k)l_q(k)p_{ij}(k)p_{iq}(k)p_{jq}(k)}{\frac{1}{2}\sum_{j=1}^{m^k}[l_j(k)]^2[p_{ij}(k)]^2 + \sum_{j=1}^{m^k}\sum_{p=j+1}^{m^k} l_j(k)l_q(k)p_{ij}(k)p_{iq}(k)}. \quad [S33]$$

The first term in both the numerator and the denominator corresponds to the link pairs for which the other end of the links point to the same row j, whereas the second terms give the contribution from link pairs connecting our node in row i to distinct rows j and q.

The average nearest neighbors degree. Finally, we mention that the degree correlations can be calculated from $p_{ij}(k)$ [and $l_i(k)$] as well. Here we derive the expression for the average nearest neighbors degree, d_{NN}, as a function of the node degree. This is one of the most simplest quantity characterizing the degree correlations: An increasing curve corresponds to an assortative network, whereas a decreasing one signals disassortative behavior. The average degree of the neighbors of a node from row i can be given as

$$d_{NN,i}^{(k)} = \frac{\sum_{j=1}^{m^k}\widehat{p}_{ij}(k)l_j(k)\langle d_j(k)\rangle}{\sum_{j=1}^{m^k}p_{ij}(k)l_j(k)}. \quad [S34]$$

The average degree of the neighbors of a node with degree d can be given as a sum over the possible $d_{NN,i}^{(k)}$, multiplied by the conditional probability $p^{(k)}(i|d)$ that the node is from row i, given that its degree is d:

$$d_{NN}^{(k)}(d) = \sum_{i=1}^{m^k} p^{(k)}(i|d)d_{NN,i}^{(k)}. \quad [S35]$$

These conditional probabilities can be obtained as follows. The number of nodes from row i with degree d is $n_i(k)\rho_i^{(k)}(d)$, whereas the total number of nodes with degree d is $n\rho^{(k)}(d)$. The probability that a node is from row i given that its degree is d is the ratio of these two:

$$p^{(k)}(i|d) = \frac{n_i(k)\rho_i^{(k)}(d)}{n\rho^{(k)}(d)} = \frac{l_i(k)\rho_i^{(k)}(d)}{\rho^{(k)}(d)}. \quad [S36]$$

By substituting [**36**] and [**34**] into [**35**] we get

$$d_{NN}^{(k)}(d) = \frac{1}{\rho^{(k)}(d)}\sum_{i=1}^{m^k} l_i(k)\rho_i^{(k)}(d)\frac{\sum_{j=1}^{m^k}p_{ij}(k)l_j(k)\langle d_j(k)\rangle}{\sum_{j=1}^{m^k}p_{ij}(k)l_j(k)}. \quad [S37]$$

1. Lovász L, Szegedy B (2006) Limits of dense graph sequences. *J Comb Theory B* 96:933–957.
2. Borgs C, Chayes J, Lovász L, Sós VT, Vesztergombi K (2008) Convergent sequences of dense graphs I: Subgraph frequencies, metric properties and testing. *Adv Math* 219:1801–1851.
3. Bianconi G (2008) The entropy of randomized network ensembles. *Europhys Lett* 81:28005.

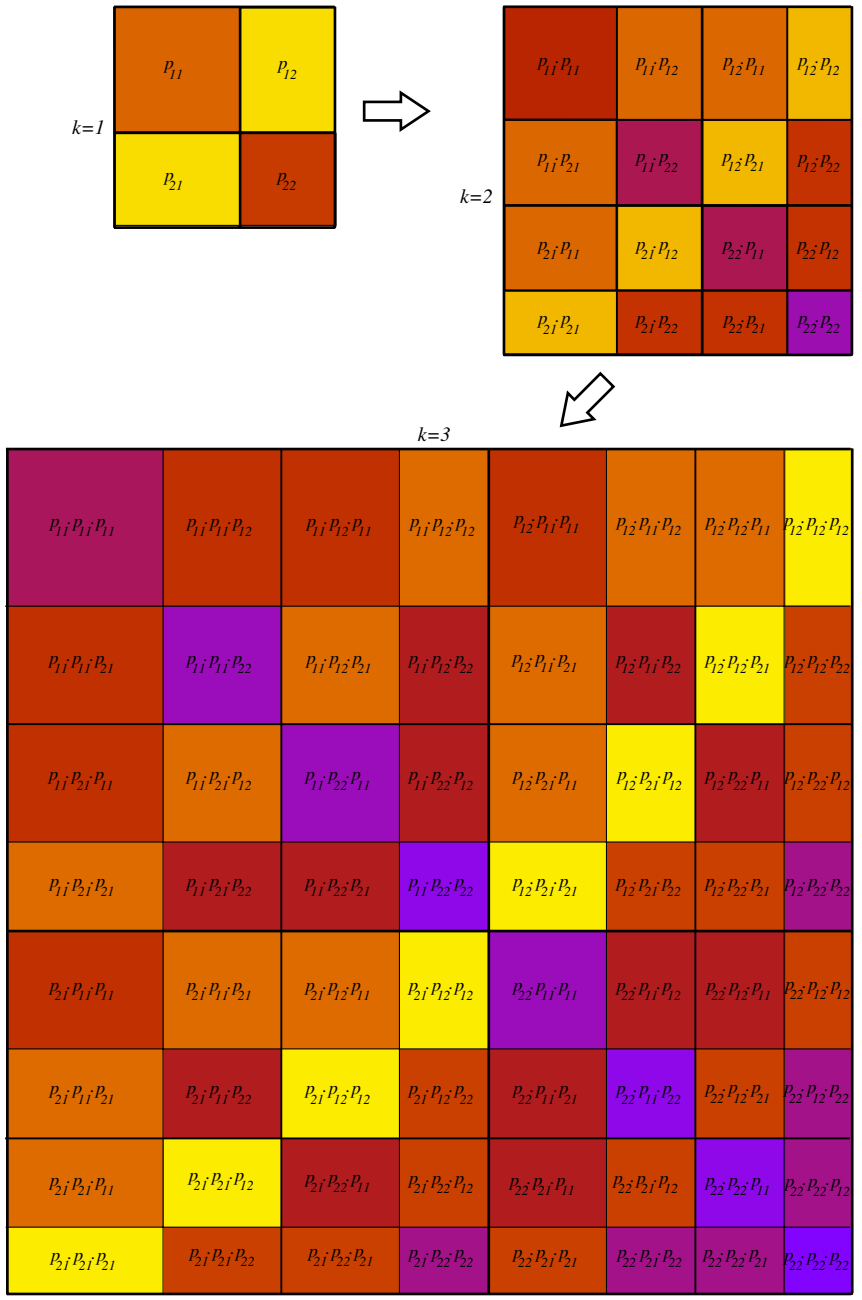

Fig. S1. The link probability measure shown in Fig. 1 of the main text with the explicit expressions of $p_{ij}(k)$ displayed as well. The generating measure is symmetric, and thus $p_{21} = p_{12}$.

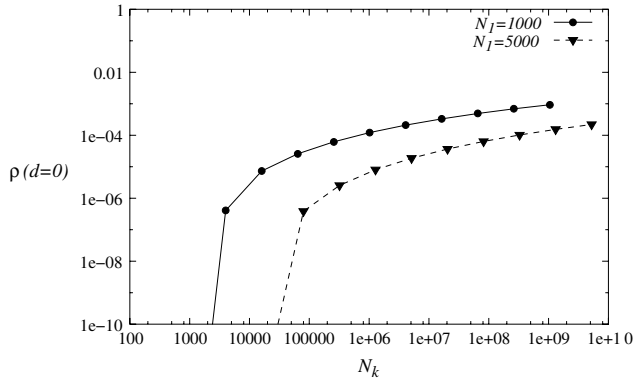

Fig. S2. The ratio of isolated nodes in function of the network size for graphs obtained from iterating random generating measures from $k = 1$ to $k = 11$, averaged over 1,000 samples. The number of nodes at iteration $k = 1$ was chosen to be $N(k = 1) = 1,000$ (circles) and $N(k = 1) = 5,000$ (triangles).

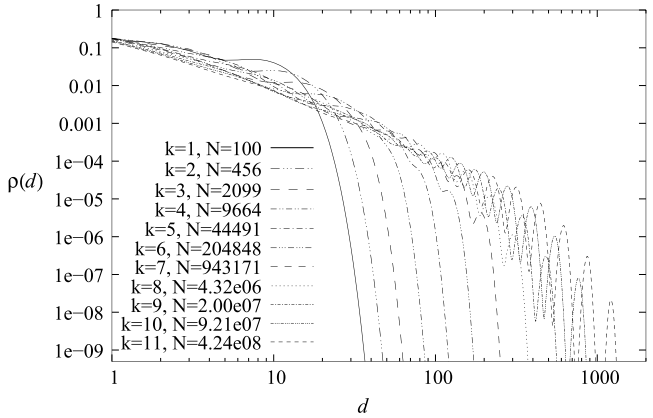

Fig. S3. The degree distribution obtained from Eq. **2** in the main text for an $m = 3$ generating measure iterated from $k = 1$, $N = 100$ to $k = 11$, $N = 4.24 \cdot 10^8$.