Subgraphs in random networks

S. Itzkovitz,[1,2] R. Milo,[1,2] N. Kashtan,[2,3] G. Ziv,[1] and U. Alon[1,2]

[1]*Department of Physics of Complex Systems, Weizmann Institute of Science, Rehovot 76100, Israel*
[2]*Department of Molecular Cell Biology, Weizmann Institute of Science, Rehovot 76100, Israel*
[3]*Department of Computer Science and Applied Mathematics, Weizmann Institute of Science, Rehovot 76100, Israel*

Understanding the subgraph distribution in random networks is important for modeling complex systems. In classic Erdős networks, which exhibit a Poissonian degree distribution, the number of appearances of a subgraph G with n nodes and g edges scales with network size as $\langle G \rangle \sim N^{n-g}$. However, many natural networks have a non-Poissonian degree distribution. Here we present approximate equations for the average number of subgraphs in an ensemble of random sparse directed networks, characterized by an arbitrary degree sequence. We find scaling rules for the commonly occurring case of directed scale-free networks, in which the outgoing degree distribution scales as $P(k) \sim k^{-\gamma}$. Considering the power exponent of the degree distribution, γ, as a control parameter, we show that random networks exhibit transitions between three regimes. In each regime, the subgraph number of appearances follows a different scaling law, $\langle G \rangle \sim N^{\alpha}$, where $\alpha = n - g + s - 1$ for $\gamma < 2$, $\alpha = n - g + s + 1 - \gamma$ for $2 < \gamma < \gamma_c$, and $\alpha = n - g$ for $\gamma > \gamma_c$, where s is the maximal outdegree in the subgraph, and $\gamma_c = s + 1$. We find that certain subgraphs appear much more frequently than in Erdős networks. These results are in very good agreement with numerical simulations. This has implications for detecting network motifs, subgraphs that occur in natural networks significantly more than in their randomized counterparts.

I. INTRODUCTION

Many natural systems are described as networks of interacting components [1–9]. Random networks have been studied as models of these complex systems. The classic model for a random network is the Erdős model [10–13], in which each of the possible edges in the network exists with probability p. There exists an analytical solution to many of the properties of Erdős networks, such as the diameter, clustering coefficient, component size distributions, and subgraph distributions [10–13]. The average number of appearances G of a subgraph with n nodes and g edges in a directed network of N nodes is

$$\langle G \rangle = \lambda \binom{N}{n} p^g (1-p)^{n(n-1)-g} \sim \lambda N^n \left(\frac{\langle K \rangle}{N} \right)^g,$$

$$\langle G \rangle \sim N^{n-g}, \tag{1}$$

assuming a fixed mean connectivity $\langle K \rangle = pN$. λ is a term of order 1 which stems from the symmetry of each subgraph. Erdős networks have been extensively used as models for analyzing real networks. An excellent example is the work of Davis, Holland, and Leinhardt on subgraphs in social networks [14–17].

Erdős networks exhibit a Poissonian degree distribution: the distribution of the number of edges per node is $P(k) = \langle k \rangle^k e^{-\langle k \rangle} / k!$. Nodes with a number of edges much higher than the mean are exponentially rare. Many naturally occurring networks, on the other hand, obey a long-tailed degree sequence, often described as a power law, $P(k) \sim k^{-\gamma}$, with γ often between 2 and 3 [18–29]. These networks, termed scale-free networks, are characterized by the existence of nodes with high degree, termed hubs (Fig. 1). The existence

of hubs dramatically influences the properties of these networks. Some of the global properties of random networks with arbitrary degree distribution, and specifically scale-free networks, have been calculated. These include sizes of connected components [3,30,31], distances [32], percolation thresholds [33–35], and clustering coefficients [36–38].

There is much current interest in the local structure of networks [5,7–9,36,39–46,53]. Recently, subgraph structure was analyzed in biological and technological networks [7,8]. It was found that these natural or designed networks contain network motifs, subgraphs that occur much more often than in an ensemble of randomized networks with the same degree sequence. In biological networks, the network motifs were suggested to be elementary building blocks which carry out key information processing functions [7,8]. In these studies, random networks generation and the enumeration of their subgraphs were performed numerically. To complement this numerical work, it would be important to theoretically characterize the subgraph distribution of random networks. Here we present approximate formulas for the average number of subgraphs in an ensemble of random networks with an
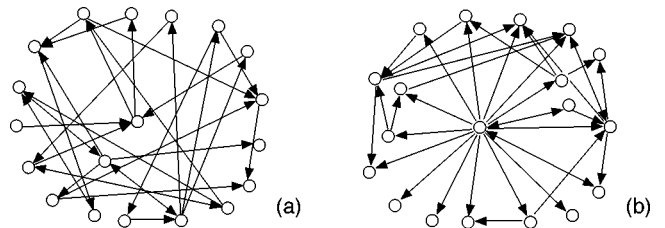

FIG. 1. Example of (a) Erdős network and (b) scale-free network ($\gamma = 2$). Mean connectivity is 1.85 in both. Notice the hub in the scale-free network.
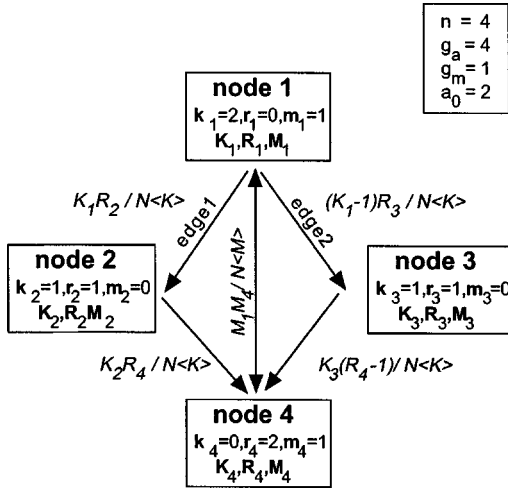
FIG. 2. A subgraph with one mutual edge and four single edges. The subgraph degree sequences $\{k_i, r_i, m_i\}$ and node degrees $\{K_i, R_i, M_i\}$ are displayed in bold. Edge probabilities are displayed in plain. Using Eq. (5), the mean subgraph number of appearances in an ensemble of random networks is $\langle G \rangle = 2 \langle K(K-1)M \rangle \langle R(R-1)M \rangle \langle RK \rangle^2 / N \langle K \rangle^4 \langle M \rangle$.

arbitrary degree sequence. In the random ensemble each node has a specified indegree, outdegree, and mutual degree. These formulas give a very good approximation for random networks which allow for multiple edges between nodes (more than one edge in a given direction), as in the well-studied configuration model [13,31,32,47,48]. We also show that they provide a reasonable approximation for networks where multiple edges are not allowed, which represent more realistically many naturally occurring networks. We apply these formulas to arrive at scaling laws for networks with a scale-free degree distribution. We find that each subgraph has its own scaling exponent, influenced by its topology. Considering the power exponent of the degree distribution, γ, as a control parameter, we show that the random networks exhibit transitions between three regimes. In each regime, the subgraph number of appearances follows a different scaling law. We find that certain subgraphs appear much more frequently than in Erdős networks.

II. NUMBER OF SUBGRAPHS: APPROXIMATE SOLUTION

The following approximation assumes sparse networks ($\langle K \rangle \ll N$). The network degree sequence is given by the outdegree $\{K_i\}_{i=1}^N$ (the number of edges outgoing from each node), indegree $\{R_i\}_{i=1}^N$ (the number of incoming edges at each node), and mutual degree $\{M_i\}_{i=1}^N$ (the number of mutual edges at each node). Mutual edges are cases where there is a pair of edges in both directions between two nodes. This property has been studied in social networks [14–17] and in the world-wide web (www) [39]. We begin by computing the probability of obtaining an n-node subgraph with g_a single edges, g_m mutual edges, subgraph outdegree sequence $\{k_j\}_{j=1}^n$, subgraph indegree sequence $\{r_j\}_{j=1}^n$, and subgraph mutual degree sequence $\{m_j\}_{j=1}^n$ in a given set of nodes.

Consider the example of Fig. 2. The probability of obtaining a directed edge from node 1 to node 2 is approximately

$$P(\text{edge } 1) = \frac{K_1 R_2}{N \langle K \rangle}, \qquad (2)$$

assuming $K_1 R_2 \ll N \langle K \rangle$ (see Appendix A). The probability of obtaining a second edge from node 1 to node 3 is

$$P(\text{edge } 2 | \text{edge } 1) = \frac{(K_1 - 1)R_3}{N \langle K \rangle}. \qquad (3)$$

This reasoning applies to all the subgraph edges. The mean number of appearances of a subgraph is found by taking the average of the resulting expression, with respect to all choices of n distinct nodes $\{\sigma_1 \cdots \sigma_n\}$, and multiplying by the number of possible choices of n nodes out of N:

$$\langle G \rangle = \frac{a N^{n - g_a - g_m}}{\langle K \rangle^{g_a} \langle M \rangle^{g_m}} \left\langle \prod_{j=1}^n \binom{K_{\sigma_j}}{k_j} \binom{R_{\sigma_j}}{r_j} \binom{M_{\sigma_j}}{m_j} \right\rangle_{\{\sigma\}}, \qquad (4)$$

where $\langle K \rangle$ is the average outdegree (equals the average indegree $\langle R \rangle$) and $\langle M \rangle$ is the average mutual edge degree. The symmetry factor a is $a_0^{-1} \Pi_{j=1}^n k_j! r_j! m_j!$, where a_0 is the number of different permutations of the nodes that give an isomorphic subgraph.

The average (4) reduces to a product of moments of different orders of the indegree, outdegree, and mutual degree distributions:

$$\langle G \rangle = \frac{a N^{n - g_a - g_m}}{\langle K \rangle^{g_a} \langle M \rangle^{g_m}} \prod_{j=1}^n \left\langle \binom{K_i}{k_j} \binom{R_i}{r_j} \binom{M_i}{m_j} \right\rangle_i, \qquad (5)$$

where the fact that each node should participate in the summation of only one term j introduces higher-order corrections which we neglect. For example, subgraph id102 (Table I), has $n = 3$ nodes, $g_a = 2$ single edges, and $g_m = 1$ mutual edge. The subgraph degree sequences are $k_j = \{1,1,0\}$, $r_j = \{0,1,1\}$, and $m_j = \{1,0,1\}$. Using Eq. (5), we find

$$\langle G \rangle = \langle \text{id102} \rangle = \frac{\langle KM \rangle \langle RM \rangle \langle RK \rangle}{\langle K \rangle^2 \langle M \rangle}. \qquad (6)$$

The approximation [Eq. (5)] is exact in the case of Erdős networks. In Erdős networks, both indegree and outdegree are Poisson distributed and independent, and Eq. (5) reduces to Eq. (1).

For nonsparse networks, a more accurate approximation takes into account the probabilities of a nonexistent edge between two nodes (see Appendix B).

We tested the equations on random networks taken with the degree sequence of real-world networks—transcription interactions in the yeast *S. cerevisiae* [8], synaptic connections between neurons in *C. elegans* [49], and world-wide-web hyperlinks between web pages in a single domain [18]. When multiple edges in the same direction are allowed, as in the configuration model, Eq. (5) is within a few percent of the numerical simulation results (Table I). We have also

TABLE I. Mean numbers of the thirteen connected directed subgraphs in an ensemble of random networks with a given degree distribution. The degree distributions are those of transcription in the yeast *S. cerevisiae* [8], synaptic connections between neurons in *C. elegans* [19], and world-wide-web hyperlinks between web pages in a single domain [18]. Shown are the theoretical values [Eq. (5)]. The values in parentheses are the percent deviations of the direct enumeration results—using the algorithms described in Ref. [8], where 1000 random networks with the same degree distributions as those of the real networks were generated and all subgraphs were counted. The left value is the percent deviation in an ensemble which allows for multiple edges, and the right value shows the deviation for an ensemble which does not allow multiple edges. Values below 0.5 were rounded to zero. In subgraphs marked with *, the theoretical values shown were obtained using the correction of Appendix B to the table equations. Subgraph id is determined by concatenating the rows of the subgraph adjacency matrix and representing the resulting vector as a binary number. The id is the minimal number obtained from all the isomorphic versions of the subgraph.

subgraph id	equation	transcription	neurons	www
6 *	$N\langle K(K-1)\rangle/2$	$1.2\times10^4(-0.16\%/-0.02\%)$	$4.3\times10^2(+2\%/+8\%)$	$4.7\times10^7(+0.06\%/+0.5\%)$
12 *	$N\langle KR\rangle$	$3.6\times10^2(+0.16\%/-0.1\%)$	$8.7\times10^2(+2.7\%/+3.0\%)$	$2.5\times10^6(+9\%/+10\%)$
14 *	$N\langle KM\rangle$	$1.9\times10^1(-0.06\%/-0.06\%)$	$8.7\times10^1(-0.15\%/+1.9\%)$	$3.8\times10^6(-0.2\%/-0.3\%)$
36 *	$N\langle R(R-1)\rangle/2$	$9.6\times10^2(-2\%/-0.03\%)$	$6.0\times10^3(-0.4\%/+0.7\%)$	$2.2\times10^8(+0.01\%/+0.1\%)$
38	$\langle K(K-1)\rangle\langle RK\rangle\langle R(R-1)\rangle/\langle K\rangle^3$	$1.3\times10^1(+1.6\%/+2.1\%)$	$1.2\times10^2(+0.6\%/-28\%)$	$3.4\times10^5(+0.7\%/-74\%)$
46	$\langle KM\rangle^2\langle R(R-1)\rangle/2\langle K\rangle^2\langle M\rangle$	$0(0\%/0\%)$	$9.3\ \ (-10\%/-57\%)$	$8.5\times10^3(-0.02\%/+8.8\%)$
74 *	$N\langle RM\rangle$	$2.9(-1.2\%/-1.8\%)$	$1.3\times10^2(+1.1\%/+1.2\%)$	$4.8\times10^6(-0.01\%/-0.01\%)$
78 *	$N\langle M(M-1)\rangle/2$	$0(0\%/0\%)$	$6.6\ \ (-0.2\%/-0.5\%)$	$2.5\times10^7(-0.4\%/-0.4\%)$
98	$\langle KR\rangle^3/3\langle K\rangle^3$	$0(0\%/0\%)$	$4.5\ \ (-40\%/-39\%)$	$3.3\times10^1(-31\%/-26\%)$
102	$\langle KM\rangle\langle RM\rangle\langle RK\rangle/\langle K\rangle^2\langle M\rangle$	$0(0\%/0\%)$	$2\ \ (-22\%/-15\%)$	$1.4\times10^2(-11\%/-4\%)$
108	$\langle RM\rangle^2\langle K(K-1)\rangle/2\langle K\rangle^2\langle M\rangle$	$0(0\%/0\%)$	$1.4\ \ (-18\%/-6\%)$	$2.9\times10^3(-11\%/-44\%)$
110	$\langle KM\rangle\langle RM\rangle\langle M(M-1)\rangle/\langle K\rangle\langle M\rangle^2$	$0(0\%/0\%)$	$0(0\%/0\%)$	$2.3\times10^3(-1.8\%/-4\%)$
238	$\langle M(M-1)\rangle^3/6\langle M\rangle^3$	$0(0\%/0\%)$	$0(0\%/0\%)$	$5\times10^4(-0.04\%/-3.6\%)$

simulated random networks in which only one edge was allowed in each direction between any two nodes. As can be seen in Table I, Eq. (5) is still within a few percent of the numerical simulation results for most subgraphs. There are some discrepancies (most notably a factor of almost 4 for subgraph id38 in the randomized world wide web network). In addition, we find good agreement between our approximation and numerical enumeration of subgraphs in simulated random networks with scale-free outdegree (Fig. 3).

III. SCALE-FREE NETWORKS

Scale-free networks have degree distributions that follow $P(k)\sim k^{-\gamma}$ at large k [18–29]. We consider directed networks in which the outgoing edge degree is scale free, while the incoming edge distribution is Poissonian. Our results can be easily extended to scale-free indegree. For simplicity, we choose the following form for the outgoing degree distribution for a network with N nodes (this function was used in Ref. [3] to fit world-wide-web data):

$$P(k)=\frac{\gamma-1}{k_0^{1-\gamma}}(k+k_0)^{-\gamma}, \quad k<N. \tag{7}$$

The mean connectivity $\langle K\rangle$ is determined by k_0.

The hub is the node with the maximal number of outgoing edges, T. The hub size distribution (Fig. 4) is

$$P(T)=NP(k=T)[P(k\leq T)]^{N-1}$$
$$=\frac{N(\gamma-1)}{k_0}(T/k_0)^{-\gamma}[1-(T/k_0)^{-\gamma+1}]^{N-1}, \tag{8}$$

assuming $T\gg k_0$. For $2<\gamma<3$, the mean hub scales as

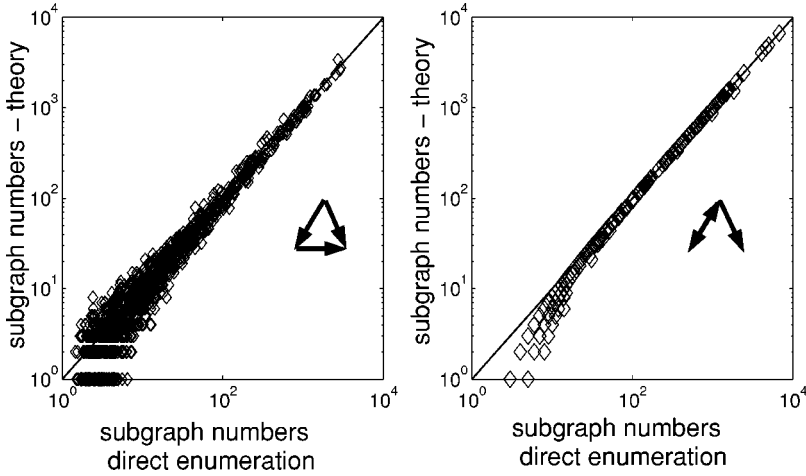$$\langle T \rangle = \int_1^{N-1} T P(T) dT \sim N^{1/(\gamma-1)}, \qquad (9)$$

where the mean is over an ensemble of random networks with the same γ and mean connectivity (see also Refs. [28,34] for an alternative method of deriving this result). At $\gamma \leq 2$, there is a condensation effect [50], where a finite fraction of the nodes have outdegree ≤ 1 and the mean hub size becomes proportional to N. Using Eq. (5), and assuming a compact distribution for the number of mutual edges, we find that the subgraph distribution is dominated by the hubs, and that the dominant term is that of the subgraph node with maximal outdegree s. The number of appearances of each subgraph can be shown to scale as

$$\langle G \rangle \sim a N^{n-g-1} \langle K \rangle^{g-s} \sum_{i=1}^{N} \binom{K_i}{s} \sim N^{\alpha}, \qquad (10)$$

where $g = g_a + 2g_m$ is the total number of edges in the subgraph [51]. We derive the scaling exponent α in the following section.
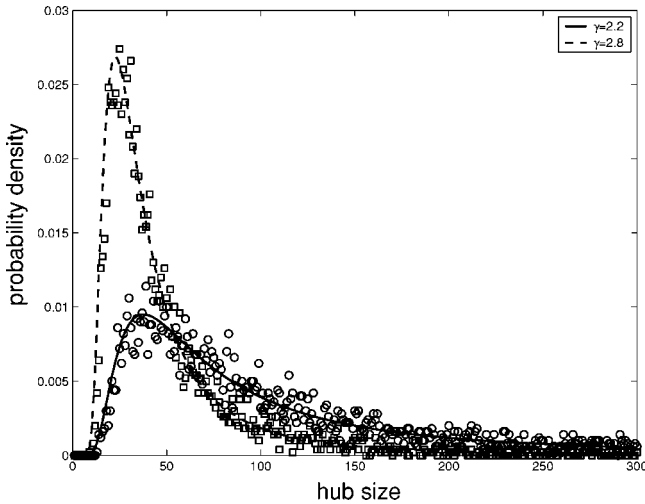

FIG. 4. Simulated and theoretical hub distribution for networks with $N=3000$ nodes, $\gamma=2.2$ ($\bigcirc$) or $\gamma=2.8$ ($\square$), and mean connectivity $\langle K \rangle = 1.2$. Lines represent theoretical calculations (8).

IV. TRANSITIONS AT DIFFERENT γ

The subgraph numbers scale as

$$\langle G \rangle \sim N^{\alpha}. \qquad (11)$$

We find three different regimes, in each of which the scaling exponent α behaves differently. Taking an ensemble average by integrating the largest term in Eq. (10) over the hub distribution (8), we get

$$\langle G \rangle \sim N^{n-g-1} \int_1^{N-1} T^s P(T) dT. \qquad (12)$$

For $\gamma \leq 2$, the network is in a condensed regime, where the hub $T = O(N)$. In this regime

$$\langle G \rangle \sim N^{n-g+s-1}. \qquad (13)$$

For $2 < \gamma < \gamma_c$, substituting Eq. (8) in Eq. (12) yields

$$\langle G \rangle \sim N^{n-g+s-\gamma+1}. \qquad (14)$$

In this regime, the tail of $P(T)$ is the dominant contribution to the integral. Finally, at values above a critical γ, another transition occurs, where α equals the scaling exponent in Erdős networks, $\alpha = n - g = \alpha_{Erdos}$. The critical γ is γ_c:

$$\gamma_c = s + 1. \qquad (15)$$

In this regime, the hubs no longer contribute significantly to the subgraph distribution. In summary, $\langle G \rangle \sim N^{\alpha}$, where α is

$$\alpha = \begin{cases} n-g+s-1, & \gamma \leq 2 \\ n-g+s-\gamma+1, & 2 < \gamma < s+1 \\ n-g, & \gamma \geq s+1. \end{cases} \qquad (16)$$

Table II shows the expected scaling exponent for the 13 connected directed 3-node subgraphs, as well as for several 4-node subgraphs. The scaling laws agree very well with the numerical results (Fig. 5). The three regimes of scaling are clearly seen. Note that the topology of each subgraph affects its scaling through the subgraph maximal outdegree s. These results can be easily extended to the case of scale-free inde-

TABLE II. The scaling exponent α of subgraph numbers for random scale-free networks with outgoing degree exponent γ. The subgraph numbers scale as $\langle G \rangle \sim N^{\alpha}$. Shown are all thirteen 3-node connected directed subgraphs and four examples of 4-node subgraphs. n is the number of nodes in the subgraph, g, is the number of edges, and s is the maximal degree within the subgraph. The exponent α has three regimes : α_{Erdos} in the "Erdős regime," when $\gamma > \gamma_c$; α_{sf} in the "scale-free regime," when $2 < \gamma < \gamma_c$; and α_{cond} in the "condensed regime," when $\gamma \lesssim 2$.

subgraph id		n nodes	g edges	s	α_{erdos} $\gamma > \gamma_c$	α_{sf} $2 < \gamma < \gamma_c$	α_{cond} $\gamma \le 2$	γ_c
6		3	2	2	1	$4 - \gamma$	2	3
12		3	2	1	1	1	1	2
14		3	3	2	0	$3 - \gamma$	1	3
36		3	2	1	1	1	1	2
38		3	3	2	0	$3 - \gamma$	1	3
46		3	4	2	-1	$2 - \gamma$	0	3
74		3	3	1	0	0	0	2
78		3	4	2	-1	$2 - \gamma$	0	3
98		3	3	1	0	0	0	2
102		3	4	2	-1	$2 - \gamma$	0	3
108		3	4	2	-1	$2 - \gamma$	0	3
110		3	5	2	-2	$1 - \gamma$	-1	3
238		3	6	2	-3	$-\gamma$	-2	3
subgraph id		n nodes	g edges	s	α_{erdos} $\gamma > \gamma_c$	α_{sf} $2 < \gamma < \gamma_c$	α_{cond} $\gamma \le 2$	γ_c
14		4	3	3	1	$5 - \gamma$	3	4
204		4	4	2	0	$3 - \gamma$	1	3
206		4	5	3	-1	$3 - \gamma$	1	4
2190		4	5	3	-1	$3 - \gamma$	1	4

gree and nondirected networks. For loops of any size in nondirected networks, the critical γ is $\gamma_c = 3$. At $\gamma > 3$, loop numbers scale as N^0. This is consistent with Ref. [45], which shows logarithmic corrections for the number of loops in Barabasi-Albert scale-free networks which have $\gamma = 3$.

V. DISCUSSION

To summarize, we have presented an approximate solution for the average number of directed connected subgraphs in an ensemble of random networks with arbitrary degree sequence. We have presented scaling formulas for the number of subgraphs in scale-free random networks, and showed that the subgraph numbers can be very different from those in Erdős random networks. In Erdős random networks, the scaling exponent is strictly determined by the number of nodes and edges of the subgraph, whereas in scale-free random networks the exact topology of the subgraph determines the scaling exponent. We showed that the scaling exponent α exhibits three different scaling laws in three regimes, depending on the control parameter γ (the power of the degree distribution). In the common case of scale-free networks with γ between 2 and 3, there are many more subgraphs which contain a node connected to more than one other node than in the corresponding Erdős networks with the same mean connectivity. For example, the feed-forward loop (id38 in Table I) is much more common for $\gamma < 3$. At $\gamma = 2.5$, the number of feed-forward loops scales as $N^{0.5}$, as opposed to N^0 in Erdős networks. On the other hand, subgraphs such as the 3-node cycle (id98 in Table I) have the same scaling, N^0, as in Erdős networks.

This study adds to our understanding of the random network models to which real-world networks are compared. It highlights the importance of using random networks that preserve the single and mutual degree sequence of the real network. Our approach may be readily extended to networks with multiple colors of edges. The present results may be useful for enumerating subgraphs in very large random networks, which are beyond the reach of current numerical algorithms.

APPENDIX A: EDGE PROBABILITIES

Here we give a more detailed derivation for the edge probabilities used in Eqs. (2) and (3). Without loss of generality, we treat a network with no mutual edges. We denote by $E = N \langle K \rangle$ the total number of edges. We begin by calculating the probability that no edge connects a source node with K outgoing edges and a target node with R incoming edges. This happens when all K edges connect to a set of nodes $\{\sigma_i\}_{i=1}^k$ which does not contain the target node:

$$p(\text{no edge}|\{\sigma_i\}) = \prod_{k=0}^{K-1} \left(1 - \frac{R}{E - R' - \sum_{i=1}^{k} R_{\sigma_i}} \right),$$
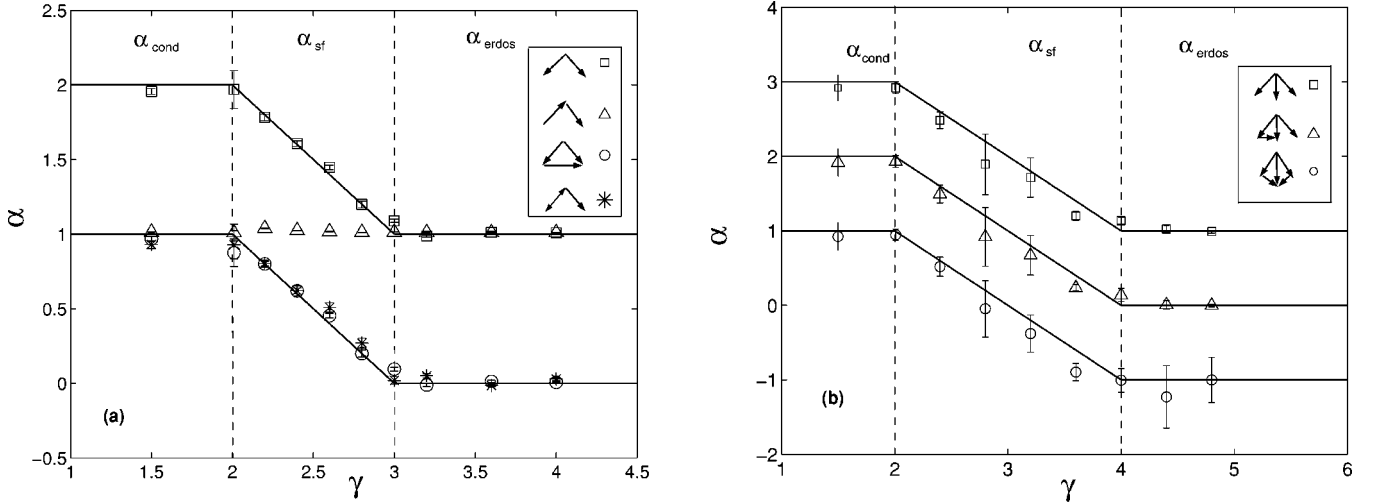
(A1)

FIG. 5. Scaling exponent of 3-node subgraphs (a) and 4-node subgraphs (b) as a function of γ. The exponent α was obtained from the slope of a log-log fit of the number of subgraphs vs network size, for nine different network sizes (30, 100, 300, 500, 1000, 1500, 2000, 2500, 3000) averaged over 5000 randomized networks for each size and outdegree power law γ. All the networks had mean connectivity $\langle K \rangle = 1.2$. The exponent α displays three regimes, $\gamma < 2$ (the condensed regime), $2 < \gamma < \gamma_c$ (the scale-free regime), $\gamma > \gamma_c$ (Erdős regime).

where R' is the indegree of the source node (we do not allow self-edges). The probability of having no edge is obtained by summing over all possible sets $\{\sigma_i\}_{i=1}^{k}$:

$$p(\text{no edge}) = \frac{1}{K!\binom{N-2}{K}} \sum_{\{\sigma\}} \prod_{k=0}^{K-1} \left(1 - \frac{R}{E - R' - \sum_{i=1}^{k} R_{\sigma_i}} \right).$$

$$(A2)$$

Assuming $\max \Sigma_{i=1}^{k} R_{\sigma_i} \ll E$, and taking the complement as the probability of an edge existing, we obtain

$$p(\text{edge}) = 1 - \left(1 - \frac{R}{N\langle K \rangle} \right)^{K} = 1 - e^{-KR/N\langle K \rangle} \sim \frac{KR}{N\langle K \rangle},$$

$$(A3)$$

where our last approximation assumes $KR \ll N\langle K \rangle$. Intuitively, this result can be understood as K attempts for the source node to connect to the target node with a probability of $R/N\langle K \rangle$ at each attempt. $R/N\langle K \rangle$ is the probability of an arbitrary edge connecting into the target node. Pairs of nodes in which KR is of the order of $N\langle K \rangle$ will contribute multiple edges in the same direction in the approximation, leading to overestimation of subgraph numbers in the simulated networks where multiple edges are not allowed (Table I).

APPENDIX B: NONSPARSE NETWORKS

In calculating the number of appearances of subgraphs in nonsparse networks, a more accurate approximation takes into account the probabilities of a nonexistent edge between two nodes. For such subgraphs, in addition to the specified subgraph, Eq. (5) counts a set of subgraphs, with the null edges replaced by single or mutual edges. The corrections for the 3-node subgraphs are

$$\langle \text{id6*} \rangle = \langle \text{id6} \rangle - \langle \text{id38} \rangle - \langle \text{id108} \rangle,$$

TABLE III. Matrix formulas for the numbers of all 3-node connected directed subgraphs. M is the adjacency matrix, S is its symmetric component, and A its asymmetric component. A' is the transposed matrix, $\widetilde{A}$ is the logical inverse of matrix A, tr$A$ is the matrix trace.

subgraph	id	formula
6		$(\sum A'A \cdot \widetilde{M} \cdot \widetilde{M}' - trA'A)/2$
12		$\sum A^2 \cdot \widetilde{M} \cdot \widetilde{M}'$
14		$\sum SA \cdot \widetilde{M} \cdot \widetilde{M}'$
36		$(\sum AA' \cdot \widetilde{M} \cdot \widetilde{M}' - trAA')/2$
38		$\sum A^2 \cdot A$
46		$(\sum AA' \cdot S)/2$
74		$\sum SA' \cdot \widetilde{M} \cdot \widetilde{M}'$
78		$(\sum S^2 \cdot \widetilde{M} \cdot \widetilde{M}' - trS^2)/2$
98		$(\sum A'^2 \cdot A)/3$
102		$\sum A^2 \cdot S$
108		$(\sum A'A \cdot S)/2$
110		$\sum S^2 \cdot A$
238		$(\sum S^2 \cdot S)/6$

$$\langle \text{id}12^* \rangle = \langle \text{id}12 \rangle - \langle \text{id}38 \rangle - \langle \text{id}102 \rangle,$$

$$\langle \text{id}14^* \rangle = \langle \text{id}14 \rangle - \langle \text{id}46 \rangle - \langle \text{id}102 \rangle - \langle \text{id}110 \rangle,$$

$$\langle \text{id}36^* \rangle = \langle \text{id}36 \rangle - \langle \text{id}38 \rangle - \langle \text{id}46 \rangle,$$

$$\langle \text{id}74^* \rangle = \langle \text{id}74 \rangle - \langle \text{id}102 \rangle - \langle \text{id}108 \rangle - \langle \text{id}110 \rangle,$$

$$\langle \text{id}78^* \rangle = \langle \text{id}78 \rangle - \langle \text{id}110 \rangle - \langle \text{id}238 \rangle, \qquad \text{(B1)}$$

where $\langle G \rangle$ represents the values obtained from Eq. (5) and $\langle G^* \rangle$ is the corrected value. Generally, for larger subgraphs the corrections made will be of an inclusion-exclusion type.

APPENDIX C: SUBGRAPH ENUMERATION

In numerically enumerating the subgraphs, we combine a dynamic programming method [8], which is applied generally for n-node subgraphs with $n \geq 4$, and a more rapid calculation, based on adjacency matrix operations, used for 3-node subgraphs. The method generalizes the results of Ref. [52]. Here we give formulas for the thirteen 3-node connected directed subgraphs based on the adjacency matrix.

The network adjacency matrix is denoted by M, where $M_{ij} = 1$ if a directed edge exists from node i to node j. We begin by dividing the network into a network containing only antisymmetric arrows, whose adjacency matrix will be denoted by A, and a network containing only mutual arrows, whose symmetric adjacency matrix will be denoted as S,

$$M = A + S. \qquad \text{(C1)}$$

We denote by AB the matrix multiplication of matrices A and B, and by $A \cdot B$ the dot multiplication. $\tilde{A}$ is the logical inverse of matrix A, where the 0 elements of A are the 1 of $\tilde{A}$ and vice versa. A' is the transpose matrix of A. A summation denotes summation of all the matrix indices. The matrix formulas for the 13 directed connected 3-node subgraphs are given in Table III. For example, id38 has two nodes which are connected by a path of two edges and a path of one edge. A^2_{ij} is the number of length 2 paths between node i and node j. Dot multiplication with matrix A and summation of the terms of the resultant matrix gives the correct count. In some of the subgraphs a correction is made for the terms on the diagonal (id6, id36, id78).

[1] S.H. Strogatz, Nature (London) **410**, 268 (2001).

[2] R. Albert and A.L. Barabasi, Rev. Mod. Phys. **74**, 47 (2002).

[3] M. Newman, S. Strogatz, and D. Watts, Phys. Rev. E **6402**, 6118 (2001).

[4] S.N. Dorogovtsev, J.F.F. Mendes, and A.N. Samukhin, e-print cond-mat/0204111.

[5] D. Watts and S. Strogatz, Nature (London) **393**, 440 (1998).

[6] L. Amaral, A. Scala, M. Barthelemy, and H. Stanley, Proc. Natl. Acad. Sci. U.S.A. **97**, 11149 (2000).

[7] S. Shen-Orr, R. Milo, S. Mangan, and U. Alon, Nat. Genet. **31**, 64 (2002).

[8] R. Milo, S. Shen-Orr, S. Itzkovitz, N. Kashtan, D. Chklovskii, and U. Alon, Science **298**, 824 (2002).

[9] S. Maslov and K. Sneppen, Science **296**, 910 (2002).

[10] P. Erdős and A. Rényi, Publ. Math. (Debrecen) **6**, 290 (1959).

[11] P. Erdős and A. Rényi, Publ. Math. Inst. Hungarian Acad. Sci. **5**, 17 (1960).

[12] P. Erdős and A. Rényi, Acta Math. Acad. Sci. Hung. **12**, 261 (1961).

[13] B. Bollobas, *Random Graphs* (Academic Press, New York 1985).

[14] P.W. Holland, S. Leinhardt, and D. Heise, *Sociological Methodology* (Jossey-Bass, San Fransisco, 1975), pp. 1–45.

[15] J.A. Davis and S. Leinhardt, in *Sociological Theories in Progress*, edited by Joseph Berger, Morris Zelditch, Jr., and Bo Anderson (Houghton Mifflin, Boston, 1972), Vol. 2, pp. 218–251.

[16] P.W. Holland and S. Leinhardt, Am. J. Sociol. **70**, 492 (1970).

[17] S. Wasserman and K. Faust, *Social Network Analysis: Methods and Applications* (Cambridge University Press, Cambridge, 1994).

[18] A.L. Barabasi and R. Albert, Science **286**, 509 (1999).

[19] S. Redner, Eur. Phys. J. B **4**, 131 (1998).

[20] M. Faloutsos, P. Faloutsos, and C. Faloutsos, Comput. Commun. Rev. **29**, 251 (1999).

[21] B.A. Huberman and L.A. Adamic, Nature (London) **401**, 131 (1999).

[22] Z. Burda, J.D. Correia, and A. Krzywicki, Phys. Rev. E **64**, 046118 (2001).

[23] A. Krzywicki, e-print cond-mat/0110574.

[24] P.L. Krapivsky, G.J. Rodgers, and S. Redner, Phys. Rev. Lett. **86**, 5401 (2001).

[25] R. Ferrer i Cancho and R.V. Solé, Proc. R. Soc. London, Ser. B **268**, 2261 (2001).

[26] S. Valverde, R. Ferrer i Cancho, and R.V. Sole, e-print cond-mat/0204344.

[27] S.N. Dorogovtsev and J.F.F. Mendes, Adv. Phys. **51**, 1079 (2002).

[28] S.N. Dorogovtsev and A.N. Samukhin, Phys. Rev. E **67**, 037103 (2003).

[29] R. Ferrer i Cancho, C. Janssen, and R.V. Sole, Phys. Rev. E **64**, 046119 (2001).

[30] W. Aiello, F. Chung, and L. Lu, Exp. Math. **10**, 53 (2001).

[31] M. Molloy and B. Reed, Combinatorics, Probab. Comput. **7**, 295 (1998).

[32] F. Chung and L. Lu, Proc. Natl. Acad. Sci. U.S.A. **99**, 15879 (2002).

[33] R. Cohen, D. Ben-Avraham, and S. Havlin, Phys. Rev. E **66**, 036113 (2002).

[34] R. Cohen, K. Erez, D. ben-Avraham, and S. Havlin, Phys. Rev. Lett. **85**, 4626 (2000).

[35] M. Newman, Phys. Rev. E **60**, 7332 (1999).

[36] M. Newman, in *Handbook of Graphs and Networks*, edited by S. BornHoldt and G. Schuster (Wiley-VCH, Berlin, 2002).

[37] S.N. Dorogovtsev, J.F.F. Mendes, and A.N. Samukhin, e-print cond-mat/0206467.

[38] E. Ravasz and A.L. Barabasi, Phys. Rev. E **67**, 026112 (2003).

[39] J. Eckmann and E. Moses, Proc. Natl. Acad. Sci. U.S.A. **99**, 5825 (2002).

[40] P. Collet and J. Eckmann, J. Stat. Phys. **108**, 1107 (2002).

[41] S. Maslov, K. Sneppen, and A. Zaliznyak, e-print cond-mat/0205379.

[42] C.A. Ouzounis and P.D. Karp, Genome Res. **10**, 568 (2000).

[43] A. Wagner and D. Fell, Proc. R. Soc. London, Ser. B **268**, 1803 (2001).

[44] N. Guelzim, S. Bottani, P. Bourgine, and F. Képès, Nat. Genet. **31**, 60 (2002).

[45] G. Bianconi and A. Capocci, Phys. Rev. Lett. **90**, 078701 (2003).

[46] J. Berg and M. Lässig, Phys. Rev. Lett. **89**, 228701 (2002).

[47] M. Molloy and B. Reed, Random Struct. Algorithms **6**, 161 (1995).

[48] E. Bender and E. Canfield, J. Comb. Theory, Ser. A **24**, 296 (1978).

[49] J.G. White, E. Southgate, J.N. Thomson, and S. Brenner, Philos. Trans. R. Soc. London, Ser. B **314**, 1 (1986).

[50] G. Bianconi and A.L. Barabasi, Phys. Rev. Lett. **86**, 5632 (2001).

[51] In subgraphs which have several nodes with the maximal degree, this approximation should still give correct scaling results as long as the probability of obtaining several large hubs in one subgraph is low. In obtaining the scaling relations, we replace the number of mutual edges with their average value $\langle K \rangle^2 / N$. This gives an incorrect value for the exact number of appearances (compare id14 vs id38 in Table I), but is valid for obtaining scaling relations.

[52] F. Harary and H.J. Kommel, J. Math. Sociol. **6**, 199 (1979).

[53] M. Newman, Phys. Rev. Lett. **89**, 208701 (2002).