Exponential random graph models for networks with community structure

Piotr Fronczak, Agata Fronczak, and Maksymilian Bujok
Faculty of Physics, Warsaw University of Technology, Koszykowa 75, PL-00-662 Warsaw, Poland
(Received 17 May 2013; published 23 September 2013)

Although the community structure organization is an important characteristic of real-world networks, most of the traditional network models fail to reproduce the feature. Therefore, the models are useless as benchmark graphs for testing community detection algorithms. They are also inadequate to predict various properties of real networks. With this paper we intend to fill the gap. We develop an exponential random graph approach to networks with community structure. To this end we mainly built upon the idea of blockmodels. We consider both the classical blockmodel and its degree-corrected counterpart and study many of their properties analytically. We show that in the degree-corrected blockmodel, node degrees display an interesting scaling property, which is reminiscent of what is observed in real-world fractal networks. A short description of Monte Carlo simulations of the models is also given in the hope of being useful to others working in the field.

DOI: 10.1103/PhysRevE.88.032810 PACS number(s): 89.75.Hc, 89.75.Da, 89.75.Kd, 02.50.−r

I. INTRODUCTION

Of particular interest in recent years has been the community structure in networks [1,2]. As a community (module or block), one understands a group of nodes that is densely connected internally but sparsely connected externally. To name a few, communities may be groups of related individuals in social networks, sets of web pages on the same topic, biochemical pathways in metabolic networks, and groups of countries in the world-trade network that signed regional trade agreements. The above examples show that the group membership is related to the function of a node in the network. For this reason, aside from the small-world effect and scale-free degree distributions, the community structure is considered one of the most important topological properties of complex networks, yet this structure is not fully understood and not well captured by network models.

One of the proposed models for networks with community structure, with a long tradition of study in the social sciences and computer science [3–8], is the so-called blockmodel. In its classical version [3], each of N vertices is assigned to one of K blocks (communities) of equal size, and undirected edges are independently drawn between pairs of nodes with probabilities that are a function only of the group membership of the nodes. The well-defined community structure of this model caused it to be the most regularly used benchmark for testing community detection algorithms [9,10].

However, apart from the built-in blocks (communities), other properties of this model, especially the Poisson-like degree distribution, indicate its shortcomings as the correct model for real networks. To our knowledge, in respect to community detection algorithms, the concern was first raised by Lancichinetti, Fortunato, and Radicchi [11]. The authors argued that real networks are characterized by heterogeneity in the distributions of node degrees and of community sizes, which is not the case in the classical blockmodel. With this motivation, Lancichinetti *et al.* have proposed an efficient numerical construction procedure for benchmark graphs that account for the desired network properties.

More recently, a similar perspective was also raised in other papers; see, e.g., Refs. [12–16]. In particular, Karrer and Newman in Ref. [12] showed that due to limitations of the traditional blockmodel, its fitting to empirical network data, which is a way of discovering community structure, may be misleading. The authors suggested how to generalize the classical blockmodel to incorporate arbitrary degree distributions. They showed that the degree-corrected counterpart dramatically outperforms the traditional blockmodel as a tool for detecting community structure.

In the following, we present an exponential random graph formulation [17–21] for both the traditional and the degree-corrected blockmodels. The aim behind is to create a general model for networks with community structure, which has a formal mathematical foundation and is easy to implement in numerical simulations. Models of networks are very important in the study of network processes and algorithms. As already indicated, community detection algorithms can be evaluated more effectively on synthetic networks with a well-defined community structure (benchmarks) than on real networks [11], because one can easily vary the model parameters and compare the recovered community structure with the predefined one. Network models are also ubiquitous in studies of different processes that takes place on networks [22], such as the spread of a disease over a social network, or the flow of traffic on communication networks. Therefore, it is quite surprising that most of the current network models disregard the issue of community structure. This paper intends to fill the gap.

The outline of the paper is as follows. In Sec. II, the classical blockmodel is considered based on exponential random graph approach. In Sec. III, its degree-corrected version is studied and various network properties are calculated. In Sec. IV, a short training in Monte Carlo simulations, to which the considered models lend themselves admirably, is given. The paper is concluded in Sec. V.

II. CLASSICAL BLOCKMODEL

Exponential random graphs (ERGs) are ensemble models, which are defined not to be a single network but a collection, $\{G\}$, of possible networks. A graph G in the ensemble is assigned the probability

$$P(G) = \frac{e^{H(G)}}{Z}, \tag{1}$$

where

$$Z = \sum_{\{G\}} e^{H(G)} \tag{2}$$

is the normalization constant (partition function), and

$$H(G) = \sum_r \theta_r m_r(G) \tag{3}$$

is called the graph Hamiltonian, with $\{m_r(G)\}$ being a collection of graph observables (which reflect desired properties of the model), and $\{\theta_r\}$ standing for model parameters (which are coupled to observables).

In this paper, when we will talk about possible realizations, $\{G\}$, of networks with community structure, we will always refer to simple graphs with N-labeled nodes and independent edges. To state precisely, a simple graph, G, has at most one link between any pair of nodes and it does not contain self-loops connecting nodes to themselves. It means that entries of its adjacency matrix, $A_{ij}(G) \equiv A_{ij}$ and $A_{ij} \in \{0,1\}$, are symmetric, $A_{ij} = A_{ji}$, and they are equal to zero on the diagonal, $A_{ii} = 0$. Therefore, by assuming statistical independence of the links, the probability of a graph, G, can be written as

$$P(G) = \prod_{i<j} p_{ij}^{A_{ij}} (1 - p_{ij})^{1-A_{ij}}, \tag{4}$$

where p_{ij} is the probability that the vertices i and j are connected.

It is remarkable, that Eq. (4) can be rewritten in the canonical form of Eq. (1),

$$P(G) = \frac{\exp\left[\sum_{i<j} \ln\left(\frac{p_{ij}}{1-p_{ij}}\right) A_{ij}\right]}{\prod_{i<j}(1-p_{ij})^{-1}}, \tag{5}$$

with the Hamiltonian defined as

$$H(G) = \sum_{i<j} \theta_{ij} A_{ij}, \tag{6}$$

where

$$\theta_{ij} = \ln\left(\frac{p_{ij}}{1-p_{ij}}\right), \tag{7}$$

is a separate parameter (the so-called Lagrange multiplier) coupling to an edge between i and j, and the partition function is given by

$$Z(\{\theta_{ij}\}) = \prod_{i<j}(1-p_{ij})^{-1} = \prod_{i<j}(1+e^{\theta_{ij}}). \tag{8}$$

Furthermore, one can show that the expected value of each matrix entry, $\langle A_{ij}\rangle$, can be calculated by differentiating the logarithm of the partition function, $\ln Z(\{\theta_{ij}\})$, with respect to θ_{ij},

$$\langle A_{ij}\rangle = \frac{\partial \ln Z}{\partial \theta_{ij}} = \frac{e^{\theta_{ij}}}{1+e^{\theta_{ij}}} = p_{ij}. \tag{9}$$

In the classical blockmodel, one assumes that the number of groups, K, and the number of vertices in each group, N_r, are known, and

$$\sum_{r=1}^{K} N_r = N, \tag{10}$$

where N is the total number of vertices in the network.[1] One also assumes that the probability that the vertices i and j are connected is a function only of their group membership,

$$p_{ij} \equiv q_{g_i g_j}, \tag{11}$$

where g_i and g_j represent the groups to which the vertices belong. With this in mind, the parameter θ_{ij}, Eq. (7), can be written as

$$\theta_{ij} = \ln\left(\frac{q_{g_i g_j}}{1-q_{g_i g_j}}\right) \equiv \omega_{g_i g_j}, \tag{12}$$

and the Hamiltonian of the classical blockmodel gets the informative form

$$H(G) = \sum_{i<j} \omega_{g_i g_j} A_{ij} = \sum_{r\leqslant s} \omega_{rs} \sum_{i<j} A_{ij}\delta_{g_i r}\delta_{g_j s} \tag{13}$$

$$= \sum_{r\leqslant s} \omega_{rs} E_{rs}(G), \tag{14}$$

where $E_{rs}(G)$ is the number of edges between the groups r and s, $E_{rr}(G)$ represents edges within the same group r, and δ_{rs} is the Kronecker delta.

The Hamiltonian obtained, Eq. (14), shows that the classical blockmodel is equivalent to the ensemble of networks with the specified average numbers of edges within and between the predefined blocks, i.e., $\langle E_{rr}\rangle$ and $\langle E_{rs}\rangle$, respectively. The partition function, Eq. (8), corresponding to this ensemble can, after an amount of algebra, be written in the convenient form

$$Z(\{\omega_{rs}\}) = \prod_{r<s}(1-q_{rs})^{-N_r N_s} \prod_r (1-q_{rr})^{-\binom{N_r}{2}} \tag{15}$$

$$= \prod_{r<s}(1+e^{\omega_{rs}})^{N_r N_s} \prod_r (1+e^{\omega_{rr}})^{\binom{N_r}{2}}, \tag{16}$$

which allows us to calculate the mentioned averages:

$$\langle E_{rr}\rangle = \frac{\partial \ln Z}{\partial \omega_{rr}} = \binom{N_r}{2} \frac{e^{\omega_{rr}}}{1+e^{\omega_{rr}}} = \binom{N_r}{2} q_{rr}, \tag{17}$$

$$\langle E_{rs}\rangle = \frac{\partial \ln Z}{\partial \omega_{rs}} = N_r N_s \frac{e^{\omega_{rs}}}{1+e^{\omega_{rs}}} = N_r N_s q_{rs}. \tag{18}$$

Equations (17) and (18) can be used to calculate internal and external degrees of a node i, i.e., $\langle k_i^{\text{int}}\rangle$ and $\langle k_i^{\text{ext}}\rangle$, which are the numbers of edges that connect the node to other nodes in the same block and nodes from different blocks, respectively. In order to do it, let us assume that the partial node's degree, $k_{i,r}^s$, is the number of edges that are incident to i, where $g_i = r$, and such that their second endpoint belongs to the block s, i.e.,

$$k_{i,r}^s = \delta_{g_i r} \sum_j A_{ij}\delta_{g_j s}. \tag{19}$$

[1] Let us note that in the classical blockmodel all communities are of the same size, see Refs. [3–6], differently from what is assumed here. A variant of the standard blockmodel with communities of different size was first introduced by Donan *et al.* in Ref. [28].

Then, the internal and external degrees can be written as

$$\langle k_i^{\text{int}} \rangle = \langle k_{i,r}^{r} \rangle, \tag{20}$$

$$\langle k_i^{\text{ext}} \rangle = \sum_{s,s\neq r} \langle k_{i,r}^{s} \rangle, \tag{21}$$

and the average node's connectivity through the whole network is

$$\langle k_i \rangle = \left\langle \sum_s k_{i,r}^{s} \right\rangle = \langle k_i^{\text{int}} \rangle + \langle k_i^{\text{ext}} \rangle. \tag{22}$$

In the classical blockmodel, Eqs. (20) and (21) simplify to:

$$\langle k_i^{\text{int}} \rangle = \delta_{g_i r} \frac{2\langle E_{rr} \rangle}{N_r} = \delta_{g_i r}(N_r - 1) q_{rr}, \tag{23}$$

and

$$\langle k_i^{\text{ext}} \rangle = \delta_{g_i r} \sum_{s,s\neq r} \frac{\langle E_{rs} \rangle}{N_r} = \delta_{g_i r} \sum_{s,s\neq r} N_s q_{rs}. \tag{24}$$

In Sec. IV, the results from this section are used to introduce and discuss the method of Monte Carlo simulations for networks with community structure.

III. DEGREE-CORRECTED BLOCKMODEL

In this section, we develop an exponential random graph approach to networks with community structure that accounts for the heterogeneity of both degree and community size. We assume that the probability, p_{ij}, that vertices i and j are connected is not only a function of their group membership (as it was the case of the classical blockmodel) but also depends on the vertices themselves. In doing so, we examine the Hamiltonian, Eq. (6), with the following set of the edge parameters:

$$\theta_{ij} = v_i + v_j + \omega_{g_i g_j}, \tag{25}$$

where the parameters $\{v_i\}$ are thought to control expected node degrees and ω_{rs} controls edges between groups $r = g_i$ and $s = g_j$ to which the nodes i and j belong.

By inserting Eq. (25) into Eq. (6), the Hamiltonian of the considered network ensemble can, after an amount of algebra, be written in two alternative forms: the first consisting of a sum over the node degrees and over the edges within and between the predefined blocks, i.e.,

$$H(G) = \sum_{i<j} (v_i A_{ij} + v_j A_{ji}) + \sum_{i<j} \omega_{g_i g_j} A_{ij} \tag{26}$$

$$= \sum_{i,j} v_i A_{ij} + \sum_{r\leqslant s} \omega_{rs} \sum_{i<j} \delta_{g_i r} \delta_{g_j s} A_{ij} \tag{27}$$

$$= \sum_i v_i k_i(G) + \sum_{r\leqslant s} \omega_{rs} E_{rs}(G), \tag{28}$$

and the second being the sum over the partial node's degrees, i.e.,

$$H(G) = \sum_{i,j} \left(v_i + \frac{\omega_{g_i g_j}}{2} \right) A_{ij} \tag{29}$$

$$= \sum_{r,s} \sum_i \left(v_i + \frac{\omega_{rs}}{2} \right) k_{i,r}^{s}(G) \tag{30}$$

$$= \sum_{r,s} \sum_i v_{i,r}^{s} k_{i,r}^{s}(G) \tag{31}$$

$$= \sum_r \sum_i v_{i,r}^{r} k_{i,r}^{r}(G) + \sum_{r\neq s} \sum_i v_{i,r}^{s} k_{i,r}^{s}(G), \tag{32}$$

where the new set of parameters,

$$v_{i,r}^{s} = \left(v_i + \frac{\omega_{rs}}{2} \right), \tag{33}$$

is introduced, such that the edge parameters, θ_{ij}, Eq. (25), can be written as

$$\theta_{ij} = v_{i,r}^{s} + v_{j,s}^{r}, \tag{34}$$

where the symmetry condition, $\omega_{rs} = \omega_{sr}$, is used. The new parameters, Eq. (33), determine the expected values of partial degrees, $\langle k_{i,r}^{s} \rangle$, Eq. (19).

The two equivalent forms of the Hamiltonian, Eqs. (28) and (32), do not only provide alternative parametrization schemes for the considered ensemble. They are also helpful in analyzing various network properties. In particular, Eq. (32) indicates that the networks may be considered as composed of K-independent blocks (i.e., K random graphs with a given node degree sequence; see Refs. [17,18]) and the K-partite graph,[2] where K is the number of blocks. Although, the noticed graph partitioning is quite obvious (especially if one recalls that the considered ensemble is the ensemble of networks with independent edges), it is also true that the partitioning is not as straightforward, if one is only acquainted with Eq. (28). On the other hand, Eq. (28) is valuable in the sense that by comparing it with the Hamiltonian of the classical blockmodel, Eq. (14), it follows directly that the ensemble parameters θ_{ij} provided by Eq. (25) do really define the degree-corrected blockmodel.

Given the understanding of the networks as composed of K-independent generalized random graphs (blocks) and the K-partite graph, the partition function of the degree-corrected blockmodel can be written in the multiplicative form:

$$Z(\{v_i + v_j + \omega_{rs}\}) = \prod_r Z_r \prod_{r>s} Z_{rs}, \tag{35}$$

where Z_r represents the partition function of the block r (see Eq. (20) in Ref. [17]),

$$Z_r = \prod_{i<j} [1 + \delta_{g_i r} \delta_{g_j r}\, e^{(v_i + v_j + \omega_{rr})}], \tag{36}$$

and Z_{rs} is the partition function of the set of edges between the blocks r and s,

$$Z_{rs} = \prod_{i<j} [1 + \delta_{g_i r} \delta_{g_j s}\, e^{(v_i + v_j + \omega_{rs})}]. \tag{37}$$

Now, the average value of the partial degree, $\langle k_{i,r}^{s} \rangle$, can be obtained by differentiating the logarithm of the partition function with respect to $v_{i,r}^{s}$, i.e.,

$$\langle k_{i,r}^{s} \rangle = \frac{\partial \ln Z}{\partial v_{i,r}^{s}} = \delta_{g_i r} \sum_j \delta_{g_j s}\; p_{ij}, \tag{38}$$

[2] K-partite graph is a graph where the vertices are partitioned into K disjoint subsets with the condition that no two vertices in the same subset are adjacent.

where p_{ij}, Eq. (9), is the probability that the vertices i and j are connected,

$$p_{ij} = \frac{\exp[v_i + v_j + \omega_{g_i g_j}]}{1 + \exp[v_i + v_j + \omega_{g_i g_j}]}. \tag{39}$$

Consequently, the average internal and external degrees characterizing nodes of the considered networks are given by, cf. Eqs. (20) and (21),

$$\langle k_i^{\text{int}} \rangle = \delta_{g_i r} \sum_j \delta_{g_j r} \frac{\exp[v_i + v_j + \omega_{rr}]}{1 + \exp[v_i + v_j + \omega_{rr}]}, \tag{40}$$

and

$$\langle k_i^{\text{ext}} \rangle = \delta_{g_i r} \sum_{s, s \neq r} \sum_j \delta_{g_j s} \frac{\exp[v_i + v_j + \omega_{rs}]}{1 + \exp[v_i + v_j + \omega_{rs}]}. \tag{41}$$

It is worth mentioning that with $\forall_i v_i = 0$, the edge parameters in the degree-corrected blockmodel, which are given by Eq. (25), simplify to the edge parameters underlying the classical blockmodel, Eq. (12). Accordingly, with $\forall_i v_i = 0$, Eqs. (40) and (41) simplify to Eqs. (23) and (24).

To proceed further, sparse networks will be considered, in which the probability of any individual edge is small, $p_{ij} \ll 1$. (The assumption is reasonable since most of real-world networks are very sparse, i.e., the number of existing edges is much smaller than the number of edges which could theoretically exist in the network.) To achieve this, one needs $e^{v_i + v_j + \omega_{g_i g_j}} \ll 1$ in Eq. (39), which results in

$$p_{ij} \simeq \exp[v_i + v_j + \omega_{g_i g_j}] = \exp\left[v_{i,r}^s\right] \exp\left[v_{j,s}^r\right]. \tag{42}$$

Equation (42) means that in a sparse degree-corrected blockmodel, the probability of an edge is simply a product of two terms, one for each of the vertices at either end of the edge. Furthermore, it turns out that these terms are simply related to the expected partial degrees of the vertices (similarly to what is found in exponential random graphs with a given node degree sequence; see, e.g., Eq. (26) in Ref. [18]). In particular, one can show that the expression for the expected partial degree can be written as (for a detailed derivation see Appendix)

$$\langle k_{i,r}^s \rangle = \exp\left[v_{i,r}^r\right] \frac{\langle E_{rs} \rangle}{\sqrt{2\langle E_{rr} \rangle}} \tag{43}$$

$$= \exp\left[v_{i,r}^s\right] \sqrt{\langle E_{rs} \rangle}, \tag{44}$$

and consequently it is straightforward to obtain the expected internal and external degrees of the nodes, cf. Eqs. (20) and (21),

$$\langle k_i^{\text{int}} \rangle = \exp\left[v_{i,r}^r\right] \sqrt{2\langle E_{rr} \rangle} \tag{45}$$

$$\langle k_i^{\text{ext}} \rangle = \exp\left[v_{i,r}^r\right] \frac{\sum_{s, s \neq r} \langle E_{rs} \rangle}{\sqrt{2\langle E_{rr} \rangle}}. \tag{46}$$

Accordingly, it is interesting to note that there is a linear dependence between the obtained degrees, i.e., for each node i the number of edges going out from i to nodes in a different block is proportional to the number of links to nodes in the same block, i.e.,

$$\langle k_{i,r}^s \rangle = \langle k_{i,r}^r \rangle \frac{\langle E_{rs} \rangle}{2\langle E_{rr} \rangle}, \tag{47}$$

and

$$\langle k_i^{\text{ext}} \rangle = \langle k_i^{\text{int}} \rangle \frac{\sum_{s, s \neq r} \langle E_{rs} \rangle}{2\langle E_{rr} \rangle}. \tag{48}$$

The scaling relation, Eq. (48), is an important feature of the degree-corrected blockmodel, because it relates the model to fractal networks as described by Song, Havlin, and Makse [23,24]. In Ref. [23], the authors argued that the self-similarity of real-world complex networks with power-law degree distributions results from the scaling property of the node degrees, which arises when the networks undergo a renormalization procedure that coarse-grains their nodes into boxes (blocks). It is remarkable that the scaling property reported by Song *et al.* is similar to the scale-transformation described by Eq. (48). This similarity comes as a surprise, especially that in this study, contrary to what is suggested in Ref. [24], the relation is not attributed to any specific network construction procedure. It is an intrinsic feature of the degree-corrected blockmodel, which is (at first and naive glance) at variance with the intuitive expectation, suggesting homogeneous distribution of connections within the whole network.

Finally, inserting Eqs. (A3) and (43) into the expression for the the connection probability, p_{ij}, between two vertices i and j, i.e., into Eq. (42), one gets that p_{ij} depends on whether the vertices belong to the same block or not. In the first case, when $g_i = g_j = r$, one has

$$p_{ij} = \frac{\langle k_{i,r}^r \rangle \langle k_{j,r}^r \rangle}{2\langle E_{rr} \rangle}. \tag{49}$$

In the second case, for $g_i = r$, $g_j = s$, and $r \neq s$, the probability is given by

$$p_{ij} = \frac{\langle k_{i,r}^s \rangle \langle k_{j,r}^r \rangle}{\langle E_{rs} \rangle}. \tag{50}$$

IV. NUMERICAL SIMULATIONS

Monte Carlo simulation [25,26] is a numerical method that is ideally suited to exponential random graphs. In what follows we briefly describe the method and apply it to the classical blockmodel and its degree-corrected counterpart.

A. Metropolis-Hastings algorithm for ERGs

Let us consider the ensemble of exponential random graphs with the Hamiltonian given by Eq. (3). Once the parameters $\{\theta_i\}$ in the Hamiltonian are specified, the probability distribution, $P(G)$, which is given by Eq. (1), makes generation of graphs correctly sampled from the ensemble straightforward using a Metropolis-Hastings-type Markov chain method. In the method, one defines a move-set in the space of graphs and then repeatedly generates moves from this set, accepting them with probability

$$p = 1 \quad \text{if} \quad P(G') > P(G), \tag{51}$$

where G' is the graph after performance of the move, and

$$p = \frac{P(G')}{P(G)}, \quad \text{if} \quad P(G') < P(G), \tag{52}$$

while rejecting them with probability $1-p$. Because of the exponential form of $P(G)$, the acceptance probability, which is given by Eq. (52), is particularly simple to calculate. It can be written as

$$p = e^{H(G')-H(G)} = e^{\Delta H} \quad \text{if} \quad \Delta H < 0, \tag{53}$$

where

$$\Delta H = \sum_r \theta_r [m_r(G') - m_r(G)]. \tag{54}$$

Let us also note that with the help of ΔH, the condition for certain acceptance of a change, i.e., Eq. (51), becomes

$$p = 1, \quad \text{if} \quad \Delta H > 0. \tag{55}$$

The choice of the right move-set depends on the set of all possible network realizations, $\{G\}$, underlying the studied ensemble. The example move-sets are: (i) addition and removal of edges between randomly chosen vertex pairs for the case of graphs, which do not have a fixed number of edges; (ii) movement of edges randomly from one place to another for the case of fixed edge numbers but variable degree sequence; (iii) edges swaps of the form $\{(v_1,w_1),(v_2,w_2)\} \to \{(v_1,w_2),(v_2,w_1)\}$ for the case of fixed degree sequence, where (v_1,w_1) denote an edge from vertex v_1 to vertex w_1. Monte Carlo numerical simulations of this type are simple to implement and appear to converge quickly allowing one to study quite large graphs.

B. Girvan-Newman benchmark

To be concrete, let us discuss the Metropolis algorithm for the classical blockmodel, which corresponds to the famous Girvan-Newman benchmark for testing community detection methods. At the beginning one assumes that the number of groups, K, and the number of vertices in each group N_r are known, and $\sum_{r=1}^{K} N_r = N$. Groups to which the nodes belong are also specified, likewise the connection probabilities, q_{rs} between and within the blocks.

Having the ensemble parameters, the ensemble construction procedure proceeds through the following steps:

(1) At the beginning one creates any simple graph (i.e., its adjacency matrix) with a given number of nodes, N. The starting configuration may be, for instance, the edgeless graph.

(2) Next, in the following time steps, one randomly chooses a matrix element, $A_{ij}(G)$, to be considered for change. For the case when $A_{ij}(G) = 1$ (0, respectively) one considers deletion (addition) of the edge, i.e., $A_{ij}(G') = 0$ (1, respectively). This corresponds to the move-set: addition and removal of edges between randomly chosen vertex pairs. Whether the change is accepted depends on ΔH; see Eq. (54). In the classical blockmodel, since $H(G) = \sum_{i<j} \omega_{g_i g_j} A_{ij}(G)$, cf. Eq. (13), one has

$$\Delta H = \pm \omega_{g_i g_j}, \tag{56}$$

with the upper (lower) sign relating to addition (deletion, respectively) of an edge. Therefore, the acceptance criteria, Eqs. (53) and (55), depend on the sign of the ensemble parameters ω_{rs}, which are given by Eq. (12).

(3) The updating of elements A_{ij} should be continued until the network observables stabilize around their mean values. In the case of the classical blockmodel, the average numbers of edges, $\langle E_{rr} \rangle$ and $\langle E_{rs} \rangle$, should place itself around the values, which are given by Eqs. (17) and (18). Once numerical simulations stabilize, graphs which appear in the course of subsequent updates of the adjacency matrix appear to be correctly sampled network realizations of the studied ensemble.

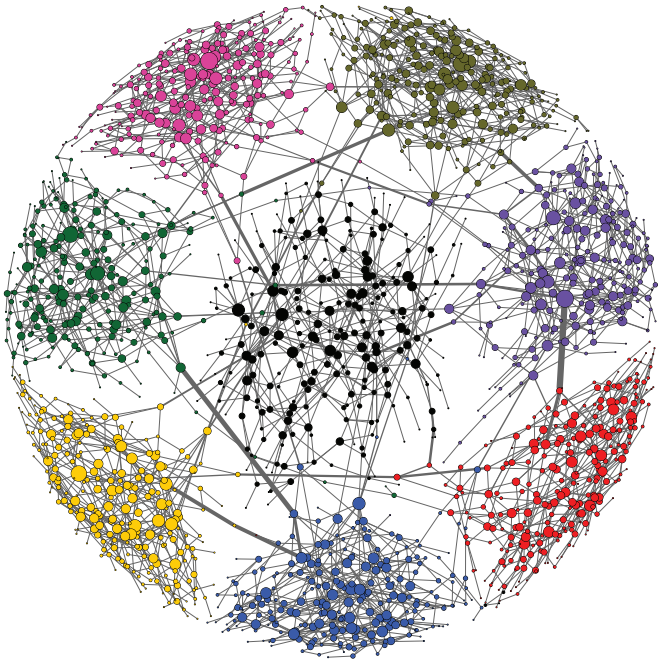

FIG. 1. (Color online) A Monte Carlo realization of the classical blockmodel with $N = 2048$ nodes divided into $K = 8$ groups of equal size $N_r = 256$. The color symbols indicate the group membership, the node size is proportional to the degree, and the edge width is proportional to the edge betweenness. To get the figure the following values of the ensemble parameters have been used: $\forall_r q_{rr} = 3/255$ and $\forall_{r \neq s} q_{rs} = 0.006/256$; see Eqs. (11) and (12). The network was visualized using Cytoscape software [27].

A realization of the classical blockmodel with $N = 2048$ nodes divided into $K = 8$ groups of equal size is shown in Fig. 1.

C. Degree-corrected blockmodel

The construction procedure of the degree-corrected blockmodel follows the same steps as in the classical blockmodel described in the previous subsection. The only difference is that the change of the Hamiltonian occurring during the addition or removal of an edge between i and j is, cf. Eq. (34),

$$\Delta H = \pm \left(v^s_{i,r} + v^r_{j,s} \right). \tag{57}$$

Therefore, to perform Monte Carlo simulations of the degree corrected blockmodel, first, one has to determine the set of parameters $\{v^s_{i,r}\}$ underlying the network ensemble with the desired (e.g., scale-free) degree distribution.

Thus, let us start with specifying the parameters $v^r_{i,r}$ which characterize the expected internal degrees of the nodes, $\langle k^{\text{int}}_i \rangle = \langle k^r_{i,r} \rangle$. In order to do it we have to assume the desired node degree sequence within each block,

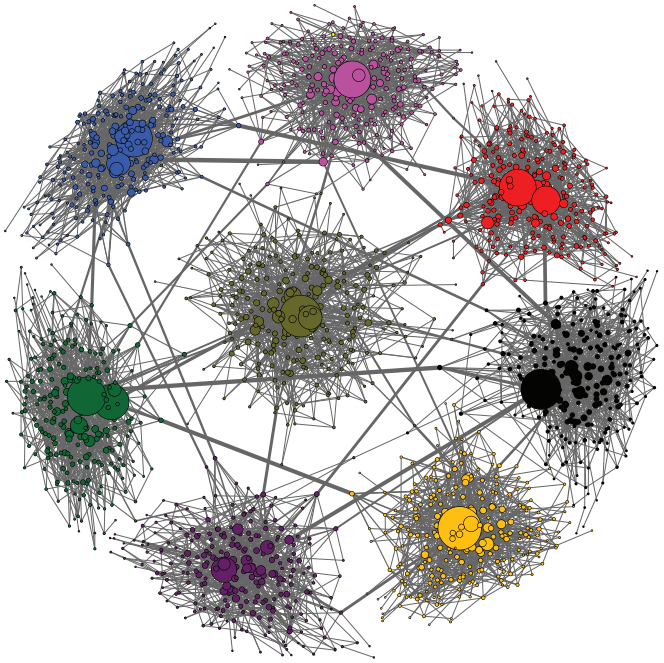

FIG. 2. (Color online) A Monte Carlo realization of the degree-corrected blockmodel with scale-free communities. The network consists of $N = 2048$ nodes divided into $K = 8$ groups of equal size. The color indicates the partition of the nodes into groups, and the size of the nodes is proportional to the degree. The edge width is proportional to the edge betweenness. To perform numerical simulations, the following values of the ensemble parameters have been used: (i) The expected internal degrees have been independently drawn from the power law distribution with the characteristic exponent $\gamma = 3$; (ii) the minimal value of the expected degree was assumed to be 2; (iii) the expected numbers of interblock connections were taken to be $\forall_{r\neq s}\langle E_{rs}\rangle = 1$. The network was visualized using Cytoscape software [27].

$\{\langle k^r_{1,r}\rangle, \langle k^r_{2,r}\rangle, \ldots, \langle k^r_{N_r,r}\rangle\}$, where the numbers $\langle k^r_{i,r}\rangle$ can be drawn from an arbitrary degree distribution. Then, the parameters, $\{v^r_{i,r}\}$, can be calculated from Eq. (45):

$$v^r_{i,r} = \ln\left[\frac{\langle k^r_{i,r}\rangle}{\sqrt{\sum_{i=1}^{N_r}\langle k^r_{i,r}\rangle}}\right]. \tag{58}$$

To determine the parameters $\{v^s_{i,r}\}$ for $r \neq s$, the average numbers of interblock connections, $\langle E_{rs}\rangle$, has to be given. Then, once $\forall_{r\neq s}\langle E_{rs}\rangle$ are known, the parameters, $\{v^s_{i,r}\}$, can be obtained from Eqs. (44), (45), and (47):

$$v^s_{i,r} = \ln\left[\frac{\langle k^s_{i,r}\rangle}{\sqrt{\langle E_{rs}\rangle}}\right] = \ln\left[\frac{\langle k^r_{i,r}\rangle\sqrt{\langle E_{rs}\rangle}}{\sum_{i=1}^{N_r}\langle k^r_{i,r}\rangle}\right] \tag{59}$$

$$= v^r_{i,r} + \ln[\sqrt{\langle E_{rs}\rangle}]. \tag{60}$$

In Fig. 2 a realization of the degree-corrected blockmodel is shown. The network has $N = 2048$ nodes, which are divided into $K = 8$ communities of equal size. The node degree distribution within the communities is scale-free, $\forall_r P(\langle k^r_{i,r}\rangle) \sim \langle k^r_{i,r}\rangle^{-\gamma}$, with the characteristic exponent $\gamma = 3$.

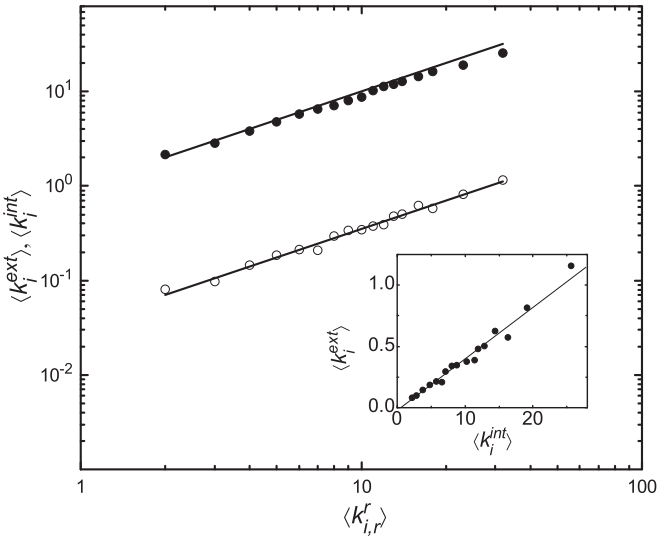

FIG. 3. Scaling relations between the internal, $\langle k_i^{\rm int}\rangle$, and external, $\langle k_i^{\rm out}\rangle$, node degrees in scale-free networks with community structure. In the main panel, linear relations of internal (open symbols) and external (solid symbols) degrees vs. the preimposed internal degrees, $\langle k^r_{i,r}\rangle$, are shown. In the inset, linear scaling between $\langle k_i^{\rm int}\rangle$ and $\langle k_i^{\rm out}\rangle$ is displayed. The scattered points represent the simulated data averaged over 100 snapshots of scale-free networks (with $\gamma = 3$) of size $N = 2048$ nodes divided into $K = 8$ groups of equal size. The smallest preimposed internal node degree was equal to 2, and the expected numbers of interblock connections were taken to be $\forall_{r\neq s}\langle E_{rs}\rangle = 5$. Solid lines either represent the identity $\langle k_i^{\rm int}\rangle = \langle k^r_{i,r}\rangle$ or Eq. (48).

Figures 3 and 4 show that the final networks generated by the proposed algorithm have topological features consistent with those that have been preimposed.

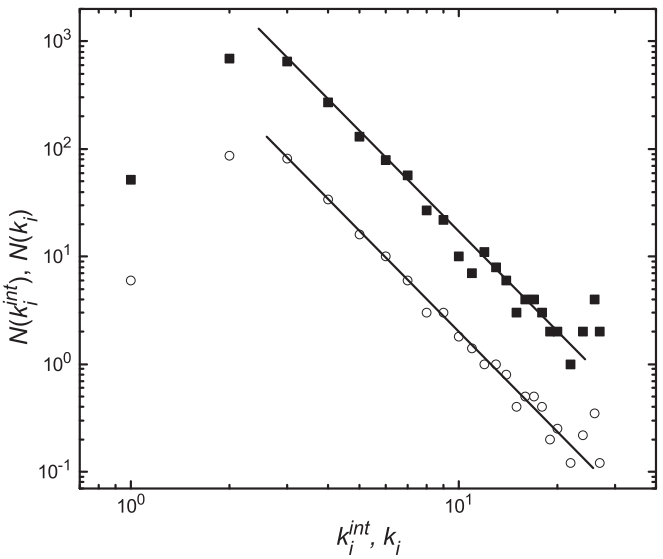

FIG. 4. Histograms of node degrees for a single community, $N(k_i^{\rm int})$ (open symbols), and for the whole network, $N(k_i)$ (solid symbols), where $k_i = k_i^{\rm int} + k_i^{\rm ext}$. The data were gathered during the same Monte Carlo simulations as those shown in Fig. 3. Solid lines represent power law with slope $\gamma = 3$.

V. SUMMARY

In this paper, we have described an exponential random graph approach to networks with community structure. We mainly built upon the idea of the blockmodel, although other ideas regarding network structure have been also exploited. Two kinds of the network structural Hamiltonians have been considered: the first one corresponding to the classical blockmodel, and the second one corresponding to its degree-corrected version. In both cases, a number of analytical predictions about various network properties was given. In particular, it was shown that in the degree-corrected blockmodel, node degrees display an interesting scaling property, that is similar to the scaling feature of the node degrees in fractal (self-similar) real-world networks. A short training in Monte Carlo simulations of the models was also given in the hope of being useful to others working in the field of networks with community structure.

ACKNOWLEDGMENT

This work was supported by the Foundation for Polish Science under Grant No. POMOST/2012-5/5.

APPENDIX

In order to derive Eqs. (43)–(46), one can start with inserting Eq. (42) into Eq. (38):

$$\langle k^s_{i,r}\rangle = \exp\left[v^s_{i,r}\right]\sum_j \exp\left[v^r_{j,s}\right]. \tag{A1}$$

Then, assuming that $s = r$ and summing the last expression over the all vertices in r, one gets

$$\sum_i \langle k^r_{i,r}\rangle = 2\langle E_{rr}\rangle = \left(\sum_i \exp\left[v^r_{i,r}\right]\right)^2, \tag{A2}$$

from which it follows that

$$\langle k^r_{i,r}\rangle = \exp\left[v^r_{i,r}\right]\sqrt{2\langle E_{rr}\rangle}, \tag{A3}$$

and

$$\langle k^s_{i,r}\rangle = \delta_{g_i r}\exp\left[v_i + \frac{\omega_{rs}}{2}\right]\sum_j \delta_{g_j s}\exp\left[v_j + \frac{\omega_{rs}}{2}\right] \tag{A4}$$

$$= \exp\left[v^r_{i,r}\right]\sqrt{2\langle E_{ss}\rangle}\exp\left[\omega_{rs} - \frac{(\omega_{rr}+\omega_{ss})}{2}\right]. \tag{A5}$$

Next, by summing Eq. (A5) over i, one finds for $r \neq s$

$$\exp\left[\frac{(\omega_{rs}-\omega_{rr})}{2} + \frac{(\omega_{rs}-\omega_{ss})}{2}\right] = \frac{\langle E_{rs}\rangle}{2\sqrt{\langle E_{rr}\rangle\langle E_{ss}\rangle}}, \tag{A6}$$

i.e.,

$$\exp\left[\frac{\omega_{rs}-\omega_{rr}}{2}\right] = \sqrt{\frac{\langle E_{rs}\rangle}{2\langle E_{rr}\rangle}}, \tag{A7}$$

and the expression for the expected partial degree, Eq. (A5), can be written as

$$\langle k^s_{i,r}\rangle = \exp\left[v^r_{i,r}\right]\frac{\langle E_{rs}\rangle}{\sqrt{2\langle E_{rr}\rangle}} \tag{A8}$$

$$= \exp\left[v^s_{i,r}\right]\sqrt{\langle E_{rs}\rangle}, \tag{A9}$$

cf. Eqs. (43) and (44).

Now, using Eqs. (A3) and (43), it is straightforward to obtain the expected internal and external degrees of the nodes,

$$\langle k^{\text{int}}_i\rangle = \exp\left[v^r_{i,r}\right]\sqrt{2\langle E_{rr}\rangle} \tag{A10}$$

$$\langle k^{\text{ext}}_i\rangle = \exp\left[v^r_{i,r}\right]\frac{\sum_{s,s\neq r}\langle E_{rs}\rangle}{\sqrt{2\langle E_{rr}\rangle}}, \tag{A11}$$

cf. Eqs. (45) and (46).

[1] S. Fortunato, Phys. Rep. **486**, 75 (2010).
[2] M. E. J. Newman, Nature Physics **8**, 25 (2012).
[3] P. W. Holland, K. B. Laskey, and S. Leinhardt, Soc. Networks **5**, 109 (1983).
[4] K. Faust and S. Wasserman, Soc. Networks **14**, 5 (1992).
[5] C. J. Anderson, S. Wasserman, and K. Faust, Soc. Networks **14**, 137 (1992).
[6] T. A. Snijders and K. Nowicki, J. Classification **14**, 75 (1997).
[7] E. M. Airoldi, D. M. Blei, S. E. Fienberg, and E. P. Xing, J. Mach. Learn. Res. **9**, 1981 (2008).
[8] A. Goldenberg, A. X. Zheng, S. E. Feinberg, and E. M. Airoldi, Found. Trends Mach. Learn. **2**, 1 (2009).
[9] M. Girvan and M. E. J. Newman, Proc. Natl. Acad. Sci. USA **99**, 7821 (2002).
[10] M. E. J. Newman and M. Girvan, Phys. Rev. E **69**, 026113 (2004).
[11] A. Lancichinetti, S. Fortunato, and F. Radicchi, Phys. Rev. E **78**, 046110 (2008).
[12] B. Karrer and M. E. J. Newman, Phys. Rev. E **83**, 016107 (2011).
[13] H.-W. Shen, X.-Q. Cheng, and J.-F. Guo, Phys. Rev. E **84**, 056111 (2011).
[14] T. P. Peixoto, Phys. Rev. E **85**, 056122 (2012).
[15] C. Seshadhri, T. G. Kolda, and A. Pinar, Phys. Rev. E **85**, 056109 (2012).
[16] S. Gregory, Physica A **391**, 2752 (2012).
[17] J. Park and M. E. J. Newman, Phys. Rev. E **70**, 066117 (2004).
[18] A. Fronczak, P. Fronczak, and J. A. Hołyst, Phys. Rev. E **73**, 016108 (2006).
[19] M. E. J. Newman, *Networks: An Introduction* (Oxford University Press, Oxford, 2010), Chap. 15.2, pp. 565–588.
[20] C. R. Shalizi and A. Rinaldo, Annals of Statistics **41**, 508 (2013).
[21] A. Fronczak, arXiv:1210.7828 [physics.soc-ph] (2012).
[22] A. Barrat and A. V. M. Barthélemy, *Dynamical Processes on Complex Networks* (Cambridge University Press, Cambridge, 2008).
[23] C. Song, S. Havlin, and H. A. Makse, Nature (London) **433**, 392 (2005).

[24] C. Song, S. Havlin, and H. A. Makse, Nature Phys. **2**, 275 (2006).
[25] T. A. B. Snijders, J. Soc. Struct. **3**, 1 (2002).
[26] M. E. J. Newman and G. T. Barkema, *Monte Carlo Methods in Statistical Physics* (Oxford University Press, Oxford, 2006).
[27] P. Shannon, A. Markiel, O. Ozier, N. S. Baliga, J. T. Wang, D. Ramage, N. Amin, B. Schwikowski, and T. Ideker, Genome Res. **13**, 2498 (2003).
[28] L. Danon, A. Díaz-Guilera, J. Duch, and A. Arenas, J. Stat. Mech.: Theory Exp. (2005) P09008.