Stochastic fluctuations and the detectability limit of network communities.

Lucio Floretta,[1] Jonas Liechti,[1] Alessandro Flammini,[2] and Paolo De Los Rios[1]

[1] *Laboratoire de Biophysique Statistique, Ecole Polytechnique Fédérale de Lausanne (EPFL), CH-1015 Lausanne, Switzerland*
[2] *School of Informatics and Computing, Indiana University, Bloomington, Indiana 47406, USA*

(Dated: June 11, 2013)

We have analyzed the detectability limits of network communities in the framework of the popular Girvan and Newman benchmark. By carefully taking into account the inevitable stochastic fluctuations that affect the construction of each and every instance of the benchmark, we come to the conclusions that the native, putative partition of the network is completely lost even before the in-degree/out-degree ratio becomes equal to the one of a structure-less Erdös-Rényi network. We develop a simple iterative scheme, analytically well described by an infinite branching-process, to provide an estimate of the true detectability limit. Using various algorithms based on modularity optimization, we show that all of them behave (semi-quantitatively) in the same way, with the same functional form of the detectability threshold as a function of the network parameters. Because the same behavior has also been found by further modularity-optimization methods and for methods based on different heuristics implementations, we conclude that indeed a correct definition of the detectability limit must take into account the stochastic fluctuations of the network construction.

In the process of introducing the most celebrated quantity *modularity* [1], Girvan and Newman (GN) also proposed a method to compare the performance of *community detection algorithms*, which is still at the basis of modern assays [2–4]. In the GN benchmark, different algorithms are tested on a set of *planted l-partition models* [5]. In a nutshell, a planted l-partition model is a network composed of l groups of n vertices that are stochastically connected with each other: the edge between any two vertices within the same group is present with probability p_{in}, whereas the edge between any two vertices belonging to different groups is present with probability p_{out}. Accordingly, the average internal degree is $\langle k_{in}\rangle = (n-1)p_{in}$, whereas the average external degree is $\langle k_{out}\rangle = (l-1)np_{out}$. To measure the extent of the community structure present in the planted l-partition model, it is customary to introduce the mixing parameter μ that is defined by the relations $\langle k_{out}\rangle = \mu\langle k_{tot}\rangle$ and $\langle k_{in}\rangle = (1-\mu)\langle k_{tot}\rangle$, where $\langle k_{tot}\rangle = \langle k_{in}\rangle + \langle k_{out}\rangle$ is the average total degree. Indeed, at $\mu = 0$ the network has l disconnected components; as μ increases, the average internal degree decreases while the average external degree toward one specific other cluster, $\langle k_{out}\rangle/(l-1)$, increases until they become equal at $\mu = (l-1)/l \equiv \mu_c^{ER}$, when the planted l-partition model is, in fact, an Erdös-Rényi graph [6]. The performance of different algorithms is then ranked according to the value of the mixing parameter beyond which they can no longer recover the *native* l clusters. Thus, the GN benchmark tries to encode the simple heuristics that communities correspond to groups of vertices that are more connected with each other than with the rest of the network, although of course a rigorous unambiguous definition of "more connected" is presently lacking.

Naively, one would expect any clustering algorithm to be successful at most up to μ_c^{ER}. However, this is not the case because the construction of a planted l-partition model is a stochastic process. Indeed, it might happen even at $\mu < \mu_c^{ER}$ that a vertex *declared* as belonging to a group has fewer connections toward its putative community than toward a different one. According to the fundamental heuristic definition of community, outlined above, this fluctuation should be interpreted as a change of membership (a *relabeling*) of the vertex. Thus, we can expect that the community structure will be badly degraded when the difference between $\langle k_{in}\rangle$ and $\langle k_{out}\rangle/(l-1)$ is comparable with its own statistical fluctuation, an argument that translates into the condition $\langle k_{in}\rangle - \langle k_{out}\rangle/(l-1) = \alpha_l\sqrt{\langle k_{tot}\rangle}$, where α_l is a positive real constant that may depend on l. Thus, we can argue that the native community structure disappears when

$$\mu > \mu_c^{ER}\left(1 - \frac{\alpha_l}{\sqrt{\langle k_{tot}\rangle}}\right) \tag{1}$$

and any algorithm trying to recover the native partition based on criteria that adhere to the heuristics should consequently fail to do so because, in essence, there is no native structure to be recognized anymore.

In order to gauge the extent to which stochastic fluctuations skew network realizations away from the native one, we have analyzed the planted l-partition models of the original GN benchmark ($l = 4$, $n = 32$, and $\langle k_{tot}\rangle = 16$). Nodes with more connections to an outer group than to their putative one were relabeled, and the procedure was iteratively repeated until no more nodes had to change their membership. For each realization, we measured the similarity between the relabeled partition $\mathcal{R}$ and the native partition $\mathcal{N}$ using the *normalized mutual information* $I_n(\mathcal{R},\mathcal{N})$ [7]. $I_n(\mathcal{R},\mathcal{N})$ is defined as

$$I_n(\mathcal{R},\mathcal{N}) \equiv \frac{I(\mathcal{R},\mathcal{N})}{H(\mathcal{R}) + H(\mathcal{N})}\ ,$$

where

$$I_n(\mathcal{R},\mathcal{N}) = \sum_{r\in\mathcal{R}}\sum_{n\in\mathcal{N}} p(r,n)\log\left(\frac{p(r,n)}{p(r)p(n)}\right) ,$$

is the mutual information of $\mathcal{R}$ and $\mathcal{N}$, $H(\mathcal{R}) = -\sum_{r\in\mathcal{R}} p(r)\log p(r)$ is the entropy of $\mathcal{R}$, and $H(\mathcal{N}) = -\sum_{n\in\mathcal{N}} p(n)\log p(n)$ is the entropy of $\mathcal{N}$. $p(r)$ is the probability that a node belongs to the community r within the relabeled partition $\mathcal{R}$, $p(n)$ is the probability that a node belongs to the community n within the native partition $\mathcal{N}$, and $p(r,n)$ is the probability that a node is simultaneously assigned to the community r within the native partition $\mathcal{R}$ and to the community n within the relabeled partition $\mathcal{N}$.

The outcome is shown in Fig. 1: because $I_n(\mathcal{R},\mathcal{N})$ decreases as μ increases and because it goes to zero well before $\mu_c^{ER} = 3/4$, a labeling closer to what a zero order heuristics would suggest is already different enough from the native partition that the latter could not be detected anymore. As modularity was conceived with the same simple heuristics in mind, we have applied three different clustering algorithms based on its maximization to the networks of Fig. 1: simulated annealing [8], that looks for the absolute maximum, and two *greedy* algorithms, fastgreedy [9] and Louvain method [10]. Both are agglomerative algorithms, but the former acts globally at each step, whereas the latter works locally and in a multistep hierarchical fashion. Interestingly, the modularity of the relabeled partition, $Q(\mathcal{R})$, can be larger than the modularity of the native one, $Q(\mathcal{N})$, in a significant region of μ values. Even more intriguingly, there the modularity found by simulated annealing roughly coincides with the modularity of the relabeled partition, $Q(\mathcal{R})$, and the similarity between the retrieved partition and the relabeled partition is larger than the similarity between the retrieved partition and the putative partition, signaling that modularity optimization and relabeling agree on the detected deviations from the native partition. Greedy methods do not follow the same trend, likely because of the roughness of the modularity landscape, that hampers these algorithms for small values of the total degree [8]; nonetheless they stop agreeing with the native partition in the same region as relabeling and simulated annealing.

We then moved to larger networks, for various values of the total average degrees and for different number of putative communities ($l = 2, 4, 8$). Once again our goal was to use relabeling to discern when stochastic fluctuations completely distort the native partition. We applied then the iterative relabeling procedure outlined above, expectedly finding that the similarity between the relabeled partition and the native one decreased as μ_c^{ER} was approached. The similarity threshold, defined as the value $\mu_c^{\mathcal{R}}$ of the mixing parameter where $I_n(\mathcal{R},\mathcal{N})$ falls below 0.01, was very well described by Eq. (1), up to corrections of order $1/\langle k_{tot}\rangle$. Furthermore, α_l approaches 1

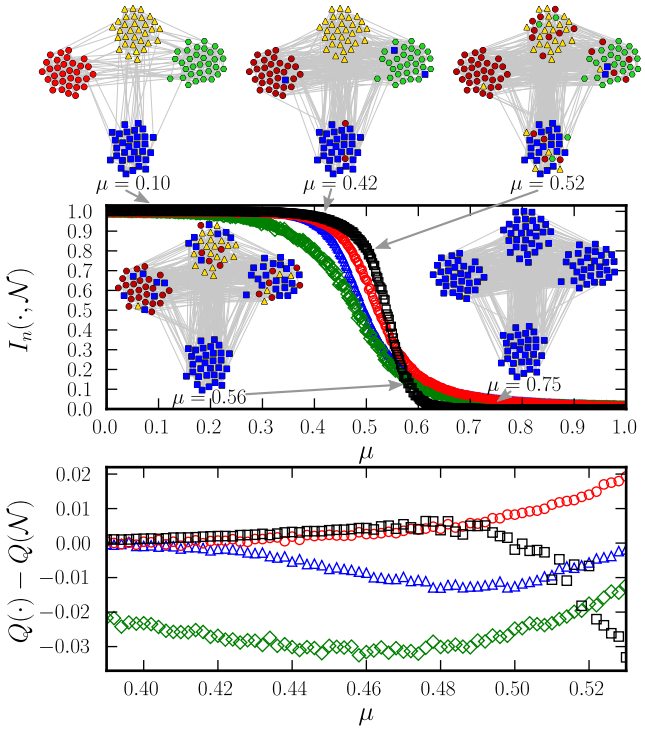

FIG. 1. Similarity (upper panel) along with the modularity difference (lower panel) between the relabeled partition (□), the partitions retrieved by simulated annealing (○ in red), by fastgreedy (◇ in green), and by the Louvain method (△ in blue) with the native partition for the original GN benchmark ($n = 32$, $l = 4$, $\langle k_{tot}\rangle = 16$). Each point is averaged over 500 realizations, the error is smaller than the marker size. Further, five networks at different μ are depicted to visualize how relabeled vertices slowly invade other putative clusters.

from above as l increases (Fig. 2, □), therefore recovering the detectability threshold

$$\hat{\mu}_c \equiv \mu_c^{ER}\left(1 - \frac{1}{\sqrt{\langle k_{tot}\rangle}}\right) \qquad (2)$$

recently derived with random matrix theory approach to modularity maximization [11] and consistent with a previous result obtained from a message passing approach to community detection over block models (solved using the the cavity method) [12, 13], which is not based on modularity. These results confirm thus that the stochastic fluctuations affecting the network construction are so strong to make the native structure disappear before μ_c^{ER}, and consistently with Eq. (1).

The functional form of Eq. (1) for $\tilde{\mu}_c^{\mathcal{R}}$ can be formally derived from the topology of the planted l-partition model and from the aforementioned definition of membership by requiring that each relabeled vertex leaves in its wake further vertices to be relabeled, thus triggering an infinite avalanche. We defined $\mathcal{P}_1^l(\mu, \langle k_{tot}\rangle)$ as the probability that a node has an internal degree that is larger or equal than its number of connections with

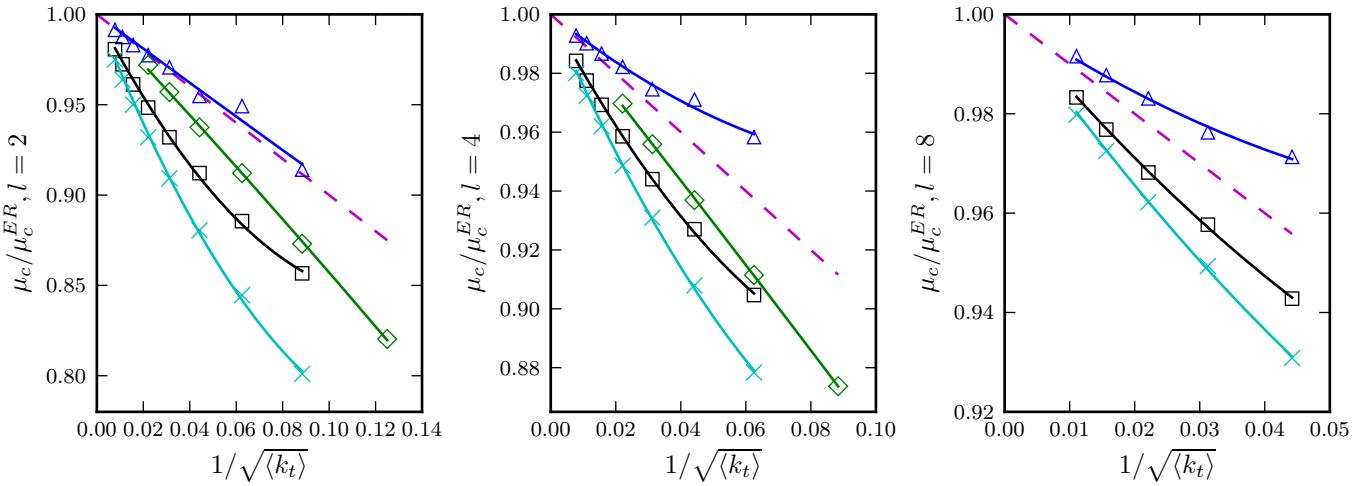

FIG. 2. Comparison between the similarity thresholds $\hat{\mu}_c$ (Eq. 2, dashed line), $\tilde{\mu}_c^{\mathcal{R}}$ (Eq. 3, $\times$), $\mu_c^{\mathcal{R}}$ ($I_n(\mathcal{A},\mathcal{R}) \approx 0.01$, $\square$), μ_c^f ($I_n(\mathcal{A},\mathcal{N}) \approx 0.01$ for fastgreedy, $\Diamond$), and μ_c^L ($I_n(\mathcal{A},\mathcal{N}) \approx 0.01$ for the Louvain method, $\triangle$) for $l = 2$ (left panel), 4 (central panel), and 8 (right panel). In the last case, the results for fastgreedy are not available because the network sizes required to exit the "glassy phase" are out of its reach. The solid lines are fits of the form $\mu_c = \mu_c^{ER}\left(1 - \alpha/\sqrt{\langle k_{tot}\rangle} + \beta/\langle k_{tot}\rangle\right)$. Errors are smaller than markers and to improve the visualization everything is divided by μ_c^{ER}.

nodes of another given group. Assuming that the network is tree-like in the same mean-field spirit of [11–13], in the limit $\langle k_{tot}\rangle \to \infty$ an avalanche is infinite if

$$\left((1-\mu)\langle k_{tot}\rangle - 1\right)\left(1 - \mathcal{P}_1^l(\mu, \langle k_{tot}\rangle)\right) = 1 \,, \qquad (3)$$

as in this limit the probability that a node with a relabeled neighbor maintains its membership converges (exponentially fast) to $\mathcal{P}_1^l(\mu, \langle k_{tot}\rangle)$. For an infinite planted l-partition model

$$\mathcal{P}_1^l(\mu, \langle k_{tot}\rangle) = e^{-\langle k_{tot}\rangle} \sum_{i=0}^{\infty} \frac{1}{i!} \left((1-\mu)\langle k_{tot}\rangle\right)^i \left(\sum_{j=0}^{i} \frac{1}{j!} \left(\frac{\mu\langle k_{tot}\rangle}{(l-1)}\right)^j\right)^{l-1} ; \qquad (4)$$

further, it is possible, although cumbersome, to prove analytically that $\mathcal{P}_1^2(\mu, \langle k_{tot}\rangle)$ depends only on the rescaled mixing parameter $(\mu - \mu_c^{ER})\sqrt{\langle k_{tot}\rangle}$ up to corrections of the form $(a + b/\langle k_{tot}\rangle)$ when $\langle k_{tot}\rangle \to \infty$. Therefore, we made the ansatz $\mathcal{P}_1^l(\mu, \langle k_{tot}\rangle) = \mathcal{F}^l\left((\mu - \mu_c^{ER})\sqrt{\langle k_{tot}\rangle}\right)$, where $\mathcal{F}^l(\cdot) : \mathbb{R} \to [0, 1]$ is a scaling function, which is also numerically confirmed for $l > 2$ (Fig. 3 for $l = 4$).

Taking these results together, the (numerical) solution of Eq. (3) yields $\tilde{\mu}_c^{\mathcal{R}}$ as a function of $\langle k_{tot}\rangle$ and it indeed shows that Eq. (1) is valid up to corrections of order $1/\langle k_{tot}\rangle$ (Fig. 2, $\times$). We can again confirm that it is a general feature of the relabeled community structure to be appreciably different from the "native" one at some $\mu < \mu_c^{ER}$ (and scaling as in Eq. (1)).

Armed with this new view of the resilience to fluctuations of the GN benchmark, we also reanalyzed the performance of modularity optimization methods *fastgreedy* and *Louvain* (we could not test simulated annealing for the sizes under scrutiny here) [4]. In spite of their differences, their behavior is qualitatively similar at low $\langle k_{tot}\rangle$, when the modularity landscape is rough, giving rise to a *glassy phase* [8]: at low values of μ the two algorithms, that just look for local modularity maxima, are not able to find partitions with a modularity as high as the native one, which on average follows the expected mean-field value (Fig. (4), upper panel; we only show the results for the Louvain method, which allows for better statistics on larger networks)

$$Q^{MF}(\mu) = \mu_c^{ER} - \mu \,. \qquad (5)$$

In that region, though, the native partition still fairly

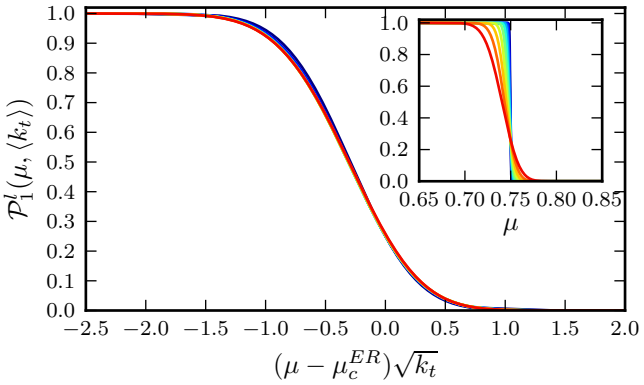

FIG. 3. Magnification of the data collapse of $\mathcal{P}_1^l(\mu, \langle k_{tot}\rangle)$ for $l = 4$ and for different values of $\langle k_{tot}\rangle$ (different colors). The inset displays the original data.

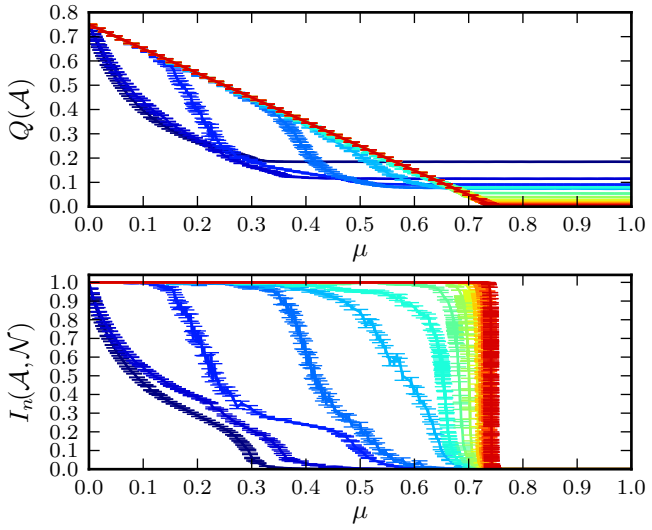

FIG. 4. Modularity (upper panel) of the partition $\mathcal{A}$ retrieved by the Louvain method and its similarity index $I_n(\mathcal{A},\mathcal{N})$ with the native partition $\mathcal{N}$ (middle panel) for the planted 4-partition models with $n = 32768$ and various $\langle k_{tot}\rangle$. Each data point is averaged over 100 realizations; only some errors are shown for clarity.

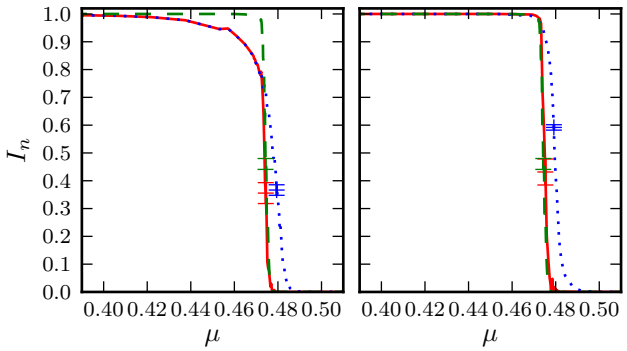

FIG. 5. Comparison close to the phase transition point among $I_n(\mathcal{A},\mathcal{R})$ (solid red line), $I_n(\mathcal{A},\mathcal{N})$ (dotted blue line), and $I_n(\mathcal{R},\mathcal{N})$ (dashed green line) for fastgreedy (left panel) and for the Louvain method (right panel) on a 2-planted partition with $n = 4096$, $\langle k_{tot}\rangle = 2048$. For clarity, only the largest standard deviation is shown.

well corresponds to the heuristics, because there the relabeling procedure involves just a very limited number of nodes, if any (data not shown). Also the similarity index $I_n(\mathcal{A},\mathcal{N})$ between the retrieved partition $\mathcal{A}$ and the native partition $\mathcal{N}$ drops to zero (in a rather erratic way) (Fig. (4), lower panel), confirming the significant difference between the partition retrieved by the algorithms and the native one. Thus, low values of $\langle k_{tot}\rangle$ (sparse networks) do indeed cause detectability problems much more serious than a simple shift of the threshold value of μ, because of the presence of multiple competing modularity maxima. In order to corroborate this interpretation, we initialized the Louvain method with the native planted l-partition, finding that the modularity of the retrieved partition never fell below Eq (5) (data not shown).

The behavior of the two algorithms differs instead outside of the glassy-phase (Fig. 5). The community structure detected by fastgreedy departs from the native partition earlier than the relabeled one, implying that the algorithms does not recover the native partition even when no nodes need to be relabeled. The Louvain method finds clusters that are instead almost identical to the ones found upon relabeling, indicating that the performance of the Louvain method is optimal.

Nevertheless, in the narrow interval defined by the dramatic drop of all similarities, the one between the relabeled partitions and the native partition, $I_n(\mathcal{R},\mathcal{N})$ decreases more rapidly and reaches zero earlier than the ones of both greedy algorithms (Fig. 5), and we quantified this difference by investigating in more detail the similarity thresholds μ_c^f and μ_c^L (now defined as the value of the mixing parameter at which $I_n(\mathcal{A},\mathcal{N})$ drops below 0.01). We found that both μ_c^f and μ_c^L are also well fitted by Eq. (1) with corrections of the order $1/\langle k_{tot}\rangle$ (Fig. 2, $\Diamond$ and $\triangle$, respectively). Moreover $\mu_c^{\mathcal{R}} < \mu_c^f < \hat{\mu}_c \leq \mu_c^L$ and they approach each other as l increases . The differences are likely due to the different ability of each procedure to take into account higher order correlations, beyond the simple one taken into account by the relabeling procedure.

Taking together these observations, we conclude that deciding whether community detection algorithms really fail is trickier than previously believed. Indeed, in order to quantify the degree of success of a clustering method, some sort of ground truth should be known about the network under scrutiny. Since this is in general not available, various benchmarks have been developed, the simplest and more widely used being the GN one, with the naive expectation that, at least for such artificial cases, the *exact* solution should be known. Unfortunately, their construction is itself affected by random fluctuations that blur their structure, making it different from the putative one. Carefully considering such statistical variations, we have shown that it is possible to change the definition of *detectable region*, recognizing that fluctuations disrupt the putative partition for values of $\mu < \hat{\mu}_c$ and well described by Eq. (1), so that one should expect all community detection algorithms to be affected, qualitatively, in the same way, with differences likely to be imputed to different ways to implement mathematically the fundamental heuristics behind the GN benchmark. Indeed, the finding that algorithms based on completely different principles (modularity-based *vs.* statistical inference) show the same threshold beyond which a network structure becomes undetectable, must be reinterpreted in the light of our results: beyond such threshold the real community structure simply bears no resemblance anymore

with the putative one.

The present results also call for a complete reconsideration of benchmarks, from the minimal GN to ones richer in structure [2–4], so that after the construction step a postprocessing procedure should be implemented to correct any skewness introduced by fluctuations. On a more philosophical note, we might even relax the stringency of benchmarking protocols, and resign to the fact that clustering is, at the end of the day, an *ill posed problem* [14].

This work was financially supported by the Swiss National Science Foundation (Grant No. 200021_125277/1).

[1] M. Girvan and M. E. J. Newman, P Natl Acad Sci Usa **99**, 7821 (2002).
[2] A. Lancichinetti, S. Fortunato, and F. Radicchi, Phys Rev E **78**, 046110 (2008).
[3] A. Lancichinetti and S. Fortunato, Phys Rev E **80**, 016118 (2009).
[4] A. Lancichinetti and S. Fortunato, Phys Rev E **80**, 056117 (2009).
[5] A. Condon and R. M. Karp, Random Struct Algor **18**, 116 (2001).
[6] P. Erdös and A. Rényi, Publ. Math. Debrecen **6**, 290 (1959).
[7] L. Danon, A. Díaz-Guilera, J. Duch, and A. Arenas, J Stat Mech-Theory E **2005**, P09008 (2005).
[8] B. H. Good, Y.-A. de Montjoye, and A. Clauset, Phys Rev E **81**, 046106 (2010).
[9] A. Clauset, M. E. J. Newman, and C. Moore, Phys Rev E **70**, 066111 (2004).
[10] V. D. Blondel, J.-L. Guillaume, R. Lambiotte, and E. Lefebvre, J Stat Mech-Theory E , P10008 (2008).
[11] R. Nadakuditi and M. E. J. Newman, Phys Rev Lett **108**, 188701 EP (2012).
[12] A. Decelle, F. Krzakala, C. Moore, and L. Zdeborová, Phys Rev Lett **107**, 065701 (2011).
[13] A. Decelle, F. Krzakala, C. Moore, and L. Zdeborová, Phys Rev E **84**, 066106 (2011).
[14] E. Domany, Physica A **263**, 158 (1999).