PHYSICAL REVIEW E **79**, 066111 (2009)

Alternative approach to community detection in networks

A. D. Medus* and C. O. Dorso†
Departamento de Física, Facultad de Ciencias Exactas y Naturales, Universidad de Buenos Aires, Pabellón 1, Ciudad Universitaria, Ciudad Autónoma de Buenos Aires 1428, Argentina

(Received 28 July 2008; revised manuscript received 21 May 2009; published 23 June 2009)

The problem of community detection is relevant in many disciplines of science and modularity optimization is the widely accepted method for this purpose. It has recently been shown that this approach presents a resolution limit by which it is not possible to detect communities with sizes smaller than a threshold, which depends on the network size. Moreover, it might happen that the communities resulting from such an approach do not satisfy the usual qualitative definition of commune; i.e., nodes in a commune are more connected among themselves than to nodes outside the commune. In this paper we present a different method for community detection in complex networks. We define merit factors based on the weak and strong community definitions formulated by Radicchi *et al.* [Proc. Natl. Acad. Sci. U.S.A. **101**, 2658 (2004)] and we show that these local definitions avoid the resolution limit problem found in the modularity optimization approach.

DOI: 10.1103/PhysRevE.79.066111 PACS number(s): 89.75.Hc, 89.75.Fb, 87.23.Ge

I. INTRODUCTION

The problem of community detection in complex networks has recently attracted the attention of researchers in different areas of scientific knowledge. This is due to the fact that it is a common practice to represent some complex systems as networks constituted by interconnected nodes.

A network G is defined by a set of nodes $\{n\}$ $(n_1,n_2,\ldots,n_n)$ and a set of links $\{l\}$ $(l_{12},l_{14}\ldots,l_{km})$. A link l_{ij} denotes a relation between node n_i and node n_j. Depending on the possible values of l_{ij} the resulting network can be of two types. If l_{ij} can only have the values 1 or 0, we will call the network unweighted; on the other hand, a network will be defined as weighted if l_{ij} can attain values different from 0 or 1, thus indicating that the relation between nodes is also characterized by a given strength. In this work we will focus on unweighted networks. We will assume that for every node n_i there exists at least another node n_j such that l_{ij} is different from 0; moreover we will consider networks such that for every conceivable pair of nodes there will be a path (i.e., a sequence of links $\{l_{ij}l_{jk}l_{km}\ldots\}$) joining them, in such a case we say that we are dealing with connected networks. We will consider that the links are undirected i.e., $l_{ij}=l_{ji}$. Further on, we will focus on sparse networks for which the number of links in $\{l\}$, L, is much less than the maximum possible number of links, $L_{\max}$ given by $L_{\max}=N(N-1)/2$, with N the total number of nodes in $\{n\}$.

Generally, complex networks contain a large number of nodes and links and it is often possible to decompose them into subgraphs called *communities* selected according to a given criterion. A community is usually defined, qualitatively, as a subgraph of the network whose nodes are more connected among them than to nodes outside the subgraph [1,2].

Community detection has a wide range of applications. The partition of a network in communities might allow us to find a specific function naturally assigned to each community, as for example in the case of metabolic networks [3]. On the other hand, community detection can help us to identify social groups in a social network or can be used to perform a coarse-graining reduction in the network to simplify subsequent analysis [4].

There are many methods to decompose a network into communities, but the widely adopted in recent years is the one proposed by Newman and Girvan [2]. These authors define a merit factor named *modularity* (Q_N) that quantifies the quality of a given m-subgraphs partition $\{C_j\}_{1\leq j\leq m}$ of the graph G, where $\cup_{j=1}^{m}C_j=G$ with $C_i\cap C_j=\varnothing$ if $i\neq j$. This quantity measures the difference between the actual fraction of internal links in each subgraph with respect to the expected value of the same quantity if nodes in the network are randomly connected keeping the degree of each one fixed. The best partition of the network is taken as the one that maximizes the modularity Q_N; in this way, the network partition problem is turned into an optimization one.

Modularity optimization is a hard problem due to the fact that the number of possible partitions of a network increases at least exponentially with its size. Indeed, it has recently been proven that this problem is nondeterministic polynomial time (NP) complete [5], and then, there is no correct polynomial-time algorithm to solve it for networks of any size. Many optimization algorithms have been developed, such as simulated annealing [3,6], extrema optimization [7], and spectral division [8], but all of them can only give an approximation to the optimum partition for large networks. In this work, we do not introduce a new Q_N optimization algorithm, but we propose different merit factors for the calculation of the partition of networks into communities.

The modularity Q_N is a *nonlocal* community definition in the sense that it is necessary to know general characteristics of the whole network in order to decide if a given subgraph of the network is a community. In a recent paper Fortunato and Barthélemy [1] have shown that this nonlocal character imposes a resolution limit, by which the minimal community size that can be detected, by modularity optimization, depends on global network parameters. Then, Q_N optimization

*admedus@gmail.com
†codorso@df.uba.ar

 ©2009 The American Physical Society

is not able to detect communities of size smaller than a given threshold.

In this work we use the weak and strong community definitions proposed by Radicchi *et al.* in [9]. We emphasize their local character and present merit factors to evaluate the quality of a given partition of a network based on these quantitative definitions of community. Then, we implement an optimization method in the spirit of simulated annealing [6,10], in order to analyze different networks using these merit factors and to show the characteristics of our approach. Finally, we show that the resolution limit problem does not appear in our approach.

The paper is organized as follows. In Sec. II we review the definitions introduced in [9] and compare them with the modularity Q_N; we also analyze the meaning of this last quantity. In Sec. II A we define the *community strength S* in the strong and weak sense and we introduce the associated merit factors. In Sec. III we apply our method to different, well known, networks. In Sec. IV we analyze the resolution limit problem for our approach. Finally, conclusions are drawn in Sec. V.

II. COMMUNITY DEFINITIONS

When thinking about communities in networks we have in mind a qualitative community definition: a community is a group of nodes in which the number of internal links, connecting nodes within the group, is larger than the number of external ones. In order to formalize this qualitative criterion we consider a graph G containing N nodes, with k_i the degree of node $i \in G$. If C is a subgraph of G with k_i^{in} and k_i^{out} the number of links of node $i \in C$ that connect it to nodes inside and outside of C, respectively. There are two quantitative community definitions introduced by Radicchi *et al.* [9]:

(i) Community in strong sense. C is a community in the strong sense if

$$k_i^{in} > k_i^{out} \quad \forall\, i \in C. \tag{1}$$

(ii) Community in weak sense. C is a community in weak sense if

$$\sum_{i \in C} k_i^{in} > \sum_{i \in C} k_i^{out}. \tag{2}$$

In words: a subgraph $C \subset G$ will be a community in the strong sense if each of its nodes has more links connecting it with nodes in C than those that connect it with other nodes not belonging to C. In a similar way, $C \subset G$ will be a community in the weak sense if the sum of the number of links that interconnect nodes inside C is larger than the sum of all links that connect nodes in C with nodes not belonging to C. These community definitions are simple, intuitive, and *local*: given a subgraph $C \subset G$ we can decide if it constitutes a community, in either strong or weak sense, without knowledge of the entire structure of G.

In order to compare the previous approach with the one proposed in [2], we briefly review the definition and meaning of the modularity Q_N. Given a m-subgraphs partition $\{C_j\}_{1\le j\le m}$ of the graph G, where $\cup_{j=1}^{m} C_j = G$, the mathematical expression of Q_N is

$$Q_N = \sum_{i=1}^{m} \left[\frac{l_i}{L} - \left(\frac{d_i}{2L} \right)^2 \right], \tag{3}$$

where l_i denotes the total number of internal links for subgraph $C_i \subset G$ and $d_i = \sum_{j \in C_i} k_j$, and $L = \frac{1}{2}\sum_{j \in G} k_j$ is the total number of links in G.

The term l_i/L in Eq. (3) denotes the actual fraction of internal links in subgraph C_i, while $d_i/2L$ can be interpreted as the probability of a link to be connected to a node in subgraph C_i. Then, $(d_i/2L)^2$ is the expected fraction of links within subgraph C_i when all nodes in G are randomly connected keeping the degree of the nodes fixed. This last ideal random picture is used to compare with the actual one because it is assumed that corresponds to a situation with no communities (although it was shown in [11] that random networks may have a community structure).

As already mentioned, the modularity Q_N was conceived as a measure of the goodness of a given partition of the network. Then, the bigger Q_N is, the better the partition is. We should notice that this merit factor implies, in turn, a community definition (which does not necessarily corresponds to the intuitive one stated above): a subgraph C_j will be a community if the actual number of links that connects nodes in C_j is bigger than the expected one when all nodes in the network are randomly connected, this is to say, when $l_i/L-(d_i/2L)^2>0$. Clearly this last condition depends on the global parameter L, then, we say that the community definition associated with Q_N is nonlocal. In what follows, we will present merit factors associated with the weak and strong community definitions.

A. Merit factors for weak and strong community definitions: Community strength

Given a graph G and a m-subgraphs partition $\{C_j\}_{1\le j\le m}$, where each subgraph $C_j \subset G$ constitutes a community according to any of the local definitions mentioned in the previous section, we want to define a quantity that measures the "quality" of each of the resulting communities. In the context of the above-mentioned local framework, this quantity must only depend on the local characteristics of the subgraph C_j. Therefore, our analysis must be circumscribed to nodes and links belonging to C_j and external links that connect nodes in C_j to nodes outside C_j. Following the weak and strong definitions of community, the more internal links a community has, with respect to the external ones, the "stronger" it will be. If $k_i = k_i^{in} + k_i^{out}$ is the degree of node $i \in C_j$, where k_i^{in} and k_i^{out} are the number of internal and external links for node i, we define the "community strength" (S) that measures the normalized difference between internal and external links for nodes in C_j:

$$S(C_j) = \sum_{i \in C_j} \frac{k_i^{in} - k_i^{out}}{2L(C_j)}, \tag{4}$$

where $L(C_j) = \frac{1}{2}\sum_{i \in C_j} k_i$. Then, $-1 \le S(C_j) \le 1$, and it achieves its maximum value 1 when $k_i^{out}=0$ $\forall i \in C_j$.

The definition of $S(C_j)$ according to Eq. (4) is valid for unweighted networks. In the case of weighted links, we have to interpret k_i as the sum of the weights of the links that connect to node i, for both k_i^{in} and k_i^{out}.

Now we introduce a merit factor Q_W for the weak community definition as the sum of $S(C_j)$ over all subgraphs $C_j \subset G$:

$$Q_W = \sum_{j=1}^{m} S(C_j) = \sum_{j=1}^{m} \sum_{i \in C_j} \frac{k_i^{in} - k_i^{out}}{2L(C_j)} \tag{5}$$

with the constraint that each subgraph $C_j \subset \{C_j\}_{1 \le j \le m}$ must satisfy the weak community definition, i.e.,

$$S(C_j) > 0 \quad \forall\, C_j \subset \{C_j\}_{1 \le j \le m}. \tag{6}$$

As in the case of Q_N: the bigger Q_W is, the better the m-subgraphs partition $\{C_j\}_{1 \le j \le m}$ of G will be, in the sense of weak community definition. Then, it is possible to implement the optimization algorithms developed for Q_N for this merit factor Q_W,

In Eq. (5), $Q_W = 1$ when all the network constitutes a single community. If as a result of the optimization process the maximum value obtained for Q_W is precisely $Q_W = 1$ and we get a single community, then the best partition of the graph corresponds to no partition. However, it is possible that one could get $S(C_j) > 0$ $\forall C_j \subset \{C_j\}_{1 \le j \le m}$ for a given m-subgraphs partition, with $m > 1$ but with $0 < Q_W < 1$. The resulting community structure of network would correspond to a suboptimal partition.

In the same spirit we now define a merit factor Q_S according to the strong community definition:

$$Q_S = \sum_{j=1}^{m} S(C_j) = \sum_{j=1}^{m} \sum_{i \in C_j} \frac{k_i^{in} - k_i^{out}}{2L(C_j)} \tag{7}$$

with the constraint

$$(k_i^{in} - k_i^{out}) > 0 \quad \forall\, i \in C_j. \tag{8}$$

Now, our definition of optimal partition can be stated in the following way:

Definition. The optimal m-subgraphs partition $\{C_j\}_{1 \le j \le m}$ of a graph G in the strong (weak) sense is that one with maximal merit factor $Q_S(Q_W)$.

In the next section we will show some examples of the application of these merit factors in network partition problems.

III. EXAMPLES

In all examples presented in this section we have used an optimization algorithm based on simulated annealing, described in [6], but for our merit factors. The optimization can be performed in two ways. In the first one the total number of communes is left as a free parameter and as a consequence the final number of communes is determined by the simulated annealing process. In the second one, the number of communes is taken as an extra constraint. We will always use the first approach unless it is explicitly stated that the number of communes is fixed. This last methodology might be used when the optimal number of communes is already known from experiments as in the case of the Zachary network.

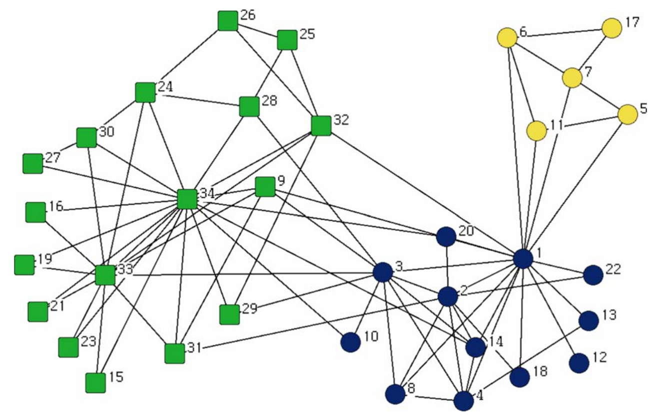

FIG. 1. (Color online) Best partition for Zachary network. Squares and circles denote the two communities obtained with our approach when the number of communities is fixed to two. This partition corresponds to the one consigned by Zachary in [12], with the exception of node 10 that appears misclassified.

A. Zachary's karate club network

We will begin with a typical case: Zachary's karate club [12], which has turned into an unavoidable example in publications about community structure. This network represents the relationships between members of a karate club at a University in the 1970s and it has been shown that it has a strong community structure in many previous studies [2,6]. Applying the optimization algorithm for the weak community definition merit factor Q_W, we obtained, for the unweighted version of Zachary network, a partition into three communities of sizes: $17(C_1)$, $12(C_2)$, and $5(C_3)$ nodes, with a value of $Q_W = 1.792$ (Fig. 1). When the number of communities was constrained to two, we obtained two communities of 17 nodes each, with $Q_W = 1.487$ (circles and squares in Fig. 1). This partition corresponds to the one observed by Zachary with the exception of node 10 that is misclassified.

With this analysis we can know, in addition, the strength $S(C_j)$ of each community C_j in the network. For the best partition of the Zachary network into three communities of $17(C_1)$, $12(C_2)$, and $5(C_3)$ nodes, we have $S(C_1) = 0.744$, $S(C_2) = 0.548$, and $S(C_3) = 0.5$, with C_1 as the strongest community. On the other hand, the partition into two communities is composed by two strong communities of 17 nodes each, with $S(C_j) = 0.744$ for each one.

When we perform the community analysis using the strong community merit factor Q_S, we obtained two communities: C_1 with 29 nodes $[S(C_1) = 0.943]$ and C_2 with five nodes $[S(C_2) = 0.5]$. In Fig. 1 it can be observed that node 10 has one internal and one external link and this situation is not allowed in the strong community definition. For this reason, the communities with 17 and 12 nodes are joined together.

B. Star network

Another testing example is the star network of Fig. 2 consisting of two interconnected stars with 11 nodes each. In the weak community picture we obtain for the optimal community structure a partition into two communities of 11 nodes each with a value of $Q_W = 1.652$. The corresponding strengths

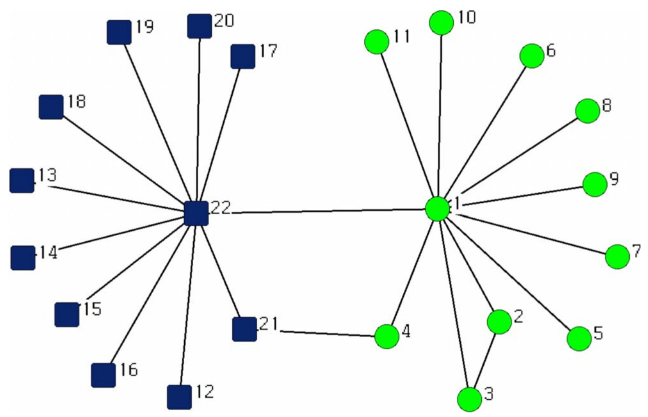

FIG. 2. (Color online) Community structure for star network obtained in our approach. The two communities are distinguished with squares and circles. For Q_N optimization we find an additional community containing nodes 4 and 21 with strength $S=0$.

are $S(C_1)=0.833$ (circles in Fig. 2) and $S(C_2)=0.818$ (squares in Fig. 2). The difference in strength between the two communities is ascribed to the extra link that connects nodes 2 and 3.

We must notice that, in the same context, there is another partition with the same value of Q_W but composed by two communities with 12 and 10 nodes. This happens when node 4 is moved from one community to the other in Fig. 2.

No partition was obtained when we use the strong community merit factor Q_S. This is due the fact that nodes 4 and 21 are singly connected and then the condition $k_i^{in}-k_i^{out}>0$ is not satisfied.

When Q_N optimization was implemented, we obtained three communities: C_1 and C_2 with ten nodes each and C_3 with two nodes, including nodes 4 and 21, with strength $S(C_3)=0$ which does not satisfy any of the quantitative community definitions reviewed in Sec. II.

C. Ring network

Another example is the ring network of Fig. 3 with 20 nodes and $k=6$. This network cannot have a community structure due to its symmetry. However, we have obtained two communities of ten nodes each, with strength $S=0.6$ by means of the weak merit factor optimization. We must notice here that the found communities are not unique, that is to say, on having applied repeatedly the algorithm different communities of the same size but involving different sequences of node indexes are obtained. This is an evidence of the absence of an underlying community structure. This unsatisfactory result is also obtained when optimizing the Q_N merit factor, but in this case the optimal partition is into three communities, two of them with seven nodes each and the third one with six nodes.

On the other hand, when we run the optimization algorithm with the strong community condition, no partition is obtained. This is true for all ring networks because a hypothetical frontier node, with $k_i^{in}=k_i^{out}$, will not satisfy the strong community condition.

D. Bottlenose dolphins network

Another social network that has attracted considerable interest is the one corresponding to the bottlenose dolphins

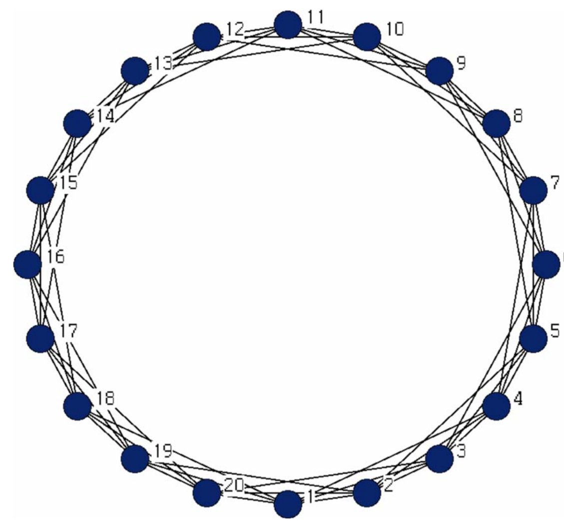

FIG. 3. (Color online) Ring network. This symmetric network does not present an intrinsic community structure. However, almost all approaches for community detection find a community structure for this example, but the identities of the nodes within each community change when the detection process is repeated.

network, which has been fully analyzed in [13] (see also [2]). This small social network is composed of 62 nodes and it is known to consist of two communities of sizes 41 and 21 nodes each. Following the approach proposed in this work we first analyze this network applying the Q_N analysis in our simulated annealing approach. The result of this analysis is the partition of the network into four communities composed by 21, 16, 13, and 12 nodes each. When we performed the optimization of the weak community definition we obtained five communities of 20, 12, 11, 10, and 9 nodes each. Finally when the dolphin network is analyzed in terms of the strong community definition we obtained the actual partition, as observed experimentally, in two communities of 41 and 21 nodes each. These last two results are displayed in Fig. 4. In this figure we show the two communities according to the strong community definition as circles (41 nodes community) and as squares (21 nodes community). The corresponding analysis according to the weak community definition further divides the previous two communities and are denoted by the different shades of gray (see caption for details) in the figure. It should be noted at this point that when the optimization of Q_W is performed with the extra constraint that the number of communes is 2 we obtain the same community structure as observed experimentally.

E. Computational generated test network

We conclude our short list of examples with the analysis of computer generated graphs that have a community structure. A word of caution should be raised at this point because the communities built into these graphs are usually of uncertain nature and being purely theoretical, the assumed community structure cannot be "verified experimentally" as in the case of the Zachary karate club or the bottlenose dolphins case.

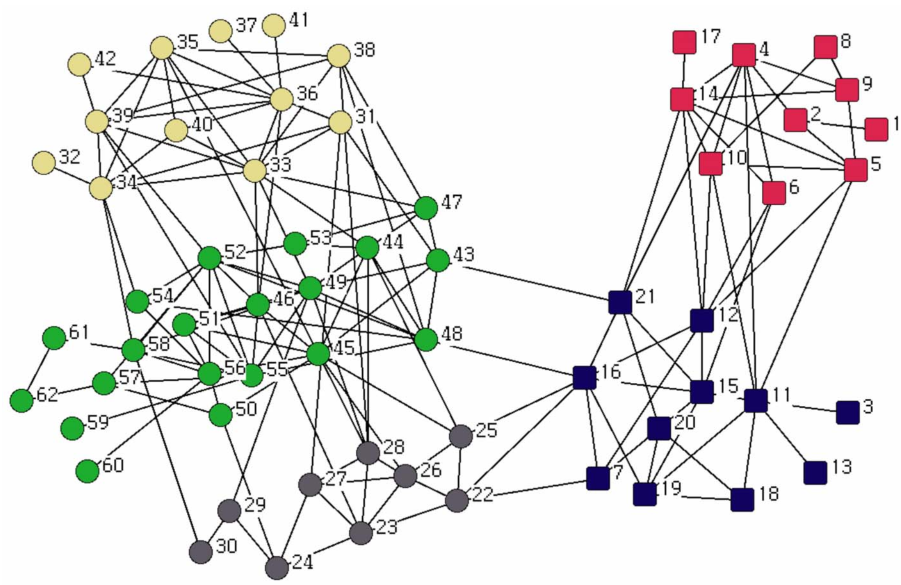

FIG. 4. (Color online) Bottlenose dolphin network. This network has a size of 62 nodes and it is known from direct observation that it has two communities. In this figure squares and circles denote the communities detected by our strong community approach and the colors (shades of gray or colors online) show the results of the weak community approach. Notice that the optimization according to Q_W merely subdivides the communities obtained through Q_S optimization [17].

1. Methodology I

In this case we use the method proposed and analyzed in [7,14]. We take a 128 nodes graph G divided in four modules $C_{j=1,\ldots,4}$ of 32 nodes each and with nodes degree $k_i = k_i^{in} + k_i^{out} = 16 \;\; \forall i \in G$. We have earlier defined $k_i^{in}(k_i^{out})$ as the number of links that connect node $i \in C_j$ to another node in (out of) C_j. When k_i^{out} is varied between 0 to 16, G goes from a strong communality graph to a quasirandom one.

In the framework of weak and strong merit factors optimization the expected four communities partition was obtained for $0 \le k_i^{out} \le 7$, where $Q_W < 1$ only for $k_i^{out} = 7$. When $k_i^{out} \ge 8$ the mean number of external links is bigger or equal to internal ones in each module and no partition was obtained.

2. Methodology II

We now use the formalism introduced in [15], in which an algorithm for generating a class of benchmark graphs that accounts for the heterogeneity in the distributions of node degrees and of community sizes was devised. The aim of this algorithm is to build a graph with more or less well defined community structure (in the caption of Fig. 5 in [15] a reference is made to “communities in the strong sense” which is not the case for this algorithm as is easily verified). It is assumed that both the degree and the community size distributions are power laws, with exponents γ and β, respectively. The number of nodes is N and the average degree is $\langle k \rangle$. One more parameter characterizing this model is the mixing parameter μ. Each node shares, in average, a fraction $1-\mu$ of its links with the other nodes of its community and a fraction μ with the other nodes of the network.

We have generated graphs according to this algorithm and we have analyzed them using the Girvan-Newman definition of community using our global optimization approach [6] and the one proposed in this work (in the weak sense). The parameters defining the graphs were chosen to be $\gamma = 2.5$, $\beta = 1.5$, $0.1 \le \mu \le 0.6$. The size of the graphs was fixed in 300 nodes and the mean degree in 8 (maximum degree=30).

The results of such a calculation are summarized in Fig. 5 (see captions for details). The quality of the partitions obtained with the recognition algorithms with respect to the communities established by the algorithm of Lancichinetti *et al.* is measured by means of the so-called *normalized mutual information* [16]. According to the results displayed in Fig. 5, our merit factor definitions outperform the Girvan-Newman approach.

IV. RESOLUTION LIMIT PROBLEM

In a recent paper Fortunato and Barthélemy [1] showed that modularity Q_N optimization fails to detect communities smaller than a certain threshold, which depends on global parameters of the network under study, as is the case of the total number of links in the network. Following [1] we can

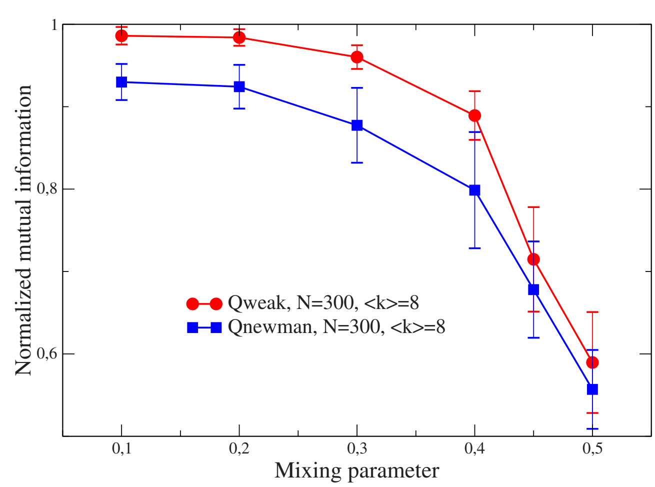

FIG. 5. (Color online) Computer generated test network. We have generated graphs according to the formalism proposed by Lancichinetti *et al.* The results of our calculation are displayed by the full circles (in red) while the results according to the Girvan-Newman approach are denoted by full squares (in blue).

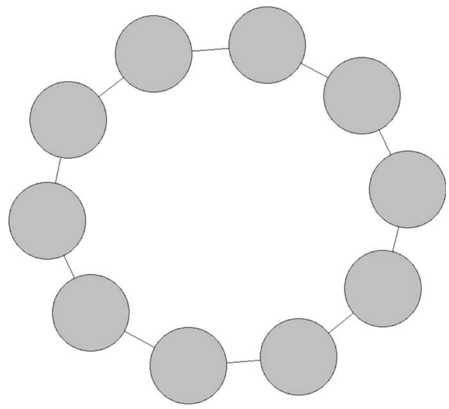

FIG. 6. Ring of cliques. Each circle represents a clique that is a totally connected subgraph with n nodes.

define a community from Eq. (3), in the framework of modularity Q_N, as a subgraph $C_i \subset G$ that satisfies

$$\frac{l_i}{L} - \left(\frac{d_i}{2L}\right)^2 > 0. \quad (9)$$

This expression can be interpreted as the community strength in this framework. We can write $d_i = 2l_i + l_i^{out}$, where l_i and l_i^{out} denote the number of internal and external links for subgraph C_i, and write l_i^{out} as a fraction of internal links $l_i^{out} = al_i$ with $a \geq 0$. Then, from Eq. (9), the following condition for the community size l_i is obtained:

$$l_i < \frac{4L}{(a+2)^2}. \quad (10)$$

The dependency on total number of links L in Eq. (9) clearly shows that the community definition, in the context of modularity Q_N, is nonlocal. In [1] the authors show that this nonlocality is the origin of the *limit resolution problem*.

On the other hand, in our approach the community definition is strictly local. Then, we can decide if a subgraph is a community without regards to the size of the entire network. To illustrate this conclusion we work out an example introduced in [1].

Let us suppose a ring of totally connected subgraphs (from now on *cliques*) in Fig. 6. Each subgraph has n nodes connected by $n(n-1)/2$ internal links and two external ones, and we have M of this subgraphs with a total number of links $L = M[n(n-1)/2+1]$. The optimal partition for strong and weak community definition frameworks is the natural one: each clique constitutes a single community. This can be easily shown by analyzing another alternative partition in which each community C_j contains $W \geq 2$ cliques.

First, we calculate the strength $S(C_j)$ for one of these subgraphs containing W cliques:

$$S(C_j) = \frac{W(n-2)(n-1) + 2(W-1)n + 2(n-2)}{W(n-1)n + 2W}. \quad (11)$$

Now, we want to compare the result of Eq. (11) with the total strength of the same subgraph C_j when each clique is taken as a single community. In this case, the strength of C_j is given by W times the strength of one clique:

$$S^*(C_j) = W\frac{(n-2)(n-1) + 2(n-2)}{n(n-1) + 2}. \quad (12)$$

Then, it is straightforward to see that $S^*(C_j) > S(C_j)$ is equivalent to

$$W(n-2)[(W-1)(n-1) + 2W] - 2(nW-2) > 0. \quad (13)$$

We have said that each clique constitutes a single community in the strong (and then, also in the weak) community definition, therefore, $n \geq 3$. Then, the condition of Eq. (13) is satisfied for all $W \geq 2$. This is to say that the optimal partition in our approach is that one for which each clique constitutes a single community. This is not the general case in the modularity Q_N framework. It was showed in [1] that, due to resolution limit problem, partitions in communities with two or more cliques can give larger values of Q_N than with single clique communities.

V. CONCLUSIONS

In this work we have proposed different merit factors to recognize communities in networks. These merit factors are more realistic than the ones currently in use in the literature because they strictly adhere to what a community is expected to be, i.e., a subset of nodes which are more connected among themselves than to the rest of the network under consideration.

We started by putting forward this qualitative definition of a community and then we reviewed the meaning of the quite popular measure of the quality of a given partition known as the modularity Q_N. As we have discussed above, the community definition associated to this quantity is nonlocal and does not necessarily correspond to the aforementioned qualitative definition. One of the consequences of the nonlocal character intrinsic to this quantity is the limit resolution problem as stated in [1].

In order to recognize communities in networks that strictly adhere to the qualitative definition, we have used (following [9]) two local community definitions: the weak one and the strong one. In order to use these definitions to recognize communities we have developed criteria to quantify the strength of a community (S). Afterward, we have defined two merit factors associated with S, which we named Q_S and Q_W. As with Q_N the problem of recognizing communities in a network is mapped onto an optimization problem; i.e., the communities in a network are the elements of the partition that maximizes Q_S or Q_W. We have performed the optimization of these merit factors on some standard networks by implementing an algorithm in the spirit of simulated annealing. The limit resolution intrinsic to the Q_N definition is not present in our approach.

It is worth noticing at this point that the solution to the detection of communities in the strong sense is also a solution in the weak sense but not necessarily optimal. On the other hand, the converse is generally not true as we stated in Sec. III.

The strong community definition tends to give larger communities because of its inability to deal with nodes that are equally shared by two highly connected subgraphs, but on the other hand has the nice property that it is the only one that gives no partition for symmetric string networks and also solves the problem of the bottle nose dolphins network exactly without constraints in the number of communities.

ACKNOWLEDGMENT

C.O.D. acknowledges partial support from CONICET through Grant No. PIP5969.

[1] S. Fortunato and M. Barthélemy, Proc. Natl. Acad. Sci. U.S.A. **104**, 36 (2007).

[2] M. E. J. Newman and M. Girvan, Phys. Rev. E **69**, 026113 (2004).

[3] R. Guimerà and L. A. N. Amaral, Nature (London) **433**, 895 (2005).

[4] A. Arenas, J. Dutch, A. Fernandez, and S. Gómez, New J. Phys. **9**, 176 (2007).

[5] U. Brandes, D. Delling, M. Gaertler, R. Görke, M. Hoefer, Z. Nikoloski, and D. Wagner, IEEE Trans. Knowl. Data Eng. **20**, 172 (2008).

[6] A. Medus, G. Acuña, and C. O. Dorso, Physica A **358**, 593 (2005).

[7] J. Duch and A. Arenas, Phys. Rev. E **72**, 027104 (2005).

[8] M. E. J. Newman, Proc. Natl. Acad. Sci. U.S.A. **103**, 8577 (2006).

[9] F. Radicchi, C. Castellano, F. Cecconi, V. Loreto, and D. Parisi, Proc. Natl. Acad. Sci. U.S.A. **101**, 2658 (2004).

[10] C. O. Dorso and J. Randrup, Phys. Lett. B **301**, 328 (1993).

[11] R. Guimerà, M. Sales-Pardo, and L. A. N. Amaral, Phys. Rev. E **70**, 025101(R) (2004).

[12] W. W. Zachary, J. Anthropol. Res. **33**, 452 (1977).

[13] D. Lusseau, Proc. R. Soc. London, Ser. B **270**, S186 (2003).

[14] B. Karrer, E. Levina, and M. E. J. Newman, Phys. Rev. E **77**, 046119 (2008).

[15] A. Lancichinetti, S. Fortunato, and F. Radicchi, Phys. Rev. E **78**, 046110 (2008).

[16] L. Danon, A. Díaz-Guilera, J. Duch, and A. Arenas, J. Stat. Mech.: Theory Exp. (2005), P09008.

[17] All figures have been drawn using NetDraw, http://www.analytictech.com.