Narrow scope for resolution-free community detection

V.A. Traag [*], **P. Van Dooren** [*] **and Y. Nesterov** [†]

[*]ICTEAM, Université Catholique de Louvain, Avenue G. Lemaître 4, B-1348 Louvain-la-neuve, Belgium, and [†]CORE, Université Catholique de Louvain, Voie du Roman Pays 34, B-1348 Louvain-la-neuve, Belgium

Submitted to Proceedings of the National Academy of Sciences of the United States of America

Detecting communities in large networks has drawn much attention over the years. While modularity remains one of the more popular methods of community detection, the so-called resolution limit remains a significant drawback. To overcome this issue, it was recently suggested that instead of comparing the network to a random null model, as is done in modularity, it should be compared to a constant factor. However, it is unclear what is meant exactly by 'resolution-free', i.e. not suffering from the resolution limit. Furthermore, the question remains what other methods could be classified as resolution-free. In this paper we suggest a rigorous definition and derive some basic properties of resolution-free methods. More importantly, we are able to prove exactly which class of community detection methods are resolution-free. Furthermore, we analyze which methods are not resolution-free, suggesting there is only a limited scope for resolution-free community detection methods. Finally, we provide such a natural formulation, and show it performs superbly.

complex networks | community detection | modularity | resolution limit | resolution-free

Introduction

The last decade has seen an incredible rise in network studies, and will likely continue to rise [1, 2]. Besides the study of properties such as degree distributions, clustering coefficients, average path length and other network characteristics [3], many complex networks exhibit some modular structure [4, 5]. These communities might represent different functions represented by the nodes (e.g. metabolic functions) or sociological communities, and have been successfully studied on a wide variety of networks, ranging from metabolic networks [6] to mobile phone networks [7] and airline transportation networks [8].

One of the most popular methods for community detection is that of modularity, introduced by Newman and Girvan [9]. The past few years suggestions have been made to extend or alter the original definition, for example, allowing detection in bipartite networks [10], networks with negative links [11], and dynamical networks [12]. Although modularity optimization seems to be able to accurately identify known community structures [13, 14], it suffers from an inherent difficulty, namely a resolution limit [15], which affects the practice of community detection [16]. This resolution limit prevents detection of smaller communities in large networks, although this effect is mitigated somewhat by the presence of a so called resolution parameter [17], which can be related to time scales of random walks on the network [18]. One example of dealing with this problem is looking at how 'stable' partitions are at various resolutions [19, 20] although the resolution problem remains intrinsic to the formulation of modularity.

Recently, a new method has been suggested that would not suffer from this resolution limit, by Ronhovde and Nussinov (RN) [20], using a reasoning similar to that of Reichardt and Bornholdt (RB) [21]. The initial claim that a method suffers from a resolution limit can be clearly demonstrated on a few cases (e.g. counterintuitive merging of weakly linked cliques), but the opposite seems more difficult to argue. That is, although there does not seem to be any dependence on global variables in the cases analyzed, perhaps more complex cases will show some issues not yet considered. Hence, a proper definition of what it means to be *resolution-free* is required, which we will develop in this paper. Furthermore, the question remains what type of community detection methods will exactly suffer from this resolution limit and which not. Intuitively, this is related to the method being 'local', whereas modularity is often denoted as a 'global' method, but these concepts remain ill-defined until now.

We will analyze the question of which community detection methods are resolution-free within the framework of the first principle Potts model as developed by Reichardt and Bornholdt [21]. Various methods can be derived from this first principle Potts model, among them modularity, and we will briefly examine them. We will suggest a very simple model, which we term the Constant Potts model (CPM), that resembles both the RB model and the RN model. It can be easily shown that the CPM is resolution-free in our definition, but it will follow immediately from the more general theorem we will prove. Arguably, the CPM is the simplest formulation of any (non-trivial) resolution-free method, and can be well interpreted.

In the next section, we will briefly examine this first principle Potts model, briefly review some models that can be derived from it, and introduce the CPM. We will then briefly explain the problem of the resolution-limit when using modularity. After having introduced the resolution-limit, we will provide the rigorous definition of resolution-free (i.e. not suffering from a resolution-limit), and show some general properties of these type of methods. We will then prove exactly which methods are resolution-free and which are not. Finally, we show the CPM method has superb performance.

Potts Model for Community Detection

First, let us introduce the notation. We consider a graph $G = (V, E)$ with $n = |V|$ nodes and $m = |E|$ edges. The adjacency matrix $A_{ij} = 1$ if there is an (ij) edge, and 0 otherwise. For weighted graphs the weight of a link is denoted by w_{ij}, while for an unweighted graph we can consider $w_{ij} = 1$. We denote the community of a node i by σ_i.

In principle, links within communities should be relatively frequent, while those between communities should be rela-

Reserved for Publication Footnotes

tively rare. Building on this idea, as done by Reichardt and Bornholdt [21], one should (i) reward internal links and external missing links; and (ii) penalize internal missing links and external links. If we consider internal (missing) links and external (missing) links to be of equal importance, we can disregard the distinction between internal and external, and focus only on the internal (missing) links. In general, this can then be written as

$$\mathcal{H} = -\sum_{ij}(a_{ij}A_{ij} - b_{ij}(1 - A_{ij}))\delta(\sigma_i, \sigma_j), \qquad [1]$$

where $\delta(\sigma_i, \sigma_j) = 1$ if $\sigma_i = \sigma_j$ and zero otherwise, and with some weights $a_{ij}, b_{ij} \geq 0$. Minimal $\mathcal{H}$ will then correspond to desirable partitions, although there is not necessarily one unique minimum. The choice of the weights a_{ij} and b_{ij} are important, and have a definite impact on what type of communities are detected. In the current literature, at least three different choices exist (and presumably some other methods may be rewritten as such), leading to three different methods for detecting communities:

Reichardt-Bornholdt (RB) $a_{ij} = w_{ij} - b_{ij}$ and $b_{ij} = \gamma_{RB}p_{ij}$ with p_{ij} the probability of a link between i and j, and γ_{RB} a resolution parameter [21].

Ronhovde-Nussinov (RN) $a_{ij} = w_{ij}$ and $b_{ij} = \gamma_{RN}$ with γ_{RN} a resolution parameter [20].

Label Propagation (LP) $a_{ij} = w_{ij}$ and $b_{ij} = 0$ without any resolution parameter, basically disregarding missing links [22].

We will briefly explicate these three different approaches.

Previous methods. Reichardt and Bornholdt introduce a new variable p_{ij} that represents the probability of a link between i and j. This is the so-called random null model used to compare to the actual network. Working out their choice of parameters, we arrive at

$$\mathcal{H}_{RB} = -\sum_{ij}(A_{ij}w_{ij} - \gamma_{RB}p_{ij})\delta(\sigma_i, \sigma_j). \qquad [2]$$

One of the most used null models is the so-called configuration model, which for unweighted (undirected) graphs is $p_{ij} = k_ik_j/2m$, where $k_i = \sum_j A_{ji}$ is the degree of node i. By using the configuration model $p_{ij} = k_ik_j/2m$ as the random null model, and setting $\gamma_{RB} = 1$ we recover the original definition of modularity [9] for unweighted graphs

$$\mathcal{Q} = \frac{1}{2m}\sum_{ij}\left(A_{ij} - \frac{k_ik_j}{2m}\right)\delta(\sigma_i, \sigma_j) \qquad [3]$$

such that $\mathcal{Q} = -\frac{1}{2m}\mathcal{H}_{RB}$. This comparison with a random null model introduces a problem with the so-called resolution limit, which prohibits finding small communities in relatively large graphs [15]. Although this effect is mitigated to some extent by the presence of the resolution parameter γ_{RB}, the problem remains inherent to the method [17]. We will come back to this issue later.

Ronhovde and Nussinov do not include such a random null model, in order to avoid issues with the resolution limit. In practice, they set $a_{ij} = w_{ij}$ and $b_{ij} = \gamma_{RN}$. Working this out we obtain

$$\mathcal{H}_{RN} = -\sum_{ij}(A_{ij}(w_{ij} + \gamma_{RN}) - \gamma_{RN})\delta(\sigma_i, \sigma_j). \qquad [4]$$

For unweighted graphs (i.e. where $w_{ij} = 1$) this reduces to

$$\mathcal{H}_{RN} = -(1 + \gamma_{RN})\sum_{ij}\left(A_{ij} - \frac{\gamma_{RN}}{1 + \gamma_{RN}}\right)\delta(\sigma_i, \sigma_j). \qquad [5]$$

Finally, the label propagation method [22] can be shown to be equivalent to the Potts model $-\sum_{ij}A_{ij}w_{ij}\delta(\sigma_i, \sigma_j)$ [23]. It can obviously be deduced from equation [1] by using the weights $a_{ij} = w_{ij}$ and $b_{ij} = 0$. This is the least interesting formulation, since there is only one global optimum, namely all nodes in a single community, which is trivial. However, the local minima could be of some interest.

It is not surprising then that these three different formulations share certain characteristics for some choice of parameters. The RB model is equivalent to the RN model up to a multiplicative constant by using an Erdös-Renyì (ER) null model, i.e. $p_{ij} = p$ and by setting $\gamma_{RN} = \gamma_{RB}p/(1 - \gamma_{RB}p)$. For $\gamma_{RN} = 0$ the RN model obviously reduces to the label propagation method.

Constant Potts model. Let us introduce an alternative method, that uses slightly different weights. By defining $a_{ij} = w_{ij} - b_{ij}$ and $b_{ij} = \gamma$, we obtain a version that is similar to both the RB and the RN model, but is simpler and more intuitive to work with. If we work this out, we obtain the rather simple expression

$$\mathcal{H} = -\sum_{ij}(A_{ij}w_{ij} - \gamma)\delta(\sigma_i, \sigma_j). \qquad [6]$$

Let us call this the Constant Potts model (CPM), with the 'constant' here referring to the comparison of A_{ij} to the constant term γ. It is clear that this is equivalent to the RN model for unweighted graphs by setting $\gamma = \frac{\gamma_{RN}}{1+\gamma_{RN}}$ and ignoring the multiplicative constant. Furthermore, it is equal to the RB model when setting $\gamma = \gamma_{RB}p$ for the ER null model. By setting $\gamma = 0$ we retrieve the label propagation method. Also, it is highly similar to an earlier Potts model suggested by Reichardt and Bornholdt [24].

Rephrasing it in terms of communities gives some intuition as to how it works. If we denote the number of edges[1] inside community c by $e_c = \sum_{ij}A_{ij}w_{ij}\delta(\sigma_i, c)\delta(\sigma_j, c)$, and by $n_c = \sum_i \delta(\sigma_i, c)$ the number of nodes in community c, we can rewrite equation [6] as

$$\mathcal{H} = -\sum_c e_c - \gamma n_c^2. \qquad [7]$$

In other words, the model tries to maximize the number of internal edges while at the same time keeping relatively small communities. The parameter γ balances these two imperatives. Moreover, for a fixed number of communities q, it favours a balanced partition, one where community sizes are as equal as possible. Hence, this suggests some connection to balanced graph partitioning, although this is not further investigated here.

In fact, the parameter γ acts as the inner and outer edge density limit. That is, suppose there is a community c with e_c edges and n_c nodes. Then it is better to split it into two communities r and s whenever

$$\frac{e_{r\leftrightarrow s}}{2n_rn_s} < \gamma, \qquad [8]$$

where $e_{r\leftrightarrow s}$ is the number of links between community r and s, which is exactly the density of links between community r and s. So, the link density between communities should be lower than γ, while the link density within communities should be

[1] Or technically, twice the number of edges in an undirected graph, or the total weight in a weighted graph.

higher than γ. This thus provides a clear interpretation of the γ parameter.

As stated earlier, for $\gamma = 0$ we retrieve the label propagation method, which has only one trivial minimum, namely all nodes in one big community, which can be easily understood when thinking of γ as the minimal intracommunity link density. In general, where $\gamma = \min_{ij} A_{ij}w_{ij}$ the optimal solution is the trivial solution of all nodes in one big community. On the other extreme, when $\gamma = \max_{ij} A_{ij}w_{ij}$, it is optimal to split all nodes in communities, i.e. such that each node forms a community in itself. In fact, communities of one node only form when $\lambda = \max_{ij} A_{ij}w_{ij}$, since otherwise it will always be beneficial to put the node in one of its neighbors' communities. Hence, for practical purposes $\min_{ij} A_{ij}w_{ij} \leq \gamma \leq \max_{ij} A_{ij}w_{ij}$.

Also, we can restrict ourselves to connected components, since disconnected components will never be together in one community. This can be easily seen since the total number of internal edges remains the same when splitting up a community that consists of two connected components, while the community sizes are decreased.

Resolution limit

The RB model, and by extension modularity, suffers from a so-called resolution limit [15]. That is, there is some dependence on global variables, which limits the size (and number) of communities that can be found. The number of communities roughly scales as $\sqrt{\gamma_{RB}m}$ with m the number of edges [17]. The RN model is claimed to be resolution-free, and should thus not show any of this unwanted dependence on global variables. The RB model is therefore considered a global optimization method (in the sense that there is some dependence on global variables, such as the number of edges m), while the RN method can be considered local (where such dependence is not present).

Traditionally the resolution limit is investigated by analyzing the counterintuitive merging of communities, for example cliques or some smaller communities that are only sparsely interconnected as displayed in Figure 1. The RB model with a configuration null model will merge two neighbouring cliques for example when [17]

$$\gamma_{RB} < \frac{q}{n_c(n_c-1)+2}, \qquad [9]$$

where q is the number of cliques and n_c is the number of nodes of a clique. So, the RB model depends on a global variable, namely the number of cliques q. This shows that modularity might be 'hiding' some smaller communities within larger communities, depending on how far you 'zoom in'. Indeed in [15] it was suggested for modularity to look at each community to consider whether it had any sub communities or not. Some related problem with modularity were noticed in [25] and more recently in [26].

The RN model on the other hand, will only join two cliques when [20]

$$\gamma_{RN} < \frac{1}{n_c^2-1}, \qquad [10]$$

which does not depend on the number of cliques q, and depends only on the 'local' variable n_c, so is argued not to suffer from any resolution limit. For the CPM suggested here, we arrive at the condition

$$\gamma < \frac{1}{n_c^2}, \qquad [11]$$

which also does not depend on the number of cliques q and can hence also said to be resolution-free. More general, CPM favors to cluster r consecutive cliques instead of $r-1$ at the point when

$$\gamma < \frac{2}{r(r-1)n_c^2}. \qquad [12]$$

Since it can be proven that cliques must always be clustered together, this defines the complete optimum over the whole range of γ, except for the clustering as individual nodes at $\gamma = 1$. A similar argument can be made using communities with heterogeneous sizes.

However, it remains somewhat unclear what is meant exactly by resolution-free in the above discussion, and the label *resolution-free* warrants a more precise definition. Consider for example that we take away the dependence on the number of links in the configuration null-model, so that we take $p_{ij} = k_i k_j$. Notice that this only corresponds to a multiplicative rescaling of γ_{RB} by $2m$. Although this no longer corresponds to a proper null-model (in the sense that $\sum_{ij} p_{ij} = m$), there's nothing preventing us from defining such weights, or taking a rescaled resolution parameter γ_{RB}. Reanalyzing the case above, we come to the conclusion that cliques are separate whenever

$$\gamma_{RB} > \frac{1}{2(n_c(n_c-1)+2)^2}, \qquad [13]$$

which unsurprisingly no longer depends on any global variables. By the argument employed in the previous section, the method should be resolution-free.

Not all problems have disappeared however. Consider that we take the subgraph consisting of only two of these cliques. Again, we can analyze when the method would merge or separate the two cliques in this subgraph. The two cliques will be joined whenever

$$\gamma_{RB} < \frac{1}{2(n_c(n_c-1)+1)^2}, \qquad [14]$$

still not depending on any global variables, so everything should be fine. Unfortunately, it is not. Combining the above two inequalities, we obtain that whenever

$$\frac{1}{2(n_c(n_c-1)+2)^2} < \gamma_{RB} < \frac{1}{2(n_c(n_c-1)+1)^2}, \qquad [15]$$

the method will separate the cliques in the larger graph, yet merge them in the subgraph.

So, again, the question remains: what does it mean to be *resolution-free*? Furthermore, what conditions should be imposed on the weights a_{ij} and b_{ij} in equation [1] for the method to be resolution-free? Would a method that takes into account the local number of triangles be resolution-free? Or would it be possible to use the shortest (weighted) path for example?

The above discussion motivates us to consider the following definition of a *resolution-free* method. The general idea is that when looking at any induced subgraph of the original graph, the partitioning results should not be changed. In order to introduce this definition, let $\mathcal{H}$ be any objective function (which we want to minimize), we then call a partition $\mathcal{C}$ for a graph $\mathcal{H}$-optimal whenever $\mathcal{H}(\mathcal{C}) \leq \mathcal{H}(\mathcal{C}')$ for any other partition $\mathcal{C}'$. We can then define *resolution-free* as follows.

Definition 1. *Let $\mathcal{C} = \{C_1, C_2, \ldots, C_q\}$ be a partition of graph G considered $\mathcal{H}$-optimal. Then the objective function $\mathcal{H}$ is called* resolution-free *if for each subgraph H induced by $\mathcal{D} \subset \mathcal{C}$, the partition $\mathcal{D}$ is also $\mathcal{H}$-optimal.*

Furthermore, some objective functions can be called additive.

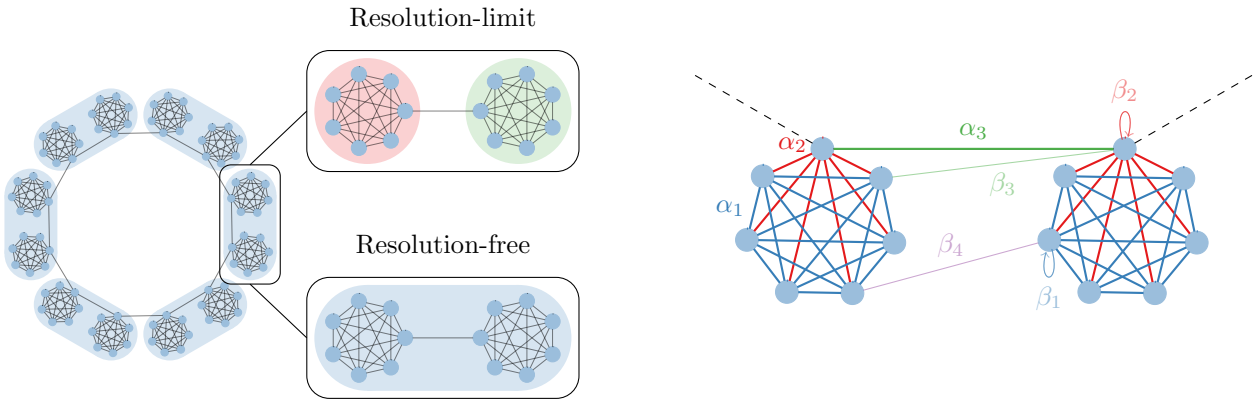

Fig. 1. The problem of the resolution-limit with modularity is usually demonstrated on a ring network of cliques. The cliques are as densely connected as possible, and as sparsely connected between them, while still retaining a connected graph. The resolution-limit is said to arise because it will merge the cliques depending on the size of the network. In fact, methods that do not suffer from the resolution limit, i.e. resolution-free methods, may merge these cliques also, but this will not depend on the size of the network. The distinguishing fact between resolution-limit and resolution-free methods is that the first will detect smaller subcommunities when applied to the subgraph, while the latter will not detect smaller subcommunities. Of course, whether the communities should consist of only cliques or of multiple joined cliques will still depend on the actual resolution of the method, which for CPM is designated by γ.

Definition 2. *An objective function $\mathcal{H}$ for a partition $\mathcal{C} = \{C_1, \ldots, C_q\}$ is called additive whenever $\mathcal{H}(\mathcal{C}) = \sum_i \mathcal{H}(C_i)$, where $\mathcal{H}(C_i)$ is the objective function defined on the subgraph H induced by C_i.*

If we have an optimal partition $\mathcal{C}$ for an additive resolution-free objective function $\mathcal{H}$, we can replace sub partitions of $\mathcal{C}$ by other optimal sub partitions.

Theorem 1. *Given an additive resolution-free objective function $\mathcal{H}$, let $\mathcal{C}$ be an $\mathcal{H}$-optimal partition of a graph G and let $H \subset G$ be the induced subgraph by $\mathcal{D} \subset \mathcal{C}$. If $\mathcal{D}'$ is an alternative optimal partition of H then $\mathcal{C}' = \mathcal{C} \setminus \mathcal{D} \cup \mathcal{D}'$ is also $\mathcal{H}$-optimal.*

Proof Define $\mathcal{C}'$ and $\mathcal{D}'$ as in the theorem. By additivity, $\mathcal{H}(\mathcal{C}') = \mathcal{H}(\mathcal{C} \setminus \mathcal{D}) + \mathcal{H}(\mathcal{D}')$, and by optimality $\mathcal{H}(\mathcal{D}') \leq H(\mathcal{D})$. Since also $\mathcal{H}(\mathcal{C}) = \mathcal{H}(\mathcal{C} \setminus \mathcal{D}) + \mathcal{H}(\mathcal{D})$ we obtain $\mathcal{H}(\mathcal{C}') \leq \mathcal{H}(\mathcal{C})$, so $\mathcal{C}'$ is also optimal. ■

Although, this might seem to contradict the NP-hardness community detection methods, this is not the case. It states that when there are two optimal partitions, any combination of those partitions are optimal, so in a certain sense, they are spanning a space of optimal partitions. It does not say whether such a partition can be easily found. Also, there might be two optimal partitions that cannot be obtained by recombining them, because all communities partly overlap with each other.

It is also possible to prove that a complete graph K_n with n nodes is never split (unless into all nodes separately).

Theorem 2. *Given a resolution-free objective function $\mathcal{H}$, the $\mathcal{H}$-optimal partition of K_n for all n is either only one community, namely all nodes, or n communities consisting each of one node.*

Proof Assume on the contrary there is an optimal partition $\mathcal{C}$ of K_n such that $1 < |\mathcal{C}| < n$. Then for any $\mathcal{D} \subset \mathcal{C}$ the subgraph H induced by $\mathcal{D}$ is a complete graph. But by assumption, $\mathcal{D}$ is then not optimal, and by resolution-free, $\mathcal{C}$ is then not optimal. Hence, inductively, the theorem must hold for all n. ■

This proves that a clique in a graph can never be split into two (or more communities), unless they are joined with other nodes. Otherwise, the partition of the subgraph would not be optimal, and by resolution-free the whole partition would not be optimal.

We can prove that CPM is *resolution-free* in this sense. By extension, the unweighted RN model and LP model are then also resolution-free. It is easy to see that the RB model using the configuration model (so also modularity) are not resolution-free in this sense, as also shown by the example in Figure 1. The CPM model is trivially shown to be additive by equation [**7**]. Perhaps it is less obvious, but the RB model (and modularity) is not additive, since it cannot be defined in terms of *indepent* contributions, i.e. the contribution $\mathcal{H}(C_i)$ per community depends on the whole graph G, instead of only on the subgraph H induced by C_i.

Since the CPM model is also related to the RB model using the ER null-model, it is tempting to conclude it is also resolution-free. Indeed, this might be said to be the case, if we choose p independently of the graph, i.e. not define it as $p = m/n(n-1)$, and simply choose it as some value $p \in \mathbb{R}$. However, we then obviously get back to the CPM model. This shows there is only a fine line between methods that are resolution-free, and those that are not.

However, we will not prove this for all methods separately. Rather, the results follow from the more general theorem we will now prove. In order to prove this more general statement, we first introduce the notion of 'local' weights. Again, building on the idea of subgraphs, we define local weights such that the weights do not change when looking to subgraphs.

Definition 3. *Let G be a graph, and let a_{ij} and b_{ij} as in equation* [**1**] *be the associated weights. Let H be a subgraph of G with associated weights a'_{ij} and b'_{ij}. Then the weights are called* local *if $a_{ij} = \lambda a'_{ij}$ and $b_{ij} = \lambda b'_{ij}$, where $\lambda = \lambda(H) > 0$ can depend on the subgraph H.*

Clearly then, the RN and CPM model have local weights, while the RB model does not. This definition says that local weights should be independent of the graph G in a certain sense. In fact, it is quite a strong requirement, as it should even hold for a single link (ij) in the subgraph where only i and j are included. That means it can't depend on any other link *but* the very link itself (excluding self-loops). Since for missing links, there is (usually) no associated weight or anything, it can only be constant. Hence, the RN model and the CPM model are one of the few sensible options available for having local variables.

Interestingly the definition of local weights does allow the dependence on some node properties. For example, consider a mobile phone network, where nodes are people, and an edge

between two people is present if they have communicated (with possibly a weight w_{ij} as the number of calls). Furthermore, suppose each node is associated to a certain geographical position. We could then take for the weights b_{ij} of missing links the geographical distance between the two nodes, since this can be defined independent of the graph, so also for the subgraph including only node i and j. Something resembling this setup has been considered in [27], where it was taken into account that the probability of a link depends highly on the distance between them [28]. Also, for multipartite graphs, the weights could depend on the class of node i and node j, thus allowing different parameters between different pairs of classes of nodes.

We can now prove the more general statement that methods using local weights are resolution-free.

Theorem 3. *The objective function $\mathcal{H}$ as defined in equation* **[1]** *is* resolution-free *if it has* local *weights.*

Proof Let $\mathcal{C}$ be the optimal partition for G with community assignments c_i, $\mathcal{D} \subset \mathcal{C}$ a subset of this partition, and H the subgraph induced by $\mathcal{D}$ with h nodes. Furthermore, we denote by d_i the community indices of $\mathcal{D}$, such that $d_i = c_i$ for $1 \leq i \leq h$ and by A' the adjacency matrix of H, so that $A_{ij} = A'_{ij}$ for $1 \leq i \leq h$. Assume $\mathcal{D}$ is not optimal for H, and that D^* is optimal, such that $\mathcal{H}(\mathcal{D}) > \mathcal{H}(D^*)$, or

$$-\sum_{ij}(A'_{ij}a'_{ij} - (1-A'_{ij})b'_{ij})\delta(d_i,d_j) > -\sum_{ij}(A'_{ij}a'_{ij} - (1-A'_{ij})b'_{ij})\delta(d^*_i,d^*_j).$$

Then define c^* by setting $c^*_i = d^*_i$ for $1 \leq i \leq h$ and $c^*_i = c_i$ for $h < i \leq n$. Then because the result is unchanged for the nodes $h < i \leq n$, we have that

$$\begin{aligned}
\Delta\mathcal{H} =& \mathcal{H}(\mathcal{C}) - \mathcal{H}(\mathcal{C}^*) \\
=& -\sum_{ij}(A_{ij} - (1-A_{ij})b_{ij})\delta(c_i,c_j) \\
&+\sum_{ij}(A_{ij} - (1-A_{ij})b_{ij})\delta(c^*_i,c^*_j) \\
=& -\sum_{ij}(A'_{ij}a_{ij} - (1-A'_{ij})b_{ij})\delta(d_i,d_j) \\
&+\sum_{ij}(A'_{ij}a_{ij} - (1-A'_{ij})b_{ij})\delta(d^*_i,d^*_j) \\
=& -\frac{1}{\lambda}\sum_{ij}(A'_{ij}a'_{ij} - (1-A'_{ij})b'_{ij})\delta(d_i,d_j) \\
&+\frac{1}{\lambda}\sum_{ij}(A'_{ij}a'_{ij} - (1-A'_{ij})b'_{ij})\delta(d^*_i,d^*_j) > 0
\end{aligned}$$

where the last step follows from the locality of the weights a_{ij} and b_{ij}. This inequality contradicts the optimality of $\mathcal{C}$. Hence, for all induced subgraphs H, the partition $\mathcal{D}$ is optimal, and the objective function $\mathcal{H}$ is *resolution-free.* ■

The converse is unfortunately not true. Consider a graph G with some weights a_{ij} and b_{ij}. Now pick your favorite graph H induced by some subpartition $\mathcal{D}$, and define the weights $a'_{ij} = a_{ij}$ and $b'_{ij} = b_{ij}$ except for one particular edge (kl), for which we set $a'_{kl} = a_{kl} + \epsilon$. Then for some $\epsilon > 0$, the original subpartition will remain optimal in H, while the weights are not local. Since the small change of the weight is *only* made when considering the graph H, all other subpartitions will always remain optimal. Of course, such a definition of the weight is rather odd, so in practice we will never use it.

However, although the converse is not true, we can say a bit more. Although the weights can be a bit different, there is not that much room for these differences. We demonstrate this on the ring network of cliques. The weights can depend only on the graph, so if G and G' are two isomorphic graphs, then $a_{ij}(G) = a_{i'j'}(G')$, where i and i' are two isomorphic nodes. This obvioulsy also applies to automorphic graphs, and thus limits the possible choices of weights. More specifically, it limits the possible choices of the weights of the ring network considerably. The weights that can be different (regardless of whether they are local or not) are illustrated in Figure 1, where the α's refer to the weights of present links, and the β's to the weight of missing links. For the convenience of the reader, we color coded (online only) the equations used in the following section, so that they match the illustration in Figure 1.

All nodes within a clique are isomorphic, except the node that connects to other cliques. So, all the edges among those $n_c - 1$ nodes are similar, and will have the same weight α_1. All edges from these $n_c - 1$ nodes to the 'outside' node will have the same weight α_2. Finally, the edge connecting two cliques is denoted by α_3. There are only relatively little links missing within a clique, namely the self-loops. The self-loop for the special 'outside' node is denote by β_2 while the self-loop for the other nodes in the cliques is denoted by β_1. Finally, there are two type of missing links between two cliques: (1) a link between the 'outside' node and a normal node denoted by β_3; and (2) a link between two normal nodes, denoted by β_4. These weights are made clear in Figure 1.

Let us now analyze when the method will not be resolution-free. That is, the cliques must be merged in some (large) graph, while for the subgraph consisting of these two merged cliques, they should be separated by the method. Or the other way around, they should be separated in some (large) graph, but merged in the subgraph.

We can write the $\mathcal{H}_s$ for all q cliques being separate as

$$\mathcal{H}_s = -q(\alpha_1(n_c-1)(n_c-2) + 2\alpha_2(n_c-1) - (n_c-1)\beta_1 - \beta_2)$$

and $\mathcal{H}_m$ for merging all two consecutive cliques as

$$\mathcal{H}_m = -\frac{q}{2}2(\alpha_1(n_c-1)(n_c-2) + 2\alpha_2(n_c-1) - (n_c-1)\beta_1 - \beta_2 + \alpha_3 - \beta_3(n_c-1) - \beta_4(n_c-1)^2)$$

Furthermore, for the induced subgraph H consisting of two consecutive cliques, we can write $\mathcal{H}'_s$ for separating the two cliques and $\mathcal{H}'_m$ for merging them, similarly as before, where α' and β' are the weights for the subgraph H. Then the method is not resolution-free if it would merge the two cliques at a higher level (i.e. when $\mathcal{H}_m < \mathcal{H}_s$) yet wouldn't merge them at smaller scale (i.e. when $\mathcal{H}'_s < \mathcal{H}'_m$), or if they would be separate at a higher level (i.e. when $\mathcal{H}_m > \mathcal{H}_s$), yet merged at a smaller scale (i.e. when $\mathcal{H}'_s > \mathcal{H}'_m$) Working out this condition for $\mathcal{H}_m < \mathcal{H}_s$ (and similarly for $\mathcal{H}_m > \mathcal{H}_s$) gives us

$$\alpha_3 > (n_c-1)(\beta_4(n_c-1) + \beta_3),$$

while for $\mathcal{H}'_s < \mathcal{H}'_m$ (and similarly for $\mathcal{H}'_s > \mathcal{H}'_m$) we obtain

$$\alpha'_3 < (n_c-1)(\beta'_4(n_c-1) + \beta'_3).$$

Combining these two inequalities for both cases we obtain

$$\alpha'_3(\beta_4(n_c-1) + \beta_3) < \alpha_3(\beta'_4(n_c-1) + \beta'_3), \quad \textbf{[16]}$$

$$\alpha'_3(\beta_4(n_c-1) + \beta_3) > \alpha_3(\beta'_4(n_c-1) + \beta'_3). \quad \textbf{[17]}$$

where either equation **[16]** or **[17]** should hold. Hence, only if the left hand side equals the right hand side, it does not constitute a counter example. Working out this equality, there

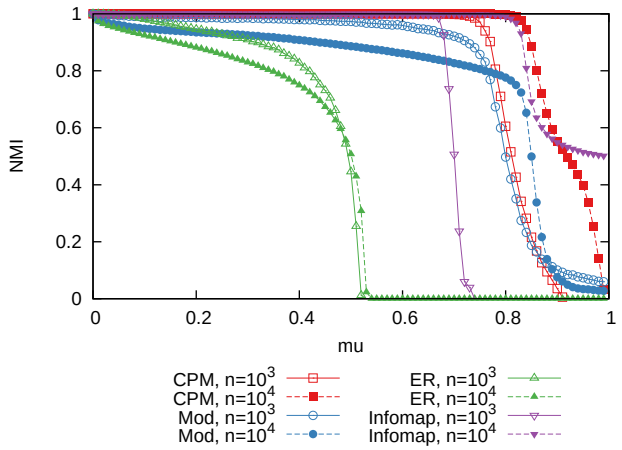

Fig. 2. Performance of different community detection methods in terms of Normalized Mutual Information (NMI) depending on mixing parameter μ, showing CPM to perform superbly. The open symbols denote results for $n = 10^3$ and the closed symbols for $n = 10^4$. Both had a degree distribution exponent of 2 (with average degree 15 and maximum degree 50) and community size distribution exponent 1 (with community sizes ranging from 20 to 100). Per value of μ 100 graphs were used to obtain this result. The resolution parameter γ used to obtain this result for CPM was analytically calculated.

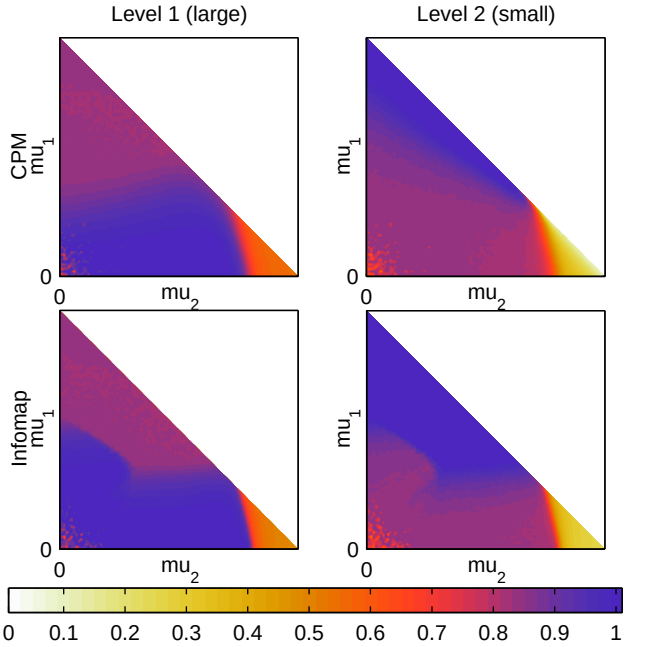

Fig. 3. Performance of CPM and Infomap on a hierarchical benchmark network in terms of Normalized Mutual Information (NMI) depending on mixing parameters μ_1 and μ_2. The networks had $n = 10^4$ nodes with a degree distribution exponent of 2 (with average degree 20 and maximum degree 50) and community size distribution exponent 1 for both small (size ranging from 10 to 50) and large communities (size ranging from 50 to 300). Per combination of parameters 10 graphs were used to obtain this result. The resolution parameters γ for the two different levels were calculated analytically.

are two possibilities. Either the weights should be local, so that

$$\frac{\alpha_3}{\alpha'_3} = \frac{\beta_3}{\beta'_3} = \frac{\beta_4}{\beta'_4} = \lambda, \qquad [18]$$

or the following equality should hold.

$$n_c - 1 = \frac{\alpha_3\beta'_3 - \alpha'_3\beta_3}{\alpha'_3\beta_4 - \alpha_3\beta'_4} \qquad [19]$$

Obviously, this again constitutes some very particular case of non-local weights. We can repeat this same procedure for other subpartition, and for other graphs, thereby forcing the weights to be of a very particular kind. This thus leaves little room for having any sensible definition of local weights.

Furthermore, it is possible to prove that the property of being additive is equivalent to having local weights if $\lambda = 1$. Hence, whenever the general first principle Potts model has local weights (with $\lambda = 1$), it is additive and resolution-free. Since there are only relatively few sensible choices available for local weights, all resolution-free methods within this framework can be easily described. That means resolution-free community detection has only a quite limited scope. In fact, the CPM seems to be the simplest non-trivial sensible formulation of any general resolution-free method, although there's some leeway for special graphs (i.e. having some node properties, such as multipartite graphs). This is not to say that methods with non-local weights (e.g. modularity, number of triangles, shortest path, betweennness) should never be used for community detection at all, they are just never resolution-free.

Performance and Application

In order to asses the performance of the proposed CPM model, we performed various tests. Using the latest suggested test networks [13] we find that the CPM model and the accompanying algorithm is both highly accurate and very efficient. More details on the efficient Louvain-like algorithm and the test procedure can be found in the Material and Methods section at the end of this article.

The benchmark networks are created by using a known community structure, i.e. a planted community structure, where both the degree distribution and the community size distribution follow some power law. We have examined both directed test networks as well as hierarchical test networks, where communities at multiple 'levels' in the data exist. The difficulty of detecting communities correctly depends on the parameter μ of having links outside its community, while for hierarchical communities, there are two such parameter μ_1 for the first level (the large communities), and μ_2 for the second level (the subcommunities).

Some of the earlier algorithms and models that showed excellent performance are the Louvain [7] method for optimizing modularity, and the Infomap method [29], of which recently an hierarchical version was proposed [30]. In Figure 2 we have displayed the results for the Louvain method (both using the configuration null model, as well as the ER null model, with default resolution parameter $\gamma_{RB} = 1$), the Infomap method, and the CPM method for test networks having $n = 10^3$ and $n = 10^4$ nodes. It can be clearly seen that CPM performs outstanding.

Most methods seem to perform well until a certain threshold, after which performance drops suddenly, suggesting either a phase transition in the community structure of the test graphs, or in the method being used. The community structure of such a test network can be said to be present, as long as the density within a community is higher than the density between communities. This is actually the case up until $\mu < \frac{n-\langle n_s\rangle}{n-1}$ where $\langle n_s\rangle$ is the expected community size, so that μ is close to 1 for both $n = 10^3$ and $n = 10^4$. Hence, the sudden drop signals the inability of the algorithm to correctly identify the communities, since the actual phase transition of the existence of communities should take place only later. For

the hierarchical version similar thresholds for the existence of communities can be provided, and in our case, this is also close to the possible maximum.

The difference in performance of the CPM model in comparison to the RB model using the ER null model is especially striking. This is no result of the method being resolution-free or not, but rather depends highly on choosing the correct resolution parameter. Obviously then, setting $\gamma = p$ is in general not a very good strategy, and for general networks one should carefully analyze at which resolution the network contains significant partitions, for example looking at stable partitions using a randomized algorithm [19, 20].

We have also performed extensive tests on hierarchical networks, where the method also performs excellent, and is able to extract the two different levels of communities effectively, as displayed in Figure 3. With increasing μ_1 is becomes more difficult to recuperate the large communities of level 1, just as it becomes increasingly difficult to detect correctly the smaller communities at level 2 with increasing μ_2. However, when μ_1 increases it becomes easier to detect smaller communities, because the average outer density for smaller communities then becomes lower (otherwise, the communities remain less clearly distinguished within one large community). On the other hand, when μ_2 is sufficiently high, the inner density of large communities and small communities becomes essentially the same, so that it becomes more difficult to detect them. Summarizing, for relatively low $\mu_2 \lessapprox 0.7$, the first (larger) level becomes more clear for low μ_1, while the second (smaller) level becomes more clear for larger μ_1. This is both the case for the Infomap method and the CPM method. The Infomap seems to be slightly better at detecting the correct communities, although the CPM method is highly competitive.

Conclusion

Several community detection methods, among which modularity, are affected by the problem of the resolution limit. In this paper we have provided a novel rigorous definition of what it means for a community detection method to be resolution (limit) free. A number of interesting (and favourable) properties of resolution-free methods are provided. Most importantly, we are able to prove exactly which community detection methods are resolution free, namely those methods that use local weights. This also clarifies the relationship between 'local' methods and the resolution limit.

Moreover, there does not seem to be much room for having resolution-free methods without local weights. Of the few possibilities available for having resolution-free community detection, the Constant Potts Model (CPM) we introduced in this paper seems to be the simplest possible formulation. We provided an intuitive interpretation of its resolution parameter in terms of inner and outer density of communities. After intensive testing, this method and its accompanying algorithm is shown to perform superbly, including being able to detect different levels in hierarchical graphs. The algorithm runs very efficiently, and can easily handle networks with several million number of edges and nodes. A rigorous definition of resolution-free community detection allows for a more articulate analysis, and induces further progress on developing novel and meaningful methods.

Materials and Methods

Louvain like Algorithm.

The algorithm we employ derives from the Louvain method [7], which also resembles the algorithm suggested by [20], and performed quite well in the context of modularity optimization [31]. We use the concept of node size, denoted by n_i for a node i, initialized to $n_i = 1$ (indeed the community size $n_c = \sum_i n_i \delta(\sigma_i, c)$ is related). We first iterate (randomly) over all nodes, and put nodes greedily into the community that minimizes equation **[6]** most. We subsequently create a new graph based on the communities, and new node sizes, and reiterate over this new smaller graph. More specifically:

1. Initialize A and set $n_i = 1$ for all nodes i.
2. Loop over nodes i (possibly randomly), remove it from its community and calculate for each community c the increase if we would put node i into community c,
$$\Delta\mathcal{H}(\sigma_i = c) = -(e_{i\leftrightarrow c} - 2\gamma n_i \sum_j n_j \delta(\sigma_j, c)), \qquad \textbf{[20]}$$
where $e_{i\leftrightarrow c} = \sum_j (A_{ij} + A_{ji})\delta(\sigma_j, c)$ is the number of edges between node i and community c. We put node i into the community c for which $\Delta\mathcal{H}(\sigma_i = c)$ is minimal. We iterate until we can no longer decrease the objective function.
3. We build a new graph $A'_{cd} = \sum_{ij} A_{ij}\delta(\sigma_i, c)\delta(\sigma_j, d)$ and node sizes $n'_c = \sum_i n_i \delta(\sigma_i, c)$. We repeat step 2 by setting $A = A'$ and $n = n'$ until the objective function can no longer be decreased.

As stated earlier n'_c is simply the number of nodes in community c, i.e. the size of community c, and that A'_{cd} is the number of links between community c and d. Merging nodes c and d in A' gives a change of

$$\Delta\mathcal{H}_{cd} = -(A'_{cd} - 2\gamma n_c n_d) \qquad \textbf{[21]}$$

which corresponds exactly with merging communities c and d in A. If there are weights present, the algorithm can be applied by using $A_{ij} = w_{ij}$ if there is an (ij) link present, and 0 otherwise.

The looping over the nodes in step 2 can be done in $O(m)$ time for m edges. Since A' is a $q \times q$ matrix, with q the number of (non-empty) communities, this reduces the problem size significantly, although the matrix A' is likely to be more dense then A, so the looping over the nodes in step 2 is then expected to be in the order of $O(q^2)$. Typically, only a few iterations over the nodes, and a few levels are required to optimize a graph. The total running time is therefore expected to be about $O(m)$.

Finally, notice that for resolution-free methods, the results should be unchanged on subgraphs. Hence, we could therefore perform the method (recursively) on subgraphs. We suggest then the following improvement. First cut the network at each recursive call, until the density of the subgraph exceeds γ. Then, we recombine the subgraphs, and loop over nodes/communities to find improvements until we can no longer increase greedily, and return to the previous recursive function call. These calls should be easily parallelized, making community detection in even larger graphs or in an on-line setting possible by using cluster computing.

Benchmark tests.

In order to asses the performance of the proposed CPM model and the suggested Louvain like algorithm, we performed various tests. Using the latest suggested benchmark networks [13] we find that the CPM model and the accompanying algorithm is both highly accurate and very efficient.

The benchmark networks are created by using a known community structure, i.e. a planted community structure. The community sizes n_s are chosen from a distribution following a power-law $\Pr(n_s = n) \sim n^{-\tau_2}$. The degrees k_i of the nodes are also chosen from a power-law distribution $\Pr(k_i = k) \sim k^{-\tau_1}$. The stubs are then connected, with probability $1 - \mu$ within a community, and with probability μ between two communities. Since the network is finite, the degree and community sizes are automatically constrained by the number of nodes n. Also, it is rather uninteresting to have communities of size 1. So, commonly a lower bound $\underline{n_s}$ and upper bound $\overline{n_s}$ on the community sizes is imposed, while for the degree the average degree $\langle k \rangle$ is specified. For the hierarchical version, there are two levels, with the communities of the second level embedded in the first level. In that version a fraction of μ_1 of the links is placed between two different macro communities

at the first level, while a fraction of μ_2 of the links are placed between the small communities of the second level (but within the same macro-community).

Instead of detecting the resolution algorithmically, we can actually calculate the proper resolution parameter value γ analytically (and therefore, beforehand). We compare our results to the standard modularity detection (i.e. using the configuration null model with resolution parameter $\gamma_{RB} = 1$) and to the ER configuration null model (with $\gamma_{RB} = 1$). Here, it becomes clear that the ER configuration null model doesn't perform well with it's standard resolution parameter, while modularity performs somewhat better. However, when choosing the correct resolution parameter, as we do, the accuracy is exceptional.

In order to calculate the correct resolution parameter, we consider the following. The resolution parameter γ acts as a sort of threshold on community inner and outer density. If we were to set γ equal to the inner density, it would be rather difficult to fulfill the condition that the inner density should be higher than that, and similarly so for γ equal to the outer density. So, we need to be as far as possible from both the inner density as well as the outer density, which would be simply the average of the two.

The inner density for a community having n_s nodes can be easily found as

$$p_{in} = \frac{(1-\mu)\langle k\rangle}{n_s - 1}, \qquad [22]$$

and the outer density (i.e. all the edges originating from a community to the outside) is

$$p_{out} = \frac{\mu\langle k\rangle}{n - n_s}, \qquad [23]$$

where n is simply the total number of nodes. The average community size $\langle n_s\rangle$ than gives us the average $\langle p_{in}\rangle$ and $\langle p_{out}\rangle$, and is proportional to

$$\langle n_s\rangle \sim \sum_{n=\underline{n_s}}^{\overline{n_s}} n n^{-\beta}, \qquad [24]$$

where $\underline{n_s}$ is the minimal community size and $\overline{n_s}$ the maximal community size. We then simply set

$$\gamma = \frac{1}{2}(\langle p_{in}\rangle + \langle p_{out}\rangle). \qquad [25]$$

For the hierarchical test networks, we can perform a similar analysis, and arrive at

$$p_1^{in} = (1-\mu_1)\langle k\rangle/(n_{s,1} - 1) \qquad [26]$$

$$p_1^{out} = \mu_1\langle k\rangle/(n - n_{s,1}) \qquad [27]$$

$$p_2^{in} = (1-\mu_1-\mu_2)\langle k\rangle/(n_{s,2} - 1) \qquad [28]$$

$$p_2^{out} = (\mu_1+\mu_2)\langle k\rangle/(n - n_{s,2}) \qquad [29]$$

and use the average of the in and outer density, similar as before, for the two different levels. Ordinarily, the communities are assumed to exist whenever $p_{in} > p_{out}$.

For comparing our results to the 'known' community structure, we use the normalized mutual information. Given two different partitions C and D, the mutual information I is defined as

$$I(C,D) = -\sum_{r,s}\frac{n_{r,s}}{n}\log n\frac{n_{r,s}}{n_r n_s} \qquad [30]$$

with $n_{r,s}$ being the number of nodes that are in community r in partition C and in community s in partition D, while n_r simply denotes the number of nodes in community r. The normalized mutual information $I_n(C,D)$ is then defined as

$$I_n(C,D) = \frac{2I(C,D)}{H(C)+H(D)}, \qquad [31]$$

where $H(C)$ indicates the entropy of a partition C, which is defined as

$$H(C) = -\sum_s \frac{n_s}{n}\log\frac{n_s}{n}. \qquad [32]$$

The normalized mutual information $0 \leq I_n(C,D) \leq 1$, with 1 indicating equivalent partitions.

ACKNOWLEDGMENTS. We acknowledge support from a grant "Actions de recherche concertes — Large Graphs and Networks" of the "Communauté Française de Belgique" and from the Belgian Network DYSCO (Dynamical Systems, Control, and Optimization), funded by the Interuniversity Attraction Poles Programme, initiated by the Belgian State, Science Policy Office. The authors would like to thank Arnaud Browet, Jean-Charles Delvenne, Renaud Lambiotte, Gautier Krings and Julien Hendrickx for helpful comments and discussion.

1. David Lazer, AS Pentland, L Adamic, S Aral, AL Barabasi, D Brewer, N Christakis, N Contractor, J Fowler, and M Gutmann. Life in the network: the coming age of computational social science. Science (New York, NY), 323(5915):721, 2009.
2. Duncan J Watts. A twenty-first century science. Nature, 445(7127):489, Jan 2007.
3. M E J Newman. The structure and function of complex networks. SIAM Review, 45:167–256, 2003.
4. Santo Fortunato. Community detection in graphs. Physics Reports, 2010.
5. Mason A Porter, Jukka-Pekka Onnela, and PJ Mucha. Communities in networks. Notices of the American Mathematical Society, 56(9):1082–1097, 2009.
6. Roger Guimerà and Luis A Nunes Amaral. Functional cartography of complex metabolic networks. Nature, 433(7028):895–900, Feb 2005.
7. Vincent D Blondel, Jean-Loup Guillaume, Renaud Lambiotte, and Etienne Lefebvre. Fast unfolding of communities in large networks. Journal of Statistical Mechanics: Theory and Experiment, 2008(10):P10008+, Oct 2008.
8. Roger Guimerà, S Mossa, A Turtschi, and L. A. N Amaral. The worldwide air transportation network: Anomalous centrality, community structure, and cities' global roles. Proc Nat Acad Sci USA, 102(22):7794–7799, May 2005.
9. M E J Newman and Michelle Girvan. Finding and evaluating community structure in networks. Phys Rev E, 69(2):026113+, Feb 2004.
10. M Barber. Modularity and community detection in bipartite networks. Phys Rev E, Jan 2007.
11. VA Traag and J Bruggeman. Community detection in networks with positive and negative links. Phys Rev E, 80(3):36115, 2009.
12. PJ Mucha, T Richardson, K Macon, Mason A Porter, and Jukka-Pekka Onnela. Community structure in time-dependent, multiscale, and multiplex networks. Science, 328(5980):876, 2010.
13. A Lancichinetti, Santo Fortunato, and Filippo Radicchi. New benchmark in community detection. Phys Rev E, 78:046110, 2008.
14. Andrea Lancichinetti, Santo Fortunato, and Filippo Radicchi. Benchmark graphs for testing community detection algorithms. Physical Review E, 78(4):046110, Oct 2008.
15. Santo Fortunato and Marc Barthélemy. Resolution limit in community detection. Proc Nat Acad Sci USA, 104:36–41, 2007.
16. Benjamin H Good, Yves-Alexandre De Montjoye, and Aaron Clauset. Performance of modularity maximization in practical contexts. Physical Review E, 81(4):046106, Apr 2010.
17. J M Kumpala, Jari Saramäki, K Kaski, and János Kertész. Limited resolution in complex network community detection with potts model approach. Eur Phys J B, 56:41–45, 2007.
18. Renaud Lambiotte, Jean-Charles Delvenne, and Mauricio Barahona. Laplacian dynamics and multiscale modular structure in networks. Arxiv preprint arXiv:0812.1770v3, Jan 2008.
19. Renaud Lambiotte. Multi-scale modularity in complex networks. Arxiv preprint arXiv:1004.4268, Jan 2010.
20. P Ronhovde and Z Nussinov. Local resolution-limit-free potts model for community detection. Phys Rev E, Jan 2010.
21. Jörg Reichardt and Stefan Bornholdt. Partitioning and modularity of graphs with arbitrary degree distribution. Phys Rev E, 76(1):015102+, 2007.
22. Usha Raghavan, Réka Albert, and Soundar Kumara. Near linear time algorithm to detect community structures in large-scale networks. Physical Review E, 76(3):036106, Sep 2007.
23. G Tibely and J Kertesz. On the equivalence of the label propagation method of community detection and a potts model approach. Phys A, 387(19-20):4982–4984, Aug 2008.
24. Jörg Reichardt and S Bornholdt. Detecting fuzzy community structures in complex networks with a potts model. Physical Review Letters, 93(21):218701, 2004.
25. Ulrik Brandes, D Delling, M Gaertler, R Görke, M Hoefer, Z Nikoloski, and D Wagner. On finding graph clusterings with maximum modularity. Lecture Notes in Computer Science, 2769:121–132, 2007.
26. Gautier Krings and Vincent D Blondel. An upper bound on community size in scalable community detection. Arxiv preprint arXiv:1103.5569, Jan 2011.
27. P Expert, T Evans, Vincent D Blondel, and Renaud Lambiotte. Beyond space for spatial networks. Arxiv preprint arXiv:1012.3409, 2010.

28. Renaud Lambiotte, Vincent D Blondel, Cristobald de Kerchove, E Huens, C Prieur, Z Smoreda, and P Van Dooren. Geographical dispersal of mobile communication networks. Phys A, 387(21):5317–5325, 2008.
29. Martin Rosvall and Carl T Bergstrom. An information-theoretic framework for resolving community structure in complex networks. Proc Nat Acad Sci USA, 104(18):7327–7331, May 2007.
30. Martin Rosvall and Carl T Bergstrom. Multilevel compression of random walks on networks reveals hierarchical organization in large integrated systems. arXiv, 2010.
31. Andrea Lancichinetti and Santo Fortunato. Community detection algorithms: A comparative analysis. Physical Review E, 80(5):056117, Nov 2009.