# Incompatibility boundaries for properties of community partitions

Arnaud Browet, Julien Hendrickx, Alain Sarlette

**Abstract**

We prove the incompatibility of certain desirable properties of community partition quality functions. Our results generalize the impossibility result of [Kleinberg 2003] by considering sets of weaker properties. In particular, we use an alternative notion to solve the central issue of the consistency property. (The latter means that modifying the graph in a way consistent with a partition should not have counterintuitive effects). Our results clearly show that community partition methods should not be expected to perfectly satisfy all ideally desired properties.

We then proceed to show that this incompatibility no longer holds when slightly relaxed versions of the properties are considered, and we provide in fact examples of simple quality functions satisfying these relaxed properties. An experimental study of these quality functions shows a behavior comparable to established methods in some situations, but more debatable results in others. This suggests that defining a notion of good partition in communities probably requires imposing additional properties.

**Index Terms**

Network theory (graphs), Partitioning algorithms, Network topology, Clustering algorithms.

◆

## 1 INTRODUCTION

In the rich literature about defining communities in graphs, two major and complementary approaches are the proposal of particular criteria defining communities and the identification of axioms that should reasonably be satisfied by such criteria, in particular by value functions that the optimal partition should maximize. When confronting the two approaches, the popular clustering criteria all fail to satisfy at least one of the reasonable axioms. For instance, the popular modularity criterion [14] is neither local (i.e. the optimal partitioning of a subset of nodes does depend on the whole graph) nor consistently improving (i.e. an optimal community partitioning for a given graph can loose its optimality when strengthening intra-community links and weakening inter-community links).

In fact Kleinberg [10] has proved that it is impossible for a function $F$ associating a community partition to any weighted graph[1] to simultaneously satisfy the three following properties, although getting two of them separately is easy:

- Richness: given any partition of the nodes $\Sigma$, there exists at least one graph for which $F$ returns $\Sigma$ as the unique[2] associated partition;
- Scale-invariance: given any graph, multiplying all the edge weight by a constant does not change the partition returned by $F$;
- Consistency: if $F$ returns a partition $\Sigma$ for a graph $G$, then it also returns $\Sigma$ for any graph $G'$ obtained by increasing the weight of intra-communities edges and decreasing the weight of inter-communities edges with respect to $\Sigma$.

The last property means that modifying a graph in a way "consistent" with its associated community partition, by increasing links inside communities and decreasing links between communities, should lead to a graph with the same community partition. Despite its natural formulation, this property is actually debatable [22], and we will see in Section 2 that it has counterintuitive consequences.

Other recent approaches have run into similar barriers, e.g. the hierarchy-based axioms in [5] would not admit scale invariance; the various subgraph-based alternatives in [15] and [9] do not return a single partition per graph, and in fact for any graph they return among others the single community encompassing the entire node set, hence failing richness.

- *A. Browet and J.M. Hendrickx are with the Université catholique de Louvain, ICTEAM institute.*

- *A.Sarlette is with INRIA Paris (QUANTIC lab) & Ghent University (Data Science Lab)*

1. This was in a framework where nodes are separated by a "distance", as opposed to nodes being connected by edges with a certain weight as considered here. The two frameworks are however equivalent provided all weights are positive and self-loops can be discarded; it suffices then indeed to take the inverse of the weights as distances.

2. Uniqueness is necessary to avoid trivial multi-valued functions associating for example all partitions to all graphs.

More examples will be covered later.

In this paper, we begin by presenting two generalizations of the result of [10] in a more flexible framework where a value function evaluates the quality of each community partition for a graph (as opposed to considering a function that directly associates a community partition to each graph). The "best" partition associated to a graph can then be obtained by optimizing this value function over all possible partitions, but "sufficiently good" partitions can also be computed when an exact solution to this generally hard optimization problem is out of reach. This approach is followed in [22], and is consistent with many well-known community partition methods, as those listed in Table 2 in Section 2.

Our first main result (Section 3) shows the impossibility of satisfying richness, scale-invariance and (together with some continuity assumption about the value function) *a much weaker form of consistency than in [10], which we argue appears more natural*. In our second impossibility result, the consistency requirement is removed altogether and replaced by some locality axiom, forbidding modifications in one part of a graph to affect the (relative) quality of communities in another part of the graph.

We further add some insight to these impossibility results. The problem can be pinpointed to the request, for richness, to make both the unique "all-encompassing" partition optimal for some graph, and the "all-singletons" partition optimal for some other graph. If we drop the all-singletons partition from the richness requirement, then value functions satisfying all the axioms can be constructed, as we show by providing actual examples (Section 4). It turns out that in previous papers which try to circumvent Kleinberg's impossibility result, this distinction is made implicitly. For instance in [1], the authors propose to circumvent Kleinberg's impossibility by imposing similar axioms on a value function for partitions instead of directly on the "clustering function" $F$. Our variant of the impossibility theorem precisely shows that this does not resolve the impossibility. However, the value function proposed in [1] is undefined for the "all-singletons" partition, and we identify that this is the key relaxation. In [2], the authors discuss and compare alternatives by incidentally restricting themselves to $k$-clustering, and [23] considers the special case of bi-partitions. In the light of our result, fixing $k$ is just a way to satisfy the sufficient relaxation, i.e. excluding either the all-singletons or the all-encompassing partition (and many others). In all those cases, this simple implicit relaxation of excluding the all-singleton partition (and possibly others) is in fact sufficient and the key step to allow satisfying Kleinberg's set of axioms. In this sense, the present paper establishes a clear and possibly practically relevant way to circumvent the impossibility. Another point is that Kleinberg's impossibility refers to graphs without self-loops. Some previous investigations, like e.g. [22], let self-loops play an instrumental role. In that case there are more options for constructing optimal graphs towards richness. We can in fact state that our example value functions, proposed in Section 4, satisfy all the axioms of Kleinberg if we allow for self-loops.

We stress that this paper implies no value judgment about clustering criteria which fail to satisfy some of the proposed axioms (including the assumption about no self-loops). To the contrary, the impossibility results clarify that it is hopeless to look for value function-based criteria satisfying all the axioms. It is hence perfectly unavoidable to select a subset of axioms, and one reasonable selection criterion is to get the most useful results. One defensible viewpoint is to exclude some partitions as a priori irrelevant, e.g. partitions containing almost only isolated nodes. Along this line, we analyze, in Section 5, the partitions obtained using one of the proposed value functions which satisfies all the axioms except strict richness for partitions involving singletons. Although our proposed value functions turn out to have strong similarities with the sum of community strengths and especially with the modular density, introduced respectively in [13], [12], they satisfy stronger axiomatic properties. Moreover, we use a more general way of normalizing the contribution of each community, allowing for some additional flexibility. The conclusion of our experimental investigation is that only particular tuning of our value function parameters leads to results that are compatible with expectations on benchmark problems. Hence, satisfying the "standard" axioms considered here is not a guarantee for more relevant results.

Towards the future, our results hence not only clarify that the set of historically proposed axioms cannot be kept in its most general form; they also highlight the need to add *compatible* axioms that would isolate a most useful set of value functions, since our own examples satisfy a minimally relaxed set of axioms, yet they still leave a design freedom among which far from all choices behave as intuitively desired.

## 2 PROPERTIES OF QUALITY FUNCTIONS

As mentioned in the introduction, there are different ways of specifying what a good partition in communities is [7], [19]. One can for example directly specify the properties that the (best) partition should have, or the algorithm to obtain it, as in the framework of [14]. We follow here the popular and more flexible option [22] of defining a quality measure of a given partition into communities for any given weighted graph. The best partition into communities is then the one maximizing the quality function. This optimization problem is often computationally challenging [3], but the use of a quality function defined for all partitions allows using heuristic methods to compute relatively good partitions.

More formally, we consider weighted symmetric graphs $G(V, W)$ without self-loops, where $V = \{1, \ldots, N\}$ is the set of nodes, $N$ being the number of nodes, and $W$ the set of weights $W_{ij} = W_{ji} \geq 0$ (with $W_{ii} = 0$). When $W_{ij} > 0$, we say that $i$ and $j$ are connected or that $(i, j) = (j, i)$ is an edge, and it will sometimes be convenient to refer to the set $E$ of

edges, which is entirely determined by $W$. A partition $\sigma$ of $V$ into communities corresponds to an assignment of each node $i \in V$ to a community label $\sigma_i \in \{1, \ldots, n_c\}$, where $n_c$ depends on the partition $\sigma$. The partition induces communities $c_1, \ldots, c_{n_c}$ defined by $c_k = \{i | \sigma_i = k\}$ and whose cardinality, i.e. the number of nodes within the community, is denoted $|c_k|$. We sometimes use the Kronecker delta $\delta(\sigma_i, \sigma_j) = 1$ if $\sigma_i = \sigma_j$, and 0 otherwise, to express whether $i$ and $j$ belong to the same community. We denote by $s_i^{int} = \sum_{j \in V} W_{ij} \, \delta(\sigma_i, \sigma_j)$ (resp. $s_i^{ext} = \sum_{j \in V} W_{ij} \, (1 - \delta(\sigma_i, \sigma_j))$) the internal (resp. external) strength of node $i$, i.e. the sum of edge weight connecting node $i$ to nodes in the same (resp. in any other) community, $s_i = \sum_{j \in V} W_{ij} = s_i^{ext} + s_i^{int}$ the total strength of node $i$, and $m = \sum_{i \in V} s_i$ which equals twice the total weight of the graph.

We then consider a *value (or quality) function* $f(G, \sigma)$, that represents the quality of the community partition $\sigma$ on the graph $G$. A classical example is Newman's Modularity [14], defined for a weighted undirected graph as $Q(G, \sigma) = \frac{1}{m} \sum_{i,j \in V} \left( W_{ij} - \frac{s_i s_j}{m} \right) \delta(\sigma_i, \sigma_j)$, which measures the difference between the actual fraction of edges falling inside the communities and the expected fraction of such edges under the configuration null model with respect to the partition. Since the graph is entirely determined by the matrix of weights $W$, we will sometimes abuse notation and refer to $f(W, \sigma)$.

Following the approach of [22], we now list some properties that are considered desirable for value functions. These properties are either taken or adapted from [22]. The first property is relatively natural towards ensuring robustness of conclusions with respect to the data, and towards providing favorable settings for community-finding algorithms.

**Property 1** (Continuity). *For any community partition $\sigma$, the value function $f$ is continuous with respect to the weights $W$.*

Thus property 1 excludes quality functions that would heavily rely on the presence or absence of an edge without considering its weight. Strictly speaking, the impossibility proofs later in the paper only require continuity "at potential optimal $W, \sigma$ combinations". As this is difficult to guarantee a priori and questionable for practical purposes, we here require full continuity.

The second property requires that only the *ratio of weights* on different edges is relevant, as far as clustering decisions are concerned.

**Property 2** (Scale invariance). *For any graph $G(V, W)$, community partitions $\sigma, \sigma'$ of $V$ and $\alpha > 0$, the following implication holds:*

$$f(W, \sigma) \geq f(W, \sigma') \Rightarrow f(\alpha W, \sigma) \geq f(\alpha W, \sigma'), \tag{1}$$

*or equivalently*[3] *$f(W, \sigma) > f(W, \sigma') \Rightarrow f(\alpha W, \sigma) > f(\alpha W, \sigma')$.*

Scale-invariance induces that the communities should not depend on an exogenous threshold value for individual edge weights. Value functions like the adaptive scale modularity [22], or the constant Potts model [21] are therefore not scale invariant.

The next property represents the fact that the community partition of one part of the graph should not be affected by the structure of other parts of the graph. It is notably not satisfied by the modularity, which suffers from the well-known resolution limit [8], [21], or in general when the clustering is influenced by the average weight in the graph. Different notions of locality can be proposed and we use the following.

**Property 3** ($k$-locality). *Given $k \in \{0, 1\}$, consider two graphs $G_1(V, W^{(1)})$ and $G_2(V, W^{(2)})$ whose restriction to a subset of nodes $V_0$ and its neighbors at distance $k$ is identical, that is: $W_{ij}^{(1)} = W_{ij}^{(2)}$ for all $i, j \in V_0$ and, if $k = 1$, for all $i, j$ for which either $i$ or $j$ belongs to $V_0$. Consider then a community partition $\sigma$ for which $C_0 = \{i \, | \, i \in V_0\}$ is a community, and another community partition $\sigma'$ exactly identical to $\sigma$ except that the community $C_0$ is split in two communities $C_{01}$ and $C_{02}$. Then there holds*

$$f(W^{(1)}, \sigma) \geq f(W^{(1)}, \sigma') \Leftrightarrow f(W^{(2)}, \sigma) \geq f(W^{(2)}, \sigma') \tag{2}$$

Zero-locality thus means that the decision of splitting a community $C_0$ into two communities only depends on the weights of edges incident to nodes within $C_0$. The weaker notion of 1-locality would allow this decision to also depend on the edges incident to one node of $C_0$ and one node outside $C_0$. A similar case appears in the definition of "locality" in [22] for instance. In the definition of [22] however, $G_1$ and $G_2$ are allowed to have different node sets $V^{(1)}, V^{(2)}$ and $C_0$ might be split in an arbitrary way in both $\sigma$ and $\sigma'$. One could also request $\{W_{ij}^{(1)} : j \in V^{(1)}\} = \{W_{ij}^{(2)} : j \in V^{(2)}\}$ for each $i \in C_0$, without requiring that the endpoints $j$ of each edge match in $G_1$ and $G_2$. Such locality notions would impose condition (2) on a *larger* class of situations with respect to our definition of Property 3. Thus our definition is a weaker property, easier to satisfy, hence providing a stronger impossibility result. When proposing cost functions that do satisfy locality in Section 4, we will show that they actually satisfy stronger notions of locality.

The next property excludes value functions for which some "relevant" partitions would never be optimal, independently of the graph. At this stage, we abstractly define the set $\Sigma$ of relevant partitions, which would typically depend on the application.

---

3. In both statements, if we have an equality on the left hand side then $\sigma$ and $\sigma'$ can be swapped, so we also need equality on the right hand side. From this, the second statement readily implies the first. Furthermore, by redefining $W' = \alpha W$ and $\alpha' = 1/\alpha$, we can reverse the first statement so equality on the right hand side also must imply equality on the left hand side, i.e. inequality on the left can only be associated to inequality on the right. Thus the first statement implies the second one.
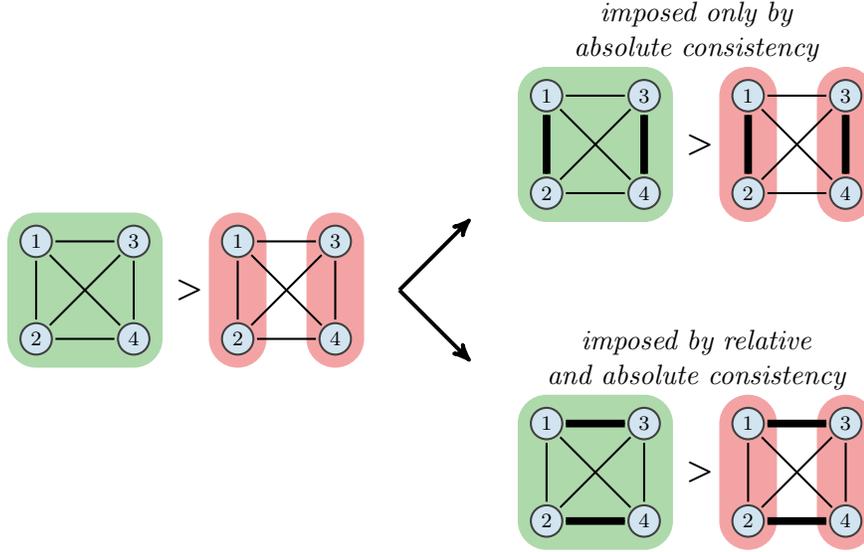
Figure 1. Illustration of the difference between (top) "absolute" consistency requirement [10] and (bottom) "relative" consistency requirement (Property 5). Symbol ">" denotes that the left partition is better than the right one.

**Property 4** (Richness, with respect to a set of partitions $\Sigma$). *For any partition $\sigma \in \Sigma$ of a set $V$ of nodes, there exists a graph $G(V, W)$, for which $\sigma$ is a strictly optimal community partition: $f(G, \sigma) > f(G, \sigma') \quad \forall \sigma' \in \Sigma \setminus \{\sigma\}$.*

Again, a stronger property could be stated by comparing to all partitions $\sigma'$ instead of only those in $\Sigma$. This makes no real difference for our results since our impossibility results in section 3 consider $\Sigma$ to be the set of all partitions, and the proposed value functions in section 4 would also satisfy this stronger version of Property 4.

The properties introduced so far do not imply that nodes inside a community should be more connected to each other than to those outside of the community, which corresponds to the general intuitive idea of community partition. In order to formalize this idea, we introduce the notion of consistent improvement as in [22]. Consider a graph $G(V, W)$ and a community partition $\sigma$. We say that $G'(V, W')$ is *a consistent improvement of $G(V, W)$ with respect to $\sigma$* if $W'_{ij} \geq W_{ij}$ for all $i, j$ for which $\sigma_i = \sigma_j$; and $W'_{ij} \leq W_{ij}$ for all $i, j$ for which $\sigma_i \neq \sigma_j$. Compared to $G$, the graph $G'$ thus has links that are stronger inside the communities defined by $\sigma$ and weaker between these communities.

Even with this notion, it turns out to be non-trivial to formalize the fact that the dependence of partitions on weights should be consistent with our intuitive idea of communities.

A natural formulation would be that an optimal partition should remain optimal for all consistent improvements of the graph with respect to this partition. This "absolute consistency" requirement is the one used in Kleinberg's impossibility result [10]. But this condition actually has counterintuitive consequences, as illustrated on Figure 1. Consider indeed a graph $G$ with a clique of four nodes $\{1, 2, 3, 4\}$ all connected one to each other by edges of similar weights. Intuitively, one would want to consider this clique as a community in the optimal partition $\sigma$. Now any graph obtained by strongly increasing the weights of the edges $(1, 2)$ and $(3, 4)$ is a consistent improvement of $G$ with respect to this partition. If these weights are sufficiently increased however, we argue that partitioning those four nodes into two communities $\{1, 2\}$ and $\{3, 4\}$ would be more natural. This would however not be allowed by "absolute consistency" (see Figure 1 top), which requires the community $\{1, 2, 3, 4\}$ to remain optimal under this consistent improvement. The impossibility proof by Kleinberg in fact relies precisely on this fact: if a partition belongs to the richness set, then none of its sub-partitions can ever be optimal.

There is however a natural workaround when using value functions: a consistent improvement should increase the value of a graph partition, but nothing forbids that the quality of another graph partition, consistent with the same improvement, increases even more and supplants the initial partition. On the example of Figure 1, since the graph modification proposed on the top right is a consistent improvement for both the partition using $\{1, 2, 3, 4\}$ and the one using two separate communities $\{1, 2\}$ and $\{3, 4\}$, this would impose no condition on their ordering. Therefore, in this paper, we only require that the ordering between partition qualities is preserved when a graph modification makes a bad partition less consistent and a good partition more consistent, as illustrated on the bottom right of Figure 1. This weaker requirement is similar to the relative monotonicity in [22] and we hence call it "relative consistency".

**Property 5** (relative consistent improvement). *Consider two graphs $G(V, W)$, $G'(V, W')$ and two partitions $\sigma, \sigma'$ of $V$. Suppose that*

   *(i) $f(G, \sigma) > f(G, \sigma')$,*

*(ii) $G'$ is a consistent improvement of $G$ with respect to $\sigma$,*
*(iii) $G$ is a consistent improvement of $G'$ with respect to $\sigma'$;*
then $f(G', \sigma) > f(G', \sigma')$.

In the following sections, we establish strict results about the impossibility to define quality functions which satisfy all the above properties. Table 2 lists how those properties are satisfied by some quality functions proposed in the literature. It is notable that none of them satisfies all properties, and we prove in Section 3 that this is indeed impossible. Note that some proposals for community partitioning involve more complex constructions than optimizing a value function. A simple variation is e.g. introducing constraints on admissible partitions in the spirit of [15], which could be reformulated as violating continuity and restricting richness. Other examples can be more difficult to (minimally) fit in the value function framework, and for the sake of simplicity we choose to not include any of them in the table below.

**Remark:** To conclude this section, we would like to insist on two points about the freedom allowed by the introduced properties.

**R1.** All our properties (with the exception of continuity) essentially state that the *ordering* of partitions according to increasing values of $f$ must be invariant under some graph transformations, but say nothing about the actual values of $f$ when the graph changes. This does not preclude that e.g. referring to property 3 (locality) the change in value from $f(W, \sigma)$ to $f(W, \sigma')$ depends on weight values outside $C_0$ and its neighbors, as long as this dependence does not affect the ordering $f(\cdot, \sigma) > f(\cdot, \sigma')$. This ordering-based approach allows more flexibility in the expression of the accepted value functions. One could of course consider analog but stronger properties about the *values* of $f$. In particular, one can verify that the following properties are sufficient but not necessary for satisfying their weaker order-based counterpart:

[value-scale invariance] For any $W, \sigma$ and $\alpha > 0$, we have $f(\alpha W, \sigma) = a(\alpha) f(W, \sigma)$ for some monotonically increasing function $a(\alpha) > 0$.

[value-$k$-locality] Consider two graphs $G_1(V, W^{(1)})$ and $G_2(V, W^{(2)})$ whose restrictions to a subset of nodes are identical as in Property 3; and partitions $\sigma, \sigma'$ which are identical except possibly on $C_0$ and for both of which $C_0$ is a union of communities. Then
$$f(G_1, \sigma) - f(G_1, \sigma') = f(G_2, \sigma) - f(G_2, \sigma').$$

[value-consistent improvement] Let $G(V, W')$ be a consistent improvement of $G(V, W)$ with respect to a community partition $\sigma$, then $f(G', \sigma) \geq f(G, \sigma)$.

**R2.** We do not impose invariance under permutation of the nodes as in [22]. This means that the value functions could depend on the particular labels of the nodes, which *does allow* using prior knowledge on the community partition. For instance, to reinforce the value of having two particular nodes $i, j$ in the same community, we can define $f_p(W, \sigma) = f(W', \sigma)$ where $W'_{1,2} = pW_{1,2}$ for some $p \gg 1$ and $W'_{i,j} = W_{i,j}$ for all other $i, j$. If $f$ satisfies some of the above properties, then the same properties are satisfied by $f_p(W, \sigma)$.

In the sequel, our impossibility results in Section 3 allow all this freedom, while the value functions which we introduce as feasibility examples in Section 4 do satisfy the stronger value-based properties of R1 and are independent of node labels. In this sense, the distinction of the "extra freedom" in R1 and R2 allows no extra margin in the impossibility and possibility results proposed in the present paper.

## 3 IMPOSSIBILITY RESULTS

We now provide two extensions to Kleinberg's result [10] about incompatible axioms for community partition based on optimizing a cost function. Our first result replaces the absolute notion of consistency used in [10] by our relative one, Property 5 (together with the continuity assumption of Property 1, related to the quality-function based framework we use). Our second result shows that even when dropping the axiom of consistent improvement, it is impossible to satisfy all the other axioms if 0-locality is required.

In Section 4, we will further identify richness as a central culprit for these impossibilities. Indeed, we will provide explicit value functions which show that the axioms can all be satisfied if richness is required with respect to all except a few particular partitions. Furthermore, we will see that these value functions satisfy all the axioms, including full richness, if the latter is allowed to exploit graphs with self-loops. The impossibility result *à la* Kleinberg indeed applies to the setting without self-loops.

The next lemma, stating that graphs for which a partition is strictly optimal can always be assumed to have positive weights $W_{ij} > 0$ on all pairs of nodes, will be useful in the sequel. The proof is available in Appendix A.

**Lemma 1.** *Let $f$ be a value function satisfying Property 1 on continuity with respect to $W$, and let $G(V, W)$ be a graph for which a partition $\sigma$ is strictly optimal: $f(G, \sigma) > f(G, \sigma')$ for all $\sigma' \neq \sigma$. Then there exists a $\delta > 0$ such that $\sigma$ is also strictly optimal for any graph $G(V, W')$ with $|W'_{ij} - W_{ij}| < \delta$ for all $i, j$.*

| | $f(W,\sigma)$ | Continuity | Scale invariance | $k$-locality | Full Richness | Rel. consistency |
|---|---|---|---|---|---|---|
| Modularity [14] | $\frac{1}{m}\sum_{i,j\in V}\left(W_{ij}-\frac{s_i s_j}{m}\right)\delta(\sigma_i,\sigma_j)$ | $y$ | $y$ | $n$ | $n^a$ | $n$ |
| Adaptative scale modularity [22] | $\sum_{k=1}^{n_c}\left[\frac{\sum_{\sigma_i=\sigma_j=k}W_{ij}}{M+\gamma v_k}-\left(\frac{v_k}{M+\gamma v_k}\right)^2\right]$ , $v_k=\sum_{\substack{\sigma_i=k\\ j\in V}}W_{ij}$ | $y$ | $\sim^b$ | $k=1$ | $\sim^c$ | $\sim^d$ |
| Constrained q-states Potts model [16] | $\alpha\sum_{\substack{i,j\in V\\ \sigma_i=\sigma_j}}[W_{ij}]-\beta\sum_{k=1}^{n_c}[|c_k|(|c_k|-1)]$ | $y$ | $n$ | $k=0$ | $y$ | $y$ |
| Constant Potts model [21] | $\sum_{\substack{i,j\in V\\ \sigma_i=\sigma_j}}[W_{ij}-\gamma]$ | $y$ | $n$ | $k=0$ | $\sim^e$ | $y$ |
| Modularity density [12] | $\sum_{k=1}^{n_c}\frac{1}{|c_k|}\sum_{i\in c_k}s_i^{int}-s_i^{ext}$ | $y$ | $y$ | $k=1$ | $n^f$ | $y$ |
| Sum of community strengths [13]$^g$ | $\sum_{k=1}^{n_c}\frac{\sum_{i\in\sigma_k}s_i^{int}-s_i^{ext}}{\sum_{i\in\sigma_k}s_i^{int}+s_i^{ext}}$ | $y$ | $y$ | $k=1$ | $n^f$ | $y$ |
| Label Propagation [20] | $\sum_{\substack{i,j\in V\\ \sigma_i=\sigma_j}}[W_{ij}]$ | $y$ | $y$ | $k=0$ | $n$ | $y$ |
| Map Equation [18] | optimal coding of random walk | $y^h$ | $y$ | $n$ | $n^f$ | $?$ |

Table 1
Properties satisfied by quality functions.

a. $\Sigma$ must exclude all partitions that contain an isolated node.
b. Yes for $M=0$
c. Yes for $\gamma=0$ at least. No for $M=0$, nor for $\gamma\geq 1$ (note that the authors reach another conclusion *with* self-loops.)
d. Yes for $\gamma\geq 2$, as claimed by authors
e. Yes for $\gamma\geq 0$
f. $\Sigma$ must exclude at least all partitions with more than one singleton community.
g. Those authors add a positivity constraint; we here consider it without constraints.
h. for connected graph

*In particular, there exists an open set of graphs $G^+(V, W^+)$ for which $\sigma$ is also strictly optimal and for which $W_{ij}^+ > 0$ for all $i \neq j$.*

Continuity is required only through this essential implication in the proofs of two main results presented in this Section.

## 3.1 Generalizing Kleinberg's impossibility result

**Theorem 1.** *There is no value function satisfying properties 1 (continuity), 2 (scale invariance), 4 (richness) with respect to all partitions, and 5 (relative consistent improvement). This holds true even if the value function is only required to be defined for a specific number of nodes $N > 1$.*

More particularly, we prove the impossibility if the richness set $\Sigma$ must contain both the partition in one community to which all nodes belong, and the partition in $N$ singleton communities.

*Proof.* We consider any set $V$ of $N > 1$ nodes and any value function satisfying the four properties mentioned in the statement of the result, and derive a contradiction.

Let $\sigma^g$ be a partition in one community to which all nodes belong, and $\sigma^s$ a partition in $N$ singleton communities. Let then $G^g(V, W^g)$ and $G^s(V, W^s)$ be graphs for which the respective partitions $\sigma^g$ and $\sigma^s$ are optimal; these graphs exist from the richness Property 4. In particular $f(W^s, \sigma^s) > f(W^s, \sigma^g)$ and $f(W^g, \sigma^g) > f(W^g, \sigma^s)$. Moreover, we can assume $W_{ij}^s, W_{ij}^g > 0$ for all $i, j$ thanks to Property 1 (continuity) which implies Lemma 1. There exists thus a $\rho > 1$ such that $\rho W_{ij}^s \geq W_{ij}^g$ for all $i \neq j$. By scale invariance, $f(W^s, \sigma^s) > f(W^s, \sigma^g)$ implies then that $f(\rho W^s, \sigma^s) > f(\rho W^s, \sigma^g)$.

Since $\rho W_{ij}^s \geq W_{ij}^g$ for all $i \neq j$, observe that $\rho W^s$ is a consistent improvement of $W^g$ for $\sigma^g$, since it corresponds to increasing the weights of edges which are all trivially within the unique community of $\sigma^g$. Similarly, $W^g$ is a consistent improvement of $\rho W^s$ for the partition $\sigma^s$ into $N$ singleton communities, as it corresponds to decreasing the weights of edges between these communities. It follows then from $f(\rho W^s, \sigma^s) > f(\rho W^s, \sigma^g)$ and Property 5 that $f(W^g, \sigma^s) > f(W^g, \sigma^g)$, in contradiction with the optimality of $\sigma^g$ for $W^g$ i.e. $f(W^g, \sigma^g) > f(W^g, \sigma^s)$. $\square$

Note that Theorem 1 does not mention any locality requirement. Before trying to find a value function satisfying locality, we must anyways first relax some of the other axioms. Possibilities in this direction are outlined in Section 4. Yet before this we present an impossibility result with another set of properties.

## 3.2 Impossibility due to locality instead of consistent improvement

In the light of the impossibility result of the previous section, it is relevant to relax some assumptions or replace them by weaker or alternative versions. One candidate property could be the consistent improvement. We have indeed seen that absolute consistency can have problematic consequences, and it could be that our version of relative consistency is still too restrictive. Besides, it is not satisfied by some famous community partition methods such as the modularity maximization (see Table 2).

This section however highlights a new impossibility, independent of any requirement related to consistent improvement, and that would therefore need to be taken into account for any property replacing consistent improvement. We remark that as consistent improvement is the only property of Section 2 that imposes the partition to correspond to our intuitive notion of communities, this impossibility result is actually relevant for any arbitrary notion of value-based partition.

**Theorem 2.** *There is no value function satisfying properties 1 (continuity), 3 for $k = 0$ (0-locality), 2 (scale invariance) and 4 (richness) with respect to all partitions. This holds true even if the value function is only required to be defined for a specific number of nodes $N > 1$.*

More particularly, we prove the impossibility if the richness set $\Sigma$ must contain a partition where two nodes form one community and a partition where these same two nodes form two singleton communities.

*Proof.* We will again show that the properties mentioned in the statement of the theorem imply a contradiction, for any set $V$ of $N > 1$ nodes.

Consider two partitions $\sigma^a, \sigma^b$ that are identical except that nodes 1 and 2 form one community in $\sigma^a$ and form two separate singleton communities in $\sigma^b$. By the richness property, there exists two graphs $G^a(V, W^a)$ and $G^b(V, W^b)$ such that $\sigma^a$ is the unique optimal partition for $G^a$ and $\sigma^b$ the unique optimal partition for $G^b$. In particular, $f(W^a, \sigma^a) > f(W^a, \sigma^b)$ and $f(W^b, \sigma^b) > f(W^b, \sigma^a)$.

It follows from the continuity property and thus Lemma 1 that the weights $W_{12}^a$ and $W_{12}^b$ of the edge connecting 1 and 2 in $G^a$ and $G^b$ respectively can both be assumed positive, so that $\rho = W_{12}^b / W_{12}^a > 0$ is well defined. By the scale-invariance property, $f(W^a, \sigma^a) > f(W^a, \sigma^b)$ implies then $f(\rho W^a, \sigma^a) > f(\rho W^a, \sigma^b)$.

By definition of $\rho$, the restriction to $\{1, 2\}$ of the graphs defined by $\rho W^a$ and $W^b$ are identical. Besides, remember that $\sigma^b$ is obtained from $\sigma^a$ by only splitting the community $\{1, 2\}$ in two singleton communities. Then because of the 0-locality property, the inequality $f(\rho W^a, \sigma^a) > f(\rho W^a, \sigma^b)$ implies $f(W^b, \sigma^a) > f(W^b, \sigma^b)$, in contradiction with the strict optimality of $\sigma^b$ for $G^b(V, W^b)$. $\square$

## 4 NEW VALUE FUNCTIONS

We have seen that none of the main value functions available in the literature satisfy all the desirable properties of Section 2, and that it is actually impossible to simultaneously satisfy their strict versions. We now show that slightly weaker versions of these properties are not mutually incompatible, by providing examples of value functions satisfying them. At this point we do not claim that these value functions lead to practically relevant community partitions; we only design them to show the possibility of simultaneously satisfying the properties of Section 2, when richness excludes a few specific partition types.

We first note that, when trying to follow a common approach in which the quality of a partition is computed by summing the individual contributions of all edges or of all edges inside communities, it appears necessary to introduce non-trivial scalings in order to simultaneously satisfy scale-invariance, locality and significant richness properties. For example, in the general framework introduced by Reichardt and Bornholdt [17], the quality function is expressed as $f(W, \sigma) = \sum_{i,j} (W_{ij} - N_{ij}) \delta(\sigma_i, \sigma_j)$ where $N_{ij}$ is the expected value of $W_{ij}$ in a given null model. This encourages grouping nodes in a community if they are joined by an edge whose weight exceeds the threshold $N_{ij}$. Now if this threshold depends on exogenous parameters (e.g. a constant), then the value function $f$ will typically not be scale-invariant. On the other hand if the $N_{ij}$ are determined endogenously involving *global* graph properties (like say the average node strength, as e.g. in the modularity [14]), then $f$ is typically non-local: a modification in any part of the graph can change the value of $N_{ij}$ and therefore affect the decision of having $i$ and $j$ in a same community or for instance as isolated nodes. This only leaves the possibility to compute $N_{ij}$ from the weights of edges incident on $i$ and $j$. A simple example of such quality function would be modularity density as defined in [12]:

$$f_{md} := \sum_k \frac{1}{|c_k|} \sum_{i \in c_k} \left( s_i^{int} - s_i^{ext} \right) . \tag{3}$$

This is a Reichardt-Bornholdt-like version of the adaptive scale modularity from [22] with $M = 0$, satisfying scale-invariance and 1-locality. It is not difficult to prove that if $\sigma$ is a partition with two isolated nodes, then $f_{md}$ can always be increased by grouping these nodes.

In the following, we show that by cleverly using community-size-dependent scalings in variations of $f_{md}$, one can obtain a class of quality functions with stronger locality or larger richness set than $f_{md}$ and than all the proposed quality functions we are aware of.

We refer the reader to Section 5 for an experimental investigation of the partitioning behavior with those value functions.

### 4.1 Sum of "average" internal strengths

Our first new value function is a re-scaled version of the sum over all communities of the internal strength of their nodes:

$$f_A(G, \sigma) = \sum_{k=1 \ldots n_c} \frac{1}{R(|c_k|)} \sum_{i \in c_k} s_i^{int} = \sum_{k=1 \ldots n_c} \frac{1}{R(|c_k|)} \sum_{i \in c_k} \sum_{j \in c_k} W_{ij}, \tag{4}$$

where $R$ is any nonnegative normalization function satisfying the following conditions: (i) $R$ is strictly increasing, and (ii) $R(a)/a$ is nonincreasing. The class of possible $R$ includes all functions $R(a) = a^\beta$ for $\beta \in (0, 1]$. For $\beta = 1$ we are just summing the *average* internal strengths of the communities.

**Theorem 3.** *The value function $f_A$ defined in* (4) *satisfies Properties 1 (continuity), 2 (scale-invariance), 3 for $k = 0$ hence also $k = 1$, Property 4 (richness) for the set $\Sigma$ of partitions with at most one singleton community, and Property 5 (relative consistent improvement).*

More precisely, $f_A$ satisfies the stronger properties of value-scale invariance, value-consistent improvement and value-0-locality mentioned in Remark R1. The locality property may in fact be extended like in [22] to graphs $G^{(1)}$ and $G^{(2)}$ with different node sets and to $\sigma$, $\sigma'$ both containing arbitrary partitions of $C_0$.

*Proof.* For a fixed $\sigma$, the value of $f_A(W, \sigma)$ is a linear combination with positive coefficients of the weights $W_{ij}$ of edges inside the communities. It is therefore clearly continuous with respect to $W$ and scale invariant, as $f_A(\alpha W, \sigma) = \alpha f_A(W, \sigma)$.

Regarding consistent improvement, consider weights $W'$ for which $W'_{ij} \geq W_{ij}$ whenever $\sigma_i = \sigma_j$ and $W'_{ij} \leq W_{ij}$ otherwise. Then there holds

$$f_A(W', \sigma) = \sum_{k=1 \ldots n_c} \frac{1}{R(|c_k|)} \sum_{i \in c_k} \sum_{j \in c_k} W'_{ij} \geq \sum_{k=1 \ldots n_c} \frac{1}{R(|c_k|)} \sum_{i \in c_k} \sum_{j \in c_k} W_{ij} = f_A(W, \sigma).$$

This proves value-consistent improvement as defined in Remark R1, which is a sufficient condition for Property 5.

Regarding locality, it is straightforward to see that the following value-locality property from Remark R1 holds and is sufficient to imply 0-locality: Given two graphs $G_1(V, W^{(1)})$ and $G_2(V, W^{(2)})$ with $W_{ij}^{(1)} = W_{ij}^{(2)}$ for all $i, j \in C_0 \subset V$, take

any partition $\sigma$ for which $C_0$ is a union of communities. Consider $\sigma'$ exactly equal to $\sigma$ except that $C_0$ might be partitioned differently into communities. Then

$$f_A(G_1, \sigma) - f_A(G_1, \sigma') = f_A(G_2, \sigma) - f_A(G_2, \sigma') \,.$$

There remains to prove that $f_A$ is rich for the set $\Sigma$ of partitions containing no more than one singleton community. For any such partition $\sigma^*$, we define a graph consisting of disjoint cliques corresponding exactly to the communities: $W_{ij} = 1$ if $\sigma_i^* = \sigma_j^*$ (and $i \neq j$) and $W_{ij} = 0$ else. The strict optimality of $\sigma^*$ for this graph is shown in Appendix B. By Lemma 1 this automatically implies that the partition $\Sigma^*$ is strictly optimal for an open set of values of $W$. $\square$

In terms of satisfying the properties of Section 2, $f_A$ is at least as good as the modular density $f_{md}$; in particular, as $f_A$ only comprises a positive term for each edge present inside a community, it satisfies the stronger property of 0-locality in conjunction with all the other axioms. This might appear surprising, considering e.g. Radicchi's proposal [15] that a community is expected to have stronger inside links than external links (the latter does not appear in $f_A$ and seems to almost exclude 0-locality). Now, the academic properties satisfied by $f_A$ do not guarantee its practical relevance, which we further investigate in Section 5. The existence of $f_A$ probably hints at an insufficiency of the listed properties towards guaranteeing relevant quality functions for community detection.

## 4.2 Penalizing external strength

From an axiomatic viewpoint, the shortcoming of $f_A$ introduced in Section 4.1 is that a partition with several singleton communities is never strictly optimal. It could however be relevant in practice to have several singleton communities. The intuitive reason for the restricted richness of $f_A$ is the following: suppose that we want nodes 1 and 2 to form two singleton communities in a partition $\sigma$. Their internal strength would by definition be 0, so the contribution of $\{1, 2\} \subset V$ to $f_A(W, \sigma)$ would be 0. On the other hand, if we join them in a two-nodes community to get the partition $\sigma'$, then their internal strength would be $W_{12}$, and the contribution of $\{1, 2\} \subset V$ to $f_A(W, \sigma')$ would be $2W_{12}$. Hence we get $f_A(W, \sigma') \geq f_A(W, \sigma)$ for any $W$, such that $\sigma$ cannot be strictly optimal.

To palliate this issue, we add in this section an incentive to keep certain weakly connected nodes in separate singleton communities. More precisely, we will penalize every edge connecting a node $i$ to other communities, with a penalization weight that depends on the size of the community to which $i$ belongs and that happens to be 0 if the community only contains $i$.

Formally, for some fixed parameter $\alpha > 0$, we define the value function

$$f_B(W, \sigma) = \sum_k \frac{1}{R(|c_k|)} \sum_{i \in c_k} \left( s_i^{int} - \alpha(|c_k| - 1) s_i^{ext} \right) \tag{5}$$

$$= \sum_k \frac{1}{R(|c_k|)} \sum_{i \in c_k} \left( \sum_{j \in c_k} W_{ij} - \alpha(|c_k| - 1) \sum_{j \notin c_k} W_{ij} \right), \tag{6}$$

where we remind that the normalization function is any positive increasing function for which $R(a)/a$ is nonincreasing.

The next theorem shows that the particular penalization weight $\alpha(|c_k| - 1)$ allows $f_B$ to be richer over a larger set $\Sigma$ than $f_A$, and than the modular density $f_{md}$ which is similar to $f_B$ but just with uniform penalization weight 1. With respect to $f_A$ this richness is however achieved at a cost, since $f_B$ only satisfies 1-locality (as $f_{md}$ does), while $f_A$ satisfies 0-locality. By Theorem 1, irrespective of locality concerns, a further extension of the richness set $\Sigma$ is not possible unless other properties are further relaxed.

**Theorem 4.** *The value function $f_B$ defined in* (5) *satisfies Properties 1 (continuity), 2 (scale-invariance), 3 (locality) with $k = 1$, Property 4 (richness) with respect to the set $\Sigma$ of partitions containing at least one community with more than one node, and Property 5 (relative consistent improvement).*

More precisely, $f_A$ satisfies the stronger properties of value-scale invariance, value-consistent improvement and value-1-locality mentioned in Remark R1. The locality property may in fact be extended like in [22] to graphs $G^{(1)}$ and $G^{(2)}$ with different node sets and to $\sigma$, $\sigma'$ both containing arbitrary partitions of $C_0$.

*Proof.* Once the partition $\sigma$ is fixed, $f(W, \sigma)$ is a linear function of the weights $W_{ij}$, it is hence trivially continuous in $W$ and scale-invariant as $f_B(\alpha W, \sigma) = \alpha f_B(W, \sigma)$.

Regarding consistent improvement, consider weights $W'$ for which $W'_{ij} \geq W_{ij}$ for all $i, j$ in the same communities and $W'_{ij} \leq W_{ij}$ for other pairs $i, j$. Then for all $i$ we have

$$\sum_{i \in C_k} \sum_{j \in C_k} W'_{ij} \geq \sum_{i \in C_k} \sum_{j \in C_k} W_{ij} \quad \text{and}$$

$$\sum_{i \in C_k} \sum_{j \notin C_k} W'_{ij} \geq \sum_{i \in C_k} \sum_{j \notin C_k} W_{ij} \;.$$

Referring to the definition (5) this straightforwardly proves value-consistent improvement as defined in Remark R1, which is a sufficient condition for Property 5.

Regarding locality, it is straightforward to see that the following value-locality property from Remark R1 holds and is sufficient to imply 1-locality: Given two graphs $G_1(V, W^{(1)})$ and $G_2(V, W^{(2)})$ with $W_{ij}^{(1)} = W_{ij}^{(2)}$ for all $i, j \in C_0 \subset V$ and for all $i, j$ for which $i$ or $j$ belongs to $C_0$, take any partition $\sigma$ for which $C_0$ is a union of communities. Consider $\sigma'$ exactly equal to $\sigma$ except that $C_0$ might be partitioned differently into communities. Then

$$f_B(G_1, \sigma) - f_B(G_1, \sigma') = f_B(G_2, \sigma) - f_B(G_2, \sigma') .$$

There remains to prove the richness property of $f_B$. We do this in two steps. For a given partition $\sigma$ of the node set $V$, let $V^* \subset V$ the nodes which do not belong to a singleton partition in $\sigma$ and $\sigma^*$ the partition of $V^*$ corresponding to $\sigma$. First, we prove that the same construction as for $f_A$, applied to the weights $W^*$ between nodes $V^*$, makes the partition $\sigma^*$ of $V^*$ optimal for $f_B$, i.e. : for any $i \in V^*$, let $W_{ij}^* = 1$ if $\sigma_i^* = \sigma_j^*$ and $W_{ij} = \delta$ otherwise, with small $\delta > 0$ selected by continuity. Second, we add (if necessary) the nodes of $V \setminus V^*$ to this construction by taking $W_{i,j} = \varepsilon$ for all $i \in V \setminus V^*$ and we show that for $\varepsilon > 0$ sufficiently small this $W$ makes $\sigma$ strictly optimal. Details are given in Appendix C. □

**Remark:** Before moving to further considerations, we must mention that *both $f_A$ and $f_B$ would satisfy all the axioms, including richness for all possible partitions of the graph nodes, if the graph was allowed to contain self-loops.* Linearity, scale-invariance and locality indeed remain trivially true if self-loops are added into $W$. Consistent improvement also still holds with the same proof. Regarding richness, we know that particular graphs, in which all self-loops have zero weight, already allow to make a large set of partitions strictly optimal. To make a partition with several isolated nodes optimal, it suffices to construct the optimal weights for the partition without those nodes, and then complete the graph by adding those nodes with each a strong self-loop and a very weak connection to any other node in the graph.

### 4.3 Relaxing consistent improvement

In Theorem 2 we have shown that even if consistent improvement is dropped, then it is still not possible to satisfy all the other criteria including 0-locality and richness with respect to all partitions. If richness is relaxed, then $f_A$ shows how also consistent improvement can be included. As a complement, the following value function shows that it is possible to be rich with respect to all partitions and satisfy 1-locality instead of 0-locality:

$$f_C(W, \sigma) = \sum_k \frac{1}{|c_k|} \left( \left( \sum_{i \in c_k} s_i^{in} \right) - \alpha |c_k|(|c_k| - 1) \min_{j \in c_k, l \in V} W_{lj} \right) .$$

One can show that $f_C$ *satisfies properties 1 (continuity), 3 with $k = 1$(1-locality), 2 (scale invariance) and 4 (richness) with respect to all partitions provided $\alpha > 1$, while it then (unavoidably, from Theorem 1) fails to satisfy Property 5 (relative consistent improvement).*

We do not provide a formal proof here because, before claiming any usefulness for $f_C$, it should be shown to satisfy some accepted criterion which would distinguish community partitions from arbitrary partitions; in the absence of such property, $f_C$ is left with the status of curiosity.

Accordingly, the main idea in the construction of $f_C$, namely enlarging the penalty term to the minimum over *all* edges connected to the community, both internal and external, is motivated by technical richness arguments and not by intuitive community characteristics. Essentially, we observe that $f_B(W, \sigma^s) = f_C(W, \sigma^s) = 0$ for all $W$ for $\sigma^s$ the all-singletons-partition. However, thanks to the modified penalty in $f_C$, with $W_{ij} = 1$ for all $i, j \in V$ we have $f_C(\sigma, W) = \sum_k (|c_k| - 1)(1 - \alpha) < 0$ for all $\sigma \neq \sigma_s$, hence the strict optimality of the all-singletons partition in that case.

#### Invariance to symmetries

To conclude this Section, let us mention that the value functions $f_A$, $f_B$ and $f_C$ are all independent of node labels, see Remark R2. They hence trivially satisfy permutation invariance as defined in [22], i.e.: if $G$ features some isometries then defining $\sigma'$ by applying one of the associated isometries to an initial partition $\sigma$ implies $f(G, \sigma') = f(G, \sigma)$.

The following Section experimentally investigates how useful the proposed value function $f_A$ and $f_B$ might be in practical clustering applications.

## 5 EXPERIMENTS

#### Benchmark and methods

We have investigated the behavior of our value functions by optimizing them for typical graphs generated using the popular LFR benchmark model [11]. This planted partitions method first defines "prior communities" on a set of nodes, and then generates a graph with edges preferentially inside these prior communities. It further ensures power-law degree and community size distributions consistent with realistic networks. We have considered graphs with $N = 1000$ nodes,
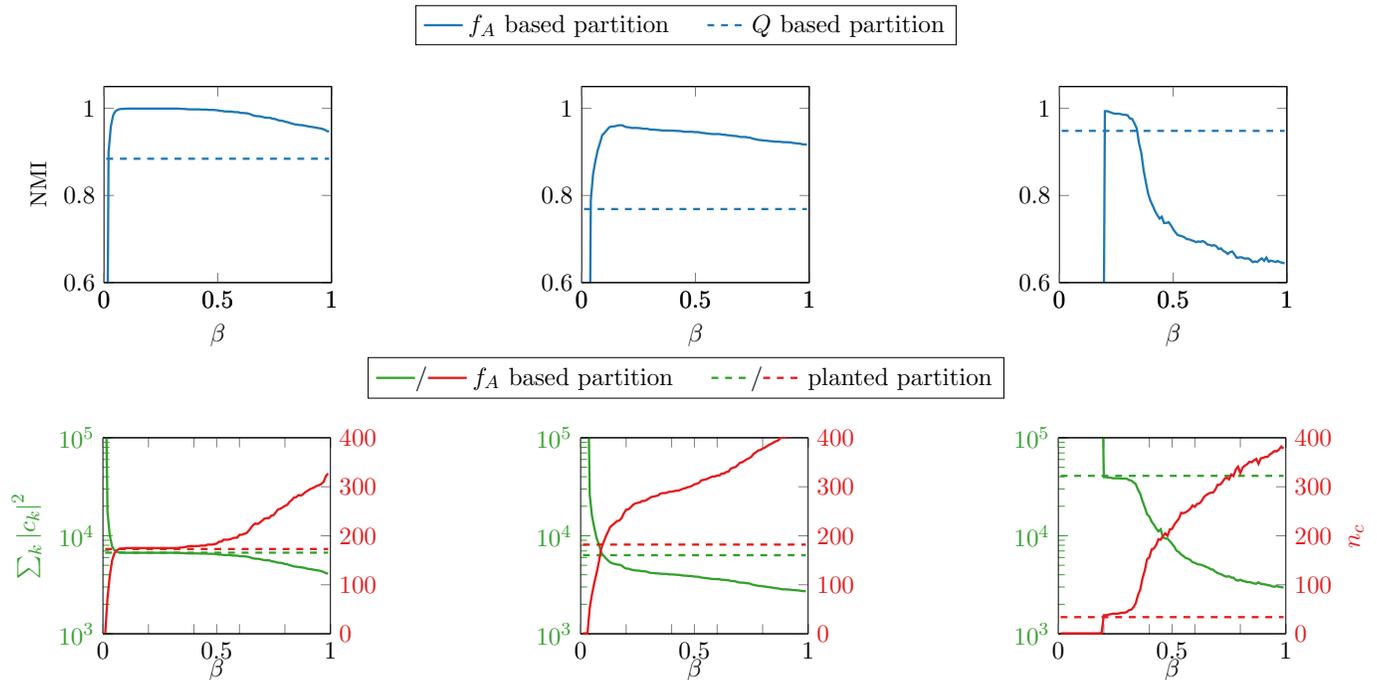
Figure 2. (color online) Quality and stability of the partitions returned by $f_A$ as a function of $\beta$ (continuous curve). The plots on the first row show the normalized mutual information (NMI) of the partitions extracted using $f_A$ with respect to the "ground truth" partition planted by the LFR benchmark model (solid line), compared to that obtained with the modularity (dashed line). The second row shows the number of communities (red, increasing with $\beta$) and their sum-of-square size (green, decreasing in $\beta$). The value of those criteria for the planted partition is also provided (dashed line). The difficulty of extracting communities increases from left to right (different $\mu_t$ and $\mu_w$ values in the LFR graph generator, see main text). For easy (left column) and medium (middle column) difficulty graphs, $f_A$ consistently outperforms modularity. For difficult graphs (right column), it is necessary to tune $\beta$ in a precise range in order to obtain, with our function $f_A$, results comparable to the modularity. This good range of $\beta$ values depends on the graph (not shown), which shows a limitation of our value function.

an average degree $\overline{k} = 25$ and community size in $[10, 100]$. Two additional parameters allow tuning the significance of the prior communities and therefore the difficulty of extracting them: $\mu_w$ the expected proportion of the strength of each node connecting it to nodes outside its prior community, and $\mu_t$ the expected proportion of edges (without taking their weight into account) connecting a node to nodes outside its prior community.

We optimize our values functions $f_A$ and $f_B$ with $R(x) = x^\beta$, for different values of $\beta \in [0, 1]$. For this purpose, we use the local optimization algorithm introduced in [4]. The latter requires self-loops at intermediate steps, which our value functions accommodate naturally, although our theoretical analysis is made without self-loops.

We evaluate the partitions extracted by optimizing our quality functions in several ways. First, we compute their normalized mutual information (NMI) [6] to the "ground truth", which we define as being the planted partition drawn by the LFR benchmark model to generate the graph. Recall that this "ground truth" is only meaningful when the method wires, on average, more edges inside the communities, i.e. $\mu_t \ll 1$, and assigns significantly larger weight to edges inside communities, i.e. for $\mu_w \leq \mu_t$; strictly speaking it is never 100% "true" except in trivial cases. In fact we are precisely seeking to better *define* communities in (almost all) non-trivial cases, and we are optimizing cost functions locally. So, although the quantitative results do most often agree with intuitive decisions about "visually reasonable or unreasonable" partitions, this comparison to ground truth must be taken with the usual grain of salt. Therefore, in complement, we have computed two indicators of stability for the partitions, namely the number of communities found $n_c$ and the sum-of-squares of the sizes of those communities $\sum_k |c_k|^2$.

## Results

Figure 2 presents the results for $f_A$ as a function of $\beta$ and compares it to those obtained with the well-established modularity criterion[14].

For easy to detect communities ($\mu_t = 0.3$ and $\mu_w = 0.1$, first column in Figure 2), sub-partitions of the planted communities are almost always found except for very small values of $\beta$. The planted communities are moreover exactly recovered for a wide range of values of $\beta$ smaller than 0.5. Selecting any value of $\beta$ should thus be expected to yield good results in such cases.
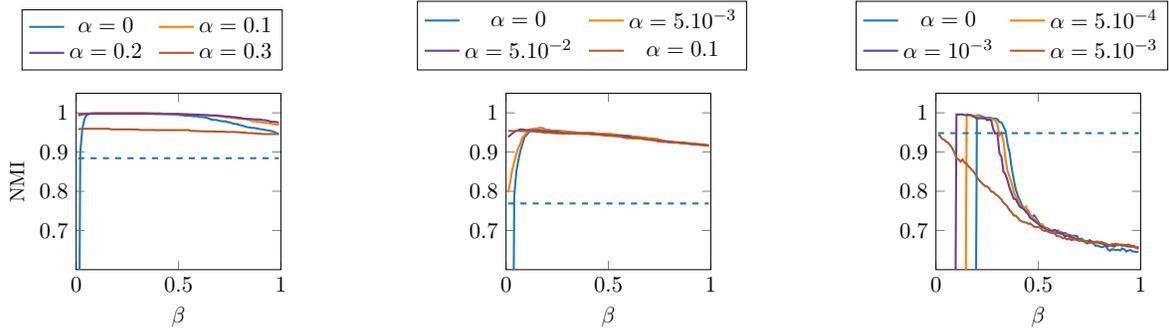
Figure 3. (color online) Quality of the partitions obtained using $f_B$, as a function of $\beta$ for different values of $\alpha$. As previously, the quality is measured using the normalized mutual information (NMI) of the partitions extracted with respect to the "ground truth" partition (solid line). The NMI obtained by the partition extracted using the modularity is also provided (dashed line). The columns correspond respectively to easy, medium and hard graphs as in Figure 2. The quality obtained by $f_A$ (i.e. $\alpha = 0$) can be outperformed with a fine tuning of the parameter $\alpha$ but the performance is highly sensitive to this parameter. This includes interdependencies between $\alpha$ and $\beta$ as can be seen in the third column.

Reasonable results are also observed for planted communities of "medium" difficulty ($\mu_t = 0.6$ and $\mu_w = 0.3$, second column of Figure 2). The extracted communities do not correspond exactly to the planted one, but the results outperform those obtained with the modularity criterion ([14], see table 2) for almost all values of $\beta$. This could be explained by the fact that modularity is known to hit a "resolution limit" which prevents it from correctly extracting communities in this situation. Our cost function $f_A$ apparently does not suffer from this problem.

The limitations of $f_A$ appear more clearly for graphs with harder-to-find communities ($\mu_t = 0.8$ and $\mu_w = 0.6$, third column of Figure 2 ). Only a very narrow range of $\beta$ values give reasonable and more or less stable results – i.e. an a priori selected value of $\beta$ would almost surely fail. This is complementary to the modularity, which precisely handles this situation well.

In conclusion, the function $f_A$ which was proposed on purely academic grounds appears to give reasonable partitions in many cases, but of course not always. An issue is the choice of $\beta$, which can strongly affect the quality of the partitions produced. Acceptable results appear not too difficult to obtain by scanning different values and analyzing the evolution of certain stability measures, i.e. selecting for example values of $\beta$ for which these measures appear "reasonable" and stable with respect to small variations of $\beta$. Such strategies would however not be compatible with our axiomatic approach. Indeed, when $\beta$ is not fixed a priori, the property of locality would almost surely be lost, and consistent improvement would not be a priori guaranteed either (those are the two properties where different graphs must be compared). There might be "soft" tuning strategies which preserve the axiomatic properties, e.g. letting $\beta$ depend on the number of nodes or replacing $R(|c_k|)$ by some slightly more complicated dependence function. Such explorations however go beyond the main purpose of the present paper.

Similar observations apply to $f_B$, see Figure 3. Reasonable results are obtained for appropriate tuning of *both* $\alpha$ and $\beta$, but the "good" tuning values depend on the graph. Since such tuning departs from our axiomatic approach, we do not include more extensive experimental results about $f_B$ in this paper.

## 6 CONCLUSION

We have considered an axiomatic approach towards defining communities as the optimum of a value function. We show that without relying on self-loops, it is not possible to satisfy a complete set of standard properties. Our main message is that this impossibility remains even when replacing Kleinberg's very strong consistent improvement requirement by a weaker and seemingly more natural form of consistency. We further show, by explicit construction, that by slightly restricting the set of a priori expected communities — e.g. excluding the case of all isolated nodes — it becomes possible to satisfy all the axioms, and reach in some situations a performance comparable to modularity. Furthermore, our constructions do satisfy the complete set of axioms when self-loops can be used to satisfy richness. These two points clarify precisely how some previous papers were able to circumvent Kleinberg's impossibility result.

This points towards several options for future research. First comes the necessity to select a subset of the standard axioms; excluding a set of a priori uninteresting partitions has been identified as one economical way to do this, but there might be others. Second and maybe more importantly, while our experimental investigations show good performance for certain parameter values, they also return extremely poor partitions for other parameter values, even though the corresponding functions still satisfy all the axioms. This demonstrates that our set of axioms is by no means sufficient to single out useful value functions, and new properties should therefore be defined. A third direction would be to extend or circumvent our impossibility result in a framework departing from value functions.

Finally, the experimental investigations with our simple value functions provide results that might be of independent interest for community partitioning. Indeed, despite their simplicity, our value functions outperform modularity in several

cases and for well-chosen parameter values. This suggests an adaptive tuning procedure for the value function parameters. Note that the axioms would no longer necessarily be satisfied when introducing such adaptive tuning.

## ACKNOWLEDGMENTS

## REFERENCES

[1] Margareta Ackerman and Shai Ben-David. Measures of clustering quality: A working set of axioms for clustering. *Advances in neural information processing systems (NIPS)*, pages 121–128, 2009.

[2] Margareta Ackerman, Shai Ben-David, and David Loker. Towards property-based classification of clustering paradigms. *Advances in Neural Information Processing Systems (NIPS)*, pages 10–18, 2010.

[3] Ulrik Brandes, Daniel Delling, Marco Gaertler, Robert Gorke, Martin Hoefer, Zoran Nikoloski, and Dorothea Wagner. On Modularity Clustering. *IEEE Transactions on Knowledge and Data Engineering*, 20(2):172–188, 2008.

[4] A. Browet, P.-A. Absil, and P. Van Dooren. Fast community detection using local neighbourhood search. *arxiv preprint 1308.6276*, August 2013.

[5] Gunnar Carlsson, Facundo Mémoli, Alejandro Ribeiro, and Santiago Segarra. Axiomatic construction of hierarchical clustering in asymmetric networks. *Proc. IEEE Int.Conf. Acoustics, Speech and Signal Processing*, pages 5219–5223, 2013.

[6] Leon Danon, Albert Díaz-Guilera, Jordi Duch, and Alex Arenas. Comparing community structure identification. *Journal of Statistical Mechanics: Theory and Experiment*, 2005(09):P09008, 2005.

[7] Santo Fortunato. Community detection in graphs. *Physics reports*, 486(3):75–174, 2010.

[8] Benjamin H. Good, Yves-Alexandre de Montjoye, and Aaron Clauset. Performance of modularity maximization in practical contexts. *Phys. Rev. E*, 81:046106, Apr 2010.

[9] Yanqing Hu, Hongbin Chen, Peng Zhang, Menghui Li, Zengru Di, and Ying Fan. Comparative definition of community and corresponding identifying algorithm. *Physical Review E*, 78:026121, 2008.

[10] Jon Kleinberg. An impossibility theorem for clustering. *Advances in neural information processing systems (NIPS)*, pages 463–470, 2003.

[11] Andrea Lancichinetti, Santo Fortunato, and Filippo Radicchi. Benchmark graphs for testing community detection algorithms. *Phys. Rev. E*, 78:046110, Oct 2008.

[12] Zhenping Li, Shihua Zhang, Rui-Sheng Wang, Xiang-Sun Zhang, and Luonan Chen. Quantitative function for community detection. *Phys. Rev. E*, 77:036109, Mar 2008.

[13] AD Medus and CO Dorso. Alternative approach to community detection in networks. *Physical Review E*, 79(6):066111, 2009.

[14] M. E. J. Newman and M. Girvan. Finding and evaluating community structure in networks. *Phys. Rev. E*, 69:026113, Feb 2004.

[15] Filippo Radicchi, Claudio Castellano, Federico Cecconi, Vittorio Loreto, and Domenico Parisi. Defining and identifying communities in networks. *Proceedings of The National Academy of Sciences*, 101(9):2658–2663, 2004.

[16] Jörg Reichardt and Stefan Bornholdt. Detecting fuzzy community structures in complex networks with a potts model. *Phys. Rev. Lett.*, 93:218701, Nov 2004.

[17] Jörg Reichardt and Stefan Bornholdt. Statistical mechanics of community detection. *Phys. Rev. E*, 74:016110, Jul 2006.

[18] Martin Rosvall and Carl T. Bergstrom. Maps of random walks on complex networks reveal community structure. *Proceedings of the National Academy of Sciences*, 105(4):1118–1123, 2008.

[19] Satu Elisa Schaeffer. Graph clustering. *Computer Science Review*, 1(1):27–64, 2007.

[20] Gergely Tibely and Janos Kertesz. On the equivalence of the label propagation method of community detection and a potts model approach. *Physica A: Statistical Mechanics and its Applications*, 387(19 - 20):4982 – 4984, 2008.

[21] V. A. Traag, P. Van Dooren, and Y. Nesterov. Narrow scope for resolution-limit-free community detection. *Phys. Rev. E*, 84:016114, Jul 2011.

[22] Twan Van Laarhoven and Elena Marchiori. Axioms for graph clustering quality functions. *Journal of Machine Learning Research*, 15(1):193–215, January 2014.

[23] Reza Bosagh Zadeh. Towards a principled theory of clustering. *13th Int.Conf. Artif.Intell. and Stat.*, 2010.

## APPENDIX A
### PROOF OF LEMMA 1

Remember that there are only finitely many possible partitions. Let $\sigma^{(2)}$ be (one of) the second best partition(s) for the graph $G(V, W)$, that is, $f(G, \sigma^{(2)}) \geq f(G, \sigma')$ for every $\sigma' \neq \sigma$, and let $\epsilon = \frac{1}{2}\left(f(G, \sigma) - f(G, \sigma^{(2)})\right)$. By continuity, for every $\sigma'$ we can pick a $\delta' > 0$ such that $||W - W'|| < \delta'$ implies $|(f(G(V, W'), \sigma') - f(G(V, W), \sigma'))| < \epsilon$. Let $\delta$ be the minimum of these $\delta'$ over all partitions, which is positive and well defined since there are finitely many of them. It follows that $||W - W'|| < \delta'$ implies $|(f(G(V, W'), \sigma') - f(G(V, W), \sigma'))| < \epsilon$ for all $\sigma'$. Using the definition of $\epsilon$, we have then for every partition $\sigma'$

$$f(W', \sigma') < f(W, \sigma') + \epsilon \leq f(W, \sigma^{(2)}) + \epsilon = f(W, \sigma) - \epsilon < f(W', \sigma),$$

which establishes the result.

## APPENDIX B
## RICHNESS OF $f_A$ IN THEOREM 3

For $\sigma^*$ a given partition, we consider the graph $G(V, W^*)$ where $W_{ij}^* = 1$ if $\sigma_i^* = \sigma_j^*$ and $W_{ij}^* = 0$ otherwise, i.e. the graph where there is one clique for each community of $\sigma^*$. We will prove that $\sigma^*$ is the strictly optimal partition of $G(V, W^*)$ according to $f_A$. The proof uses two claims.

*Claim 1: Consider a community $c_0$ of some arbitrary partition $\sigma$, and $\sigma'$ a partition identical to $\sigma$ except that $c_0$ is split into two communities $c_{01}$ and $c_{02}$. If $W_{ij}^* = 0$ for all $i \in c_{01}$ and $j \in c_{02}$ (i.e. no clique contains a node from both communities) then $f_A(\sigma', W^*) \geq f_A(\sigma, W^*)$, with equality holding if and only if $\sum_{i \in c_0} \sum_{j \in c_0} W_{ij} = 0$.*

*Proof.* (of Claim 1): Define $\bar{s}_k^{(int)} = \frac{1}{R(|c_k|)} \sum_{i \in c_k} \sum_{j \in c_k} W_{ij}^*$ the normalized internal strength of a community $c_k$. Then we have by definition

$$f_A(\sigma, W^*) - \bar{s}_0^{(int)} = f_A(\sigma', W^*) - \bar{s}_{01}^{(int)} - \bar{s}_{02}^{(int)} .$$

If $W_{ij}^* = 0$ for all $i \in c_{01}$ and $j \in c_{02}$ then we have $\sum_{j \in c_{0a}} W_{ij} = \sum_{j \in c_0} W_{ij}$ for all $i$ in $c_{0a}$ and $a \in \{0, 1\}$. Therefore, $R(|c_0|)\bar{s}_0^{(int)} = R(|c_{01}|)\bar{s}_{01}^{(int)} + R(|c_{02}|)\bar{s}_{02}^{(int)}$. From there we have

$$\bar{s}_0^{(int)} = \frac{R(|c_{01}|)}{R(|c_0|)}\bar{s}_{01}^{(int)} + \frac{R(|c_{02}|)}{R(|c_0|)}\bar{s}_{02}^{(int)} \leq \bar{s}_{01}^{(int)} + \bar{s}_{02}^{(int)}, \tag{7}$$

where we have used $|c_0| > \max\{|c_{01}|, |c_{02}|\}$ and the fact that $R$ is increasing, so that $\max\{\frac{R(|c_{01}|)}{R(|c_0|)}, \frac{R(|c_{02}|)}{R(|c_0|)}\} < 1$. It follows thus also that the inequality in (7) is strict unless $\bar{s}_{01}^{(int)} = \bar{s}_{02}^{(int)} = 0$, that is, $W_{ij} = 0$ for all $i, j \in c_0$. $\qquad\square$

*Claim 2: Consider a partition $\sigma$ where $W_{ij}^* = 1$ for all $i, j \in c_0$ and define a partition $\sigma'$ identical to $\sigma$ except that $c_0$ is split into two communities $c_{01}$ and $c_{02}$. Then $f_A(\sigma, W^*) > f_A(\sigma', W^*)$.*

*Proof.* (of Claim 2): With the same notation as in the proof of Claim 1, we still have

$$f_A(\sigma, W^*) - \bar{s}_0^{(int)} = f_A(\sigma', W^*) - \bar{s}_{01}^{(int)} - \bar{s}_{02}^{(int)} .$$

Since $W_{ij}^* = 1$ for all $i, j \in c_{01} \cup c_{02} = c_0$ we readily get $\bar{s}_{01}^{(int)} = \frac{|c_{01}|}{R(|c_{01}|)}(|c_{01}| - 1)$, $\bar{s}_{02}^{(int)} = \frac{|c_{02}|}{R(|c_{02}|)}(|c_{02}| - 1)$, and

$$\bar{s}_0^{(int)} = \frac{|c_0|}{R(|c_0|)}(|c_0| - 1) = \frac{|c_0|}{R(|c_0|)}(|c_{01}| + |c_{02}| - 1).$$

Remembering that $R(a)/a$ is nonincreasing and that $|c_0| = |c_{01}| + |c_{02}|$, we obtain

$$\bar{s}_{01}^{(int)} + \bar{s}_{02}^{(int)} \leq \frac{|c_0|}{R(|c_0|)}(|c_{01}| - 1) + \frac{|c_0|}{R(|c_0|)}(|c_{02}| - 1) < \frac{|c_0|}{R(|c_0|)}(|c_{01}| + |c_{02}| - 1) = \bar{s}_0^{(int)},$$

so that $f_A(\sigma, W^*) > f_A(\sigma', W^*)$. $\qquad\square$

We can now prove the strict optimality of $\sigma^*$. Consider a partition $\sigma^{(1)}$ and denote its communities by $c_k^{(1)}$. Define $\sigma^{(2)}$ the partition where each $c_k^{(1)}$ has been further partitioned into the largest possible communities $c_{k1}^{(2)}, c_{k2}^{(2)}, \dots c_{kn_k}^{(2)}$ for which $W_{ij}^* = 1$ whenever $\sigma_i^{(2)} = \sigma_j^{(2)}$. Note that as we require the communities to remain as large as possible, we also have $W_{ij}^* = 0$ if $\sigma_i^{(1)} = \sigma_j^{(1)}$ but $\sigma_i^{(2)} \neq \sigma_j^{(2)}$. Therefore the transition from $\sigma^{(1)}$ to $\sigma^{(2)}$ can be obtained by repeated application of a transition from $\sigma$ to $\sigma'$ as set out in Claim 1. Thus $f_A(\sigma^{(2)}, W^*) \geq f_A(\sigma^{(1)}, W^*)$.

From $\sigma^{(2)}$, we can obtain the original partition $\sigma^*$ by repeated application of a transition from $\sigma'$ to $\sigma$ (note the reversed order!) as set out in Claim 2. We hence have $f_A(\sigma^*, W^*) \geq f_A(\sigma^{(2)}, W^*)$.

Hence overall $f_A(\sigma^*, W^*) \geq f_A(\sigma^{(2)}, W^*) \geq f_A(\sigma^{(1)}, W^*)$ for any $\sigma^{(1)}$. Let us now examine in which cases we can have equality. To have $f_A(\sigma^*, W^*) = f_A(\sigma^{(2)}, W^*)$ we need $\sigma^{(2)} = \sigma^*$. To have $f_A(\sigma^{(2)}, W^*) = f_A(\sigma^{(1)}, W^*)$ we need for each $c_k^{(1)} \in \sigma^{(1)}$:

- either $n_k = 1$ i.e. $c_k^{(1)}$ has not been further partitioned in the process of obtaining $\sigma^{(2)}$, which also means that $W_{ij}^* = 1$ for all $i, j \in c_k^{(1)}$;
- or $\sum_{i \in c_k^{(1)}} \sum_{j \in c_k^{(1)}} W_{ij} = 0$, which also implies that the $c_{ki}^{(2)}$ are all singleton communities for all $i$.

The second situation is excluded if $\sigma^{(2)} = \sigma^*$ and $\sigma^*$ has no more than one singleton community. Hence we must be in the first situation for every $c_k^{(1)} \in \sigma^{(1)}$, which means that $\sigma^{(2)}$ is obtained from $\sigma^{(1)}$ by applying no modification, and then $\sigma^*$ is obtained from $\sigma^{(2)}$ by applying no modification; i.e. that $\sigma^{(1)} = \sigma^*$. For any $\sigma^{(1)} \neq \sigma^*$ we hence have $f_A(\sigma^*, W^*) > f_A(\sigma^{(1)}, W^*)$, proving the strict optimality of $\sigma^*$.

## APPENDIX C
## RICHNESS OF $f_B$ IN THEOREM 4

To establish the richness of $f_B$ with respect to all partitions containing at least one community with more than one node, we will use the following Lemma.

**Lemma 2.** *Let $\sigma_*$ be a strictly optimal partition for $f_B$ for a graph $G(V, W)$ on $N$ nodes where $\sum_j W_{ij} > 0$ for all $i \in V$, i.e. every node is incident to at least one edge with positive weight. Build the graph $G^{+1}(V^{+1}, W^{+1})$ by adding a node $N + 1$ to $G$ and one single edge of weight $\epsilon$ connecting node $N + 1$ to some node of $V$, and let $\sigma_*^{+1}$ be the partition of $V^{+1}$ obtained by adding the singleton community $\{N + 1\}$ to $\sigma_*$.*

*Then, there exists an $\epsilon^* > 0$ such that if $\epsilon < \epsilon^*$, then $\sigma_*^{+1}$ is a strictly optimal partition for $G^{+1}$. As a consequence, $\sigma_*^{+1}$ and $G^{+1}$ also satisfy the hypothesis of the present Lemma (and can thus play the role of $\sigma_*$ and $G$ in a further iteration of it).*

*Proof.* (of Lemma 2) Any partition $\sigma^{+1}$ of $G^{+1}$ can be obtained from some particular partition $\sigma$ of $G$ by either adding the node $N + 1$ to $\sigma$ as a singleton community, or merging node $N + 1$ into an existing community of $\sigma$. We start by examining these two cases assuming $\epsilon = 0$, i.e. $W_{(N+1)j} = 0$ for all $j \in V$, while $s_i^{ext} + s_i^{int} = \sum_{j \in V} W_{ij} > 0$ for all $i \in V$.

First consider a partition $\sigma^{+1}$ of $G^{+1}$ which results from the addition of node $N + 1$ to some community $c_k$ of $\sigma$. Since $W^{+1}$ just contains additional zero entries with respect to $W$ and $s_i^{ext} + s_i^{int} > 0$ for all $i \in V$, it is straightforward to observe that

$$f_B(G^{+1}, \sigma^{+1}) - f_B(G, \sigma) = \left( \frac{1}{R(|c_k| + 1)} - \frac{1}{R(|c_k|)} \right) \left( \sum_{i \in c_k} s_i^{int} \right)$$
$$+ \left( \frac{|c_k|}{R(|c_k| + 1)} - \frac{|c_k| - 1}{R(|c_k|)} \right) \left( -\alpha \sum_{i \in c_k} s_i^{ext} \right). \tag{8}$$

(With a slight abuse of notation, in $G^{+1}$ we still denote $c_k$ the set of nodes that constituted the community $c_k$ in partition $\sigma$ of the graph $G$.) In the first term of (8) we have $\frac{1}{R(|c_k| + 1)} < \frac{1}{R(|c_k|)}$ because $R$ is increasing. For the second term, observe that since $R(a)$ is increasing and $R(a)/a$ is nonincreasing, the following inequality holds for any $a > 0$:

$$\frac{a - 1}{R(a)} = \frac{a}{R(a)} - \frac{1}{R(a)} < \frac{a + 1}{R(a + 1)} - \frac{1}{R(a + 1)} = \frac{a}{R(a + 1)}.$$

In particular, $\frac{|c_k|}{R(|c_k| + 1)} > \frac{|c_k| - 1}{R(|c_k|)}$. Thus for $\sigma \neq \sigma_*$ we have $f_B(G^{+1}, \sigma^{+1}) \leq f_B(G, \sigma) < f_B(G, \sigma_*)$. For $\sigma = \sigma_*$, since we have excluded $\sum_{i \in c_k} s_i^{int} = \sum_{i \in c_k} s_i^{ext} = 0$, (8) is strictly negative and thus in all cases we have $f_B(G^{+1}, \sigma^{+1}) < f_B(G, \sigma_*)$.

Next consider the partition $\sigma^{+1}$ of $G^{+1}$ consisting of $\sigma$ plus the singleton community $\{N + 1\}$, which would thus be $\sigma_*^{+1}$ if $\sigma = \sigma_*$. The definition of $f_B$ readily implies that $f_B(G^{+1}, \sigma^{+1}) = f_B(G, \sigma)$. In particular, we $f_B(G^{+1}, \sigma^{+1}) = f_B(G, \sigma) < f_B(G, \sigma_*)$ and thus again $f_B(G^{+1}, \sigma^{+1}) < f_B(G, \sigma_*)$, whenever $\sigma \neq \sigma_*$. The only remaining case, for $\sigma = \sigma_*$, gives $f_B(G^{+1}, \sigma_*^{+1}) = f_B(G, \sigma_*)$ and must thus be the optimal one. This establishes the result for $\epsilon = 0$.

The case of sufficiently small $\epsilon > 0$ follows then directly from Lemma 1. This ensures that node $N + 1$ is also incident to an edge with positive weight, and thus $G^{+1}$ and $\sigma_*^{+1}$ also satisfy the assumptions of the Lemma. $\qquad \square$

To prove that a target partition $\sigma$ of $V$ can be made optimal under $f_B$ for some well-chosen weights $W$, we build as for $f_A$ the graph with $W_{ij} = 1$ if $\sigma_i = \sigma_j$, otherwise $W_{ij} = 0$. Define $G^*(V^*, W^*)$ and $\sigma_*$ respectively the graph and partition where all the nodes that form singleton communities in $\sigma$ have been dropped. This graph is non-empty because we have assumed $\sigma$ contains at least one community with more than one node.

With respect to $f_A$, the value function $f_B$ just adds a penalty for links between different communities, so that $f_B(W', \sigma') \leq f_A(W', \sigma')$ for any $W'$ and $\sigma'$. For our particular construction of $W^*$ associated to $\sigma^*$, we have $s_i^{ext} = 0$ for all $i$, hence $f_B(W^*, \sigma^*) = f_A(W^*, \sigma^*)$. Note that $\sigma^*$ contains no singleton community so we can apply the proof of Appendix B, showing that $\sigma^*$ is strictly optimal for $W^*$ under $f_A$, i.e. $f_A(W^*, \sigma^*) > f_A(W^*, \sigma')$ for all $\sigma' \neq \sigma^*$. We then get

$$f_B(W^*, \sigma^*) = f_A(W^*, \sigma^*) > f_A(W^*, \sigma') \geq f_B(W^*, \sigma')$$

for all $\sigma' \neq \sigma^*$, proving strict optimality of partition $\sigma^*$ for the graph weights $W^*$ also under $f_B$. By Lemma 1, partition $\sigma^*$ will still be optimal under $f_B$ for a slightly modified $W^{*\epsilon}$ where $W_{ij}^{*\epsilon} > 0$ for all $i, j \in V^*$.

The graph $G^{*\epsilon}(V^*, W^{*\epsilon})$ obtained in this way satisfies the assumptions of Lemma 2 with optimal partition $\sigma^*$. The existence of a graph for which the target partition $\sigma$ is strictly optimal follows then from a repeated application of that Lemma, adding singleton communities one by one until $\sigma$ is reached.