Incompatibility boundaries for properties of community partitions

Arnaud Browet, Julien Hendrickx, Alain Sarlette

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 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.**

I. Introduction

In the rich literature about defining communities in graphs, two major and complementary approaches are the proposal of specific 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 lose 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 Σ, there exists at least one graph for which F returns Σ as the unique[2] associated partition;
- Scale-invariance: given any graph, multiplying all the edge weights by a constant does not change the partition returned by F;
- Consistency: if F returns a partition Σ for a graph G, then it also returns Σ for any graph G' obtained by increasing the weight of intra-communities edges and decreasing the weight of inter-communities edges with respect to Σ.

Despite its natural formulation, this last property is actually debatable [19], and we will see in Section II 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. 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 [19], and is consistent with many well-known community partition methods, as those listed in Table 1 in the supplementary material.

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

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 IV). 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

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 (ELIS dept.)

[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 equivalent provided all weights are positive and self-loops can be discarded; it suffices then 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.

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 [20] 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 is in fact sufficient and the key step towards satisfying Kleinberg's set of axioms. In this sense, the present paper establishes a clear and possibly relevant way to circumvent the impossibility. Another point is that Kleinberg's impossibility refers to graphs without self-loops. Some previous investigations, like [19], let self-loops play an instrumental role. This gives more options for constructing optimal graphs towards richness. We can in fact state that our example value functions, proposed in Section IV, satisfy all the axioms of Kleinberg if we allow 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 unavoidable to select a subset of axioms, and one reasonable selection criterion is to get the most useful results — why not, to exclude some partitions as a priori irrelevant, e.g. partitions containing almost only isolated nodes. Along this line, we analyze, in Section V, the partitions obtained using one of our proposed value functions which satisfy all the axioms except strict richness for partitions involving singletons. Although these 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 allow some additional flexibility in the way of normalizing the contribution of each community. 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.

II. Properties of quality functions

As mentioned in the introduction, there are different ways of specifying what a good partition in communities is, see e.g. [7], [17] . 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 of defining a *quality measure* of a given partition for any given weighted graph. The best partition into communities is then the one maximizing the quality function [19]. 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 $V = \{1, \dots, N\}$ is the set of nodes, N 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 σ of V into communities corresponds to an assignment of each node $i \in V$ to a community label $\sigma_i \in \{1, \dots, n_c\}$, where n_c depends on the partition σ. The partition induces communities $c_1, \dots, c_{n_c}$ 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 $m = \sum_{i \in V} s_i$ the sum of all total strengths, which equals twice the total weight of the graph.

We then consider a *value (or quality) function* $f(W, \sigma)$, that represents the quality of the community partition σ on the graph G. A classical example is Newman's Modularity [14], defined for a weighted undirected graph as $Q(W, \sigma) = \frac{1}{m} \sum_{i,j \in V} \left(W_{ij} - \frac{s_i\, s_j}{m}\right) \delta(\sigma_i, \sigma_j)$, 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.

Following the approach of [19], we now list some properties that are considered desirable for value functions. These properties are either taken or adapted from [19]. 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 σ, 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, σ 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 towards partition decisions.

Property 2 (Scale invariance)**.** *For any graph $G = (V, W)$,*

community partitions σ, σ' of V and $\tau > 0$, the following implication holds:

$$f(W,\sigma) \geq f(W,\sigma') \Rightarrow f(\tau W,\sigma) \geq f(\tau W,\sigma'), \quad (1)$$

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

Scale-invariance induces that the communities should not depend on an exogeneous threshold value for individual edge weights. Value functions like the adaptive scale modularity [19], or the constant Potts model [18] 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], [18], 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: $W^{(1)}_{ij} = W^{(2)}_{ij}$ 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 σ for which $C_0 = \{i \,|\, i \in V_0\}$ is a community, and another community partition σ' exactly identical to σ except that the community C_0 is split in two communities C_{01} and C_{02}. Then it holds*

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

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 [19] for instance. In the definition of [19], 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 σ and σ'. One could also request $\{W^{(1)}_{ij} : j \in V^{(1)}\} = \{W^{(2)}_{ij} : j \in V^{(2)}\}$ 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 IV, 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 Σ of relevant partitions, which would typically depend on the application.

Property 4 (Richness, with respect to a set of partitions Σ)**.** *For any partition $\sigma \in \Sigma$ of a set V of nodes, there exists a graph $G = (V,W)$, for which σ 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 σ' instead of only those in Σ. This makes no real difference for our results since our impossibility results in Section III consider Σ to be the set of all partitions, and the proposed value functions in section IV 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 [19]. Consider a graph $G = (V,W)$ and a community partition σ. We say that $G' = (V,W')$ is *a consistent improvement of G with respect to σ* 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 σ 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 mutually connected by edges of similar weights. Intuitively, one would want to consider this clique as a community in the optimal partition σ. 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 Fig.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 improvement of this definition 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

[3] In both statements, if we have an equality on the left hand side then σ and σ' 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 $W' = \tau W$ and $\tau' = 1/\tau$, 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.

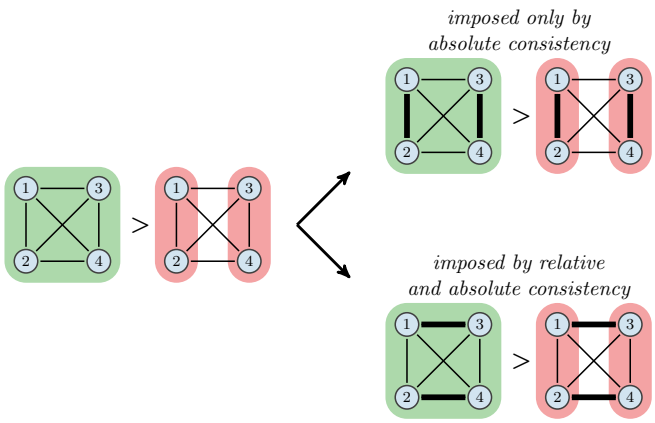

Figure 1. Illustration of the difference between (top) "absolute" consistency requirement [10] and (bottom) "relative" consistency requirement (Property 5). Symbol ">" denotes that the left partition is better than the right one.

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 [19] and we hence call it "relative consistency".

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

(i) $f(W, \sigma) > f(W, \sigma')$,

(ii) G' is a consistent improvement of G with respect to σ,

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

In the following sections, we establish strict results about the impossibility to define quality functions which satisfy all the above properties. Table 1 in the supplementary material lists how those properties are satisfied by some quality functions proposed in the literature. 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 with value functions 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 them in the table.

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 weights 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, σ and $\tau > 0$, we have $f(\tau W, \sigma) = a(\tau)\, f(W, \sigma)$ for some monotonically increasing function $a(\tau) > 0$.

[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 σ, σ' which are identical except possibly on C_0 and for both of which C_0 is a union of communities. Then $f(W^{(1)}, \sigma) - f(W^{(1)}, \sigma') = f(W^{(2)}, \sigma) - f(W^{(2)}, \sigma')$.

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

R2. We do not impose invariance under permutation of the nodes as in [19]. 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 $1, 2$ 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 III allow all this freedom, while the value functions which we introduce as feasibility examples in Section IV 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.

III. 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 IV, 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 in the supplementary material.

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

In particular, there exists an open set of graphs $G^+ = (V, W^+)$ for which σ 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 two main results presented in this Section.

A. 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 Σ 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 σ^g be a partition in one community to which all nodes belong, and σ^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 σ^g and σ^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 $W^s_{ij}, W^g_{ij} > 0$ for all i, j thanks to Property 1 (continuity) which implies Lemma 1. There exists thus a $\rho > 1$ such that $\rho W^s_{ij} \geq W^g_{ij}$ 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 $\rho W^s_{ij} \geq W^g_{ij}$ for all $i \neq j$, observe that ρW^s is a consistent improvement of W^g for σ^g, since it corresponds to increasing the weights of edges which are all trivially within the unique community of σ^g. Similarly, W^g is a consistent improvement of ρW^s for the partition σ^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 σ^g for W^g i.e. $f(W^g, \sigma^g) > f(W^g, \sigma^s)$. □

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 IV. Yet before this we present an impossibility result with another set of properties.

B. 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 1 in the supplementary material).

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 II 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 Σ 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 σ^a, σ^b that are identical except that nodes 1 and 2 form one community in σ^a and form two separate singleton communities in σ^b. By the richness property, there exist two graphs $G^a = (V, W^a)$ and $G^b = (V, W^b)$ such that σ^a is the unique optimal partition for G^a and σ^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^a_{12} and W^b_{12} of the edge connecting 1 and 2 in G^a and G^b respectively can both be assumed positive, so that $\rho = W^b_{12}/W^a_{12} > 0$ 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 ρ, the restriction to $\{1, 2\}$ of the graphs defined by ρW^a and W^b are identical. Besides, remember that σ^b is obtained from σ^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 σ^b for $G^b(V, W^b)$. □

IV. New value functions

We have seen that none of the main value functions available in the literature satisfy all the desirable properties of Section II, 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 II, 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 [16], 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 [19] with $M = 0$, satisfying scale-invariance and 1-locality. It is not difficult to prove that if σ 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 V for an experimental investigation of the partitioning behavior with those value functions.

A. 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:

$$\begin{aligned} f_A(W,\sigma) &= \sum_{k=1\dots n_c} \frac{1}{R(|c_k|)} \sum_{i\in c_k} s_i^{int} \\ &= \sum_{k=1\dots n_c} \frac{1}{R(|c_k|)} \sum_{i\in c_k}\sum_{j\in c_k} W_{ij}, \end{aligned} \tag{4}$$

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 Σ 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 [19] to graphs $G^{(1)}$ and $G^{(2)}$ with different node sets and to σ, σ' both containing arbitrary partitions of C_0.

Proof. For a fixed σ, 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 $f_A(\tau W,\sigma) = \tau f_A(W,\sigma)$.

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 it holds

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

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 σ for which C_0 is a union of communities. Consider σ' exactly equal to σ except that C_0 might be partitioned differently into communities. Then

$$f_A(W^{(1)},\sigma) - f_A(W^{(1)},\sigma') = f_A(W^{(2)},\sigma) - f_A(W^{(2)},\sigma') .$$

There remains to prove that f_A is rich for the set Σ of partitions containing no more than one singleton community. For any such partition σ^*, 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 σ^* for this graph is shown in Appendix B in the supplementary material. By Lemma 1 this automatically

implies that the partition Σ^* is strictly optimal for an open set of values of W. □

In terms of satisfying the properties of Section II, 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 V. The existence of f_A probably hints at an insufficiency of the listed properties towards guaranteeing relevant quality functions for community detection.

B. Penalizing external strength

From an axiomatic viewpoint, the shortcoming of f_A introduced in Section IV-A 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 σ. 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 σ', 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 σ 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.

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

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

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

where we remind that the normalization function is any positive increasing function for which $R(a)/a$ is nonincreasing. The factor $(|c_k|-1)$ for the penalization of external edges is not the unique one that works for our proof below. It is designed to be less severe with small communities; in particular, the penalty is 0 if the community only contains a single node. One can also understand this choice as penalizing an external edge by exactly the number of internal connections that the node could have made in the given partition.

The next theorem shows that this penalization allows f_B to be richer over a larger set Σ 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 Σ 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 Σ 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 [19] to graphs $G^{(1)}$ and $G^{(2)}$ with different node sets and to σ, σ' both containing arbitrary partitions of C_0.

Proof. Once the partition σ 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 $f_B(\tau W,\sigma) = \tau f_B(W,\sigma)$.

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 $W^{(1)}_{ij} = W^{(2)}_{ij}$ 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 σ for which C_0 is a union of communities. Consider σ' exactly equal to σ except that C_0 might be partitioned differently into communities. Then

$$f_B(W^{(1)},\sigma) - f_B(W^{(1)},\sigma') = f_B(W^{(2)},\sigma) - f_B(W^{(2)},\sigma') \,.$$

There remains to prove the richness property of f_B. We do this in two steps. For a given partition $\sigma \in \Sigma$ of the node set V, let $V^* \subseteq V$ be the set of nodes that do not belong to a singleton partition in σ and σ^* the partition of V^* corresponding to σ. First, we prove that the same construction as for f_A, applied to the weights W^* between nodes in V^*, makes the partition σ^* 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 a 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 σ strictly optimal. Details are given in Appendix B in the supplementary material. □

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.

C. 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 for graphs on $N \geq 3$ nodes, *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. [4]

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 σ^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 $f_C(W,\sigma) = \sum_k(|c_k|-1)(1-\alpha) < 0$ for all $\sigma \neq \sigma_s$, hence the strict optimality of the all-singletons partition in that case.

[4] This curiosity, however, shows that the intuition on the two-nodes graph, that a single scale-invariant edge cannot serve to distinguish whether the nodes should be merged or not, does not carry any further to $N > 2$.

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 [19], i.e.: if G features some isometries then defining σ' by applying one of the associated isometries to an initial partition σ implies $f(W,\sigma') = f(W,\sigma)$.

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

V. Experiments

Benchmark and methods

We have investigated the behavior of our value functions by optimizing them for typical graphs generated using a weighted version of 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, 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: μ_w the expected proportion of the strength of each node connecting it to nodes outside its prior community, and μ_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 $\sum_k |c_k|^2$.

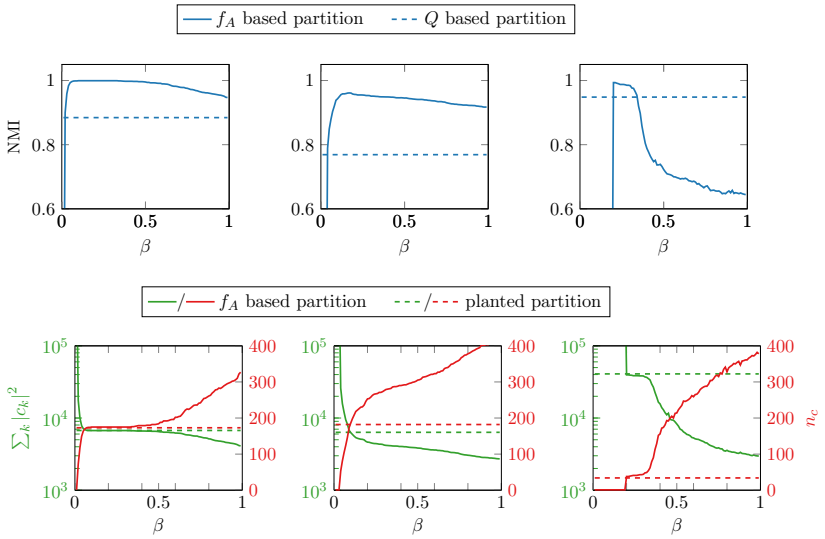

Figure 2. (color online) Quality and stability of the partitions returned by f_A as a function of β (continuous curve). The plots on the first row show the normalized mutual information (NMI) of the partitions extracted using f_A with respect to the "ground truth" partition planted by the LFR benchmark model (solid line), compared to that obtained with the modularity (dashed line). The second row shows the number of communities (red, increasing with β) and their sum-of-square size (green, decreasing in β). 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 μ_t and μ_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 β in a precise range in order to obtain, with our function f_A, results comparable to the modularity. This good range of β values depends on the graph (not shown), which shows a limitation of our value function.

In summary, the parameters for the LFR benchmark in our experiments are: number of nodes: 1000; average degree: 25; max degree: 100; min community size: 10; max community size: 100. We use three "types of graph" from which it is more ore less obvious to extract communities. The corresponding parameters are, for easy: $\mu_t = 0.3$, $\mu_w = 0.1$; medium: $\mu_t = 0.6$, $\mu_w = 0.4$; hard: $\mu_t = 0.8$, $\mu_w = 0.6$.

Results

Figure 2 presents the results for f_A as a function of β 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 β. The planted communities are moreover exactly recovered for a wide range of values of β smaller than 0.5. Selecting any value of β should thus be expected to yield good results in such cases.

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 1 in the supplementary material) for almost all values of β. 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 appears not to 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 β values give reasonable and more or less stable results – i.e. an a priori selected value of β 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 β, 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 β for which these measures appear "reasonable" and stable with respect to small variations of β. Such strategies would however not be compatible with our axiomatic approach. Indeed, when β 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 β 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* α and β, 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.

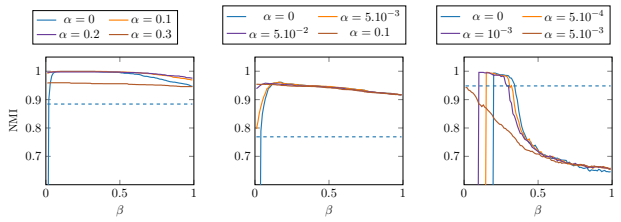

Figure 3. (color online) Quality of the partitions obtained using f_B, as a function of β for different values of α. As previously, the quality is measured using the normalized mutual information (NMI) of the partitions extracted with respect to the "ground truth" partition (solid line). The NMI obtained 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 α but the performance is highly sensitive to this parameter.This includes interdependencies between α and β as can be seen in the third column.

VI. 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.

Acknowledgment

The authors want to thank Vincent Traag, Vincent Blondel and Renaud Lambiotte for occasional but very helpful discussions on this topic, in addition to drawing our attention to it and providing useful references. This paper presents research results of the Belgian Network DYSCO (Dynamical Systems, Control, and Optimization), funded by the Inter-university Attraction Poles program, initiated by the Belgian State, Science Policy Office. QUANTIC is a joint lab involving INRIA Paris (host institution), Mines ParisTech, Ecole Normale Superieure Paris, Université Pierre & Marie Curie Paris and CNRS.

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. Statistical mechanics of community detection. *Phys. Rev. E*, 74:016110, Jul 2006.

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

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

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

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