An axiomatic study of objective functions for graph clustering

Twan van Laarhoven tvanlaarhoven@cs.ru.nl
Institute for Computing and Information Sciences, Radboud University Nijmegen, The Netherlands

Elena Marchiori elenam@cs.ru.nl
Institute for Computing and Information Sciences, Radboud University Nijmegen, The Netherlands

Editor: ??

Abstract

We investigate axioms that intuitively ought to be satisfied by graph clustering objective functions. Two tailored for graph clustering objectives are introduced, and the four axioms introduced in previous work on distance based clustering are reformulated and generalized for the graph setting. We show that modularity, a standard objective for graph clustering, does not satisfy all these axioms. This leads us to consider adaptive scale modularity, a variant of modularity, that does satisfy the axioms. Adaptive scale modularity has two parameters, which give greater control over the clustering. Standard graph clustering objectives, such as normalized cut and unnormalized cut, are obtained as special cases of adaptive scale modularity. We furthermore show that adaptive scale modularity does not have a resolution limit. In general, the results of our investigation indicate that the considered axioms cover existing 'good' objective functions for graph clustering, and can be used to derive an interesting new family of objectives.

Keywords: graph clustering, modularity, axiomatization

1. Introduction

Following the work by Kleinberg (2002) there have been various contributions to the theoretical foundation and analysis of clustering, such as axiomatization of quality functions (Ackerman and Ben-David, 2008), of criteria for comparing clusterings (Meila, 2005), uniqueness theorems for specific types of clustering (Zadeh and Ben-David, 2009; Ackerman and Ben-David, 2013; Carlsson et al., 2013), taxonomy of clustering paradigms (Ackerman et al., 2010), and characterization of diversification systems (Gollapudi and Sharma, 2009).

Kleinberg focused on clustering functions, which are functions from a distance function to a clustering. He showed that there are no clustering functions that simultaneously satisfy three intuitive axioms: scale invariance, monotonicity and richness. Ackerman and Ben-David (2008) continued on this work, and showed that the impossibility result does not apply when formulating these axioms in terms of quality functions instead of clustering functions.

Both of these previous works are formulated in terms of distance functions over a fixed domain. In this paper we focus on weighted graphs, where the weight of an edge indicates the strength of a connection. Essentially, the weight of an edge is the inverse of the distance between its two nodes.

Graphs provide additional freedoms over distance functions. In particular, it is possible for two points to be completely unrelated, indicated by a weight of 0. These zero-weight edges in turn make it natural to consider graphs over different sets of nodes as part of a larger graph. Secondly, we can allow for self loops. Self loops can indicate internal edges in a node. This notation is used for instance by Blondel et al. (2008), where a graph is contracted based on a fine-grained clustering.

In the setting where edges with weight 0 are possible, Kleinberg's impossibility result does not apply. This can be seen by considering the connected components of a graph. This is a clustering function that satisfies all three of Kleinberg's axioms: scale invariance, monotonicity (also called consistency) and richness (see Section 4.2).

The rest of this paper is organized as follows: Section 2 gives basic definitions. Next, section 3 discusses different ways in which axioms could be formulated.

In Section 4 of this paper we propose an axiomatic framework for graph clustering objective functions that consists of six axioms: the (adaption of the) four axioms by Ackerman and Ben-David (2008) (permutation invariance, scale invariance, richness and monotonicity) and two additional axioms specific for the graph setting: continuity and the locality.

Then, in Section 5, we show that modularity (Newman, 2006a) does not satisfy the monotonicity and locality axioms.

This result motivates the analysis of variants of modularity, leading to the derivation of a new parametric objective function in Section 6, that satisfies all axioms. This objective, which we call adaptive scale modularity, has two parameters, M and γ which can be tuned to control the resolution of the clustering. We show that objective functions similar to normalized cut and unnormalized cut are obtained in the limit when M goes to zero and to infinity, respectively. Furthermore, setting γ to 0 yields a parametric objective function similar to that proposed by Reichardt and Bornholdt (2004).

1.1 Related Work

Previous work on axioms for clustering objective functions has focused mostly on hierarchical clustering and on weakest and strongest link style objective functions (Kleinberg, 2002; Ackerman and Ben-David, 2008; Zadeh and Ben-David, 2009; Carlsson et al., 2013). All of these are framed in terms of distance (or similarity and dissimilarity) functions. This style of clustering is not commonly used for network community detection (see e.g. an overview in Fortunato, 2010).

Bubeck and Luxburg (2009) studied statistical consistency of clustering methods. They introduced the so-called nearest neighbor clustering and showed its consistency also for standard graph based objective functions, such as normalized cut, ratio cut, and modularity. Here we do not focus on properties of methods to optimize clustering objectives, but on natural properties that objective functions for graph clustering should satisfy.

Related works on graph clustering objective functions mainly focus on the so-called resolution limit, that is, the tendency of an objective function to prefer either small or large clusters. In particular, Fortunato and Barthélemy (2007) proved that modularity may not detect clusters smaller than a scale which depends on the total size of the network and on the degree of interconnectedness of the clusters. Van Laarhoven and Marchiori (2013) showed that the resolution limit is the most important difference between objective functions in graph clustering optimized using local search optimization.

To mitigate the resolution limit phenomenon, the objective may be extended with a so-called resolution parameter. For example, Reichardt and Bornholdt (2006) proposed a formulation of network community detection based on principles from statistical mechanics. This interpretation leads to the introduction of a family of objectives with a parameter that allows to control the clustering resolution. In Section 6.1 we will show that this extension is a special case of adaptive scale modularity.

Traag et al. (2011) formalized the notion of resolution-free objective functions, that is, not suffering from the resolution limit, and provided a characterization of this class of objectives. Their notion is essentially an axiom, and we will discuss the relation to our axioms in Section 4.1.1.

2. Definitions and Notation

A *symmetric weighted graph* is a pair (V, E) of a finite set V of nodes and a function $E : V \times V \to \mathbb{R}_{\geq 0}$ of edge weights, where $E(i, j) = E(j, i)$ for all $i, j \in V$. Edges with larger weights represent stronger connections, so missing edges can get weight 0. We explicitly allow for self loops, that is, nodes for which $E(i, i) > 0$.

A *clustering* C of a graph $G = (V, E)$ is a partition of its nodes. That is, $\bigcup C = V$ and for all $c_1, c_2 \in C$, $c_1 \cap c_2 \neq \emptyset$ if and only if $c_1 = c_2$. When two nodes i and j are in the same cluster in clustering C, i.e. when $i, j \in c$ for some $c \in C$, then we write $i \sim_C j$. Otherwise we write $i \not\sim_C j$.

A clustering C is a *refinement* of a clustering D, written $C \sqsubseteq D$, when for every cluster $c \in C$ there is a cluster $d \in D$ such that $c \subseteq d$.

A *graph clustering objective function* Q is a function from graphs G and clusterings of G to the real numbers. We adopt the convention that higher objective values indicate 'better' clusterings. As a generalization, we will sometimes work with parameterized *families of objective functions*. A single objective function can be seen as a family with no parameters.

Let $G_1 = (V_1, E_1)$ and $G_2 = (V_2, E_2)$ be two graphs and let $V_a \subseteq V_1 \cap V_2$ be a subset of the common nodes. We say that the graphs *agree on* V_a if $E_1(i, j) = E_2(i, j)$ for all $i, j \in V_a$. We say that the graphs also *agree on the neighborhood of* V_a If

- $E_1(i, j) = E_2(i, j)$ for all $i \in V_a$ and $j \in V_1 \cap V_2$,
- $E_1(i, j) = 0$ for all $i \in V_a$ and $j \in V_1 \setminus V_2$, and
- $E_2(i, j) = 0$ for all $i \in V_a$ and $j \in V_2 \setminus V_1$.

This means that for nodes in V_a the weights and endpoints of incident edges are exactly the same in the two graphs.

3. On the Form of Axioms

There are three different ways to state axioms for clustering:

1. As a property of clustering functions, as in Kleinberg (2002). For example, scale invariance of a clustering function f would be written as "$f(G) = f(\alpha G)$, for all

graphs G, $\alpha > 0$". I.e. the optimal clustering is invariant under scaling of edge weights.

2. As a property of the values of an objective function Q, as in Ackerman and Ben-David (2008). For example "$Q(G, C) = Q(\alpha G, C)$, for all graphs G, all clustering C of G, and $\alpha > 0$". I.e. the objective value is invariant under scaling of edge weights.

3. As a property of the relation between objective values of different clustering, or equivalently, as a property of an ordering of clusterings for a particular graph. For example "$Q(G, C) \geq Q(G, D) \Rightarrow Q(\alpha G, C) \geq Q(\alpha G, D)$".I.e. the 'better than' relation for clusterings is invariant under scaling of edge weights.

The third form is slightly more flexible than the other two. Any objective function that satisfies axioms in the second style will also satisfy the corresponding axioms in the third style, but the converse is not true. Note also that if D is not restricted in an axiom in the third style, then one can take $f(G) = \text{argmax}_C\, Q(G, C)$ to obtain a clustering function and an axiom in the first style.

Most axioms are more easily stated and proved in the second, absolute, style. Therefore, we adopt the second style unless doing so requires us to make specific choices.

4. Axioms for Graph Clustering Objective Functions

In Ackerman and Ben-David (2008) the authors give four axioms for distance based clustering objective functions. These axioms can easily be adapted to the graph setting.

The first axiom, permutation invariance, requires that an objective function does not depend on the identity of nodes, but only on the weight of edges between them. We might also call this permutation invariance.

Axiom 1 (Permutation invariance) *A graph clustering objective function Q is* permutation invariant *if for all graphs $G = (V, E)$ and all isomorphisms $f : V \to V'$, it is the case that $Q(G, C) = Q(f(G), f(C))$; where f is extended to graphs and clusterings by $f(C) = \{\{f(i) \mid i \in c\} \mid c \in C\}$ and $f((V, E)) = (V', (i, j) \mapsto E(f^{-1}(i), f^{-1}(j)))$.*

The second axiom, scale invariance, requires that the objective value doesn't change when edge weights are scaled uniformly. Ackerman and Ben-David (2008) defined scale invariance by insisting that the objective value stays equal when distances are scaled. In contrast, in Puzicha et al. (1999) the objective value should scale proportional with the scaling of distances. We generalize both of these previous definitions by only considering the relations between objective values.

Another consideration is that families of objective functions often have extra parameters, such as a threshold or target scales, which are also scale dependent. To accommodate such families, we allow the parameter values to change with the scaling of edge weights. This makes intuitive sense when one thinks in terms of units: a graph with edges in "m/s" can be scaled to a graph with edges in "km/h", but then the parameters of an objective function might also have to be scaled from "m/s" to "km/h".

Axiom 2 (Scale invariance) *A family of graph clustering objective function Q_P parameterized by $P \in \mathcal{P}$ is* scale invariant *if for all constants $P \in \mathcal{P}$ and $\alpha > 0$ there is a $P' \in \mathcal{P}$ such that for all graphs $G = (V, E)$, and all clusterings C_1, C_2 of G, $Q_P(G, C_1) \leq Q_P(G, C_2)$ if and only if $Q_{P'}(\alpha G, C_1) \leq Q_{P'}(\alpha G, C_2)$. Where $\alpha G = (V, (i, j) \mapsto \alpha E(i, j))$ is a graph with edge weights scaled by a factor α.*

Thirdly, we want to rule out trivial objective functions. This is done by requiring richness, i.e. that by changing the graph weights any clustering can be made optimal for that objective.

Axiom 3 (Richness) *A graph clustering objective function Q is* rich *if for all sets V and all non-trivial partitions C^* of V, there is a graph $G = (V, E)$ such that C^* is the Q-optimal clustering of V, i.e.* $\operatorname{argmax}_C Q(G, C) = C^*$.

The last axiom that Ackerman and Ben-David consider is by far the most interesting. Intuitively, we expect that when the edges within a cluster are strengthened, or when edges between clusters are weakened, that this does not decrease the objective value. Formally such a change of a graph is a consistent improvement.

Definition 1 (Consistent improvement) *Let $G = (V, E)$ be a graph and C a clustering of G. A graph $G' = (V, E')$ is a C-*consistent improvement of *G if for all nodes i and j, $E'(i, j) \geq E(i, j)$ whenever $i \sim_C j$ and $E'(i, j) \leq E(i, j)$ whenever $i \not\sim_C j$.*

We say that an objective function that does not decrease under consistent improvement is monotonic. In previous work this axiom is often called consistency.

Axiom 4 (Monotonicity) *A graph clustering objective function Q is* monotonic *if for all graphs G, all clusterings C of G and all C-consistent improvements G' of G it is the case that $Q(G', C) \geq Q(G, C)$.*

4.1 Locality

In the graph setting it also becomes natural to look at combining different graphs. With distance functions this is impossible, since it is not clear what the distance between nodes from the two different sets should be. But for graphs we can take the edge weight between nodes not in both graphs to be zero, which is the case when the graphs agree on the neighborhood of some set.

Consider adding nodes to one side of a large network, then we would not want the clustering on the other side of the network to change if there is no direct connection. For example, if a new protein is discovered in yeast, then the clustering of unrelated proteins in humans should remain the same. Similarly, we can consider any two graphs with disjoint node sets as one larger graph. Then the objective values of clusterings of the two original graphs should relate directly to objective values on the combined graph.

In general, local changes to a graph should have only local consequences to a clustering. Or in other words, the contribution to the objective value for a cluster should only depend on nodes in the neighborhood of that cluster.

Axiom 5 (Locality) *A graph clustering objective function Q is* local *if for all graphs $G_1 = (V_1, E_1)$ and $G_2 = (V_2, E_2)$ that agree on a set V_a and its neighborhood, and for all clusterings C_a, D_a of V_a, C_1 of $V_1 \setminus V_a$ and C_2 of $V_2 \setminus V_a$, if $Q(G_1, C_a \cup C_1) \geq Q(G_1, D_a \cup C_1)$ then $Q(G_2, C_a \cup C_2) \geq Q(G_2, D_a \cup C_2)$.*

Any objective function that has a preference for a fixed number of clusters will not be local. On the other hand, an objective function that is written as a sum over clusters, where each summand depends only on properties of nodes and edges in one cluster, is local.

4.1.1 Relation to Resolution-Limit-Free Objective Functions

Traag et al. (2011) introduced the notion of *resolution-limit-free* objective functions, which is similar to locality. They then showed that resolution-limit-free objectives do not suffer from the resolution limit as described by Fortunato and Barthélemy (2007). Their definition is as follows.

Definition 2 (Resolution-limit-free) *Call a clustering C of a graph G Q-optimal if for all clustering C' of G we have that $Q(G, C) \geq Q(G, C')$. Let C be a Q-optimal clustering of a graph G_1. Then the objective function Q is called resolution-limit-free if for each subgraph G_2 induced by $D \subset C$, the partition D is also Q-optimal.*

There are three differences compared to our locality axiom. First of all, Definition 2 refers only to the optimal clustering, not to the objective value, i.e. it is an axiom in the style of Kleinberg. Secondly, locality does not require that G_2 be a subgraph of G_1. Locality is stronger in that sense. Thirdly, and perhaps most importantly, in the subgraph G_2 induced by $D \subset C$, edges from a node in D to nodes not in D will be removed. That means that while G_1 and G_2 agree on the set of common nodes, they do not also agree on their neighborhood. So in this sense locality is weaker than resolution-limit-freedom.

The notion of resolution-limit-free objectives was born out of the need to avoid the resolution limit of graph clustering. And indeed locality is not enough to guarantee that an objective is free from this resolution limit.

We could look at a stronger version of locality, which replaces agreement on the neighborhood of a set V_a by plain agreement on that set. Such a *strong locality* axiom would imply resolution-limit-freedom. However, it is a very strong axiom in that it rules out many sensible objective functions. In particular, a strongly local objective function can not depend on the weight of edges entering or leaving a cluster, since that weight can be different in another graph that agrees only on that cluster.

The solution used by Traag et al. is to use the number of nodes instead of the volume of a cluster. In this way they obtain a resolution-limit-free variant of the Potts model by Reichardt and Bornholdt (2004), which they call the constant Potts model.

4.2 Continuity

In the context of graphs, perhaps the most intuitive clustering function is finding the connected components of a graph. As an objective function, we could write

$$Q_{\text{coco}}(G, C) = \mathbf{1}[C \text{ contains exactly the connected components of } G].$$

This objective function is clearly permutation invariant, scale invariant, rich, and local. Since a consistent change can only remove edges between clusters and add edges within clusters, the coco objective is also monotonic.

In fact, all of Kleinberg's axioms (reformulated in terms of graphs) also hold for this objective, which seems to refute their impossibility result. However, the impossibility proof can not be directly transfered to graphs, because it involves a multiplication and division by a maximum distance. In the graph setting this would be multiplication and division by a minimum edge weight, which can be zero.

Still, despite connected components satisfying all axioms, it is not a very useful objective function. In many real-world graphs, most nodes are part of one giant connected component (Bollobás, 2001). We would also like the clustering to be influenced by the weight of edges, not just by their existence. A natural way to rule out such degenerate objective functions is to require continuity.

Axiom 6 (Continuity) *An objective function Q is* continuous *if a small change in the graph leads to a small change in the objective value. Formally, Q is continuous if for every $\epsilon > 0$ and every graph $G = (V, E)$ there exists a $\delta > 0$ such that for all graphs $G' = (V, E')$, if $E(i, j) - \delta < E'(i, j) < E(i, j) + \delta$ for all nodes i and j, then $Q(G', C) - \epsilon < Q(G, C) < Q(G', C) + \epsilon$ for all clusterings C of G.*

Connected components clustering is not continuous, because adding an edge with a small weight δ between clusters changes the connected components, and hence dramatically changes the objective value.

Continuous objective functions have an important property in practice, in that they provide a degree of robustness to noise. A clustering that is optimal with regard to a continuous objective will still be close to optimal after a small change to the graph.

5. Modularity

For graph clustering one of the most popular objective functions is modularity (Newman and Girvan, 2004),

$$Q_{\text{modularity}}(G, C) = \sum_{c \in C} \left(\frac{w_c}{v_V} - \left(\frac{v_c}{v_V} \right)^2 \right). \tag{1}$$

In this expression $v_c(G) = \sum_{i \in c} \sum_{j \in V} E(i, j)$ is the volume of a cluster, while $w_c(G) = \sum_{i,j \in c} E(i, j)$ is the within cluster weight. v_V is the volume of the entire graph. We leave the argument G implicit for readability.

It is easy to see that modularity is permutation invariant, scale invariant and continuous.

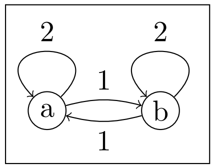
Theorem 1 *Modularity is rich.*

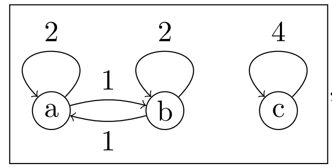
The proof of Theorem 1 is in appendix A.

An important aspect of the modularity objective is that volume and within weight are normalized with respect to the total volume of the graph. This ensures that the objective function is scale invariant, but it also means that the objective value can change in unexpected ways when the total volume of the graph changes. This leads us to Theorem 2.

Theorem 2 *Modularity is not local.*

Proof Consider the graphs

$G_1 =$ 

$G_2 =$ 

,

which agree on the set $V_a = \{a, b\}$. Note that we draw the graphs as directed graphs, to make it clear that each undirected edge is counted twice for the purposes of volume and within cluster weight. Now take the clusterings $C_a = \{\{a\}, \{b\}\}$ and $D_a = \{\{a, b\}\}$ of V_a; $C_1 = \{\}$ of $V_1 \setminus V_a$; and $C_2 = \{\{c\}\}$ of $V_2 \setminus V_a$. Then

$$Q_{\text{modularity}}(G_1, C_a \cup C_1) = 1/6 > 0 = Q_{\text{modularity}}(G_1, D_a \cup C_1),$$

while

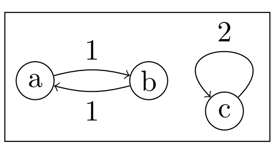
$$Q_{\text{modularity}}(G_2, C_a \cup C_2) = 23/50 < 24/50 = Q_{\text{modularity}}(G_2, D_a \cup C_2).$$

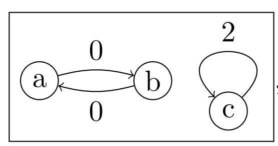
This counterexample shows that modularity is not local. ■

Even without changing the node set, changes in the total volume can be problematic, as shown by the following theorem.

Theorem 3 *Modularity is not monotonic.*

Proof Consider the graphs

$G =$ 

$G' =$ 

,

and the clustering $C = \{\{a\}, \{b\}, \{c\}\}$. G' is a C-consistent improvement of G, because the weight of a between-cluster edge is decreased. The modularity of C in G is $Q_{\text{modularity}}(G, C) = 1/8$, while the modularity of C in G' is $Q_{\text{modularity}}(G', C) = 0$. So modularity can decrease with a consistent change of a graph, and hence it is not a monotonic objective function. ■

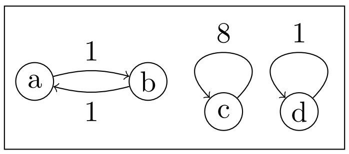
Monotonicity might be too strong a condition. When the goal is to find a clustering of a single graph, we are not actually interested in the absolute value of an objective function. Rather, what is of interest is the optimal clustering, and which changes to the graph preserve this optimum. At a smaller scaler, we can look at the relation between two clusterings. If C is better then D on a graph G, then on what other graphs is C better then D?

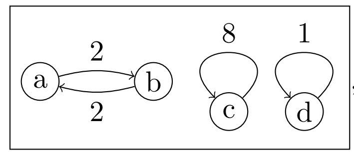
We therefore define a relative version of monotonicity, in the hopes that modularity does satisfy this weaker version.

Definition 3 (Relative monotonicity) *An objective function Q is* relatively monotonic *if for all graphs G and G' and clusterings C and D, if G' is a C-consistent improvement of G and G is a D-consistent improvement of G' and $Q(G,C) \geq Q(G,D)$ then $Q(G',C) \geq Q(G',D)$.*

Theorem 4 *Modularity is not relatively monotonic.*

Proof Take the graphs

$G =$ 

$G' =$ 

,

and the clusterings $C = \{\{a,b,c\},\{d\}\}$ and $D = \{\{a\},\{b\},\{c,d\}\}$. G' is a C-consistent improvement of G, because the weight of a within cluster edge is increased. G is a D-consistent improvement of G', because the weight of a between cluster edge is decreased. However $Q_{\text{modularity}}(G,C) = 20/121 > 16/121 = Q_{\text{modularity}}(G,D)$ while $Q_{\text{modularity}}(G',C) = 24/169 < 28/121 = Q_{\text{modularity}}(G',D)$. This counterexample shows that modularity is not relatively monotonic. ∎

6. Adaptive Scale Modularity

The problems with modularity stem from the fact that the total volume can change when changes are made to the graph. It is therefore natural to look at a variant of modularity where the total volume is replaced by a constant M,

$$Q_{M\text{-fixed}}(G,C) = \sum_{c \in C} \left(\frac{w_c}{M} - \left(\frac{v_c}{M}\right)^2 \right).$$

This objective is obviously local. It is also a scale invariant family parameterized by M.

We might hope that fixed scale modularity would be monotonic, because it doesn't suffer from the problem where changes in the edge weights affect the total volume. Unfortunately, fixed scale modularity has problems when the volume of a cluster starts to exceed $M/2$. In that case, increasing the weight of within cluster edges starts to decrease the fixed scale modularity. Looking at a cluster c with volume $v_c = w_c + b_c$,

$$\frac{\partial Q_{M\text{-fixed}}(G,C)}{\partial w_c} = \frac{1}{M} - \frac{2v_c}{M^2}. \tag{2}$$

This derivative is negative when $2v_c > M$, so in that case increasing the weight of a within-cluster edge will decrease the objective value. Hence fixed scale modularity is not monotonic.

The above argument also suggests a possible solution: add $2v_c$ to the normalization factor M. Or more generally, add γv_c with $\gamma \geq 2$, which leads to the objective function

$$Q_{M,\gamma}(G,C) = \sum_{c \in C} \left(\frac{w_c}{M + \gamma v_c} - \left(\frac{v_c}{M + \gamma v_c}\right)^2 \right). \tag{3}$$

This *adaptive scale modularity* objective function is clearly still permutation invariant, continuous and local. For every γ it is also a scale invariant family parameterized by M. Additionally, we have the following two theorems:

Theorem 5 *Adaptive scale modularity is rich for all $M \geq 0$ and $\gamma \geq 1$.*

Theorem 6 *Adaptive scale modularity is monotonic for all $M \geq 0$ and $\gamma \geq 2$.*

The proofs of these theorems can be found in appendices B and C.

This shows that adaptive scale modularity satisfies all six axioms we have defined. This shows that our extended set of axioms is consistent.

6.1 Relation to Other Objective Functions

Interestingly, in the limit as M goes to 0, the adaptive-scale objective function becomes similar to normalized cut with an added constant,

$$Q_{0,\gamma}(G, C) = \frac{1}{\gamma} \sum_{c \in C} \Big(\frac{w_c}{v_c} - \frac{1}{\gamma} \Big).$$

This 0-adaptive modularity is also scale invariant as a single objective function.

Conversely, when M goes to infinity the objective value goes to 0. However, the objective function approaches unnormalized cut in behavior:

$$\lim_{M \to \infty} M \cdot Q_{M,\gamma}(G, C) = \sum_{c \in C} w_c.$$

This expression is similar to the Constant Potts model (CPM) by Traag et al. (2011),

$$Q_{\text{cpm}}(G, C) = \sum_{c \in C} \Bigg(w_c - \gamma n_c^2 \Bigg). \tag{4}$$

In contrast to the objective functions discussed thus far, CPM uses the number of nodes instead of volume to control the size of clusters. Like adaptive scale modularity, the constant Potts model satisfies all six axioms.

As stated before, the fixed scale and adaptive scale modularity objective functions are a scale invariant family; they are not scale invariant for a fixed value of M (except for $M = 0$). This is not a large problem in practice, since scale invariance is often sacrificed to overcome the resolution limit of modularity (Fortunato and Barthélemy, 2007). In fact, fixed scale modularity is proportional to the objective function introduced by Reichardt and Bornholdt (2004),

$$Q_{\text{RB}}(G, C) = \sum_{c \in C} \Big(w_c - \gamma_{\text{RB}} \frac{v_c^2}{v_V} \Big) = M \cdot Q_{M\text{-fixed}}(G, C),$$

with $M = v_V / \gamma_{\text{RB}}$.

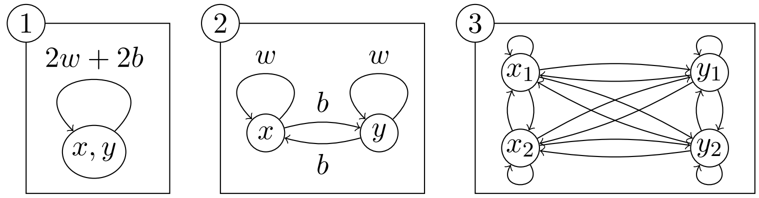


Figure 1: An illustration of the possible outcomes when clustering a two-clique network. Clusters are indicated by circles. In outcome (3), the vertical edges each have weight $w/4$, while the horizontal and diagonal ones have weight $b/4$.

6.2 Resolution Limit Analysis

There has been a lot of interest in the so called resolution limit of modularity. This problem can be illustrated with a simple graph that consists of a ring of cliques, where each clique is connected to the next one with a single edge. We would like the clusters in the optimal clustering to correspond to the cliques in the ring. It was observed by Fortunato and Barthélemy (2007) that, as the number of cliques in the ring increases, at some point the clustering with the highest modularity will have multiple cliques per cluster.

This resolution problem stems from the fact that the behavior of modularity depends on the total volume of the graph. Both the fixed scale and adaptive scale modularity objective functions instead have a parameter M, and hence do not suffer from this problem. In fact, any local objective function will not have a resolution limit in the sense of Fortunato and Barthélemy. A similar observation was made by Traag et al. (2011) in the context of modularity like objective functions.

In real situations graphs are not uniform as in the ring-of-cliques model. But we can still take simple uniform problems as a building block for larger and more complex graphs, since for local objective functions the rest of the network doesn't matter. Therefore we will look at a simple problem with two subgraphs of varying sizes connected by a varying number of edges. More precisely, we take two cliques each with within weight w, connected by edges with weight b. The total volume of this (sub)graph is then $2w + 2b$.

There are three possible outcomes when clustering such a two-clique network: (1) the optimal solution has a single cluster; (2) the optimal solution has two clusters, corresponding to the two cliques; (3) the optimal solution has more than two clusters, splitting the cliques apart. See Figure 1 for an illustration. Which of these outcomes is desirable depends on the circumstances.

Another heterogeneous resolution limit model was proposed by Lancichinetti and Fortunato (2011). In this situation there are two cliques of equal size connected by a single edge, and a random subgraph. Now the ideal solution would be to find three clusters, one for each clique and one for the random subgraph. The optimal split of the random subgraph will roughly cut it in half, with a fixed fraction of the volume being between the two clusters (Reichardt and Bornholdt, 2007). So this model can be considered as a combination of two instances of our simpler problem, one for the two cliques and one for the random subgraph[1]. Hence, we want outcome (2) for the cliques, and outcome (1) for the random subgraph.

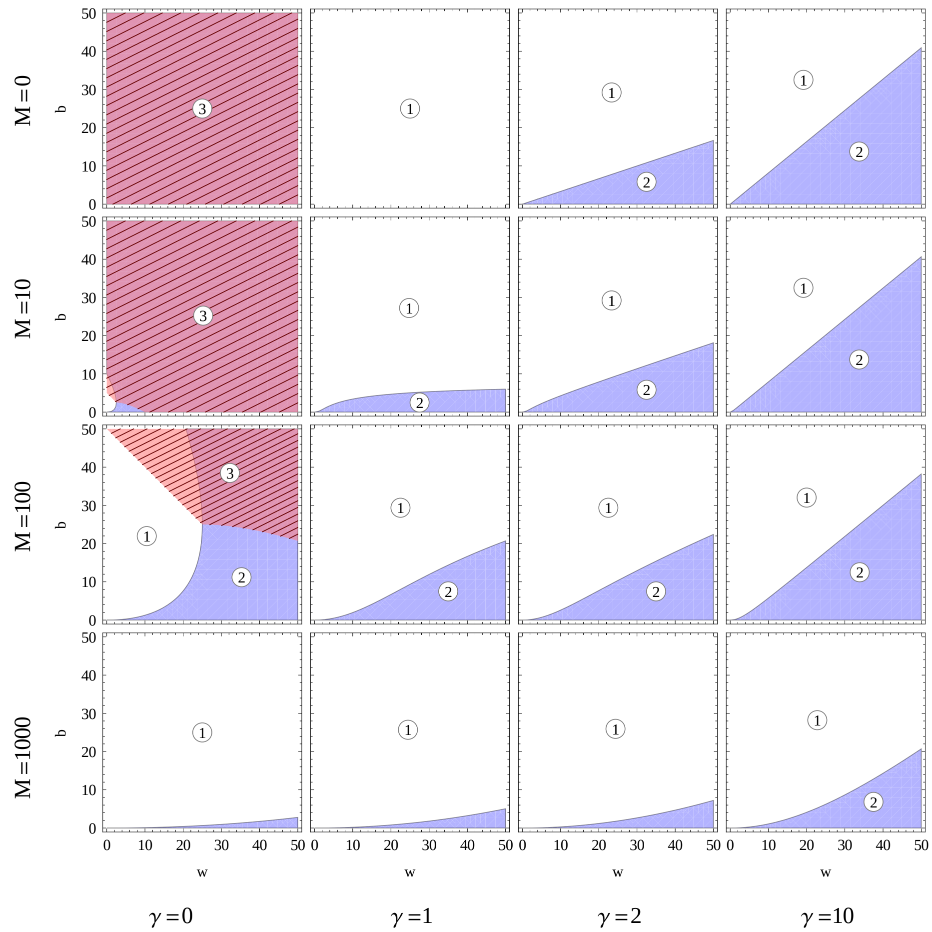


Figure 2: The behavior of $Q_{M,\gamma}$ for varying parameter values. The graph consists of two subgraphs with w internal weight each, connected by an edge with weigh b. Hence the volume of the total graph is $2w + 2b$. In region (1) the optimal clustering has a single cluster, In region (2) (light blue) the optimal clustering separates the subgraphs. In region (3) (red, hatched) the subgraphs themselves will be split apart.

In Figure 2 we show which graphs give which outcomes for adaptive scale modularity with various parameter settings. The first column, $\gamma = 0$, is of particular interest, since it corresponds to fixed scale modularity and hence also to Q_{RB} and to modularity in certain graphs. In the third row we can see that when $2v = 2w + 2b > M = 100$ the cliques are split apart. This is precisely the region in which monotonicity no longer holds. Overall,

1. Lancichinetti and Fortunato includes edges between the cliques and the random subgraph to ensure that the entire network is connected, these edges are not relevant to the problem

the parameter M has the effect of determining the scale; each row in this figure is merely the previous row magnified by a factor 10. Increasing M has the effect of merging small clusters. On the other hand, the γ parameter controls the slope of the boundary between outcomes (1) and (2), i.e. the fraction of edges that should be within a cluster. This is most clearly seen when $M = 0$, while otherwise the effect of M dominates for small clusters.

7. Conclusion and Open Questions

In this paper we presented an axiomatic study of graph clustering objective functions. We showed that modularity does not satisfy the monotonicity axiom. This motivated the derivation of a new parametric objective function, adaptive scale modularity, that satisfies all axioms and has standard graph clustering objectives as special cases. Results of an experimental resolution limit analysis indicated the capability of adaptive scale modularity to model the natural community structure in diverse graph settings. Furthermore we investigated resolution-limit-free objectives as defined in (Traag et al., 2011).

Our paper did not address questions such as finding a best objective function (Almeida et al., 2011), or selecting a significant resolution scale (Traag et al., 2013). The aim was to provide necessary conditions about what a good objective function is, in order to rule out and/or to improve objective functions. The proposed axioms and the introduction of adaptive scale modularity are an effort in this direction.

We also did not address the question of finding a clustering with an optimal objective value. Finding the optimal value of objective functions such as modularity is NP-hard (Brandes et al., 2008), but several heuristic algorithms have been developed. One class of algorithms uses a divisive approach, see for instance Newman (2006b); Ruan and Zhang (2008). For such a tactic to be valid, an optimal or close to optimal clustering of a subgraph should also be a near optimal clustering of the entire graph. This is ensured by locality.

In this work we have only looked at non-negative weights, undirected graphs, and only at hard partitioning. An extension to graphs with negative weights, to directed graphs and to overlapping clusters remains to be investigated. Another open problem is how to use these axioms for reasoning about objectives and clustering algorithms.

Acknowledgments

This work has been partially funded by the Netherlands Organization for Scientific Research (NWO) within the NWO project 612.066.927.

Appendix A. Proof of Theorem 1 (Modularity is Rich)

The proofs of richness rely on clique graphs,

Definition 4 (Clique graph) *Let V be a set of nodes, C be a partition of V, and k be a positive constant. The* clique graph of C with edge weight k *is defined as* $G = (V, E)$ *where* $E(i, j) = k$ *if* $i \sim_C j$ *and* $E(i, j) = 0$ *otherwise.*

Proof

Let V be a set of nodes and $C \neq \{V\}$ be a clustering of V. Let $G = (V, E)$ be a clique graph of C with edge weight 1. Note that $E(i,i) = 1$, so any possible cluster will have a positive volume. Let D be a clustering of G with maximal modularity.

Suppose that there is a cluster $d \in D$ that contains $i, j \in d$ with $i \not\sim_C j$. Then we can split the cluster into $d_1 = \{k \in d \mid k \sim_C i\}$ and $d_2 = \{k \in d \mid k \not\sim_C i\}$. Because there are no edges between nodes in d_1 and nodes in d_2, it is the case that $w_d = w_{d_1} + w_{d_2}$. Both d_1 and d_2 are non-empty and have a positive volume, so $v_d^2 = (v_{d_1} + v_{d_2})^2 < v_{d_1}^2 + v_{d_2}^2$. Therefore $Q_{\text{modularity}}(G, D) < Q_{\text{modularity}}(G, D \setminus \{d\} \cup \{d_1, d_2\})$. So D does not have maximal modularity, which is a contradiction.

Suppose, on the other hand that all clusters $d \in D$ are a subset of some cluster in C, i.e. D is a refinement of C. Then either $D = C$, or there are two clusters $d_1, d_2 \in D$ that are both a subset of the same cluster $c \in C$. In the latter case we can combine the two clusters into $d = d_1 \cup d_2$. The within weight of this combined cluster is $w_d = |d|^2 = w_{d_1} + w_{d_2} + 2|d_1||d_2|$. The squared volume of the combined cluster is $v_d^2 = |d|^2|c|^2 = v_{d_1}^2 + v_{d_2}^2 + 2|d_1||d_2||c|^2$. So this changes increases the modularity by

$$\begin{aligned}
&Q_{\text{modularity}}(G, D \setminus \{d_1, d_2\} \cup \{d\}) - Q_{\text{modularity}}(G, D) \\
&\quad = 2|d_1||d_2|/v_V - 2|d_1||d_2||c|^2/v_V^2 \\
&\quad = 2|d_1||d_2|(v_V - |c|^2)/v_V^2 > 0,
\end{aligned}$$

which contradicts the assumption that D has maximal modularity. Therefore the only optimal clustering of G is C. Note that the above inequality only holds when $|c|^2 = v_c < v_V$, which is the case because $C \neq \{V\}$.

When $C = \{V\}$, a clique graph will not work; because both $\{V\}$ and the clustering that assigns half the nodes to one cluster, and half to another have modularity equal to 0. In this case, instead define $G = (V, E)$ by $E(i, j) = 1$ if $i \neq j$ and 0 if $i = j$. Then the modularity for C is $q(G, \{V\}) = 0$. Any cluster d in a clustering D will have $v_d = |d|(|V|-1)$ and $w_d = |d|(|d| - 1)$. Therefore the contribution of this cluster to the objective value is $-|d|(|V| - |d|)/(|V|^2(|V| - 1))$, which is negative when $|d| < |V|$. So the modularity of any clustering other than $\{V\}$ will be negative, hence $\{V\}$ is the only optimal clustering.

Since for every C we can construct a graph where C is the only optimal clustering, modularity is rich.

■

Appendix B. Proof of Theorem 5 (Adaptive Scale Modularity is Rich)

Denote by $f_C(d)$ the largest fraction of any cluster from C that is contained in a cluster d.

$$f_C(d) = \max_{c \in C} \frac{|c \cap d|}{|c|}.$$

For any clustering D we have that

$$\sum_{d \in D} f_C(d) = \sum_{d \in D} \max_{c \in C} \frac{|c \cap d|}{|c|} \leq \sum_{d \in D} \sum_{c \in C} \frac{|c \cap d|}{|c|} = |C|. \tag{5}$$

And since $f_C(d) \leq 1$ for all clusters d, we also have that

$$\sum_{d \in D} f_C(d) \leq |D|. \tag{6}$$

Lemma 7 *For a clique graph of C it is the case that $w_d/v_d \leq f_C(d)$.*

Proof Given a cluster d and a clique graph G of C with weight $k > 0$, the volume of d is

$$v_d = \sum_{c \in C} k|c \cap d||c|,$$

and the within cluster weight is

$$w_d = \sum_{c \in C} k|c \cap d|^2.$$

Therefore

$$w_d \leq \sum_{c \in C} k|c \cap d||c| f_C(d) = v_d f_C(d).$$

And hence $w_d/v_d \leq f_C(d)$. ■

Lemma 8 *Let G be the clique graph of a clustering C with weight k, and let $0 < \beta < 1$ be a constant. Then $\sum_{d \in D}(w_d/v_d - \beta) = (1-\beta)|C|$ if $D = C$, while $\sum_{d \in D}(w_d/v_d - \beta) < (1-\beta)|C| - \epsilon$ if $D \neq C$, where $\epsilon = \min(\beta, 1-\beta, 1/|V|)/2$.*

Proof Suppose that $D = C$, then for every cluster $c \in C$, $w_c = v_c = k|c|^2$, and so

$$\sum_{c \in C}\Big(\frac{w_d}{v_d} - \beta\Big) = (1-\beta)|C|.$$

Otherwise, $D \neq C$. Assume that $\sum_{d \in D}(w_d/v_d - \beta) \geq (1-\beta)|C| - \min(\beta, 1/|V|)/2$. By Lemma 7,

$$\begin{aligned}
&|C| - \beta(|C|+1) \\
<&|C| - \beta|C| - \epsilon \\
\leq&\sum_{d \in D}(\frac{w_d}{v_d} - \beta) \\
\leq&\sum_{d \in D}(f_C(d) - \beta) \\
\leq&|C| - \beta|D|.
\end{aligned}$$

Since $\beta > 0$, this implies that $|D| < |C| + 1$.

Additionally, since $f_C(d) \leq 1$ for all clusters $d \in D$,

$$\begin{aligned}
&(1-\beta)(|C|-1) \\
<&(1-\beta)|C| - \epsilon \\
\leq&\sum_{d\in D}(f_C(d)-\beta) \\
\leq&(1-\beta)|D|
\end{aligned}$$

Since $\beta < 1$, this implies that $|D| > |C| - 1$. Hence $|D| = |C|$.

Suppose that $f_C(d) < 1$ for some $d \in D$, which implies that $|c \cap d| < |c|$. Because edges are discrete, this can only happen when $|c \cap d| \leq |c| - 1$ for all clusters c. And the size of clusters is bounded by $|c| \leq |V|$. Hence $f_C(d) \leq (|V|-1)/|V| = 1 - 1/|V|$. And since for all other clusters d', $f_C(d') \leq 1$, we then have

$$\begin{aligned}
&\sum_{d\in D}(f_C(d)-\beta) \\
\leq&(1-\beta)|D| - 1/|V| \\
<&(1-\beta)|C| - \epsilon \\
\leq&\sum_{d\in D}(w_d/v_d-\beta) \\
\leq&\sum_{d\in D}(f_C(d)-\beta),
\end{aligned}$$

which is a contradiction. Hence, it must be the case that $f_C(d) = 1$ for all clusters $d \in D$. By the definition of f_C this means that for every d there is a cluster $c \in C$ such that $|c \cap d| = |c|$, and therefore $c \subseteq d$. Since the clusters are disjoint and $|D| = |C|$, this implies that $D = C$. Which is a contradiction, so $\sum_{d\in D}(w_d/v_d - \beta) < (1-\beta)|C| - \epsilon$. ■

When $M = 0$, the adaptive scale modularity reduces to $w_d/(\gamma v_d) - |D|/\gamma^2$, and the above lemma is enough to prove richness. For non-zero values of M, we can get 'close enough' by choosing large enough edge weights. This is formalized in the following lemma.

Lemma 9 *Let d be a cluster in a clustering of a clique graph of C with weight k. Then*

$$\frac{w_d}{v_d} - \beta - \beta M/k \leq q(d)/\beta \leq \frac{w_d}{v_d} - \beta + 2\beta^2 M/k,$$

where

$$q(d) = \frac{w_d}{M + v_d/\beta} - \Big(\frac{v_d}{M + v_d/\beta}\Big)^2$$

denotes the contribution of d to the M-adaptive modularity objective value.

Proof Since clusters are non-empty, and in a clique graph $E(i,i) = k$, it follows that $v_d \geq w_d \geq k$. So

$$\begin{aligned}
&q(d)/\beta \\
=&\frac{\beta M w_d + v_d w_d - \beta v_d^2}{(\beta M + v_d)^2} \\
=&\frac{w_d}{v_d} - \beta + \frac{\beta^2 M(\beta M + 2v_d) - \beta^2 M^2 w_d/v_d - \beta M w_d}{(\beta M + v_d)^2} \\
\leq&\frac{w_d}{v_d} - \beta + \frac{\beta^2 M(\beta M + 2v_d)}{(\beta M + v_d)^2} \\
\leq&\frac{w_d}{v_d} - \beta + \frac{2\beta^2 M(\beta M + 2v_d)}{(\beta M + v_d)(\beta M + 2v_d)} \\
=&\frac{w_d}{v_d} - \beta + \frac{2\beta^2 M}{\beta M + v_d} \\
\leq&\frac{w_d}{v_d} - \beta + \frac{2\beta^2 M}{k}.
\end{aligned}$$

And since $w_d \leq v_d$,

$$\begin{aligned}
&q(d)/\beta \\
=&\frac{w_d}{v_d} - \beta + \frac{\beta^2 M(\beta M + 2v_d) - \beta^2 M^2 w_d/v_d - \beta M w_d}{(\beta M + v_d)^2} \\
\geq&\frac{w_d}{v_d} - \beta - \frac{\beta^2 M^2 + \beta M v_d}{(\beta M + v_d)^2} \\
=&\frac{w_d}{v_d} - \beta - \frac{\beta M}{\beta M + v_d} \\
\geq&\frac{w_d}{v_d} - \beta - \frac{\beta M}{k}.
\end{aligned}$$

■

Combining these lemmas yields the proof of the general theorem:

Proof Given a clustering C. Define $\beta = 1/\gamma$. If $\gamma > 1$ then $0 < \beta < 1$. Pick $k > 3|V|\beta^2 M/\epsilon$ where ϵ is defined as in Lemma 8.

Let G be the clique graph of C with weight k. Let $D \neq C$ be a clustering of G. Then by Lemmas 8 and 9,

$$\begin{aligned}
&Q_{M,\gamma}(G,D)/\beta \\
=&\sum_{d\in D} q(d) \\
\leq&\sum_{d\in D}(w_d/v_d - \beta + 2\beta^3 M/k) \\
\leq&(1-\beta)|C| + 2|D|\beta^3 M/k - \epsilon \\
\leq&(1-\beta)|C| + 2|V|\beta^2 M/k - \epsilon \\
<&(1-\beta)|C| - |V|\beta^2 M/k \\
\leq&(1-\beta)|C| - |C|\beta^2 M/k \\
=&\sum_{c\in C}(w_c/v_c - \beta + \beta^2 M/k) \\
\leq&Q_{M,\gamma}(C)/\beta.
\end{aligned}$$

Hence the objective value is maximal for C. Since there is a clique graph and k for every clustering, adaptive scale modularity is rich. ■

Appendix C. Proof of Theorem 6 (Adaptive Scale Modularity is Monotonic)

Proof

Given a constants $M > 0$ and $\gamma \geq 2$, a graph G and a clustering C of G. Let $c \in C$ be any cluster. Writing the volume of c as $v_c = w_c + b_c$, the contribution of this cluster to the objective value of G is $q(w_c, b_c)$ where

$$q(w,b) = \frac{w}{M+\gamma w+\gamma b} - \Big(\frac{w+b}{M+\gamma w+\gamma b}\Big)^2.$$

The partial derivatives of q are

$$\begin{aligned}
\frac{\partial q(w,b)}{\partial w} &= \frac{M^2 + (\gamma-2)M(w+b) + \gamma b(M+\gamma w+\gamma b)}{(M+\gamma w+\gamma b)^3} \geq 0 \\
\frac{\partial q(w,b)}{\partial b} &= -\frac{\gamma w M + (w+b)(M+\gamma^2 w)}{(M+\gamma w+\gamma b)^3} \leq 0.
\end{aligned}$$

This means that q is a monotonically non-decreasing function in w and a non-increasing function in b.

For any graph G' that is a C-consistent change of G, it holds that $w'_c \geq w_c$ and $b'_c \leq b_c$. So $q(w'_c, b'_c) \geq q(w_c, b_c)$. And therefore $Q_{M,\gamma}(G', C) \geq Q_{M,\gamma}(G, C)$. So adaptive scale modularity is monotonic. ■

References

Margareta Ackerman and Shai Ben-David. Measures of clustering quality: A working set of axioms for clustering. In Daphne Koller, Dale Schuurmans, Yoshua Bengio, and Léon Bottou, editors, *NIPS*, pages 121–128. Curran Associates, Inc., 2008.

Margareta Ackerman and Shai Ben-David. A characterization of linkage-based hierarchical clustering. *Journal of Machine Learning Research*, 2013.

Margareta Ackerman, Shai Ben-David, and David Loker. Towards property-based classification of clustering paradigms. In John D. Lafferty, Christopher K. I. Williams, John Shawe-Taylor, Richard S. Zemel, and Aron Culotta, editors, *NIPS*, pages 10–18. Curran Associates, Inc., 2010.

Helio Almeida, Dorgival Guedes, Wagner Meira Jr., and Mohammed J. Zaki. Is there a best quality metric for graph clusters? In Dimitrios Gunopulos, Thomas Hofmann, Donato Malerba, and Michalis Vazirgiannis, editors, *Machine Learning and Knowledge Discovery in Databases*, volume 6911 of *Lecture Notes in Computer Science*, pages 44–59. Springer Berlin Heidelberg, 2011. ISBN 978-3-642-23779-9.

Vincent D. Blondel, Jean-Loup Guillaume, Renaud Lambiotte, and Etienne Lefebvre. Fast unfolding of communities in large networks. *J. Stat. Mech. Theory Exp.*, 2008(10):P10008, 2008. ISSN 1742-5468. doi: 10.1088/1742-5468/2008/10/P10008. URL `http://dx.doi.org/10.1088/1742-5468/2008/10/P10008`.

Béla Bollobás. *The Evolution of Random Graphs – the Giant Component*, pages 130–159. Cambridge University Press, 2001. ISBN 9780521797221.

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. ISSN 1041-4347. doi: 10.1109/TKDE.2007.190689.

Sébastien Bubeck and Ulrike von Luxburg. Nearest neighbor clustering: A baseline method for consistent clustering with arbitrary objective functions. *J. Mach. Learn. Res.*, 10: 657–698, June 2009. ISSN 1532-4435. URL `http://dl.acm.org/citation.cfm?id=1577069.1577092`.

Gunnar Carlsson, Facundo Mémoli, Alejandro Ribeiro, and Santiago Segarra. Axiomatic construction of hierarchical clustering in asymmetric networks. *CoRR*, abs/1301.7724, 2013.

Santo Fortunato. Community detection in graphs. *Physics Reports*, 486:75–174, 2010. URL `http://arxiv.org/abs/0906.0612`.

Santo Fortunato and Marc Barthélemy. Resolution limit in community detection. *Proc. Natl. Acad. Sci. USA*, 104(1):36–41, 2007. doi: 10.1073/pnas.0605965104.

Sreenivas Gollapudi and Aneesh Sharma. An axiomatic approach for result diversification. In *Proceedings of the 18th international conference on World wide web*, pages 381–390, 2009.

Jon M. Kleinberg. An impossibility theorem for clustering. In Suzanna Becker, Sebastian Thrun, and Klaus Obermayer, editors, *NIPS*, pages 446–453. MIT Press, 2002. ISBN 0-262-02550-7.

Andrea Lancichinetti and Santo Fortunato. Limits of modularity maximization in community detection. *Phys. Rev. E*, 84:066122, December 2011. doi: 10.1103/PhysRevE.84.066122. URL http://dx.doi.org/10.1103/PhysRevE.84.066122.

Marina Meila. Comparing clusterings: an axiomatic view. In *Proceedings of the 22nd international conference on Machine learning*, pages 577–584. ACM, 2005.

Mark E. J. Newman. Modularity and community structure in networks. *Proc. Natl. Acad. Sci. USA*, 103(23):8577–8582, 2006a. doi: 10.1073/pnas.0601602103. URL http://www.pnas.org/content/103/23/8577.abstract.

Mark E. J. Newman. Finding community structure in networks using the eigenvectors of matrices. *Phys. Rev. E*, 74(3):036104, July 2006b. doi: 10.1103/PhysRevE.74.036104. URL http://dx.doi.org/10.1103/PhysRevE.74.036104.

Mark E. J. Newman and Michelle Girvan. Finding and evaluating community structure in networks. *Phys. Rev. E*, 69:026113, Feb 2004. doi: 10.1103/PhysRevE.69.026113. URL http://pre.aps.org/abstract/PRE/v69/i2/e026113.

Jan Puzicha, Thomas Hofmann, and Joachim M. Buhmann. A theory of proximity based clustering: Structure detection by optimization. *Pattern Recognition*, 33:617–634, 1999.

Jörg Reichardt and Stefan Bornholdt. Detecting fuzzy community structures in complex networks with a potts model. *Phys. Rev. Lett.*, 93:218701, 2004. doi: 10.1103/PhysRevLett.93.218701.

Jörg Reichardt and Stefan Bornholdt. Statistical mechanics of community detection. *Physical Review E*, 74(1):016110, 2006.

Jörg Reichardt and Stefan Bornholdt. Partitioning and modularity of graphs with arbitrary degree distribution. *Phys. Rev. E*, 76:015102, Jul 2007. doi: 10.1103/PhysRevE.76.015102. URL http://link.aps.org/doi/10.1103/PhysRevE.76.015102.

Jianhua Ruan and Weixiong Zhang. Identifying network communities with a high resolution. *Phys. Rev. E*, 77:016104, Jan 2008. doi: 10.1103/PhysRevE.77.016104. URL http://link.aps.org/doi/10.1103/PhysRevE.77.016104.

Vincent A. Traag, Paul Van Dooren, and Yurii E. Nesterov. Narrow scope for resolution-limit-free community detection. *Phys. Rev. E*, 84:016114, Jul 2011. doi: 10.1103/PhysRevE.84.016114. URL http://link.aps.org/doi/10.1103/PhysRevE.84.016114.

Vincent A. Traag, Gautier Krings, and Paul Van Dooren. Significant scales in community structure. *Submitted*, Jun 2013. URL http://arxiv.org/abs/1306.3398.

Twan van Laarhoven and Elena Marchiori. Graph clustering with local search optimization: The resolution bias of the objective function matters most. *Phys. Rev. E*, 87:012812, Jan 2013. doi: 10.1103/PhysRevE.87.012812. URL http://link.aps.org/doi/10.1103/PhysRevE.87.012812.

Reza Bosagh Zadeh and Shai Ben-David. A uniqueness theorem for clustering. In *Proceedings of the Twenty-Fifth Conference on Uncertainty in Artificial Intelligence*, UAI '09, pages 639–646, Arlington, Virginia, United States, 2009. AUAI Press. ISBN 978-0-9749039-5-8. URL http://dl.acm.org/citation.cfm?id=1795114.1795189.