IRTG 1792 Discussion Paper 2019-006

Adaptive Nonparametric Community Detection

Larisa Adamyan [*]
Kirill Efimov [*]
Vladimir Spokoiny [* *2]

[*] Humboldt-Universität zu Berlin, Germany
[*2] Weierstrass Institute Berlin, Germany

This research was supported by the Deutsche Forschungsgesellschaft through the International Research Training Group 1792 "High Dimensional Nonstationary Time Series".

http://irtg1792.hu-berlin.de
ISSN 2568-5619

Adaptive Nonparametric Community Detection

Larisa Adamyan ADAMYANL@HU-BERLIN.DE
Humboldt University Berlin
IRTG 1792, Spandauer Str. 1,
10178 Berlin, Germany

Kirill Efimov EFIMOVKQ@HU-BERLIN.DE
Humboldt University Berlin
IRTG 1792, Spandauer Str. 1,
10178 Berlin, Germany

Vladimir Spokoiny SPOKOINY@WIAS-BERLIN.DE
Weierstrass Institute Berlin, Mohrenstr. 39, 10117 Berlin, Germany
IITP RAS, HSE, Skoltech, Moscow

Editor:

Abstract

Understanding the topological structure of real world networks is of huge interest in a variety of fields. One of the way to investigate this structure is to find the groups of densely connected nodes called communities. This paper presents a new non-parametric method of community detection in networks called Adaptive Weights Community Detection. The idea of the algorithm is to associate a *local* community for each node. On every iteration the algorithm tests a hypothesis that two nodes are in the same community by comparing their local communities. The test rejects the hypothesis if the density of edges between these two local communities is lower than the density inside each one. A detailed performance analysis of the method shows its dominance over state-of-the-art methods on well known artificial and real world benchmarks.

Keywords: Adaptive weights, Gap coefficient, Graph clustering, Nonparametric, Overlapping communities

1. Introduction

Discovering the topological structure of real world networks is a fundamental problem which arises in a variety of fields. Networks can be of a very different nature like brain activity, web applications, protein interactions, social networks. One of the way of exploring the structure of networks is finding communities inside. A community is a group of nodes which are more densely connected to each other than to the rest of the network Girvan and Newman (2002) . Implicit definition of a community as well as methods for finding communities vary from field to field. This diversity is natural as far as one can not expect for networks from different fields to obey the same generating process. In real world graphs where nodes represent entities and edges represent the relations between them, communities can relate to groups of entities of the same interest. The modern graphs can consist of up to several millions of nodes which makes it crucial to develop scalable community detection

algorithms. For a comprehensive survey on main approaches we refer to Fortunato (2010), Fortunato and Hric (2016).

Graph partitioning methods aim to divide a network into k communities of predefined size such that the number of edges between communities (so called cut size) is minimal. This group includes Kernighan-Lin algorithm Kernighan and Lin (1970), one of the earliest methods proposed and still frequently applied, and many others, see e.g. Pothen (1997). When there is no preliminary knowledge about the network structure, which is usually the case, the graph partitioning methods requiring fixing the number of communities are non effective.

Often real networks display hierarchical structure, where small communities are included in larger communities. In this case *hierarchical* clustering algorithms can be used, Friedman et al. (2001). Having a defined similarity measure between nodes, these algorithms cluster nodes either iteratively merging clusters with high similarity, or splitting clusters by removing edges connecting nodes with low similarity. The divisive hierarchical algorithm Girvan and Newman, Girvan and Newman (2002); Newman and Girvan (2004), is one of the most popular algorithms in the field of modern community detection. It iteratively removes intercommunity edges until the resulting partition reaches maximum modularity. Intercommunity edges are identified based on high edge betweenness value, which expresses the number of shortest paths between nodes that run along the edge. The algorithm has a complexity $O(n^3)$ on a sparse graph. The quality function *modularity*, introduced by Newman and Girvan, Newman (2006), has became one of the most popular metrics for measuring the goodness of a partition and gave a boost to algorithms based on modularity maximization, e.g. Blondel et al. (2008), Clauset et al. (2004), Noack and Rotta (2009). Modularity is based on the comparison between the number of edges in each community and the number of edges one would expect in the null model, where the edges are randomly distributed, while preserving the same degree sequence as in the original graph.

Louvain is one of the well known modularity optimization methods introduced by Blondel et al. (2008) which is a greedy multiphase algorithm used for extracting community structure of large networks. It starts with assigning to each node in a network its own community, then at the first phase it sequentially moves each node from its community to another one trying to maximize the weighted modularity. At the second phase it merges all nodes from the same community into one supernode yielding a new graph with loops and weighted edges. The weight between new nodes equals to the number of edges in the old graph between corresponding communities. The procedure is then iterated, until modularity computed with respect to the original graph does not grow any more. This method has linear complexity in the number of edges of the graph.

Many clustering algorithms explicitly or implicitly try to optimize some quality measure such as normalized cut, Shi and Malik (2000), or modularity, White and Smyth (2005). The class of *spectral* graph clustering methods partition a network by using the eigenvectors of the similarity matrix or it's derivatives, Von Luxburg (2007). The idea is that eigenvector components corresponding to nodes in the same community should have similar values, if communities are well identified. Two popular algorithms Shi and Malik (2000), Ng et al. (2002) focus on the eigenvectors of the normalized Laplacian. The drawback of spectral methods is that they require a fixed k number of clusters, besides, the computation of the first k eigenvectors of a Laplacian matrix in large graphs is computationally expensive.

Methods based on *statistical inference* attempt to fit a generative model to the actual graph structure based on hypotheses on connectivity properties of nodes. Newman and Leicht (2007) uses Bayesian inference to derive the best fit of a model to the data, expressed by likelihood that is maximized by means of the expectation-maximization method Dempster et al. (1977). Its drawback is the need to specify the number of groups, which is usually unknown.

Rosvall and Bergstrom viewed the problem of finding the best partition of a graph as the problem of optimally compressing the information needed to describe the process of how information flows across the graph. Random walk is chosen as a proxy of information diffusion. The intuitive idea is that a random walker will be "trapped" inside communities for a long time rather than walking between communities. The best graph partition is the one yielding the Minimum Description Length of an infinite random walk, Rissanen (1978), Grünwald et al. (2005). This optimisation of the description length is carried out with a combination of greedy search and simulated annealing. This method known as *Infomap* is frequently used for community detection in large networks and has linear complexity in the number of edges in the network.

In this work we introduce a novel method for detecting communities called Adaptive Weights Community Detection AWCD. The idea of the algorithm is to associate a *local* community for each node. These communities may intersect. Technically, for each node i its local cluster is described in terms of binary weights w_{ij} and includes only nodes j with $w_{ij} = 1$. Thus, the community structure of a graph is described using the matrix of weights W which is recovered from the original graph. The weights are computed by an iterative multiscale procedure. On every iteration the algorithm tests a hypothesis that two nodes are in the same community by comparing their local communities. The test rejects the hypothesis if the density of edges between these two local communities is lower than the density inside each one. The idea originates from the propagation-separation approach introduced in Polzehl and Spokoiny (2006a) for nonparametric smoothing. A similar idea of hypothesis testing is used in the nonparametric clustering technique AWC Efimov et al. (2017). However, an extension to graph clustering is a non-trivial task. Here the construction of the test is entirely different. The main advantage of the proposed procedure is that it is fully adaptive and does not require any preliminary information about the number of communities. We show the performance of AWCD on real world datasets and artificial LFR benchmark, the latter became a standard among all the new algorithms for community detection. The implementation of the method is publicly available on `http://quantlet.de`. More details about the platform can be found in Borke and Härdle (2017).

2. Graph clustering

Consider the graph G with the set of vertices $V = \{v_1, v_2, \ldots, v_n\}$ and the set of edges E. In the following we assume that the graph G is undirected and unweighted. The adjacency matrix of the graph is the matrix $\mathcal{Y} = (Y_{ij})_{i,j=1,\ldots,n}$, where $Y_{ij} = 1$ if there is an edge between the nodes v_i and v_j, otherwise $Y_{ij} = 0$. We set $Y_{ii} = 0$ for all nodes v_i.

2.1 Stochastic block model

Stochastic block model is widely employed as a canonical model to study the statistical and computational properties of random graphs, hence is used in many methods trying to recover the community structure of networks. Consider a stochastic graph model

$$Y_{ij} \sim \text{Bernoulli}(\theta_{ij}) .$$

Here θ_{ij} are the underlying edge probabilities. A block model assumes a piecewise constant structure of these probabilities. Namely, let the set of vertices V is split into non-overlapping communities $\mathcal{C}_1, \ldots, \mathcal{C}_M$. The block structure means that the probabilities θ_{ij} only depend on the community of the indices i and j. Namely, assume that for some set $(\theta_{\xi\eta})$, it holds

$$Y_{ij} \sim \text{Bernoulli}(\theta_{\xi\eta}), \quad i \in \mathcal{C}_\xi , j \in \mathcal{C}_\eta , i \neq j.$$

Also set $Y_{ii} \equiv 0$ for all i. Block structure implicitly assumes that the diagonal probabilities $\theta_{\xi\xi}$ are significantly larger than the cross diagonal entries $\theta_{\xi\eta}$ for $\xi \neq \eta$.

2.2 Weighted communities

A block structure is very transparent and appealing but can be too restrictive. In particular, it does not allow to model overlapping communities. An extension can be given by the notion of weighted communities. Let Ξ be the index set for all considered communities, $\Xi = \{1, \ldots, n\}$. In what follows we suppose that the community structure is described by the binary matrix W with entries $w_{\xi\eta}$, $\xi, \eta \in \Xi$. By technical reasons, we set $w_{\xi\xi} \equiv 0$ for all ξ. For each ξ, define the set of its friends $\mathcal{C}_\xi$ which consists of all other nodes η with $w_{\xi\eta} = 1$:

$$\mathcal{C}_\xi = \{\eta \colon w_{\xi\eta} = 1\} .$$

In the case of non-overlapping communities, the matrix W has a block structure (except the diagonal terms). In general, we allow for overlapping communities and they do overlap significantly. Our basic assumption is that the edge probability in our graph is determined by the community structure. Namely, for each $\xi, \eta \in \Xi$, let $\theta^*_{\xi\eta}$ be the related parameter, that is,

$$I\!P(Y_{ij} = 1) = \theta^*_{\xi\eta} \quad i \in \mathcal{C}_\xi , j \in \mathcal{C}_\eta, \quad i \neq j.$$

By N_ξ we denote the volume of $\mathcal{C}_\xi$:

$$N_\xi \stackrel{\text{def}}{=} \sum_\eta w_{\xi\eta} . \tag{1}$$

Also denote by $N_{\xi\eta}$ the number of possible connections between $\mathcal{C}_\xi$ (friends of ξ) and $\mathcal{C}_\eta$ (friends of η):

$$N_{\xi\eta} \stackrel{\text{def}}{=} N_\xi N_\eta .$$

Apart of these quantities consider the one-neighbourhood $\boldsymbol{u}_\xi$ of ξ, which is given by a vector $\boldsymbol{u}_\xi = (u_{\xi\eta})$ describing for each $\eta \in \Xi$, the number of connections between η and all friends of ξ :

$$\boldsymbol{u}_\xi^\top \stackrel{\text{def}}{=} \boldsymbol{w}_\xi^\top \mathcal{Y} = \Big(\sum_\zeta w_{\xi\zeta} Y_{\zeta\eta} \Big)_{\eta \in \Xi} .$$

The community η is in the one-vicinity of ξ, if $u_{\xi\eta} > 0$. By $\mathcal{U}$ we denote the matrix with the rows $\boldsymbol{u}_\xi^\top$, $\xi \in \Xi$. The number of edges connecting $\mathcal{C}_\xi$ and $\mathcal{C}_\eta$ can be written as

$$S_{\xi\eta} \stackrel{\text{def}}{=} S(\mathcal{C}_\xi, \mathcal{C}_\eta) = \boldsymbol{w}_\xi^\top \mathcal{Y} \boldsymbol{w}_\eta = \boldsymbol{u}_\xi^\top \boldsymbol{w}_\eta .$$

Then we can write the matrix $S = (S_{\xi\eta})$ as

$$S = \mathcal{U} W = W \mathcal{Y} W.$$

The proposed procedure consequently repeats two steps: given the community structure in terms of the matrix of weights W, one-vicinity matrix $\mathcal{U}$, and the values N_ξ, it recomputes the estimates $\widetilde{\theta}_{\xi\eta}$ of the parameters $\theta^*_{\xi\eta}$:

$$\widetilde{\theta}_{\xi\eta} \stackrel{\text{def}}{=} \operatorname*{argmax}_\theta L_{\xi\eta}(\theta) = \frac{S_{\xi\eta}}{N_{\xi\eta}} .$$

Further, with the given family of estimates $\widetilde{\theta}_{\xi\eta}$, it reconsiders the whole community structure by the mean of statistical tests on homogeneity and gap. These statistical tests are discussed further in details in the section 5.

For each ξ, η, consider the testing problem whether there is a significant difference between the parameters $\theta^*_{\xi\xi}$, $\theta^*_{\eta\eta}$, and $\theta^*_{\xi\eta}$, or one can combine two communities into one with a joint parameter θ^*. This leads to the formulation of the null hypothesis:

$$H_0 : \theta_{\xi\xi} = \theta_{\eta\eta} = \theta_{\xi\eta}$$
$$H_1 : \mathit{alternative}$$

Let $\xi \vee \eta$ denote the union of two communities.

The null hypothesis leads to the maximum likelihood estimation (MLE):

$$\widetilde{\theta}_{\xi\vee\eta} \stackrel{\text{def}}{=} \operatorname*{argmax}_\theta \big\{ L_{\xi\xi}(\theta) + L_{\eta\eta}(\theta) + L_{\xi\eta}(\theta) \big\} = \frac{S_{\xi\xi} + S_{\eta\eta} + S_{\xi\eta}}{N_{\xi\xi} + N_{\eta\eta} + N_{\xi\eta}} ,$$

and the likelihood ratio (LR) test statistic reads as

$$T_{\xi\eta} \stackrel{\text{def}}{=} \sup_{\theta_{\xi\xi}, \theta_{\eta\eta}, \theta_{\xi\eta}} \big\{ L_{\xi\xi}(\theta_{\xi\xi}) + L_{\eta\eta}(\theta_{\eta\eta}) + L_{\xi\eta}(\theta_{\xi\eta}) \big\} - \sup_\theta \big\{ L_{\xi\xi}(\theta) + L_{\eta\eta}(\theta) + L_{\xi\eta}(\theta) \big\}.$$

The Lemma 2 in the section 5.1 shows that

$$T_{\xi\eta} = N_{\xi\xi} \mathcal{K}(\widetilde{\theta}_{\xi\xi}, \widetilde{\theta}_{\xi\vee\eta}) + N_{\eta\eta} \mathcal{K}(\widetilde{\theta}_{\eta\eta}, \widetilde{\theta}_{\xi\vee\eta}) + N_{\xi\eta} \mathcal{K}(\widetilde{\theta}_{\xi\eta}, \widetilde{\theta}_{\xi\vee\eta}) \tag{2}$$

where $\mathcal{K}(\theta_1, \theta_2)$ is the Kullback-Leibler divergence between two Bernoulli distributions with parameters θ_1 and θ_2:

$$\mathcal{K}(\theta_1, \theta_2) = \theta_1 \log \frac{\theta_1}{\theta_2} + (1 - \theta_1) \log \frac{1 - \theta_1}{1 - \theta_2}.$$

The null hypothesis has to be rejected if this quantity exceeds some critical value λ: $T_{\xi\eta} > \lambda$. The Wilks phenomenon Fan et al. (2001) claims that $T_{\xi\eta}$ is nearly χ^2_2 under the null hypothesis $\theta^*_{\xi\xi} = \theta^*_{\eta\eta} = \theta^*_{\xi\eta}$ for large samples $N_{\xi\xi}$, $N_{\eta\eta}$, and $N_{\xi\eta}$. This gives some hints about the choice of the hyperparameter λ.

3. Procedure

Initialization The starting point of the algorithm is the trivial community structure: one community ξ per point, where the friends of ξ are given by the adjacency matrix. This means that $W^{(0)} = \mathcal{Y}$. The size $N^{(0)}_\xi$ of the starting community ξ is equal to the degree of this node. The one-vicinity matrix is $\mathcal{U}^{(0)} = W^{(0)}\mathcal{Y} = \mathcal{Y}^2$. The matrix $S^{(0)} = \mathcal{U}^{(0)}W^{(0)} = \mathcal{Y}^3$ describes the paths of length 3 from each node.

The further step of the procedure requires to compute the weights $w^{(k+1)}_{\xi\eta}$ for any ξ and any η from its one-vicinity.

Update of the community structure The procedure subsequently updates the whole community structure at each step starting from the initialization. Suppose that the community structure is fixed after k steps of the algorithm in terms of the symmetric matrix $W^{(k)}$ of weights $w^{(k)}_{\xi\eta}$ for each pair $\xi, \eta \in \Xi$. At the step $k+1$ we recompute all the weights $w^{(k)}_{\xi\eta}$ for η from the one-vicinity of ξ given by $\boldsymbol{u}^{(k)}_\xi$ which are columns of the one-vicinity matrix $\mathcal{U}^{(k)} = W^{(k)}\mathcal{Y}$.

As before, for each ξ, let $N^{(k)}_\xi$ stand for the number of friends of $\xi \in \Xi$ at step k; see (1). Now we set $w^{(k+1)}_{\xi\xi} = 0$ and for each η with $\boldsymbol{u}^{(k)}_{\xi\eta} > 0$, we compute the new weight $w^{(k+1)}_{\xi\eta}$ and update the other elements of the community structure as follows. Define $N^{(k)}_{\xi\eta} = N^{(k)}_\xi N^{(k)}_\eta$. Further, set

$$S^{(k)} = \mathcal{U}^{(k)}W^{(k)} = W^{(k)}\,\mathcal{Y}\,W^{(k)} \tag{3}$$

which means

$$S^{(k)}_{\xi\eta} = \boldsymbol{w}^{(k)\top}_\xi \mathcal{Y}\, \boldsymbol{w}^{(k)}_\eta = \boldsymbol{u}^{(k)\top}_\xi \boldsymbol{w}^{(k)}_\eta = \sum_\zeta u^{(k)}_{\xi\zeta} w^{(k)}_{\eta\zeta}, \qquad \xi, \eta \in \Xi.$$

Only nodes ζ with positive values $u^{(k)}_{\xi\zeta}$ or $w^{(k)}_{\eta\zeta}$ are involved in this computation. The new estimates of the edge density parameters are

$$\widetilde{\theta}^{(k)}_{\xi\eta} = \frac{S^{(k)}_{\xi\eta}}{N^{(k)}_{\xi\eta}},$$

$$\widetilde{\theta}^{(k)}_{\xi\vee\eta} = \frac{S^{(k)}_{\xi\xi} + S^{(k)}_{\eta\eta} + S^{(k)}_{\xi\eta}}{N^{(k)}_{\xi\xi} + N^{(k)}_{\eta\eta} + N^{(k)}_{\xi\eta}}.$$

The test statistics $T^{(k)}_{\xi\eta}$ is given by (2). This requires only few arithmetic operations for each pair ξ, η. The new weight $w^{(k+1)}_{\xi\eta}$ is the test which compares the test statistics with the corresponding threshold parameter λ. For each pair ξ, η with $u^{(k)}_{\xi\eta} \neq 0$, update

$$w^{(k+1)}_{\xi\eta} = I\!I(T^{(k)}_{\xi\eta} < \lambda).$$

Hence, for each pair $\xi, \eta \in \Xi^{(k)}$ we only need a finite number of operations for computing $w^{(k+1)}_{\xi\eta}$, and only communities η from one-vicinity of ξ have to be rechecked.

Without a statistical test, the procedure would replace the community $\mathcal{C}^{(k)}_{\xi}$ by its one-vicinity at each step k. A statistical test makes this extension in a more sophisticated way with the aim to restrict such an expansion only to the true underlying community.

Tuning the parameter λ The procedure has only one parameter λ, which influences its performance. Large λ values result in a conservative gap test allowing to skip small gaps between groups. This leads to aggregation of such groups and may result a partition with big communities. In the contrary, small λ values increase the sensitivity of the method towards small density between local communities and may lead to artificial segmentation of a graph. We propose to fix the parameter λ which maximises the modularity.

3.1 Complexity

Initialization. A sparse structure of the graph with a bounded degree $m_i \leq \overline{m}$ for each node i yields the same row sparsity $\overline{m}$ of $\mathcal{Y}$ and allows to compute each value $N^{(0)}_i$ by at most $\overline{m}$ operations. Moreover, each one-vicinity vector $\boldsymbol{u}^{(0)}_i$ describes all paths of length 2 from i and is also computable in finite time $\overline{m}^2$. The one-vicinity matrix $\mathcal{U}^{(0)} = \mathcal{Y}^2$ can be computed in $n\overline{m}^2$ time and it can still be considered as sparse, through the sparsity $\overline{m}^2$ is much less prominent compared to the original adjacency matrix $\mathcal{Y}$. Computing the matrix $S^{(0)} = \mathcal{U}^{(0)}W^{(0)} = \mathcal{Y}^3$ requires $n\overline{m}^3$ time.

Complexity of the step k can be easily bounded via the sparsity $\overline{m}$ of the network and the max-volume $\overline{N}^{(k)} = \max_\xi N^{(k)}_\xi$ of the communities at step k. The most computationally intensive part of the procedure is to compute the matrix $S^{(k)}$ from (3). This can be done in $n\,\overline{m}(\overline{N}^{(k)})^2$ time. One can see that the procedure remains scalable provided that the size of each community does not grow over a certain value.

Note that for real world networks, the "small world phenomenon" yields that the edge matrix $\mathcal{Y}^k$ becomes dense for k larger than 4 or 5. In other words, each node can find

a path to almost any other node in the network by crossing 4-5 edges. If $W^{(k)}$ inherits such a dense structure, the procedure can become computationally infeasible. To prevent communities from uncontrolled size growth, the procedure checks on a possible gap between any two communities. This allows to control the complexity of the method as well.

4. Propagation effect

Here we discuss the propagation property of the AWCD algorithm in the homogeneous case of just one community. Let the probability of an edge in the graph be the same for all pairs (i, j):

$$\mathbb{P}\big(Y_{ij} = 1\big) = \theta^*. \tag{4}$$

We show that the choice of the parameter λ as $\lambda = \mathtt{C} \log n$ for a specific absolute constant $\mathtt{C}$ ensures the propagation property, which means a small probability of artificial partitioning of a homogeneous region.

Theorem 1 *Suppose* (4). *If* $\lambda \geq 4 \log n$, *then* $\mathbb{P}\big(w_{ij}^{(k)} = 0\big) \leq 1/n$.

Proof See Appendix C 6.5. ■

5. Two communities and statistical tests

Consider a stochastic block model with edges Y_{ij} and two communities $\mathcal{C}_1$ and $\mathcal{C}_2$ given by weights w_{ij}. The value $w_{ij} = 1$ means that i, j are in the same block, $w_{ij} = 0$ for the opposite event. Let also θ_{11} and θ_{22} denote the edge probability within each community while θ_{12} be the edge probability between communities, that is,

$$\mathbb{P}(Y_{ij} = 1) = \theta_{\xi\xi} \quad \text{if } i, j \in \mathcal{C}_\xi \qquad \mathbb{P}(Y_{ij} = 1) = \theta_{12} \quad \text{if } i \in \mathcal{C}_\xi, j \notin \mathcal{C}_\xi$$

for $\xi = 1, 2$.

Split the data Y_{ij} into three groups:

$$\mathcal{Y}_{11} \stackrel{\text{def}}{=} \big\{Y_{ij} \colon i, j \in \mathcal{C}_1\big\},$$

$$\mathcal{Y}_{22} \stackrel{\text{def}}{=} \big\{Y_{ij} \colon i, j \in \mathcal{C}_2\big\},$$

$$\mathcal{Y}_{12} \stackrel{\text{def}}{=} \big\{Y_{ij} \colon i \in \mathcal{C}_1, j \in \mathcal{C}_2\big\}.$$

Note that each edge in $\mathcal{Y}_{11}$ and $\mathcal{Y}_{22}$ is counted twice, this is useful especially for overlapping communities.

5.1 LR test of homogeneity

The parameters $\boldsymbol{\theta} = (\theta_{11}, \theta_{22}, \theta_{12})$ can be estimated by the maximum likelihood method:

$$\widetilde{\boldsymbol{\theta}} = \operatorname*{argmax}_{\boldsymbol{\theta}} L(\boldsymbol{\theta}) = \big(\widetilde{\theta}_{11}, \widetilde{\theta}_{22}, \widetilde{\theta}_{12}\big)$$

with

$$\widetilde{\theta}_{\xi\eta} \stackrel{\text{def}}{=} \operatorname*{argmax}_{\boldsymbol{\theta}} L_{\xi\eta}(\theta_{\xi\eta}) = \operatorname*{argmax}_{\theta_{\xi\eta}} \sum_{i,j \in \mathcal{Y}_{\xi\eta}} \ell(Y_{ij}, \theta_{\xi\eta})$$

for $\xi, \eta = 1, 2$. Here $\ell(y, \theta)$ is the log-likelihood function for the Bernoulli family:

$$\ell(y, \theta) = y \log(\theta) + (1 - y) \log(1 - \theta).$$

It obviously holds

$$\widetilde{\theta}_{\xi\eta} = \frac{S_{\xi\eta}}{N_{\xi\eta}}$$

with

$$N_{\xi\eta} \stackrel{\text{def}}{=} \sum_{i,j \in \mathcal{Y}_{\xi\eta}} 1, \qquad S_{\xi\eta} \stackrel{\text{def}}{=} \sum_{i,j \in \mathcal{Y}_{\xi\eta}} Y_{ij}\,, \quad \xi, \eta = 1, 2.$$

Now consider the null hypothesis that two communities $\mathcal{C}_{11}$ and $\mathcal{C}_{22}$ can be joined into one, that is $\theta_{11} = \theta_{22} = \theta_{12}$. We will denote this value by $\theta_{1\vee 2}$. The estimator of $\theta_{1\vee 2}$ under the null reads as

$$\begin{aligned} \widetilde{\theta}_{1\vee 2} &\stackrel{\text{def}}{=} \operatorname*{argmax}_{\theta} \bigl\{ L_{11}(\theta) + L_{22}(\theta) + L_{12}(\theta) \bigr\} \\ &= \frac{1}{N_{11} + N_{22} + N_{12}} \sum_{i,j} Y_{ij} = \frac{S_{11} + S_{22} + S_{12}}{N_{11} + N_{22} + N_{12}}\,. \end{aligned}$$

Note that we double the inner edges in $\mathcal{C}_1$ and $\mathcal{C}_2$. This does not affect the values $\widetilde{\theta}_{11}$ and $\widetilde{\theta}_{22}$, however it matters for $\widetilde{\theta}_{1\vee 2}$. In fact, this doubling balances the impact of inner connections from $\mathcal{Y}_{11} \cup \mathcal{Y}_{22}$ and inter connections from $\mathcal{Y}_{12}$. The population counterpart of $\widetilde{\theta}_{1\vee 2}$ is

$$\theta^*_{1\vee 2} \stackrel{\text{def}}{=} \frac{1}{N_{11} + N_{22} + N_{12}} \sum_{i,j} I\!\!E Y_{ij} = \frac{N_{11}\theta^*_{11} + N_{22}\theta^*_{22} + N_{12}\theta^*_{12}}{N_{11} + N_{22} + N_{12}}$$

Now let us introduce the likelihood ratio test of the null hypothesis against the alternative that the communities are different. The related likelihood ratio test statistic reads as:

$$T_{12} \stackrel{\text{def}}{=} \sup_{\theta_{11}, \theta_{22}, \theta_{12}} \bigl\{ L_{11}(\theta_{11}) + L_{22}(\theta_{22}) + L_{12}(\theta_{12}) \bigr\} - \sup_{\theta} \bigl\{ L_{11}(\theta) + L_{22}(\theta) + L_{12}(\theta) \bigr\}.$$

Also define the statistical distance between the null and the alternative set:

$$\overline{T}_{12} \stackrel{\text{def}}{=} \sup_{\theta_{11}, \theta_{22}, \theta_{12}} I\!\!E \bigl\{ L_{11}(\theta_{11}) + L_{22}(\theta_{22}) + L_{12}(\theta_{12}) \bigr\} - \sup_{\theta} I\!\!E \bigl\{ L_{11}(\theta) + L_{22}(\theta) + L_{12}(\theta) \bigr\}.$$

Lemma 2 *It holds*

$$T_{12} = N_{11}\mathcal{K}(\widetilde{\theta}_{11}, \widetilde{\theta}_{1\vee 2}) + N_{22}\mathcal{K}(\widetilde{\theta}_{22}, \widetilde{\theta}_{1\vee 2}) + N_{12}\mathcal{K}(\widetilde{\theta}_{12}, \widetilde{\theta}_{1\vee 2}).$$
$$\overline{T}_{12} = N_{11}\mathcal{K}(\theta^*_{11}, \theta^*_{1\vee 2}) + N_{22}\mathcal{K}(\theta^*_{22}, \theta^*_{1\vee 2}) + N_{12}\mathcal{K}(\theta^*_{12}, \theta^*_{1\vee 2}).$$

where $\theta^*_{\xi\eta}$ *are the true values and* $\mathcal{K}(\theta_1, \theta_2)$ *is the Kullback-Leibler divergence between two Bernoulli distributions with parameters* θ_1 *and* θ_2 *:*

$$\mathcal{K}(\theta_1, \theta_2) = \theta_1 \log \frac{\theta_1}{\theta_2} + (1-\theta_1) \log \frac{1-\theta_1}{1-\theta_2}\,.$$

Proof See Polzehl and Spokoiny (2006b). ■

Moreover, the Wilks phenomenon Fan et al. (2001) claims that under the null hypothesis $\theta^*_{11} = \theta^*_{22} = \theta^*_{12}$, the T_{12} is nearly χ^2 with two degrees of freedom for large samples N_{11}, N_{22}, and N_{12}.

5.2 Test on similarity

Below we differentiate two departures from the null hypothesis. The first one compares the intercommunity parameters θ_{11} and θ_{22}. If they are significantly different, the null is rejected. The corresponding test is entirely based on $\mathcal{Y}_{11}$ and $\mathcal{Y}_{22}$. Similarly to Polzehl and Spokoiny (2006b), the LR-test statistic reads as

$$\mathbb{D}_{12} = N_{11}\mathcal{K}\big(\widetilde{\theta}_{11}, \widetilde{\theta}_{1+2}\big) + N_{22}\mathcal{K}\big(\widetilde{\theta}_{22}, \widetilde{\theta}_{1+2}\big).$$

where

$$\widetilde{\theta}_{1+2} \stackrel{\text{def}}{=} \frac{S_{11} + S_{22}}{N_{11} + N_{22}}\,.$$

This statistic is sensitive provided that the value

$$\mathbb{D}^*_{12} \stackrel{\text{def}}{=} N_{11}\mathcal{K}\big(\theta^*_{11}, \theta^*_{1+2}\big) + N_{22}\mathcal{K}\big(\theta^*_{22}, \theta^*_{1+2}\big)$$

with

$$\theta^*_{1+2} \stackrel{\text{def}}{=} \frac{N_{11}\theta^*_{11} + N_{22}\theta^*_{22}}{N_{11} + N_{22}}$$

is sufficiently large; Polzehl and Spokoiny (2006b).

5.3 Test on a gap

Now we consider a different situation when the communities are similar, that is, the value $\mathbb{D}^*$ is not sufficiently large. Then we assume that $\theta^*_{11} = \theta^*_{22}$ and introduce the gap coefficient which measures the ratio between θ^*_{11} and θ^*_{12}. The gap means that θ^*_{12} is significantly

smaller than θ^*_{11}. To quantify the notion of significance, we consider the statistical likelihood ratio test of "gap" between two local communities. Consider a one sided test with a composite null $H_0 : \theta^*_{12} < \theta^*_{1+2}$.

Then the corresponding test statistic reads as

$$\mathbb{G}_{12} \stackrel{\text{def}}{=} \Big\{ N_{1+2}\mathcal{K}\big(\widetilde{\theta}_{1+2}, \widetilde{\theta}_{1\vee 2}\big) + N_{12}\mathcal{K}\big(\widetilde{\theta}_{12}, \widetilde{\theta}_{1\vee 2}\big) \Big\} \times \Big\{ \mathbb{I}\big(\widetilde{\theta}_{1+2} > \widetilde{\theta}_{12}\big) - \mathbb{I}\big(\widetilde{\theta}_{1+2} < \widetilde{\theta}_{12}\big) \Big\},$$

where

$$N_{1+2} \stackrel{\text{def}}{=} N_{11} + N_{22}, \quad \widetilde{\theta}_{1+2} \stackrel{\text{def}}{=} \frac{S_{11} + S_{22}}{N_{11} + N_{22}}, \quad \widetilde{\theta}_{1\vee 2} \stackrel{\text{def}}{=} \frac{S_{11} + S_{22} + S_{12}}{N_{11} + N_{22} + N_{12}}.$$

See Lemma 4 in the Appendix B for the proof.

Again, the test is sensitive if the following value is sufficiently large:

$$\mathbb{G}^*_{12} \stackrel{\text{def}}{=} \big\{ N_{1+2}\mathcal{K}\big(\theta^*_{1+2}, \theta^*_{1\vee 2}\big) + N_{12}\mathcal{K}\big(\theta^*_{12}, \theta^*_{1\vee 2}\big) \big\}\, \mathbb{I}\big(\theta^*_{1+2} > \theta^*_{12}\big).$$

with

$$\theta^*_{1+2} \stackrel{\text{def}}{=} \frac{N_{11}\theta^*_{11} + N_{12}\theta^*_{22}}{N_{11} + N_{12}}, \quad \theta^*_{1\vee 2} \stackrel{\text{def}}{=} \frac{N_{11}\theta^*_{11} + N_{22}\theta^*_{22} + N_{12}\theta^*_{12}}{N_{11} + N_{22} + N_{12}}.$$

5.4 Alternative test on a gap

The aforementioned test works very well on classical artificial benchmarks and is easier for theoretical development. Here we consider an alternative approach to test the gap between communities to face some specific situations which arise in real world graphs. Suppose for communities $\mathcal{C}_1, \mathcal{C}_2$ we have the next situation: $\theta^*_{11} < \theta^*_{12} < \theta^*_{22}$. This means that the community $\mathcal{C}_1$ is more connected to the community $\mathcal{C}_2$ than its own nodes are connected to each other. In case of social networks, the $\mathcal{C}_2$ would consist of the popular members and $\mathcal{C}_1$ would consist of their followers or fans. It is arguable what the algorithm should do in this special case: separate $\mathcal{C}_1$ and $\mathcal{C}_2$ or join them. We believe this depends on the nature of the network and in many scenarios we will want to merge them. Hence, we consider a "gap" between communities, when the observed density of edges between communities is lower than any density inside communities. We separate $\mathcal{C}_1$ and $\mathcal{C}_2$ if we observe a gap between them. The reformulated null hypothesis in this case has the form:

$$H_0 : \theta^*_{12} < \min(\theta^*_1, \theta^*_2) \tag{5}$$

Lemma 3 *The likelihood ratio test statistics T for the hypothesis* (5)*, has the form*

$$T = \min(T_A, T_B)$$
$$\textit{if } \widetilde{\theta}_{12} < \min(\widetilde{\theta}_1, \widetilde{\theta}_2), \ \textit{otherwise } T \le 0,$$

where

$$T_A \stackrel{\text{def}}{=} N_1\mathcal{K}(\widetilde{\theta}_1, \widetilde{\theta}_{1\vee 12}) + N_{12}\mathcal{K}(\widetilde{\theta}_{12}, \widetilde{\theta}_{1\vee 12}),$$
$$T_B \stackrel{\text{def}}{=} N_2\mathcal{K}(\widetilde{\theta}_2, \widetilde{\theta}_{2\vee 12}) + N_{12}\mathcal{K}(\widetilde{\theta}_{12}, \widetilde{\theta}_{2\vee 12}),$$

Proof See Appendix A 6.5. ■

6. Evaluation

In this section we investigate the performance of AWCD algorithm on artificial and real world data. The standard benchmarks considered in the literature are Stochastic Block Model (SBM) and a more sophisticated LFR benchmark first proposed by Lancichinetti et al. (2008). The latter is intended to mimic properties of real world graphs such as heterogeneity in the distributions of node degree and community size. SBM is a pretty simple set up which makes it possible to develop theoretical results. The evaluation of community detection algorithms is difficult for real world graphs because there are only few data sets available with a known true community structure. To be more precise, for real world graph we know only meta information such as node attributes. Considering the graph partition based on this meta information as true community structure is questionable, see Peel et al. (2017), Hric et al. (2014).

We compare the performance of AWCD with the most popular community detection algorithms Infomap by Rosvall and Bergstrom (2008) based on map equations and greedy modularity optimization method Louvain by Blondel et al. (2008), both featured in comparative analysis of many methods on SBM/LFR benchmarks and on real world datasets. See the overview of the comparison between various algorithms in Hric et al. (2014), Yang et al. (2016a), Yang et al. (2016b), Lancichinetti and Fortunato (2009). Lancichinetti and Fortunato (2009) showed that Infomap outperforms many other community detection algorithms exactly on LFR benchmark, allowing to consider the Infomap as a state-of-the-art algorithm for this benchmark. The implementation and detailed description of Infomap is available on the website `www.mapequation.org`. For the Louvain algorithm we used the *python-louvain* python package from `https://python-louvain.readthedocs.io`.

6.1 AWCD setup

Here we fix the setup of the algorithm showing the best results among aforementioned versions of test statistics. We briefly describe this setup with notations coming from the previous section.

0. Initialization: $W^{(0)} = \mathcal{Y}$, $k = 0$
1. $\mathcal{U}^{(k)} = W^{(k)}\mathcal{Y}$
2. For each pair i, j such that $u_{ij}^{(k)} \neq 0$ update w_{ij} of matrix $W^{(k)}$,

$$w_{ij}^{(k+1)} = I\!\!I(T_{ij}^{(k)} < \lambda).$$

where T_{ij} is the classical gap test statistics from section 5.3 for the synthetic data and alternative gap test statistics from section 5.4 for real world data.

The implementation of the AWCD method including a fast version for sparse graphs is publicly available on `https://github.com/QuantLet/awcd`.

6.2 Evaluation criteria

Modularity is the standard evaluation criteria for community detection methods first proposed by Newman (2006). The main idea behind modularity function is that in a good graph division most of the edges lie inside the groups and only a few of them between groups. Modularity doesn't require knowledge about ground truth and can be optimized directly like in the popular Louvain algorithm Blondel et al. (2008). Newman (2016) showed that modularity maximization is equivalent to the maximum likelihood approach applied to a special case of the stochastic block model.

Consider a graph G with the adjacency matrix $\mathcal{Y} = (Y_{ij})_{i,j=1,\dots,n}$, where $Y_{ij} \in \{0,1\}$. Suppose the set of vertices is divided into non-overlapping communities $\mathcal{C}_1, \dots, \mathcal{C}_M$, and θ_i is the index of the community containing the vertex i. The most commonly used definition of modularity is:

$$Q = \frac{1}{2m} \sum_{ij} \left(Y_{ij} - \frac{k_i k_j}{2m} \right) \delta_{\theta_i \theta_j},$$

where m is the number of edges in the graph, k_i, k_j are the degrees of nodes i, j and δ is the Kronecker delta.

When the true community structure is known one can directly measure the similarity between the detected community structure and the true partition. This can be done in many ways. One of the popular measures is *Normalized Mutual Information* NMI Danon et al. (2005). Suppose the true community structure is $C^* = \{C_m^*\}_{m=1}^M$ and the estimated community structure is $C = \{C_l\}_{l=1}^L$. Then NMI is defined as:

$$NMI(\mathcal{C}, \mathcal{C}^*) = \frac{2I(C, C^*)}{H(C) + H(C^*)} = \frac{\sum_{ml} n_{ml} \log \frac{n\, n_{ml}}{n_m^*\, n_l}}{\sqrt{\sum_m n_m^* \log(n_m^*/n)\ \sum_l n_l \log(n_l/n)}}.$$

where $I(C, C^*)$ is the mutual information of two partitions, $H(C)$ is the entropy, and values:

$$n_{ml} = |C_m^* \cap C_l|, \quad n_m^* = |C_m^*|, \quad n_l = |C_l|.$$

6.3 LFR benchmark

LFR benchmark proposed by Lancichinetti et al. (2008) is a well known network model with community structure. In this model the node degrees are distributed according to a power law with exponent τ_1; the community sizes also obey a power law distribution, with exponent τ_2. Define by k_i^{in} and k_i^{out} the internal and external degree of the node i correspondingly. k_i^{in} is the number of edges between node i and the nodes in the

same community, k_i^{out} is the number of edges between node i and the nodes from other communities. Define by μ the mixing parameter

$$\mu \stackrel{\text{def}}{=} \frac{k_i^{out}}{k_i^{in} + k_i^{out}}.$$

In LFR graphs each node shares a fraction μ of its edges with the other nodes of its community and a fraction μ with the nodes of the other communities; $0 < \mu < 1$ is the mixing parameter. And the condition $\mu < (N - n_c^{max})/N$ guaranties that all communities are well defined, with n_c^{max} being the size of the largest community. For large networks with $n_c^{max} \ll N$, we have $\mu < 1$. The software for generating such graphs is available on `https://sites.google.com/site/santofortunato/inthepress2`. To generate LFR graphs, the next parameters of networks were used: the average node degree equal to 20, the maximum degree equal to 50, the exponent of the degree distribution equal to 2 and the community size distribution equal to 1. The remaining parameters are the minimum s_{min} and maximum s_{max} size of communities. It is usual to differentiate two cases: small (LFR S) and big (LFR B) size of communities as follows: $s_{min} = 10, s_{max} = 50$ for LFR S and $s_{min} = 20, s_{max} = 100$ for LFR B.

6.4 Results for artificial data

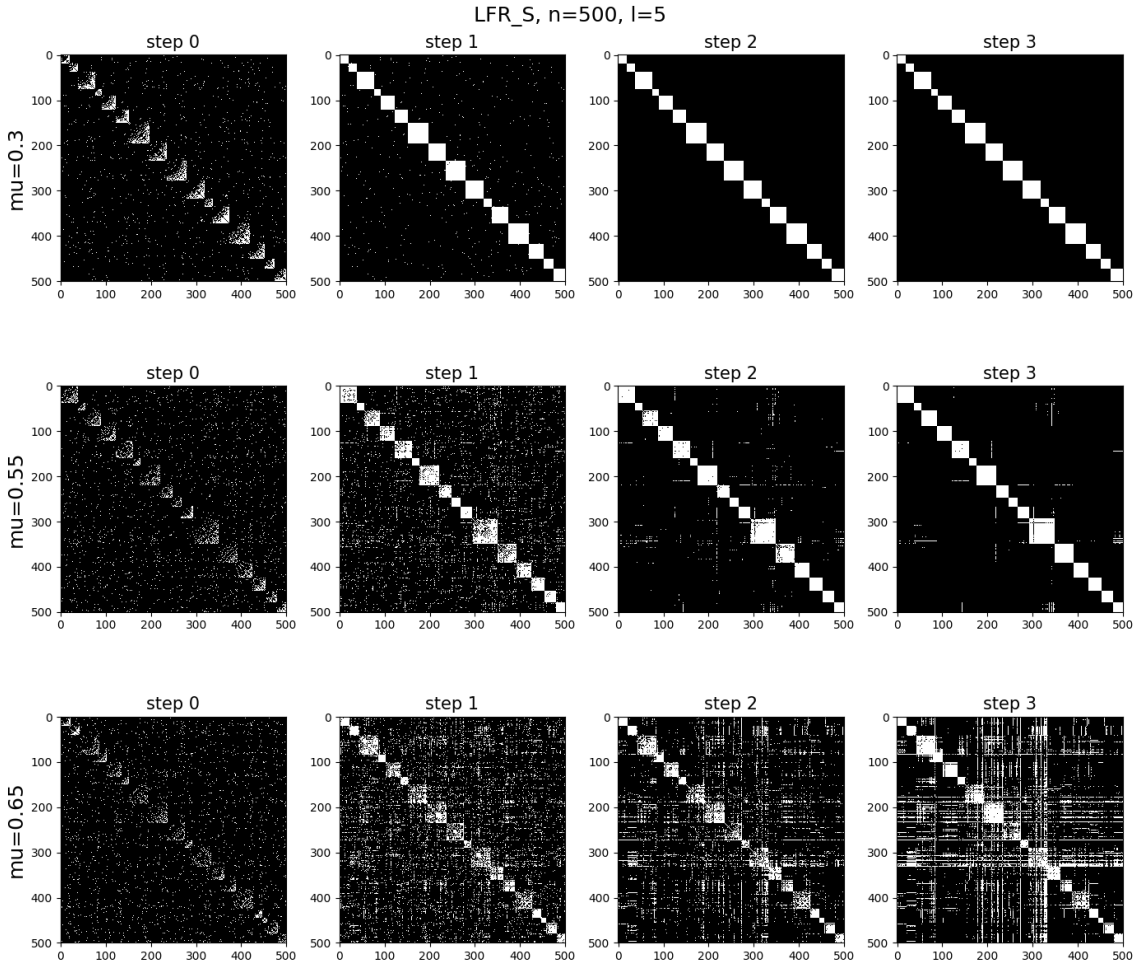
To grasp the idea of AWCD through visual illustration, we demonstrate the evolution of the detected community structure of a graph during different steps of the algorithm. As an example, on the Figure 1 we show what is happening during the each step of AWCD algorithm on LFR S benchmark. The Figure 1 has three rows, each row corresponds to one realization of a graph with $n = 500$ and the value of the mixing parameter $\mu \in \{0.3, 0.55, 0.65\}$, the value is shown on the left. For each realization, from left to right we plot its matrix of weights $W^{(k)}$ for steps $k = 0, 1, 2, 3$. In this demonstration the parameter λ is fixed by $\lambda = 5$ without any preliminary optimization.

As in the case with $\mu = 0.3$ it is seen that the communities are easily identified even on the step 1. During the step 2 the procedure converges by removing the remaining connections between communities. Hence, after the further step 3 the matrix of weights remains unchanged.

The second graph realization with $\mu = 0.55$ corresponds to the breaking point when the algorithm starts to make mistakes. At the step 1 it sets up the majority of the intra-community connections, as well as additionally connecting nodes between communities. Further, on the step 2, the algorithm shapes the identified community structure by deleting these inter-community connections. And the final cleaning is done at the step 3.

The last graph realization with $\mu = 0.65$ is a complex case, however the algorithm manages to identify correctly some of the communities.

After demonstrating the behaviour of the procedure with the specified parameters, we now test the performance of the algorithm by varying λ and the mixing parameter μ. We also increase the benchmark size by setting the number of nodes $n = 1000$. To measure the performance we compute the average NMI over 10 runs, that is for each value of μ we generate 10 graphs and run AWCD with different values of λ on these graphs. The results



AWCD

Figure 1: AWCD results ($\lambda = 5$) for LFR S graph with $\mu = 0.35, 0.55, 0.65$. *From left in each row:* $W^{(0)}, W^{(1)}, W^{(2)}, W^{(3)}$

are shown on the Figure 2. Consider the left plot for LFR S benchmark. Here besides AWCD with $\lambda = 0, 1, 2$ we plot the performance of Louvain and Infomap algorithms.

One can see from the left plot that for small values of the mixing parameter $\mu < 0.6$ there is no much difference in the performance. All the algorithms correctly identify the community structure. The NMI values of the algorithms start to decline after further increasing the parameter μ. As the choice of the best performing λ varies depending on μ, we show this dependence on the Figure 3. Here for each fixed $\mu \in \{0.4, 0.6, 0.65, 0.7, 0.8\}$ we run AWCD and plot the NMI as a function of λ. It can be seen that the choice of λ for $\mu < 0.65$ is not significant for the performance of AWCD.

To understand how modularity helps to choose the best performing λ, consider the Figure 4, where NMI and modularity Q are plotted as functions of λ. The figure consists of 3 LFR S graph realisations with fixed μ values. Our approach is to choose the λ value, which maximizes the modularity. As might be seen, the λ value maximizing modularity does not always coincide with the one providing maximum value of NMI. Nevertheless, it leads to a close to the optimum NMI value . To test the proposed approach we plot the performance of AWCD with λ chosen by modularity maximization. The results are shown on the Figure 5, together with results for Infomap and Louvain algorithms. As we have seen, for $\mu < 0.6$ all the algorithms produce good results. When $\mu \in [0.6, 0.7]$ the performance of Infomap drops significantly resulting a low NMI value for $\mu = 0.7$ and failing to identify any proper community partition. However, AWCD and Louvain still keep to detect communities with a slight advantage of AWCD. After $\mu > 0.7$ the performance of Louvain inevitably worsens , on the contrary AWCD continues to produce results with acceptable level of NMI.

Comparing Figures 2 and 5 one can conclude that the modularity maximization technique helps in choosing λ in almost optimal way.

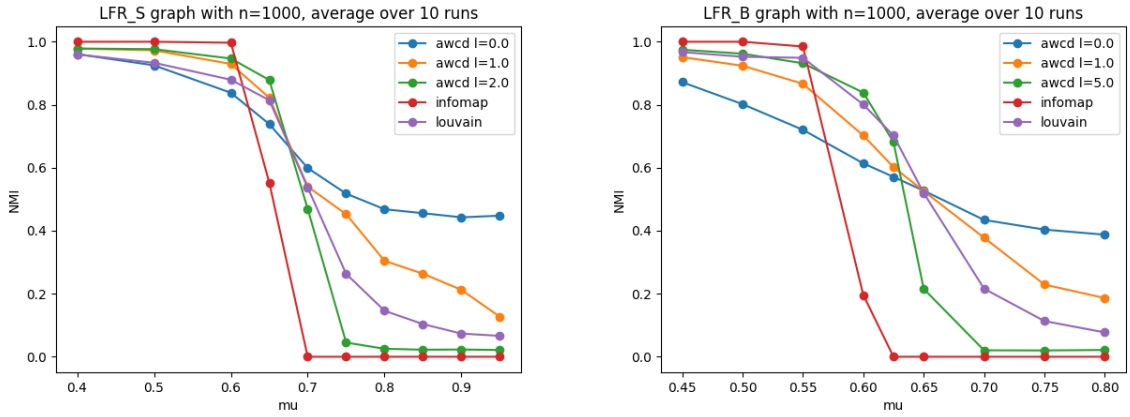


Figure 2: NMI for Louvain, Infomap and AWCD with different λ on LFR bencmhark with $n = 1000$. Average over 10 run. *Left:* LFR S, $\lambda \in \{0, 1, 2\}$. *Right:* LFR B, $\lambda \in \{0, 1, 5\}$.

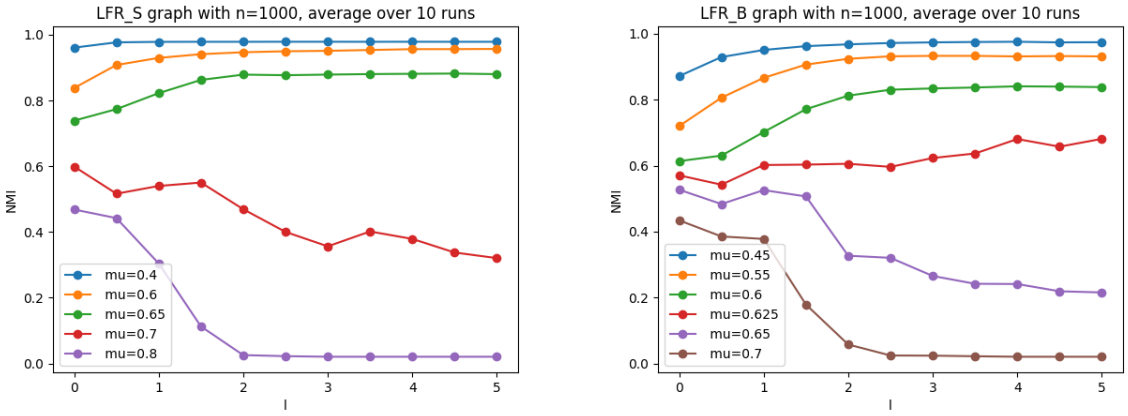


Figure 3: NMI as a function of λ. Each line is an average of AWCD results for 10 graphs with fixed μ, see the legend. *Left:* LFR S, n = 1000. *Right:* LFR B, n = 1000.

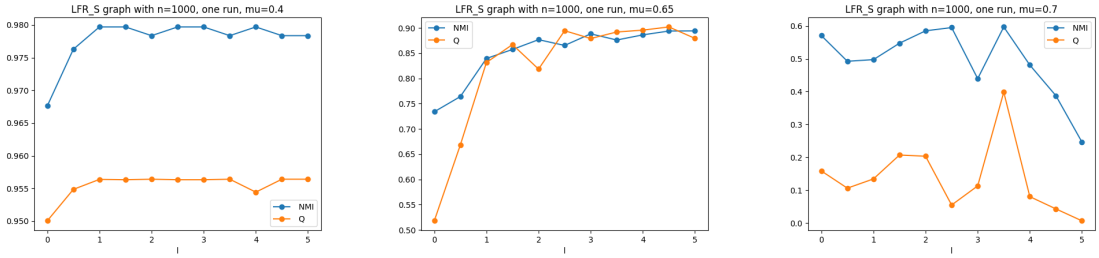


Figure 4: NMI and modularity for one realisation of LFR S graph with $n = 1000$. *From left:* $\mu = 0.4$, $\mu = 0.65$, $\mu = 0.7$.

6.5 Results for real world data

Karate club Zachary (1977) is one of the most popular real world networks used in the evaluation of community detection algorithms. The data consists of 34 nodes representing the members of a university karate club. The edges connect those members who has interacted outside the activities of the club. At some time point a conflict between the club president and the instructor occurred. This led to a division of the members into two separate groups. A question arises if it is possible to reconstruct this division by observing the original network.

We run AWCD on this network with different values of parameter λ and calculate for each λ the modularity and sum of weights, see the left plot of the Figure 6. For the final answer we choose λ with maximum modularity Q, that is $\lambda = 0$. As we know the "true" community structure of this network, we also calculate and plot on the Figure 6 the NMI

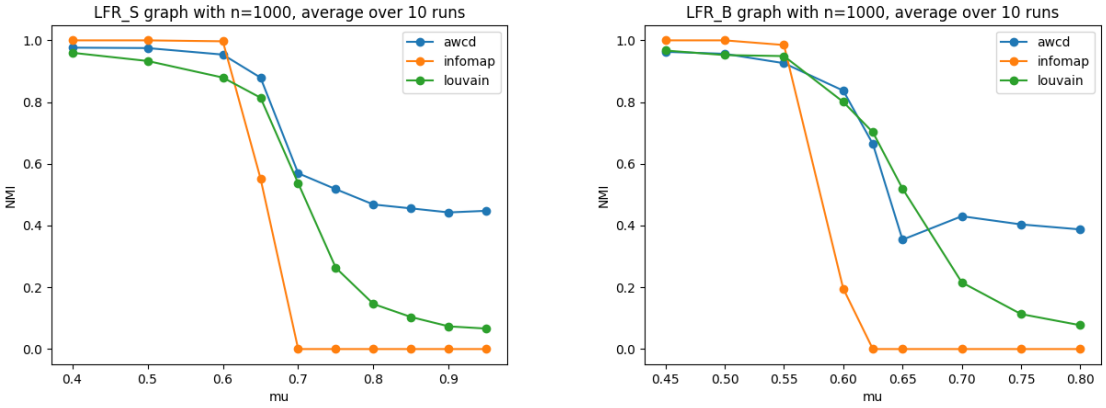


Figure 5: NMI for AWCD with choice of λ maximizing the modularity, $n = 1000$ *Left:* LFR S, *Right:* LFR B.

score for each λ. One can see from the plot, that $\lambda = 0$ also leads to the maximum value of NMI.

The right plot of the Figure 6 shows the resulted community structure of the network found by AWCD. Here each detected community is plotted with its own color. To see the difference between the true community structure, we plotted the network in a way, that the nodes within one true community have the same size. The plot shows that the community supporting the president (node 34) is detected perfectly and the community supporting the instructor is divided further into two groups.

Dolphins dataset Lusseau (2003) is another well known network, consisting of 62 bottle nose dolphins. A pair of animals were considered connected (by an edge in the network) when they were seen together more often than expected by chance. To partition this network we follow the same approach we used for the Karate club dataset. We run AWCD algorithm for a range of λ and compute modularity, see Figure 7. The highest modularity is achieved with $\lambda = 0$ resulting a community structure shown on the Figure 8 left. On the Figure 7 besides NMI and modularity we also plot the sum of weights. Another interesting insight in the plot of modularity is the area within range of $\lambda \in [2, 9]$, where the modularity remains unchanged forming a so called plateau. This plateau indicates that there is a huge gap between detected communities, hence we need to drastically increase λ to overcome this gap and start to merge nodes from different communities.

This structure with a gap is shown on Figure 8 right. Indeed, the gap is easy to see even by looking at the graph. Furthermore this structure repeats almost exactly the community structure proposed by Lusseau (2003) except for the node $SN89$ which originally belongs to the upper green community.

Polbooks dataset Krebs (2004) is a network of 105 books about US politics sold by Amazon.com. The edge between two nodes indicates that the books were frequently purchased together. Books are labelled as "liberal","neutral" or "conservative". We consider

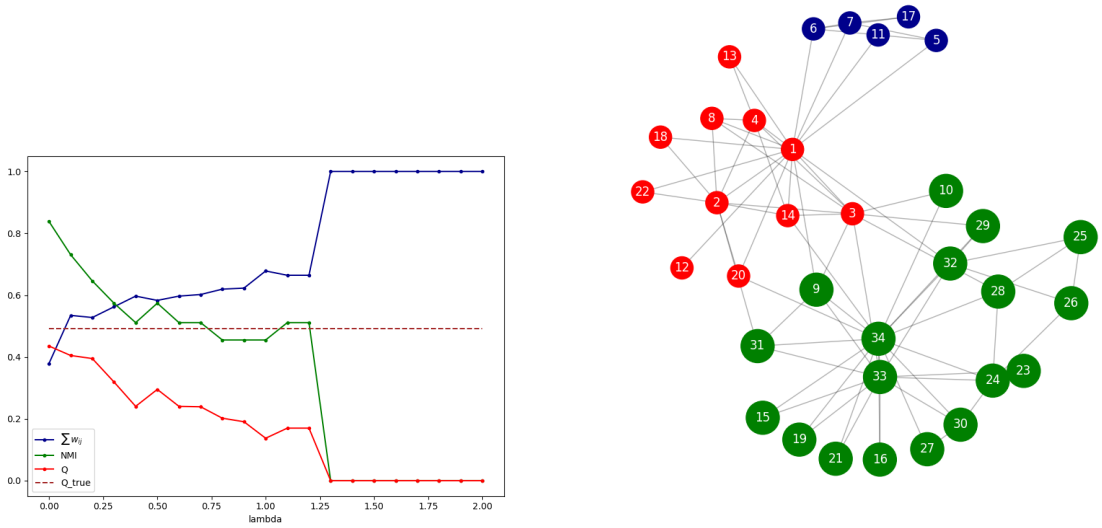


Figure 6: AWCD results for Karate Club network. Left: NMI and modularity. Right: Graph partition.The size of circles represents the true partition, and the color corresponds to the community structure detected by AWCD with $\lambda = 0$.

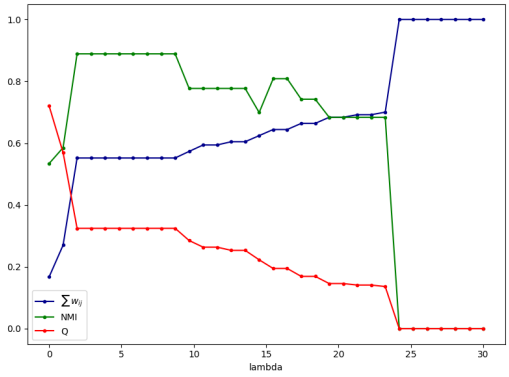


Figure 7: NMI and modularity of AWCD for dolphins dataset.

this labelling as the true community labelling. The left plot of Figure 9 shows the performance metrics of AWCD with different values of parameter λ on this dataset. From this plot we fix $\lambda = 4$, which maximises the modularity, and run AWCD with it. The next plots on Figure 9 show the true matrix of weights, the observed edge matrix and the detected partition by AWCD. Again the chosen λ results in almost optimum value of NMI.

Polblogs dataset Political blogs represented as a network of hyperlinks between weblogs on US politics, recorded in 2005 by Adamic and Glance (2005). Links are all front-page hyperlinks at the time of the crawl. Groups are assigned "liberal" or "conservative" either

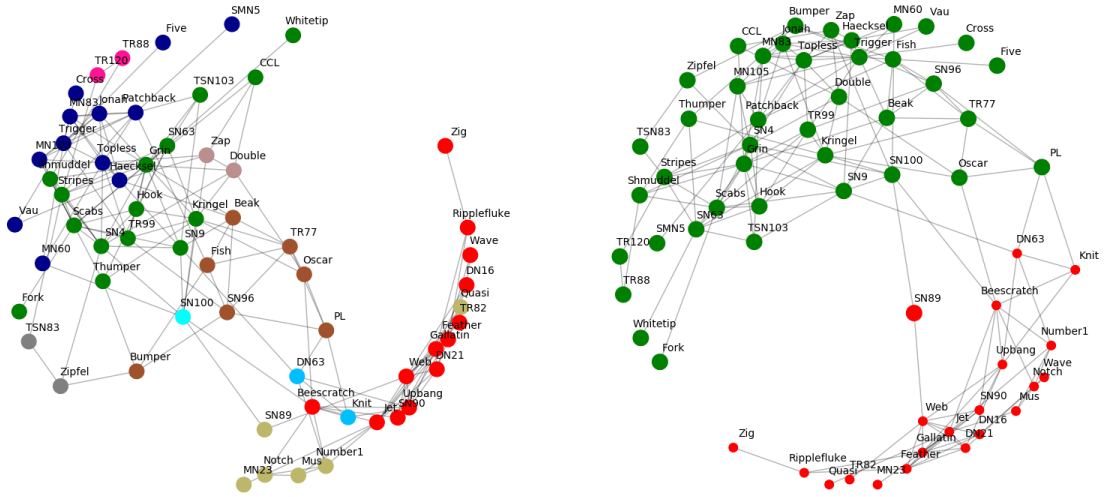


Figure 8: Dolphins dataset, colors represent the AWCD partition with $\lambda = 0$ (left), and $\lambda = 5$ (right).

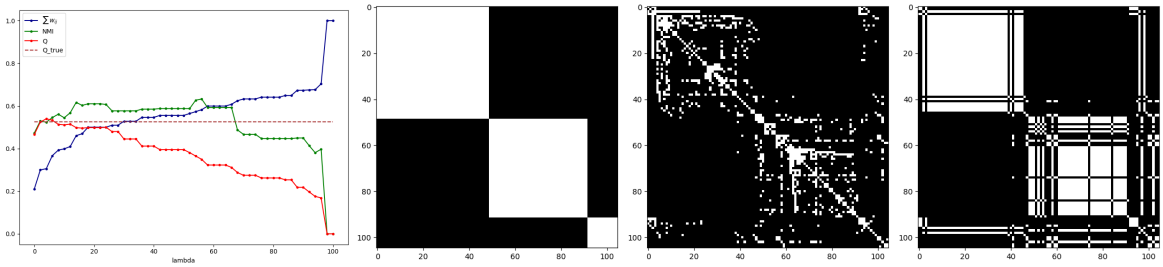


Figure 9: AWCD results for polbooks network. From left: NMI and modularity, the true community structure, the observed adjacency matrix, the detected partition with $\lambda = 4$.

blog directories or occasional self-evaluation. The network consists of 1222 nodes. The left plot on Figure 10 shows the NMI and modularity values of community structure detected by AWCD with different parameter λ values. The next plots of the Figure 10 show the true community structure, edge matrix and the result of our algorithm for optimal λ. As one can see from the adjacency matrix, the network has a huge gap between communities, and one needs a very large λ to merge everything into one community. There is no significant difference in NMI between smaller λ, but the one taken from modularity maximization leads to almost optimum NMI value, similarly to previous datasets.

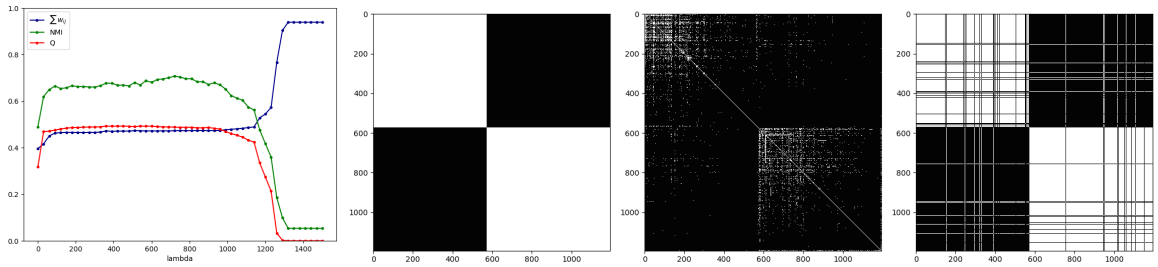


Figure 10: AWCD results for Polblogs network. From left: NMI and modularity, the true community structure, the observed adjacency matrix, the detected partition with $\lambda = 400$.

Football network Girvan and Newman (2002) represents the schedule of an American college football games. Nodes in the graph are the football teams and edges represent regular-season games between the two teams they connect. The network consists of 115 nodes divided in 12 conferences, which are considered as the true community structure. The left plot of Figure 11 shows the performance metrics of AWCD with different λ. To partition this dataset we run AWCD with $\lambda = 0.5$ providing maximum modularity. The true matrix of weights, the observed edge matrix and the AWCD result are shown on the next plots of Figure 11. Again the chosen λ results in almost optimum value of NMI.

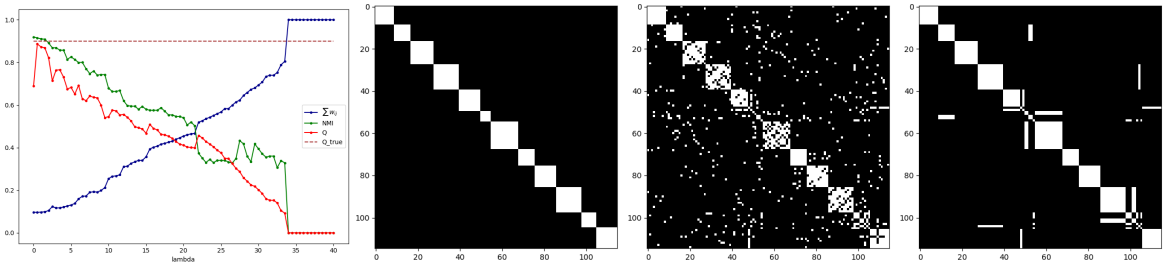


Figure 11: AWCD results for Football network. From left: NMI and modularity, the true community structure, the observed adjacency matrix, the detected partition with $\lambda = 0.5$.

Conclusion

Across many fields, from physiology to nutrition to marketing to providing services, detecting and mapping communities is critical in understanding the structure of complex networks. In this work we introduced a new method for detecting communities called Adaptive Weights Community Detection AWCD. The procedure of detecting communities exploits the idea of automatically extracting the structural information in the observed graph in terms of adaptive weights and utilizes this information for graph partitioning. The method is computationally feasible and scalable for applying on huge graphs. The numerical study

of the method showed a state-of-the-art performance of the method in comparison with existing well known procedures.

Acknowledgments

Financial support in part from the Deutsche Forschungsgemeinschaft via the IRTG 1792 "High Dimensional Non Stationary Time Series", Humboldt-Universität zu Berlin , in part from the Research Unit 1735 "Structural Inference in Statistics: Adaptation and Efficiency" and in part from the Russian Science Foundation (project no. 14 50 00150) is gratefully acknowledged.

Appendix A. Proof of Lemma 3

Consider a testing problem with the null hypothesis

$$H_0 : \theta_{12}^* < \min(\theta_1^*, \theta_2^*)$$
$$H_1 : alternative$$

The LR test statistics is calculated via

$$T = \sup \mathcal{L}(\theta_1, \theta_2, \theta_{12}|H_0) - \sup \mathcal{L}(\theta_1, \theta_2, \theta_{12}|H_1),$$

where

$$\mathcal{L}(\theta_1, \theta_2, \theta_{12}) = \mathcal{L}_1(\theta_1) + \mathcal{L}_2(\theta_2) + \mathcal{L}_{12}(\theta_{12}), \quad \theta_1, \theta_2, \theta_{12} \in I\!R.$$

Then

$$T = \sup_{\theta_{12} < \min(\theta_1, \theta_2)} (\mathcal{L}_1(\theta_1) + \mathcal{L}_2(\theta_2) + \mathcal{L}_{12}(\theta_{12})) - \sup_{\theta_{12} \geq \min(\theta_1, \theta_2)} (\mathcal{L}_1(\theta_1) + \mathcal{L}_2(\theta_2) + \mathcal{L}_{12}(\theta_{12})). \tag{6}$$

Set $\sup_\theta \mathcal{L}_1(\theta) = \mathcal{L}_1(\widetilde{\theta}_1)$, $\sup_\theta \mathcal{L}_{12}(\theta) = \mathcal{L}_{12}(\widetilde{\theta}_{12})$ and $\sup_\theta \mathcal{L}_2(\theta) = \mathcal{L}_2(\widetilde{\theta}_2)$. Without loss of generality we assume that $\widetilde{\theta}_1 < \widetilde{\theta}_2$.

$\widetilde{\theta}_{12} \leq \widetilde{\theta}_1 \leq \widetilde{\theta}_2$

This is a natural "gap" case when the observed density of edges between communities is lower than any density inside communities. Then the test statistics T from (6) can be rewritten as

$$T = \mathcal{L}_1(\widetilde{\theta}_1) + \mathcal{L}_2(\widetilde{\theta}_2) + \mathcal{L}_{12}(\widetilde{\theta}_{12}) - \sup_{\theta_{12} \geq \min(\theta_1, \theta_2)} (\mathcal{L}_1(\theta_1) + \mathcal{L}_2(\theta_2) + \mathcal{L}_{12}(\theta_{12})).$$

Suppose we know the point of maximum $(\theta_1^*, \theta_{12}^*, \theta_2^*)$ in the right part of previous equation, that is $\theta_1 = \theta_1^*$, $\theta_{12} = \theta_{12}^*$, $\theta_2 = \theta_2^*$. Consider two cases.

First if $\theta_1^* \le \theta_{12}^*$ and $\theta_2^* \neq \widetilde{\theta}_2$. Then due to concavity of likelihood function another triplet $(\theta_1^*, \theta_{12}^*, \widetilde{\theta}_2)$ will lead to a higher value of $\mathcal{L}(\theta_1, \theta_2, \theta_{12})$. It means that from $\theta_1^* \le \theta_{12}^*$ it follows $\theta_2^* = \widetilde{\theta}_2$. In addition again due to concavity of likelihood function as far as $\widetilde{\theta}_{12} \le \widetilde{\theta}_1$ it follows that $\theta_1^* = \theta_{12}^* = \widehat{\theta}$:

$$\theta_2^* = \widetilde{\theta}_2, \quad \theta_1^* = \theta_{12}^* = \widehat{\theta} \in [\widetilde{\theta}_{12}, \widetilde{\theta}_1]$$

Another case $\theta_2^* \le \theta_{12}^*$ by following the same argumentation will lead to:

$$\theta_1^* = \widetilde{\theta}_1, \quad \theta_2^* = \theta_{12}^* = \widehat{\theta} \in [\widetilde{\theta}_{12}, \widetilde{\theta}_2].$$

In the end we can rewrite our test statistics

$$\begin{aligned} T = \inf \big(& \mathcal{L}_1(\widetilde{\theta}_1) + \mathcal{L}_{12}(\widetilde{\theta}_{12}) - \mathcal{L}_{1\vee 12}(\widetilde{\theta}_{1\vee 12}), \\ & \mathcal{L}_2(\widetilde{\theta}_2) + \mathcal{L}_{12}(\widetilde{\theta}_{12}) - \mathcal{L}_{2\vee 12}(\widetilde{\theta}_{2\vee 12})\big), \end{aligned} \tag{7}$$

where

$$\begin{aligned} \mathcal{L}_1(\widetilde{\theta}_1) &= S_1 \log \widetilde{\theta}_1 + (N_1 - S_1)\log(1-\widetilde{\theta}_1), \\ \mathcal{L}_{12}(\widetilde{\theta}_{12}) &= S_{12} \log \widetilde{\theta}_{12} + (N_{12} - S_{12})\log(1-\widetilde{\theta}_{12}), \\ \mathcal{L}_{1\vee 12}(\widetilde{\theta}_{1\vee 12}) &= S_{1\vee 12} \log \widetilde{\theta}_{1\vee 12} + (N_{1\vee 12} - S_{1\vee 12})\log(1-\widetilde{\theta}_{1\vee 12}) = \\ &= (S_1 + S_{12}) \log \widetilde{\theta}_{1\vee 12} + (N_1 - S_1 + N_{12} - S_{12})\log(1-\widetilde{\theta}_{1\vee 12}). \end{aligned}$$

Then the left side in (7)

$$\mathcal{L}_1(\widetilde{\theta}_1) + \mathcal{L}_{12}(\widetilde{\theta}_{12}) - \mathcal{L}_{1\vee 12}(\widetilde{\theta}_{1\vee 12}) = S_1 \log \frac{\widetilde{\theta}_1}{\widetilde{\theta}_{1\vee 12}} + (N_1 - S_1)\log \frac{1-\widetilde{\theta}_1}{1-\widetilde{\theta}_{1\vee 12}} +$$

$$+ S_{12} \log \frac{\widetilde{\theta}_2}{\widetilde{\theta}_{1\vee 12}} + (N_{12} - S_{12})\log \frac{1-\widetilde{\theta}_{12}}{1-\widetilde{\theta}_{1\vee 12}} =$$

$$= N_1[\widetilde{\theta}_1 \log \frac{\widetilde{\theta}_1}{\widetilde{\theta}_{1\vee 12}} + (1-\widetilde{\theta}_1)\log \frac{1-\widetilde{\theta}_1}{1-\widetilde{\theta}_{1\vee 12}}] + N_{12}[\widetilde{\theta}_{12} \log \frac{\widetilde{\theta}_{12}}{\widetilde{\theta}_{1\vee 12}} + (1-\widetilde{\theta}_{12})\log \frac{1-\widetilde{\theta}_{12}}{1-\widetilde{\theta}_{1\vee 12}}] =$$

$$= N_1 \mathcal{K}(\widetilde{\theta}_1, \widetilde{\theta}_{1\vee 12}) + N_{12}\mathcal{K}(\widetilde{\theta}_{12}, \widetilde{\theta}_{1\vee 12}) \stackrel{\text{def}}{=} T_A$$

Similarly, the right part

$$\mathcal{L}_2(\widetilde{\theta}_2) + \mathcal{L}_{12}(\widetilde{\theta}_{12}) - \mathcal{L}_{2\vee 12}(\widetilde{\theta}_{2\vee 12}) = N_2 \mathcal{K}(\widetilde{\theta}_2, \widetilde{\theta}_{2\vee 12}) + N_{12}\mathcal{K}(\widetilde{\theta}_{12}, \widetilde{\theta}_{2\vee 12}) = T_B$$

Therefore the test statistics T in (7) in the case when $\widetilde{\theta}_{12} < \widetilde{\theta}_1 < \widetilde{\theta}_2$ is equal to

$$T = \min(T_A, T_B).$$

Appendix B. LR test for the classical gap case

Lemma 4 *The likelihood ratio test for the hypothesis*

$$H_0 : \theta_1^* < \theta_2^*$$
$$H_1 : alternative$$

is written as

$$T = \Big\{ N_1 \mathcal{K}(\widetilde{\theta}_1, \widetilde{\theta}_{1+2}) + N_2 \mathcal{K}(\widetilde{\theta}_2, \widetilde{\theta}_{1+2}) \Big\} \times \Big\{ \mathbb{1}\big(\widetilde{\theta}_2 > \widetilde{\theta}_1\big) - \mathbb{1}\big(\widetilde{\theta}_2 < \widetilde{\theta}_1\big) \Big\}. \tag{8}$$

Proof Test statistics is calculated via

$$T = \sup \mathcal{L}(\theta_1, \theta_2 | H_0) - \sup \mathcal{L}(\theta_1, \theta_2 | H_1),$$

where

$$\mathcal{L}(\theta_1, \theta_2) = \mathcal{L}_1(\theta_1) + \mathcal{L}_2(\theta_2), \quad \theta_1, \theta_2 \in I\!R.$$

Then

$$T = \sup_{\theta_1 < \theta_2} (\mathcal{L}_1(\theta_1) + \mathcal{L}_2(\theta_2)) - \sup_{\theta_1 \geq \theta_2} (\mathcal{L}_1(\theta_1) + \mathcal{L}_2(\theta_2)).$$

We will follow the same notations as in previous chapters: $\sup_\theta \mathcal{L}_1(\theta) = \mathcal{L}_1(\widetilde{\theta}_1)$, $\sup_\theta \mathcal{L}_2(\theta) = \mathcal{L}_2(\widetilde{\theta}_2)$.

First consider the case when $\widetilde{\theta}_1 < \widetilde{\theta}_2$. Then

$$T = \mathcal{L}_1(\widetilde{\theta}_1) + \mathcal{L}_2(\widetilde{\theta}_2) - \sup_{\theta_1 \geq \theta_2} (\mathcal{L}_1(\theta_1) + \mathcal{L}_2(\theta_2)).$$

As far as $\mathcal{L}_1'(\theta) > 0$ for $\theta < \widetilde{\theta}_1$ and $\mathcal{L}_1'(\theta) < 0$ for $\theta > \widetilde{\theta}_1$, and the same holds for the $\mathcal{L}_2'$, then

$$\sup_{\theta_1 \geq \theta_2} (\mathcal{L}_1(\theta_1) + \mathcal{L}_2(\theta_2)) = \sup_{\theta} (\mathcal{L}_1(\theta) + \mathcal{L}_2(\theta)) = \mathcal{L}_{1+2}(\theta_{1+2})$$

Following the same logic as in the proof of the Theorem 3, it holds

$$T = \mathcal{L}_1(\widetilde{\theta}_1) + \mathcal{L}_2(\widetilde{\theta}_2) - \mathcal{L}_{1+2}(\widetilde{\theta}_{1+2}) = N_1 \mathcal{K}(\widetilde{\theta}_1, \widetilde{\theta}_{1+2}) + N_2 \mathcal{K}(\widetilde{\theta}_2, \widetilde{\theta}_{1+2}).$$

Symmetrically in the case $\widetilde{\theta}_2 < \widetilde{\theta}_1$ by following the same argumentation it holds

$$T = \mathcal{L}_{1+2}(\widetilde{\theta}_{1+2}) - \mathcal{L}_1(\widetilde{\theta}_1) - \mathcal{L}_2(\widetilde{\theta}_2).$$

Hence, by combining both cases we get (8). ■

Appendix C. Proof of Theorem 1

We proceed by induction in k. The goal is to show that all the weights $w_{ij}^{(k)}$ computed at each step k are equal to one. This would mean that any community η from $\mathcal{C}_\xi^{(k)}$ and from its one-vicinity $\boldsymbol{u}_\xi^{(k)}$ will get the weight one and hence, included in $\mathcal{C}_\xi^{(k+1)}$.

First consider the step after initialization. For notational convenience we suppress the upper index (0) for all considered objects. For any node ξ, its friends list $\mathcal{C}_\xi$ is described by the ξ-row of $\mathcal{Y}$. It is important that ξ itself is not included in this set. Let $\mathcal{Y}_\xi$ be the set of edges from the node ξ. Let now η be another node from one-vicinity of $\mathcal{C}_\xi$. This means that $\mathcal{C}_\xi$ and η are connected by some edge(s). Denote by $\mathcal{Y}_{\xi\xi}$ the set of edges between nodes in $\mathcal{C}_\xi$, similarly for $\mathcal{Y}_{\eta\eta}$, and by $\mathcal{Y}_{\xi\eta}$ the set of edges connecting $\mathcal{C}_\xi$ and $\mathcal{C}_\eta$. The key observation is that the sets $\mathcal{Y}_\xi$ and $\mathcal{Y}_\eta$ do not overlap with $\mathcal{Y}_{\xi\xi}$, $\mathcal{Y}_{\xi\eta}$, and $\mathcal{Y}_{\eta\eta}$ yielding their independence. Note also that $N_{\xi\xi} = N_\xi^2 = |\mathcal{Y}_\xi|^2$ and only depends on $\mathcal{Y}_\xi$, while $S_{\xi\xi}$ is defined by summation over $\mathcal{Y}_{\xi\xi}$. Therefore, the conditional distribution of $S_{\xi\xi}$ given $\mathcal{Y}_\xi$ coincides with the unconditional one and we can proceed as if the set $\mathcal{Y}_\xi$ were deterministic. The same is true for $S_{\xi\eta}$ and $S_{\eta\eta}$ when conditioning w.r.t. $\mathcal{Y}_\xi$ and $\mathcal{Y}_\eta$. Therefore, the results from Polzehl and Spokoiny (2006a) for i.i.d. Bernoulli random sums apply here:

$$I\!P\big(N_{\xi\xi}\mathcal{K}(\widetilde{\theta}_{\xi\xi},\theta^*) > \mathfrak{z}\big) = I\!P\Big(N_{\xi\xi}\mathcal{K}(\widetilde{\theta}_{\xi\xi},\theta^*) > \mathfrak{z} \,\Big|\, \mathcal{Y}_\xi, \mathcal{Y}_\eta\Big) \le 2\mathrm{e}^{-\mathfrak{z}}$$

and similarly for $\widetilde{\theta}_{\xi\eta}$ and $\widetilde{\theta}_{\eta\eta}$. Moreover, similar bounds apply to $\widetilde{\theta}_{\xi+\eta}$ and to $\widetilde{\theta}_{\xi\vee\eta}$. The choice $\mathfrak{z} = 3\log n$ allows to bound all considered test statistics simultaneously on a set of overholming probability $2/n$. Therefore, ignoring an event of a very small probability $2/n$, we can assume that

$$\begin{aligned} N_{\xi\eta}\mathcal{K}(\widetilde{\theta}_{\xi\eta},\theta^*) &\le \mathfrak{z}, \\ N_{\xi+\eta}\mathcal{K}(\widetilde{\theta}_{\xi+\eta},\theta^*) &\le \mathfrak{z}, \\ N_{\xi\vee\eta}\mathcal{K}(\widetilde{\theta}_{\xi\vee\eta},\theta^*) &\le \mathfrak{z} \end{aligned}$$

for all considered pairs ξ, η.

Suppose the induction assumption for the step $k-1$. Then at the step k, the adaptive vicinity of each vertex V_i coincides with the non-adaptive ball around V_i described by the condition $d_{ij} \le h_k$. The related test statistic $T_{ij}^{(k)}$ fulfills.

References

Lada A Adamic and Natalie Glance. The political blogosphere and the 2004 us election: divided they blog. In *Proceedings of the 3rd international workshop on Link discovery*, pages 36–43. ACM, 2005.

Vincent D Blondel, Jean-Loup Guillaume, Renaud Lambiotte, and Etienne Lefebvre. Fast unfolding of communities in large networks. *Journal of statistical mechanics: theory and experiment*, 2008(10):P10008, 2008.

Lukas Borke and Wolfgang K Härdle. Github api based quantnet mining infrastructure in r. 2017.

Aaron Clauset, Mark EJ Newman, and Cristopher Moore. Finding community structure in very large networks. *Physical review E*, 70(6):066111, 2004.

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

Arthur P Dempster, Nan M Laird, and Donald B Rubin. Maximum likelihood from incomplete data via the em algorithm. *Journal of the royal statistical society. Series B (methodological)*, pages 1–38, 1977.

Kirill Efimov, Larisa Adamyan, and Vladimir Spokoiny. Adaptive nonparametric clustering. *arXiv preprint arXiv:1709.09102*, 2017.

Jianqing Fan, Chunming Zhang, and Jian Zhang. Generalized likelihood ratio statistics and wilks phenomenon. *Annals of statistics*, pages 153–193, 2001.

Santo Fortunato. Community detection in graphs. *Physics reports*, 486(3-5):75–174, 2010.

Santo Fortunato and Darko Hric. Community detection in networks: A user guide. *Physics Reports*, 659:1–44, 2016.

Jerome Friedman, Trevor Hastie, and Robert Tibshirani. *The elements of statistical learning*, volume 1. Springer series in statistics New York, NY, USA:, 2001.

Michelle Girvan and Mark EJ Newman. Community structure in social and biological networks. *Proceedings of the national academy of sciences*, 99(12):7821–7826, 2002.

Peter D Grünwald, In Jae Myung, and Mark A Pitt. *Advances in minimum description length: Theory and applications*. MIT press, 2005.

Darko Hric, Richard K Darst, and Santo Fortunato. Community detection in networks: Structural communities versus ground truth. *Physical Review E*, 90(6):062805, 2014.

Brian W Kernighan and Shen Lin. An efficient heuristic procedure for partitioning graphs. *The Bell system technical journal*, 49(2):291–307, 1970.

Valdis Krebs. Books about us politics. *unpublished, http://www. orgnet. com*, 2004.

Andrea Lancichinetti and Santo Fortunato. Community detection algorithms: a comparative analysis. *Physical review E*, 80(5):056117, 2009.

Andrea Lancichinetti, Santo Fortunato, and Filippo Radicchi. Benchmark graphs for testing community detection algorithms. *Physical review E*, 78(4):046110, 2008.

David Lusseau. The emergent properties of a dolphin social network. *Proceedings of the Royal Society of London B: Biological Sciences*, 270(Suppl 2):S186–S188, 2003.

Mark EJ Newman. Modularity and community structure in networks. *Proceedings of the national academy of sciences*, 103(23):8577–8582, 2006.

Mark EJ Newman and Michelle Girvan. Finding and evaluating community structure in networks. *Physical review E*, 69(2):026113, 2004.

Mark EJ Newman and Elizabeth A Leicht. Mixture models and exploratory analysis in networks. *Proceedings of the National Academy of Sciences*, 104(23):9564–9569, 2007.

MEJ Newman. Community detection in networks: Modularity optimization and maximum likelihood are equivalent. *arXiv preprint arXiv:1606.02319*, 2016.

Andrew Y Ng, Michael I Jordan, Yair Weiss, et al. On spectral clustering: Analysis and an algorithm. *Advances in neural information processing systems*, 2:849–856, 2002.

Andreas Noack and Randolf Rotta. Multi-level algorithms for modularity clustering. In *International Symposium on Experimental Algorithms*, pages 257–268. Springer, 2009.

Leto Peel, Daniel B Larremore, and Aaron Clauset. The ground truth about metadata and community detection in networks. *Science advances*, 3(5):e1602548, 2017.

Joerg Polzehl and Vladimir Spokoiny. Propagation-separation approach for local likelihood estimation. *Probability Theory and Related Fields*, 135(3):335–362, 2006a.

Jörg Polzehl and Vladimir Spokoiny. Propagation-separation approach for local likelihood estimation. *Probability Theory and Related Fields*, 135(3):335–362, 2006b.

Alex Pothen. Graph partitioning algorithms with applications to scientific computing. In *Parallel Numerical Algorithms*, pages 323–368. Springer, 1997.

Jorma Rissanen. Modeling by shortest data description. *Automatica*, 14(5):465–471, 1978.

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

Jianbo Shi and Jitendra Malik. Normalized cuts and image segmentation. *IEEE Transactions on pattern analysis and machine intelligence*, 22(8):888–905, 2000.

Ulrike Von Luxburg. A tutorial on spectral clustering. *Statistics and computing*, 17(4): 395–416, 2007.

Scott White and Padhraic Smyth. A spectral clustering approach to finding communities in graphs. In *Proceedings of the 2005 SIAM international conference on data mining*, pages 274–285. SIAM, 2005.

Liang Yang, Xiaochun Cao, Dongxiao He, Chuan Wang, Xiao Wang, and Weixiong Zhang. Modularity based community detection with deep learning. In *IJCAI*, pages 2252–2258, 2016a.

Zhao Yang, René Algesheimer, and Claudio J Tessone. A comparative analysis of community detection algorithms on artificial networks. *Scientific Reports*, 6:30750, 2016b.

Wayne W Zachary. An information flow model for conflict and fission in small groups. *Journal of anthropological research*, 33(4):452–473, 1977.

IRTG 1792 Discussion Paper Series 2019

For a complete list of Discussion Papers published, please visit http://irtg1792.hu-berlin.de.

001 "Cooling Measures and Housing Wealth: Evidence from Singapore" by Wolfgang Karl Härdle, Rainer Schulz, Taojun Xie, January 2019.

002 "Information Arrival, News Sentiment, Volatilities and Jumps of Intraday Returns" by Ya Qian, Jun Tu, Wolfgang Karl Härdle, January 2019.

003 "Estimating low sampling frequency risk measure by high-frequency data" by Niels Wesselhöfft, Wolfgang K. Härdle, January 2019.

004 "Constrained Kelly portfolios under alpha-stable laws" by Niels Wesselhöfft, Wolfgang K. Härdle, January 2019.

005 "Usage Continuance in Software-as-a-Service" by Elias Baumann, Jana Kern, Stefan Lessmann, February 2019.

006 "Adaptive Nonparametric Community Detection" by Larisa Adamyan, Kirill Efimov, Vladimir Spokoiny, February 2019.

IRTG 1792, Spandauer Strasse 1, D-10178 Berlin
http://irtg1792.hu-berlin.de

This research was supported by the Deutsche Forschungsgemeinschaft through the IRTG 1792.