RC25453 (WAT1403-050) March 14, 2014
Mathematics

IBM Research Report

Community Detection with the Weighted Parsimony Criterion

Andrea Bettinelli
Università degli Studi di Milano
via Bramante 65
Crema, Italy

Pierre Hansen
GERAD
HEC
Montréal, Canada
and
École Polytechnique
F-91128 Palaiseau
France

Leo Liberti
IBM Research Division
Thomas J. Watson Research Center
P.O. Box 218
Yorktown Heights, NY 10598
USA
and
École Polytechnique
F-91128 Palaiseau
France

Community detection with the weighted parsimony criterion

Andrea Bettinelli · Pierre Hansen · Leo Liberti

January 20, 2014

Abstract Community detection in networks has been studied extensively in the last decade. Many criteria, expressing the quality of the partitions obtained, as well as a few exact algorithms and a large number of heuristics have been proposed. The parsimony criterion consists in minimizing the number of edges added or removed from the given network in order to transform it into a set of disjoint cliques. Recently Zhang, Qiu and Zhang have proposed a weighted parsimony model in which a weight coefficient is introduced to balance the numbers of inserted and deleted edges. These authors propose rules to select a good value of the coefficient, use simulated annealing to find optimal or near-optimal solutions and solve a series of real and artificial instances. In the present paper, an algorithm is proposed for solving exactly the weighted parsimony problem for all values of the parameter. This algorithm is based on iteratively solving the problem for a set of given values of the parameter using a row generation algorithm. This procedure is combined with a search procedure to find all lowest breakpoints of the value curve (i.e., the weighted sum of inserted and deleted edges). Computational results on a series of artificial and real world networks from the literature are reported. It appears that several partitions for the same network may be informative and that the set of solutions usually contains at least one intuitively appealing partition.

Keywords community detection, complex networks, parsimony

Andrea Bettinelli
DTI, Università degli Studi di Milano, via Bramante 65, Crema, Italy
E-mail: andrea.bettinelli@unimi.it

Pierre Hansen
GERAD, HEC, Montréal, Canada
Also at LIX, École Polytechnique, F-91128 Palaiseau, France
E-mail: pierre.hansen@gerad.ca

Leo Liberti
LIX, École Polytechnique, F-91128 Palaiseau, France
E-mail: liberti@lix.polytechnique.fr

1 Introduction

Networks, or graphs, are powerful and versatile tools in the study of complex systems arising in the natural and social sciences as well as in engineering and medicine. A network consists of a set of vertices and a set of edges. Edges are pairs of vertices which are graphically represented by lines joining them. Vertices are associated with entities (people, members of an organization, countries, crossroads, atoms, ...) and edges with relationship between them (friendship, cooperation, conflict, existence of connections such as road or electrical line, bonds in a molecule,...).

A ubiquitous problem is to find communities in networks. In general terms a set of vertices is a community if inner edges joining two of its vertices are more dense than outer edges joining one of its vertices to another outside of it. The set of communities (or *clusters*) in a network are always assumed to be pairwise disjoint, thus forming a partition of the vertices. There are many ways to make the concept of community precise and many criteria, or indices to evaluate the value of a community. In turn, numerous heuristics and a few exact algorithms have been proposed for finding an optimal or near-optimal partition of the vertices of a network into communities. Probably the most studied criterion is the *modularity* due to Newman and Girvan [16]. It is defined for each community as the difference between the number of inner edges and the expected number of inner edges keeping the distribution of degrees fixed. The modularity of a partition is the sum of the modularities of its clusters. This popular criterion has been subject to some criticism [4, 7, 9, 15]. The main concern is the resolution limit: small clusters may be "absorbed by" larger ones even if they are very dense (and should therefore be considered as separate clusters). Another approach consists in removing or adding edges to the network until some criterion is satisfied. In the multicut problem one aims at removing the smallest number of edges in order to transform a connected network into k connected components. The main difficulty of this approach is that this connected components may often be reduced to single entities [5]. A related problem corresponds to the parsimony criterion i.e., remove or add the smallest number of edges in order to transform the network into a disjoint collection of cliques [10]. As observed by Zhang et al. [19] this criterion tends to favor small communities; in order to address this shortcoming, these authors propose to modify the parsimony criterion by introducing a weight parameter to balance the contribution to the objective of edges which are deleted and edges which are inserted. The values assigned to this parameter are given by one of three possible formulæ which depend on the density of the network and its clustering coefficient. Near-optimal solutions (or optimal solutions but without a proof of optimality) are obtained with a simulated annealing heuristic. Several artificial and real-world networks are studied and results compared for some of them with those obtained with the modularity [16] and modularity density criteria [13].

In the present paper, we extend the work of Zhang et al. in three ways: (i) we study the properties of the parametric curve of weighted parsimony values; (ii) we present an algorithm for finding the set of all optimal partitions for all values of the parameter. More precisely, we partition the parameter range into a finite set of intervals, to each of which there corresponds a unique optimal weighted parsimony value (associated to one or more optimal partitions). A similar approach was proposed for modularity clustering in [3]. (iii) We apply this algorithm to the same artificial and real examples studied by Zhang et al. and some more besides, showing that considering several partitions instead of a single one can be more informative.

The motivation for our paper is to investigate how the weight parameter influences the optimal clustering in the weighted parsimony clustering criterion. Previous work on this criterion (Zhang et al. [19]) only propose heuristic choices of this parameter. The new contribution of our paper is to provide a systematic method for finding an optimal partition for all values of the parameter. Clustering techniques are ubiquitous in big data technology: our work is relevant to every application of data science. Section 2 (particularly 2.4) provides the core of our theoretical study. We chose to present it in discoursive form, rather than using a theorem/proof approach, to be consistent with the rest of the literature in community detection (see e.g. the papers about modularity clustering in Physical Review E); but the line of reasoning is completely formal and rigorous.

The paper is organized as follows. Definitions and notations are given in the next section, followed by a parametric linear integer programming formulation and an algorithm for solving it. Artificial and real instances are considered in Section 3. The problem of choosing a best value for the parameter is briefly discussed in Section 4. Section 5 concludes the paper.

2 Algorithm

2.1 Definitions and notations

Let $G = (V, E)$ be an undirected graph, or network, with a set of vertices V and a set of edges E. If several edges join the same pair of vertices they are called *multiple edges* and the graph G is a multigraph. An edge joining a vertex to itself is called a *loop*. A simple graph has no loops nor multiple edges. The number of vertices of G is usually denoted by n and called its order. The number of edges of G is usually denoted by m and called its size. The degree k_i of a vertex is equal to the number of vertices it is incident to, or, in other words, to the number of its neighbors. A vertex of degree 1 as well as its only incident edge are called pendant. The density d of a network without loops or multiple edges is the ratio of its number of edges to the maximum possible number of edges i.e., $d = \frac{2m}{n(n-1)}$. Let U be a subset of V, the cutset of U is the set of edges in E with exactly one endpoint in U; a cutset is trivial if $U = \varnothing$ or $U = V$. A graph is connected if all nontrivial cutsets are nonempty. A clique is a subgraph having an edge between any two distinct vertices.

2.2 Mathematical programming formulation

Let $G = (V, E)$ be a simple graph with adjacency matrix A, where $A_{ij} = 1$ if vertices i and j are joined by an edge and to 0 otherwise. We can then express the weighted parsimony problem for a given value of the weight w for the given graph as follows:

$$\begin{aligned}
\min\ P_w = w \sum_{i=1}^{n-1} \sum_{j=i+1}^{n} A_{ij}(1 - x_{ij}) + (1-w) \sum_{i=1}^{n-1} \sum_{j=i+1}^{n} (1 - A_{ij}) x_{ij} & \\
s.t.\ x_{ij} + x_{jk} - x_{ik} \leq 1 & \qquad \forall 1 \leq i < j < k \leq n \\
x_{ij} - x_{jk} + x_{ik} \leq 1 & \qquad \forall 1 \leq i < j < k \leq n \\
-x_{ij} + x_{jk} + x_{ik} \leq 1 & \qquad \forall 1 \leq i < j < k \leq n \\
x_{ij} \in \{0, 1\} & \qquad \forall 1 \leq i < j \leq n
\end{aligned} \tag{1}$$

where the binary variables x_{ij} are equal to 1 if vertices i and j are in the same clique and 0 otherwise. So if an edge (i, j) of G is removed, $1 - x_{ij}$ is equal to 1 with a contribution to the objective value of w, and if an edge (i, j) is inserted in G, x_{ij} is equal to 1 with a contribution to the objective function of $1 - w$. The set of feasible solutions of (1) corresponds exactly to all partitions into cliques of the vertex set V. Each such partition corresponds to an equivalence relation on the entities. Indeed, the corresponding relation satisfies reflexivity (we can assume $x_{ii} = 1$ for each entity i as x_{ii} does not appear in (1)), symmetry (since (1) only mentions indices i, j with $i < j$, we may set $x_{ij} = x_{ji}$ for $i > j$, as we only consider undirected networks) and transitivity (encoded by the constraints of (1)). This problem is a parametric Integer Linear Program (ILP), where the parameter w is allowed to vary in the interval $[0, 1]$. Several algorithms for clique partitioning problems, whose formulation has the same constraints as above, have been proposed in the Combinatorial Optimization literature. Among these a well-known one is the row generation algorithm of Grötschel and Wakabayashi [10, 11]. The problem (1) has $O(n^3)$ constraints and $O(n^2)$ variables. After relaxing the integrality constraints the numerous transitivity constraints are adjoined a batch at a time. When all of those constraints are satisfied, if the solution is integer the algorithm terminates. Otherwise more sophisticated constraints may be adjoined to the formulation, or a Branch-and-Bound (BB) procedure might be called. This algorithm allows solution of instances with up to 150 entities in "reasonable" time.

2.3 Properties of the parametric curve of weighted parsimony values

At least one solution of (1) is the result of the minimization of a parametric linear function on the (unknown) convex hull $\mathcal{H}$ of its integer solutions: in other words, a linear program on $\mathcal{H}$, which is a polyhedron with integer extreme points. It is well-known [6] that linear programs attain at least one of their optima at extreme points of the polyhedron defined by their constraints. If we fix the variables of (1) to an integer extreme point vector $\bar{x}$ of $\mathcal{H}$, the objective function of (1) becomes a linear function of w. To each extreme point there corresponds therefore a linear function $P_w(\bar{x})$ in w. For any w, the optimal solution of (1) is on the lower envelope of this set of linear functions; i.e., on a concave piecewise linear function.

It follows that there is a sequence of consecutive intervals of w (possibly reduced to a point) such that, for each successive interval, there is a solution of (1) which is optimal in the whole interval. The problem is then to determine all breakpoints of the curve P_w in function of w, i.e., the lowest points of intersection of the lines

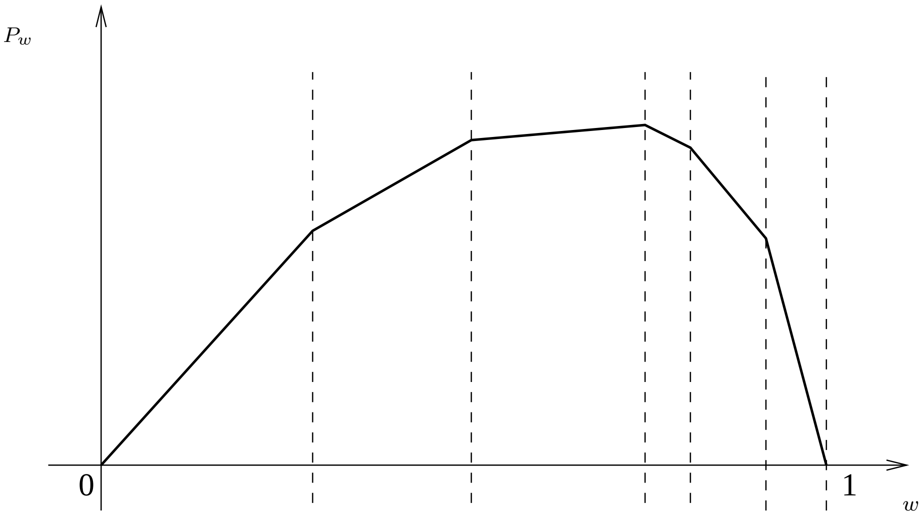


Fig. 1 P_w is a concave piecewise linear function.

$P_w(\bar{x})$ as functions of w for a given partition $\bar{x}$, as w ranges between 0 and 1 (see Fig. 1).

2.4 Theorectical analysis

At a generic iteration t of our algorithm, we have a value w_t, a corresponding partition x^t and its weighted parsimony value $P^t = P_{w_t}(x^t)$. We determine whether x^t is optimal for w by solving (1) for $w = w_t$, and update x^t and P^t accordingly. Next, we determine whether w_t is the next breakpoint after w_{t-1}: we compute the intersection w^* of the two lines at w_{t-1} and w_t defined respectively by $P_w(x^{t-1})$ and $P_w(x^t)$ (see Fig.2, left) and a corresponding optimal partition x^* with weighted parsimony P^*, using (1) for $w = w^*$. Now there are three cases for P^*: (a) it is at the top end of the interval of possible values; (b) it is at the bottom end; (c) it lies between the two interval endpoints (see Fig.2, right).

In case (a), w^* is the next breakpoint after w_{t-1} and w_t is the next breakpoint after w^*: for suppose there were a different breakpoint $\tilde{w}$ between w_{t-1} and w^*, then its optimal parametric parsimony value $\tilde{P}$ would be greater than $P' = P_{\tilde{w}}(x^{t-1})$; this would mean that x^{t-1} is a better partition than the one corresponding to $\tilde{P}$, contradicting optimality of $\tilde{P}$ (see Fig. 3). The argument when $\tilde{w}$ lies between w^* and w_t is similar.

In case (b), w^* is not a breakpoint, so w_t is the next breakpoint after w_{t-1}: for, suppose it were not, then the next breakpoint after w_{t-1} would be smaller than w_t, say $\tilde{w}$ with associated optimal parametric parsimony value $\tilde{P}$. This breakpoint would define a nonconcave piecewise linear function P_w, as shown in Fig. 4.

In case (c), w^* may or may not be a breakpoint. In such cases, we update $w_t = w^*$ and repeat.

In our approach, we find values of w corresponding to putative breakpoints in increasing order of w (we allow backtracking as explained above). In general, in

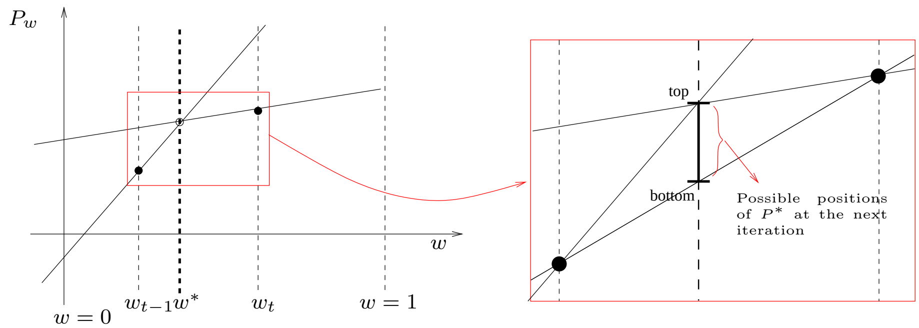


Fig. 2 Finding the next breakpoint. The optimal weighted parsimony value must be on the emphasized segment in the right hand side frame.

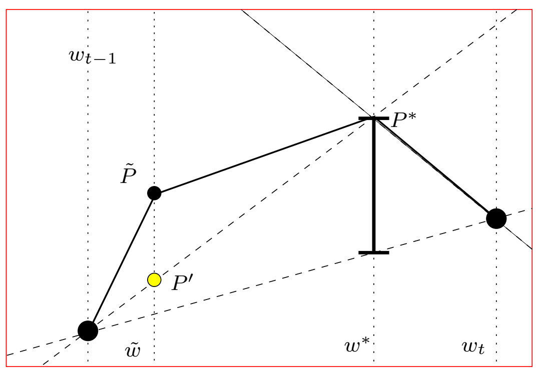


Fig. 3 A proof sketch for case (a).

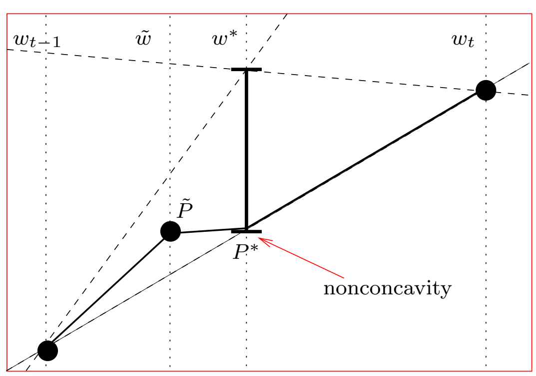


Fig. 4 A proof sketch for case (b).

order to find a value $w_t > w_{t-1}$ at the next iteration, we use an agglomerative approach: we find the smallest value of w for which it is worthwhile to merge two communities. Consider then two communities C_r and C_s. When merging them the change in weighted parsimony will be

$$\Delta P_w = w(-\sum_{i\in C_r}\sum_{j\in C_s} A_{ij}) + (1-w)(|C_r||C_s| - \sum_{i\in C_r}\sum_{j\in C_s} A_{ij})$$

where the first term corresponds to the edges of G which were previously deleted and are now inserted again, and the second term corresponds to those edges added between a vertex of C_r and one of C_s. The above formula can be simplified:

$$\Delta P_w = -w|C_r||C_s| + |C_r||C_s| - \sum_{i\in C_r}\sum_{j\in C_s} A_{ij}.$$

So a merging of two clusters of the current partition will only be worthwhile if

$$w \geq \frac{|C_r||C_s| - \sum_{i\in C_r}\sum_{j\in C_s} A_{ij}}{|C_r||C_s|}. \quad (2)$$

The right hand side of this last expression will be used as a tentative value for the parameter of weighted parsimony to be a breakpoint. Then the arguments presented above will be applied to explore the interval of parameter values between the two last breakpoints.

The steps of the exact algorithm are as follows:

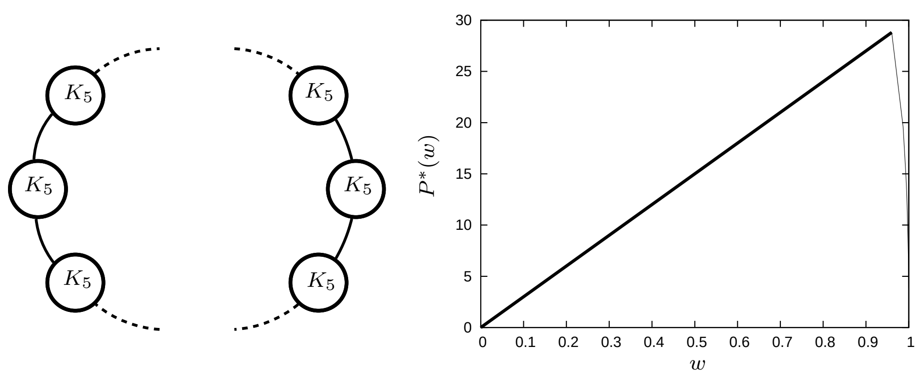
1. **Initialization.** Set $t = 1$ and $w_t = 0$; consider the initial solution x^t with n communities, each containing one entity, and a value $P^t = 0$.
2. **Tentative optimal solution.** If x^t has a single community print all values of w_t, P^t and the corresponding partitions x^t, then stop. Otherwise, increase t by 1. Consider the set of all pairs (C_r, C_s) of communities in the previous partition x^{t-1}. Compute the new tentative value w^t using (2). Let C_{r^*} and C_{s^*} be the two communities to be merged at level w_t. Obtain x^t by replacing C_{r^*} and C_{s^*} by their union in x^{t-1} and compute the new value $P_{w_t}(x^t) = \sum_{i\in V}\sum_{j\in V} A_{ij}(1-x_{ij}) + (1-w)\sum_{i\in V}\sum_{j\in V}(1-A_{ij})x_{ij}$.
3. **Optimality test**. Find the next breakpoint w_t after w_{t-1} using the arguments above, and update x^t and P^t. Then return to 2.

The algorithm terminates when all entities are in the same community, which will always be the case when $w_t = 1$. Termination is guaranteed because $w_t \leq w_{t-1}$, there is only a finite number of breakpoints and in case $w_t = w_{t-1}$ the algorithm does not cycle. We may have $w_t = w_{t-1}$ in two cases. The first is when the solution found by the agglomerative method is optimal for $w = w_{t-1}$. The value of the parameter may not change even for several iterations, but the number of communities is reduced by 1 at each iteration. The second case in which we may have $w_t = w_{t-1}$ is after a backtrack. Let $w^* > w_{t-1}$ be the value of the parameter corresponding to the putative breakpoint proposed by the agglomerative method (or obtained in a previous backtrack iteration) and x^* a corresponding optimal solution with parametric weighted parsimony P^*. We have $w_t = w_{t-1}$ if and only if x^* is optimal for values of the parameter in the interval $[w_{t-1}, w^*]$. This means there are no breakpoints between w_{t-1} and w^*. As a consequence, at iteration $t+1$ the agglomerative method, applied to x^*, will propose a putative breakpoint corresponding to a value of the parameter larger or equal to w^* and the actual breakpoint w_{t+1} found after the optimality check and the eventual backtracking phase will be $w_{t+1} \geq w^* > w_t$, otherwise the optimality of x^* for $w = w^*$ is contradicted.

3 Experiments

3.1 Artificial networks

We first tested our algorithm on four artificial networks from the literature. These networks where designed to verify whether a given method detects some “obvious” communities. The first one [7] consists in a ring of cliques, each joined to the next by a single edge. Specifically, we consider a ring of 30 cliques with 5 vertices each. This network was designed to illustrate the resolution limit of the modularity criterion. The function $P^*(w)$ is presented in Fig. 5. For $w = 0$ edges can be deleted without cost. The solution with minimum value 30 is obtained and corresponds to the partition into 30 cliques of order 5. It remains optimal for a very large interval of values of w i,e, $[0, 0.96]$. Then 15 pairs of cliques are merged one at a time, giving partitions into 29 to 15 communities. They have a common value of 28.8 and are optimal only for the interval reduced to the point $w = 0.9600$ except for the last one which is valid for the interval $[0.9600, 0.9867]$. The next partition consists in 10 communities of 15 vertices each i.e., merging 3 cliques at a time, 10 edges are removed and 730 edges are added. This partition is optimal for a small interval of value of w i.e., $[0.9867, 0.9933]$. The next partition into 8 communities is obtained by merging 4 communities 6 times and 2 communities twice. It is optimal for $0.9933 \leq w \leq 0.9950$. The remaining partitions are obtained by merging cliques in the most equal possible way (see Tab. 1 for details); so the weighted parsimony algorithm finds very quickly the structure of this network as well as a corresponding interval of values of the parameter.

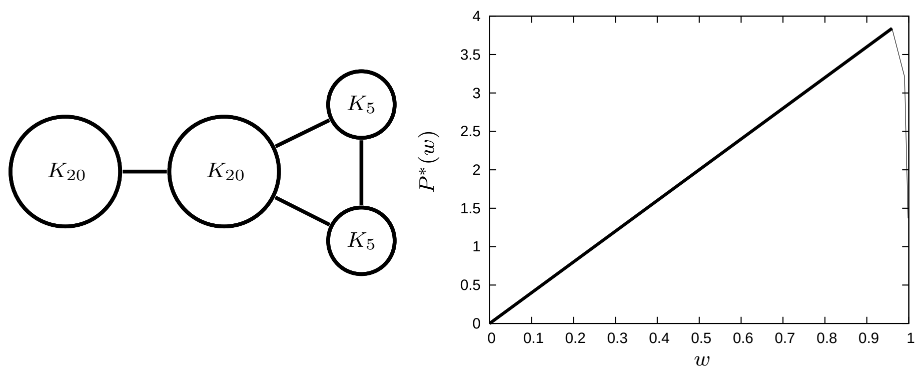


(a) (b)

Fig. 5 Optimal parametric curve for the ring of cliques.

The second network [7] consists of two cliques on 20 vertices joined by an edge and two smaller cliques on 5 vertices both joined by an edge between themselves and by an edge to the same large clique (see Fig. 6(a)). For this network there are only 4 weighted parsimony optimal partitions. As in the previous example the partition for $w = 0$ captures the structure of the network as it consists of the four cliques separated. Again the solution is optimal for $0.0 \leq w \leq 0.96$. At the next iteration the two small cliques are merged and this solution is optimal for w

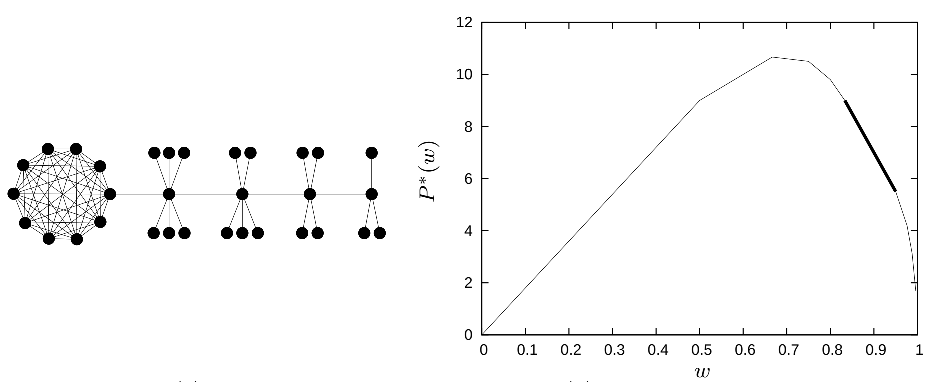
between 0.9600 and 0.9900. Then one of the two large cliques is merged with the union of the two small ones for a value of $P^* = 3.2100$ and finally all cliques are merged into one at $w = 0.9983$.



(a) (b)

Fig. 6 Optimal parametric curve for the 4 cliques network.

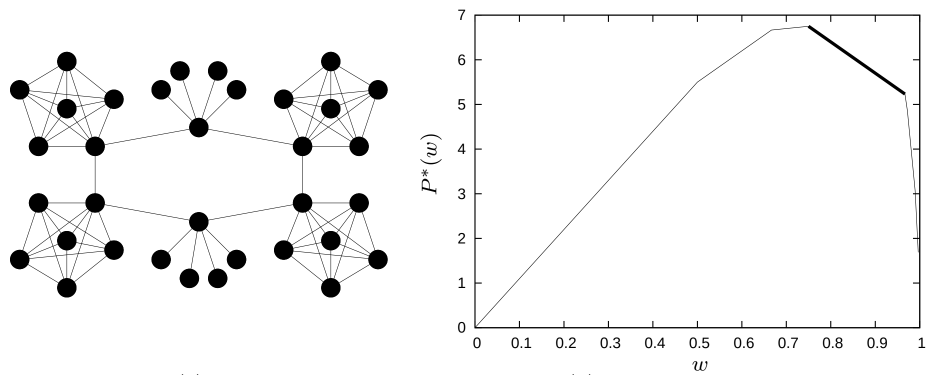
The next two artificial networks are built from cliques and stars connected by a chain or a cycle. The third network consists of a clique on 10 vertices and stars on 7, 6, 5, and 4 vertices each joined to the next one by an edge. This network was introduced in Ref. [2] in order to show the limits of parametric modularity methods based on Potts model [17]. The optimal partitions obtained are listed in Tab. 3. The optimal partition 15, into five clusters, has a value of 9, 4 edges being removed and 34 added. It is optimal for the very large interval $[0.8333, 0.9500]$.



(a) (b)

Fig. 7 Optimal parametric curve for the 5 modules network.

The fourth artificial network consists of four cliques on 6 vertices and two stars on 5 vertices joined by a 6-cycle (see Fig. 8(a)). It was introduced in Ref. [19] to illustrate the difference between the results obtained by parsimony and by weighted parsimony. Characteristics of the optimal partitions are given in Tab. 4. The partition into 6 communities captures exactly the structure of this network. It has a value of 6.75, which coincides with the value of the optimal partition into 7 communities, 5 edges are removed and 12 inserted.



(a) (b)

Fig. 8 Optimal parametric curve for the network with 4 cliques and 2 stars.

To study further the partition of graphs built from cliques and stars we consider a fifth network consisting of two cliques on 4 vertices joined by an edge and both of them joined by an edge to the center of a star on 18 vertices (see Fig. 9). This time the "obvious" structure consisting of the two 4-cliques and the star does not coincide with any optimal partition. For three communities the optimal solution corresponds to a community of 8 vertices obtained by joining the two small cliques, one isolated vertex and a star on 17 vertices, 3 edges are removed and 135 added. The solution is optimal for the w belonging to the interval $[0.9375, 0.9412]$. The partition into 4 communities consists of two cliques on 4 vertices, an isolated point and a star on 17 vertices, 4 edges are removed and 120 inserted. It is optimal for $w = 0.9375$.

3.2 Real world networks

The first example is the well-known karate club network of Zachary [18]. It has 34 vertices and 78 edges, corresponding to members of the club and friendship relations between them. At some time during Zachary's investigation a dispute arose between the administrator and the karate instructor and the club broke into two. It is a challenge for community detection criteria algorithms and heuristics to predict this bipartition from the previous data collected in the network. There are 26 optimal solutions and corresponding intervals. They are listed in Tab. 6. Observe that the number of communities is not monotonous in the parameter: indeed there are two partitions into 2 communities for disjoint intervals of w, and three partitions into 3 communities. The first partition into 2 communities (represented in Fig. 10(a)) is optimal for $0.9375 \leq w \leq 0.9569$. It almost reproduces

Iter.	w_{min}	$P^*(w)$	n. of comm.	edges removed	edges inserted
1	0.0000	0.0000	30	30	0
2	0.9600	28.8000	29	29	24
3	0.9600	28.8000	28	28	48
4	0.9600	28.8000	27	27	72
5	0.9600	28.8000	26	26	96
6	0.9600	28.8000	25	25	120
7	0.9600	28.8000	24	24	144
8	0.9600	28.8000	23	23	168
9	0.9600	28.8000	22	22	192
10	0.9600	28.8000	21	21	216
11	0.9600	28.8000	20	20	240
12	0.9600	28.8000	19	19	264
13	0.9600	28.8000	18	18	288
14	0.9600	28.8000	17	17	312
15	0.9600	28.8000	16	16	336
16	0.9600	28.8000	15	15	360
17	0.9867	19.6000	10	10	730
18	0.9933	14.8000	8	8	1028
19	0.9950	13.1000	7	7	1227
20	0.9960	11.8800	6	6	1476
21	0.9973	9.9200	5	5	1850
22	0.9983	8.2087	4	4	2424
23	0.9989	6.6162	3	3	3348
24	0.9995	4.7840	2	2	5222
25	0.9996	3.8560	1	0	10845

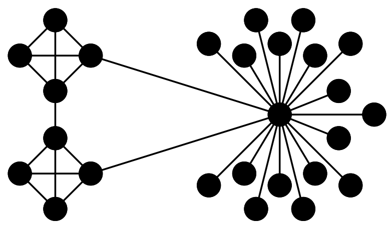
Table 1 Ring of cliques. Values of the optimal solution for parametric weighted parsimony.

Iter.	w_{min}	$P^*(w)$	n. of comm.	edges removed	edges inserted
1	0.0000	0.0000	4	4	0
2	0.9600	3.8400	3	3	24
3	0.9900	3.2100	2	1	222
4	0.9983	1.3683	1	0	821

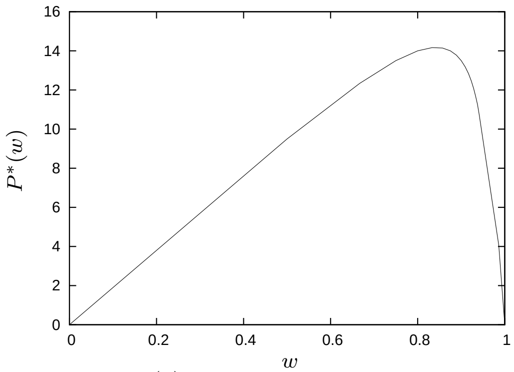
Table 2 Four cliques (2 large, 2 small). Values of the optimal solution for parametric weighted parsimony.

Iter.	w_{min}	$P^*(w)$	n. of comm.	edges removed	edges inserted
1	0.0000	0.0000	19	18	0
2	0.5000	9.0000	18	17	1
3	0.5000	9.0000	17	16	2
4	0.5000	9.0000	16	15	3
5	0.5000	9.0000	15	14	4
6	0.6667	10.6667	14	13	6
7	0.6667	10.6667	13	12	8
8	0.6667	10.6667	12	11	10
9	0.6667	10.6667	11	10	12
10	0.7500	10.5000	10	9	15
11	0.7500	10.5000	9	8	18
12	0.7500	10.5000	8	7	21
13	0.8000	9.8000	7	6	25
14	0.8000	9.8000	6	5	29
15	0.8333	9.0000	5	4	34
16	0.9500	5.5000	4	3	53
17	0.9762	4.1905	3	2	94
18	0.9878	3.1220	2	1	175
19	0.9961	1.6824	1	0	429

Table 3 Five modules. Values of the optimal solution for parametric weighted parsimony.



(a)



(b)

Fig. 9 Optimal parametric curve for the network with 2 cliques and 1 star.

Iter.	w_{min}	$P^*(w)$	n. of comm.	edges removed	edges inserted
1	0.0000	0.0000	12	11	0
2	0.5000	5.5000	11	10	1
3	0.5000	5.5000	10	9	2
4	0.6667	6.6667	9	8	4
5	0.6667	6.6667	8	7	6
6	0.7500	6.7500	7	6	9
7	0.7500	6.7500	6	5	12
8	0.9667	5.2333	5	4	41
9	0.9667	5.2333	4	3	70
10	0.9722	4.8611	3	2	105
11	0.9896	3.0729	2	1	200
12	0.9965	1.6886	1	0	488

Table 4 Four cliques and two stars. Values of the optimal solution for parametric weighted parsimony.

the observed split. Only the member 10 is misclassified, as was the case for several previous heuristics. As noted in Ref. [19], this entity was also considered as belonging to both clusters by several fuzzy partitioning methods [12,20,21]. The second partition into 2 communities, also represented in Fig. 10(b), is optimal for $0.9643 \leq w \leq 0.9724$ and it exhibits a small and dense cluster with 5 entities and attached to the remaining part by a cut vertex. The first partition into three communities, represented on Fig. 10(c), is optimal for $0.9091 \leq w \leq 0.9310$. It can be viewed as the intersection of the two partitions into 2 clusters as it corresponds to the first partition into 2 after the isolation of the small cluster found by the second partition into 2. The second and third partition into three communities are similar to the two partitions into two communities except for the fact that member 12 forms now an isolated community by himself. Partition into larger numbers of communities often present communities reduced to a single or small number of vertices.

A second real world network is the main component of the collaboration network of scientists at the Santa Fe Institute [8], a widely used test example for communities detection methods. It consists of 118 vertices and 200 edges. The optimal parametric curve is reported in Fig. 16 and the intervals are listed in

Iter.	w_{min}	$P^*(w)$	n. of comm.	edges removed	edges inserted
1	0.0000	0.0000	19	19	0
2	0.5000	9.5000	18	18	1
3	0.6667	12.3333	17	17	3
4	0.7500	13.5000	16	16	6
5	0.8000	14.0000	15	15	10
6	0.8333	14.1667	14	14	15
7	0.8571	14.1429	13	13	21
8	0.8750	14.0000	12	12	28
9	0.8889	13.7778	11	11	36
10	0.9000	13.5000	10	10	45
11	0.9091	13.1818	9	9	55
12	0.9167	12.8333	8	8	66
13	0.9231	12.4615	7	7	78
14	0.9286	12.0714	6	6	91
15	0.9333	11.6667	5	5	105
16	0.9375	11.2500	4	4	120
17	0.9375	11.2500	3	3	135
18	0.9412	10.7647	2	2	151
19	0.9861	4.0694	1	0	293

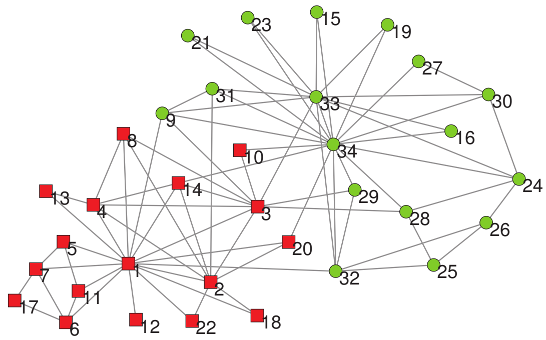
Table 5 Two K_4 cliques and a star with 18 entities. Values of the optimal solution for parametric weighted parsimony.

Table 7. A reasonable partition into 9 communities is found for values of w in $[0.9643, 0.9722]$. It is represented in Fig. 11(a). For w in $[0.9792, 0.9875]$ we obtain a partition into 6 communities which is very close to the one obtained by maximizing modularity (it differs for two vertices). The next interval $[0.9875, 0.9877]$ corresponds to another partition into 6 communities (see Fig. 11(c)). Optimal solutions for larger values of the parameters are obtained by merging communities of this partition.

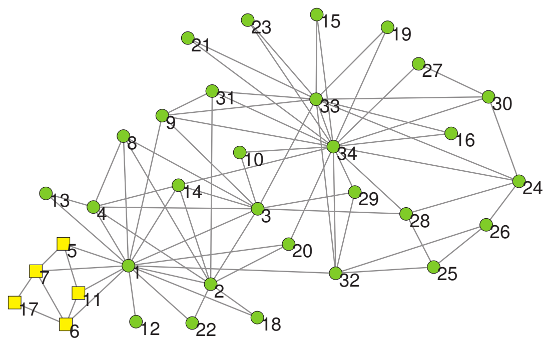
The third example is the game schedule of the 2000 season of Division I of the US college football league [8]. The 115 vertices represent the teams, while edges correspond to 613 games played between the two teams they connect during the year. Teams are grouped in 12 conferences of 8 to 12 teams each. Usually, games between members of the same conference are more frequent than games between teams of different conferences.

Three of the conferences are correctly identified even for $w = 0$: Atlantic Coast, Big East and Mountain West. The community corresponding to the Atlantic Coast conference appears in an optimal solution for values of the parameter in the range $[0, 0.9222]$. Two teams of the IA Independents conference join the community of the Big East for $w = 0.200$; the resulting community remains unchanged until it is merged with the Atlantic Coast for $w = 0.9222$. As noted in other papers [20,13, 19], the members of the IA Independents have more links with teams of the other conferences than internal edges, so it is not surprising if the five vertices of this conference are distributed to other communities. The Mountain West conference is exactly isolated for $0 \leq w \leq 0.8958$.

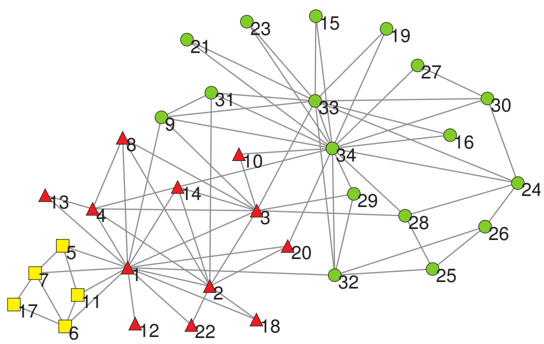
A partition into 13 communities is found for values of the parameter in $[0.6667, 0.6923]$; 7 conferences are correctly identified and 12 vertices are misclassified. The next 2 intervals, $[0.6923, 0.8913]$ and $[0.8913, 0.8958]$, corresponds to partitions into 12 and 11 communities respectively. In both of them 6 conferences are correctly iso-



(a) First partition into two communities.



(b) Second partition into 2 communities.

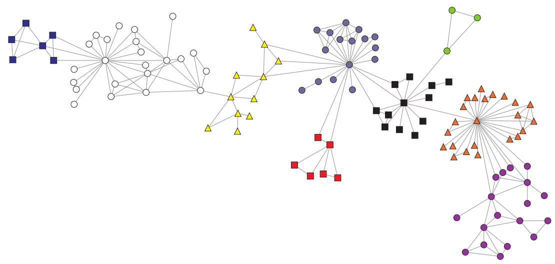


(c) First partition into three communities.

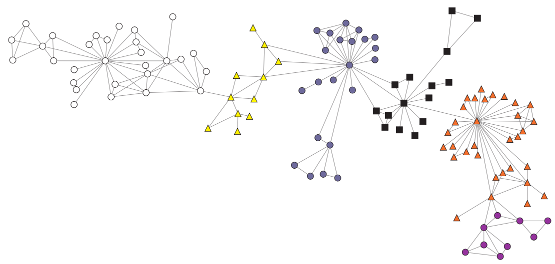
Fig. 10 Partitions of the Zachary network. (a) is optimal for values of w in the interval $[0.9375, 0.9569]$, (b) in $[0.9643, 0.9724]$ and (c) in $[0.9091, 0.9310]$.

lated and 11 vertices are misclassified. The partition into 11 clusters is represented in Fig. 12. All three seems reasonable partitions for the network.

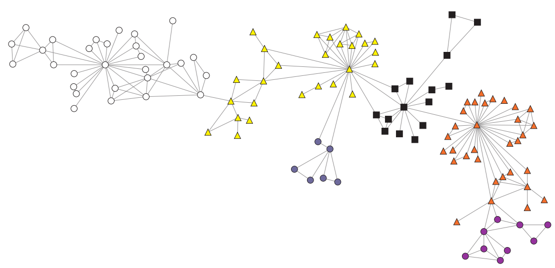
A fourth real-world network is the dolphins social network reported by Lousseau et al. [14]. It has 62 vertices and 159 edges. A partition into 2 groups of predominantly male and female dolphins respectively was described by Luosseau. There are 37 optimal partitions and in this case the number of clusters is monotonous up to 19 communities. Moreover there is a single optimal partition into 1 to 6



(a) Partition into 9 communities.



(b) First partition into 6 communities.



(c) Second partition into 6 communities.

Fig. 11 Partitions of the scientific collaboration network. (a) is optimal for values of w in the interval $[0.9643, 0.9722]$, (b) in $[0.9792, 0.9875]$ and (c) in $[0.9875, 0.9877]$.

communities. The partition into two communities observed by Zhang et. al [19], represented in Fig.13, is almost identical to the one described by Luosseau. Only dolphin 40 is misclassified and is connected to one vertex only of both communities. A recent overlapping community detection heuristic consider it to be shared between the two groups. The optimal partition in three to six groups consists of the partition into 2 with one to 4 “degenerate” communities, composed of single dolphins, detached from the largest community. They correspond to the 4 pending edges in that community. Partitions into a larger number of communities contain several 1 dolphin degenerate communities or other small communities. The smallest community among the partition into two remains untouched until the partition into 10 communities. This suggest that ties between members of that community are stronger than ties among members of the other one. This is an evident example of the additional information obtained by analysing the partitions obtained for different values of the w parameter, instead of looking at a single partition.

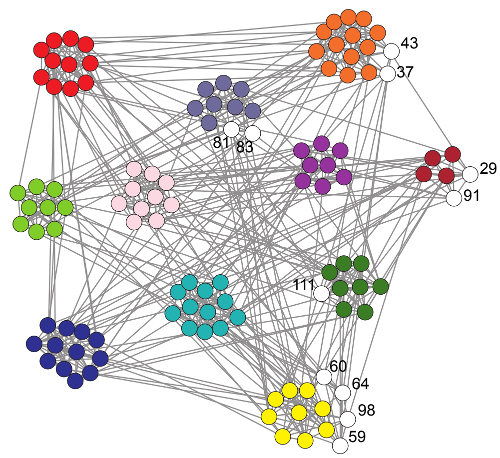


Fig. 12 College football network. Partition into 11 communities.

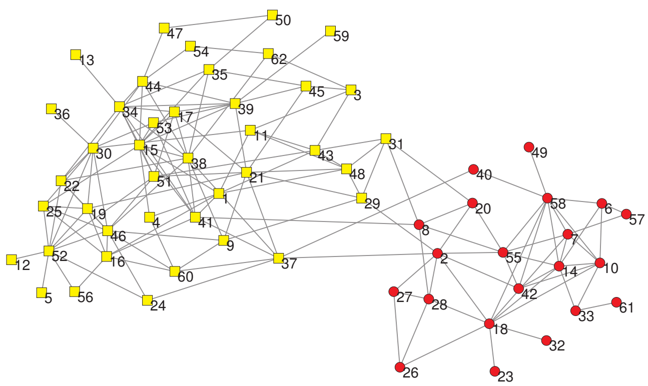


Fig. 13 Dolphin network. Partition into two communities.

We also considered a often studied real network which was not included into those reported on by Zhang et al. [19]. It describes the interactions between the characters in Victor Hugo's masterpiece *Les misérables* and has 77 vertices and 257 edges. There are 54 optimal partitions. The single partition into 2 communities separates neatly a community of 10 characters which are bishop Myriel and people he encountered during his long life. This partition is optimal for $0.9923 \leq w \leq 0.9955$. This appears to be the most obvious split and the 10 characters community remains unchanged until the partition into 18 communities. In the next partition Napoleon forms a one character degenerate community. Other partitions into small number of communities again have several one character or two characters degenerate communities. After sometimes the largest community splits into two. This happens in the first partition into 6 communities which is optimal for $w = 0.9804$ only. This partition is presented in Fig. 14. For this network it appears that the weighted parsimony criterion captures part of the structures but also exhibits many small communities, as did the (unweighted) parsimony criterion. Nevertheless, by look-

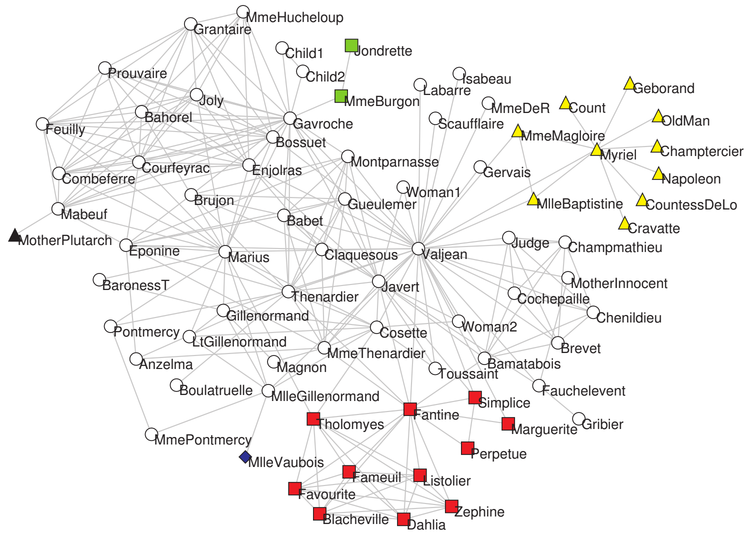


Fig. 14 *Les Misérables* network. First partition into six communities.

ing at all the optimal partitions we are able to gain information on the strenght of some communities.

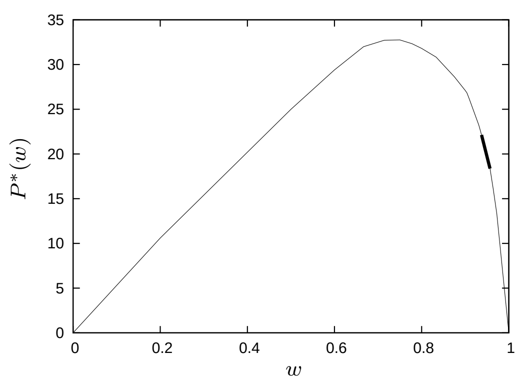


Fig. 15 Optimal parametric curve for the Zachary network.

4 Choosing a value for the parameter

Zhang, Qiu and Zhang [19] do not determine solutions for all values of the parameter $\bar{w} = 1 - w$ for the weighted parsimony criterion. Instead, they give three formulæ for the choice of the best value of $\bar{w}$. To this effect they first argue that $\bar{w}$ should increase with the average density of the network. More precisely they introduce the value $D_e = m/n$ (1/2 of the average degree) and normalize it to

Iter.	w_{min}	$P^*(w)$	n. of comm.	edges removed	edges inserted
1	0.0000	0.0000	20	53	0
2	0.2000	10.6000	19	49	1
3	0.5000	25.0000	18	47	3
4	0.6000	29.4000	17	45	6
5	0.6667	32.0000	16	43	10
6	0.6667	32.0000	15	41	14
7	0.6667	32.0000	14	39	18
8	0.6667	32.0000	13	37	22
9	0.7143	32.7143	12	35	27
10	0.7143	32.7143	11	33	32
11	0.7500	32.7500	10	31	38
12	0.7500	32.7500	9	29	44
13	0.7778	32.3333	8	27	51
14	0.8000	31.8000	7	26	55
15	0.8333	30.8333	6	23	70
16	0.8333	30.8333	6	22	75
17	0.8750	28.6250	6	21	82
18	0.9000	27.1000	5	20	91
19	0.9038	26.8269	4	15	138
20	0.9091	26.1818	3	14	148
21	0.9310	23.2414	4	12	175
22	0.9333	22.8667	3	11	189
23	0.9375	22.1250	2	10	204
24	0.9569	18.3621	3	5	315
25	0.9643	16.0714	2	4	342
26	0.9724	13.3241	1	0	483

Table 6 Zachary network. Values of the optimal solution for parametric weighted parsimony.

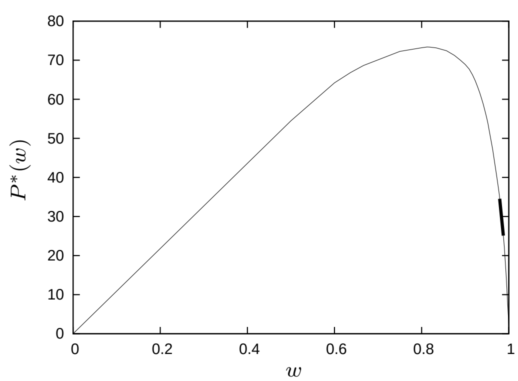


Fig. 16 Optimal parametric curve for the scientific collaboration network.

$\bar{D}_e = 1 - 1/D_e$ that is $0 \leq \bar{D}_e \leq 1$. Similarly these authors argue that $\bar{w}$ should increase with the clustering coefficient C. Recall that $C_i = 2K_i/(k_i(k_i - 1))$ for a vertex i, where K_i is the number of edges joining neighbors of a vertex i and k_i is the degree of i (or in other words its number of neighbors), and $C = (1/n)\sum_{i=1}^{n} C_i$ for the network G itself. Formulae for the optimal value of $\bar{w}$ according to interval of values for $\bar{D}_e$ and C are given for three cases and are reproduced in Tab. 11

Note that no formula is proposed for the case $\bar{D}_e <= 0.5$ and $C <= 0.5$. which does not arise for any of the five problems considered. While indeed seems

Iter.	w_{min}	$P^*(w)$	n. of comm.	edges removed	edges inserted
1	0.0000	0.0000	65	109	0
2	0.5000	54.5000	64	107	2
3	0.5000	54.5000	63	106	3
4	0.5000	54.5000	60	103	6
5	0.6000	64.2000	59	101	9
6	0.6000	64.2000	58	95	18
7	0.6000	64.2000	57	93	21
8	0.6364	66.8182	56	89	28
9	0.6667	68.6667	55	87	32
10	0.6667	68.6667	54	86	34
11	0.6667	68.6667	53	85	36
12	0.6667	68.6667	52	84	38
13	0.6667	68.6667	52	83	40
14	0.7500	72.2500	51	82	43
15	0.7500	72.2500	50	80	49
16	0.7500	72.2500	49	78	55
17	0.7500	72.2500	49	77	58
18	0.8000	73.2000	48	76	62
19	0.8125	73.3750	48	73	75
20	0.8182	73.3636	47	71	84
21	0.8333	73.1667	46	70	89
22	0.8333	73.1667	45	68	99
23	0.8571	72.4286	44	67	105
24	0.8571	72.4286	43	66	111
25	0.8571	72.4286	43	63	129
26	0.8750	71.2500	42	62	136
27	0.8750	71.2500	41	61	143
28	0.8750	71.2500	41	60	150
29	0.8889	70.0000	40	59	158
30	0.9000	68.9000	39	58	167
31	0.9000	68.9000	38	56	185
32	0.9091	67.7273	37	55	195
33	0.9091	67.7273	36	54	205
34	0.9091	67.7273	31	50	245
35	0.9167	66.2500	30	49	256
36	0.9167	66.2500	29	48	267
37	0.9167	66.2500	29	46	289
38	0.9231	64.6923	28	45	301
39	0.9231	64.6923	27	44	313
40	0.9231	64.6923	27	43	325
41	0.9286	63.1429	26	41	351
42	0.9333	61.6667	25	40	365
43	0.9333	61.6667	24	39	379
44	0.9333	61.6667	24	38	393
45	0.9375	60.1875	23	37	408
46	0.9412	58.8235	22	36	424
47	0.9412	58.8235	21	35	440
48	0.9412	58.8235	21	34	456
49	0.9444	57.4444	20	33	473
50	0.9474	56.1579	19	32	491
51	0.9474	56.1579	18	31	509
52	0.9500	54.9000	17	30	528
53	0.9500	54.9000	16	29	547
54	0.9500	54.9000	15	28	566
55	0.9524	53.6190	14	27	586
56	0.9524	53.6190	13	26	606
57	0.9545	52.3636	12	25	627
58	0.9545	52.3636	11	24	648
59	0.9630	47.1111	10	21	726
60	0.9643	46.1786	9	20	753
61	0.9722	40.3611	8	19	788
62	0.9762	37.3095	8	16	911
63	0.9783	35.4565	7	13	1046
64	0.9792	34.5208	6	11	1140
65	0.9875	25.1125	6	10	1219
66	0.9877	24.9259	5	8	1379
67	0.9896	22.2813	4	5	1664
68	0.9939	15.0545	3	2	2156
69	0.9993	3.5474	2	1	3547
70	0.9997	2.1232	1	0	6703

Table 7 Scientific collaboration network. Values of the optimal solution for parametric weighted parsimony.

Iter.	w_{min}	$P^*(w)$	n. of comm.	edges removed	edges inserted
1	0.0000	0.0000	22	311	0
2	0.2000	62.2000	21	307	1
3	0.2000	62.2000	20	287	6
4	0.2500	76.2500	19	272	11
5	0.3333	98.0000	18	252	21
6	0.5000	136.5000	17	234	39
7	0.5000	136.5000	16	216	57
8	0.5000	136.5000	15	214	59
9	0.6250	155.8750	14	208	69
10	0.6667	161.6667	13	194	97
11	0.6923	164.1538	12	190	106
12	0.8913	180.8696	11	185	147
13	0.8958	181.0417	10	180	190
14	0.9063	180.9375	10	177	219
15	0.9074	180.8889	9	167	317
16	0.9167	179.5000	8	154	460
17	0.9222	177.8000	7	147	543
18	0.9394	171.0000	6	137	698
19	0.9444	168.1667	5	131	800
20	0.9444	168.1667	5	125	902
21	0.9545	160.3182	4	107	1280
22	0.9616	152.0242	3	88	1756
23	0.9721	134.5260	3	88	1756
24	0.9736	130.4960	2	61	2697
25	0.9808	111.6923	2	58	2850
26	0.9816	109.4083	1	0	5942

Table 8 College football network. Values of the optimal solution for parametric weighted parsimony.

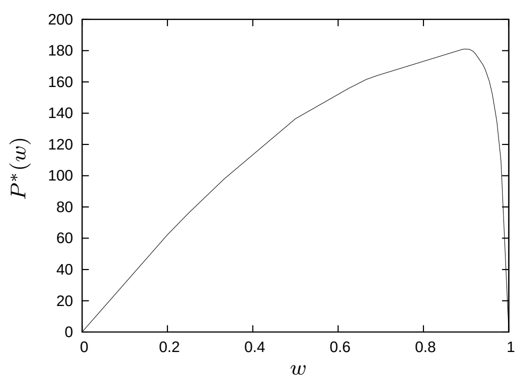


Fig. 17 Optimal parametric curve for the college football network.

plausible that $\bar{w}$ should increase with $\bar{D}_e$ and C no *a priory* justification is given for the precise form of the three complicated formulæ of Tab. 11, although they give good results *a posteriori* i.e, point values in the intervals corresponding to plausible partitions. It is possible to still have this property with much simpler formulæ e.g. linear expressions in $\bar{D}_e$ and C. For the first two problems, ring of cliques and star-shapes and karate club, $\bar{D}_e \geq 0.5$ and $C \geq 0.5$, the intervals of values for w corresponding to the best partitions are $[0.75, 0.9667]$ and $[0.9091, 0.9310]$.

Iter.	w_{min}	$P^*(w)$	n. of comm.	edges removed	edges inserted
1	0.0000	0.0000	29	103	0
2	0.2000	20.6000	28	99	1
3	0.2500	25.5000	27	96	2
4	0.4444	43.7778	26	91	6
5	0.5000	48.5000	25	90	7
6	0.5000	48.5000	24	83	14
7	0.6000	55.4000	24	81	17
8	0.6250	57.0000	23	78	22
9	0.6667	59.3333	22	72	34
10	0.7143	61.1429	21	70	39
11	0.7500	62.2500	20	68	45
12	0.7500	62.2500	19	67	48
13	0.7500	62.2500	20	66	51
14	0.7647	62.4706	19	62	64
15	0.7857	62.4286	18	59	75
16	0.8235	61.8235	18	56	89
17	0.8704	60.2778	17	49	136
18	0.8889	58.6667	16	46	160
19	0.9015	57.2273	16	33	279
20	0.9167	53.5000	15	31	301
21	0.9286	50.2857	14	30	314
22	0.9286	50.2857	14	29	327
23	0.9375	47.6250	13	28	342
24	0.9412	46.4706	12	27	358
25	0.9444	45.3889	11	26	375
26	0.9474	44.3684	10	25	393
27	0.9500	43.4000	9	24	412
28	0.9503	43.2795	9	16	565
29	0.9531	41.7344	8	13	626
30	0.9706	31.0294	7	12	659
31	0.9706	31.0294	7	11	692
32	0.9722	29.9167	6	10	727
33	0.9730	29.3784	5	9	763
34	0.9737	28.8421	4	8	800
35	0.9744	28.3077	3	7	838
36	0.9750	27.7750	2	6	877
37	0.9930	12.0697	1	0	1732

Table 9 Dolphin. Values of the optimal solution for parametric weighted parsimony.

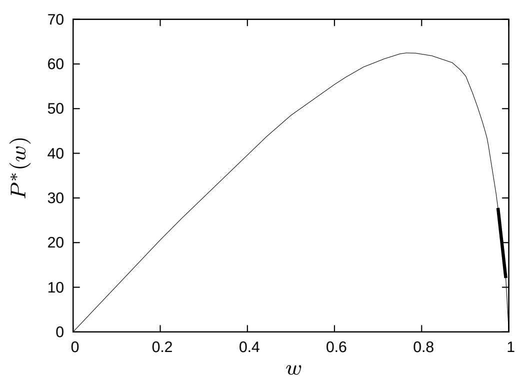


Fig. 18 Optimal parametric curve for the dolphin network.

Consider then the following linear program in the variables w_1, w_2, y_1 and y_2 as

Iter.	w_{min}	$P^*(w)$	n. of comm.	edges removed	edges inserted
1	0.0000	0.0000	35	118	0
2	0.2222	26.2222	35	111	2
3	0.2857	33.1429	35	106	4
4	0.3077	35.3846	35	97	8
5	0.4167	45.0833	34	90	13
6	0.5000	51.5000	33	89	14
7	0.6667	64.0000	32	88	16
8	0.6667	64.0000	31	85	22
9	0.7500	69.2500	30	84	25
10	0.7500	69.2500	29	83	28
11	0.7500	69.2500	29	82	31
12	0.7778	70.6667	26	74	59
13	0.8000	71.0000	25	73	63
14	0.8000	71.0000	24	72	67
15	0.8333	71.1667	23	71	72
16	0.8571	71.1429	22	70	78
17	0.8571	71.1429	21	68	90
18	0.8750	70.7500	20	67	97
19	0.8750	70.7500	19	66	104
20	0.8750	70.7500	19	65	111
21	0.8889	70.1111	18	64	119
22	0.9000	69.5000	17	63	128
23	0.9107	68.8036	17	48	281
24	0.9167	67.4167	16	47	292
25	0.9259	65.1481	15	45	317
26	0.9333	63.1333	18	41	373
27	0.9355	62.4194	17	39	402
28	0.9375	61.6875	16	37	432
29	0.9545	54.9545	15	34	495
30	0.9667	49.3667	14	27	698
31	0.9750	43.7750	14	25	776
32	0.9767	42.4651	13	24	818
33	0.9767	42.4651	13	23	860
34	0.9778	41.6000	12	22	904
35	0.9783	41.1739	11	21	949
36	0.9787	40.7447	10	20	995
37	0.9792	40.3125	9	19	1042
38	0.9796	39.8776	8	18	1090
39	0.9800	39.4400	7	17	1139
40	0.9804	39.0000	6	16	1189
41	0.9804	39.0000	13	15	1239
42	0.9815	37.6667	12	14	1292
43	0.9815	37.6667	12	13	1345
44	0.9821	36.7857	11	12	1400
45	0.9825	36.3509	10	11	1456
46	0.9828	35.9138	9	10	1513
47	0.9831	35.4746	8	9	1571
48	0.9833	35.0333	7	8	1630
49	0.9836	34.5902	6	7	1690
50	0.9839	34.1452	5	6	1751
51	0.9841	33.6984	4	5	1813
52	0.9844	33.2500	3	4	1876
53	0.9923	18.4000	2	3	2005
54	0.9955	11.9642	1	0	2672

Table 10 Les Misérables. Values of the optimal solution for parametric weighted parsimony.

well as slacks variables for lower and upper bounds s^l and s^u

$$\begin{aligned}
&\min s \\
&s.t.\ w_1 = 0.5641y_1 + 0.5706y_2 \\
&\quad 0.9091 \leq w_1 \leq 0.9310 \\
&\quad w1 + s_1^l - s_1^u = 0.92005 \\
&\quad w2 = 0.5405y_1 + 0.6443y_2 \\
&\quad 0.75 \leq w2 \leq 0.9667 \\
&\quad w_2 + s_2^u - s_2^l = 0.85835 \\
&\quad s \geq s_1^l \\
&\quad s \geq s_2^l \\
&\quad s \geq s_1^u \\
&\quad s \geq s_2^u
\end{aligned} \tag{3}$$

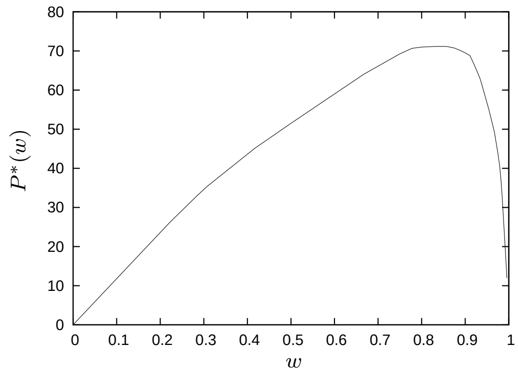


Fig. 19 Optimal parametric curve for *Les Misérables* network.

$\bar{d}_e$	C	$\bar{w}$
≥ 0.5	≥ 0.5	$\bar{w} = \frac{1}{2}(\bar{D}_e)^2 C$
≤ 0.5	≥ 0.5	$\bar{w} = \frac{1}{2}(\bar{D}_e)^{\frac{1}{\bar{D}_e}} C$
≥ 0.5	≤ 0.5	$\bar{w} = \frac{1}{2}(\bar{D}_e)^2 C^{\frac{0.5}{C}}$

Table 11 Formulae proposed by Zhang et al. [19] to choose a value for the parameter $\bar{w}$.

The aim of this program is to find values for w_1 and w_2 as close as possible to the midpoints of the corresponding intervals, i.e., $w_1 = 0.92005$ and $w_2 = 0.85835$. In this case the solution is such that all the departures from those values are equal to 0. Similar results are obtained for the scientific collaboration network (the only instance considered with $\bar{D}_e \leq 0.5$ and $C \geq 0.5$) and for the football and dolphins network ($\bar{D}e \geq 0.5$ and $C \leq 0.5$). One finds $w_3 = 0.96825$ for the scientific collaboration network. Finally $w_4 = 0.89355$ and $w_5 = 0.984$ for the football and dolphins networks. Note that this approach can be extended to more than two instances at a time but the probability of finding a feasible solution decreases rapidly.

5 Conclusion

In a recent paper Zhang, Qiu and Zhang [19] extended the parsimony criterion for detecting community structures to weighted parsimony. This gives partitions with fewer communities, which often appear to be more plausible and informative than those obtained with the usual parsimony criterion. After formulating the weighted parsimony problem, these authors propose a simulated annealing heuristic for obtaining an optimal or near-optimal solution for a particular value of the parameter. Three formulæ for choosing the value of the parameter according to the values of the average distance and the clustering coefficient are presented .

This paper further explores communities detection in networks according to the weighted parsimony criterion. The curve of weighted parsimony values is shown to be a piecewise concave function of the parameter with a finite (and usually moderate) number of breakpoints. A parametric integer program is proposed for finding all breakpoints of this curve as well as the corresponding optimal partitions and the intervals of values in which they remain optimal. Experimental results confirms those of Zhang et al. [19] and also show that several partitions into a small number of communities may be of interest. The question of choosing *a priori* a good value for the parameter is also discussed. At the end of the paper Zhang et al. mention as future work the study of the theoretical properties of the weight parameter w, the robustness of the simulated annealing heuristic and solving larger instances. We believe that the characterization of the curve of optimal values of the parametric weighted parsimony model is an important step regarding the first point. Similarly the proposed parametric integer program provides optimal solutions for all values of the parameter. This is done with a guarantee of optimality unlike previous heuristics. Finally, the size of the problems solved is similar (and often the same) as those previously considered. Clearly solving larger instances would be of interest. One possible way to do so would be to replace in the proposed parametric algorithm the clique partitioning routine [11, 10] by a stabilized column generation one [1].

References

1. D. Aloise, S. Cafieri, G. Caporossi, P. Hansen, , L. Liberti, and S. Perron. Column generation algorithms for exact modularity maximization in networks. *Physical Review E*, 82:046112, 2010.
2. A. Arenas, A. Fernàndez, and S. Gòmez. Analysis of the structure of complex networks at different resolution levels. *New Journal of Physics*, 10:053039, 2008.
3. A Bettinelli, P. Hansen, and L. Liberti. Algorithm for parametric community detection in networks. *Phys. Rev. E*, 86:016107, 2012.
4. S. Cafieri, P. Hansen, and L. Liberti. Loops and multiple edges in modularity maximization of networks. *Physical Review E*, 81(4):46102, 2010.
5. E. Dahlhaus, D. S. Johnson, C. H. Papadimitriou, P. D. Seymour, and M. Yannakakis. The complexity of multiterminal cuts. *SIAM Journal on Computing*, 23:864–894, 1994.
6. G.B. Dantzig. *Linear Programming and Extensions.* Princeton University Press, Princeton, NJ, 1963.
7. S. Fortunato and M. Barthelemy. Resolution limit in community detection. *PNAS USA*, 104:36, 2007.
8. M. Girvan and M.E.J. Newman. Community structure in social and biological networks. *Proc. Natl. Acad. Sci.*, 99:7821–7826, 2002.
9. B.H. Good, Y.-A. de Montjoye, and A. Clauset. Performance of modularity maximization in practical contexts. *Physical Review E*, 81(4):046106, 2010.
10. M. Grötschel and Y. Wakabayashi. A cutting plane algorithm for a clustering problem. *Mathematical Programming B*, 45:59–96, 1989.
11. M. Grötschel and Y. Wakabayashi. Facets of the clique partitioning polytope. *Mathematical Programming*, 47:367–387, 1990.
12. Andrea Lancichinetti, Santo Fortunato, and Filippo Radicchi. Benchmark graphs for testing community detection algorithms. *Physical Review E*, 78:046110, 2008.
13. Z. Li, S Zhang, RS WAng, XS Zhang, and L Chen. Quantitative function for community detection. *Phys. Rev E*, 77:036109, 2008.
14. D. Lusseau, K. Schneider, O. J. Boisseau, P. Haase, E. Slooten, and S. M. Dawson. The bottlenose dolphin community of doubtful sound features a large proportion of long-lasting associations. *Behavioral Ecology and Sociobiology*, 54:396–405, 2003.
15. C.P. Massen and J.P.K. Doye. Identifying communities within energy landscapes. *Physical Review E*, 71(046101), 2005.
16. M. E. J. Newman and M. Girvan. Finding and evaluating community structure in networks. *Physical Review E*, 69:026113, 2004.
17. J. Reichardt and S. Bornholdt. Statistical mechanics of community detection. *Physical Review E*, 74:016110, 2006.

18. W. Zachary. *Journal of Anthropological Research*, 33:452, 1977.
19. J. Zhang, Y Qiu, and X.-S. Zhang. Detecting community structure: from parsimony to weighted parsimony. *J Sys Sci Complex*, 23:1024–1036, 2010.
20. S. Zhang, R. S. Wang, and X. S. Zhang. Identication of overlapping community structure in complex networks using fuzzy c-means clustering. *Physica A*, 374:483–490, 2007.
21. S. Zhang, R. S. Wang, and X. S. Zhang. Uncovering fuzzy community structure in complex networks. *Phys. Rev. E*, 76:046103, 2007.