



**HAL**  
open science

## Extracting Modular-based Backbones in Weighted Networks

Zakariya Ghalmane, Chantal Cherifi, Hocine Cherifi, Mohammed El Hassouni

► **To cite this version:**

Zakariya Ghalmane, Chantal Cherifi, Hocine Cherifi, Mohammed El Hassouni. Extracting Modular-based Backbones in Weighted Networks. *Information Sciences*, 2021, 576, pp.454-474. <10.1016/j.ins.2021.06.087>. <hal-04094168>

**HAL Id: hal-04094168**

**<https://hal.science/hal-04094168v1>**

Submitted on 10 May 2023

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



HAL Authorization

# Extracting Modular-based Backbones in Weighted Networks

Zakariya Ghalmane<sup>a,\*</sup>, Chantal Cherifi<sup>b</sup>, Hocine Cherifi<sup>c</sup> and Mohammed El Hassouni<sup>d</sup>

<sup>a</sup>LIRMM, University of Montpellier, Montpellier, France

<sup>b</sup>DISP Lab, University of Lyon 2, Lyon, France

<sup>c</sup>LIB EA 7534, University of Burgundy, Dijon, France

<sup>d</sup>FLSH, Mohammed V University in Rabat, Rabat, Morocco

---

## ARTICLE INFO

*Keywords:*  
Backbone  
Modular structure  
Weighted network

## ABSTRACT

Networks are an adequate representation for modeling and analyzing a great variety of complex systems. However, understanding networks with millions of nodes and billions of connections can be pretty challenging due to memory and time constraints. Therefore, selecting the relevant nodes and edges of these large-scale networks while preserving their core information is a major issue. In most cases, the so-called backbone extraction methods are based either on coarse-graining or filtering approaches. Coarse-graining techniques reduce the network size by gathering similar nodes into super-nodes, while filter-based methods eliminate nodes or edges according to a statistical property. In this work, a filter-based method is proposed and investigated. It uses the overlapping community structure to build the backbone in weighted networks. While most filtering techniques rely on link features to extract the backbone, the proposed method exploits both nodes and links. It takes advantage of the network communities through their main features (overlapping nodes, hubs, and bridging connections) to select influential edges and nodes while preserving the ability of the information dissemination of the original network. The so-called "Modular filtering backbone" combines two components. The first one is the network connecting the overlapping nodes and the top connected nodes (also called the hubs). One discards the edges with the lowest weights as long as connected components are maintained. The second component uses the network of the inter-community links with the nodes at their extremities. The disparity filter algorithm allows preserving only its most crucial connections. An extensive investigation is performed on a set of real-world weighted networks of various sizes and a wide range of origin. Results show the advantage of the proposed method over alternative filtering-based methods used for comparative purposes. Furthermore, this sheds new light on the most relevant parts of empirical networks hidden by their complexity.


---

## 1. Introduction

Scholars have studied complex networks to understand systems involving interacting components. They have permeated a broad spectrum of fields, including biological, transportation, social, and technological systems (1; 2). Exploring the structural properties of complex systems is one of the primary purposes of complex networks. Some common characteristics are the small-world property, the scale-free behavior of the degree distribution and the community structure (3; 4; 5; 6). As real-world systems become increasingly large, identifying and analyzing meaningful information is more and more a challenging issue. Little redundant intricacies usually submerge this relevant information. Thus, it is of prime interest to find the right trade-off between reducing the size of the network while preserving a decent amount of relevant information. Recently, there has been growing attention to the extraction of backbones in large-scale networks. Some works are devoted to mono-mode networks (7; 8; 9; 10) while others are tailored to bipartite networks (11; 12; 13; 14; 15). This work focuses on one-mode networks. One may classify existing methods into two categories: coarse-graining and filter-based approaches. Coarse-graining methods (16; 17; 18) gather nodes with some shared specific characteristics. They represent such communities or groups as a single node to reduce the network while maintaining its properties. Usually, the common property held by the reassembled nodes is that they are densely interconnected compared to nodes outside of their community. In (18), the authors describe a coarse-graining approach that maintains random walks. It groups nodes with roughly the same neighbors. Indeed, from the perspective of a random walk, they cannot be identified. Filter-based approaches adopt a bottom-up strategy to uncover the

---

\*Corresponding author

 zakariya.ghalmane@um5s.net.ma (Z. Ghalmane); Chantal.BonnerCherifi@univ-lyon2.fr (C. Cherifi); hocine.cherifi@u-bourgogne.fr (H. Cherifi); mohamed.elhassouni@flsh.um5.ac.ma (M. El Hassouni)

ORCID(s): 0000-0002-2440-2886 (Z. Ghalmane); 0000-0001-9124-4921 (H. Cherifi); 0000-0002-6741-4799 (M. El Hassouni)

backbone (19; 20; 21; 22; 23; 24; 25; 26). A statistical feature of a node or a link is specified. It is then used as a criterion for retaining or discarding nodes or links from the original network. As a consequence, only relevant nodes and links according to the defined property compose the backbone. Filter-based approaches use either global, local, or both types of information to filter nodes and links. Link betweenness (19) is used as a global measure in some filter-based methods. This strategy retains in the backbone the links with betweenness above a threshold. The link salience is another approach relying on a global measure (20). It depends on the shortest-path tree, which summarizes the shortest paths from a source vertex to the rest of the network. To this end,  $S$ , the average shortest path tree matrix, is defined. The matrix element  $s_{ij}$  refers to the number of shortest-path trees to which the edge  $(i, j)$  belongs. Then, edges with values above the threshold are included in the backbone.

The Motif-based Global Threshold method (Motif-GT) is another global filtering approach proposed by J. Cao *et al.* (8). It takes its inspiration from motifs disparity in a network. A motif  $M$  is an interconnection pattern appearing in a network  $G$  at a frequency significantly higher than in its random version. The Motif-GT algorithm proceeds as follows. First, a motif weight matrix is computed. It contains the number of occurrences of each edge  $e_{ij}$  that appears in a given motif. Second, the weighted adjacency matrix and the motif weight matrix are added to form the new weights of the network. Finally, all edges with a weight below a given global threshold are removed. Motif-based Disparity Filter (Motif-DF) is a variant of the previous method. Like Motif-GT, it computes new weights for the network links based on the motif weight and the adjacency matrices. After that, rather than using a thresholding process, it relies on the disparity filter algorithm to preserve the relevant edges. J. Cao *et al.* consider motifs consisting of three nodes in their experiments. Results show that Motif-GT outperforms simple thresholding of the original adjacency matrix. Furthermore, the comparative evaluation of the disparity filter with Motif-DF highlights the greater effectiveness of the latter. Some filter-based methods use local measures for extracting backbones. J. Chalupa *et al.* (21; 22) use the k-core which is a familiar hierarchical measure as a filter. In weighted networks, all the connections with weights lighter than  $w$  are discarded from the original network to obtain the w-backbone (23). However, considering only the strength is not enough to apprehend the weighted structure of nodes. In this context, Serrano *et al.* proposed the disparity filter backbone (24). This technique is an enhancement of the link weight-based extraction approach. It exploits a null model with uniformly distributed weights of connections. One estimates the probability  $\alpha_{ij}$  indicating the compatibility of the normalized weight of each edge  $(i, j)$  with the null model. Then, a significance level  $\alpha$  is defined. The links satisfying the condition  $(\alpha_{ij} < \alpha)$  are preserved in the backbone. This method aims to select links with a significant fraction of each given node's local strength and weight magnitude. The third category of filtering method uses a mixture of local and global measures. Ronda J *et al.* (25) propose the h-backbone. The algorithm contains three phases. Let's define the bridge measure of an edge as its betweenness divided by the size of the network. Moreover,  $h_b$  stands for the highest number such that there are  $h_b$  edges, each with a bridge value that is at least equal to  $h_b$ . In the initial phase, edges with a bridge value greater or equal to  $h_b$  are selected to form the h-bridge. Additionally,  $h_s$  stands for the highest number such that there are  $h_s$  edges, each with a strength value (weight) that is at least equal to  $h_s$ . The following phase consists of selecting edges with a weight larger or similar to  $h_s$  to form the h-strength. Then, the final step consists of fusing the links extracted from the former two steps to compose the h-backbone. This method extracts relevant edges connecting the network through the h-bridge. Additionally, it includes high-strength edges located in the network core through the h-strength. Community structure is a ubiquitous characteristic of complex networks. Real-world networks are generally organized around modules, communities, or groups. Modeling and examining the modular structure of networks is an active area of research (27; 28; 29; 30; 31; 32; 33). However, there is no consensus on this property. The most popular definition considers the community structure as a partition of the network into tightly linked subgroups of nodes loosely interconnected. These modules can also overlap. Indeed, a node may be shared by several communities. This pattern is widespread in social networks (34).

Although numerous research works are leveraging the community structure (35; 36; 37; 38; 39; 29; 40; 41; 42), filter-based extraction methods have ignored this feature. This work fills the gap. Indeed, it is essential to emphasize the effect of overlapping nodes, hubs, and bridges on the dynamics and structure of modular networks. Hubs are known to be the most dominant nodes in the network. They have a tremendous influence in their communities as they are the most connected nodes of the network. In addition, overlapping nodes are located at the heart of the network. Previous experiments (43; 44) have shown that they are more densely connected than nodes situated in the non-overlapping areas. Thus, they can exert a significant influence on nodes in various communities. Finally, bridges are the nodes connecting the different communities of the network. They are the glue that holds the parts of the network together. Thus, they have a big global influence on the network. After all, hubs, overlapping nodes, and bridges are very crucial building blocks of the network.

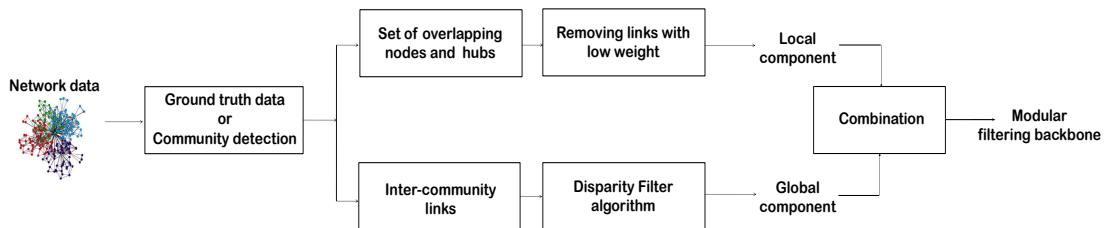
In previous work, we proposed the so-called overlapping nodes and hub backbone model for modular weighted network(26). It is composed of the overlapping nodes and the hubs of the original network. Although this backbone compares favorably with the disparity filter, it relies only on the importance of the nodes in the network. We believe that it misses some critical links in the networks that are not related to overlapping nodes and hubs. To overcome this limitation, we introduce and investigate a backbone extraction method for modular weighted networks. The proposed method highlights critical nodes in the communities (hubs and overlapping nodes) and the most essential edges bridging the communities. To our knowledge, this is the first attempt to extract backbones using all the major elements of the community structure. This approach aims to preserve the most connected nodes and the most relevant edges in terms of weight and information dissemination (high betweenness) while preserving the information content of the original network (high entropy). Additionally, when the community structure is known, the proposed method relies on local information, making it very efficient. The Modular filtering backbone is based on the union of a local and a global component of the original network. The local component is the set of overlapping nodes and hubs with their connected links. Links with the lowest weights are then discarded as long as the components do not split.

The global component contains all the inter-community links and their attached nodes. One discard all the links connecting nodes from the same communities (intra-community links) and the subsequently isolated nodes to extract it. Afterward, it is pruned using the disparity filter algorithm. The goal is to preserve only the network's most relevant inter-community links. Note that, unlike most alternative filter-based approaches, the proposed solution considers links and nodes' importance. Additionally, the proposed method relies on local information if the community structure is known. To assess the effectiveness of the proposed method, we perform several experiments on real-world networks originating from diverse domains (social, collaboration, co-appearance, biological, transportation, and technological) and of various sizes. It includes an extensive comparative evaluation involving five alternative filtering methods. The overlapping nodes and hub method and the disparity filter method are based on a local measure. The Motif-GT and the Motif-DF are two global approaches. Finally, the h-backbone method is based on local and global measures. We use various measures to assess the similarity between the proposed backbone extracting process and the alternative ones, such as the proportion of common nodes, the rank-biased overlap, and the correlation. Moreover, their effectiveness is evaluated through quality measures (the average betweenness, average weight, and entropy). Experimental results demonstrate that the Modular filtering backbone outperforms the alternative filtering methods.

The rest of the paper is organized as follows. Section 2 presents the Modular filtering backbone algorithm and a toy example to illustrate the process. Data and methods used in the experiments are described in Section 3. Results of the comparative evaluation with five alternative methods are reported in section 4. Finally, conclusions are presented in section 5.

## 2. Modular filtering backbone

### 2.1. General scheme



**Figure 1:** The main steps to extract the Modular filtering backbone.

The Modular filtering backbone extraction process relies on the idea that overlapping nodes, hubs, and link bridges (inter-community links) are the most influential elements in a network. Figure 1 represents the methodology used to extract it in weighted networks with an overlapping community structure. At first, one needs to know the overlapping community structure. Thus, the first step consists of uncovering it if the ground truth is not available. Then, one can extract the Modular filtering backbone.

It combines two components. The first subnetwork, called the local component, contains the overlapping nodes

and hubs and their stronger interactions. Indeed, one removes the edges with the lowest weights as long as components stay connected. It allows retaining only the most important links of this network. The second sub-network is called the global component. It regroups all the inter-community links of the network and their attached nodes. Its extraction from the original network consists of removing all the intra-community links and the remaining isolated nodes. Note that the intra-community links connect nodes in the same community while the inter-community links play the role of bridges between the communities. Then, the disparity filter algorithm (24) is used to keep only the most pertinent bridge connections of the global component. Preserving the most critical inter-community links in this component allows keeping the largest information flow exchange between the different communities in the network. Finally, those two components are combined. The parameter  $s$  controls the size of the Modular filtering backbone. It represents the proportion of nodes extracted from the original network. The ratio of the local and global components retained in the Modular filtering backbone is related to the strength of the community structure. In networks with a well-defined community structure, the local component retains a higher proportion of nodes and links. Indeed, the global component size is small due to the limited amount of inter-community links. In networks with a loose community structure, the global component selects a higher proportion of nodes and links because of its bigger size and importance.

## 2.2. Algorithm

The procedure followed for extracting the Modular filtering backbone is specified in more detail in this subsection. Let's consider a network  $G(V, E)$ .  $V = \{v_1, v_2, \dots, v_n\}$  denotes the set of nodes while  $E = \{(v_i, v_j) \mid v_i, v_j \in V\}$  indicates the set of edges. The Modular filtering backbone consists of the union of both the local and global components. The algorithm for its extraction is made of six steps:

**Step 1:** If the community structure of the network is unknown, an overlapping community detection algorithm is used to uncover it.

**Step 2:** Extract the local component  $G_l(V_l, E_l)$  of size  $n_l = |V_l|$  using the algorithm introduced in the following subsection 2.2.1.

**Step 3:** Extract the global component  $G_g(V_g, E_g)$  of size  $n_g = |V_g|$  using the algorithm introduced in the following subsection 2.2.1.

**Step 4:** Extract a sub-network  $G_{s_l}(V_{s_l}, E_{s_l})$  from the local component of size  $n_{s_l}$  ( $n_{s_l} \leq n_l$ ). To do so, nodes of the local component are sorted in decreasing order according to the weighted degree (45). Nodes with the lowest degrees are excluded.

**Step 5:** Extract a sub-network  $G_{s_g}(V_{s_g}, E_{s_g})$  from the global component of size  $n_{s_g}$  ( $n_{s_g} \leq n_g$ ). To do so, edges of the global component are sorted according to their strength. Edges with the lowest weights are discarded to obtain the sub-network  $G_{s_g}$ .

**Step 6:** Form the Modular filtering backbone such that  $G_b = G_{s_l} \cup G_{s_g}$  of size  $n_b$  ( $n_b \leq n_l + n_g$ ).

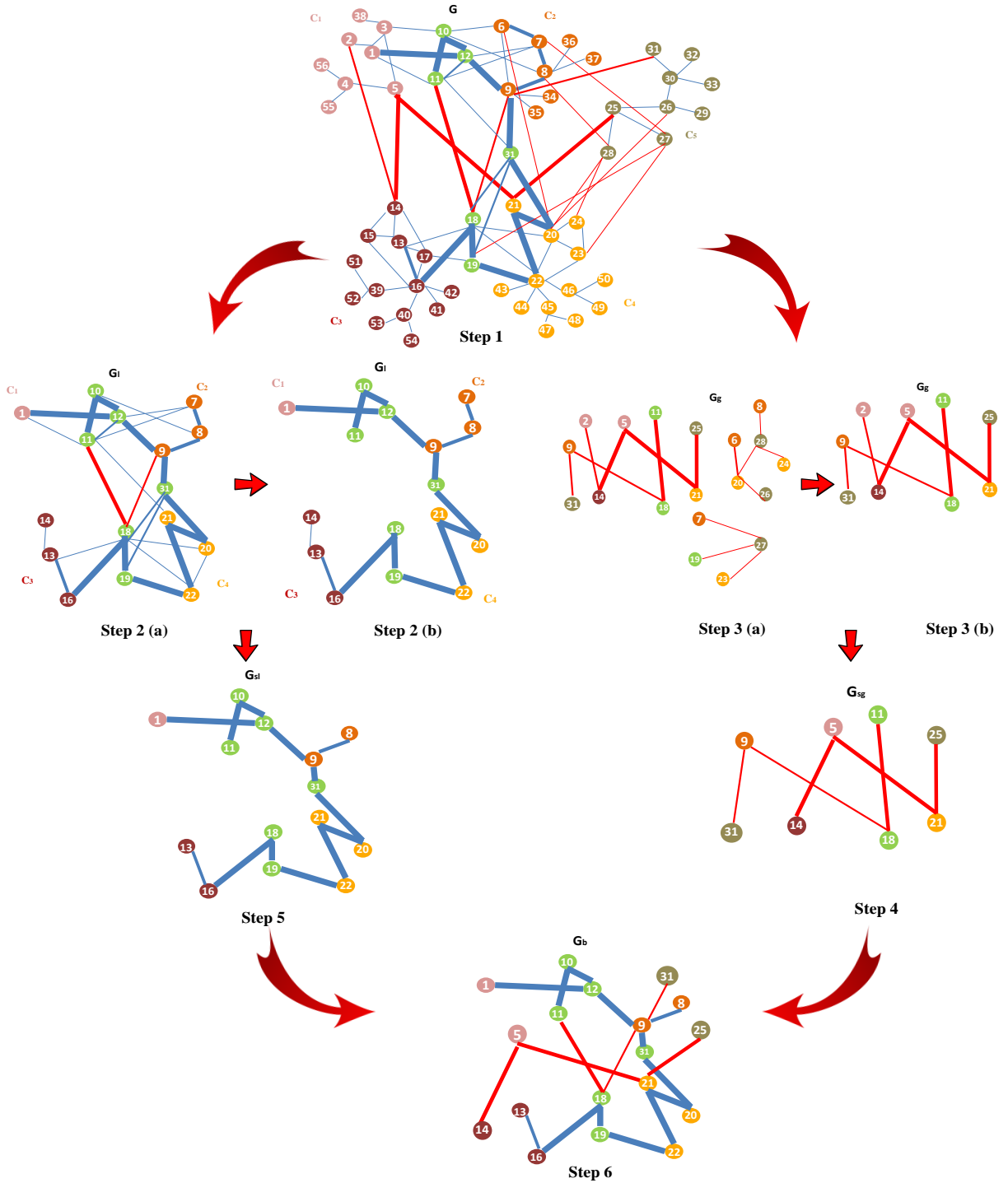
The size of the Modular filtering backbone  $n_b$  is an input parameter of the algorithm. Therefore, one can use different strategies to combine the two components sub-networks. We use the following strategy in the experimental process. If both components have a sufficient size (larger than half of the size of the generated backbone, i.e.,  $n_l > n_b/2$  and  $n_g > n_b/2$ ), the nodes of the Modular filtering backbone are equally selected from each component. In other words, the values of the sizes of the reduced version of the local and global components ( $n_{s_l}$  and  $n_{s_g}$ ) are equal ( $n_b/2$ ). Otherwise, if the size of one component is smaller than  $n_b/2$ , the Modular filtering backbone is composed by selecting all the nodes of the smallest component. Other nodes are extracted from the other component until a backbone of size  $n_b$  is formed.

Figure 2 illustrates the steps of the algorithm in a toy example.

### 2.2.1. Extracting the local component

The local component is extracted as follows:

## Extracting Modular-based Backbones in Weighted Networks



**Figure 2:** The extraction process of the Modular filtering backbone. Node colors represent the different communities. Overlapping nodes are in green. The intra-community links are blue, while the inter-community links are red. The size of links is proportional to their strength. The parameter  $n_b$  controlling the backbone size is set to the value 18 (32% of the original network).

**Step 1:** Form the local component based on the set of overlapping nodes and hubs. The overlapping nodes subset is denoted  $V_o = \{v_1^o, v_2^o, \dots, v_m^o\} \subset V$  where  $m$  represents its size. The set of hubs is noted  $V_h = \{v_1^h, v_2^h, \dots, v_t^h\} \subset V$  where  $t$  stands for its size. The local component is the union of the set of hubs and the set of overlapping nodes. To form this sub-network, all nodes that do not belong to any of the two sets are removed. The local component is denoted  $G_l(V_l, E_l)$ , where  $V_l = V_o \cup V_h$  and  $E_l = \{(v_i^l, v_j^l) \setminus v_i, v_j \in V_l\}$  are its set of vertices and links respectively. The formation of this component is illustrated in Figure 2 (Step 2 (a)).

**Step 2:** Eliminate the links with the lightest weights to reduce the local sub-network size. To this end, sort the links in weight descending order. After that, remove the links with the lowest weights as long as the components do not split into sub-components. This step is illustrated in the toy example in Figure 2 (Step 2 (b)).

### 2.2.2. Extracting the global component

The global component is extracted as follows:

**Step 1:** Form the global component. It is defined as the union of the connected nodes  $M_y$  obtained by discarding all the intra-community links from the network and the isolated nodes following this operation. The global component is denoted  $G_g = \bigcup_{y=1}^p M_y$ , where  $p = |G_g|$  is the number of the uncovered modules. Each module  $M_y$  is denoted as  $M_y(V_y, E_y)$ , where  $V_y = \{v_i^y \setminus v_i \in V\}$  refers to its set of nodes and  $E_y = \{(v_i^{y_1}, v_j^{y_2}) \setminus v_i, v_j \in V \text{ and } y_1 = y_2\}$  refers to the set of the inter-community links, while  $v_i^y$  stands for any node  $v_i$  belonging to the sub-component  $M_y$ . The formation of this component is illustrated in Figure 2 (Step 3 (a)).

**Step 2:** Reduce the global component using the disparity filter algorithm. It allows preserving only the most relevant inter-community links of the network. Figure 2 (Step 3 (b)) presents the global component of the original network after processing by the disparity filter algorithm.

### 2.3. Toy example

Figure 2 illustrates the steps of the algorithm to extract the Modular filtering backbone of a toy network. In this network, there are five well-defined communities (represented in Step 1). Nodes in the same communities have the same color. Overlapping nodes are in green. Intra-community links are in blue, and inter-community links are in red. This toy network is composed of 56 nodes. In this example, we extract a backbone containing 18 nodes. It represents 32% of the size of the original network. To this end, the parameter  $n_b$  controlling the size of the generated backbone is set to the value 18.

Step 2 of this figure shows the extraction of the local component. Step 2 (a) represents the sub-network containing the overlapping nodes and hubs. The set of hubs is formed according to the weighted degree. Top-ranked nodes are selected. The size of this set is identical to the size of the set of the neighbors of the overlapping node. In this example, the overlapping nodes have ten neighbors. Consequently, the set of hubs contains the 10 top-ranked nodes of this network. Then, this local component size is reduced in Step 2 (b). Links with the lowest weights are removed as long as the connected component does not split. The component contains 16 nodes. Step 3 illustrates the extraction of the global component. One selects all the inter-community links with their respective nodes in Step 3 (a). This sub-network has three components containing 9, 6, and 4 nodes. It is then reduced using the disparity filter algorithm as shown in Step 3 (b). The final global component contains eight nodes. After that, one combines the two components such that the size of the backbone is equal to 18 nodes. As the size of the global component is smaller than  $9 (n_b/2)$ , it is not reduced in Step 4. The ten other nodes needed to form the backbone are selected from the local component build in Step 2 (b). Nodes with the lowest weighted degree are discarded from the local component, as illustrated in Step 5. Finally, Step 6 of Figure 2 represents the composition of the Modular filtering backbone. It is the union of the reduced version of the local and global components.

This toy example shows that the local component preserves the most relevant nodes and connections. However, this sub-network does not contain any node from community  $C_5$ . Indeed, this peripheral community has only nodes with low weight. Moreover, the local component also misses essential inter-community links connecting nodes from different communities and combining the local with the global component to address these drawbacks. As shown in Figure 2, the Modular filtering backbone contains nodes originating from all the network communities. Indeed, nodes 25 and 31 of community  $C_5$  are included in this backbone. It also contains the main inter-community links bridging the communities. Therefore, the Modular filtering backbone seems more respectful of the original network topology.

### 3. Materials and methods

#### 3.1. Methods

We use real-world networks originating from various fields (co-appearance, social, collaboration, transportation, and technological networks) in the experimental process to cover a broad range of situations. We use the Speaker-Listener Label Propagation Detection Algorithm (SLPA) (46) to reveal the overlapping community structure of the networks. Indeed, previous results demonstrate that the community structures uncovered by various community detection algorithms are usually relatively consistent (5). In other words, the influence of the community detection algorithm due to the variations of the community structure is quite limited on the extracted backbones.

We perform a comparative evaluation of the proposed backbone extractor with five alternative methods (overlapping nodes and hub, disparity filter, Motif-GT, Motif-DF, and h-backbone) using various evaluation metrics. In the first part of the experiments, the proportion of common nodes and the rank-biased overlap allow us to evaluate the similarity between the extracted backbones. Moreover, Pearson and Kendall Tau correlation coefficients are estimated to measure their statistical relationship. The second part of the experiments compares the backbone extractors' effectiveness using the average link betweenness, average link weight, and entropy. The first measure indicates the amount of information that the edges of the backbone can broadcast. The average link weight reflects their ability to preserve the most critical edges. Finally, the entropy measures the quantity of information kept in each extracted backbone.

#### 3.2. Data

A set of real-world networks is collected from different fields (co-appearance, social, collaboration, transportation, and technological networks) to conduct a series of experiments. Their size ranges from hundreds to thousands of nodes and edges. These datasets are available on "toreopsahl.com/datasets" and "github.com/zakariyaGH/Datasets".

**Zachary's Karate Club network.** This social network represents the friendship among members of a US university karate club. The nodes are the members of the club, and the connections reflect their friendship. The weight of each link ranges between 0 and 8. It refers to the strength of the relationship of a pair of individuals.

**Intra-Organisational network.** It represents a consulting firm. Nodes are to the 46 workers of the firm. There is a connection between two workers when they have shared advice or information. Weights range between 0 and 5. They reflect how frequently, over a three-month cycle, a worker referred to his colleague for guidance or information on topics related to the job.

**Freeman's EIES network.** It includes 48 scientists focusing on social networks. Nodes represent researchers. The edges represent their relation. The weight of the edges ranges from 0 to 4. It is equal to 0 if the two persons are unfamiliar. It is equal to 1 if the researcher has heard of this person but didn't meet her. The value 3 refers to a friendship relationship, and 4 reflects a strong personal friendship.

**Train Bombing network.** It is a social network of terrorism suspects implicated in the train bombing on 11 March 2004 in Madrid. It was collected based on newspaper articles. These terrorists are represented by nodes, while a connection between two suspected terrorists exists if they were in touch. The weights reflect how solid their connection was (a co-participating in training camps or previous attacks, or just a friendship relationship).

**Les Miserables network.** It is a network of co-appearances that includes the characters of the famous novel 'Les Miserables'. Nodes represent the characters. A connection occurs between two characters if they are mentioned in the same chapter. The weights indicate the number of occurrences.

**Game of thrones network.** It collects the co-appearance network of the characters of the Game of Thrones series. Nodes are the characters of the book 'A Storm of Swords'. A connection occurs between a pair of characters if their names are mentioned within 15 words of each other in the text. The number of such co-appearances weights the links.

**Caenorhabditis elegans worm's neural network (C. Elegans Neural).** This network contains 306 neurons represented by nodes. A connection exists when at least one synapse or gap junction exists between neurons. The weights reflect the number of gap junctions and synapses.

**Facebook-like Social network.** It is an online student social network from the University of California. Users involved in this network are students who have at least one message sent or received. The weights indicate the number of characters of the messages.

**Facebook-like Forum network.** This network collects users sharing messages in a forum. This platform contains 899 members and 522 subjects. Nodes are the platform members. The edges represent the messages exchanged. The weights represent the number of messages posted by a member on a given topic.

**US Power Grid network.** It is the power grid network of the western states of the USA. Nodes refer to the

substations, generators, and transformers. The connections represent high voltage lines. The weights represent the cumulative electrical charge in the various lines.

**Scientific Collaboration network.** This network includes authors of papers published in the Condensed Matter section of arXiv E-Print Archive. Nodes indicate the authors. A link exists between a pair of authors if they co-author a manuscript. The weights denote the number of common manuscripts.

**US Top-500 Airport network.** This network represents the 500 busiest commercial airports of the United states. Nodes stands for the airports, and an edge exists between two airports if a flight was scheduled between them in 2002. The number of seats on these flights corresponds to the weights of the edges.

**US Airport network.** This network represents the commercial airports of the United States. The flights considered in this dataset are scheduled in 2010.

**Openflights network.** This dataset contains non-US-based airports. It was downloaded from ‘‘Openflights.org’’. In this network, The weights represent the number of routes between two airports.

### 3.3. Evaluation measures

We compare the backbones using three types of evaluation measures (similarity, correlation, and effectiveness). Brief definitions follow.

#### 3.3.1. Similarity measure: Proportion of common nodes

The proportion of common nodes measures the similarity between two backbone extractors. It examines if the backbones select a similar set of nodes. This basic similarity measure estimates the number of elements belonging to two distinct sets. The proportion of common nodes  $A_n$  among two sets of vertices of equal size  $n$  is defined as follows:

$$A_n = \frac{|X \cap Y|}{n} \quad (1)$$

Where  $X$  and  $Y$  represents two different sets of vertices.

#### 3.3.2. Similarity measure: Rank-biased overlap

The rank-biased overlap (RBO) (47) is another similarity measure allowing to weight the importance of common nodes of two backbones. It measures the fraction of common elements between a pair of ranking sets while continuously increasing their depths. The nodes are ranked according to their weighted degree. For each depth, a different weight is assigned to the overlap.  $X$  and  $Y$  are two ranked sets sorted in decreasing order. The RBO  $r$  between  $X$  and  $Y$  is defined by:

$$r(X, Y) = \sum_d w_d \cdot A_d \quad (2)$$

Where  $A_d$  indicates the proportion of common elements between  $X$  and  $Y$  at depth  $d$ . It is given by:

$$A_d = \frac{|X_{:d} \cap Y_{:d}|}{d} \quad (3)$$

$X_{:d}$  designates the set of rankings belonging to the range of  $[1, d]$ .  $|X_{:d} \cap Y_{:d}|$  is the set of the common elements (overlap) between  $X$  and  $Y$  at depth  $d$ .  $w_d$  represents the weight at depth  $d$ . It is defined by:

$$w_d = (1 - p) \cdot p^{d-1} \quad \text{where:} \quad \sum_d w_d = 1 \quad (4)$$

The weights can be set through  $p$ . This parameter value is in the range  $[0, 1]$ . The top-ranked elements of the set have more weight if  $p$  has a low value. The comparison between the two sets is deeper when this parameter gets close to 1. Moreover, RBO values range between 0 and 1. The value of 1 is obtained when the two sets are identical. It is equal to 0 when  $X$  and  $Y$  are totally disjoint.

### 3.3.3. Correlation measure: Pearson correlation

This correlation coefficient evaluates the agreement between the two sets. It measures the monotonic association between the set of degrees of the extracted backbones. The Pearson correlation  $\rho$  of two sets  $X$  and  $Y$  is given by:

$$\rho(X, Y) = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^n (x_i - \bar{x})^2 \sum_{i=1}^n (y_i - \bar{y})^2}} \quad (5)$$

Where:

$$\bar{x} = \frac{\sum_{i=1}^n x_i}{n}; \quad \bar{y} = \frac{\sum_{i=1}^n y_i}{n} \quad (6)$$

$X$  and  $Y$  refer to two sorted sets of weighted degrees for nodes belonging to the two compared backbone extractor.

### 3.3.4. Correlation measure: Kendall Tau correlation

The Kendall Tau correlation coefficient is a measure of the ranking agreement between two sets. It is used to quantify the monotonic association between the set of ranks of the different extracted backbones. Let  $X$  and  $Y$  be the ranking sets regarding the nodes of two backbones of size  $n$ , where  $X = (x_1, x_2, \dots, x_n)$  and  $Y = (y_1, y_2, \dots, y_n)$ . Here, the ranking sets are computed based on the weighted degree of nodes belonging to two different backbones. The two ranks  $(x_i, y_i)$  and  $(x_j, y_j)$  are concordant when  $x_i < x_j$  and  $y_i < y_j$ , or  $x_i > x_j$  and  $y_i > y_j$ . The pair of ranks are discordant when  $x_i < x_j$  and  $y_i > y_j$ , or  $x_i > x_j$  and  $y_i < y_j$ . They are neither concordant nor discordant when  $x_i = x_j$  or  $y_i = y_j$ . The Kendall Tau correlation coefficient  $\rho_\tau$  between two ranking sets  $X$  and  $Y$  of size  $n$  is defined as follows:

$$\rho_\tau(X, Y) = \frac{n_c - n_d}{\frac{1}{2}n(n-1)} \quad (7)$$

Where  $n_c$  and  $n_d$  are respectively the amount of concordant and discordant pairs.

The correlation coefficient values range in  $[-1, 1]$ . If it is positive, then there is a positive monotonic association between the two sets. That means that their values tend to increase or decrease simultaneously. If the value of the correlation coefficients is less than 0, there is a negative monotonic association between the two sets. In this case, values of one set tend to increase when the values of the other decrease. In addition, the pair of sets have no monotonic association when the value of the correlation coefficients is equal to 0. Moreover, the two sets have a moderate strength relation when the coefficient ranges between 0.5 and 0.8. This relationship is very strong when the coefficient is ranging between 0.8 and 1.

### 3.3.5. Effectiveness measure: Average betweenness

This measure refers to the quantity of information flow crossing all the edges of a given backbone. It allows determining which backbone may act as the best information gateway in the originating network. The higher the average link betweenness, the better the ability of the backbone to disseminate information. The betweenness is a measure of centrality based on the shortest paths. It quantifies the number of shortest paths of each pair of nodes passing through a given node. Here, we focus on link betweenness. The betweenness of a link can be similarly defined. It quantifies the number of times a connection acts as a bridge in the shortest path between two nodes. The betweenness centrality of a link  $e$  belonging to a network  $G$  is defined by:

$$\beta(e) = \sum_{v_s, v_t \in G} \frac{\sigma_{st}(e)}{\sigma_{st}} \quad (8)$$

Where  $\sigma_{st}$  counts the number of shortest paths connecting nodes  $v_s$  and  $v_t$ .  $\sigma_{st}(e)$  represents the number of shortest paths connecting nodes  $v_s$  and  $v_t$  and passing through the link  $e$ . The average betweenness is obtained by dividing the betweenness centrality of the total links of a given backbone  $M$ . It is defined as follows:

$$\langle \beta \rangle = \frac{\sum_{e_{ij} \in M} \beta(e_{ij})}{m} \quad (9)$$

Where  $e_{ij}$  is a link connecting two nodes  $v_i$  and  $v_j$ .  $m$  represents the size of the set of links of the backbone  $M$ .

### 3.3.6. Effectiveness measure: Average weight

This measure computes the average link weight of a network. It is calculated for all the links of a given backbone. The average link weight quantifies the thickness of the connections of the extracted backbones. The higher the value, the more relevant the selected links.

Let's consider that  $w_{ij}$  denotes the weight of the link connecting the nodes  $v_i$  and  $v_j$ . The average distance  $\langle w \rangle$  of a given backbone  $M$  is defined as:

$$\langle w \rangle = \frac{\sum_{v_i, v_j \in M} w_{ij}}{m} \quad (10)$$

Where  $m$  represents the size of the backbone  $M$ .

### 3.3.7. Effectiveness measure: Entropy

The entropy measures the amount of information delivered by an information source. It indicates how much information a backbone can preserve from the original network. More information is preserved from the original network if this measure has a higher value. Generally speaking, the entropy represents the average quantity of information generated by a probabilistic stochastic source of data (48). Usually, it quantifies the uncertainty in the system. Let's consider a system  $X$  composed of  $n$  elements, each element  $x_i$  having a probability  $P_i$  to appear. The entropy  $H$  is defined as follows:

$$H(X) = \sum_{i=1}^n P_i \log\left(\frac{1}{P_i}\right) = - \sum_{i=1}^n P_i \log(P_i) \quad (11)$$

In this work, the entropy is computed based on the degree distribution of the nodes of each backbone.

## 4. Results and discussion

This section reports and analyzes the comparative evaluation results of the proposed Modular filtering backbone extractor with five alternative methods under test. The same methodology is followed for similarity and effectiveness analysis. First, we compare the proposed method to the overlapping nodes and hub method. Indeed, the proposed technique is an extension of the former one. Next, the Modular Filtering method is compared to the other alternative methods (disparity filter, Motif-GT, Motif-DF, and h-backbone).

Parameters are set such that the size of the resulting backbones is identical (30% of the original network size). Indeed, it allows a fair comparison. Furthermore, for all the networks, the significance level  $\alpha$  of the disparity filter algorithm used in the global component to form the Modular filtering backbone is set to the value 0.5 for the same reason.

### 4.1. Similarity analysis

We report an extensive similarity analysis conducted to compare the various generated backbones for the networks under test. It is based on the proportion of common nodes and the rank-biased overlap measures.

#### 4.1.1. Proportion of common nodes

Here, the fraction of common nodes selected by different backbone extraction approaches are estimated. This experiment aims to check if the backbones pick a similar set of nodes. At first, let's consider the backbones extracted in small size networks. The set of nodes in these networks can be examined visually. The figs. 3 and 5 represent respectively Les Miserables and Game of thrones networks. Their respective backbones are reported in figs. 4 and 6. In these figures, (a) is the overlapping nodes and hub backbone, (b) is the Modular filtering backbone, (c) is the disparity filter backbone, (d) is the Motif-GT backbone, (e) is the Motif-DF backbone, and (f) is the h-backbone. The color of the vertices indicates the communities to which they belong. The overlapping nodes are colored in gray. The thickness of the edges is proportional to their weight and the size of nodes to their weighted degree (45).

First of all, let us compare the Modular filtering and the overlapping nodes and hub backbones. One can see that these two backbones share a large number of nodes. Around 95% of nodes are common in both backbones for Les

**Table 1**

Estimated parameters of the empirical networks under test.  $N$  is the total number of nodes of the network.  $\mu$  is the mixing parameter. It is equal to the weight of the inter-community links divided by the weight of the overall links of the network.  $A_n$  is the fraction of common nodes between two sets of nodes of size  $n$ . Note that MF stands for the Modular filtering backbone, OH stands for the Overlapping nodes and Hubs backbone, and DF stands for the Disparity Filter backbone. Motif-GT refers to the motif-based global threshold method, and Motif-DF refers to the motif-based disparity filter. Each reported value is the average of 10 SLPA simulation runs. The standard deviation is not in this table due to its small values ranging between 0% and 2%.

Networks	$N$	$\mu$	$A_n(\%)$				
			MF-OH	MF-DF	MF - Motif-GT	MF - Motif-DF	MF - h-backbone
Zachary's karate club	34	0.08	90.00	81.82	80.00	50.00	77.77
Intra-organisational	46	0.11	81.82	61.54	63.95	56.66	65.61
Freeman's EIES	48	0.08	81.81	72.72	71.64	53.33	76.37
Train bombing	62	0.05	97.60	94.73	84.21	66.66	64.70
Les Miserables	77	0.09	95.67	83.80	76.26	57.15	70.84
Game of thrones	107	0.07	96.87	65.62	75.00	52.94	65.94
C.elegans Neural	306	0.05	93.25	60.95	77.89	59.58	61.50
US Top-500 Airport	500	0.08	98.66	57.10	84.00	57.66	55.83
US Airoport	1574	0.09	95.95	74.69	76.13	58.90	69.57
Openflights	7976	0.09	88.21	71.41	60.17	56.45	63.53
Facebook-like Forum	899	0.31	80.79	57.03	50.56	50.16	45.54
Facebook-like Social	1899	0.40	80.44	72.53	55.55	51.20	50.00
US Power Grid	4941	0.12	97.49	65.89	78.64	59.32	64.09
Scientific Collaboration	16726	0.97	92.48	56.75	71.15	54.33	60.91

**Table 2**

Estimated parameters of the empirical networks under test.  $A_t$  is the proportion of the highly connected nodes of a given set of nodes, where  $t$  is equal to 10% of the top-ranked nodes of the network ( $t = 0.1 * N$ ). Note that MF stands for the Modular filtering backbone, OH stands for the Overlapping nodes and Hubs backbone, and DF stands for the Disparity Filter backbone. Motif-GT refers to the motif-based global threshold method, and Motif-DF refers to the motif-based disparity filter. Each reported value is the average of 10 SLPA simulation runs. The standard deviation is not in this table due to its very small values ranging from 0% to 2%.

Networks	$A_t(\%)$					
	MF	OH	DF	Motif-GT	Motif-DF	h-backbone
Zachary's karate club	100.00	100.00	100.00	100.00	100.00	100.00
Intra-organisational	100.00	100.00	75.00	75.00	45.00	75.00
Freeman's EIES	100.00	100.00	80.00	100.00	47.00	53.33
Train bombing	100.00	100.00	100.00	100.00	46.00	100.00
Les Miserables	100.00	100.00	93.33	100.00	55.90	100.00
Game of thrones	100.00	100.00	90.00	100.00	90.00	90.00
C.elegans Neural	100.00	100.00	75.86	51.72	48.27	55.17
US Top-500 Airport	100.00	100.00	94.00	100.00	94.00	68.00
US Airoport	100.00	100.00	91.53	83.92	59.22	100.00
Openflights	100.00	100.00	82.00	67.80	55.90	65.14
Facebook-like Forum	100.00	100.00	70.78	61.61	57.74	73.26
Facebook-like Social	100.00	100.00	91.53	79.86	53.74	68.47
US Power Grid	100.00	100.00	72.26	63.28	50.85	65.18
Scientific Collaboration	92.32	97.05	68.19	67.44	57.05	63.32

Miserables and Game of thrones networks, as illustrated in figs. 4 and 6 respectively. Indeed, both extraction methods select the set of overlapping nodes and the most connected nodes of the network (hubs). The Modular filtering backbone combines these nodes forming its local component with the global component. The latter is composed of bridge nodes and links. However, the global component has a very small size in these networks. Indeed, with their mixing parameter values that do not exceed 0.1, a well-defined community structure characterizes these networks. Note that the mixing parameter represents the proportion of inter-community links in a non-weighted network. It is equal to the weight of the inter-community links divided by the weight of the total connections. Therefore, it measures the strength of the community structure of a given weighted network. A limited number of bridges quantified by a small mixing parameter value characterize networks exhibiting a well-defined community structure. In this situation, the Modular filtering

and the overlapping nodes and hub backbones extract approximately the same set of nodes. Generally speaking, one can notice that both Modular filtering and overlapping nodes and hub backbones preserve more relevant nodes and links from the original network. For instance, in *Les Misérables*' novel, Jean Valjean is the main character. This ex-convict followed throughout his life by Javert (the inspector). During the narration of this storytelling, Jean Valjean is firmly connected to his adopted daughter Cosette besides Marius (her future husband). Cosette's mother, Fantine, Mr, and Mme Thenardier, and Gavroche are also essential characters of this novel. Interestingly, the proposed backbone and the overlapping nodes and hub extractor algorithm manage to select all these characters. One can draw the same conclusions in the Game of thrones network. Both the Modular Filtering and the overlapping nodes and hub backbones contain the main characters. Indeed, these backbones preserve most of the top-connected nodes and crucial links.

Differences are more pronounced when comparing the backbones extracted by the proposed method with those obtained using other alternative methods (disparity filter, Motif-GT, Motif-DF, and h-backbone). The proportion of common nodes exhibits systematically smaller values. figs. 4 and 6 illustrate this behavior common to all networks. These alternative filtering methods concentrate on links while the Modular filtering and the overlapping nodes and hub backbones focus more on nodes. Indeed, some main characters do not appear in these alternative backbones. In "*Les Misérables*" network, for instance, the disparity filter backbone does not contain Marius. Cosette is also missing in this backbone and the Motif-DF backbone. Gavroche is another main character who the Motif-GT and the h-backbone do not select. Moreover, these backbones select some secondary characters. Besides, the connection between Javert and Jean Valjean should not have been removed from the backbone, knowing that they have co-appeared in several chapters. Yet, the disparity filtering, Motif-DF, and the h-backbone methods miss this extremely relevant connection. Although Catelyn Stark is the main character in the Game of Thrones saga series, she is not selected by these competing filtering-based methods. Furthermore, one can notice that the Modular filtering backbone is the only approach that succeeds in selecting nodes from all the communities of the original *Les Misérables* and Game of Thrones networks. Indeed, nodes belonging to the small community colored in yellow in "*Les Misérables*" network do not appear in the overlapping nodes and hub backbone, as shown in figs. 3 and 4. Such is the case for nodes belonging to the community colored in purple in the Game of thrones network as shown in figs. 5 and 6. Therefore, the Modular filtering backbone is the only extraction method that manages to select quite relevant nodes from various network communities.

Table 1 reports the results for all the networks under study. One may consider that the similarity between two backbones is high if the fraction of their identical nodes exceeds 50%. The proportion of common nodes between the proposed backbone and the overlapping nodes and hub backbone ranges from 80% to 98% for all the different networks. Therefore, these two backbone extractors generate very similar sets of nodes. However, differences are more pronounced for networks with high mixing parameter values. It is the case in the Facebook-like Forum and Facebook-like Social networks. For these networks, the two backbones have fewer nodes in common. It is due to the larger size of the global component. Networks with a weak community structure have more bridge nodes and inter-community links. Whatever, even if it decreases when the mixing parameter increases, the overlap between the two backbones is overall quite meaningful. autoreft1 reports the proportion of common nodes between the Modular filtering backbone and the other alternative methods (disparity filter, Motif-GT, Motif-DF, and h-backbone). Results reveal a relatively more minor overlap with the proposed backbone extractor. Indeed, the Modular filtering backbone extractor tends to uncover meaningful nodes and links while competing methods focus on selecting edges. The disparity filter method targets the most relevant edges while also selecting some edges with small weights to preserve the original network's link weight distribution. The Motif-based methods tend to select edges incorporated in the highest number of motifs. The h-backbones searches for edges with high h-index and link betweenness. Following a different approach, the proposed method selects nodes from the local component (overlapping nodes and hubs) and edges from the global component (bridges). It explains the relatively low overlap between the Modular filtering backbone and the backbones uncovered by the competing methods. Overall, the Modular filtering backbone presents the highest proportion of common nodes with the Overlapping nodes and hubs backbone, followed by the disparity filter backbone. The Motif-GT comes in third place while the h-backbone comes fourth. Finally, the smallest overlap is with the Motif-DF backbone.

We also evaluate the proportion of the top-ranked nodes of the original networks preserved in the various backbone extractors. First, we compute the reference set of top-10% nodes of the original network based on their weighted degree. Then, one calculates the proportion of common nodes between the reference set and the set of the top nodes derived from the backbone extractors. Experimental results reported in Table 2 show that this proportion always reaches 100% for both Modular filtering and overlapping nodes and hub backbones. Thus, the sets are identical. This observation is valid for all the real-world networks under test except for the scientific collaboration network that shows an overlap higher than 90%.

**Table 3**

Estimated values of the rank-biased overlap  $r$  between the sets of nodes of two backbones. MF stands for the Modular filtering backbone. OH stands for the Overlapping nodes and Hubs backbone. DF stands for the Disparity Filter backbone. Motif-GT refers to the motif-based global threshold method. Motif-DF refers to the motif-based disparity filter. The parameter  $p$  defines the weights of the elements of the two sets. Equal weight is assigned to all the elements of both sets when  $p$  has a high value (close to 1). Each value is the average of 10 SLPA simulation runs. The standard deviation is not reported in this table (it is around  $10^{-3}$ ).

Networks	$r$														
	MF-OH			MF-DF			MF - Motif-GT			MF - Motif-DF			MF - h-backbone		
	$p=0.5$	$p=0.8$	$p=0.98$	$p=0.5$	$p=0.8$	$p=0.98$	$p=0.5$	$p=0.8$	$p=0.98$	$p=0.5$	$p=0.8$	$p=0.98$	$p=0.5$	$p=0.8$	$p=0.98$
Zachary's karate club	1.00	0.99	0.92	0.93	0.92	0.83	0.98	0.91	0.81	0.99	0.95	0.90	0.98	0.92	0.87
Intra-organisational	1.00	0.98	0.85	0.83	0.75	0.65	0.70	0.50	0.53	0.73	0.55	0.46	0.97	0.85	0.71
Freeman's EIES	1.00	0.99	0.92	0.92	0.87	0.82	0.95	0.77	0.57	0.72	0.57	0.52	0.83	0.58	0.54
Train bombing	1.00	1.00	1.00	0.93	0.93	0.86	0.99	0.99	0.88	0.63	0.54	0.51	0.99	0.93	0.70
Les Miserables	1.00	1.00	0.97	0.93	0.92	0.86	0.90	0.89	0.51	0.77	0.55	0.50	0.89	0.75	0.54
Game of thrones	1.00	1.00	0.98	0.93	0.90	0.70	0.99	0.99	0.83	0.97	0.70	0.66	0.99	0.96	0.65
C.elegans Neural	1.00	1.00	0.99	0.93	0.90	0.63	0.95	0.77	0.53	0.84	0.75	0.42	0.99	0.90	0.62
US Top-500 Airport	1.00	1.00	0.99	0.98	0.98	0.87	0.99	0.99	0.98	0.99	0.99	0.91	0.89	0.99	0.77
US Airport	1.00	1.00	0.98	0.98	0.97	0.91	0.91	0.81	0.58	0.76	0.68	0.56	0.93	0.84	0.61
Openflights	1.00	1.00	0.98	0.95	0.94	0.83	0.94	0.78	0.64	0.71	0.65	0.52	0.97	0.89	0.57
Facebook-like Forum	1.00	1.00	0.96	0.93	0.92	0.67	0.88	0.70	0.51	0.64	0.50	0.40	0.88	0.70	0.50
Facebook-like Social	1.00	1.00	0.95	0.93	0.92	0.89	0.85	0.75	0.52	0.53	0.51	0.46	0.86	0.69	0.51
US Power Grid	1.00	1.00	0.98	0.92	0.89	0.86	0.97	0.79	0.57	0.69	0.53	0.49	0.87	0.75	0.58
Scientific Collaboration	1.00	1.00	0.94	0.94	0.93	0.88	0.94	0.82	0.53	0.76	0.54	0.51	0.89	0.79	0.61

Comparisons between the top-ranked nodes preserved in each backbone are also performed. To do so, the set of 10% of the top-connected nodes of the original network is defined according to their weighted degree. Then, the proportion of the common nodes between this set and the set of the top-ranked nodes of the same size derived from the other backbones is computed. Experimental results reported in Table 2 show that this proportion reaches always 100% for both Modular filtering and overlapping nodes and hub backbones. Thus, the sets are identical. This observation is valid for all the real-world networks under test except for the scientific collaboration network that shows nevertheless an overlap higher than 90%. The overlapping nodes and hub backbone selects mainly the hubs. As expected, it preserves a slightly higher proportion of highly connected nodes. Comparison with the other backbone extractors (disparity filter, Motif-GT, Motif-DF, and h-backbone) shows a lower overlap. Indeed, focusing more on links, they select a lower fraction of the top-ranked nodes. Overall, the Modular filtering, together with the overlapping nodes and hub method, manages to preserve the most connected nodes (with a slight preference for the overlapping nodes and hub backbone). They are followed by the disparity filter and the Motif-GT in the fourth place. The h-backbone ranks fifth, and the Motif-DF backbone comes in the last position.

#### 4.1.2. Rank-biased overlap

We also use the rank-biased overlap RBO to quantify the similarity between the backbones. The node sets are identical if its value is equal to 1 and the overlap decreases when it tends to 0. This measure attaches diverse weights to the elements of the sets. The parameter  $p$  allows controlling those weights. When it has a low value, more weight is assigned to compare the top-ranked elements of the two sets. Otherwise, the higher the  $p$  value, the more comparable the importance given to the elements of the sets. The RBO quantifies the similarity between the set of nodes of the Modular filtering backbone and its alternatives. Table 3 shows the experimental results of the RBO evaluation. First, let's compare the Modular Filtering with the overlapping nodes and hub backbone. It appears that there is a significant overlap between the sets of nodes of both backbones. This result is valid when all the elements ( $p = 0.98$ ) have the same weight. In this case, RBO varies between 81% and 99%. These proportions increase if the top-ranked elements have more weights ( $p = 0.5$ ). In this case, the value of the RBO is always equal to 100% for all the networks under test. figs. 4 and 6 illustrates these results for small networks. Note that each backbone contains the top-connected nodes of the original network. Additionally, Table 3 shows that the Facebook-like Forum and the Facebook-like Social exhibit slightly lower RBO values. It is because they have a bigger global component.

Thus, more bridge nodes and inter-community links are extracted from this component, decreasing the overlap between the two backbones. RBO exhibits relatively lower values when comparing the proposed backbone extractor with the disparity filter, Motif-GT, Motif-DF, and the h-backbone. These values increase when more weight is assigned

**Table 4**

Pearson correlation  $\rho$  between the sets of nodes of two backbones. MF stands for the Modular filtering backbone. OH stands for the Overlapping nodes and Hubs backbone. DF stands for the Disparity Filter backbone. Motif-GT refers to the motif-based global threshold method. Motif-DF refers to the motif-based disparity filter. Each value is the average of 10 SLPA simulation runs. The standard deviation is not reported in this table (around  $10^{-3}$ ).

Networks	$\rho$				
	MF-OH	MF-DF	MF - Motif-GT	MF - Motif-DF	MF - h-backbone
Zachary's karate club	0.99	0.96	0.98	0.70	0.92
Intra-organisational	0.92	0.88	0.77	0.66	0.87
Freeman's EIES	0.88	0.57	0.86	0.65	0.75
Train bombing	0.85	0.79	0.89	0.57	0.89
Les Miserables	0.87	0.78	0.81	0.69	0.80
Game of thrones	0.96	0.89	0.89	0.87	0.89
C.elegans Neural	0.93	0.79	0.86	0.76	0.89
US Top-500 Airport	0.99	0.95	0.89	0.89	0.80
US Airoport	0.95	0.91	0.82	0.68	0.84
Openflights	0.96	0.90	0.85	0.64	0.87
Facebook-like Forum	0.80	0.94	0.79	0.58	0.79
Facebook-like Social	0.82	0.90	0.77	0.48	0.77
US Power Grid	0.98	0.87	0.87	0.62	0.78
Scientific Collaboration	0.93	0.86	0.85	0.68	0.80

to the top-ranked elements ( $p = 0.5$ ). Similar to the previous measures, the lowest RBO values appear with the Modular Filtering and Motif-DF backbone. Indeed, these backbones contain some peripheral nodes. These nodes which do not have a significant influence on the network are excluded from the Modular Filtering and the overlapping nodes and hub backbones. Therefore, these findings reinforce the observations made in the previous subsection.

#### 4.1.3. Summary

To sum up, there is a high similarity between the sets of nodes extracted by the Modular filtering and the overlapping nodes and hub backbones. However, this similarity decreases slightly when the mixing proportion parameter of the network increases (in networks with a non-cohesive community structure). Additionally, the top-ranked nodes selected by both backbones are identical for all the networks under test. Comparisons with other backbone extractors under test show that these backbones sometimes select less relevant nodes. Globally, one can classify these backbones according to their similarity to the Modular filtering backbone. The overlapping nodes and hub backbone is the most similar to the proposed method. The disparity filter backbone method follows. Motif-GT comes in third place, and the h-backbone method comes just after. The Motif-DF method is the least similar to the proposed method.

## 4.2. Correlation Evaluation

The correlation of the set of nodes of the various backbones is measured using Pearson and Kendall Tau coefficients. Equation 5 represents the expression of the first metric. It assesses the degree of accordance between two sets of nodes. Here, it is computed between the sets of the weighted degrees of nodes revealed by the backbones. The sets are sorted in decreasing order. The Kendall Tau metric evaluates the statistical relationship between the ranks of these sets. One computes the sets of ranks using the weighted degree centrality. The values of the correlation coefficients are ranging from -1 to 1. The value 1 indicates an absolute positive correlation between the sets of nodes. There is a complete negative correlation if the value is equal to -1. We compute the Pearson and Kendall Tau coefficients for all the empirical networks. Results are reported in tables 4 and 5 respectively. Let's compare the Modular filtering backbone and the overlapping nodes and hub backbone first. Results show that there is always a very high correlation value between the set of nodes selected by the Modular filtering backbone and the overlapping nodes and hub backbone. Indeed, all the values are close to 1, especially for networks with a small mixing parameter value. The correlation coefficients are slightly smaller in networks with a loose community structure (great mixing parameter). Overall, one can say that the sets of nodes of both backbones have a strong relationship. Let's turn now to the comparisons between the proposed method and the other alternatives. Globally, correlation exhibits relatively smaller values for all the empirical networks than the ones estimated between the proposed backbone and the overlapping nodes and hub backbone. Indeed, these alternative methods contains nodes with very low weighted degrees in some networks. Consequently, one observes a slightly weaker monotonic relationship between the Modular filtering and the disparity filter, Motif-GT, h-backbone

**Table 5**

Kendall Tau correlation  $\rho_\tau$  between the sets of nodes of two backbones. MF stands for the Modular filtering backbone. OH stands for the Overlapping nodes and Hubs backbone. DF stands for the Disparity Filter backbone. Motif-GT refers to the motif-based global threshold method. Motif-DF refers to the motif-based disparity filter. Each value is the average of 10 SLPA simulation runs. The standard deviation is not reported in this table (around  $10^{-3}$ ).

Networks	$\rho_\tau$				
	MF-OH	MF-DF	MF - Motif-GT	MF - Motif-DF	MF - h-backbone
Zachary's karate club	0.98	0.95	0.97	0.92	0.97
Intra-organisational	0.99	0.98	0.81	0.69	0.92
Freeman's EIES	0.99	0.97	0.90	0.68	0.78
Train bombing	1.00	0.97	0.94	0.60	0.94
Les Miserables	0.99	0.97	0.85	0.73	0.84
Game of thrones	0.95	0.92	0.94	0.92	0.94
C.elegans Neural	0.99	0.97	0.90	0.79	0.94
US Top-500 Airport	0.99	0.97	0.94	0.94	0.84
US Airoport	0.99	0.99	0.86	0.72	0.88
Openflights	0.99	0.98	0.89	0.67	0.92
Facebook-like Forum	0.90	0.97	0.83	0.60	0.83
Facebook-like Social	0.90	0.98	0.80	0.50	0.81
US Power Grid	0.92	0.86	0.92	0.65	0.82
Scientific Collaboration	0.99	0.97	0.89	0.72	0.84

**Table 6**

The relative difference of the average link betweenness  $\Delta r_{<\beta>}$  between the Modular filtering backbone and the alternative ones. MF stands for the Modular filtering backbone. OH stands for the Overlapping nodes and Hubs backbone. DF stands for the Disparity Filter backbone. Motif-GT refers to the motif-based global threshold method. Motif-DF refers to the motif-based disparity filter.  $\Delta r = (R_0 - R)/R_0$ , where  $R$  is the value of an effectiveness measure for a given backbone extraction method and  $R_0$  represents the measure considered as a reference. The MF backbone represents the reference method. A positive value of  $\Delta r$  means that the MF backbone is more efficient than the alternative one. Each value is the average of 10 SLPA simulation runs. The standard deviation is omitted in this table due to its small value (around  $10^{-3}$ ).

Networks	$\Delta r_{<\beta>}$				
	MF-OH	MF-DF	MF - Motif-GT	MF - Motif-DF	MF - h-backbone
Zachary's karate club	8.00	70.89	10.66	66.67	3.85
Intra-organisational	15.00	76.92	21.05	53.33	9.52
Freeman's EIES	2.92	28.18	8.46	17.50	2.17
Train bombing	12.36	36.05	17.11	32.45	11.11
Les Miserables	1.51	94.29	65.85	78.95	1.49
Game of thrones	81.13	47.69	39.13	41.18	31.51
C.elegans Neural	23.33	60.87	32.14	51.02	15.63
US Top-500 Airport	12.29	79.45	75.45	78.64	4.80
US Airoport	16.50	87.85	17.24	73.09	8.51
Openflights	14.29	78.57	60.00	61.65	12.04
Facebook-like Forum	16.72	78.67	52.27	59.71	12.79
Facebook-like Social	15.71	90.07	41.85	67.98	9.43
US Power Grid	3.13	98.00	23.75	60.71	1.33
Scientific Collaboration	8.79	88.57	45.05	82.44	4.32

and the Motif-DF. Note that the Modular filtering backbone and the Motif-DF displays the lowest correlation values. As expected, these results corroborate the findings of the previous analysis.

### 4.3. Effectiveness analysis

We investigate the effectiveness of the proposed backbone and its alternatives using the average link betweenness, the average link weight, and the entropy. Results are reported in tables 6, 7 and 8 for all the networks under test. These tables represent the relative difference  $\Delta r$  of the three measures computed between the Modular filtering backbone and the alternatives. Note that a positive value of  $\Delta r$  means that the proposed backbone is more effective than the alternative.

**Table 7**

The relative difference of the link weights  $\Delta r_{\langle w \rangle}$  between the Modular filtering backbone and the alternative ones. MF stands for the Modular filtering backbone. OH stands for the Overlapping nodes and Hubs backbone. DF stands for the Disparity Filter backbone. Motif-GT refers to the motif-based global threshold method. Motif-DF refers to the motif-based disparity filter.  $\Delta r = (R_0 - R)/R_0$ , where  $R$  is the value of an effectiveness measure for a given backbone extraction method and  $R_0$  represents the measure considered as a reference. The MF backbone represents the reference method. A positive value of  $\Delta r$  means that the MF backbone is more efficient than the alternatives. Each value is the average of 10 SLPA simulation runs. The standard deviation is omitted in this table due to its value (it ranges from 0.1 to 0.3).

Networks	$\Delta r_{\langle w \rangle}$				
	MF-OH	MF-DF	MF - Motif-GT	MF - Motif-DF	MF - h-backbone
Zachary's karate club	2.89	13.02	-15.64	9.54	5.64
Intra-organisational	-3.46	2.29	-54.02	1.83	1.36
Freeman's EIES	-1.19	16.82	-36.01	4.17	10.62
Train bombing	-7.25	4.07	-47.54	2.40	4.92
Les Miserables	-9.61	18.18	-13.33	15.66	11.90
Game of thrones	-4.64	5.54	-23.51	4.15	2.13
C.elegans Neural	-3.38	4.68	-31.11	1.54	2.80
US Top-500 Airport	-8.48	17.65	-9.09	11.11	7.67
US Airoport	-9.09	11.11	-9.29	9.62	5.37
Openflights	-10.23	13.64	-15.89	9.41	5.48
Facebook-like Forum	-5.15	18.18	-31.01	11.71	8.69
Facebook-like Social	-4.19	8.86	-6.24	3.73	2.48
US Power Grid	-4.03	3.79	-18.37	2.94	2.86
Scientific Collaboration	-1.90	1.03	-10.87	0.66	0.06

**Table 8**

The relative difference of the entropy  $\Delta r_H$  computed between the Modular filtering backbone and the alternative ones. MF stands for the Modular filtering backbone. OH stands for the Overlapping nodes and Hubs backbone. DF stands for the Disparity Filter backbone. Motif-GT refers to the motif-based global threshold method. Motif-DF refers to the motif-based disparity filter.  $\Delta r = (R_0 - R)/R_0$ , where  $R$  is the value of an effectiveness measure for a given backbone extraction method and  $R_0$  represents the measure considered as a reference. The MF backbone represents the reference method. A positive value of  $\Delta r$  means that the MF backbone is more efficient than the alternative one. Each value is the average of 10 SLPA simulation runs. The standard deviation is omitted in this table due to its small value (it ranges from 0.1 to 0.3).

Networks	$\Delta r_H$					
	MF-Original	MF-OH	MF-DF	MF - Motif-GT	MF - Motif-DF	MF - h-backbone
Zachary's karate club	-16.87	7.83	8.92	13.85	8.28	15.54
Intra-organisational	-17.00	13.36	15.67	18.76	14.85	20.63
Freeman's EIES	-21.08	2.72	3.76	11.32	3.43	35.70
Train bombing	-21.49	11.58	15.03	19.80	13.73	22.24
Les Miserables	-16.49	1.65	3.73	9.26	2.93	29.63
Game of thrones	-20.77	24.92	28.79	34.72	27.52	49.08
C.elegans Neural	-17.37	81.01	88.55	81.14	84.53	71.66
US Top-500 Airport	-31.12	31.33	32.65	38.10	28.74	47.27
US Airoport	-30.52	51.56	54.66	64.28	53.26	79.66
Openflights	-49.56	7.80	8.89	11.38	8.87	16.36
Facebook-like Forum	-44.25	6.95	10.26	18.65	8.20	31.62
Facebook-like Social	-48.36	7.80	8.89	14.09	8.53	28.45
US Power Grid	-42.41	18.25	20.66	32.01	20.22	39.39
Scientific Collaboration	-47.46	18.03	19.22	24.49	18.67	33.32

#### 4.3.1. Average betweenness

The average link betweenness designates the amount of information flow passing through all the edges of a given backbone. A higher value means that the diffusion of information in the backbone is more effective. One can notice in Table 6 that the relative difference of the average link betweenness between the Modular filtering and the overlapping nodes and hub backbones is positive whatever the network. The gain of the proposed backbone over the overlapping nodes and hub backbone ranges from 1.51% to 81%. Comparisons with the disparity filter, Motif-GT, and Motif-DF methods reveal higher gains. However, the values of the relative difference of the average link betweenness between the proposed method and the h-backbone are very small. Indeed, this backbone extraction method uses the link be-

tweenness and the h-index to select the relevant edges. That is why the h-backbone has a very high average link betweenness as compared to the other alternatives. Moreover, the Modular filtering backbone manages to select links and nodes, allowing the connection of the different modules of the network. The majority of these links come from the global component. These inter-community links and bridge nodes play a major role in diffusing the information across communities. It is the reason why the links selected by the proposed backbone are a better information gateway of the original network as compared to the alternative backbone extractors. Generally speaking, Modular filtering is the most effective backbone extractor in terms of link betweenness, followed by the h-backbone. Then, the overlapping nodes and hub backbone, the Motif-GT, the Motif-DF follow. The disparity filter backbone is the last one.

#### 4.3.2. Average weight

Let's now turn to the average link weight. The higher its value, the more relevant the links extracted from the original network. Table 7 reports the relative difference of the average link weight between the Modular filtering and the alternative methods. One can notice that the values related to the proposed method and the overlapping nodes and hub backbones are very often negative. The overlapping nodes and hub backbone generally possess connections with a higher strength than the proposed one. The gain ranges from 1.19% to 10.23%. It retains the relationships with high weights principally, while the proposed methods give more importance to links between communities even if their weights are smaller.

This behavior is even more pronounced when comparing with the Motif-GT backbone extractor. Indeed, Motif-GT selects edges belonging to a maximum number of motifs. These edges tend to have the highest weights of the network. One can notice in small-size networks that the Modular filtering backbone has many links with small weights. These links come from the global component. They act as inter-community or bridge links. Although their weights are small, they are essential to irrigate all the network communities. That explains why the proposed backbone has a lower average link weight than both the overlapping nodes and hub and the Motif-GT backbones. However, comparisons with the other alternative backbone extractors are more convincing. Indeed, the relative difference of this measure between the Modular filtering and the remaining alternative backbones is always positive. The highest gain is over the disparity filter backbone. It ranges from 4.03% to 18.18%. Indeed, the disparity filter misses some very relevant links. Remember that there is no connection between Javert and Jean Valjean in "Les Misérables" disparity filter backbone. Thus, the proposed extraction method retains more relevant links than the disparity filter, the Motif-DF, and the h-backbone. Overall, the Motif-GT is the most effective method for keeping connections with the highest weights. The Overlapping nodes and hub backbone follows it. The Modular filtering backbone ranks third. The h-backbone comes in the fourth position. The Motif-DF backbone follows. Finally, the disparity filter backbone is less effective. Another advantage worth noticing is that the backbone extracted by the Modular filter is formed with the main connected components of the original network. In contrast, the backbones generated by some alternative methods (such as disparity filter) can contain several components. The sub-network topology developed by the proposed method, therefore, preserves more the connectivity of the original network.

#### 4.3.3. Entropy

Finally, we compare the entropy of the different backbones. This measure indicates the quantity of information contained in an information source. Here, the entropy is computed based on the degree distribution of the nodes of each backbone. A higher value of this measure means that nodes selected by the backbone preserve more information of the original network. The relative difference of the entropy between the Modular filtering backbone and the original network is reported in Table 8. As expected their values are always negative. This is because the backbone is a reduced representation of the original network, and therefore it contains less information.

Table 8 reports the relative difference of entropy between the proposed backbone and the alternative (the overlapping nodes and hub, disparity filter, Motif-GT, Motif-DF and the h-backbone). For all the networks, the values are always positive. This demonstrates the better ability of the Modular filtering extractor to preserve the information of the original network. The highest gain is over the Motif-GT backbone. It reaches even values greater than 64%. As mentioned in the previous subsection, the Modular filtering backbone extracts relevant nodes in almost all the communities of the network. It appears visually in Les Misérables and Game of Thrones networks in figs. 4 and 6. Indeed, nodes with all the various colors are preserved by the Modular filtering backbone while some colors do not appear in the alternative backbone. Moreover, it selects also many important links connecting multiple modules of the original network. Therefore, the Modular filtering backbone retains more information about the original network as compared to the alternative ones. It is followed by the overlapping nodes and hub backbone. The Motif-DF comes just after.

Then, the disparity filter is in the fourth position. The Motif-GT and the h-backbones are the less effective backbone extractors in terms of entropy.

#### 4.3.4. Summary

We use the average link betweenness, the average link weight, and the entropy to evaluate the performance of the backbone extraction methods. Results show that the Modular filtering backbone is the most effective in terms of average link betweenness and entropy. Motif-GT exhibits the best performance in terms of average link weights. Note, however, that Motif-GT has a very high computational complexity. It is around  $O(n^k)$  ( $n$  is the size of the network and  $k$  is the size of the motif). It is one of its main drawbacks. In contrast, the proposed method uses only local information at the level of the overlapping nodes. Therefore, it depends only on the complexity of the used community detection algorithm. That is what makes Modular Filtering a very efficient method also in terms of computational complexity.

## 5. Conclusion

Large-scale network analysis and understanding is quite a challenging issue. Therefore, it is essential to find a way to select the genuinely significant nodes and links to obtain a simplified and accurate portrayal of such networks. A local approach for the extraction of backbones in weighted networks based on overlapping nodes, hubs, and inter-community links is proposed and investigated in this study.

It is called the ‘Modular filtering backbone’ method. It extracts backbones based on the main elements of the community structure (overlapping nodes, hubs, and bridging edges). This local-based filtering approach aims at selecting the most influential nodes and edges, allowing the propagation of the most significant amount of information in the network. The combination of the local and the global components of the original network form the Modular filtering backbone. The local component regroups the subnetworks containing the overlapping nodes and the hubs. Then, one removes the links with the lowest weights as long as the components stay connected. The global component contains the inter-community links and their attached nodes. After that, the disparity filter algorithm is used to extract only the most relevant inter-community links of the network.

We perform a comparative analysis of the proposed extraction method with five alternative backbone extraction methods (the overlapping nodes and hub backbone, disparity filter, Motif-GT, Motif-DF, and the h-backbone) real-world networks. The proportion of common nodes, the rank-biased overlap, and the correlation are used. The first two measurements performed on the Modular filtering backbone and the overlapping nodes and hub backbone show a substantial similarity between these extraction methods. At the same time, the correlation indicates that there is a solid association between the two sets of nodes. It implies that both backbones are pretty similar. However, this similarity gets lower when the mixing parameter value increases.

One observes a lower similarity when comparing with the other backbone extractors. Generally, there is a smaller overlap. The lower correlation values also corroborate this behavior. Results show that the Modular filtering backbone is the most effective method compared to all its alternatives in terms of link betweenness and entropy. It is not the case for the average link weights. The overlapping nodes and hub backbone and the Motif-GT outperform the proposed one. Indeed, they rely mainly on betweenness, while the proposed backbone extractor incorporates low-weight inter-community links that tie the various network communities. Besides, the proposed method has lower computational complexity than the global-based approaches (Motif-GT and Motif-DF).

## A. My Appendix

### CRedit authorship contribution statement

**Zakariya Ghalmane:** Conceptualization of this study, Methodology, Software, Writing - Original draft preparation. **Chantal Cherifi:** Conceptualization of this study, Methodology, Writing - Original draft preparation. **Hocine Cherifi:** Conceptualization of this study, Methodology, Writing - Original draft preparation. **Mohammed El Hassouni:** Conceptualization of this study, Methodology, Writing - Original draft preparation.

### Declarations of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

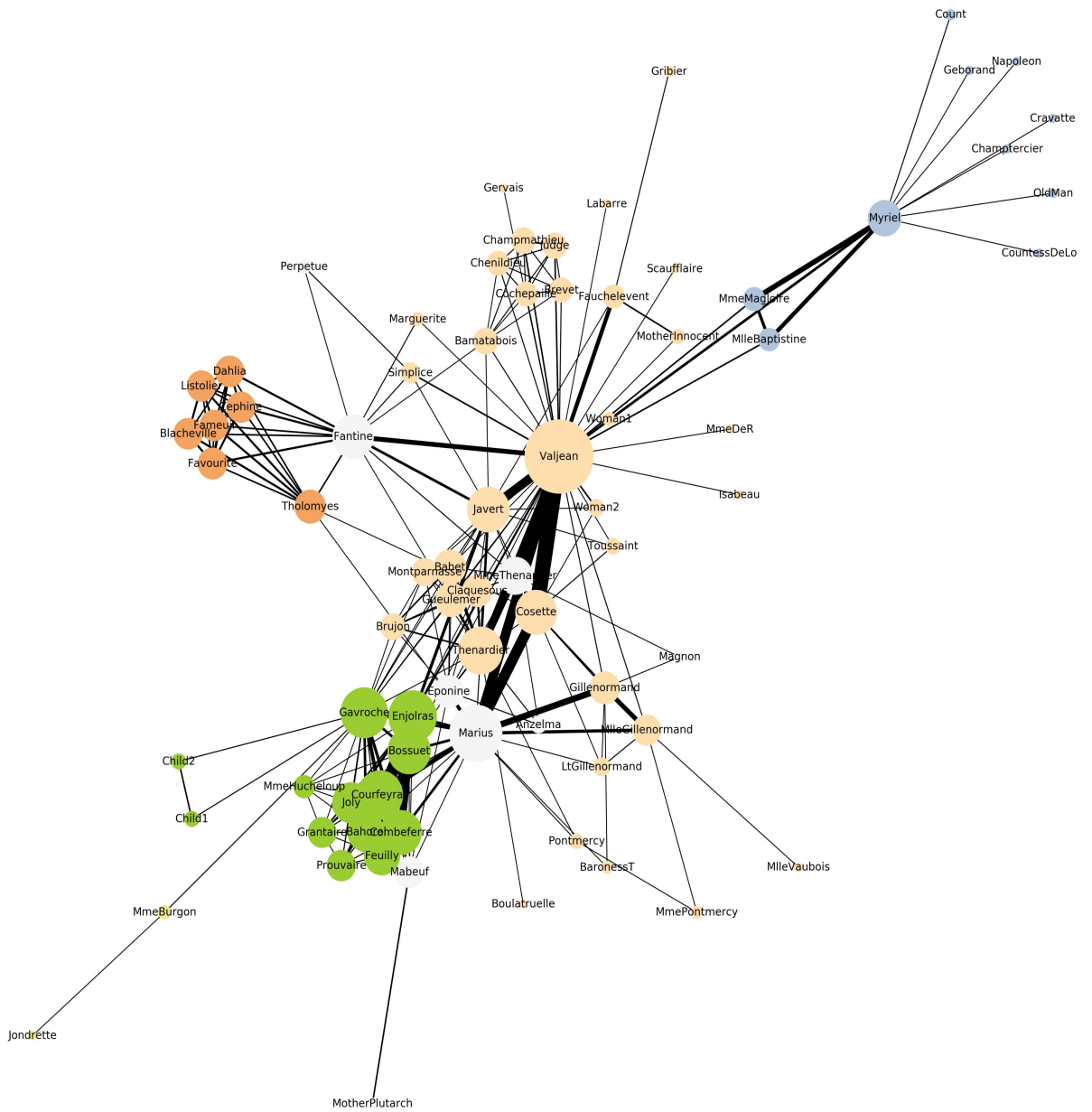
## Acknowledgment

This research is carried out as a part of the project « Plateforme logicielle d'intégration de stratégies d'immunisation contre la pandémie COVID-19 » funded by the grant of the Hassan II Academy of Sciences and Technology of Morocco.

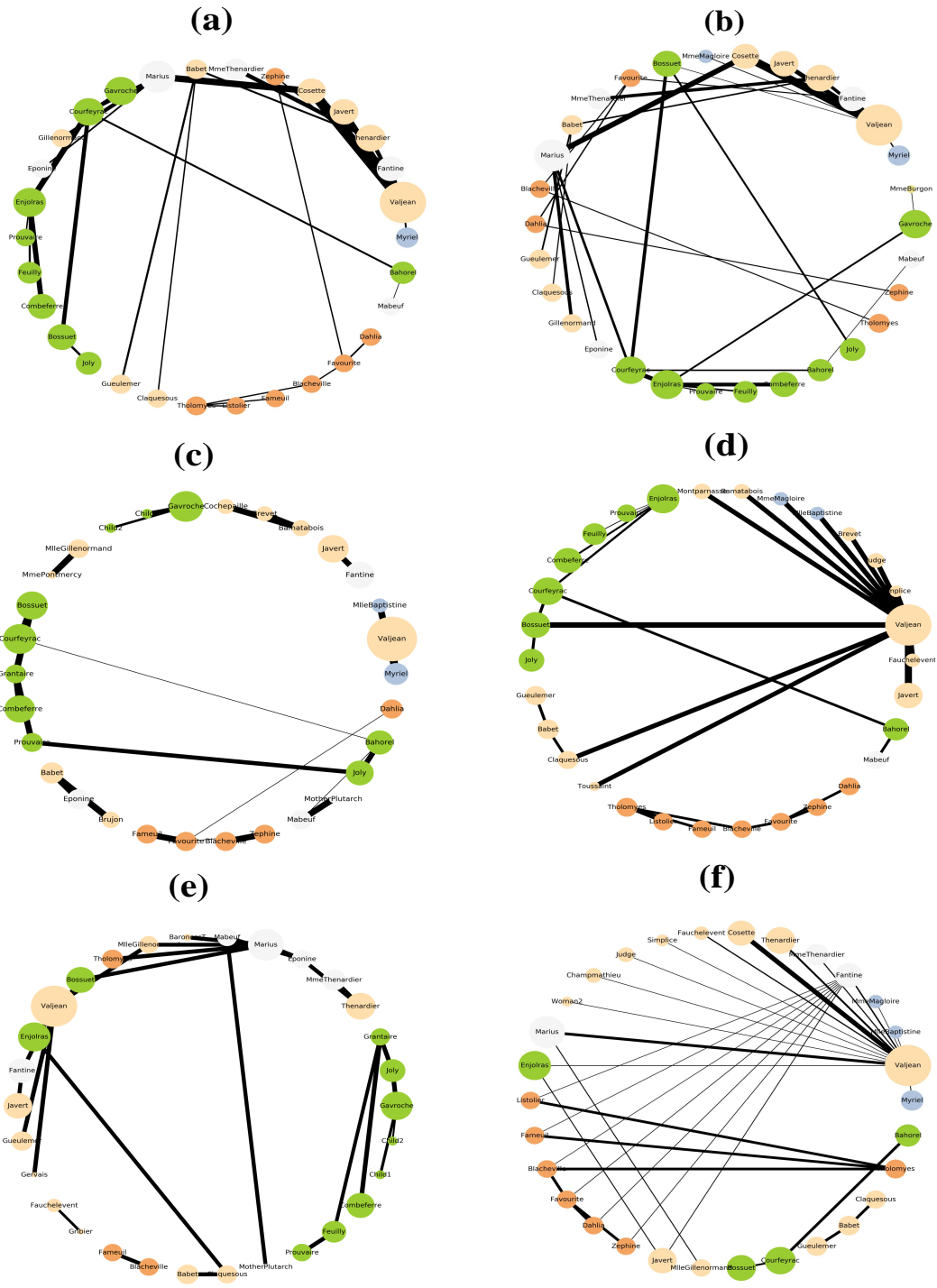
## References

- [1] Chu CC, Iu HHC. Complex networks theory for modern smart grid applications: A survey. *IEEE Journal on Emerging and Selected Topics in Circuits and Systems*. 2017;7(2):177–191.
- [2] Bassett DS, Sporns O. Network neuroscience. *Nature neuroscience*. 2017;20(3):353–364.
- [3] Yoshida T, Yamada Y. A Community Structure-Based Approach for Network Immunization. *Computational Intelligence*. 2017;33(1):77–98.
- [4] Gong K. Local immunization based on Degree-Community-Bridge-Find in heterogeneity community networks. In: *International Conference on Logistics Engineering, Management and Computer Science (LEMCS 2014)*. Atlantis Press; 2014.
- [5] Ghalmane Z, Cherifi C, Cherifi H, El Hassouni M. Exploring hubs and overlapping nodes interactions in modular complex networks. *IEEE Access*. 2020;.
- [6] Ghalmane Z, El Hassouni M, Cherifi C, Cherifi H. k-Truss Decomposition for Modular Centrality. In: *2018 9th International Symposium on Signal, Image, Video and Communications (ISIVC)*. IEEE; 2018. p. 241–248.
- [7] van den Heuvel MP, Kahn RS, Goñi J, Sporns O. High-cost, high-capacity backbone for global brain communication. *Proceedings of the National Academy of Sciences*. 2012;109(28):11372–11377.
- [8] Cao J, Ding C, Shi B. Motif-based functional backbone extraction of complex networks. *Physica A: Statistical Mechanics and its Applications*. 2019;526:121123.
- [9] da Fontoura Costa L. The hierarchical backbone of complex networks. *Physical review letters*. 2004;93(9):098702.
- [10] Zhan XX, Hanjalic A, Wang H. Information diffusion backbones in temporal networks. *Scientific reports*. 2019;9(1):1–12.
- [11] Domagalski R, Neal Z, Sagan B. backbone: An R Package for Backbone Extraction of Weighted Graphs. *arXiv preprint arXiv:191212779*. 2019;.
- [12] Zweig KA, Kaufmann M. A systematic approach to the one-mode projection of bipartite graphs. *Social Network Analysis and Mining*. 2011;1(3):187–218.
- [13] Neal Z. The backbone of bipartite projections: Inferring relationships from co-authorship, co-sponsorship, co-attendance and other co-behaviors. *Social Networks*. 2014;39:84–97.
- [14] Neal Z. Identifying statistically significant edges in one-mode projections. *Social Network Analysis and Mining*. 2013;3(4):915–924.
- [15] Jo WS, Park J, Luhur A, Kim BJ, Ahn YY. Extracting hierarchical backbones from bipartite networks. *arXiv preprint arXiv:200207239*. 2020;.
- [16] Boyack KW, Klavans R, Börner K. Mapping the backbone of science. *Scientometrics*. 2005;64(3):351–374.
- [17] Chen M, Li L, Wang B, Cheng J, Pan L, Chen X. Effectively clustering by finding density backbone based-on kNN. *Pattern Recognition*. 2016;60:486–498.
- [18] Gfeller D, De Los Rios P. Spectral coarse graining of complex networks. *Physical review letters*. 2007;99(3):038701.
- [19] Goh KI, Salvi G, Kahng B, Kim D. Skeleton and fractal scaling in complex networks. *Physical review letters*. 2006;96(1):018701.
- [20] Grady D, Thiemann C, Brockmann D. Robust classification of salient links in complex networks. *Nature communications*. 2012;3(1):1–10.
- [21] Carmi S, Havlin S, Kirkpatrick S, Shavitt Y, Shir E. A model of Internet topology using k-shell decomposition. *Proceedings of the National Academy of Sciences*. 2007;104(27):11150–11154.
- [22] Chalupa J, Leath PL, Reich GR. Bootstrap percolation on a Bethe lattice. *Journal of Physics C: Solid State Physics*. 1979;12(1):L31.
- [23] Zhang X, Zhu J. Skeleton of weighted social network. *Physica A: Statistical Mechanics and its Applications*. 2013;392(6):1547–1556.
- [24] Serrano MÁ, Boguná M, Vespignani A. Extracting the multiscale backbone of complex weighted networks. *Proceedings of the national academy of sciences*. 2009;106(16):6483–6488.
- [25] Zhang RJ, Stanley HE, Fred YY. Extracting h-Backbone as a Core Structure in Weighted Networks. *Scientific reports*. 2018;8(1):1–7.
- [26] Ghalmane Z, Cherifi C, Cherifi H, El Hassouni M. Extracting backbones in weighted modular complex networks. *Scientific Reports*. 2020;10(1):1–18.
- [27] Cherifi H, Palla G, Szymanski BK, Lu X. On community structure in complex networks: challenges and opportunities. *Applied Network Science*. 2019;4(1):1–35.
- [28] Xu Y. A spectral method to detect community structure based on the communicability modularity. *Physica A: Statistical Mechanics and its Applications*. 2020;537:122751.
- [29] Fortunato S, Hric D. Community detection in networks: A user guide. *Physics Reports*. 2016;659:1–44.
- [30] Guo K, He L, Chen Y, Guo W, Zheng J. A local community detection algorithm based on internal force between nodes. *Applied Intelligence*. 2020;50(2):328–340.
- [31] Guimera R, Sales-Pardo M, Amaral LA. Classes of complex networks defined by role-to-role connectivity profiles. *Nature physics*. 2007;3(1):63–69.
- [32] Tulu MM, Hou R, Younas T. Identifying Influential Nodes Based on Community Structure to Speed up the Dissemination of Information in Complex Network. *IEEE ACCESS*. 2018;6:7390–7401.
- [33] Kitromilidis M, Evans TS. Community Detection with Metadata in a Network of Biographies of Western Art Painters. *arXiv preprint arXiv:180207985*. 2018;.
- [34] Palla G, Derényi I, Farkas I, Vicsek T. Uncovering the overlapping community structure of complex networks in nature and society. *nature*. 2005;435(7043):814.

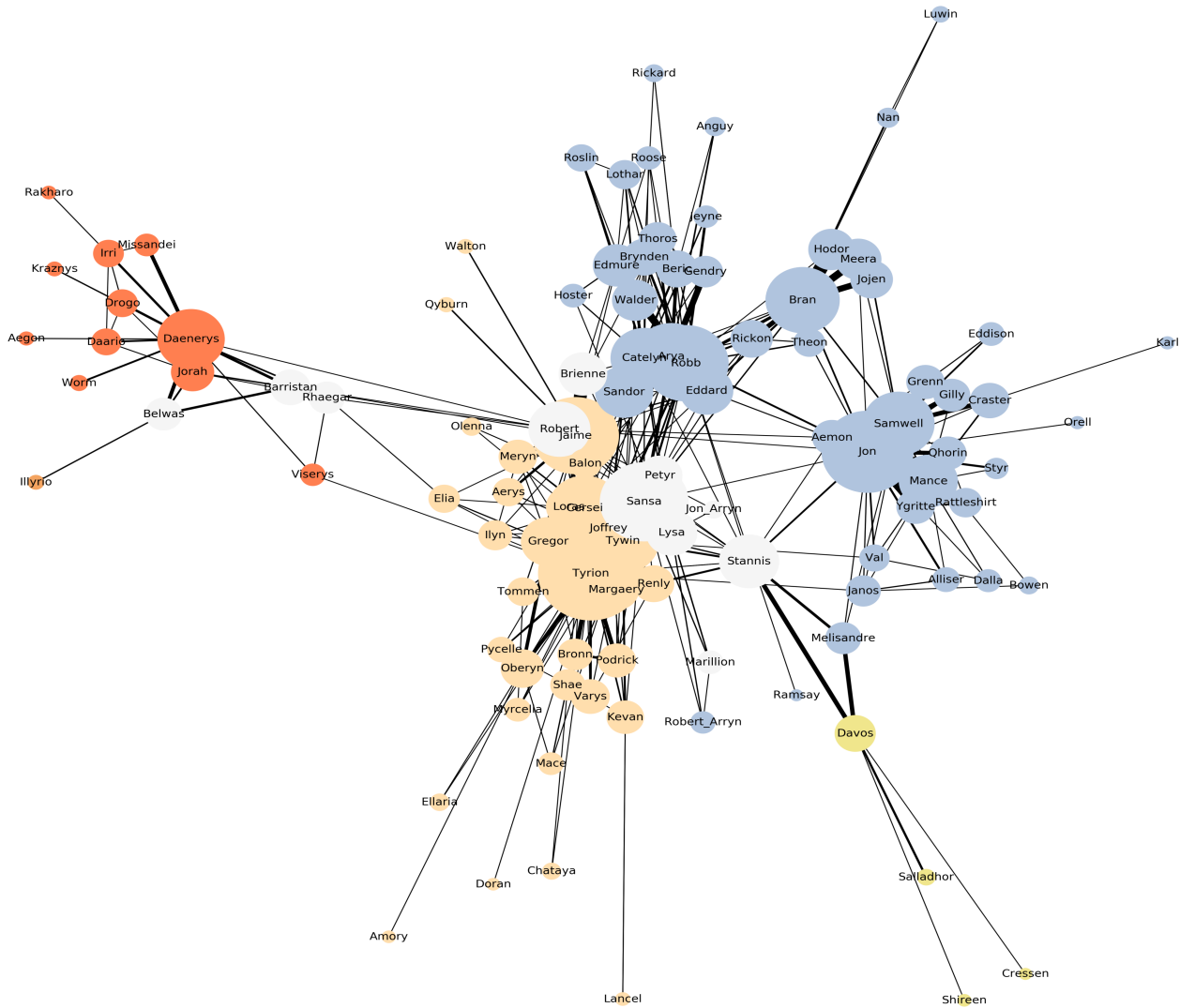
- [35] Xu EH, Hui PM. Uncovering complex overlapping pattern of communities in large-scale social networks. *Applied Network Science*. 2019;4(1):1–16.
- [36] Ghalmane Z, El Hassouni M, Cherifi H. Betweenness centrality for networks with non-overlapping community structure. In: 2018 IEEE workshop on complexity in engineering (COMPENG). IEEE; 2018. p. 1–5.
- [37] Jiang H, Liu Z, Liu C, Su Y, Zhang X. Community detection in complex networks with an ambiguous structure using central node based link prediction. *Knowledge-Based Systems*. 2020; p. 105626.
- [38] Kumar M, Singh A, Cherifi H. An Efficient Immunization Strategy Using Overlapping Nodes and Its Neighborhoods. In: Companion of the The Web Conference 2018 on The Web Conference 2018. International World Wide Web Conferences Steering Committee; 2018. p. 1269–1275.
- [39] Sathyakala M, Sangeetha M. A weak clique based multi objective genetic algorithm for overlapping community detection in complex networks. *Journal of Ambient Intelligence and Humanized Computing*. 2020; p. 1–11.
- [40] Ghalmane Z, El Hassouni M, and Cherifi H. Immunization of networks with non-overlapping community structure. *Social Network Analysis and Mining*, 9(1):45, 2019.
- [41] Dao VL, Bothorel C, Lenca P. Community structure: A comparative evaluation of community detection methods. *Network Science*. 2020;8(1):1–41.
- [42] Gupta N, Singh A, Cherifi H. Community-based immunization strategies for epidemic control. In: Communication Systems and Networks (COMSNETS), 2015 7th International Conference on. IEEE; 2015. p. 1–6.
- [43] Yang J, Leskovec J. Structure and overlaps of ground-truth communities in networks. *ACM Transactions on Intelligent Systems and Technology (TIST)*. 2014;5(2):1–35.
- [44] Kudelka M, Ochodkova E, Zehnalova S, Plesnik J. Ego-zones: non-symmetric dependencies reveal network groups with large and dense overlaps. *Applied Network Science*. 2019;4(1):81. doi:10.1007/s41109-019-0192-6.
- [45] Opsahl T, Agneessens F, Skvoretz J. Node centrality in weighted networks: Generalizing degree and shortest paths. *Social networks*. 2010;32(3):245–251.
- [46] Xie J, Szymanski BK, Liu X. Slpa: Uncovering overlapping communities in social networks via a speaker-listener interaction dynamic process. In: 2011 IEEE 11th International Conference on Data Mining Workshops. IEEE; 2011. p. 344–349.
- [47] Webber W, Moffat A, Zobel J. A similarity measure for indefinite rankings. *ACM Transactions on Information Systems (TOIS)*. 2010;28(4):1–38.
- [48] Gray RM. *Entropy and information theory*. Springer Science & Business Media; 2011.



**Figure 3:** Representation of Les Misérables network. Nodes with the same color belong to the same community, and the overlapping nodes are in grey. The size of the nodes is proportional to their weighted degree. The size of links is proportional to their weights. The community structure is revealed using the SLPA detection algorithm.

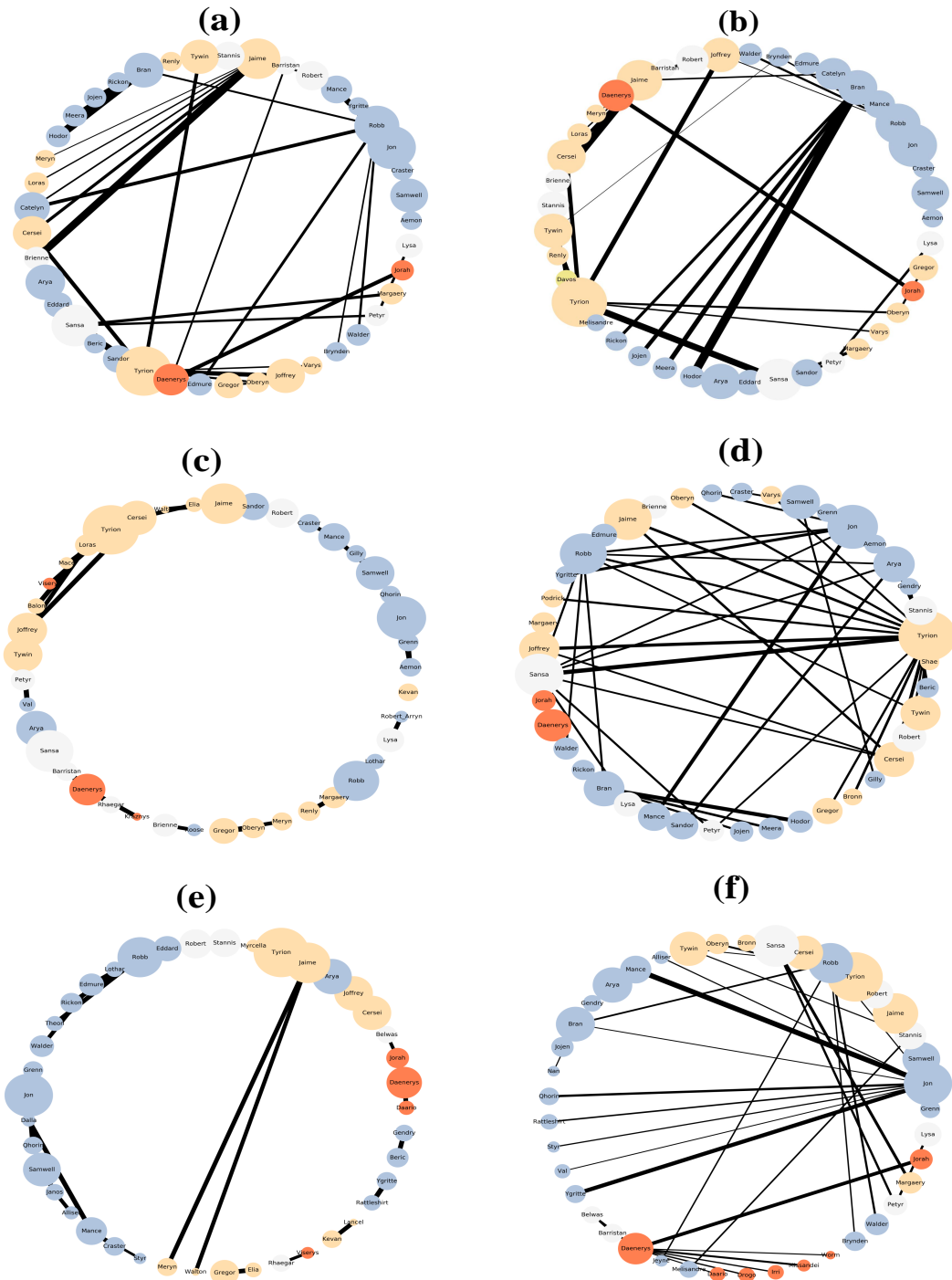


**Figure 4:** The backbone extraction of different methods for Les Misérables network. (a) Modular filtering backbone, (b) Overlapping nodes and hubs backbone, (c) Disparity filter backbone, (d) Motif-GT backbone, (e) Motif-DF backbone, and (f) h-backbone. Nodes with the same color belong to the same community, and the overlapping nodes are in grey. The size of the nodes is proportional to their weighted degree. The size of links is proportional to their weights. The community structure is revealed using the SLPA detection algorithm.



**Figure 5:** Representation of the Game of thrones network. Nodes with the same color belong to the same community, and the overlapping nodes are in grey. The size of the nodes is proportional to their weighted degree. The size of links is proportional to their weights. The community structure is revealed using the SLPA detection algorithm.

## Extracting Modular-based Backbones in Weighted Networks



**Figure 6:** The backbone extraction of different methods for the Game of thrones network. (a) Modular filtering backbone, (b) Overlapping nodes and hubs backbone and, (c) Disparity filter backbone, (d) Motif-GT backbone, (e) Motif-DF backbone, and (f) h-backbone. Nodes with the same color belong to the same community, and the overlapping nodes are in grey. The size of the nodes is proportional to their weighted degree. The size of links is proportional to their weights. The community structure is revealed using the SLPA detection algorithm.