A backbone extraction method for complex weighted networks

Zakariya Ghalmane[1,3], Chantal Cherifi[2], Hocine Cherifi[3], and Mohammed El Hassouni[1,4]

[1] LRIT, Mohammed V University, Rabat, Morocco,
[2] DISP Lab, University of Lyon 2, Lyon, France
[3] LE2I, University of Burgundy, Dijon, France
[4] FLSH, Mohammed V University, Rabat, Morocco
zakaria.ghalmane@gmail.com

1 Introduction

Network science provides effective tools to model and analyze complex systems. However, the increasing size of real-world networks becomes a major obstacle in order to understand their structure and topological features. Therefore, reducing the original network into a smaller one while preserving its information is an important issue. Extracting the so-called backbone of a network is a very challenging problem that is generally solved either by coarse-graining or filter-based methods. Coarse-graining methods reduce the network size by grouping similar nodes, while filter-based methods map the network by discarding nodes or edges based on a statistical property. In this work, we propose and investigate one filter-based method exploiting the overlapping community structure in order to extract the backbone in weighted networks. In the proposed method, called "overlapping nodes ego backbone", the backbone is formed simply from the set of overlapping nodes and their neighbors. It is based on the results of a previous study [1] showing that the overlapping nodes are neighbors of the highly connected nodes (hubs) of the network. Indeed, hubs and overlapping nodes are located at the heart of the network. In this method, the links with low weights are removed from the network as long as a backbone with a single connected component is preserved. Experiments have been performed on real-world weighted networks originating from various domains (social, co-appearance, collaboration, biological, and technological) and different sizes. Results of the comparison with the disparity filter method [2], which is considered as one of the most influential alternative filtering methods demonstrate the greater ability of the proposed backbone extraction method to uncover the most relevant parts of the network.

2 Methods

In this study, the backbone of weighted networks is extracted using a filter-based algorithm. It is called the "overlapping nodes ego backbone" which consists of defining the set of overlapping nodes and their neighbors as the network backbone. Let's consider a network $G(V,E)$, where $V = \{v_1, v_2, ..., v_n\}$ and $E = \{(v_i, v_j) \setminus v_i, v_j \in V\}$ denotes

Table 1. N is the size of the network. on is the fraction of overlapping nodes. A_n represents the estimated values the proportion of common nodes of two backbones. A_t is the proportion of the 10% of the most highly ranked nodes belonging to the backbone. $<\beta>$ is the average node betweenness. OE stands for the overlapping nodes ego backbone, while DF stands for the disparity filter backbone. Each value is the average of 10 SLPA simulation runs. The standard deviation is excluded from this table due to its very small value (it is around 10^{-3}).

Network	N	$A_n(\%)$	$A_t(\%)$		$<\beta>$	
		OE-DF	**OE**	**DF**	**OE**	**DF**
Zachary's karate club	34	68	100	100	0.088	0.079
Intra-organisational	46	84.61	100	50	0.028	0.013
Freeman's EIES	48	80	100	80	0.014	0.011
Train bombing	62	94.73	100	100	0.077	0.047
Les Miserables	77	85.11	100	93.33	0.057	0.035
Game of thrones	107	68.75	100	90	0.053	0.034
C.elegans Neural	306	64.04	100	75.86	0.012	0.009
Facebook-like Forum	899	61.71	100	70.78	0.005	0.004
Facebook-like Social	1899	75.04	100	91.53	0.003	0.002
US Power Grid	4941	61.03	100	72.26	0.008	0.005
Scientific Collaboration	16726	57.83	97.05	68.19	0.008	0.006

the set of nodes and edges respectively. The algorithm of the model used to extract the backbone is detailed as follows:

Step 1: If not known, the community structure of the network using a given overlapping community detection algorithm is uncovered.

Step 2: Form the overlapping nodes ego sub-network. It is made of the union of the set of overlapping nodes and the set of their neighbors. It is obtained by removing all nodes that do not belong to one of the two sets.

Step 3: Remove the edges with low weights from the overlapping ego sub-network. To do so, edges are sorted in decreasing order according to their weights. Then, edges with low weights are removed as long as the sub-network component is not split into two components. This ensures that the backbone is formed with a single component.

Step 4: Control the size of the overlapping ego backbone with a parameter s. This parameter allows to preserve only the top-ranked nodes of this backbone. To this end, all the nodes of the overlapping nodes ego backbone are sorted in decreasing order according to the weighted degree centrality [3]. After that, according to the value of this threshold parameter, the nodes with low degrees are removed from the network.

3 Results

A set of experiments is performed to compare the overlapping nodes ego backbone with the disparity filter which is one of the most effective alternative backbones. Results

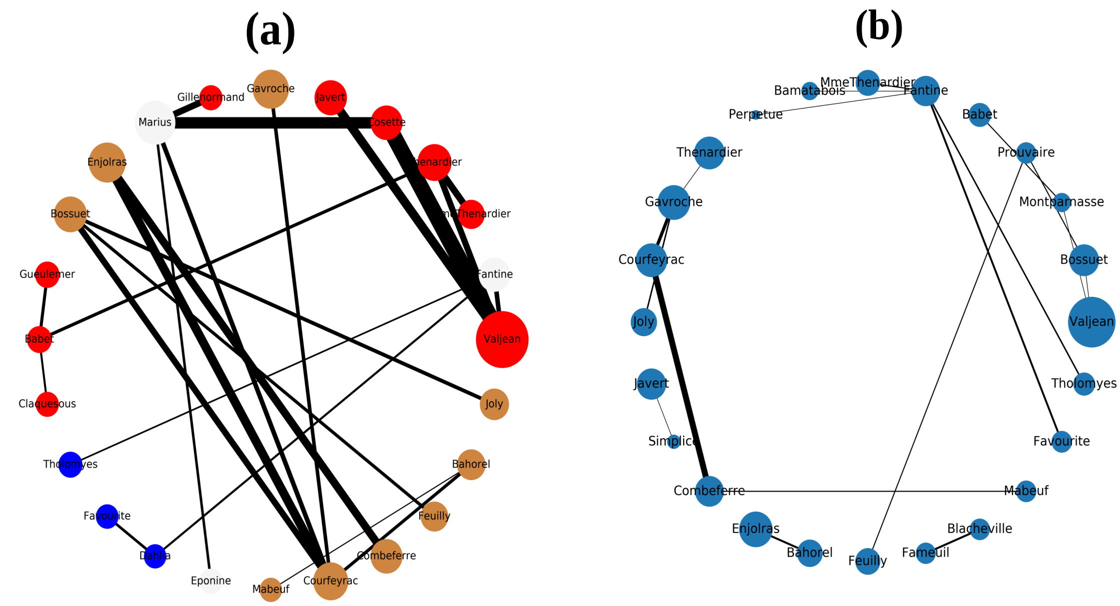

Fig. 1. The backbone extraction of different methods for Les Miserables network. (a) overlapping nodes ego backbone, (b) disparity filter backbone. Nodes are highlighted in different colors according to the community they belong to. Nodes with the same color belong to the same community while those in gray represent the overlapping nodes. The size of the nodes is proportional to their weighted degree, while the size of links is proportional to their weights. The community structure is revealed using the SLPA detection algorithm.

of the comparative evaluation of the backbone extraction methods are presented and analyzed. Samples of various real-world networks (social, co-appearance, biological, technological and collaboration networks) have been used in order to cover a wide range of situations. The Speaker-Listener Label Propagation Algorithm (SLPA) [4] is used to uncover the overlapping community structure of the networks. Note that in all the experiments, the size of the backbone is limited to 30% of the size of the original network by setting the parameter s to 0.3. Accordingly, the parameter α is tuned in order to obtain a disparity filter backbone of the same size.

At first, the proportion of common nodes extracted by the two backbone extraction methods is computed. It is defined as the fraction of the size of the intersection between the two sets divided by their size. Our goal is to check if they extract the same set of nodes or not. Table 1 reports the proportion of common nodes computed between the proposed backbone and the disparity filter backbone. Its values exhibit usually medium values while it has high values in some networks. The main difference lies in the fact that the disparity filter concentrates on links while the proposed method is based on nodes in its extraction process. This can be visualized in Les Miserables network illustrated in Figure 1. In this figure, the overlapping nodes ego backbone is represented in (a), while the disparity filter backbone is reported in (b). The color of nodes refers to the communities to which they belong. The overlapping nodes are colored in grey. The size of the nodes is proportional to their weighted degree and the thickness of the links is proportional to their weights. One can notice from this figure that the disparity filter backbone misses some very important nodes such as "Marius" and "Cosette" that are considered as one of the main characters in the Victor Hugo's novel. Thus, the backbone obtained by the proposed approach preserves almost all high-connectivity nodes and essential connections.

Furthermore, we also check if the top-ranked nodes of the original network are preserved in the backbones. The set of the 10% top-ranked nodes according to their weighted degree is computed in the original network. It is compared to the set of top-ranked nodes of the same size extracted from the backbones. Results reported in Table 1 show that both sets are identical most of the time for the proposed backbone. This remark is not valid for the disparity filter backbone which preserves a smaller proportion of the top-ranked nodes. It focuses on links rather than on nodes. That is the main difference between the proposed backbone and the disparity filter backbone that explain its lower ability to preserve hubs.

In the second part of the experiments, the performance of the two backbone extraction methods is compared by measuring the average betweenness. Table 1 reports the results for all networks under test. The average betweenness indicates how much information can pass through the nodes of the backbone. The values of the average betweenness of the proposed backbone are higher than the ones computed in the disparity filter backbone. This implies that the nodes extracted by the overlapping nodes ego backbone act as a better information gateway of the original network as compared to the disparity filter backbone. Its performance is higher than the disparity filter backbone for all the tested empirical networks. This confirms its ability to preserve nodes playing a major role in the network.

References

1. Ghalmane, Z., Cherifi, C., Cherifi, H., & El Hassouni, M. (2020). Exploring Hubs and Overlapping Nodes Interactions in Modular Complex Networks. IEEE Access, 8, 79650-79683.
2. Serrano, M. Á., Boguná, M., & Vespignani, A. (2009). Extracting the multiscale backbone of complex weighted networks. Proceedings of the national academy of sciences, 106(16), 6483-6488.
3. Opsahl, T., Agneessens, F., & Skvoretz, J. (2010). Node centrality in weighted networks: Generalizing degree and shortest paths. Social networks, 32(3), 245-251.
4. Xie, J., Szymanski, B. K., & Liu, X. (2011, December). Slpa: Uncovering overlapping communities in social networks via a speaker-listener interaction dynamic process. In 2011 ieee 11th international conference on data mining workshops (pp. 344-349). IEEE.