

Received April 12, 2020, accepted April 23, 2020, date of publication April 28, 2020, date of current version May 12, 2020.

Digital Object Identifier 10.1109/ACCESS.2020.2991001

Exploring Hubs and Overlapping Nodes Interactions in Modular Complex Networks

ZAKARIYA GHALMANE^{1,3}, CHANTAL CHERIFI², HOCINE CHERIFI³,
AND MOHAMMED EL HASSOUNI^{1,4}

¹LRIT, URAC, Rabat IT Center, Mohammed V University, Rabat 10500, Morocco

²DISP Laboratory, University Lyon 2, 69007 Lyon, France

³LIB EA 7534, University of Burgundy, 21078 Dijon, France

⁴FLSH, Mohammed V University, Rabat 10500, Morocco

Corresponding author: Zakariya Ghalmane (zakaria.ghalmane@gmail.com)

ABSTRACT In real-world networks, nodes are usually organized into modules or communities of densely connected nodes. In situations where nodes can belong to multiple communities we say that the communities overlap, and the nodes shared by more than one community are called the overlapping nodes. This occurs especially in social networks where an individual belongs to various social groups and organizations such as working circles, family, friendship or virtual groups on the Internet. Complex networks are known to have a heavy tail degree distribution. Indeed, they are organized with a vast majority of nodes with few interactions and a small set of highly connected nodes called hubs. In this paper, our goal is to study the relationship between the overlapping nodes and the hubs. Indeed, we suspect that the hubs are in the vicinity of the overlapping nodes. If this assumption is confirmed, it gives a new perspective on how the communities are organized and of the crucial importance of the overlapping nodes. In an attempt to investigate the ubiquity of this property, we perform series of experiments on various real-world networks with overlapping community structure. Results show that the hubs represent always a large proportion of the one-step neighbors of overlapping nodes. These results may have implications in various contexts. For example, searching for the hubs in large networks can be done starting from the overlapping nodes. Furthermore, this study may also provide new directions for designing new community detection algorithms.

INDEX TERMS Complex networks, community structure, overlapping nodes, hubs.

I. INTRODUCTION

In the study of complex networks, such as social, biological and information networks, many different topological features have been observed to occur commonly. One of the main common characteristic is the degree distribution of the nodes. It is well described by a non-homogeneous distribution with a heavy tail. This results in the majority of nodes share a low amount of connections and a small number of remaining nodes that have a large number of connections. The latter ones commonly referred as hubs [1] tend to be extremely influential. Indeed, in transportation networks, for example, the underlying topology based on a small number of hubs

allows efficient travel services. Hubs play also the role of super-spreaders in the context of epidemic spreading.

Community structure is another characteristic frequently observed. Indeed, the majority of real-world networks are found to be naturally partitioned into multiple modules or communities. Many studies have been conducted to model and to analyse the community structure of a network [2]–[14]. Until now, there is no consensual definition of the community structure. The most influential informal definition considers the community structure as a partition of the networks into groups of vertices which are densely interconnected while being loosely connected with vertices of other communities. In other words, a community is a subset of highly connected nodes sparsely connected with the rest of the network. This assumption has been challenged. Indeed, other works [15]–[22] have shown that real-world

The associate editor coordinating the review of this manuscript and approving it for publication was Muhammad Zakarya¹.

networks can display overlapping and nested communities. While non-overlapping communities are made of modules where nodes belong to a single community, in overlapping communities, nodes can belong to more than one community. This is typical of social networks, where individuals can belong to various groups at the same time such as family, school, hobby and so on. Our goal here is to go deeper into the understanding of the interplay between the microscopic and the mesoscopic organization of real-world complex networks. Indeed, the heavy-tail degree distribution of the networks is characterized by the emergence of a small set of highly connected nodes. Furthermore, overlapping nodes is one of the main organizing principles of modular networks. Investigating the relations between those two types of nodes is essential to develop an effective analysis and modeling tools for real-world networks. Analyzing the topological properties of the community structure is of prime interest. Indeed, it allows a better understanding of one of the main organizing principles encountered in real-world networks. It may also provide clues about the emergence of the community structure in real-world networks. Despite the interest of the issue, there are few works that dive deeply into the subject, in particular in the case of overlapping community structure. The main contribution of this paper is to characterize the relationship between the overlapping nodes and the hubs. Our aim is to give a clear answer about the relations between the hubs and the overlapping nodes in networks with overlapping community structure. To our knowledge, this is the first attempt to conduct a systematic study in order to check if hubs lie in the vicinity of overlapping nodes, and if this is a ubiquitous property of real-world complex networks. Based on a series of experiments (using different evaluation measures) performed on a variety of real-world networks originating from various fields, we show that the overlapping nodes are neighbors of the hubs. Exploring this property is a challenging issue from the perspective of the network structure. Indeed, our analysis sheds more light on how the communities are organized.

The rest of the paper is organized as follows. In Section 2, related work about the topological properties of the overlapping community structure of real-world networks presented. Section 3 introduces the data and the community detection algorithms used to perform the experiments. Section 4 is devoted to the presentation of the various measures and methods used to compare the set of overlapping nodes with the set of hubs. In Section 5, we report and discuss our empirical findings. Finally, conclusions are given in section 6.

II. RELATED WORK

In a seminal paper, Palla *et al.* [15] show that real-world networks can display significant overlap between communities. They introduce four relevant quantities in order to characterize the overlapping community structure in large networks. The membership number of a node quantifies the number of communities to which it belongs. The overlap size between two communities is the number of nodes they share. The community degree is the number of communities overlapping

with it. Finally, the community size is the number of nodes of a given community. In their work, they investigate the distributions of these four quantities. Their results show that the community size exhibits a power-law distribution. These latter show also that the power-law distribution presents a good fit for the overlap size as well as the membership number. However, the community degree distribution exhibits a different behavior. We can distinguish two parts: an exponential decay followed by a power-law tail.

Another important study has been conducted by Yang *et al.* [23] in order to characterize the overlapping areas of the community structure. The authors show that nodes belonging to the overlapping zones between communities are more densely connected than those belonging to the non-overlapping areas of the network. This behavior has been observed in a series of experiments conducted on six large networks of various origins. These results contradict the conventional assumption that the overlapping nodes are more sparsely connected than the non-overlapping parts of the communities. In their work, they study the edge probability of a pair of overlapping nodes as a function of the number of their shared communities. Results show that the edge probability between two overlapping nodes increases with the number of communities they have in common. In their experiments, they also investigate if the most connected node in each community belongs to the overlapping zone. More precisely, they study the probability that a hub belongs to the community overlap as a function of the overlap size. Results show that community hubs are not central in a community. They actually tend to reside in the overlapping zone. In addition, the probability that a hub belongs to the overlap area increases linearly with the overlap size. Another work reported in [8] studies the relationship between the transitivity and the community structure strength measured by the network modularity. Extensive experiments show that transitivity increases accordingly with the community structure strength. Furthermore, if a weak community structure is associated with a low transitivity value, the opposite is not true. A network with a close to zero transitivity can still have a well-defined community structure.

More recently Kudelka *et al.* [24] presented a new perspective on the problem of group detection bridging the gap between structural and ground-truth communities. Using the non-symmetric structural similarity between pairs of nodes, they introduce an algorithm to detect groups referred as zones. Their approach allows highlighting the prominent nodes responsible for large zone overlaps. Results of their investigations on real-world networks clearly show the existence of large and dense overlaps of detected groups.

Besides these works on the topology of the community structure, there have been some other attempts to characterize the overlapping nodes mainly in the context of diffusion dynamics [25]–[28]. Indeed, overlapping nodes play a highly relevant role in the network due to their ability to reach multiple communities. We recently reported their importance on the epidemic spreading process [29], and how they can

be used in order to design effective immunization strategies. Indeed, when an epidemic outbreak occurs, it is not possible to immunize every individual through vaccination due to the limited amount of resources or time. Designing effective immunization strategies is therefore crucial to control the epidemic spreading and to reduce the cost of vaccine resources. Classically, deterministic targeted immunization strategies select the most influential spreaders to immunize according to a given centrality measure and an immunization budget. Nodes are ranked in the decreasing order of their centrality value and immunized in that order until the budget of the vaccine is over. We show that overlapping nodes deserve special treatment in deterministic immunization strategies. Indeed, these nodes have a higher level of both local influence and global influence as compared to nodes belonging to a single community.

Works in the same vein have been reported in [27], but in the random immunization strategy context. In order to mitigate the epidemic outbreaks in networks with overlapping structure, Kumar *et al.* proposed the so-called *OverlapNeighborhood* strategy. It immunizes randomly selected neighbors of the overlapping nodes. This local strategy is agnostic about the global structure of the network. It requires only to locate the overlapping nodes. It is, therefore, more appropriate for large scale networks than a deterministic strategy that requires to rank all the nodes according to a centrality measure. The main idea of the *OverlapNeighborhood* strategy is that there is a high probability that overlapping nodes are neighbors of high degrees nodes. Thus, once the overlapping nodes are identified, one can target the hubs in their neighborhood for immunization. Experiments performed with four empirical networks showed that it is almost as effective as the deterministic degree strategy where nodes are ranked in decreasing order of their degree. Therefore *OverlapNeighborhood* can select highly connected nodes for immunization. It performs sometimes as well as the betweenness strategy while using less information about the overall network structure. Remember that usually random immunization strategies are less effective than deterministic ones. These results corroborate the importance of the overlapping nodes. The fact that the random *OverlapNeighborhood* strategy is as effective as the deterministic degree centrality strategy suggests that the same set of nodes are targeted in both case. Consequently, there is a high probability that the hubs targeted by the deterministic strategy are immediate neighbors of the overlapping nodes.

To summarize, there is a great deal of work on complex networks topological properties. It is commonly admitted that they are organized around a small set of highly connected nodes, and that the overall structure is modular with nodes belonging to multiple communities. However, how these two types of nodes interact is still an open question. In this paper, we investigate the assumption that the overlapping nodes are the neighbors of the hubs. This idea is inspired by the effectiveness of the *OverlapNeighborhood* immunization strategy as compared to the degree deterministic strategy.

Note that, to our knowledge, no previous work that systematically explores the relationship between overlapping nodes and hubs has been reported.

III. DATA AND METHODS

In this section, we present briefly the data under investigation and the community detection algorithm used in the study. Indeed, as there is no information about the true community structure of these networks, different community detection algorithms are used to uncover it. This allows us to check the sensitivity of the results to the variations of the community structure linked to the algorithms.

A. DATASET DESCRIPTION

The selected networks come from a variety of domains (social, co-appearance, collaboration and e-commerce networks). Their size varies from hundreds to thousands of nodes and millions of edges to cover a wide range of situations. A short description of these networks follows, for more information refer to [34].

1) ZACHARY'S KARATE CLUB

It is a social network of 34 members of a karate club at a US university in 1970. The nodes represent the members of the club while the edges represent their friendship.

2) DOLPHINS

It is a social network of frequent associations between 62 dolphins in a community living in Doubtful Sound, New Zealand.

3) TERRORIST

It is a social network of known social association of the hijackers responsible for the September 11th, 2001 terrorist attacks.

4) ECOLOGY

This network represents the interactions among species and organisms within an ecosystem.

5) LES MISERABLES

It is a co-appearance network of characters in Victor Hugo's novel 'Les Miserables'. The nodes in this network represent the characters of this book and an edge between two nodes exists if they appear in the same chapter of the novel.

6) GAME OF THRONES

It is a co-appearances network of the characters of the Game of Thrones series. An edge between two nodes exists if they appear within 15 words of each other in the text.

7) ADJNOUN

This network contains the common adjective and noun adjacencies for the novel "David Copperfield" of Charles Dickens. An edge exists if two pairs of words occur in adjacent positions in the text of the book.

8) AMERICAN COLLEGE FOOTBALL

It is a network of American football games between Division IA colleges during regular season Fall 2000.

9) EUROROAD

This network represents the international road network located mostly in Europe. It is an undirected network where nodes stand for cities and an edge between two nodes denotes that they are connected by a road.

10) CRIME

This network contains persons who appeared in at least one crime case as either a suspect, a victim, a witness or both a suspect and victim at the same time.

11) YEAST PROTEIN INTERACTION

It is a network of the protein interactions contained in yeast. Nodes represent proteins while edges stand for their metabolic interaction.

12) EGO-FACEBOOK

The ego-Facebook network is collected from survey participants using the Facebook app.

13) ca-GRQC

It is a collaboration network which has been collected from the e-print arXiv. It covers scientific collaborations between authors of papers submitted to the General Relativity and Quantum Cosmology category.

14) GNUTELLA PEER-TO-PEER

It is a sequence of snapshots of the Gnutella peer-to-peer file sharing network from August 2002. The nodes represent the hosts in the Gnutella network while the edges are connections between the Gnutella hosts.

15) WIKIPEDIA VOTE

This network contains all the Wikipedia voting data from the inception of Wikipedia until January 2008. The nodes represent wikipedia users and an edge from node i to node j means that user i voted on user j .

16) ca-HepTh

It is a collaboration network extracted from arXiv. It covers scientific collaborations between authors of the High Energy Physics - Theory category. The network contains an edge from i to j if an author i co-authored a paper with author j .

17) ca-AstroPh

This network represents scientific collaborations between author's papers submitted to Astro Physics category. It contains an undirected edge from i to j if an author i co-authored a paper with author j .

18) FACEBOOK LARGE PAGE-PAGE

It is a network of verified Facebook sites collected in November 2017 and restricted to pages from four categories (television shows, politicians, companies and organizations).

The nodes stand for official Facebook pages while the edges represent mutual likes between sites.

19) ca-CondMat

It is a collaboration network extracted from arXiv. It covers scientific collaborations between authors of the Condense Matter category. The network contains an edge from i to j if an author i co-authored a paper with author j .

20) ENRON EMAIL

It is a communication network where nodes represent email addresses. A given node i is connected to j if the address i sent at least one email to address j .

21) AMAZON CO-PURCHASING NETWORK

This network is collected from Amazon web site. If a product i is frequently co-purchased with product j , the graph contains an undirected edge from i to j .

22) DBLP

The DBLP computer science bibliography contains the list of research papers in computer science. Nodes represent authors and edges connect nodes that have co-authored a paper.

The basic topological properties of these networks are given in Table 1. There is eight small networks with sizes ranging from 34 to 115 nodes. Zachary's karate club network is the one with the smallest size. Seven big empirical networks are also used with sizes ranging from 9877 to 334863 nodes. Amazon and DBLP are the networks with the biggest sizes. Six networks have a medium size. All the networks have a relatively high clustering coefficient value ranging from 0.258 to 0.633. They are disassortative except for American college football and the collaboration networks. The networks of medium and big sizes have a very small density, while it is relatively high for small networks. The basic properties of these empirical networks are very typical of what is generally observed in many real-world situations.

B. COMMUNITY DETECTION ALGORITHMS

Community structure is a common property of many real-world networks [29], [36]–[44]. Research on community detection is very active and a plethora of new algorithms based on various definitions are regularly published. In this paper, three influential overlapping community detection algorithms belonging to two different classes are used. Lancichinetti Fortunato Method and EAGLE (LFME) [45] is a local expansion and optimization algorithm. The Speaker-listener Label Propagation Algorithm (SLPA) together with the Democratic Estimate of the Modular Organization of a Network (DEMON) [46] belong to the label propagation category. More details are given in recent surveys about community detection methods [6], [7].

1) LANCICHINETTI FORTUNATO METHOD AND EAGLE (LFME)

This algorithm is a combination between the Lancichinetti Fortunato Method (LFM) [47] and the agglomerative hierarchical clustering based on maximal clique (EAGLE) [48]

TABLE 1. Common Topological properties of the real-world networks under study. N is the total number of nodes. E is the number of edges. $\langle k \rangle$ is the average degree. k_{max} is the max degree. C is the average clustering coefficient. α is the degree assortativity of the network. μ represents the density.

Network	N	E	$\langle k \rangle$	k_{max}	C	α	μ
Zachary's karate club	34	78	4.59	16	0.448	-0.475	0.139
Dolphins	62	159	5.12	12	0.258	-0.043	0.084
Terrorist	62	152	4.9	22	0.486	-0.08	0.08
Ecology	69	885	25.65	64	0.552	-0.296	0.377
Les Miserables	77	254	6.59	36	0.499	-0.165	0.086
Game of thrones	107	352	6.57	36	0.551	-0.13	0.062
Adjnoun	112	425	7.58	49	0.172	-0.129	0.068
American college football	115	613	10.66	12	0.403	0.162	0.093
Euroroad	1174	1417	2.41	10	0.033	0.126	0.002
Crime	1380	1476	2.13	25	0.0058	-0.161	0.0042
Yeast Protein Interaction	2361	7182	6.08	65	0.051	-0.084	0.002
ego-Facebook	4039	88234	43.69	1045	0.605	0.063	0.01
ca-GrQc	4158	13428	5.53	81	0.529	0.659	0.001
Gnutella peer-to-peer	6301	20777	6.59	97	0.01	0.035	0.001
Wikipedia vote	7115	100762	28.32	1065	0.14	-0.083	0.003
ca-HepTh	9877	25998	5.26	65	0.471	0.267	0.0005
ca-AstroPh	18772	198110	21.1	504	0.63	0.205	0.001
Facebook Large Page-Page	22012	170947	15.53	708	0.323	0.082	0.0007
ca-CondMat	23133	93497	8.083	279	0.633	0.135	0.0003
Enron Email	36692	183831	10.02	1383	0.497	-0.11	0.0002
DBLP	317080	1049866	6.62	343	0.632	0.266	$2.088 e^{-5}$
Amazon	334863	925872	5.52	529	0.396	-0.058	$1.651 e^{-5}$

algorithm. It is commonly known that communities are usually overlapping and hierarchical. Yet, the majority of algorithms investigate these two properties in a separate way. The Lancichinetti Fortunato Method and EAGLE (LFME) use the fitness function optimization of the LFM algorithm to identify with accuracy the overlapping communities. In addition, the hierarchical structure is defined using EAGLE algorithm. The LFME algorithm begins by picking a random seed node as the original member of a community. Then, nodes are added to this community until a fitness function is locally maximal. After uncovering one community, the same process is applied to another seed node until all the vertices of the network are assigned to at least one community. The second stage of this algorithm consists on selecting a specific similarity function. After that, the pair of communities with the maximum similarity are merged into one community. Then, the similarity function is computed between the new community and other communities. This process stops when we obtain just one community. Thus, the whole process can form a dendrogram. This means that the hierarchical structure is uncovered. Through this dendrogram, the overlapping modularity is computed. Finally, the dendrogram is cut when the overlapping modularity has a maximum value.

2) SPEAKER-LISTENER LABEL PROPAGATION ALGORITHM (SLPA)

SLPA [35] is an extension of the Label Propagation Algorithm (LPA). While in LPA, each node holds only a single label that is iteratively updated by adopting the majority label in the neighborhood, in SLPA each node possesses a memory containing multiple labels. Starting from a node selected as a listener, its neighbors send out a label following certain speaking rules. The listener selects one label according to a listening rule and adds it to its memory. Once all the nodes

have been visited, the communities are extracted from the node's memory converted into a probability distribution of labels that defines the membership degree to communities.

3) DEMOCRATIC ESTIMATE OF THE MODULAR ORGANIZATION OF A NETWORK (DEMON)

DEMON [46] tends to affect a node to the most frequent community by the application of a label propagation algorithm on its neighbors sub-graphs. In other words, for each node, their neighbors vote for its community membership. All the votes are then combined to construct the overlapping community structure.

IV. EVALUATION MEASURES

We suspect that the hubs are neighbors with the overlapping nodes. In order to investigate this hypothesis, we need to form and compare these two sets. To do so, first of all, the overlapping community structure of real-world networks is uncovered using an overlapping community detection algorithm. The overlapping nodes and the set of their neighbors are extracted. Note that n is the size of the neighbors of the overlapping nodes. Then, the set of hubs is formed using the n highest degree nodes of the network (Top n nodes). Strictly speaking, the size of the set of hubs is not necessarily of the same size that the set of the neighbors of the overlapping nodes. Nevertheless, this choice is conservative. It is motivated by the fact that some similarity measures need to be computed on sets of the same size. In fact, some nodes in the hub set may not have a high degree enough to be called hubs, however, rather than speaking of this set as "the set of the top degree nodes", we choose to call it the set of hubs for short. As our purpose is to investigate the similarities between these two sets, we compute classical measures such as the proportion of common nodes of the two sets, the Jaccard

Algorithm 1 Extraction of the Set of Hubs

Input : Graph $G(V, E)$
Output: Set of neighbors of the overlapping nodes X and set of hubs Y

- 1 Extract the set of overlapping nodes S_o using a given overlapping community detection algorithm
- 2 Initialize the size of X the set of neighbors of the overlapping nodes $n \leftarrow 0$
- 3 **for** each $v \in S_o$ **do**
- 4 Add all the neighbors of the node v to the set X
- 5 $n \leftarrow n + 1$
- 6 **end**
- 7 Sort the set of neighbors of the overlapping nodes X in decreasing order according to their degree
- 8 Sort all the nodes of the network in decreasing order according to their degree
- 9 Add the top n nodes of the network to the set of hubs Y
- 10 Return X, Y

Index, the Rank-biased overlap and the correlation measures between the two sets (Pearson and Spearman). We also compare their degree distribution. Secondly, we form the sub-network restricted to the overlapping nodes, the hubs and their links (the so-called Overlap-Hub network), in order to compare some of its topological properties (diameter, mean shortest path) with the original network. Indeed, we can expect that the distances are smaller in the Overlap-Hub network as compared to the original network if the proportion of hubs that are immediate neighbors of the overlapping node is high enough.

A. COMPARISON BETWEEN THE SET OF NEIGHBORS OF THE OVERLAPPING NODES AND THE SET OF HUBS

Various measures are computed on real-world data to compare the set of neighbors of the overlapping nodes and the set of hubs. They are defined as follows:

1) PROPORTION OF COMMON NODES IN THE SET OF NEIGHBORS OF THE OVERLAPPING NODES AND THE SET OF HUBS

This measure assesses how many hubs are directly connected to the overlapping nodes. The set of hubs is computed according to the algorithm 1. At first, all the neighbors of the overlapping nodes are added to the set X . Next, all the repeated elements are removed from the set X . The size of the set of neighbors of the overlapping nodes n is defined as ($n = |X|$). After that, the set of hubs is computed. All the nodes of the network are sorted in a decreasing order according to their degree. Finally, the set of hubs Y is defined as the top n nodes of the network.

The proportion of common nodes in the set of neighbors of the overlapping nodes and the set of hubs (proportion of hubs for short) is defined as the ratio of the size of the intersection of the two sets to the size of the neighborhood

of the overlapping nodes set. For a given network G , it is given by:

$$A_n = \frac{|X \cap Y|}{n} \tag{1}$$

where X and Y are the set of neighbors of the overlapping nodes and the set of hubs respectively. n is the size of the neighborhood of the overlapping nodes. If the hubs are randomly distributed, this proportion should be very small. We consider that if this value is greater than 50%, a high proportion of the hubs are neighbors of the overlapping nodes.

2) RANK-BIASED OVERLAP

The measure, presented in [30], quantifies how well two ranking sets are in agreement with each other. It computes the fraction of overlapping elements of the two rankings while incrementally increasing their depths. Furthermore, the overlap of each rank in this measure has a fixed weight. It is set by a parameter that gives more weight to the top elements of the set if it has a low value. Consequently, the top of the set gets a higher weight than the tail. Let's X and Y be two infinite rankings. The set of elements ranges from position 1 to position d in set X is denoted as $X_{:d}$. The proportion of the overlap of two sets X and Y at depth d can be defined as:

$$A_d = \frac{|X_{:d} \cap Y_{:d}|}{d} \tag{2}$$

where $|X_{:d} \cap Y_{:d}|$ is the size of the intersection between sets X and Y at depth d . The Rank-biased overlap between two sets X and Y is defined as follows:

$$r(X, Y) = \sum_d w_d * A_d \tag{3}$$

where w_d is the weight at position d , which is equal to: $w_d = (1 - p) * p^{d-1}$, so that $\sum_d w_d = 1$. The Rank-biased overlap belongs to the interval $[0, 1]$. The value 0 means that the sets X and Y are disjoint, while 1 means that they are identical. The parameter p determines the weights of the elements. The smaller p , the more top-weighted the metric. When $p = 0$, only the top-ranked item is considered, and the RBO score is either zero or one. When p approaches 1, the evaluation becomes deep in both sets.

3) JACCARD INDEX

The Jaccard index is a statistic used to measure the similarities between two finite sample sets. It is formally known as the ratio of the size of the intersection between the sets to the size of their union. It is defined as follows:

$$J(X, Y) = \frac{|X \cap Y|}{|X \cup Y|} = \frac{|X \cap Y|}{|X| + |Y| - |X \cap Y|} \tag{4}$$

where X and Y are two sets. The Jaccard index can take values ranging between 0 and 1. It is equal to 0 when there is no overlap and 1 when there is a complete overlap between the sets.

4) CORRELATION

The correlation is used to measure the association between the set of neighbors of the overlapping nodes and the set of hubs. Two different types of correlation coefficients are employed:

a: PEARSON CORRELATION

The Pearson’s correlation coefficient is a common measure of association between two vectors. Let’s consider X and Y the vectors of degrees of nodes belonging respectively to the ordered set of neighbors of the overlapping nodes and the ordered set of hubs. The Pearson correlation coefficient ρ between the two vectors X and Y can be obtained 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:

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

b: SPEARMAN CORRELATION

Spearman’s correlation coefficient is a rank-based version of the Pearson’s correlation coefficient. We compute the Pearson correlation between the ranked set of neighbors of the overlapping nodes and the ranked set of the hubs of the same size. The ranks are computed according to the degree of nodes. The Spearman correlation coefficient denoted ρ_s between the two vectors X and Y can be written as follows:

$$\begin{aligned} \rho_s(X, Y) &= \frac{\sum_{i=1}^n (\text{rank}(x_i) - \overline{\text{rank}(x)})(\text{rank}(y_i) - \overline{\text{rank}(y)})}{\sqrt{\sum_{i=1}^n (\text{rank}(x_i) - \overline{\text{rank}(x)})^2 \sum_{i=1}^n (\text{rank}(y_i) - \overline{\text{rank}(y)})^2}} \end{aligned} \quad (7)$$

where $\text{rank}(x_i)$ and $\text{rank}(y_i)$ represent the ranks of the elements x_i and y_i respectively. Both correlation coefficients range from -1 to $+1$. The values of the two vectors tend to increase or decrease simultaneously (a positive monotonic association) results in $\rho > 0$, and the values of one vector tend to increase when the values of the other decrease (negative monotonic association) results in $\rho < 0$. The absence of a monotonic association between the two vectors results in ρ equals to 0. In addition, the correlation is moderate if the coefficient ranges between 0.5 and 0.7. The correlation is significant if the coefficient ranges between 0.7 and 1.0.

5) DEGREE DISTRIBUTION

Real-world networks have some common topological features that distinguish them from random graphs. The most popular property is the heavy-tailed degree distribution [31]. The degree of a node represents the number of connections the node has to other nodes in the network (number of neighbors). Thus, the degree distribution $P(k)$ of a network is estimated as the proportion of nodes in the network with degree k . It is a relevant characteristic of networks which indicates the overall pattern of connections. For a large amount of real-world networks, it often follows a heavy-tailed distribution

such as a power-law [32]. The power-law distribution can be denoted as $P(k) = k^{-\gamma}$, where γ is a positive exponent. The exponent value ranges usually between 2 and 3 according to several experimental studies [32], [33]. The power-law exponent γ can be employed to characterize the networks. In order to estimate the degree distribution of a network, a well-known approach is to fit a power-law distribution and to find its power-law exponent.

B. COMPARISON BETWEEN THE OVERLAP-HUB NETWORK AND THE GLOBAL NETWORK

Our aim is to compare the main properties of the global network and the sub-network called the ‘‘Overlap-Hub network’’. It is formed with the overlapping nodes, the hubs and their interactions. To this end, we analyze different properties. We measure the average distance, the median distance as well as the diameter of the Overlap-Hub network. These values are compared to their counterpart in the global network. Before describing these properties, we provide a formal definition of the Overlap-Hub network.

1) OVERLAP-HUB NETWORK

Let’s consider a network $G(V, E)$, where $V = \{v_1, v_2, \dots, v_n\}$ and $E = \{(v_i, v_j) \mid v_i, v_j \in V\}$ denotes the set of nodes and edges respectively. The set of its overlapping nodes is denoted $V_o = \{v_1^o, v_2^o, \dots, v_m^o\} \subset V$ where m is the number of the overlapping nodes. The set of hubs is denoted $V_h = \{v_1^h, v_2^h, \dots, v_k^h\} \subset V$ where k represents the numbers of hubs. The Overlap-Hub network is formed from the union of the set of overlapping nodes and the set of hubs. It is obtained by removing all nodes that do not belong to one of the two sets. The Overlap-Hub network is denoted $S(V_q, E_q)$, where $V_q = V_o \cup V_h$ and $E_q = \{(v_i^q, v_j^q) \mid v_i, v_j \in V_q\}$ are respectively its set of nodes and edges. Figure 1 illustrates the representation of the Global and the Overlap-Hub network for the Karate club network.

2) DISTANCES

Three topological properties are introduced in this subsection. They are computed on both the original network and the Overlap-Hub network.

a: AVERAGE DISTANCE

A shortest path between two vertices in a network is a path with the minimum number of links. The average distance computes the average shortest path lengths in a network. Let’s consider that $d(v_i, v_j)$ denotes the distance or the length of the shortest path between nodes v_i and v_j . The average distance l of a given network G is defined as:

$$l = \frac{1}{n(n-1)} * \sum_{i \neq j} d(v_i, v_j) \quad (8)$$

where n is the number of nodes of network G .

b: MEDIAN DISTANCE

The median distance is the value separating the higher half from the lower half of the length of shortest paths in

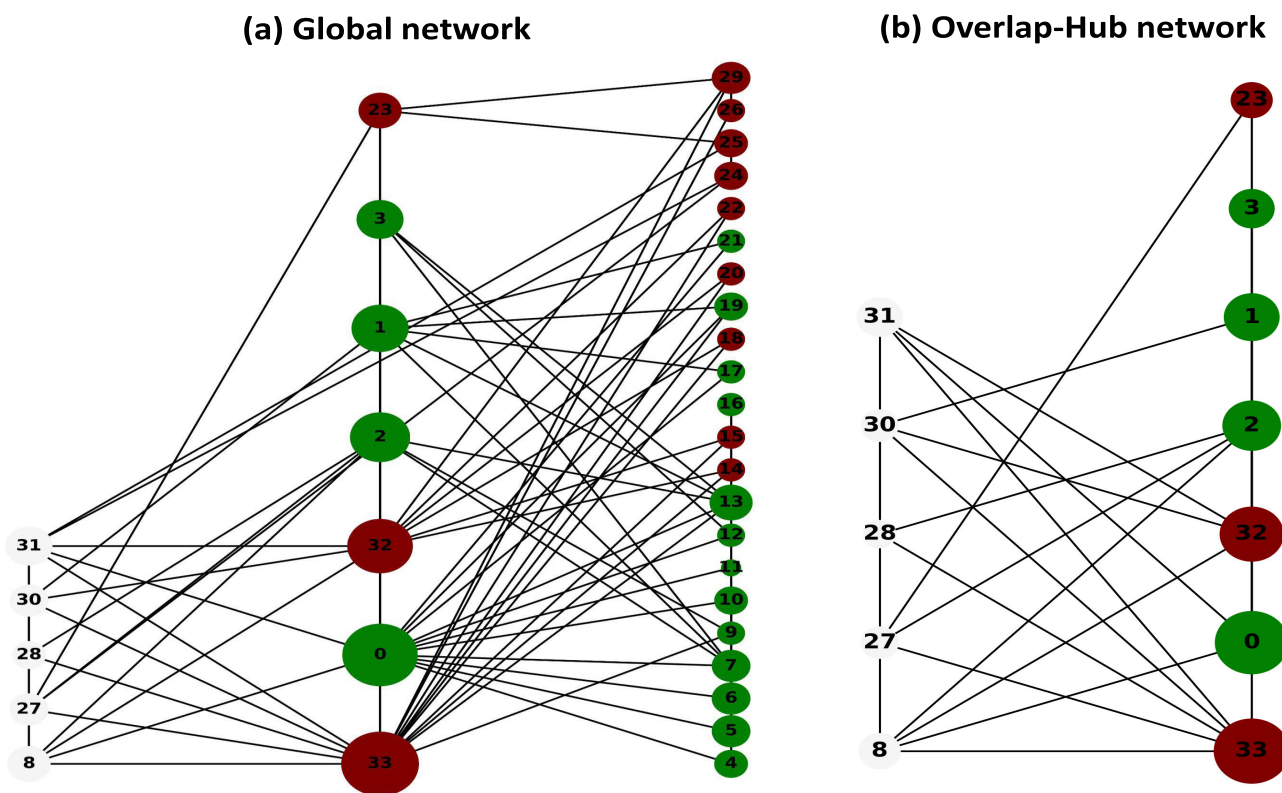


FIGURE 1. Representation of the Global and Overlap-Hub networks formed from Karate club data. Nodes are highlighted in different colors according to the community they belong to. Nodes with the same color belong to the same community. Those in gray represent the overlapping nodes. We note that the left column of the global network represents the overlapping nodes, the nodes in the middle column are the hubs while those located in the right column represent the remaining nodes. The size of the nodes is proportional to their degree. Overlapping nodes of the Overlap-Hub networks are at the left while hubs are at the right. The community structure is revealed using the SLPA detection algorithm.

the network. The basic advantage of computing the median distance as compared to the average distance is that it may give a better idea of a more meaningful shortest path value. Indeed, the median distance is not skewed by a small proportion of extremely small or large values. To compute this value, the shortest paths between all the pairs of nodes of the network are sorted in an increasing order. Then, the middle value is picked as the median distance if there is an odd number of shortest paths values. Otherwise, the median distance is defined as the average of the two middle values.

c: DIAMETER

The diameter of a given network represents the largest distance between any two nodes in the network. A small diameter denotes that the nodes of the network are tightly connected. The diameter δ of a network is defined as follows:

$$\delta = \max_{i,j} d(v_i, v_j) \tag{9}$$

V. RESULTS AND DISCUSSION

In this section, we report and discuss our empirical findings.

Various evaluation criteria are used to analyze the relationship between the overlapping nodes and the hubs. At first, we compare the set of neighbors of the overlapping nodes with the set of hubs. To this end, the proportion of common nodes in the two sets and the Rank-biased overlap measures are computed. The correlation is also used to measure the statistical association between the two sets. Moreover, the degree distribution of both sets is estimated and compared. In the second part of our experiment, keeping only the overlapping nodes, the hubs and their connections from the original networks, the so-called Overlap-Hub networks are formed. Topological properties of the Overlap-Hub networks are then compared with their original counterpart. To do so, we analyze some popular topological network properties such as the average and the median distance for both networks. Furthermore, the influence of both the neighborhood size and the membership degree of the overlapping nodes are investigated. In the following, results are reported first based on the SLPA uncovered community structure. Then, in order to evaluate the impact of the community detection algorithms, comparisons are performed with the community structures uncovered by LFME and DEMON.

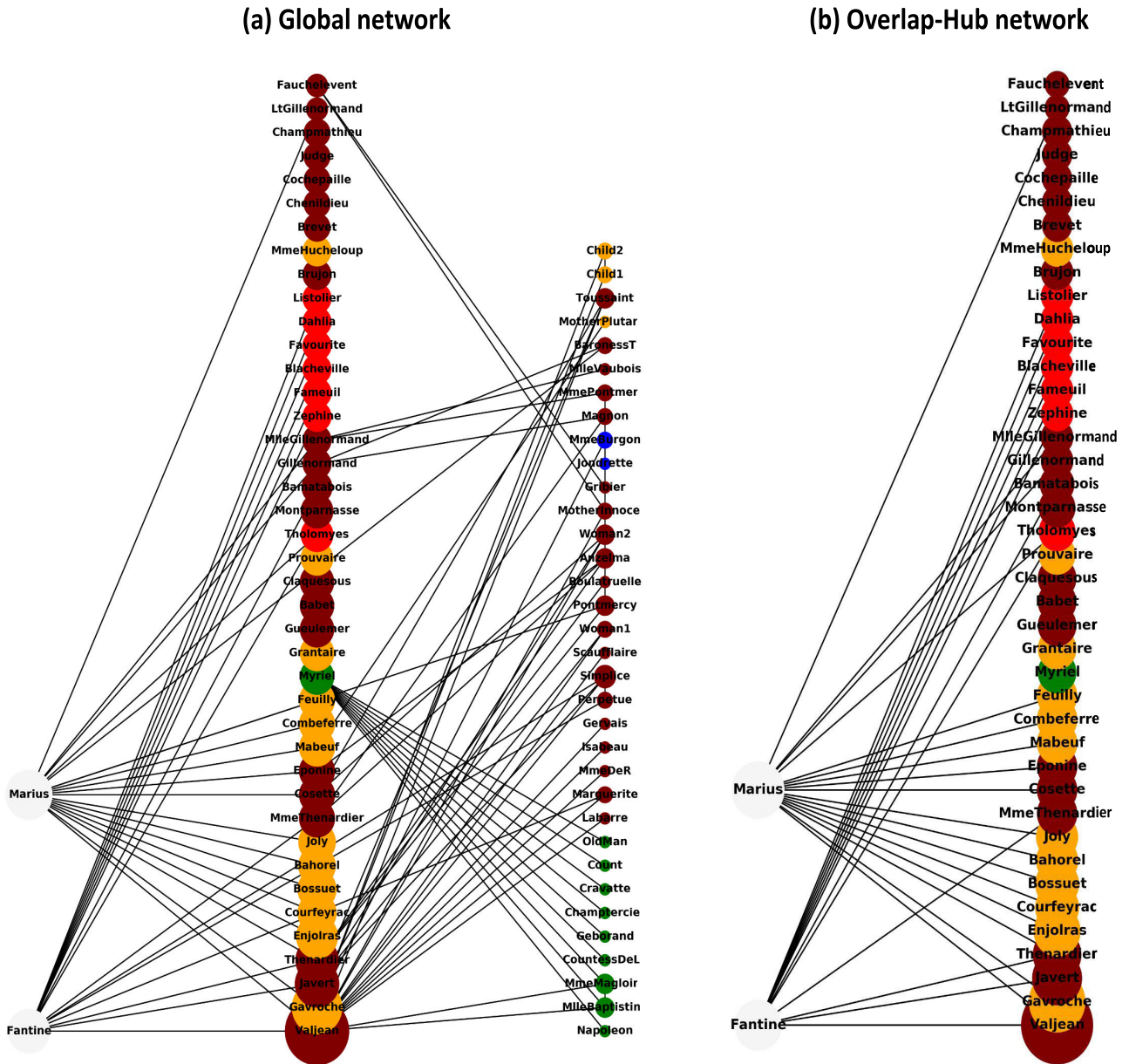


FIGURE 2. Representation of the Global and Overlap-Hub networks formed from Les Misérables network. Nodes are highlighted in different colors according to the community they belong to. Nodes with the same color belong to the same community while those in gray represent the overlapping nodes. We note that the left column of the global network are the overlapping nodes, the nodes in the middle column are the hubs while those located in the right column represent the rest of the nodes. The left column of the Overlap-Hub network are the overlapping nodes, the nodes in the middle column are the hubs. The size of the nodes is proportional to their degree. The community structure is revealed using the SLPA detection algorithm.

A. COMPARISON BETWEEN THE SET OF NEIGHBORS OF THE OVERLAPPING NODES WITH THE SET OF HUBS

1) PROPORTION OF COMMON NODES IN THE SET OF NEIGHBORS OF THE OVERLAPPING NODES AND THE SET OF HUBS

In this experiment, we compute the proportion of identical nodes in the set of hubs and set of the neighbors of the overlapping nodes. A high value of this proportion needs to be uncovered in order to validate the assumption that

the two sets are similar. Let’s first consider the small size networks. Indeed, in this case, we can compare the two sets visually. The figs. 1 to 8 illustrate the tripartite representation of Zachary’s Karate Club, Les Misérables, Game of thrones, Dolphin, Terrorist, Ecology, Adjnoun and American college football networks respectively. The grey nodes of the left column are the overlapping nodes, the nodes in the middle column are the hubs and those located in the right column represent the rest of the nodes. Note that the size of nodes is

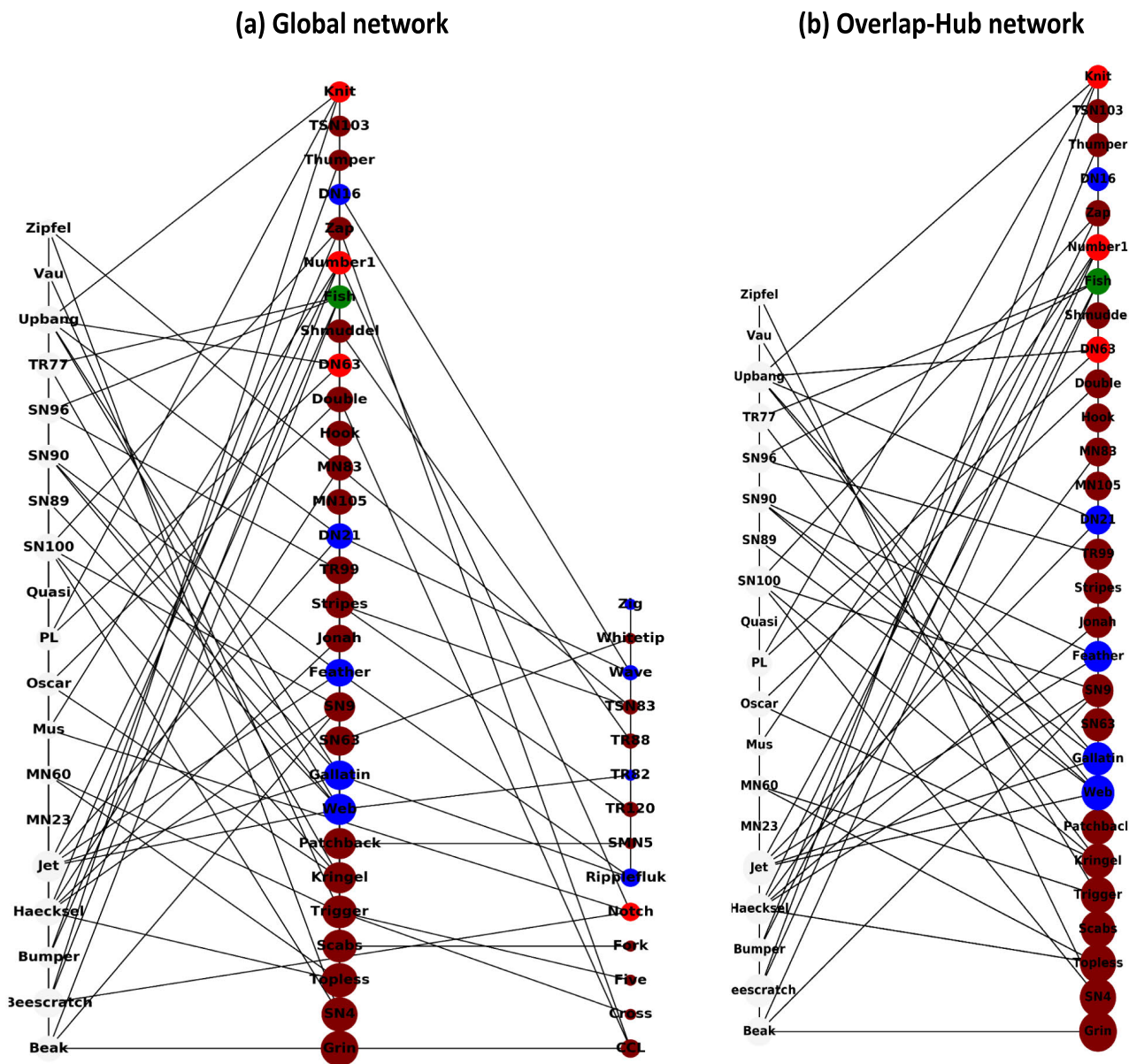


FIGURE 3. Representation of the Global and Overlap-Hub networks formed from Dolphins network. Nodes are highlighted in different colors according to the community they belong to. Nodes with the same color belong to the same community while those in gray represent the overlapping nodes. We note that the left column of the global network are the overlapping nodes, the nodes in the middle column are the hubs while those located in the right column represent the rest of the nodes. The left column of the Overlap-Hub network are the overlapping nodes, the nodes in the middle column are the hubs. The size of the nodes is proportional to their degree. The community structure is revealed using the SLPA detection algorithm.

proportional to their degree. These networks have a number of overlapping communities ranging from 2 to 5. One can notice that all the overlapping nodes are connected to the hubs and that these hubs belong to different communities. As shown in all the figures, the overlapping nodes are densely connected with the hubs while being sparsely connected with the rest of the nodes. Another way to compare both sets is to rank the nodes in both sets in degree decreasing value and to compute the maximum size of the sets where there is always 100% of the nodes that are common in both sets.

This gives us a conservative value as compared to the proportion of common nodes. However, it allows to see if we can reasonably accept the assumption that the hubs are the neighbors of the overlapping nodes. Indeed, we know that the proportion of hubs in a network with a heavy-tailed degree distribution is quite small. If the top 10% of the nodes are common in both sets there is a high probability that they include all the hubs of the network. Table 2 reports the set of overlapping nodes, the set of the neighbors of the overlapping nodes as well as the set of hubs for these small networks.

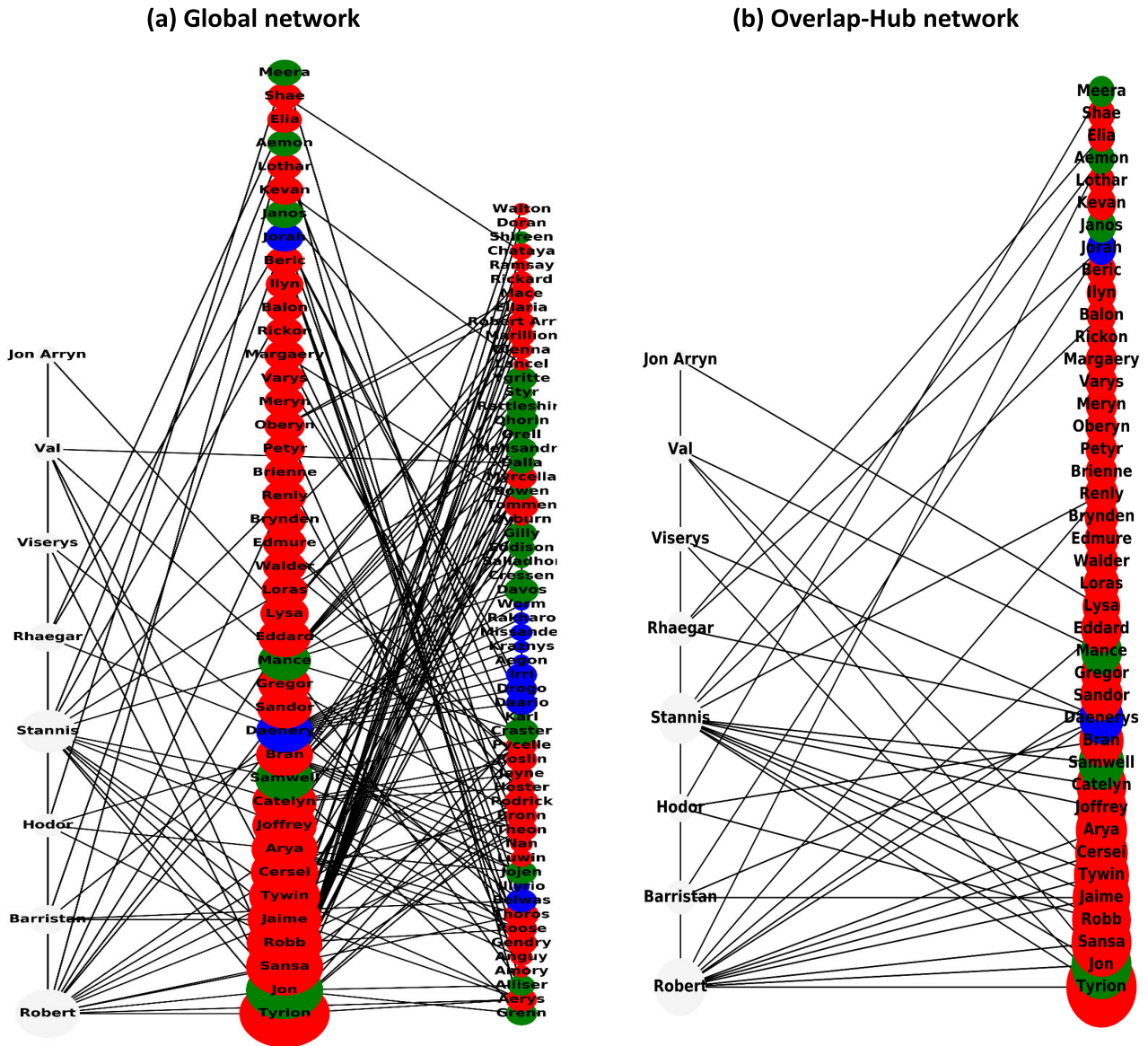


FIGURE 4. Representation of the Global and Overlap-Hub networks formed from Game of thrones network. Nodes are highlighted in different colors according to the community they belong to. Nodes with the same color belong to the same community while those in gray represent the overlapping nodes. We note that the left column of the global network are the overlapping nodes, the nodes in the middle column are the hubs while those located in the right column represent the rest of the nodes. The left column of the Overlap-Hub network are the overlapping nodes, the nodes in the middle column are the hubs. The size of the nodes is proportional to their degree. The community structure is revealed using the SLPA detection algorithm.

The nodes are ranked in decreasing order of their degree values. Labels that are common to the set of neighbors of the overlapping nodes and the set of hubs are written in blue. One can notice that there is a high overlap between these two sets. Furthermore, nodes that are different in the two sets are the ones with the lowest degree values. In the Karate Club network, the top 5 high degree nodes are in the overlapping nodes neighborhood. As the network is made of 34 nodes, this value represents 14.7% of the nodes with the highest degree. In Les Miserables, it is the 23 top high degree nodes out of 77 nodes of the network that belong to the overlapping

nodes neighborhood. It is, therefore, 29.8% of nodes with the highest degree which are in the vicinity of the overlapping nodes. The top 16 high degree nodes out of the 107 nodes of the Game of Throne network are neighbors of the overlapping nodes. This represents 14.9% of the highest degree nodes. In the Dolphin network, the node Jet (node ranked with the fifth-highest degree) does not belong to the neighborhood of the overlapping nodes. So, we can say that the top 4 high degree nodes are in the vicinity of the overlapping nodes. This corresponds to 6.8% of the nodes with the highest degree in the network. For the Ecology network, the top 6 high degree

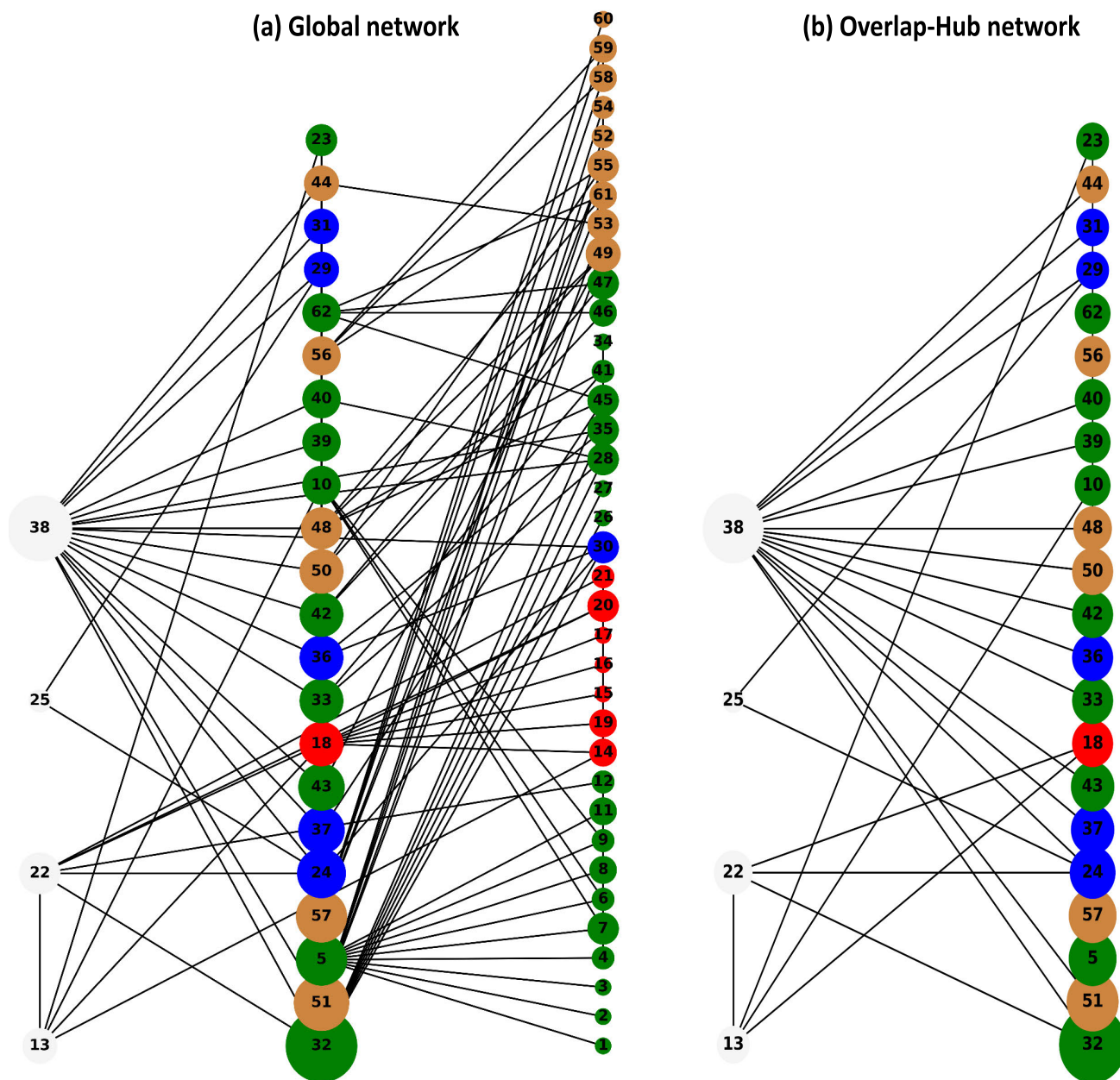


FIGURE 5. Representation of the Global and Overlap-Hub networks formed from Terrorist network. Nodes are highlighted in different colors according to the community they belong to. Nodes with the same color belong to the same community while those in gray represent the overlapping nodes. We note that the left column of the global network are the overlapping nodes, the nodes in the middle column are the hubs while those located in the right column represent the rest of the nodes. The left column of the Overlap-Hub network are the overlapping nodes, the nodes in the middle column are the hubs. The size of the nodes is proportional to their degree. The community structure is revealed using the SLPA detection algorithm.

nodes out of 69 nodes of the network are neighbors of the overlapping nodes. This represents 9% of the highest degree nodes. In addition, in Adjnoun, the top 15 high degree nodes out of 112 belong all to the neighborhood of the overlapping nodes except for the node ‘little’ which is one of the overlapping nodes. So, this represents 12.5% of the highest degree nodes. In Terrorist network, not all the highest-ranked nodes belong to the neighborhood of the overlapping nodes. If we consider the top six nodes with the highest degrees,

we can notice that only three nodes (‘32’, ‘51’ and ‘24’) are neighbors with the overlapping nodes. However, the node ‘38’ which is the second most connected node is one of the overlapping nodes. In the American college football, only the top 6 high degree nodes out of 115 nodes are neighbors of the overlapping nodes (5.21%). This small percentage is due to the fact that this network does not exhibit a heavy-tailed degree distribution. Thus, the majority of nodes in the American college football have the same degree. In some

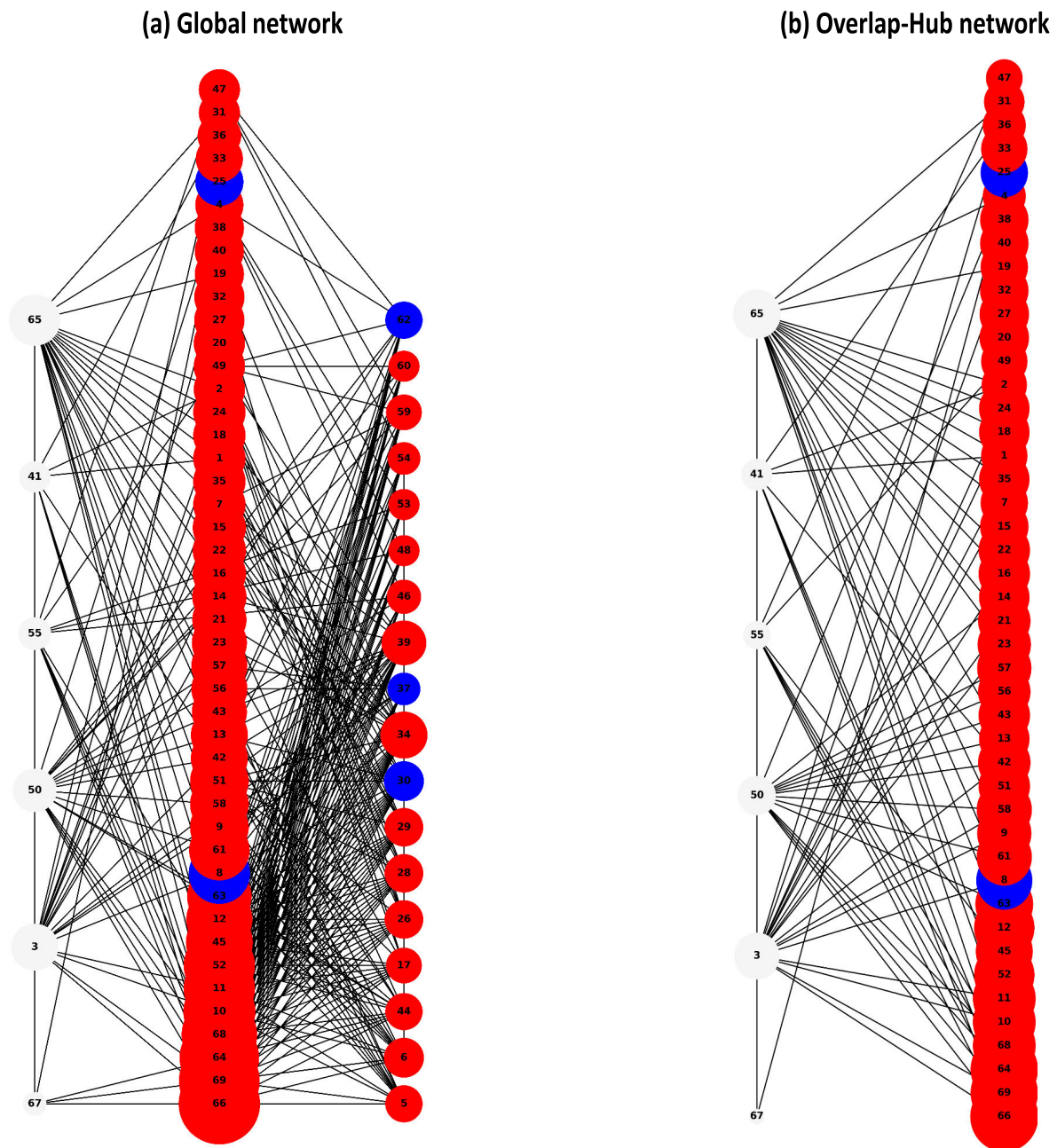


FIGURE 6. Representation of the Global and Overlap-Hub networks formed from Ecology network. Nodes are highlighted in different colors according to the community they belong to. Nodes with the same color belong to the same community while those in gray represent the overlapping nodes. We note that the left column of the global network are the overlapping nodes, the nodes in the middle column are the hubs while those located in the right column represent the rest of the nodes. The left column of the Overlap-Hub network are the overlapping nodes, the nodes in the middle column are the hubs. The size of the nodes is proportional to their degree. The community structure is revealed using the SLPA detection algorithm.

networks, it is also noticed that the set of hubs and the set of overlapping nodes have some nodes in common. Labels that are common to these two sets are written in red. In Terrorist network, two nodes belonging to the set of overlapping nodes ('38' and '22') are also considered as hubs. This corresponds to 8% of the hubs. Furthermore, the node '38' is the second-ranked node of the network. The same for Ecology network,

two overlapping nodes belong to the set of hubs. They represent 4.44% of the hubs. In Adjnoun network, six overlapping nodes are also considered as hubs. They represent 11.11% of the hubs. Additionally, one of these nodes (node 'little') is the most connected node of the network. That being said, there is a very high proportion of hubs that belongs to the neighborhood of the overlapping nodes in small networks.

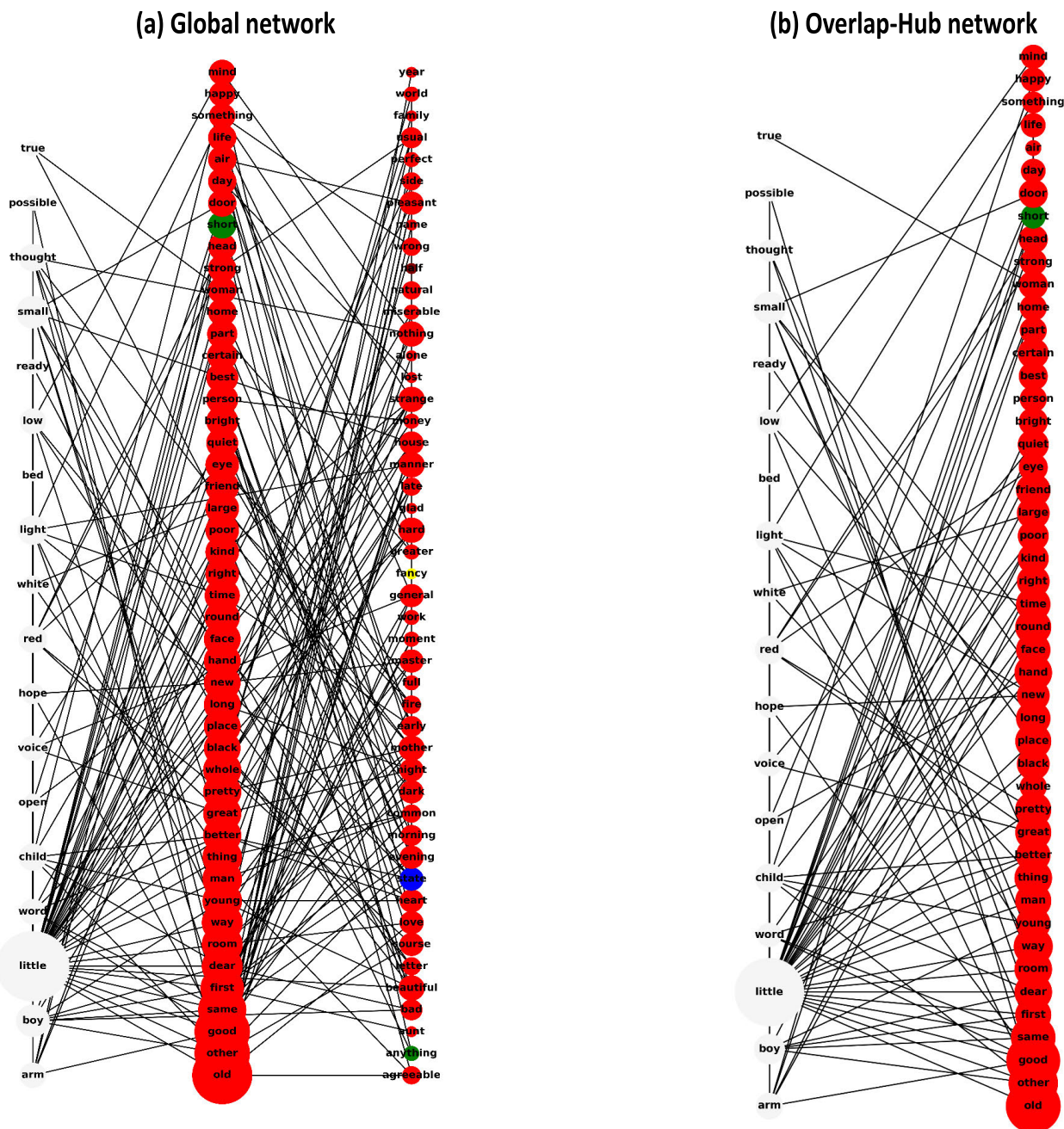


FIGURE 7. Representation of the Global and Overlap-Hub networks formed from Adjnoun network. Nodes are highlighted in different colors according to the community they belong to. Nodes with the same color belong to the same community while those in gray represent the overlapping nodes. We note that the left column of the global network are the overlapping nodes, the nodes in the middle column are the hubs while those located in the right column represent the rest of the nodes. The left column of the Overlap-Hub network are the overlapping nodes, the nodes in the middle column are the hubs. The size of the nodes is proportional to their degree. The community structure is revealed using the SLPA detection algorithm.

Another quite smaller proportion belongs to the set of overlapping nodes. The hubs excluded from this area are the ones with low degrees. All these results are a strong indication that hubs are neighbors of the overlapping nodes.

Results for all the networks under test are reported on Table 3. We can consider that the fraction of common nodes in both sets is high if its value is higher than 50%. If we look

at all the networks, the proportion of hubs belonging to the overlapping nodes neighborhood is ranging from 59.25% to 87.27%. It is always above 50%. However, the highest value is for small size networks. As the network size grows, there are more and more peripheral hubs that are not in the vicinity of the overlapping nodes. Indeed, overlapping nodes are in the core of the communities. Moreover, as for small networks,

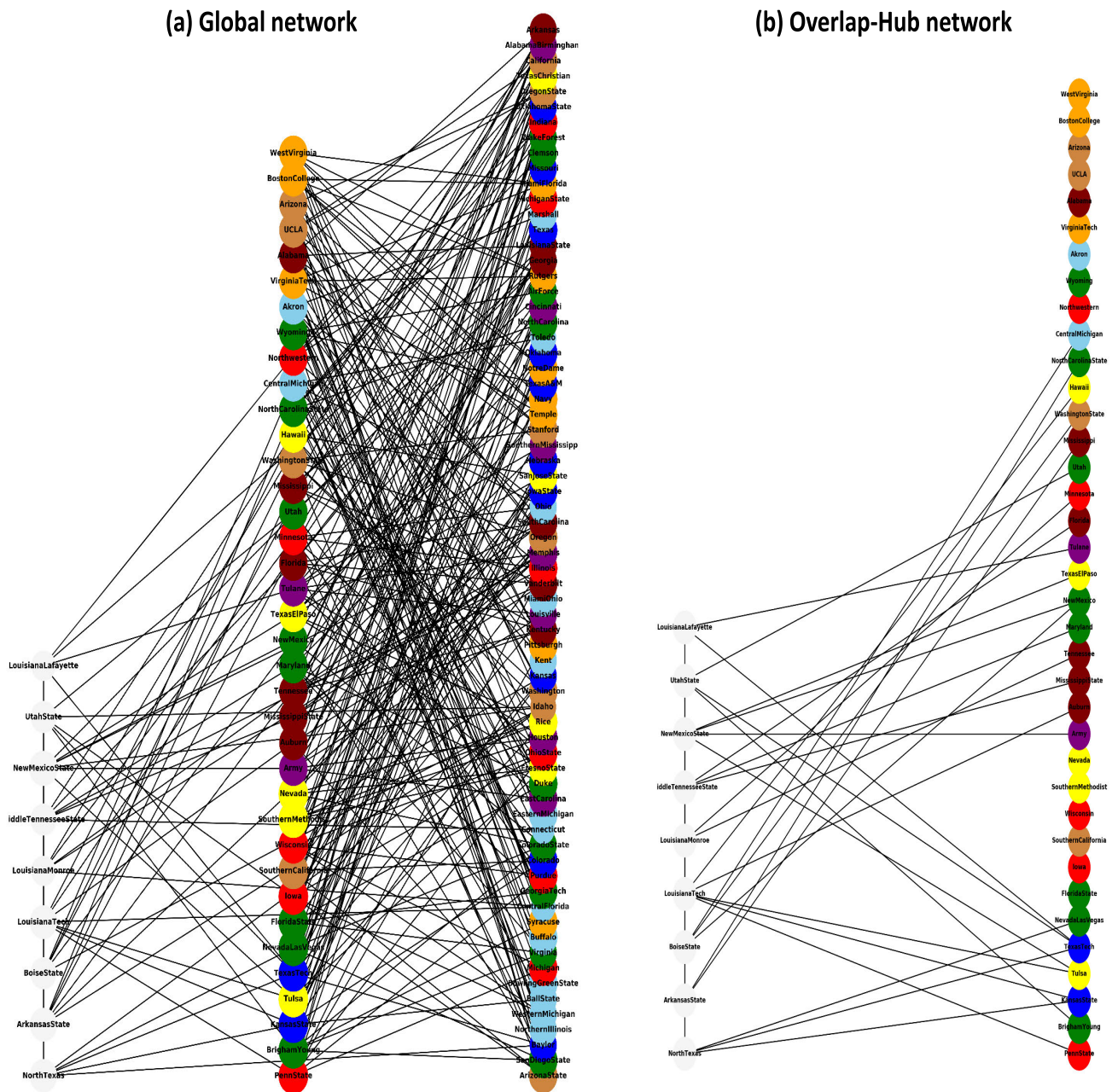


FIGURE 8. Representation of the Global and Overlap-Hub networks formed from American college football network. Nodes are highlighted in different colors according to the community they belong to. Nodes with the same color belong to the same community while those in gray represent the overlapping nodes. We note that the left column of the global network are the overlapping nodes, the nodes in the middle column are the hubs while those located in the right column represent the rest of the nodes. The left column of the Overlap-Hub network are the overlapping nodes, the nodes in the middle column are the hubs. The size of the nodes is proportional to their degree. The community structure is revealed using the SLPA detection algorithm.

we also check if the top-ranked nodes of the set of hubs are neighbors to the overlapping nodes. To do so, we compute the proportion of the most connected nodes located in the neighborhood of the overlapping nodes. We note that the top-ranked nodes represent 10% of nodes with the highest degrees of the network. Experimental results show that the proportion of the top-ranked hubs is always very high for most of the networks. That means that the nodes belonging

to the set of hubs with the highest degrees are neighbors of the overlapping nodes. Additionally, it can be noticed that the set of hubs and the set of the overlapping nodes have some nodes in common. This is consistent with the results of Yang *et al.* study [23] showing that the overlapping nodes are densely connected. However, the proportion of common nodes is relatively small as compared to the proportion of common nodes in the set of the neighborhood of overlapping

TABLE 2. The set of neighbors of overlapping nodes and the set of hubs for small networks. Nodes are sorted in decreasing order according to their degree.

Network	Overlapping nodes	Neighbors of overlapping nodes	Hubs
Karate club	8, 27, 28, 30, 31	33, 0, 32, 2, 1, 23, 25, 24	33, 0, 32, 2, 1, 3, 13, 23
Les Miserables	Fantine, Marius	Valjean, Gavroche, Javert, Thenardier, Enjolras, Bossuet, Courfeyrac, Joly, Bahorel, Mabeuf, Cosette, MmeThenardier, Eponine, Combeferre, Feuilly, Tholomyes, Bamatabois, Zephine, MlleGillenormand, Favourite, Blacheville, Gillenormand, Fameuil, Dahlia, Listolier, Simplice, LtGillenormand, Pontmercy, Marguerite, BaronessT	Valjean, Gavroche, Javert, Thenardier, Enjolras, Bossuet, Courfeyrac, Joly, Bahorel, Mabeuf, Cosette, MmeThenardier, Eponine, Combeferre, Feuilly, Tholomyes, Bamatabois, Zephine, MlleGillenormand, Favourite, Blacheville, Gillenormand, Fameuil, Myriel, Grantaire, Gueulemer, Babet, Claquesous, Prouvaire, Montparnasse
Dolphins	Beak, CCL, Double, Five, Haecksel, Jonah, Kringel, MN105, MN60, MN83, Oscar, Patchback, PL, SMN5, SN100, SN89, SN9, Thumper, Topless, Trigger, Vau, Zap, Zipfel	Grin, SN4, Scabs, Web, Beescratch, SN63, Stripes, TR99, TR77, Hook, SN96, Shmuddel, Fish, DN63, Knit, TSN103, Bumper, TSN83, Cross	Grin, SN4, Scabs, Web, Jet, Beescratch, SN63, Gallatin, Stripes, TR99, Feather, Upbang, TR77, Hook, SN96, DN21, Shmuddel, Fish, DN63
Terrorist	38, 22, 13, 25	32, 51, 24, 37, 43, 18, 33, 36, 42, 50, 48, 10, 39, 40, 29, 31, 44, 23, 30, 20, 28, 35, 14, 12, 21	32, 38, 51, 5, 57, 24, 37, 43, 18, 33, 36, 42, 50, 22, 48, 10, 39, 40, 56, 62, 13, 29, 31, 44, 23
Ecology	65, 3, 67, 50, 55, 41	66, 69, 64, 68, 10, 11, 45, 12, 63, 8, 61, 9, 58, 42, 13, 43, 56, 57, 23, 21, 14, 16, 22, 7, 35, 1, 18, 24, 2, 49, 19, 4, 36, 31, 47, 6, 62, 5, 44, 59, 17, 46, 37, 53, 48, 60	66, 69, 64, 68, 10, 11, 52, 45, 12, 63, 8, 61, 9, 58, 51, 42, 13, 43, 56, 57, 23, 21, 14, 16, 22, 15, 7, 35, 1, 18, 24, 2, 49, 65, 20, 27, 32, 19, 40, 38, 4, 25, 3, 36, 31, 47
Game of thrones	Robert, Barristan, Hodor, Stannis, Rhaegar, Viserys, Val, Jon Arryn	Tyrion, Jon, Sansa, Robb, Jaime, Tywin, Cersei, Arya, Joffrey, Catelyn, Samwell, Bran, Daenerys, Sandor, Mance, Eddard, Lysa, Renly, Jorah, Balon, Meera, Elia, Aemon, Thoros, Belwas, Jojen, Melisandre, Aerys, Dalla	Tyrion, Jon, Sansa, Robb, Jaime, Tywin, Cersei, Arya, Joffrey, Catelyn, Samwell, Bran, Daenerys, Sandor, Mance, Eddard, Gregor, Lysa, Loras, Renly, Walder, Edmure, Brynden, Brienne, Petyr, Oberyn, Meryn, Varys, Margaery
Adjnoun	little, small, boy, light, red, thought, arm, word, child, open, voice, hope, white, bed, low, ready, possible, true	old, other, good, same, first, dear, room, way, young, man, thing, better, great, pretty, black, place, long, new, hand, face, round, time, right, kind, poor, large, friend, eye, quiet, bright, home, woman, strong, head, short, door, something, happy, mind, manner, beautiful, nothing, strange, mother, hard, pleasant, love, general, house, state, bad, anything, miserable, money	little, old, other, good, same, first, dear, room, way, young, man, thing, better, great, pretty, whole, black, place, long, new, hand, face, round, time, right, kind, small, poor, boy, large, friend, eye, quiet, bright, person, best, light, certain, part, home, child, red, woman, strong, head, short, thought, door, day, air, life, something, happy, mind
American college football	NorthTexas, ArkansasState, BoiseState, LouisianaTech, LouisianaMonroe, MiddleTennesseeState, NewMexicoState, UtahState, LouisianaLafayette	PennState, BrighamYoung, KansasState, Tulsa, TexasTech, NevadaLasVegas, Army, Auburn, MississippiState, Tennessee, Maryland, NewMexico, TexasElPaso, Tulane, Florida, Minnesota, Utah, CentralMichigan, SouthCarolina, ArizonaState, Oklahoma, Texas, TexasChristian, Illinois, Memphis, Mississippi, WashingtonState, Hawaii, NorthCarolinaState, AlabamaBirmingham, Baylor, MiamiFlorida, Arkansas, Georgia, Idaho, CentralFlorida, Connecticut	PennState, BrighamYoung, KansasState, Tulsa, TexasTech, NevadaLasVegas, FloridaState, Iowa, SouthernCalifornia, Wisconsin, SouthernMethodist, Nevada, Army, Auburn, MississippiState, Tennessee, Maryland, NewMexico, TexasElPaso, Tulane, Florida, Minnesota, Utah, Mississippi, WashingtonState, Hawaii, NorthCarolinaState, CentralMichigan, Northwestern, Wyoming, Akron, VirginiaTech, Alabama, UCLA, Arizona, BostonCollege, WestVirginia

nodes and the set of hubs. Therefore, the majority of hubs are neighbors of the overlapping nodes, and a small proportion of them are also overlapping nodes.

2) JACCARD INDEX

We compute the Jaccard index between the set of neighbors of the overlapping nodes and the set of hubs. This measure is used to emphasize the similarity between both sets.

Results are reported in Table 3. One can notice that the Jaccard index displays high values for the vast majority of the networks. Only 5 networks out of the 22 understudy exhibit a Jaccard index value below 50%. Results are well correlated with the previous observations. Values are smaller as compared to the proportion of common nodes in both sets. This is due to the fact that the size of the neighborhood of the overlapping nodes is sometimes very large. Its size is

TABLE 3. Estimated parameters of real-world networks under study for the SLPA community detection algorithm. N is the size of the network. μ is the density. on is the fraction of overlapping nodes. S is the size of the neighborhood of the overlapping nodes. A_n is the proportion of hubs in the neighborhood of the overlapping nodes while σ_A is its standard deviation. J represents the Jaccard index between the set of neighbors of the overlapping nodes and the set of hubs while σ_J is its standard deviation. A_t (%) is the proportion of the top connected nodes in the neighborhood of the overlapping nodes, where t is equal to 10% of the most highly ranked nodes of the network ($t = 0.1 * n$). A_n^o is the proportion of hubs belonging to the set of overlapping nodes. We note that each proportion value is the average of 10 SLPA simulation runs.

Network	N	μ	on (%)	S (%)	A_n (%)	σ_A (%)	J (%)	σ_J (%)	A_t (%)	A_n^o (%)
Zachary's karate club	34	0.139	15.1	23.52	77.6	3.82	50	3.38	100	1.9
Dolphins	62	0.084	24	35.48	77.89	6.73	64	2.28	83.33	3.92
Terrorist	62	0.08	6.45	43.54	77.77	4.11	63.63	5.18	50	8
Ecologie	69	0.377	8.69	66.46	74	5.32	58.73	2.91	100	4.44
Les Miserables	77	0.086	2.5	44.15	81.06	3.34	68.62	2.71	100	1.85
Game of thrones	107	0.062	7.4	33.64	67.73	3.79	52.54	1.83	100	6.62
Adjnoun	112	0.068	16.07	49.1	73.21	3.21	71.08	5.15	91	10.96
American college football	115	0.093	7.82	33.17	65.21	2.8	57	3.51	54.54	8
Euroroad	1174	0.002	26	38.5	68.96	3.6	45.6	4.9	84.61	11.88
Crime	1380	0.0042	26.17	52.16	76.83	2.5	57.54	3.7	90.24	7.74
Yeast Protein Interaction	2361	0.002	31.85	55.82	80.27	6.17	61.42	5.82	95.76	11.44
ego-Facebook	4039	0.01	3.1	17.5	59.25	1.95	44.41	3.87	50.38	1.43
ca-GrQc	5242	0.001	26.43	33.19	64.04	2.26	49	2.9	65.12	8.5
Gnutella peer-to-peer	6301	0.001	14.33	41.27	65.66	7.05	47.28	2.15	70.47	3.98
Wikipedia vote	7115	0.003	13.66	38.14	71.99	5.23	55.26	2.48	92.68	8.33
ca-HepTh	9877	0.0005	31.36	45.2	75.32	0.41	57.53	3.36	90.88	13.13
ca-AstroPh	18772	0.001	11.18	51.59	72.98	2.96	56.38	3.84	93.44	3.94
Facebook Large Page-Page	22012	0.0007	14.26	49.1	63.56	3.78	46.2	2.66	74.92	3.07
ca-CondMat	23133	0.0003	29.37	51.69	79.57	1.42	64.87	3.79	97.92	2.07
Enron Email	36692	0.0002	35	48.88	73.54	1.96	58.36	2.25	98.88	5.92
DBLP	317080	$2.088 e^{-5}$	15.87	30.08	74.64	1.77	60.38	3.14	93.71	5.07
Amazon	334863	$1.651 e^{-5}$	24.25	41.13	71.37	1.14	57.8	2.36	90.93	7.98

even greater than 65% in some small networks. Consequently, some nodes belonging to the set of hubs have small degrees, since it has the same size as the set of neighbors of the overlapping nodes. So, there is a high chance that these nodes with a low degree do not take part in the neighborhood of the overlapping nodes. This explains the dissimilarities between the set of neighbors of the overlapping nodes and the set of hubs.

3) RANK-BIASED OVERLAP

The Rank-biased overlap (*RBO*) is also computed in order to compare the set of neighbors of the overlapping nodes and the set of hubs. This measure quantifies the fraction of the overlapping elements of two sets while incrementally increasing their depths. Two sets are similar if their *RBO* value is equal to 1, whereas they are disjoint if the value is equal to 0. The *RBO* is computed between the set of neighbors of the overlapping nodes and the set of hubs while assigning different weights to the elements for both sets. The parameter p determines the weights of the elements (see Subsection II.A). More importance is given to the comparison of the top elements of the two sets when the p value is small. Results of the *RBO* computation for various p values are reported in Table 4. One can notice that the *RBO* values are quite high when almost all the elements have the same weight ($p = 0.98$). Indeed, the overlap between the two sets ranges between 83.4% and 99.9%. This value increases when more weight is given to the top-ranked elements of the sets. The overlap between both sets ranges from 98.9% to 100% in this case ($p = 0.5$). These results confirm the previous observations reported in Table 2. Indeed, almost all the top-ranked nodes

are colored in blue. Thus, these nodes belong to the set of neighbors of the overlapping nodes. Therefore, almost all the most highly connected nodes of the network are located in the neighborhood of the overlapping nodes. To conclude, the set of neighbors of the overlapping nodes is generally very similar to the set of hubs. This validates the assumption that hubs are neighbors of the overlapping nodes. Furthermore, the top-ranked nodes according to their degree are always in the vicinity of the overlapping nodes.

4) CORRELATION

In this section, we report results on the correlation between the set of hubs and the set of neighbors of the overlapping nodes. First, the Pearson correlation coefficient, given in Equation 5, is used to measure the degree of relationship between both sets. It is computed using the degrees of these sets sorted in decreasing order. Secondly, we use also the Spearman correlation to measure the strength of the monotonic relationship between the ranks of the set of neighbors of the overlapping nodes and the set of hubs. The ranks of the sets are computed according to the degree of nodes. We note that the value of both correlation measures varies between -1 (perfect negative correlation) and 1 (perfect positive correlation). To illustrate the process, Table 5 represents the correlation between the set of neighbors of the overlapping nodes and the set of hubs for Karate club network. At first, we define both sets based on algorithm 1. Pearson correlation is then computed between the degrees of nodes belonging to these two sets. The nodes are sorted in decreasing order of their degree and each degree is associated with a rank. We note that the highest degree has a rank equal to 1.

TABLE 4. Rank-biased overlap r between the set of neighbors of the overlapping nodes and the set of hubs. p determines the weights of the elements. More weights is given to the first elements of both sets if p has a small value. Each proportion value is the average of 10 SLPA simulation runs. The standard deviation of RBO values are omitted from this table because the ranges are very small (close to 0).

Network	r				
	$p=0.5$	$p=0.7$	$p=0.8$	$p=0.9$	$p=0.98$
Zachary's karate club	0.997	0.986	0.971	0.948	0.923
Les Miserables	0.999	0.999	0.999	0.989	0.91
Dolphins	0.989	0.963	0.945	0.934	0.956
Game of thrones	0.999	0.999	0.998	0.990	0.978
American college football	0.996	0.965	0.916	0.831	0.834
Terrorist	0.823	0.769	0.685	0.663	0.663
Ecologie	0.997	0.986	0.973	0.955	0.928
Adjnoun	0.999	0.999	0.998	0.99	0.981
Euroroad	0.836	0.764	0.857	0.884	0.92
Crime	0.999	0.999	0.999	0.997	0.98
Yeast Protein Interaction	0.999	0.999	0.999	0.999	0.996
ego-Facebook	0.998	0.982	0.952	0.884	0.844
ca-GrQc	0.999	0.999	0.996	0.977	0.838
Gnutella peer-to-peer	0.923	0.861	0.837	0.813	0.759
Wikipedia vote	1	0.999	0.999	0.999	0.998
ca-HepTh	0.999	0.999	0.999	0.994	0.984
ca-AstroPh	1	0.999	0.999	0.999	0.999
ca-CondMat	1	0.999	0.999	0.999	0.999
Facebook Large Page-Page	0.999	0.999	0.999	0.998	0.979
Enron Email	1	0.999	0.999	0.999	0.999
Amazon	0.999	0.999	0.999	0.996	0.961
DBLP	0.999	0.999	0.985	0.975	0.919

TABLE 5. The correlation between the set of neighbors of the overlapping nodes and the set of hubs for Karate club network. ρ is the Pearson correlation while ρ_s is the Spearman correlation.

	Overlapping nodes	Neighbors of overlapping nodes	Hubs	ρ	ρ_s
Label of nodes	28, 27, 30, 8, 31	33, 0, 32, 2, 1, 23, 25, 24	33, 0, 32, 2, 1, 3, 13, 23	0.994	0.986
Degree of nodes	3, 4, 4, 5, 6	17, 16, 12, 10, 9, 5, 3, 3	17, 16, 12, 10, 9, 6, 5, 5		
Rank of nodes	9, 8, 8, 7, 6	1, 2, 6, 8, 9, 13, 15, 15	1, 2, 6, 8, 9, 12, 13, 13		

Spearman correlation is then computed based on the set of ranks of neighbors of the overlapping nodes and the set of ranks of hubs according to Equation 7. It is noticed from Table 5 that nodes belonging to the set of neighbors of the overlapping nodes as well as the set of hubs have about the same degrees and ranks (the overlapping elements of both sets are in the blue color). That explains why the values of both Pearson and Spearman correlation are very high (close to 1). Therefore, there is a very strong monotonic relationship between the set of neighbors of the overlapping nodes and the set of hubs in Karate club network.

Results of Pearson and Spearman correlation for all the other empirical networks are reported in Table 6. One can notice that the values of the correlation measures are most of the time greater than 0.9 for different networks whatever their size and origin. Therefore, it clearly indicates that there is a very strong relationship between the set of hubs and the set of neighbors of the overlapping nodes for all the networks under test. Indeed, the degrees, as well as the ranks of nodes belonging to the set of neighbors of the overlapping nodes, decrease in the same way as those belonging to the set of hubs. Globally, both sets have nearly the same degrees and ranks. These results corroborate the findings made using the previous measures. Thus, one can say that a large amount of nodes belong to both sets.

TABLE 6. Correlation between the set of neighbors of the overlapping nodes and the set of hubs. ρ is the Pearson correlation while ρ_s is the Spearman correlation. Each proportion value is the average of 10 SLPA simulation runs.

Network	ρ	σ_ρ	ρ_s	σ_{ρ_s}
Zachary's karate club	0.991	0.0018	0.98	0.004
Les Miserables	0.979	0.0013	0.972	0.0024
Dolphins	0.973	0.0028	0.944	0.0122
Game of thrones	0.99	0.003	0.99	0.0032
American college football	0.804	0.0146	0.81	0.0081
Terrorist	0.969	0.0011	0.974	0.0055
Ecologie	0.974	0.0062	0.989	0.0012
Adjnoun	0.992	0.0052	0.983	0.0076
Euroroads	0.921	0.0096	0.922	0.006
Crime	0.985	0.0037	0.983	0.0021
Yeast Protein Interaction	0.998	0.0055	0.998	0.0019
ego-Facebook	0.985	0.0044	0.98	0.0062
ca-GrQc	0.969	0.007	0.976	0.0033
Gnutella peer-to-peer	0.859	0.0002	0.853	0.0069
Wikipedia vote	0.999	0.0079	0.999	0.0003
ca-HepTh	0.99	0.0053	0.99	0.0024
ca-AstroPh	0.999	0.0047	0.999	0.0009
ca-CondMat	0.989	0.0076	0.99	0.0041
Facebook Large Page-Page	0.999	0.0027	0.999	0.001
Enron Email	0.983	0.0124	0.98	0.0043
Amazon	0.97	0.0019	0.966	0.0036
DBLP	0.99	0.0036	0.987	0.0015

5) HUB DISTRIBUTION ACCORDING TO THE DEGREE OF OVERLAPPING NODES

One interesting question is to know if hubs tend to link more with overlapping nodes with a high degree. In order to

TABLE 7. α_1 and α_2 represent the estimated Power-Law exponent of the degree distribution for respectively the set of neighbors of the overlapping nodes and the set of hubs. Each proportion value is the average of 10 SLPA simulation runs. The standard deviation values are omitted in this table because their values are close to 0.

Network	α_1	α_2
ego-Facebook	2.03	2.01
ca-GrQc	2.28	2.19
ca-HepTh	2.42	2.54
ca-CondMat	2.61	2.58
Enron Email	2.41	2.39

answer it, we compute the proportion of hubs in the neighborhood of the overlapping nodes as a function of the degree of overlapping nodes. Results for the various networks are reported in Figure 9. One can see from these figures that the distributions are quite uniform. This result is valid for almost all the overlapping nodes whatever their degree and for all the networks. Therefore, we can conclude that hubs are linked to overlapping nodes quite uniformly.

6) DEGREE DISTRIBUTION

Here, we examine the degree distribution of both the set of neighbors of the overlapping nodes and the set of hubs. Our purpose is to check if they present the same empirical distribution. The degree distribution can be appropriately described as a power-law ($P(k) = k^{-\alpha}$) for a wide number of networks [32], [49]. The value of the exponent of the power-law ranges between 2 and 3 according to many experimental studies [32]. Figure 10 represents the cumulative degree distribution of neighbors of the overlapping nodes as well as the hubs for the various empirical networks. These figures show that the cumulative degree distribution of both sets displays the same behavior. In addition, the neighbors of the overlapping nodes as well as the hubs have approximately the same cumulative degree distribution. This is more apparent namely for Condense matter collaboration and Enron Email networks. Moreover, Figure 11 reports the empirical degree distribution of the sets of neighbors of the overlapping nodes together with the hubs for all the real-world networks. It is shown in this figure that the two sets exhibit similar degree distribution (described by a heavy tail degree distribution). Their estimated distribution under the power-law hypothesis is also represented in this figure. The maximum likelihood estimators [31] is used to compute the exponent values which are reported in Table 7. For all the networks, it is noticed that the estimated exponents of the theoretical distribution for neighbors of the overlapping nodes and hubs have quite the same values. Therefore, one can conclude from this experiment that the power-law seems to be a suitable distribution for both sets. Additionally, their exponents have about the same values.

B. COMPARISON BETWEEN THE GLOBAL NETWORK AND THE OVERLAP-HUB NETWORK

In this experiment, we compare the mean and median distance values of the Overlap-Hub network (containing only the

overlapping nodes and the hubs) with those of the original network. Figs. 1 to 8 (b) are the bipartite representation of the Overlap-Hub network for respectively Karate club, Les Misérables, Dolphin and Game of Thrones networks. Nodes in the left column (colored in gray) are the overlapping nodes while the nodes in the right column are the hubs of the overall network presented in Figs. 1 to 8 (a). The distance of the Overlap-Hub network is quantified by its average and median measure. It tells us how much the overlapping nodes are close to the hubs. The smaller the value of this measure, the more likely the overlapping nodes are neighbors to the hubs. Table 11 reports the average and the median distance of the Overlap-Hub network (distance between the overlapping nodes and the hubs) together with the same measures for the global network.

Results show that the average distance values for the Overlap-Hub network are quite small. They are lower than the average distance of the global network. In small networks (Karate club, Les Misérables, Dolphin, Game of Thrones and American football network), one can also notice that the average distance of the Overlap-Hub network is slightly lower than the one of the global network. In these networks, the density of links is quite high as compared to the density of large networks. On top of that, small networks have also a relatively small diameter. So, the nodes are tightly connected even if we consider the entire network. That what makes the average shortest path for the global network has also a small value. This explains the small differences between the values of these measures computed on both networks.

In medium and large networks, the average distance of the Overlap-Hub network has always a relatively very small value as compared to the same measure computed on the global network. These networks have very low links density while having a large diameter. That explains why the average distance between overlapping nodes and hubs is much smaller than the average distance of the global network. This means that the overlapping nodes and the hubs are tightly connected as compared to the rest of the network. Moreover, results also show that the median distance displays the same behavior as for the average distance. Indeed, the median distance of the Overlap-Hub network is slightly lower or equal to the one of the global network for all the networks with small sizes. However, the difference is more significant in medium and large networks. Therefore, this confirms that each overlapping node can be neighbor to a large amount of hubs.

C. INFLUENCE OF THE NEIGHBORHOOD SIZE OF OVERLAPPING NODES

In the previous experiments, we checked if the hubs are immediate neighbors of the overlapping nodes. Now, we want to know if there are more hubs in the vicinity of the overlapping nodes. To do so, we increase the size of the neighborhood of the overlapping nodes by increasing the size of the max distance to define the neighborhood. This parameter represents the maximum number of links that must be taken from an overlapping node to reach the perimeter of its neighborhood.

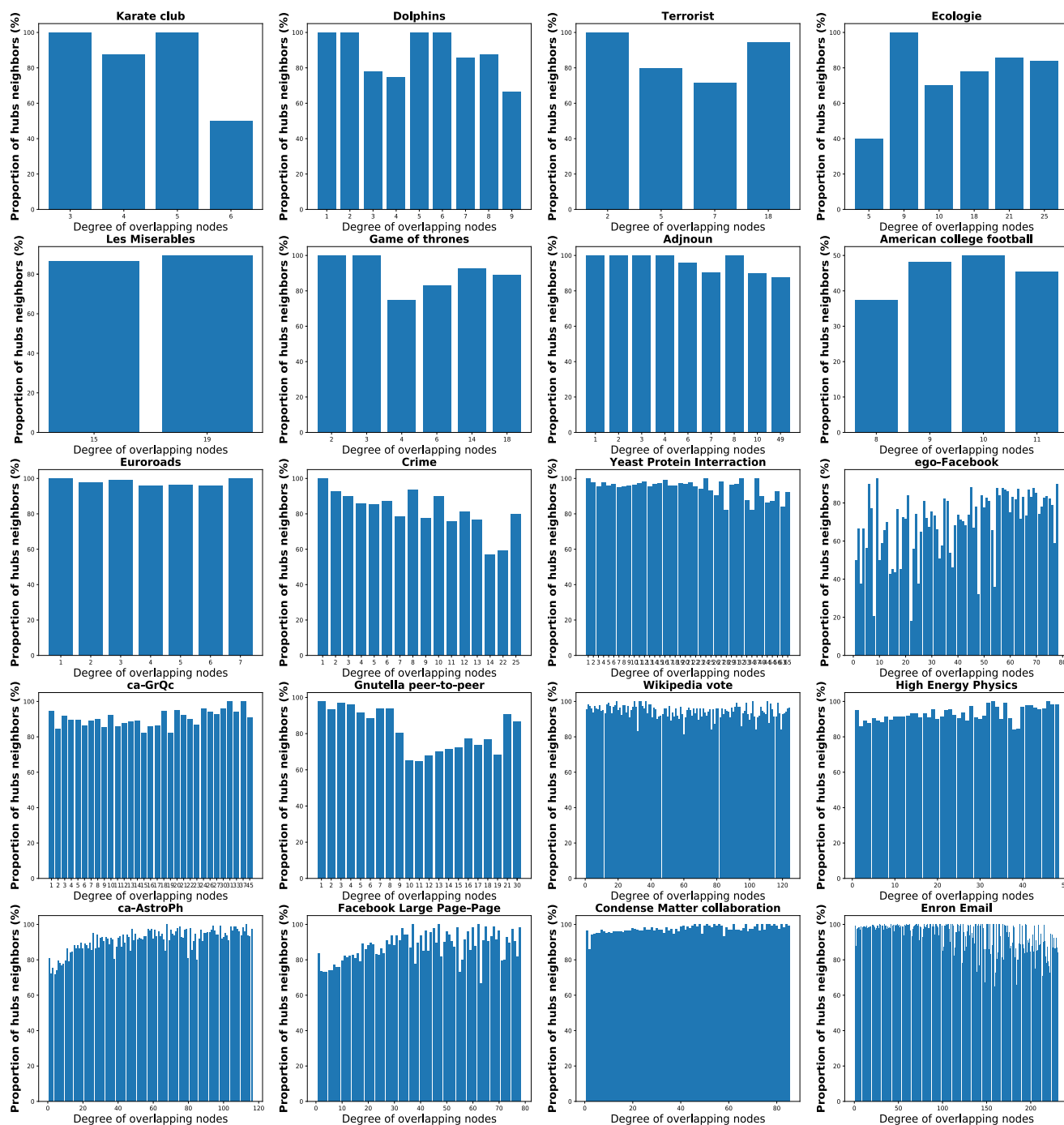


FIGURE 9. Proportion of hubs in the neighborhood of overlapping nodes as a function of their degree for real-world networks under study.

Thus, we select only the immediate neighbors of the overlapping nodes if we set the distance to 1. By incrementing the distance, we increase also the circle of the overlapping nodes' neighborhood. The purpose of this investigation is to examine the influence of the size of the overlapping neighborhood on the proportion of hubs belonging to this area. Table 10 reports the proportion of hubs in the neighborhood of overlapping nodes as a function of the distance. One can see

that the values of the proportion of common nodes in the set of hubs and the set of neighbors of overlapping nodes hubs exhibit the same behavior. They increase with the distance. Note that it is equal to 100% for small size networks when the distance is only incremented once. This is because these networks have a small diameter. In networks of medium and big sizes, the proportion of hubs does not reach its maximum value until setting the distance to 5 or 6. This is due to their

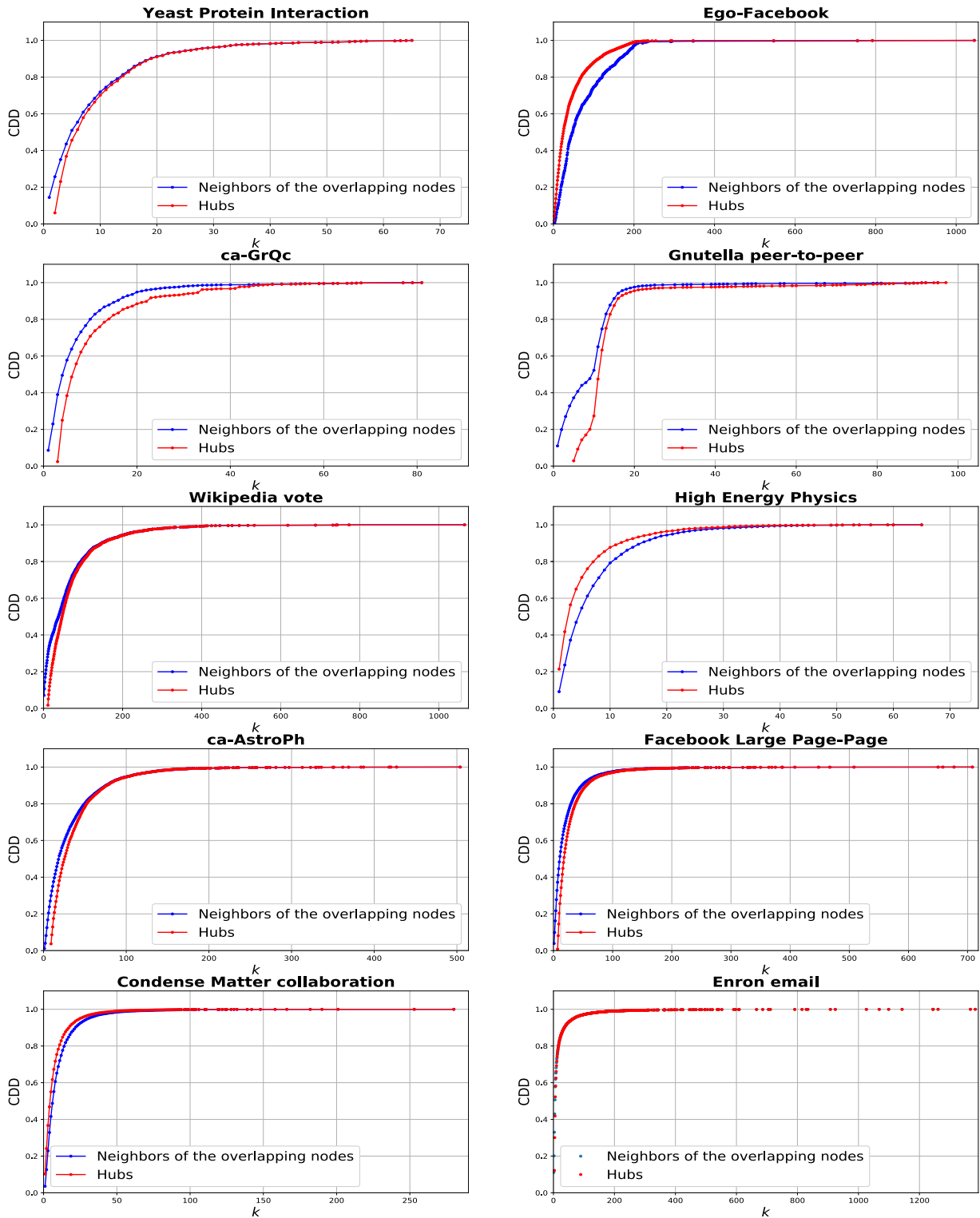


FIGURE 10. Cumulative degree distribution of neighbors of the overlapping nodes (blue) and the hubs (red) for the networks under study.

relatively larger diameter. We note the maximum value of the proportion of hubs in the neighborhood of overlapping nodes is usually less than 100%. This is because some hubs belong also to the set of overlapping nodes. Actually, by increasing the size of the neighborhood of the overlapping nodes, there

is a high chance that it contains more hubs belonging to different communities. That is the reason why the proportion of hubs always increases while increasing the size of the neighborhood of the overlapping nodes whatever the network origin and size.

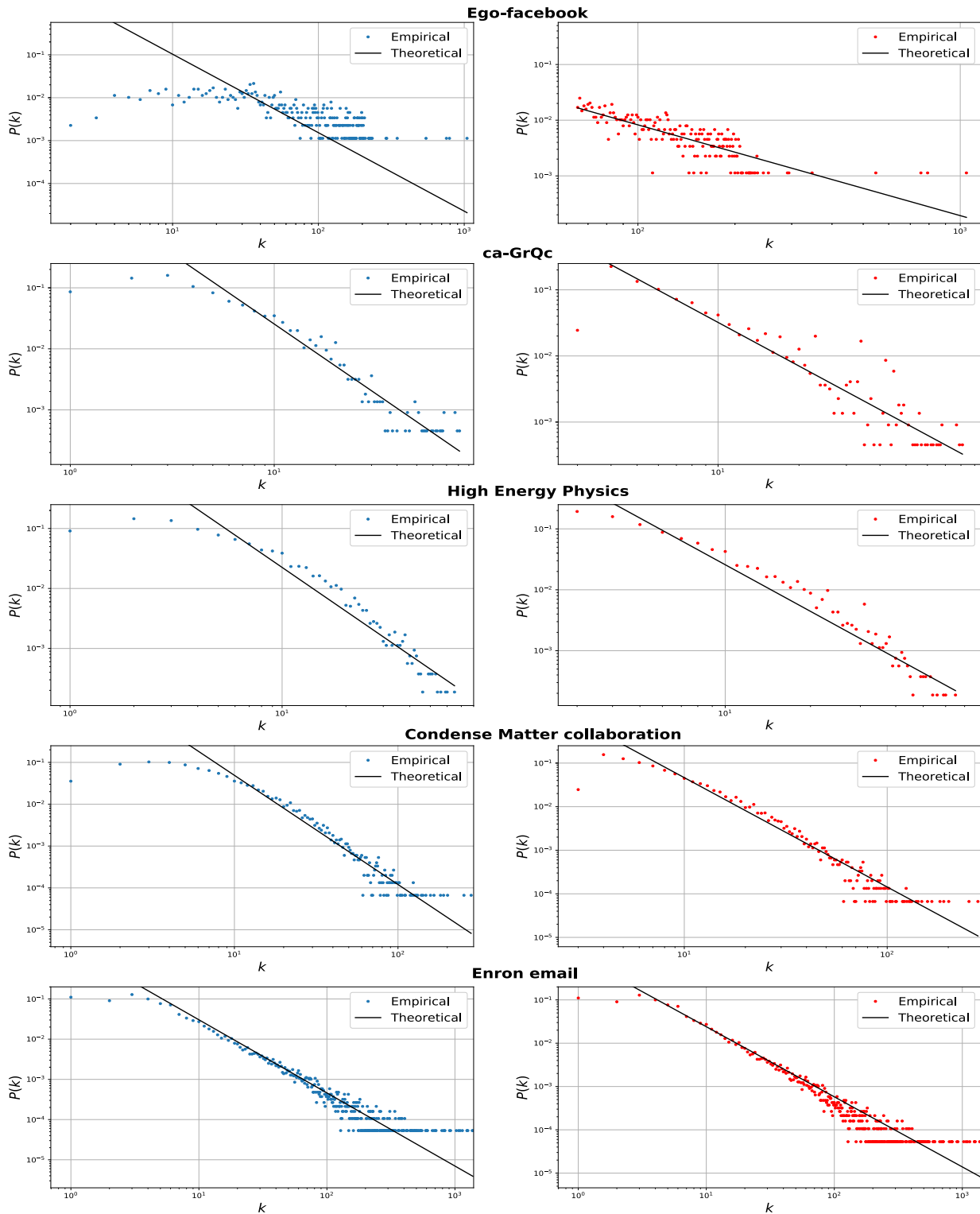


FIGURE 11. Log-log representation of degree distribution for different real-wold networks. The empirical degree distribution of neighbors of the overlapping nodes is in blue. It is in red for the hubs. Power-Law estimates are in black.

D. INFLUENCE OF THE MEMBERSHIP NUMBER OF THE OVERLAPPING NODES

The aim of this experiment is to study the influence of the membership number of overlapping nodes on the proportion

of hubs in their neighborhood. Remember that the membership number is the number of communities to which the overlapping node belongs. Table 8 reports the example of Dolphin networks. This network has thirteen overlapping nodes with

TABLE 8. Proportion of hubs in the neighborhood of overlapping nodes as a function of their membership for Dolphin network. om is the membership degree of the overlapping nodes. A_n is the proportion of hubs in the neighborhood of the overlapping nodes.

Network	om	Overlapping nodes	Neighbors of overlapping nodes	Hubs	A_n (%)
Dolphins	2	CCL, Double, Jonah, MN83, PL, SN100, SN9, Thumper, Topless, Trigger, Vau, Zap, Zipfel	Grin, SN4, Scabs, Beescratch, SN63, TR99, TR77, SN96, Shmuddel, DN63 , TSN103, Bumper, Knit, TSN83, Cross	Grin, SN4, Scabs, Web, Jet, Beescratch, SN63 , Gallatin, Stripes, TR99 , Feather, Upbang, TR77 , Hook, SN96 , DN21, Shmuddel , Fish, DN63	66.6
	3	Beak, Kringel, SN89, Five, Haecksel, MN105, MN60, Oscar, Patchback, SMN5	Grin, SN4, Scabs, Web, Beescratch, SN63, Stripes, TR99, TR77, Hook, SN96, Fish	Grin, SN4, Scabs, Web, Jet, Beescratch, SN63 , Gallatin, Stripes, TR99 , Feather, Upbang, TR77, Hook, SN96 , DN21, Shmuddel, Fish , DN63	100

a membership degree equal to 2, while it has eleven overlapping nodes with a membership degree equal to 3. It is shown in this table that the overlapping nodes with the lower membership ($om = 2$) has also a lower proportion of hubs in their neighborhood. However, this proportion reaches 100% in the neighborhood of the overlapping nodes with the higher membership ($om = 3$). Table 9 reports the proportion of hubs in the neighborhood of overlapping nodes as a function of their membership number for the other networks under study. In some networks, all the overlapping nodes have the same membership number (Karate club, Les Miserable, Game of thrones, Dolphins and ego-Facebook). They are not represented in this table. First, note that the values of the overlap with the set of hubs generally increases with the membership number value. In ca-GrQc, for instance, the proportion of hubs is around 86% if the overlapping nodes belong to only two different communities, while it is 100% when the membership of the overlapping nodes is equal to 6. Indeed, if the overlapping node belongs to several communities, it is more likely to be neighbor to a high number of hubs belonging to these multiple modules. In other words, the proportion of hubs in the neighborhood of an overlapping node is higher if it belongs to a higher number of communities. However, as the size of the networks increases, differences are less pronounced.

E. INFLUENCE OF THE COMMUNITY DETECTION ALGORITHMS

In this section, we report the same series of experiments conducted on all the empirical networks of various sizes using alternative overlapping community detection algorithms to SLPA. Both Lancichinetti Fortunato Method and EAGLE (LFME) [45] and Democratic Estimate of the Modular Organization of a Network (DEMON) [46] detection algorithms are used in these experiments. Our main goal is to check the validity of our hypothesis while using a different overlapping community detection algorithm.

1) COMPARISON OF THE UNCOVERED COMMUNITY STRUCTURES

Here, we analyze five measures to highlight the differences between the revealed communities by the 3 community detection algorithms used in the experiments. The Normalized

Mutual Information (NMI), the mixing parameter estimate, the number of communities, the fraction of overlapping nodes and the overlapping modularity measure are computed to compare the community structures uncovered by SLPA, LFME and DEMON. Results for all the networks are reported in Table 12.

In the community detection literature, [50] the Normalized Mutual Information (NMI) is commonly used to compare two community structures. Its value is close to 1 if the community structures are very similar and close to 0 if they do not share any information. In small networks, the NMI between the three algorithms has medium values. Yet, it is slightly higher between SLPA and LFME algorithms. This shows that there are significant similarities between the communities revealed by the three algorithms. In medium and large networks, the NMI has relatively small values for all the algorithms. It is higher between SLPA and LFME detection algorithms. This indicates that in this type of networks, the community structures are quite different. We also report the mixing parameter in this table. This measure represents the fraction of the inter-community links of nodes. It determines the strength of the community structure. For small values, the communities are well defined. For all the networks, the mixing parameter has small values for the three detection algorithms. However, its values are quite close between SLPA and LFME algorithms while it is higher for DEMON algorithm. Thus, the community structure defined by both SLPA and LFME algorithms has similar strengths. Moreover, for small networks, the three algorithms detect nearly the same number of communities. This confirms the similarity of the community structure in this type of networks. In medium and large networks, we observe a larger variation of this measure. The proportion of overlapping nodes is another evaluated parameter. The results show that LFME detects the lower fraction of overlapping nodes followed by SLPA algorithm while this measure is quite high for DEMON algorithm.

Furthermore, the modularity is also used to quantify the quality of the community structure. It assesses the internal connectivity of the identified communities as compared to a random network with no community structure. Note that the overlapping modularity values are low for all the three detection algorithms. However, SLPA has the highest values followed by LFME while DEMON has the lowest values.

TABLE 9. Proportion of hubs in the neighborhood of overlapping nodes as a function of their membership. om is the membership degree of the overlapping nodes. $on(\%)$ is the proportion of overlapping nodes. A_n is the proportion of hubs in the neighborhood of overlapping nodes. Each proportion value is the average of 10 SLPA simulation runs. The standard deviation values are omitted in this table because of their small values.

Network	om	$on(\%)$	$A_n(\%)$
Dolphin	2	13.63	66.6
	3	10.37	100
American college football	2	11.3	77.32
	3	4.3	89.52
Euroroads	2	20.52	97.65
	3	5.45	98.51
	4	3	100
Crime	2	22.43	91.02
	3	3.74	92.29
Yeast Protein Interaction	2	22.91	97.18
	3	7.07	97.42
	4	1.27	98.61
	5	0.5	99.11
	6	0.1	100
ca-GrQc	2	18.49	86.28
	3	5.14	88.16
	4	1.39	89.01
	5	0.81	89.21
	6	0.6	100
Gnutella peer-to-peer	2	14.14	88.81
	3	0.19	92.67
Wikipedia vote	2	13.5	92.42
	3	0.16	95.73
High Energy Physics	2	22.55	76.38
	3	6.27	80.95
	4	1.72	90.04
	5	0.4	90.44
	6	0.25	90.66
ca-AstroPh	2	10.1	78.14
	3	0.4	80.88
	4	0.68	100
Facebook Large Page-Page	2	11.84	96.83
	3	1.92	97.06
	4	0.5	97.57
Condense Matter collaboration	2	19.4	94.1
	3	7.2	94.31
	4	1.91	94.4
	5	0.39	94.72
	6	0.24	95.42
	7	0.13	97.5
Enron Email	2	28.66	96.83
	3	5.77	96.91
	4	0.52	97.06
	5	0.05	97.57
DBLP	2	8.88	72.44
	3	1.9	80.96
	4	1.42	85.74
	5	1.11	88.28
	6	0.96	92.67
	7	0.95	93.48
	8	0.63	98.3
Amazon	2	15.04	78.15
	3	2.91	85.30
	4	2.18	85.74
	5	1.69	92.18
	6	1.45	96.31
7	0.97	98.3	

Therefore, according to this measure for all the various networks, SLPA is the most accurate algorithm, followed by LFME then DEMON.

2) COMPARISON OF THE DISCOVERED OVERLAPPING NODES

Here, we compare the sets of overlapping nodes defined by the various detection algorithms. Our purpose is to check if all the community detection algorithms uncover the same set of overlapping nodes. To do so, we measure the proportion of the overlaps between the sets two by two using the Jaccard index as well as the Rank-biased overlap. Let's start with the small networks that can be visualized. Figure 12 and 13 present the tripartite representation of respectively Karate club and Les Miserables networks using (a) SLPA (b) LFME (c) DEMON algorithms. We note that the left column of the network represents the overlapping nodes, the nodes in the middle column are the hubs while those located in the right column are the remaining nodes. In Karate club, the sets of overlapping nodes detected by SLPA, LFME and DEMON have only one node in common (node 30), while two overlapping nodes are shared by SLPA and DEMON (nodes 30 and 8). In this network, the overlap proportion between SLPA and LFME is higher than the one between SLPA and DEMON. The same behavior is noticed in Les Miserables. The sets of overlapping nodes detected by SLPA, LFME and DEMON have only two nodes in common (Fantine and Marius). In these networks, DEMON algorithm detects a larger number of overlapping nodes as compared to the two other algorithms. In addition, it is also noticed that some nodes defined as hubs with SLPA and LFME algorithms belong to the set of overlapping nodes for DEMON algorithm. That is why the overlaps between the sets of overlapping nodes for SLPA and LFME algorithms are higher than when DEMON algorithm is used. This is true in Karate club and Les Miserables networks. Table 13 reports the proportion of overlaps between the sets of overlapping nodes for all real-world networks when different detection algorithms are used. Results show that the overlaps between the sets of overlapping nodes for the different algorithms are small. They do not exceed 34%. However, unlike the previous networks (Karate club and Les Miserables), SLPA and DEMON algorithms have generally more nodes in common as compared to the LFME algorithm. Table 13 reports also the Rank-biased overlap between these sets. In this measure, we set the parameter p to the value 0.5 to give more weights to the top-ranked nodes. The values of the RBO confirm the findings obtained by the previous results. They display small values. However, they are higher between SLPA and DEMON algorithms. That means that there are usually higher similarities between the sets of overlapping nodes generated by SLPA and DEMON algorithms. These two algorithms detect a large proportion of overlapping nodes while LFME detects a smaller proportion. Therefore, the sets of overlapping nodes defined by SLPA and DEMON algorithms have more nodes in common than the one defined by LFME algorithm.

3) COMPARISON OF THE DISCOVERED NEIGHBORS OF THE OVERLAPPING NODES

Here, we repeat the same experiments between the sets of neighbors of the overlapping nodes defined by SLPA, LFME

TABLE 10. Proportion of hubs in the neighborhood of overlapping nodes as a function of the distance. The distance is the maximum number of links from an overlapping node to reach the perimeter of its neighborhood. A_n is the proportion of hubs in the neighborhood of overlapping nodes. A_n^o is the proportion of hubs belonging to the set of overlapping nodes. S represents the size of the neighborhood of the overlapping nodes. Each proportion value is the average of 10 SLPA simulation runs. The standard deviation values are omitted in this table because of their small values ranging from 0 to 3.7.

Network	d=1			d=2			d=3			d=4			d=5			d=6		
	$A_n(\%)$	$A_n^o(\%)$	$S(\%)$	$A_n(\%)$	$A_n^o(\%)$	$S(\%)$	$A_n(\%)$	$A_n^o(\%)$	$S(\%)$	$A_n(\%)$	$A_n^o(\%)$	$S(\%)$	$A_n(\%)$	$A_n^o(\%)$	$S(\%)$	$A_n(\%)$	$A_n^o(\%)$	$S(\%)$
Zachary's karate club	77.6	9.9	23.52	85.75	14.24	85.29	-	-	-	-	-	-	-	-	-	-	-	-
Dolphins	77.89	3.92	35.48	86.81	13.18	69.35	-	-	-	-	-	-	-	-	-	-	-	-
Terrorist	77.77	8	43.54	93.1	6.89	93.54	-	-	-	-	-	-	-	-	-	-	-	-
Ecology	74	4.44	66.46	83.02	16.98	97.1	-	-	-	-	-	-	-	-	-	-	-	-
Les Miserables	81.06	1.85	44.15	97.33	2.66	97.40	-	-	-	-	-	-	-	-	-	-	-	-
Game of thrones	67.73	6.62	33.64	91.91	8.08	92.52	-	-	-	-	-	-	-	-	-	-	-	-
Adjnoun	73.21	10.96	49.1	85.1	14.89	83.92	-	-	-	-	-	-	-	-	-	-	-	-
American college football	77.07	8	33.17	87.5	12.5	83.47	-	-	-	-	-	-	-	-	-	-	-	-
Euroroads	68.96	11.88	38.5	75.8	13.82	40.71	79.42	20.58	58.77	-	-	-	-	-	-	-	-	-
Crime	76.83	7.74	52.16	81.87	18.13	73.22	-	-	-	-	-	-	-	-	-	-	-	-
Yeast Protein Interaction	80.27	11.44	55.82	84.21	15.79	61.07	-	-	-	-	-	-	-	-	-	-	-	-
ego-Facebook	59.25	1.43	17.5	96.77	3.22	96.83	-	-	-	-	-	-	-	-	-	-	-	-
ca-GrQc	64.04	8.5	33.19	73.27	16.65	54.61	76.57	23.43	59.76	-	-	-	-	-	-	-	-	-
Gnutella peer-to-peer	65.66	3.98	41.27	87.43	12.56	85.52	-	-	-	-	-	-	-	-	-	-	-	-
Wikipedia vote	71.99	8.33	38.14	89.19	10.81	85.56	-	-	-	-	-	-	-	-	-	-	-	-
ca-HepTh	75.32	13.13	45.2	77.91	14.21	54.86	78.6	15.19	56.11	79.57	16.43	57.4	80.46	16.56	59.5	82.61	16.39	61.11
ca-AstroPh	72.98	3.94	51.59	87.08	10.42	83.58	88.66	11.34	85.18	-	-	-	-	-	-	-	-	-
Facebook Large Page-page	63.56	3.07	49.1	84.79	10.14	80.08	87.92	12.08	85.57	-	-	-	-	-	-	-	-	-
ca-CondMat	79.57	2.07	51.69	83.84	6.94	54.36	85.66	7.16	56.17	88.2	8.48	57.19	90.92	9.08	60.03	-	-	-
Enron Email	73.54	5.92	48.88	81.2	9.1	55.42	84.34	11.99	56.03	87.15	12.85	58.59	-	-	-	-	-	-
Dblp	74.64	5.3	30.08	76.43	6.36	36.09	78.11	7.63	54.14	79.27	9.15	75.2	81.56	10.07	89.12	83.84	16.16	96.25
Amazon	71.37	7.98	41.13	73.51	9.57	49.35	75.71	11.49	74.03	79.28	12.64	81.66	85.66	14.34	87.98	-	-	-

TABLE 11. Distance measures of the empirical networks. δ_G represents the diameter of the global network while δ is the diameter of the Overlap-Hub network. l_G represents the average distance of the global network while l is the average distance between the overlapping nodes and the hubs. l_{Gm} is the median value of the average distance of the global network while l_m is the median value of the average distance of the Overlap-Hub network. Each proportion value is the average of 10 SLPA simulation runs. The standard deviation values are omitted in this table because the ranges are very small.

Network	δ_G	δ	l_G	l	l_{Gm}	l_m
Zachary's karate club	5	2	2.408	1.616	2	2
Dolphins	8	5	3.356	2.614	3	3
Terrorist	5	4	2.945	2.203	3	2
Ecology	3	3	1.635	1.418	3	3
Les Miserables	5	3	2.641	1.971	3	2
Game of thrones	6	4	2.903	2.572	3	3
Adjnoun	5	4	2.535	2.201	3	2
American college football	4	3	2.508	2.488	3	3
Euroroads	16	9	6.35	3.576	5	2
Crime	10	6	5.04	3.601	3	2
Yeast Protein Interaction	8	5	4.376	2.909	3	2
ego-Facebook	8	5	3.692	2.906	5	2
ca-GrQc	17	8	6.049	1.972	6	3
Gnutella peer-to-peer	9	5	4.642	2.2445	3	2
Wikipedia vote	7	4	3.247	1.96	3	2
ca-HepTh	18	10	5.945	2.351	6	3
ca-AstroPh	14	9	4.194	2.76	3	2
Facebook Large Page-Page	9	6	4.165	2.046	3	2
ca-CondMat	15	13	5.352	2.232	6	3
Enron Email	11	8	4.025	2.136	6	3
DBLP	21	16	7.437	3.971	4	3
Amazon	44	31	6.209	3.677	4	3

and DEMON algorithms. We measure the proportion of overlaps as well as the Rank-biased overlap between these sets two-by-two.

It can be noticed from Figure 12 and 13 that the sets of neighbors of the overlapping nodes revealed by SLPA and LFME have the majority of nodes in common for both Karate club and Les Miserables networks. DEMON has less nodes in common with SLPA and LFME algorithms in these networks. This is because the set of neighbors of the overlapping nodes revealed by DEMON is very large as compared

to the other algorithms. Table 14 reports the proportion of overlaps between the sets of neighbors of overlapping nodes for different empirical networks when various community detection algorithms are used. Results show that the overlap between the sets of neighbors of overlapping nodes for the different algorithms is usually high. They can even reach 78%. Yet, unlike in Karate club and Les Miserables networks, SLPA and DEMON algorithms have generally higher proportion of overlap as compared to the LFME algorithm. Indeed, both SLPA and DEMON detect a large proportion of

TABLE 12. The estimated number of communities N_c , fraction of overlapping nodes on , the mixing parameter μ , the overlapping modularity Q and the Normalized Mutual Information (NMI) in the networks under study for community structures uncovered by SLPA, DEMON and LFME algorithms.

Network	N_c			$on(\%)$			μ			Q			NMI		
	SLPA	DEMON	LFME	SLPA	DEMON	LFME	SLPA	DEMON	LFME	SLPA	DEMON	LFME	SLPA/DEMON	SLPA/LFME	DEMON/LFME
Zachary's karate club	2	3	3	15.1	14.7	8.82	0.051	0.115	0.106	0.205	0.101	0.196	0.322	0.444	0.496
Les Miserables	5	5	7	2.5	14.28	3.89	0.074	0.098	0.083	0.207	0.103	0.221	0.343	0.761	0.395
Dolphins	4	9	5	26	30.03	9.67	0.018	0.226	0.139	0.204	0.086	0.191	0.367	0.368	0.221
Game of thrones	3	3	6	7.47	15.82	7.47	0.034	0.198	0.073	0.213	0.176	0.185	0.396	0.424	0.312
American college football	7	6	6	15.6	16.52	6.08	0.169	0.256	0.218	0.188	0.131	0.188	0.328	0.38	0.272
ego-Facebook	12	57	91	3.16	9.16	5.79	0.016	0.118	0.092	0.219	0.182	0.205	0.211	0.262	0.209
ca-GrQc	676	519	844	25.3	25.27	18.35	0.102	0.207	0.123	0.199	0.098	0.186	0.22	0.615	0.27
ca-HepTh	1044	1109	1001	31.7	28.15	11.75	0.188	0.3	0.24	0.18	0.081	0.168	0.183	0.382	0.198
ca-CondMat	1889	3169	1701	29.37	40.6	9.34	0.234	0.087	0.282	0.174	0.098	0.164	0.215	0.351	0.277
Enron Email	1503	1811	1101	35	36.23	19.16	0.114	0.083	0.15	0.143	0.011	0.156	0.241	0.402	0.235

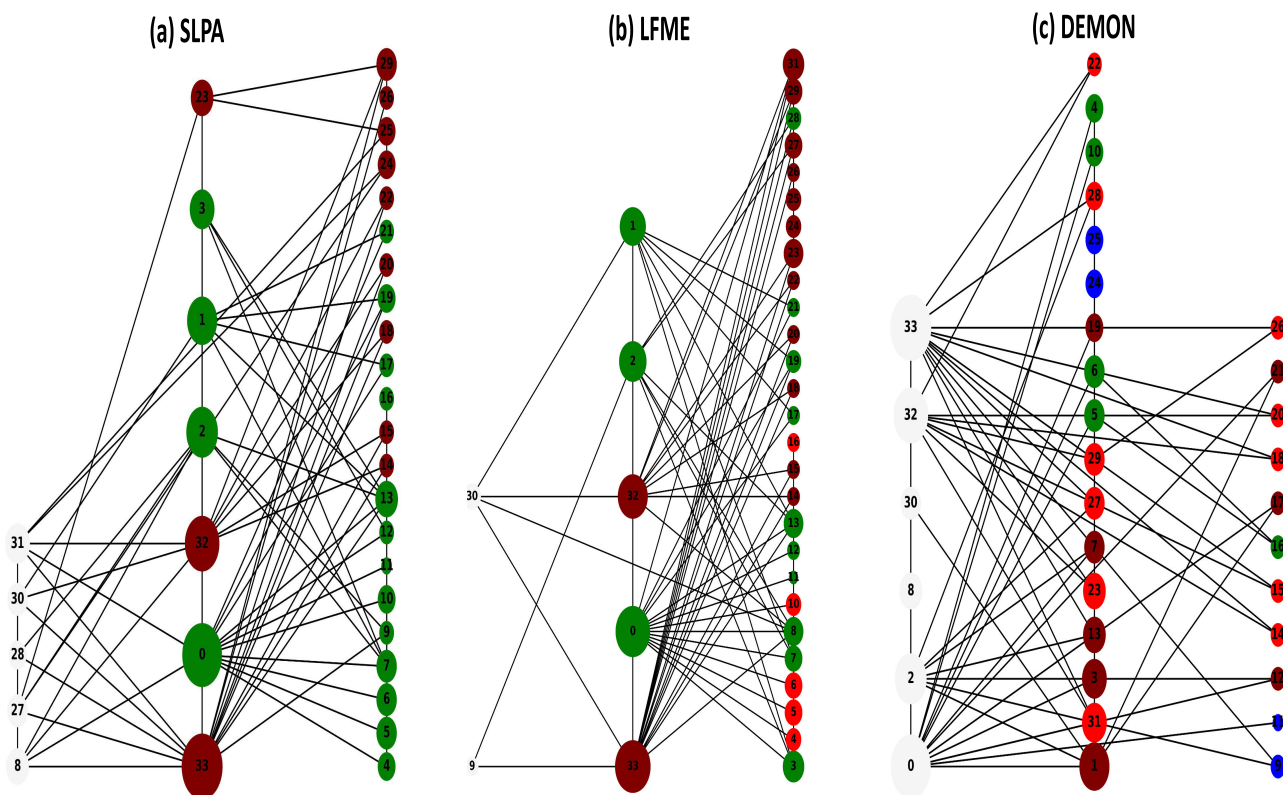


FIGURE 12. Tripartite representation of Karate network using (a) SLPA (b) LFME (c) DEMON. Nodes are highlighted in different colors according to the community they belong to. Nodes with the same color belong to the same community while those in gray represent the overlapping nodes. We note that the left column of the network are the overlapping nodes, the nodes in the middle column are the hubs while those located in the right column represent the rest of the nodes. The size of the nodes is proportional to their degree.

overlapping nodes. Thus, their neighborhood of the overlapping nodes is also large as compared to LFME algorithm. Table 14 reports also the Rank-biased overlap between the sets of neighbors of the overlapping nodes. The parameter p is set to the value 0.5 to give more weights to the top-ranked nodes. It is noticed that the RBO exhibits also very high values (close to 1). That means that the most highly connected nodes belong to all the sets of neighbors of the overlapping nodes defined by the three community detection algorithms.

4) COMPARISON OF THE DISCOVERED OVERLAP-HUB NETWORKS

In this part of the experiment, we measure the similarities between the discovered sets of the nodes forming the

Overlap-Hub network (overlapping nodes and hubs) when SLPA, LFME and DEMON algorithms are used. To do so, we compute the Jaccard index as well as the Rank-biased Overlap between the revealed sets.

It can be noticed from Figure 12 and 13 that the nodes of the Overlap-Hub network (left and middle column) uncovered by the three community detection algorithms in Karate club and Les Miserables networks have a majority of nodes in common. Indeed, some nodes can be identified as overlapping nodes by some algorithms and as hubs by others. For instance, nodes 33, 0 and 32 in Karate club network are considered as hubs with SLPA and LFME algorithms and as overlapping nodes with DEMON algorithm. The same behavior appears in Les Miserables network for nodes Valjean, Gavroche, Javert and Enjolras. They are considered as hubs

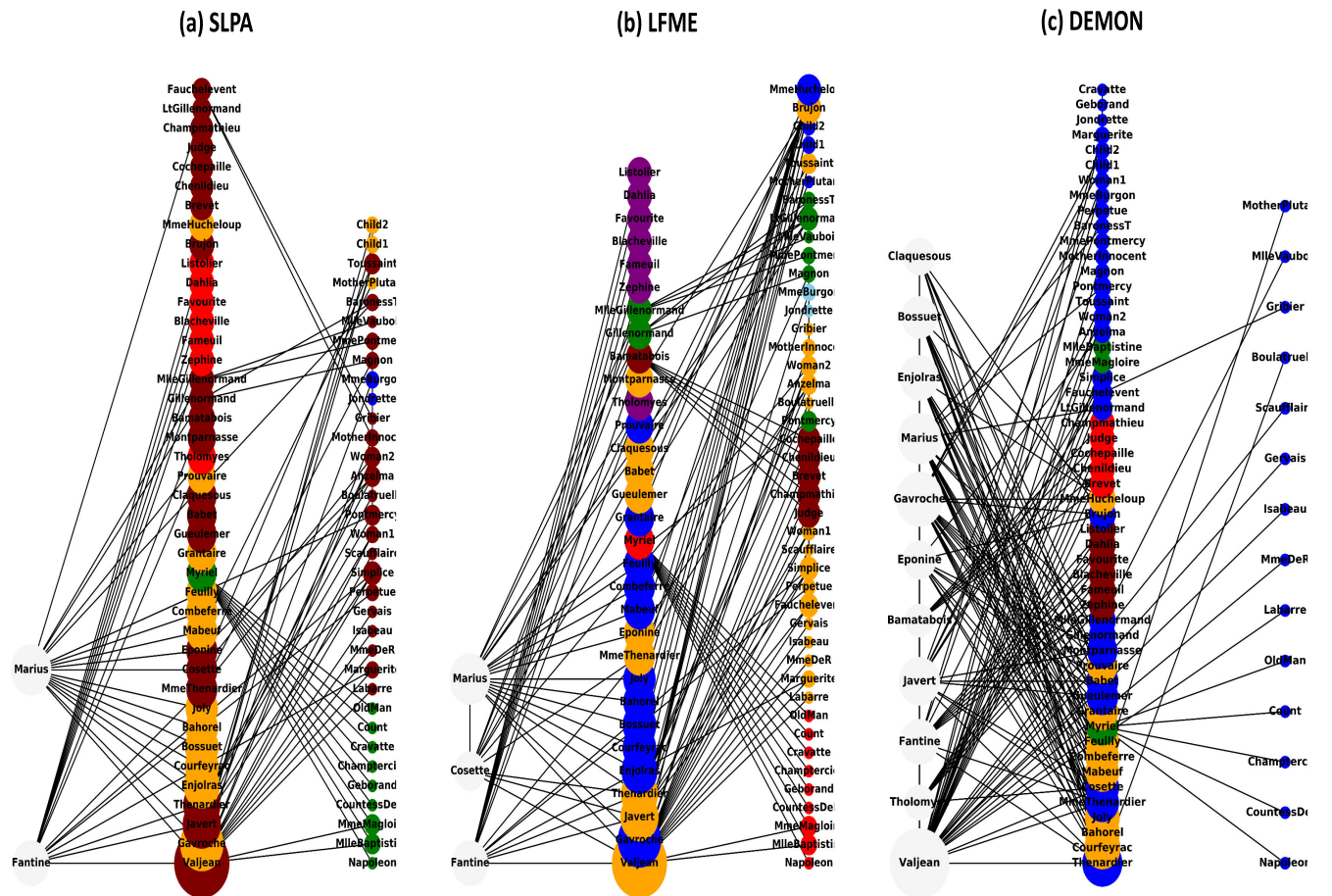


FIGURE 13. Tripartite representation of Les Miserables network using (a) SLPA (b) LFME (c) DEMON. Nodes are highlighted in different colors according to the community they belong to. Nodes with the same color belong to the same community while those in gray represent the overlapping nodes. We note that the left column of the network are the overlapping nodes, the nodes in the middle column are the hubs while those located in the right column represent the rest of the nodes. The size of the nodes is proportional to their degree.

for SLPA and LFME algorithms and as overlapping nodes for DEMON algorithm. One can find the explanation in the study reported by Yang *et al.* [23]. Indeed, the authors show that the overlapping nodes are more densely connected than the other nodes of the network. Thus, they can be considered as hubs by some community detection algorithms. Table 15 reports the Jaccard index between the sets of overlapping nodes and hubs revealed by different community detection algorithms. Experimental results show that the Jaccard index values are very high (close to 1). The values are even higher between SLPA and DEMON algorithm. Therefore, the Overlap-Hub networks defined by the different algorithms have a majority of nodes in common. Table 15 reports also the Rank-biased overlap between the same sets for the three tested algorithms. We note that the parameter p is set to the value 0.5 to give more weights to the top-ranked nodes. It can be noticed that the RBO values are in most of the time equal to 1. These values consolidate the results obtained using the Jaccard index. It shows that the Overlap-Hub networks uncovered by SLPA, LFME and DEMON algorithms are very similar.

5) COMPARISON BETWEEN THE SET OF NEIGHBORS OF THE OVERLAPPING NODES AND THE SET OF HUBS

In this section, we use the same evaluation criteria to examine the relationship between the overlapping nodes and the hubs using LFME and DEMON algorithms. At first, the set of neighbors of the overlapping nodes and the set of hubs are compared. To do so, the previous measures (proportion of hubs, Rank-biased overlap and correlation) are computed. We also analyze their degree distribution.

Table 16 represents the proportion of hubs in the neighborhood of the overlapping nodes. As for SLPA, in all the studied networks, the proportion of hubs in the overlapping neighborhood has always high values for both LFME and DEMON. However, the fraction of hubs neighbors of the overlapping nodes is higher for DEMON algorithm. This is due to the higher number of the overlapping nodes revealed by this detection algorithm. In this case, we have a larger neighborhood of the overlapping nodes. Thus, there is a higher chance that it contains more hubs. Table 17 reports the Rank-biased overlap for LFME and DEMON algorithms. Note that

TABLE 13. Overlap between the sets of overlapping nodes discovered by SLPA, LFME and DEMON for different real-world networks. J represents the Jaccard Index. r is the Ranked-biased Overlap. Its parameter p is set to the value 0.5 to give more weight to the top-ranked nodes.

Network		J			r		
		SLPA	LFME	DEMON	SLPA	LFME	DEMON
Karate Club	SLPA	1	0.16	0.22	1	0.2	0.25
	LFME	0.16	1	0.14	0.2	1	16.12
	DEMON	0.22	0.14	1	0.25	16.12	1
Les Miserables	SLPA	1	0.66	0.18	1	0.85	0.13
	LFME	0.66	1	0.16	0.85	1	0.09
	DEMON	0.18	0.16	1	0.13	0.09	1
Dolphins	SLPA	1	0.25	0.31	1	0.724	0.905
	LFME	0.25	1	0.13	0.724	1	0.738
	DEMON	0.31	0.13	1	0.905	0.738	1
Game of thrones	SLPA	1	0.066	0.23	1	0.097	0.1
	LFME	0.066	1	0.083	0.097	1	0.075
	DEMON	0.23	0.083	1	0.1	0.075	1
A. college football	SLPA	1	0.04	0.075	1	0.088	0.114
	LFME	0.04	1	0	0.088	1	0.087
	DEMON	0.075	0	1	0.114	0.087	1
ego-Facebook	SLPA	1	0.04	0.08	1	0.03	0.09
	LFME	0.04	1	0.023	0.03	1	0.029
	DEMON	0.08	0.023	1	0.09	0.029	1
ca-GrQc	SLPA	1	0.17	0.21	1	0.11	0.01
	LFME	0.17	1	0.15	0.11	1	0.19
	DEMON	0.21	0.15	1	0.01	0.19	1
High Energy Physics	SLPA	1	0.13	0.24	1	0.07	0.03
	LFME	0.13	1	0.14	0.07	1	0.029
	DEMON	0.24	0.14	1	0.03	0.029	1
Condense Matter	SLPA	1	0.117	0.295	1	0.1	0.069
	LFME	0.117	1	0.118	0.1	1	0.068
	DEMON	0.295	0.118	1	0.069	0.068	1
Enron Email	SLPA	1	0.212	0.321	1	0.218	0.307
	LFME	0.212	1	0.335	0.218	1	0.207
	DEMON	0.321	0.335	1	0.307	0.207	1

the RBO values are quite high for all the community detection algorithms. As for the previous measure, DEMON has always slightly higher values. In addition, the values of the RBO get higher when more weight is given to the top-ranked nodes. That means that the most highly connected nodes of the network belong to the neighborhood of the overlapping nodes for all the community detection algorithms. We also compute both Pearson and Spearman correlation for both detection algorithms. The results are illustrated in Table 18. For all the empirical networks, the values of the two coefficients are most of the time very close to 1 when LFME and DEMON are employed. Therefore, there is a strong monotonic relationship between the set of neighbors of the overlapping nodes and the set of hubs for the three detection algorithms. Globally, DEMON gives better results (in terms of the proportion of hubs, RBO and correlation) as compared to the other algorithms. However, this algorithm defines a very large set of neighbors of the overlapping nodes. It can reach sometimes 71% of the size of the network. In this case, several nodes can be identified as hubs even if they have a small degree. This is because the set of hubs has the same size as the set of neighbors of the overlapping nodes. Thus, LFME identifies a more meaningful set of hubs followed by SLPA.

The degree distribution of the set of neighbors of the overlapping nodes and the set of hubs is also studied for LFME and DEMON algorithms. Figure 14 presents the

TABLE 14. Overlap between the sets of neighbors of the overlapping nodes discovered by SLPA, LFME and DEMON for different real-world networks. J represents the Jaccard Index. r is the Ranked-biased Overlap. Its parameter p is set to the value 0.5 to give more weight to the top-ranked nodes.

Network		J			r		
		SLPA	LFME	DEMON	SLPA	LFME	DEMON
Karate Club	SLPA	1	0.444	0.303	1	0.998	0.809
	LFME	0.444	1	0.161	0.998	1	0.807
	DEMON	0.303	0.161	1	0.809	0.807	1
Les Miserables	SLPA	1	0.711	0.682	1	0.999	0.997
	LFME	0.711	1	0.539	0.999	1	0.997
	DEMON	0.682	0.539	1	0.997	0.997	1
Dolphins	SLPA	1	0.541	0.78	1	0.989	0.997
	LFME	0.541	1	0.557	0.989	1	0.989
	DEMON	0.78	0.557	1	0.997	0.989	1
Game of thrones	SLPA	1	0.416	0.444	1	0.999	0.999
	LFME	0.416	1	0.5	0.999	1	0.999
	DEMON	0.444	0.5	1	0.999	0.999	1
A. college football	SLPA	1	0.405	0.636	1	0.996	0.996
	LFME	0.405	1	0.425	0.996	1	0.999
	DEMON	0.636	0.425	1	0.996	0.999	1
ego-Facebook	SLPA	1	0.312	0.511	1	0.998	1
	LFME	0.312	1	0.378	0.998	1	0.998
	DEMON	0.511	0.378	1	1	0.998	1
ca-GrQc	SLPA	1	0.437	0.555	1	0.999	0.999
	LFME	0.437	1	0.437	0.999	1	0.999
	DEMON	0.555	0.437	1	0.999	0.999	1
High Energy Physics	SLPA	1	0.412	0.62	1	0.999	0.999
	LFME	0.412	1	0.381	0.999	1	0.999
	DEMON	0.62	0.381	1	0.999	0.999	1
Condense Matter	SLPA	1	0.41	0.717	1	1	1
	LFME	0.41	1	0.379	1	1	1
	DEMON	0.717	0.379	1	1	1	1
Enron Email	SLPA	1	0.415	0.634	1	1	1
	LFME	0.415	1	0.343	1	1	1
	DEMON	0.634	0.343	1	1	1	1

degree distribution of neighbors of the overlapping nodes and the hubs for LFME and DEMON algorithms. It can be noticed that the two sets exhibit the same behavior for both algorithms. They display a heavy-tailed degree distribution. The estimated exponents under the power-law hypothesis are also computed and reported in Table 19. Results show that the exponents of the set of neighbors of the overlapping nodes and the set of hubs are very close whatever the used community detection algorithm. Furthermore, we also compute the proportion of hubs while increasing the size of the neighborhood of the overlapping nodes when LFME and DEMON are used. Results reported in Table 21 show that, as for SLPA, the proportion of hubs increases when we increase the neighborhood of the overlapping nodes for both algorithms. Moreover, Table 20 represents the proportion of hubs as a function of the membership of the overlapping nodes. Results show also that the proportion of hubs increases as the membership of the overlapping nodes gets higher for all the community detection algorithms.

6) COMPARISON BETWEEN THE GLOBAL NETWORK AND THE OVERLAP-HUB NETWORK

In the second part of this experiment, we compare also the Overlap-Hub network and the global network using LFME and DEMON algorithms. Table 22 represents the average and the median distance of both networks for LFME and DEMON

TABLE 15. Overlap between the sets of the union of the overlapping nodes and the hubs discovered by SLPA, LFME and DEMON for different real-world networks. J represents the Jaccard Index. r is the Ranked-biased Overlap. Its parameter p is set to the value 0.5 to give more weight to the top-ranked nodes.

Network		J			r		
		SLPA	LFME	DEMON	SLPA	LFME	DEMON
Karate Club	SLPA	1	1	1	1	0.993	0.999
	LFME	1	1	0.857	0.993	1	0.993
	DEMON	1	0.857	1	0.999	0.993	1
Les Miserables	SLPA	1	1	1	1	1	1
	LFME	1	1	1	1	1	0.999
	DEMON	1	1	1	1	0.999	1
Dolphins	SLPA	1	0.944	1	1	0.999	0.999
	LFME	0.944	1	1	0.999	1	0.999
	DEMON	1	1	1	0.999	0.999	1
Game of thrones	SLPA	1	0.959	1	1	0.999	0.999
	LFME	0.959	1	0.959	0.999	1	0.999
	DEMON	1	0.959	1	0.999	0.999	1
A. college football	SLPA	1	0.968	1	1	1	1
	LFME	0.968	1	0.953	1	1	1
	DEMON	1	0.953	1	1	1	1
ego-Facebook	SLPA	1	0.97	1	1	0.999	0.999
	LFME	0.97	1	0.952	0.999	1	0.999
	DEMON	1	0.952	1	0.999	0.999	1
ca-GrQc	SLPA	1	0.879	1	1	1	1
	LFME	0.879	1	0.93	1	1	1
	DEMON	1	0.93	1	1	1	1
High Energy Physics	SLPA	1	0.936	1	1	1	1
	LFME	0.936	1	0.976	1	1	1
	DEMON	1	0.976	1	1	1	1
Condense Matter	SLPA	1	0.952	1	1	1	1
	LFME	0.952	1	0.989	1	1	1
	DEMON	1	0.989	1	1	1	1
Enron Email	SLPA	1	0.958	1	1	1	1
	LFME	0.958	1	1	1	1	1
	DEMON	1	1	1	1	1	1

TABLE 16. Estimated parameters of the real-world networks for LFME and DEMON detection algorithms. on is the number of overlapping nodes. S is the size of the neighborhood of the overlapping nodes. A_n is the proportion of hubs in the neighborhood of the overlapping nodes while σ_A is its standard deviation. We note that each proportion value is the average of 10 simulation runs.

Network	LFME				DEMON			
	$on(\%)$	$S(\%)$	$A_n(\%)$	σ_A	$on(\%)$	$S(\%)$	$A_n(\%)$	σ_A
Zachary's karate club	8.82	14.7	80	5.56	14.7	71.17	88.04	5.99
Les Miserables	3.89	35.32	76.47	4.96	14.26	61.81	89.82	3.97
Dolphins	9.67	30.96	75	4.48	30.03	62.41	73.75	3.2
Game of thrones	7.47	42.61	73.68	1.83	16.93	53.82	71	4.74
American college football	6.07	40.34	75.86	4.54	16.52	58.43	73.8	5.4
ego-Facebook	5.78	32.26	71.83	9.14	9.16	43.95	74.67	8.41
ca-GrQc	15.23	30.13	64.01	7.27	25.27	49.18	84.82	7.77
ca-HepTh	11.75	27.56	65.17	8.58	28.15	56.36	84.26	5.07
ca-CondMat	9.37	28.9	63.02	6.22	40.6	67.6	71.01	8.77
Enron Email	19.14	30.38	70.42	2.27	35.23	69.8	73.77	7.27

TABLE 17. Rank-biased overlap r between the set of neighbors of the overlapping nodes and the set of hubs. This measure is computed using LFME and DEMON detection algorithms. p determines the weights of the elements. More weights is given to the first elements of both sets if p has a small value. Each proportion value is the average of 10 simulation runs. The standard deviation of RBO values are omitted in this table because the ranges are very small (close to 0).

Network	r									
	p=0.5		p=0.7		p=0.8		p=0.9		p=0.98	
	LFME	DEMON	LFME	DEMON	LFME	DEMON	LFME	DEMON	LFME	DEMON
Zachary's karate club	0.805	0.999	0.794	0.999	0.778	0.999	0.769	0.991	0.772	0.955
Les Miserables	0.997	1	0.986	0.999	0.969	0.999	0.931	0.999	0.852	0.990
Dolphins	0.998	0.999	0.986	0.999	0.966	0.999	0.934	0.999	0.914	0.998
Game of thrones	0.999	0.999	0.999	0.999	0.997	0.998	0.989	0.984	0.979	0.953
American college football	0.999	0.996	0.992	0.965	0.976	0.916	0.938	0.834	0.921	0.831
ego-Facebook	0.999	1	0.999	0.999	0.999	0.999	0.999	0.999	0.998	0.999
ca-GrQc	0.999	1	0.999	0.999	0.999	0.999	0.998	0.999	0.986	0.999
ca-HepTh	0.999	1	0.999	0.999	0.999	0.999	0.999	0.999	0.996	0.999
ca-CondMat	1	1	0.999	0.999	0.999	0.999	0.999	0.999	0.999	0.999
Enron Email	1	1	0.999	0.999	0.999	0.999	0.999	0.999	0.998	0.999

algorithms. We aim to measure how much the overlapping nodes are close to the hubs when different community detection algorithms are used. In small networks, as for SLPA,

the average and median distance of the Overlap-Hub network is slightly better than the one of the global network when LFME and DEMON are used. This is due to the high density

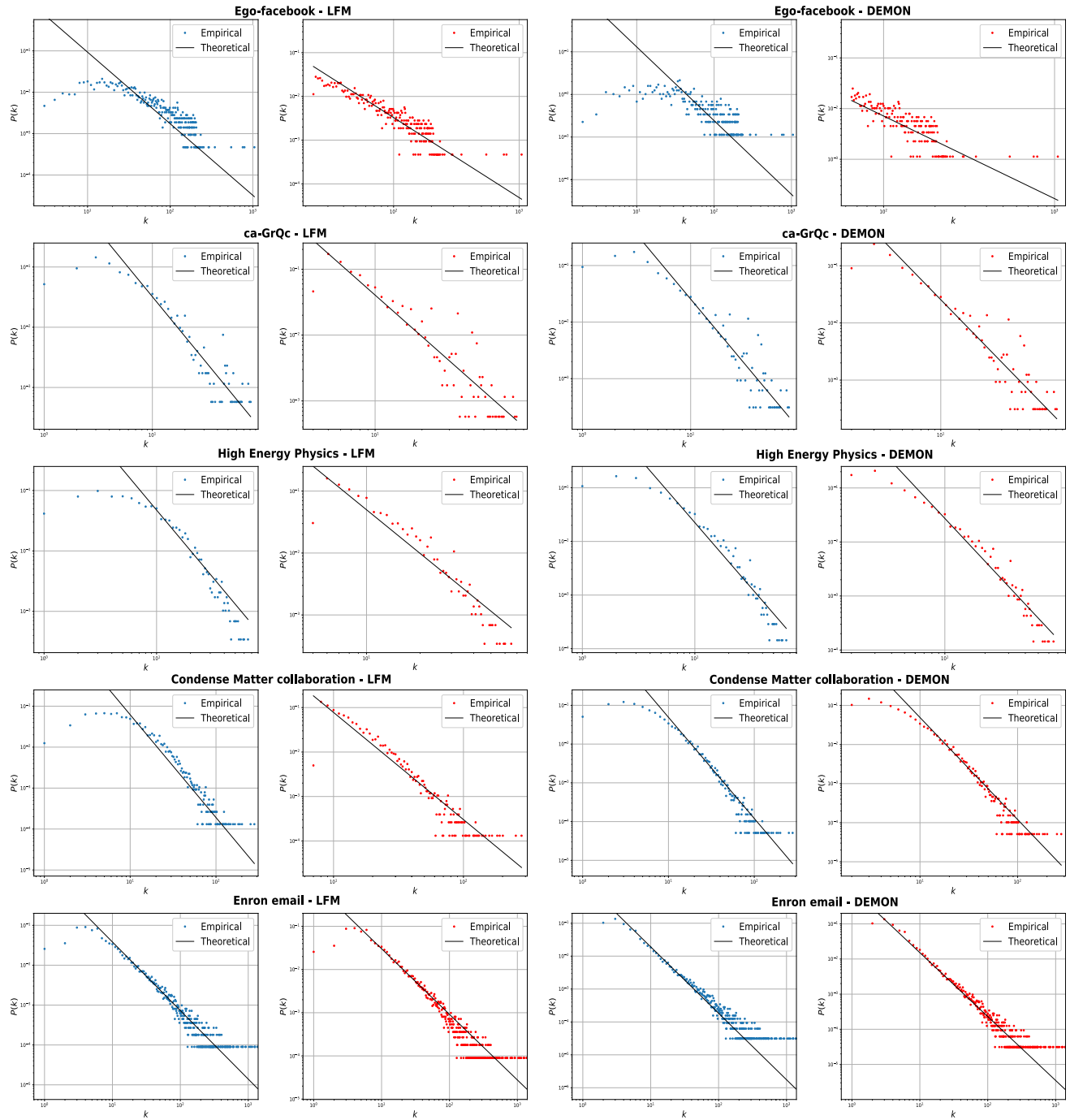


FIGURE 14. Log-log representation of degree distribution for real-wold networks under study using LFME and DEMON detection algorithms. The empirical degree distribution of neighbors of the overlapping nodes is in blue, it is in red for the Hubs. Power-Law estimates representation are in black.

of links in these types of networks. In medium and large networks, the average and the median distance between overlapping nodes and hubs are much smaller compared to those of the global network for both algorithms. However, the values of the average and median distance of the Overlap-Hub network get higher when DEMON algorithm is used. Indeed, this algorithm uncovers a large number of overlapping nodes

having a large neighborhood size. Thus, the Overlap-Hub network has also a larger size in this case. That explains why the average and the median distance of the Overlap-Hub network are higher for DEMON algorithm. Globally, the distance of the Overlap-Hub network is much smaller than the one of the global network whatever the community detection algorithm used. Therefore, the overlapping nodes

TABLE 18. Correlation between the set of neighbors of the overlapping nodes and the set of hubs for both LFME and DEMON detection algorithms. ρ stands for the average value of Pearson correlation while ρ_s stands for the average value of Spearman correlation. Each proportion value is the average of 10 simulation runs.

Network	LFME				DEMON			
	ρ	σ_ρ	ρ_s	σ_{ρ_s}	ρ	σ_ρ	ρ_s	σ_{ρ_s}
Zachary's karate club	0.919	0.0038	0.987	0.0065	0.997	0.0078	0.993	0.003
Les Miserables	0.979	0.0097	0.983	0.0051	0.99	0.0011	0.99	0.0079
Dolphins	0.972	0.0041	0.972	0.0084	0.987	0.0056	0.987	0.0021
Game of thrones	0.997	0.0077	0.996	0.0054	0.994	0.0096	0.989	0.0039
American college football	0.85	0.0096	0.86	0.001	0.824	0.0023	0.87	0.0087
ego-Facebook	0.99	0.0081	0.99	0.0011	0.99	0.0051	0.995	0.0069
ca-GrQc	0.982	0.0078	0.985	0.0098	0.998	0.0092	0.998	0.0019
ca-HepTh	0.996	0.0081	0.996	0.0031	0.998	0.0031	0.998	0.0047
ca-CondMat	0.997	0.0021	0.997	0.0052	0.99	0.009	0.999	0.0088
Enron Email	0.99	0.004	0.99	0.002	0.99	0.0065	0.997	0.0067

TABLE 19. α_1 and α_2 represent the estimated Power-Law exponent of the degree distribution for respectively the set of neighbors of the overlapping nodes and the set of hubs. Each proportion value is the average of 10 simulations for LFME and DEMON detection algorithms. The standard deviation values are omitted from this table because their values are close to 0.

Network	LFME		DEMON	
	α_1	α_2	α_1	α_2
ego-Facebook	1.91	2.27	1.97	2.05
ca-GrQc	2.12	2.06	2.27	2.23
ca-HepTh	2.29	2.38	2.41	2.69
ca-CondMat	2.54	2.41	2.66	2.64
Enron Email	2.35	2.23	2.64	2.53

and the hubs are tightly connected as compared to the rest of the network when the communities are uncovered by all the three community detection algorithms.

7) SUMMARY

To summarize, the community detection algorithms uncover similar communities in small networks. Indeed, they have a medium values of the NMI and close values of the mixing parameter and number of communities. In medium and large networks, their community structure can be quite different. They have small NMI values (relatively higher values between SLPA and LFME), and a different number of communities. Furthermore, we also use the overlapping modularity to measure the quality of the community structure revealed by the algorithms. The modularity values using the different algorithms are small with the lowest values for DEMON algorithm. Thus, SLPA is the most accurate detection algorithm followed by LFME, then DEMON algorithm. Moreover, we measure the overlap proportion between the sets of overlapping nodes defined by the three community detection algorithms. Results show that the overlap proportion between the algorithms is small. However, it is most of the time higher between SLPA and DEMON algorithms. The same experiment is performed between the sets of neighbors of the overlapping nodes. Results show there is generally a large intersection of these sets for all the community detection algorithms. Yet, there are smaller overlaps between LFME and both SLPA and DEMON algorithms. This is because LFME usually detects a smaller number of overlapping

TABLE 20. Proportion of hubs in the neighborhood of overlapping nodes as a function of their membership. om is the membership degree of the overlapping nodes. $on(\%)$ is the proportion of overlapping nodes. A_n is the proportion of hubs in the neighborhood of overlapping nodes. Each proportion value is the average of 10 simulations for LFME and DEMON detection algorithms. The standard deviation values are omitted in this table because of their small values.

Network	LFME			DEMON		
	om	$on(\%)$	$A_n(\%)$	om	$on(\%)$	$A_n(\%)$
Les Miserables	-	-	-	2	10.38	94.44
	-	-	-	3	2.59	100
	-	-	-	4	1.29	100
Dolphins	-	-	-	2	19.35	97.31
	-	-	-	3	8.06	100
	-	-	-	4	3.22	100
Game of thrones	-	-	-	2	15.88	93.58
	-	-	-	3	1.05	94.44
	-	-	-	3	1.73	88.88
American college football	2	4.34	56.34	2	14.79	86.74
	3	1.73	76.56	3	1.73	88.88
	3	0.29	57.28	-	-	-
ego-Facebook	3	5.49	67.99	-	-	-
	2	13.22	83.69	2	15.19	96.94
	4	6.53	97.53	4	6.53	97.53
ca-GrQc	3	1.25	87.5	5	1.67	98.78
	6	0.82	98.65	6	0.82	98.65
	9	0.34	100	9	0.34	100
ca-HepTh	2	11.07	82.36	2	19.29	97.56
	6	8.02	97.82	6	8.02	97.82
	8	0.53	98.15	8	0.53	98.15
ca-CondMat	3	0.65	92.27	10	0.27	99.12
	4	0.02	87.5	16	0.04	100
	2	8.96	79.18	2	21.62	98.83
ca-CondMat	4	13.66	99.23	4	13.66	99.23
	8	4.91	100	8	4.91	100
	12	0.34	100	12	0.34	100
Enron Email	4	0.04	100	18	0.06	100
	2	15.8	93.76	2	11.76	79.11
	6	8.64	82.09	6	8.64	82.09
Enron Email	3	2.92	98.19	12	7.26	100
	4	0.4	99.03	18	4.54	100
	5	0.02	100	23	3.03	100

nodes having then a smaller neighborhood size. Additionally, the most highly connected nodes of the network belong to the sets of neighbors of the overlapping nodes revealed by the three community detection algorithms.

Besides comparing the community detection algorithms, we also perform the same experiments using LFME and DEMON algorithms to check the robustness of the results obtained using the SLPA algorithm. We first compare the set of neighbors of the overlapping nodes and the set of hubs. The proportion of hubs, Rank-biased overlap, correlation and the degree distribution of both sets are analyzed. Experimental results show that all the measures display quite similar behavior as compared to the ones obtained based on SLPA. Generally, we get slightly better results when DEMON algorithm is used even though it is the less accurate community detection algorithm. Indeed, this algorithm detects a large amount of overlapping nodes. Thus, the ego-network of overlapping nodes has a big size for all the networks (it can reach 71% of the network). This is why the results are better when DEMON algorithm is used. However, we obtain close results when LFME and SLPA algorithms are employed. Secondly, we compare the Overlap-Hub network and the

TABLE 21. Proportion of hubs in the neighborhood of overlapping nodes as a function of the distance. The distance is the maximum number of links from an overlapping node to reach the perimeter of its neighborhood. A_n is the proportion of hubs in the neighborhood of overlapping nodes. A_n^o is the proportion of hubs belonging to the set of overlapping nodes. S represents the size of the neighborhood of the overlapping nodes. Each proportion value is the average of 10 simulation runs for LFME and DEMON detection algorithms. The standard deviation values are omitted in this table because of their small values.

Network	Algorithm	d=1			d=2			d=3			d=4			d=5			d=6		
		A_n (%)	A_n^o (%)	S (%)	A_n (%)	A_n^o (%)	S (%)	A_n (%)	A_n^o (%)	S (%)	A_n (%)	A_n^o (%)	S (%)	A_n (%)	A_n^o (%)	S (%)	A_n (%)	A_n^o (%)	S (%)
Karate club	LFME	80	0	14.7	96.77	3.22	87.17	96.875	3.125	94.11	-	-	-	-	-	-	-	-	-
	DEMON	88.04	2.5	73.52	96.34	3.66	82.35	-	-	-	-	-	-	-	-	-	-	-	-
Les Miserables	LFME	76.47	2.12	41.55	95.945	4.054	96.10	-	-	-	-	-	-	-	-	-	-	-	-
	DEMON	89.82	5.89	67.53	93.34	6.66	85.71	-	-	-	-	-	-	-	-	-	-	-	-
Dolphins	LFME	75	3.54	35.48	89.28	10.71	90.32	-	-	-	-	-	-	-	-	-	-	-	-
	DEMON	73.75	20	46.77	77.66	22.34	69.35	-	-	-	-	-	-	-	-	-	-	-	-
Game of thrones	LFME	73.68	4.11	50.46	91.83	8.16	91.58	-	-	-	-	-	-	-	-	-	-	-	-
	DEMON	71	4.92	50.46	89.77	10.23	83.17	-	-	-	-	-	-	-	-	-	-	-	-
College football	LFME	75.86	4.71	40.08	96.29	3.71	93.91	-	-	-	-	-	-	-	-	-	-	-	-
	DEMON	73.8	7.27	58.26	87.5	12.5	83.47	-	-	-	-	-	-	-	-	-	-	-	-
ego-Facebook	LFME	71.83	3.08	48.97	94.45	5.54	94.2	-	-	-	-	-	-	-	-	-	-	-	-
	DEMON	74.67	2.95	53.86	96.77	3.23	96.83	-	-	-	-	-	-	-	-	-	-	-	-
ca-GrQc	LFME	64.01	7.16	28.6	77.98	14.68	64.04	79.79	14.26	65.90	85.74	14.26	67.78	-	-	-	-	-	-
	DEMON	84.82	4.22	41.66	87.8	5.2	58.98	93.45	4.87	60	95.45	4.55	63.97	-	-	-	-	-	-
ca-HepTh	LFME	65.17	5.05	23.65	81.46	13.24	72.76	82.64	15.05	75.17	83.48	15.46	76	84.54	15.46	78.71	-	-	-
	DEMON	84.26	11.7	42.43	87.01	11.67	58.98	87.9	11.18	59.62	88.73	11.07	60.79	-	-	-	-	-	-
ca-CondMat	LFME	63.02	5.33	27.91	83.34	8.04	68.96	86.07	9.45	80.54	87.57	9.2	82.66	88.81	9.17	83	90.85	9.15	84.01
	DEMON	71.01	20.1	44.02	75.7	19.08	51.73	80.33	17.63	55.88	83.1	16.9	61.93	-	-	-	-	-	-
Enron Email	LFME	70.42	12.92	19.64	75.99	16.5	66.34	79	16.36	72.69	80.05	16.34	74.69	83.79	16.21	76.72	-	-	-
	DEMON	73.77	11.87	42.08	77.68	13.95	46.97	83.24	12.39	48.01	85.19	13.44	49.92	87.47	12.53	52.24	-	-	-

TABLE 22. Distance measures of the empirical networks. δ_G represents the diameter of the global network while δ is the diameter of the Overlap-Hub network. l_G represents the average distance of the global network while l is the average distance between the overlapping nodes and the hubs. l_{Gm} is the median value of the average distance of the global network while l_m is the median value of the average distance of the Overlap-Hub network. Each proportion value is the average of 10 simulation runs for LFME and DEMON detection algorithms. The standard deviation values are omitted in this table because the ranges are very small.

Network	Algorithm	δ_G	δ	l_G	l	l_{Gm}	l_m
Karate Club	LFME	5	2	2.408	1.5	2	2
	DEMON	5	3	2.641	1.686	3	2
Les Miserables	LFME	5	3	2.641	2.062	3	2
	DEMON	5	5	3.356	2.734	3	3
Dolphins	LFME	8	4	2.903	2.208	3	3
	DEMON	6	5	2.903	2.608	3	3
Game of thrones	LFME	4	3	2.508	2.315	3	3
	DEMON	4	4	2.508	2.957	3	3
ego-Facebook	LFME	8	5	3.692	2.848	5	2
	DEMON	8	6	3.692	3.037	5	4
ca-GrQc	LFME	17	9	6.049	2.102	6	3
	DEMON	17	11	6.049	3.599	6	4
ca-HepTh	LFME	18	10	5.945	2.152	6	3
	DEMON	18	12	5.945	3.365	6	4
ca-CondMat	LFME	15	12	5.352	2.119	6	3
	DEMON	15	14	5.352	3.266	6	4
Enron Email	LFME	11	8	4.025	2.011	6	3
	DEMON	11	10	4.025	3.083	6	5

global network by computing their average and median distance using LFME and DEMON algorithms. Results show that the distance of the Overlap-Hub network is always much smaller than the same measure computed on the global network. Yet, this difference gets slightly smaller when DEMON algorithm is used. That being said, both experiments show that the overlapping nodes are neighbors with a large proportion of hubs. Therefore, our hypothesis seems valid whatever the community detection algorithm used to uncover the community structure.

VI. CONCLUSION

The community structure is one of the main topological features of a vast majority of real-world networks. Unveiling their properties is of great interest in order to gain a better understanding of the structure and dynamics of complex networks. In this paper, we characterize the relationship between the overlapping nodes and the highly connected nodes of the networks (hubs). The overlapping nodes can belong to multiple communities. Thus, we believe that there is a high chance that they can be neighbors with the highly connected nodes in their respective communities. In order to characterize the relation between the hubs and the overlapping nodes, a series of experiments have been performed on a set of real-world networks of different sizes and origins. At first, we compare the set of neighbors of overlapping nodes and the set of hubs using some evaluation measures (the proportion of hubs, Rank-biased overlap and correlation). Extensive investigations show that there is a big overlap between these two sets. It appears that a high proportion of the hubs are

one-step neighbors of the overlapping nodes. This confirms the assumption that overlapping nodes are neighbors of the highly connected nodes of the network. Results show also that the Ranked-biased Overlap between the set of neighbors of the overlapping nodes and the set of hubs has very high values. It confirms that there is a great similarity between the two sets. Additionally, there is a very high correlation between the set of neighbors of the overlapping nodes and the hubs. Moreover, the degree distribution analysis of both sets shows that they display comparable empirical distribution. They exhibit a power-law degree distribution with very close exponent values. Secondly, the global network has been compared with its sub-network formed by the overlapping nodes and the hubs (the so-called Overlap-Hub network). The average and median shortest paths of both networks have been compared in order to examine how overlapping nodes are close to the hubs. Experimental results show that the distance between overlapping nodes and hubs is much smaller as compared to the overall network.

Furthermore, the influence of the neighborhood size of the overlapping nodes on the proportion of hubs has been investigated. Results show that using a n -steps neighborhood allows finding a larger proportion of hubs in this neighborhood. In addition, we also look at the influence of the membership degree of overlapping nodes on the proportion of hubs in their neighborhood. Results show that if overlapping nodes belong to a higher number of communities, the proportion of hubs in their neighborhood becomes higher. We also performed the same experiments using LFME and DEMON detection algorithms to uncover the communities. Results of the investigations show that the conclusions are quite comparable. Even if the community structures are not the same, the evaluation measures of the proximity of overlapping nodes and hubs have overall close values. This is quite valid for all the networks under test. This confirms that the overlapping nodes are neighbors with a large amount of hubs.

Results of our investigations are relevant in multiple settings. Indeed, our analysis sheds light on the organization of complex networks and provides new directions for research on community detection. This work can also help to elaborate new strategies to target the most influential nodes in modular networks.

REFERENCES

- [1] P. De Meo, E. Ferrara, G. Fiumara, and A. Provetti, "On facebook, most ties are weak," 2012, *arXiv:1203.0535*. [Online]. Available: <http://arxiv.org/abs/1203.0535>
- [2] H. Cherifi, G. Palla, B. K. Szymanski, and X. Lu, "On community structure in complex networks: Challenges and opportunities," *Appl. Netw. Sci.*, vol. 4, no. 1, pp. 1–35, Dec. 2019.
- [3] S. Fortunato and D. Hric, "Community detection in networks: A user guide," *Phys. Rep.*, vol. 659, pp. 1–44, Nov. 2016.
- [4] M. Jebabli, H. Cherifi, C. Cherifi, and A. Hamouda, "User and group networks on YouTube: A comparative analysis," in *Proc. IEEE/ACS 12th Int. Conf. Comput. Syst. Appl. (AICCSA)*, Nov. 2015, pp. 1–8.
- [5] M. Jebabli, H. Cherifi, C. Cherifi, and A. Hammouda, "Overlapping community structure in co-authorship networks: A case study," in *Proc. 7th Int. Conf. u-e-Service, Sci. Technol.*, Dec. 2014.
- [6] M. Jebabli, H. Cherifi, C. Cherifi, and A. Hamouda, "Community detection algorithm evaluation with ground-truth data," *Phys. A, Stat. Mech. Appl.*, vol. 492, pp. 651–706, Feb. 2018.
- [7] J. Xie, S. Kelley, and B. K. Szymanski, "Overlapping community detection in networks: The state-of-the-art and comparative study," *ACM Comput. Surv.*, vol. 45, no. 4, p. 43, 2013.
- [8] G. K. Orman, V. Labatut, and H. Cherifi, "An empirical study of the relation between community structure and transitivity," 2012, *arXiv:1207.3234*. [Online]. Available: <https://arxiv.org/abs/1207.3234>
- [9] G. K. Orman, V. Labatut, and H. Cherifi, "On accuracy of community structure discovery algorithms," *J. Conver. Inf. Technol.*, vol. 6, no. 11, pp. 283–292, Nov. 2011.
- [10] V. Labatut and J.-M. Balasque, "Detection and interpretation of communities in complex networks: Practical methods and application," in *Computational Social Networks*. London, U.K.: Springer, 2012, pp. 81–113, doi: 10.1007/978-1-4471-4048-1_4.
- [11] G. K. Orman, V. Labatut, and H. Cherifi, "Comparative evaluation of community detection algorithms: A topological approach," *J. Stat. Mech., Theory Exp.*, vol. 2012, no. 08, 0, Art. no. P08001.
- [12] G. K. Orman, V. Labatut, and H. Cherifi, "Towards realistic artificial benchmark for community detection algorithms evaluation," *Int. J. Web Based Communities*, vol. 9, no. 3, pp. 349–370, 2013.
- [13] N. Lu, W. Luo, L. Ni, H. Jiang, and W. Ding, "Extending CDFR for overlapping community detection," in *Proc. 1st Int. Conf. Data Intell. Secur. (ICDIS)*, Apr. 2018, pp. 200–206.
- [14] L. Ni, W. Luo, W. Zhu, and B. Hua, "Local overlapping community detection," *ACM Trans. Knowl. Discovery Data*, vol. 14, no. 1, pp. 1–25, Feb. 2020.
- [15] G. Palla, I. Derényi, I. Farkas, and T. Vicsek, "Uncovering the overlapping community structure of complex networks in nature and society," *Nature*, vol. 435, no. 7043, pp. 814–818, Jun. 2005.
- [16] M. Kudelka, E. Ochodkova, S. Zehnalova, and J. Plesnik, "Ego-zones: Non-symmetric dependencies reveal network groups with large and dense overlaps," *Appl. Netw. Sci.*, vol. 4, no. 1, p. 81, Dec. 2019.
- [17] Y. Li, K. He, K. Kloster, D. Bindel, and J. Hopcroft, "Local spectral clustering for overlapping community detection," *ACM Trans. Knowl. Discovery Data*, vol. 12, no. 2, pp. 1–27, Mar. 2018.
- [18] M. Huang, G. Zou, B. Zhang, Y. Liu, Y. Gu, and K. Jiang, "Overlapping community detection in heterogeneous social networks via the user model," *Inf. Sci.*, vol. 432, pp. 164–184, Mar. 2018.
- [19] G. Bello-Organ, S. Salcedo-Sanz, and D. Camacho, "A multi-objective genetic algorithm for overlapping community detection based on edge encoding," *Inf. Sci.*, vol. 462, pp. 290–314, Sep. 2018.
- [20] J. J. Whang, D. F. Gleich, and I. S. Dhillon, "Overlapping community detection using neighborhood-inflated seed expansion," *IEEE Trans. Knowl. Data Eng.*, vol. 28, no. 5, pp. 1272–1284, May 2016.
- [21] T. Ma, Y. Wang, M. Tang, J. Cao, Y. Tian, A. Al-Dhelaan, and M. Al-Rodhaan, "LED: A fast overlapping communities detection algorithm based on structural clustering," *Neurocomputing*, vol. 207, pp. 488–500, Sep. 2016.
- [22] X. Wen, W.-N. Chen, Y. Lin, T. Gu, H. Zhang, Y. Li, Y. Yin, and J. Zhang, "A maximal clique based multiobjective evolutionary algorithm for overlapping community detection," *IEEE Trans. Evol. Comput.*, vol. 21, no. 3, pp. 363–377, Jun. 2017.
- [23] J. Yang and J. Leskovec, "Structure and overlaps of communities in networks," 2012, *arXiv:1205.6228*. [Online]. Available: <http://arxiv.org/abs/1205.6228>
- [24] M. Kudelka, E. Ochodkova, S. Zehnalova, and J. Plesnik, "Ego-zones: Non-symmetric dependencies reveal network groups with large and dense overlaps," *Appl. Netw. Sci.*, vol. 4, no. 1, p. 81, Dec. 2019, doi: 10.1007/s41109-019-0192-6.
- [25] L. Hébert-Dufresne, A. Allard, J.-G. Young, and L. J. Dubé, "Global efficiency of local immunization on complex networks," *Sci. Rep.*, vol. 3, no. 1, p. 2171, Dec. 2013.
- [26] F. Taghavian, M. Salehi, and M. Teimouri, "A local immunization strategy for networks with overlapping community structure," *Phys. A, Stat. Mech. Appl.*, vol. 467, pp. 148–156, Feb. 2017.
- [27] M. Kumar, A. Singh, and H. Cherifi, "An efficient immunization strategy using overlapping nodes and its neighborhoods," in *Proc. Companion The Web Conf. Web Conf. (WWW)*, 2018, pp. 1269–1275.
- [28] H. Wei, Z. Pan, G. Hu, L. Zhang, H. Yang, X. Li, and X. Zhou, "Identifying influential nodes based on network representation learning in complex networks," *PLoS ONE*, vol. 13, no. 7, Art. no. e0200091.

- [29] Z. Ghalmane, C. Cherifi, H. Cherifi, and M. E. Hassouni, "Centrality in complex networks with overlapping community structure," *Sci. Rep.*, vol. 9, no. 1, pp. 1–29, Dec. 2019.
- [30] W. Webber, A. Moffat, and J. Zobel, "A similarity measure for indefinite rankings," *ACM Trans. Inf. Syst.*, vol. 28, no. 4, pp. 1–38, Nov. 2010.
- [31] A. Clauset, C. R. Shalizi, and M. E. J. Newman, "Power-law distributions in empirical data," *SIAM Rev.*, vol. 51, no. 4, pp. 661–703, Nov. 2009.
- [32] M. E. Newman, "The structure and function of complex networks," *SIAM Rev.*, vol. 45, no. 2, pp. 167–256, 2003.
- [33] S. Boccaletti, V. Latora, Y. Moreno, M. Chavez, and D.-U. Hwang, "Complex networks: Structure and dynamics," *Phys. Rep.*, vol. 424, nos. 4–5, pp. 175–308, 2006.
- [34] R. A. Rossi and N. K. Ahmed, "The network data repository with interactive graph analytics and visualization," in *Proc. 29th AAAI Conf. Artif. Intell.*, 2015, pp. 4292–4293.
- [35] J. Xie, B. K. Szymanski, and X. Liu, "SLPA: Uncovering overlapping communities in social networks via a speaker-listener interaction dynamic process," in *Proc. IEEE 11th Int. Conf. Data Mining Workshops*, Dec. 2011, pp. 344–349.
- [36] H. Cherifi, "Chapter eleven. Non-overlapping community detection," in *Complex Networks and Their Applications*. 2014, p. 320.
- [37] N. Gupta, A. Singh, and H. Cherifi, "Community-based immunization strategies for epidemic control," in *Proc. 7th Int. Conf. Commun. Syst. Netw. (COMSNETS)*, Jan. 2015, pp. 1–6.
- [38] D. Chakraborty, A. Singh, and H. Cherifi, "Immunization strategies based on the overlapping nodes in networks with community structure," in *Proc. Int. Conf. Comput. Social Netw.* London, U.K.: Springer, 2016, pp. 62–73, doi: 10.1007/978-3-319-42345-6_6.
- [39] Z. Ghalmane, M. El Hassouni, C. Cherifi, and H. Cherifi, "K-truss decomposition for modular centrality," in *Proc. 9th Int. Symp. Signal, Image, Video Commun. (ISIVC)*, Nov. 2018, pp. 241–248.
- [40] Z. Ghalmane, M. El Hassouni, C. Cherifi, and H. Cherifi, "Centrality in modular networks," *EPJ Data Sci.*, vol. 8, no. 1, p. 15, Dec. 2019.
- [41] N. Gupta, A. Singh, and H. Cherifi, "Centrality measures for networks with community structure," *Phys. A, Stat. Mech. Appl.*, vol. 452, pp. 46–59, Jun. 2016.
- [42] Z. Ghalmane, M. E. Hassouni, and H. Cherifi, "Betweenness centrality for networks with non-overlapping community structure," in *Proc. IEEE Workshop Complex. Eng. (COMPENG)*, Oct. 2018, pp. 1–5.
- [43] M. Kitromilidis and T. S. Evans, "Community detection with metadata in a network of biographies of western art painters," 2018, *arXiv:1802.07985*. [Online]. Available: <http://arxiv.org/abs/1802.07985>
- [44] Z. Ghalmane, M. E. Hassouni, and H. Cherifi, "Immunization of networks with non-overlapping community structure," *Social Netw. Anal. Mining*, vol. 9, no. 1, p. 45, Dec. 2019.
- [45] M. Wang, S. Yang, and L. Wu, "Improved community mining method based on LFM and EAGLE," *Comput. Sci. Inf. Syst.*, vol. 13, no. 2, pp. 515–530, 2016.
- [46] M. Coscia, G. Rossetti, F. Giannotti, and D. Pedreschi, "DEMON: A local-first discovery method for overlapping communities," in *Proc. 18th ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining*, 2012, pp. 615–623.
- [47] A. Lancichinetti, S. Fortunato, and J. Kertész, "Detecting the overlapping and hierarchical community structure in complex networks," *New J. Phys.*, vol. 11, no. 3, 2009, Art. no. 033015.
- [48] H. Shen, X. Cheng, K. Cai, and M.-B. Hu, "Detect overlapping and hierarchical community structure in networks," *Phys. A, Stat. Mech. Appl.*, vol. 388, no. 8, pp. 1706–1712, Apr. 2009.
- [49] P. Holme, "Rare and everywhere: Perspectives on scale-free networks," *Nature Commun.*, vol. 10, no. 1, p. 1016, Dec. 2019.
- [50] L. Danon, A. Díaz-Guilera, J. Duch, and A. Arenas, "Comparing community structure identification," *J. Stat. Mech., Theory Exp.*, vol. 2005, no. 09, 2005, Art. no. P09008.



networks, community structure, graph theory, and social computing.



member of GDR-MACS, IMU, IXXI, and a member of the Steering Committee of the International Conference of COMPLEX NETWORKS.



since 1999. He has published more than 200 scientific articles in international refereed journals and conference proceedings. His recent research interests include computer vision and complex networks. He held leading positions in more than 15 international conference organization (the General Chair and the Program Chair). He has served on more than 100 Program Committee. He is the Founder of the International Conference on Complex Networks and Their Applications. He is an Editorial Board Member of the *Computational Social Networks* (Springer) and *PLOS One*, and the Editor-In-Chief of the *Journal of Applied Network Science* (Springer).



networks, quality assessment, and mesh processing, especially on medical imaging, biometry, and QoE applications. He has published more than 150 scientific articles in international refereed journals and conference proceedings.

...