Identifying Influential Nodes Using Overlapping Modularity Vitality

Stephany Rajeh
University of Burgundy
Dijon, France
Email: stephany.rajeh
@u-bourgogne.fr

Marinette Savonnet
University of Burgundy
Dijon, France
Email: marinette.savonnet
@u-bourgogne.fr

Eric Leclercq
University of Burgundy
Dijon, France
Email: eric.leclercq
@u-bourgogne.fr

Hocine Cherifi
University of Burgundy
Dijon, France
Email: hocine.cherifi
@u-bourgogne.fr

Abstract—It is of paramount importance to uncover influential nodes to control diffusion phenomena in a network. In recent works, there is a growing trend to investigate the role of the community structure to solve this issue. Up to now, the vast majority of the so-called community-aware centrality measures rely on non-overlapping community structure. However, in many real-world networks, such as social networks, the communities overlap. In other words, a node can belong to multiple communities. To overcome this drawback, we propose and investigate the "Overlapping Modularity Vitality" centrality measure. This extension of "Modularity Vitality" quantifies the community structure strength variation when removing a node. It allows identifying a node as a hub or a bridge based on its contribution to the overlapping modularity of a network. A comparative analysis with its non-overlapping version using the Susceptible-Infected-Recovered (SIR) epidemic diffusion model has been performed on a set of six real-world networks. Overall, Overlapping Modularity Vitality outperforms its alternative. These results illustrate the importance of incorporating knowledge about the overlapping community structure to identify influential nodes effectively. Moreover, one can use multiple ranking strategies as the two measures are signed. Results show that selecting the nodes with the top positive or the top absolute centrality values is more effective than choosing the ones with the maximum negative values to spread the epidemic.

Index Terms—Complex Networks, Centrality, Overlapping Community Structure, Influential Nodes, SIR model

I. INTRODUCTION

Amidst the decisiveness of curbing the spread of epidemics, improving marketing and awareness campaigns, or maintaining network connectivity, identifying influential nodes in the dynamics of networks is a challenging issue. Centrality measures are one of the main approaches to rank nodes according to their importance in the network. Classical centrality measures rely on local or global information on the network topology [1]. For example, Degree centrality measures the number of links of a node. It relies only on the knowledge of the node's neighborhood. In contrast, Betweenness centrality uses information of the whole network. Indeed, it measures the frequency of a node lying in the shortest paths between all other nodes. Generally, local measures are computationally efficient but less accurate, while it is the opposite for global measures. These two types of measures are also combined to get the best of both worlds [2], [3].

The modular organization of real-world networks is one of the essential topological features. They are formed with modules of densely connected nodes that have rare connections with the other modules. Nonetheless, nodes can belong to several modules. Indeed, many networks, such as social, collaboration, biological, and infrastructural networks, are characterized by an overlapping community structure [4]–[7]. In these networks, the overlapping nodes may have multiple functions. For example, a person may adhere to several organizations. A scientist may conduct interdisciplinary research with various scientific groups. A protein may have several functions in different complexes. Also, an airport may have internal and external flights. Indeed, an overlapping community structure is frequent in many real-world situations.

Even though real-world networks are generally modular, classical centrality measures discard this important feature. Most of the centrality measures proposed in the literature are agnostic about the community structure, such as Degree centrality, Betweenness centrality, and k-core [8]–[11]. A new emergent research area exploits the community structure to identify influential nodes [12]–[19]. They quantify the local and global influence of a node through its intra-community links and inter-community links [20]. The former connects nodes in the same community, while the latter connects nodes in different communities. Nonetheless, these measures consider that a node belongs to a unique community. For example, Community-based Mediator [15] targets influential nodes based on the entropy of their intra-community links and inter-community links. There are no overlapping nodes. Indeed, if a node is overlapping, its centrality measure needs to be evaluated depending on its multiple communities. One can build various strategies accordingly. Another example is Comm Centrality [16]. It targets hubs and bridges simultaneously but gives more weight to bridges. Here too, a node belongs to one and only one community. Few works

have shown that incorporating information on the overlapping community structure results in identifying influential nodes more effectively [21]–[23].

A recent work investigates the behavior of a set of non-overlapping community-aware centrality measures. Results show that Modularity Vitality [12] outperforms its alternatives. This signed community-aware centrality measure quantifies the modularity variation when one removes a node in the network. It can pinpoint bridges or hubs depending on the ranking scheme. Indeed, eliminating hubs in a community decreases modularity. In contrast, removing bridges between communities increases the modularity. As it has proved to be quite effective, we propose to extend it to networks with an overlapping community structure. The SIR infectious spreading model is used to evaluate the performance of "Overlapping Modularity Vitality" compared to its non-overlapping version. Simulations are performed on six real-world networks originating from three domains: infrastructural networks, collaboration networks, and online social networks. Results demonstrate that Overlapping Modularity Vitality outperforms its non-overlapping counterpart. Moreover, when considering the absolute value of both measures (i.e., targeting hubs and bridges simultaneously), the difference in the performance of the two measures version is even higher.

The paper is organized as follows. The non-overlapping and overlapping versions of Modularity Vitality are given in section II. The data and the tools used in the evaluation process are presented in section III. In section IV, experimental results are given. A discussion is developed in section V. Finally, section VI concludes the article.

II. Non-overlapping and Overlapping Modularity Vitality

In this section, we present the definitions of non-overlapping Modularity Vitality based on Newman's modularity and its overlapping version. First, suppose that $G(V, E)$ is an undirected and unweighted graph where V is the set of nodes of size $N = |V|$ and $|E|$ is the total number of edges. The connections between the nodes are described in the graph's adjacency matrix A. The graph G is partitioned into $C = \{c_1, c_2, ..., c_k, ...c_{|C|}\}$ communities where c_k is k-th community and $|C|$ is the total number of communities. In a network with a non-overlapping community structure, node v_i can belong to one and only one community. In a network with an overlapping community structure, node v_i can belong to several communities. That being said, its strength of belonging differs from one community to another. Hence, each node v_i is characterized by a belonging coefficient vector $(\alpha_{v_i,c_1}, \alpha_{v_i,c_2}, ..., \alpha_{v_i,c_k}, ..., \alpha_{v_i,c_{|C|}})$ and assuming the following conditions hold: $0 \leq a_{v_i,c_k} \leq 1; \forall v_i \in V; \forall c_k \in C$ and $\sum_{c_k \in C} a_{v_i,c_k} = 1$.

A. Non-overlapping Modularity Vitality

Modularity Vitality [12] is based on Newman's modularity [24]. A lot of community detection algorithms use this property as an objective function [25]–[28]. It measures the difference between the actual connections within a community to the connections with wiring at random. The larger the value of modularity, the stronger the community structure. Hence, developing a community-aware centrality measure that is directly related to the community structure strength is ideal. Newman's modularity is defined as follows:

$$Q(G) = \sum_{c_k \in C} \left[\frac{|E_{c_k}^{in}|}{|E|} - \left(\frac{2|E_{c_k}^{in}| + |E_{c_k}^{out}|}{2|E|} \right)^2 \right] \quad (1)$$

where:
- $|E_{c_k}^{in}| = \frac{1}{2} \sum_{v_i, v_j \in c_k} A_{v_i, v_j}$, the intra-community edges of c_k
- $|E_c^{out}| = \sum_{v_i \in c_k} \sum_{v_j \in C - c_k} A_{v_i, v_j}$, the inter-community edges of c_k

Using the above definition of modularity, Modularity Vitality can distinguish hubs from bridges based on their contribution to the overall modularity of a network. Indeed, hubs increase the modularity while bridges decrease it. It is defined as follows:

$$\alpha_{MV}(v_i) = Q(G) - Q(G \setminus \{v_i\}) \quad (2)$$

where:
- $Q(G)$ is the network's modularity based on Newman
- $Q(G \setminus \{v_i\})$ is the network's modularity based on Newman after the removal of node v_i

Since Modularity Vitality is signed, we can use three ranking strategies. The first ranking strategy ranks nodes in decreasing order from the highest positive to the lowest negative value. This strategy targets hubs first. The second orders nodes from negative to positive centrality values. Consequently, it targets bridges first. Finally, nodes are ordered according to their absolute centrality value. It allows targeting hubs and bridges simultaneously based on their contribution to the network's overall modularity.

B. Overlapping Modularity Vitality

Several extensions to Newman's modularity for overlapping community structures have been proposed. One can divide them into two groups: fuzzy modularity and non-fuzzy modularity [29], [30]. In the former, a node belongs to several communities based on a belonging coefficient. In the latter, the relationship is binary. This paper uses a simple fuzzy version of modularity to quantify the quality of the overlapping community structure. It is defined as follows:

$$a_{v_i, c_k} = \frac{1}{O_{v_i}} \quad (3)$$

where O_{v_i} is the number of communities shared by node v_i. In other words, if a node belongs to one community, its belonging coefficient is 1. Node shared by two communities, gets a belonging coefficient of 0.5 for each of the communities, and so on.

Based on the belonging coefficient vector for each node v_i, this information can be incorporated within $|E_{c_k}^{in}|$ and $|E_{c_k}^{out}|$, where:

$$|E_{c_k}^{in}| = \frac{1}{2} \sum_{v_i, v_j \in c_k} \frac{a_{v_i, c_k} + a_{v_j, c_k}}{2} A_{v_i, v_j} \quad (4)$$

$$|E_{c_k}^{out}| = \sum_{v_i \in c_k} \sum_{v_j \in C - c_k} \frac{a_{v_i, c_k} + (1 - a_{v_j, c_k})}{2} A_{v_i, v_j} \quad (5)$$

Hence, the main difference between the overlapping version and the non-overlapping version is that the adjacency matrix (A_{v_i, v_j}) contains the weights of the average of belonging coefficients of the nodes. Indeed, it has been shown that the definition of modularity holds for a non-binary adjacency matrix [31]. The Overlapping Modularity Vitality is defined as follows:

$$\alpha_{OMV}(v_i) = Q_o(G) - Q_o(G \setminus \{v_i\}) \quad (6)$$

where:

- $Q_o(G)$ is the network's overlapping modularity
- $Q_o(G \setminus \{v_i\})$ is the network's overlapping modularity after the removal of node v_i

Similar to its non-overlapping version, Overlapping Modularity Vitality is also signed. Consequently, we also consider three ranking strategies as defined for its non-overlapping counterpart.

III. DATA AND TOOLS

This section presents the real-world networks used, the community detection algorithms, as well as the SIR evaluation process, and the subsequent evaluation criterion to compare the community-aware centrality measures.

A. Data

We perform a comparative evaluation on six real-world networks originating from three domains: infrastructural, collaboration, and online social networks. Table I reports their basic topological characteristics.

1) Infrastructural Networks: We consider two air transportation networks, EU Airlines [32], and U.S. Airports [33]. In these networks, the nodes represent European and U.S. airports, respectively. The nodes are connected if there's a direct flight between them.

2) Collaboration Networks: The AstroPh [34] network is extracted from the e-print arXiv. Nodes represent authors who have submitted their papers to the Astrophysics category. The nodes are connected if two people have co-authored a paper. In New Zealand Collaboration [32], the nodes are scientific institutions in New Zealand such as universities and organizations. If Scopus lists a minimum of one publication with authors in any two institutions, the nodes are connected.

3) Online Social Networks: Hamsterster [33] is an online social pet platform. Nodes represent users of the platform, and connections represent online friendships. In the DNC Emails [32], nodes represent members of the Democrat National

Committee. If the members have exchanged emails, they are connected.

B. Community Detection Algorithms

As the network's community structure is unknown, one uses a community detection algorithm to uncover it. Infomap [35] reveals non-overlapping communities while Speaker-Listener Label Propagation Algorithm (SLPA) [36] unveils overlapping communities. These influential algorithms have proved their effectiveness in estimating the mesoscopic scale [37]–[39].

1) Infomap: extracts a non-overlapping community structure by minimizing the description of a random walk. Densely connected modules that are sparsely connected characterize many real-world networks. Therefore, it is likely for a random walker to stay longer within the modules than jumping from one module to another. One uses Huffman coding with a prefix code and a suffix code. The prefix code refers to the module. The suffix code is assigned to a node within the modules. The compression of the description of the random walker unveils the community structure.

2) SLPA: extracts overlapping communities. It is inspired by how humans acquire and spread opinions. As a start, each node has a unique label. Then, each node acts as a listener and selects one of the labels sent to it from its neighbors (i.e., speakers). The listener updates its memory based on the label frequency it listens to. When all the nodes are visited, the label frequency in their memory is converted to a probability distribution. This distribution characterizes the membership degree for each of the communities. The distribution of the membership of each node is further processed to extract the communities. A threshold r allows discarding the membership to a specific community if it is below the threshold. In the experiments, we set its value to 0.01.

C. Susceptible-Infected-Removed Model

One uses the epidemiological model Susceptible-Infected-Recovered (SIR) [40] to assess the effectiveness of the community-aware centrality measures. In this model, nodes can be in either of these states: Susceptible (S), Infected (I), or Recovered (R). Initially, all nodes are susceptible except for a given proportion (f_o) of the top nodes ranked according to a specific centrality measure. These nodes are infected. Each infected node can infect its susceptible neighbor with a probability λ. Simultaneously, each infected node can recover with a probability γ. The propagation proceeds until all nodes are either recovered or still in the susceptible state. At this point, one computes the outbreak size, which is the number of nodes in the recovered state (R). This outbreak size dictates the spreading effectiveness of a specific centrality measure for each f_o considered. Centrality measures aim to maximize this value. In the experiments, 100 SIR simulations are performed and averaged for each fraction of initially infected nodes (f_o) for the community-aware centrality measures under study.

TABLE I

BASIC TOPOLOGICAL PROPERTIES OF THE NETWORKS. N IS THE NUMBER OF NODES. $|E|$ IS THE NUMBER OF EDGES. $<k>$ IS THE AVERAGE DEGREE. ζ IS THE TRANSITIVITY. Q IS THE NON-OVERLAPPING MODULARITY. Q_o IS THE OVERLAPPING MODULARITY. $on(\%)$ IS THE FRACTION OF OVERLAPPING NODES. m IS THE AVERAGE NUMBER OF COMMUNITY MEMBERSHIPS OF THE NODES. THE OVERLAPPING PROPERTIES ARE BASED ON THE COMMUNITY STRUCTURE UNCOVERED BY SLPA. * MEANS THE LARGEST CONNECTED COMPONENT OF THE NETWORK IS TAKEN IF IT IS DISCONNECTED.

Network	N	$|E|$	$<k>$	ζ	Q	Q_o	$on(\%)$	m
EU Airlines	417	2,953	14.16	0.304	0.109	0.741	0.062	2.154
U.S. Airports	500	2,980	11.92	0.351	0.161	0.731	0.118	2.152
DNC Emails*	849	10,384	24.46	0.548	0.416	0.593	0.285	2.004
New Zealand*	1,463	4,246	5.80	0.063	0.401	0.524	0.364	2.163
Hamsterster*	1,788	12,476	13.49	0.090	0.391	0.648	0.251	2.247
AstroPh*	17,903	196,972	22.00	0.317	0.563	0.208	0.569	2.669

D. Evaluation Measure

We use the Degree centrality as a baseline to compare the spreading outbreak size of community-aware centrality measures. The relative difference is defined as:

$$\Delta R = \frac{R_c - R_b}{R_b} \tag{7}$$

where:

- R_c denotes the outbreak size using a specific community-aware centrality measure c
- R_b represents the outbreak size using the baseline Degree centrality

If the community-aware centrality measure is more effective than the baseline, ΔR is positive. Otherwise, ΔR is negative.

IV. EXPERIMENTAL RESULTS

This section reports the comparative evaluation of the effectiveness of the non-overlapping Modularity Vitality and its overlapping version.

A. Targeting Hubs

First, we examine how Modularity Vitality in its non-overlapping and overlapping versions compare when the fraction of initially infected nodes (f_o) is chosen based on the ranks ordered from the highest positive centrality values. In other words, initially infected nodes are local hubs as they contribute positively to modularity. The left column of figure 1 shows the relative difference of the outbreak size (ΔR) as a function of the fraction of initially infected nodes (f_o) where the nodes are ranked based on the top positives (hubs-first). One observes three typical behaviors.

The first case is illustrated by EU Airlines, U.S. Airports, Hamsterster, and AstroPh. In these networks, Overlapping Modularity Vitality (α_{OMV}) outperforms its non-overlapping version (α_{MV}). However, the outperformance of α_{OMV} starts after infecting 9% of the nodes in EU Airlines and AstroPh and after 5% in U.S. Airports and Hamsterster. This behavior demonstrates that infecting local hubs is more beneficial than infecting hubs located in the overlap when resources are limited. Above this fraction, the outperformance of α_{OMV} compared to α_{MV} can reach up to 5%. In this situation,

hubs in the overlap matter. The second typical behavior is illustrated by DNC Emails. In this network, the curves of the relative difference of the outbreak size (ΔR) for both α_{OMV} and α_{MV} essentially intersect when $f_o > 0.10$. In other words, α_{OMV} and α_{MV} perform similarly. Finally, the third behavior is illustrated by the New Zealand Collaboration network. In this network, the non-overlapping Modularity Vitality (α_{MV}) outperforms its overlapping version (α_{OMV}) when $f_o > 0.26$. Below this value, the curves mainly coincide.

B. Targeting Bridges

The performances of the non-overlapping Modularity Vitality (α_{MV}) and its overlapping version (α_{OMV}) are investigated on the SIR model when the fraction of initially infected nodes (f_o) contains the nodes ordered from the highest negative centrality values. In other words, the initially infected nodes are bridges since they contribute negatively to modularity. The middle column of figure 1 shows the relative difference of the outbreak size (ΔR) as a function of the fraction of initially infected nodes (f_o) where the nodes are ranked based on the top negatives (bridges-first). Once again, one can distinguish three behaviors.

EU Airlines and U.S. Airports illustrate the first behavior. In these networks, the curves of the relative difference of the outbreak size (ΔR) for both α_{OMV} and α_{MV} essentially intersect at all fractions of initially infected nodes. Moreover, they don't exhibit a higher performance compared to the baseline. They even perform worse than the baseline in the U.S. Airports network after infecting 20% of the nodes. Indeed, ΔR is negative in this case. Hamsterster and AstroPh illustrate the second behavior. In these two networks, α_{MV} results in a higher epidemic outbreak than its overlapping version when $f_o > 0.27$ for Hamsterster and when $f_o > 0.24$ for AstroPh. Its outperformance can reach up to 4% when compared to α_{OMV}. When f_o is less than these two values, the curves generally coincide. Finally, the third behavior is illustrated by the networks DNC Emails and New Zealand Collaboration. Here, the Overlapping Modularity Vitality shows a slightly better performance than its non-overlapping version. However, similar to Hamsterster and AstroPh, when f_o is less than 0.16 in DNC Emails, both measures share a similar relative difference of the epidemic outbreak size.
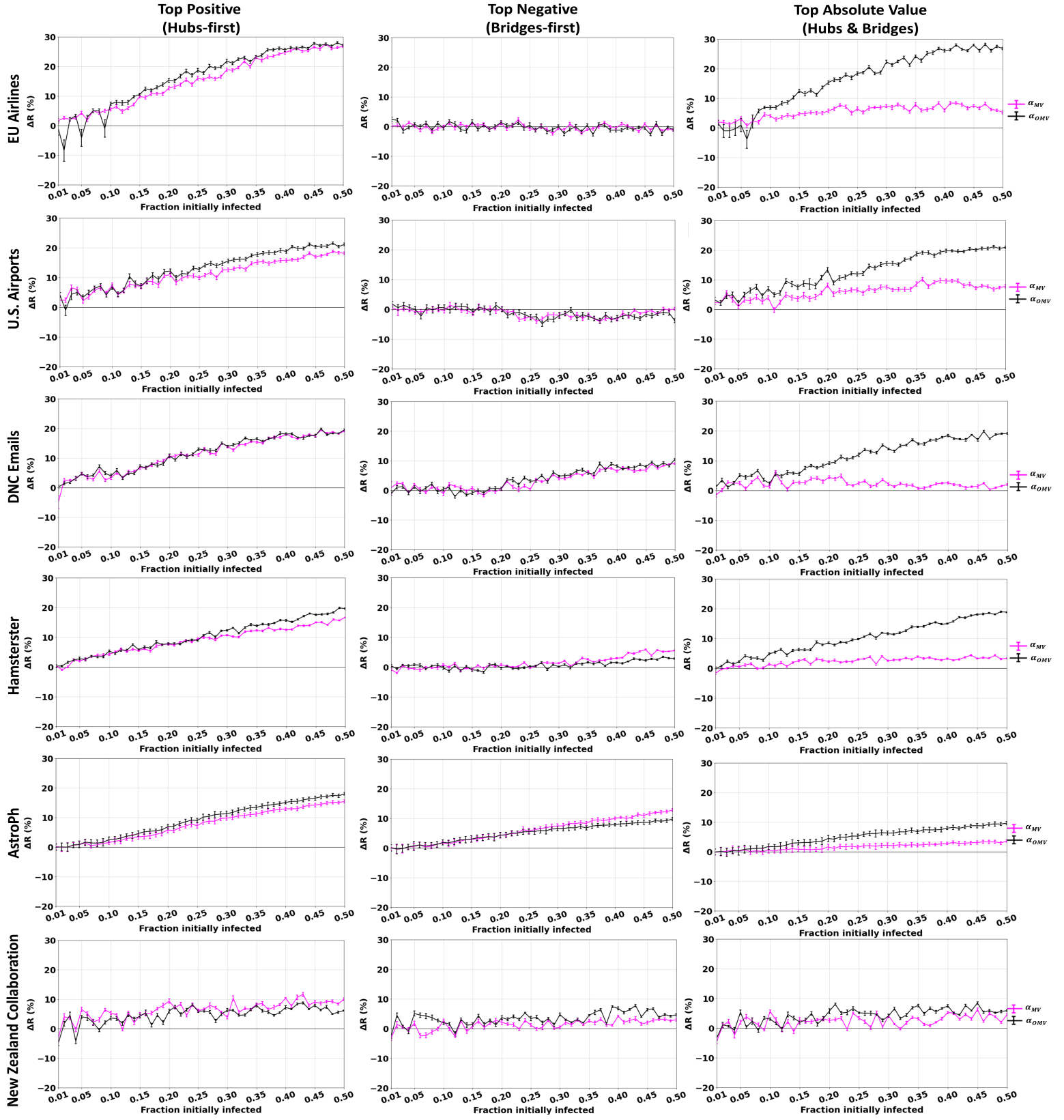
Fig. 1. Relative difference of the outbreak size (ΔR) as a function of the fraction of initially infected nodes (f_o) on the six real-world networks used. The community-aware centrality measures are Modularity Vitality (α_{MV}) and Overlapping Modularity Vitality (α_{OMV}). On the left, middle, and right of the figure, nodes are ordered from top positive centrality values, top negative centrality values, and top absolute values, respectively.

C. Targeting Hubs and Bridges

In the final experiment, we investigate the performances of the non-overlapping Modularity Vitality (α_{MV}) and its overlapping version (α_{OMV}) when the fraction of initially infected nodes (f_o) uses the absolute value of the centrality measures ranks. In other words, the initially infected nodes target hubs and bridges simultaneously, based on their global effect on modularity. The right column of figure 1 shows the relative difference of the outbreak size (ΔR) as a function of the fraction of initially infected nodes (f_o) where the nodes are ranked based on the absolute value (hubs and bridges).

Infecting nodes based on their global effect, regardless of their nature (hubs or bridges), the Overlapping Modularity Vitality (α_{OMV}) outperforms its non-overlapping counterpart in all of the networks under study. Indeed, the outperformance of α_{OMV} compared to α_{MV} can reach up to 28% in terms of the relative difference of the outbreak size (ΔR) in EU Airlines. Up to 19% in DNC Emails and Hamsterster. Up to 13% in U.S. Airports. Up to 7% in AstroPh. Finally, up to 5% in New Zealand Collaboration. The network where α_{MV} and α_{OMV} are the closest in terms of their relative difference is New Zealand Collaboration. In the other networks, the difference between α_{MV} and α_{OMV} increases as f_o increases. Indeed, one can see that when the fraction of initially infected nodes is low, the difference between α_{OMV} and α_{MV} is minimal. This illustrates the effectiveness of incorporating information from the overlapping community structure and not discarding the influence of hubs on behalf of bridges (and vice versa). In other words, both hubs and bridges play a crucial role in epidemic spreading. Targeting both of them leads to a higher diffusion.

V. Discussion

In this work, we introduce an extension of Modularity Vitality (α_{MV}), a community-aware centrality measure designed for networks with non-overlapping community structure to overlapping communities. The Overlapping Modularity Vitality (α_{OMV}) addresses a critical gap of many realistic situations where a node might belong to several communities. The extension we consider uses a simple weighting scheme based on the reciprocal number of communities shared by a node. This information is incorporated in the adjacency matrix of the network. One computes the weight by averaging the strength of membership to different communities of the overlapping nodes. This strategy is inspired by one of the various strategies discussed in [29]. So doing, hubs shared by several communities contribute the most to the overlapping modularity of a network. Conversely, bridge nodes belonging to several communities have a substantial negative contribution to the network's overlapping modularity.

Since both community-aware centrality measures can target bridges and hubs, we investigate the effectiveness of three different ranking strategies in a SIR epidemic scenario. First, nodes are ranked in decreasing order of their centrality. This strategy targets hubs at first. With this ranking scheme, when the fraction of initially infected nodes (f_o) is less than 9%,

there is no difference between the overlapping and the non-overlapping centrality measures (α_{MV} and α_{OMV}). These results corroborate similar findings reported in [41]. Indeed, extensive analysis shows that most hubs are direct neighbors of overlapping nodes. Hence, the epidemic outbreak effect is not tremendously different since local hubs and overlapping hubs are in the vicinity of each other. The difference is more pronounced when we infect more than 9% of the nodes. Here, α_{OMV} targets overlapping hub nodes that are more strategically positioned in a network, resulting in a higher epidemic outbreak.

Second, we investigate the strategy consisting of infecting nodes with the highest negative contribution to the non-overlapping (Q) and overlapping modularity (Q_o). In other words, we now target bridges. We notice that non-overlapping Modularity Vitality (α_{MV}) and Overlapping Modularity Vitality (α_{OMV}) are less effective. Indeed, even a negative relative difference of the outbreak size (ΔR) is observed. In addition, the highest relative difference of the outbreak size amounts to 13% when compared to the baseline for α_{MV} and 10% for α_{OMV}. It occurs in the AstroPh network. These values are almost half of the ones observed when the ranking scheme targets hubs first. Indeed, when hubs are infected first, the highest relative difference of the outbreak size reaches 27% when compared to the baseline for the Overlapping Modularity Vitality and 26% for the non-overlapping Modularity Vitality. Hence, targeting bridges is not as effective as targeting hubs. Moreover, one can see in DNC Emails and New Zealand Collaboration networks that α_{OMV} has a slightly higher performance than α_{MV}. For Hamsterster and AstroPh, it is the opposite. Yet in both cases, when the fraction of initially infected nodes (f_o) is less than 0.27, it is hard to differentiate between α_{MV} and α_{OMV}. These results suggest that bridges identified by the two measures are very similar or close to each other. Consequently, they result in identical relative differences in the outbreak size.

Finally, we investigate infecting the nodes based on the absolute value of the centrality values. This strategy simultaneously targets hubs and bridges. In this case, the Overlapping Modularity Vitality (α_{OMV}) outperforms its counterpart on all networks under study. Moreover, the relative difference of the epidemic outbreak size reaches up to 28% in EU Airlines. This is the highest value reached among all networks throughout the three different ranking schemes. The outperformance of α_{OMV} demonstrates the importance of considering the roles of overlapping hubs and overlapping bridges simultaneously rather than targeting each in silo.

It is worth mentioning deviations from the general trend. In some networks, using non-overlapping Modularity Vitality (α_{MV}) leads to a higher epidemic spreading compared to its overlapping version (α_{OMV}). It happens when targeting hubs first in the New Zealand Collaboration network and when targeting bridges first in Hamsterster and AstroPh. These networks are characterized with a high proportion of overlapping nodes ($on(\%) = 0.364, 0.251, 0.561$, respectively). In these cases, α_{OMV} might not be able to differentiate between

strategically positioned hubs and bridges. However, when targeting both (i.e., taking the absolute value), α_{OMV} shows its merit. It indicates the importance of targeting hubs and bridges simultaneously rather than opting for one or the other. However, these preliminary results need more investigation on networks with different proportions of overlapping nodes.

VI. CONCLUSION

Many real-world networks naturally have a community structure made of communities tightly connected internally and loosely connected externally. In a lot of realistic situations, nodes can belong to several communities. These nodes, called overlapping nodes, frequently occur in social networks, collaboration networks, biological networks, ecological networks, to name a few. For example, one protein may contribute to various functions, one human may belong to multiple institutions, and one researcher may work in several scientific domains.

Identifying influential nodes within these networks is a crucial issue. Indeed, in networks with a community structure, nodes can act as hubs (locally influential) or bridges (globally influential). This problem gave rise to several propositions in the literature. However, Modularity Vitality is the only one directly linked to the quality of the community structure. It can target hubs and bridges by quantifying the node's contribution to the modularity. In this work, we propose an extension of Modularity Vitality incorporating information about the overlap in communities. Using SIR simulations, we show that Overlapping Modularity Vitality outperforms its counterpart designed for non-overlapping communities. The difference between both measures is higher when one uses the absolute value of the centrality to rank the nodes. These results demonstrate the importance of incorporating information about the overlapping communities to identify influential nodes better. Moreover, results show that it is more effective to target hubs or hubs and bridges simultaneously rather than targeting bridges.

In future work, we plan to investigate the various alternative definitions of overlapping modularity. As in many situations, the network's community structure is unknown; one needs to evaluate the robustness of the results to community structure variation linked to the community detection algorithm. Furthermore, we plan to relate the overall performance of the Overlapping Modularity Vitality with the network topological properties.

REFERENCES

[1] L. Lü, D. Chen, X.-L. Ren, Q.-M. Zhang, Y.-C. Zhang, and T. Zhou, "Vital nodes identification in complex networks," *Physics Reports*, vol. 650, pp. 1–63, 2016.

[2] A. Ibnoulouafi, M. El Haziti, and H. Cherifi, "M-centrality: identifying key nodes based on global position and local degree variation," *Journal of Statistical Mechanics: Theory and Experiment*, vol. 2018, no. 7, p. 073407, 2018.

[3] D. Bucur, "Top influencers can be identified universally by combining classical centralities," *Scientific reports*, vol. 10, no. 1, pp. 1–14, 2020.

[4] 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, 2005.

[5] J. Xie, S. Kelley, and B. K. Szymanski, "Overlapping community detection in networks: The state-of-the-art and comparative study," *ACM Computing Surveys*, vol. 45, no. 4, pp. 1–35, 2013.

[6] M. Jebabli, H. Cherifi, C. Cherifi, and A. Hammouda, "Overlapping community structure in co-authorship networks: a case study," in *2014 7th International Conference on u-and e-Service, Science and Technology*. IEEE, 2014, pp. 26–29.

[7] M. Jebabli, H. Cherifi, C. Cherifi, and A. Hamouda, "User and group networks on youtube: A comparative analysis," in *2015 IEEE/ACS 12th International Conference of Computer Systems and Applications (AICCSA)*. IEEE, 2015, pp. 1–8.

[8] K. Das, S. Samanta, and M. Pal, "Study on centrality measures in social networks: a survey," *Social network analysis and mining*, vol. 8, no. 1, pp. 1–11, 2018.

[9] F. Bloch, M. O. Jackson, and P. Tebaldi, "Centrality measures in networks," *Available at SSRN 2749124*, 2019.

[10] S. Rajeh, M. Savonnet, E. Leclercq, and H. Cherifi, "Interplay between hierarchy and centrality in complex networks," *IEEE Access*, vol. 8, pp. 129 717–129 742, 2020.

[11] M. Kitsak, L. K. Gallos, S. Havlin, F. Liljeros, L. Muchnik, H. E. Stanley, and H. A. Makse, "Identification of influential spreaders in complex networks," *Nature physics*, vol. 6, no. 11, pp. 888–893, 2010.

[12] T. Magelinski, M. Bartulovic, and K. M. Carley, "Measuring node contribution to community structure with modularity vitality," *IEEE Transactions on Network Science and Engineering*, vol. 8, no. 1, pp. 707–723, 2021.

[13] Z. Ghalmane, M. El Hassouni, and H. Cherifi, "Immunization of networks with non-overlapping community structure," *SNAM*, vol. 9, no. 1, pp. 1–22, 2019.

[14] Z. Ghalmane, M. El Hassouni, C. Cherifi, and H. Cherifi, "Centrality in modular networks," *EPJ Data Science*, vol. 8, no. 1, p. 15, 2019.

[15] M. M. Tulu, R. Hou, and T. Younas, "Identifying influential nodes based on community structure to speed up the dissemination of information in complex network," *IEEE Access*, vol. 6, pp. 7390–7401, 2018.

[16] N. Gupta, A. Singh, and H. Cherifi, "Centrality measures for networks with community structure," *Physica A: Statistical Mechanics and its Applications*, vol. 452, pp. 46–59, 2016.

[17] R. Guimera and L. A. N. Amaral, "Functional cartography of complex metabolic networks," *nature*, vol. 433, no. 7028, pp. 895–900, 2005.

[18] Z. Zhao, X. Wang, W. Zhang, and Z. Zhu, "A community-based approach to identifying influential spreaders," *Entropy*, vol. 17, no. 4, pp. 2228–2252, 2015.

[19] S.-L. Luo, K. Gong, and L. Kang, "Identifying influential spreaders of epidemics on community networks," *arXiv:1601.07700*, 2016.

[20] S. Rajeh, M. Savonnet, E. Leclercq, and H. Cherifi, "Characterizing the interactions between classical and community-aware centrality measures in complex networks," *Scientific Reports*, vol. 11, no. 1, pp. 1–15, 2021.

[21] Z. Ghalmane, C. Cherifi, H. Cherifi, and M. El Hassouni, "Centrality in complex networks with overlapping community structure," *Scientific reports*, vol. 9, no. 1, pp. 1–29, 2019.

[22] D. Chakraborty, A. Singh, and H. Cherifi, "Immunization strategies based on the overlapping nodes in networks with community structure," in *International conference on computational social networks*. Springer, 2016, pp. 62–73.

[23] F. Taghavian, M. Salehi, and M. Teimouri, "A local immunization strategy for networks with overlapping community structure," *Physica A: Statistical Mechanics and its Applications*, vol. 467, pp. 148–156, 2017.

[24] M. E. Newman, "Modularity and community structure in networks," *Proceedings of the national academy of sciences*, vol. 103, no. 23, pp. 8577–8582, 2006.

[25] V. D. Blondel, J.-L. Guillaume, R. Lambiotte, and E. Lefebvre, "Fast unfolding of communities in large networks," *Journal of statistical mechanics: theory and experiment*, vol. 2008, no. 10, p. P10008, 2008.

[26] A. Clauset, M. E. Newman, and C. Moore, "Finding community structure in very large networks," *Physical review E*, vol. 70, no. 6, p. 066111, 2004.

[27] L. Yang, X. Cao, D. He, C. Wang, X. Wang, and W. Zhang, "Modularity based community detection with deep learning." in *IJCAI*, vol. 16, 2016, pp. 2252–2258.

[28] R. Shang, J. Bai, L. Jiao, and C. Jin, "Community detection based on modularity and an improved genetic algorithm," *Physica A: Statistical Mechanics and its Applications*, vol. 392, no. 5, pp. 1215–1231, 2013.

[29] M. Chen and B. K. Szymanski, "Fuzzy overlapping community quality metrics," *Social Network Analysis and Mining*, vol. 5, no. 1, pp. 1–14, 2015.

[30] A. Lázár, D. Abel, and T. Vicsek, "Modularity measure of networks with overlapping communities," *EPL (Europhysics Letters)*, vol. 90, no. 1, p. 18001, 2010.

[31] M. E. Newman, "Analysis of weighted networks," *Physical review E*, vol. 70, no. 5, p. 056131, 2004.

[32] T. P. Peixoto, "The netzschleuder network catalogue and repository," 2020, [Online]. Available: https://networks.skewed.de/.

[33] J. Kunegis, "Handbook of network analysis [konect–the koblenz network collection]," *arXiv preprint arXiv:1402.5500*, 2014.

[34] R. A. Rossi and N. K. Ahmed, "The network data repository with interactive graph analytics and visualization," in *AAAI*, 2015.

[35] M. Rosvall and C. T. Bergstrom, "Maps of random walks on complex networks reveal community structure," *PNAS*, vol. 105, no. 4, pp. 1118–1123, 2008.

[36] J. Xie, B. K. Szymanski, and X. Liu, "Slpa: Uncovering overlapping communities in social networks via a speaker-listener interaction dynamic process," in *2011 ieee 11th international conference on data mining workshops*. IEEE, 2011, pp. 344–349.

[37] G. K. Orman, V. Labatut, and H. Cherifi, "Comparative evaluation of community detection algorithms: a topological approach," *Journal of Statistical Mechanics: Theory and Experiment*, vol. 2012, no. 08, p. P08001, 2012.

[38] M. Jebabli, H. Cherifi, C. Cherifi, and A. Hamouda, "Community detection algorithm evaluation with ground-truth data," *Physica a: Statistical mechanics and its applications*, vol. 492, pp. 651–706, 2018.

[39] V.-L. Dao, C. Bothorel, and P. Lenca, "Community structure: A comparative evaluation of community detection methods," *arXiv preprint arXiv:1812.06598*, 2018.

[40] R. M. Anderson and R. M. May, "Population biology of infectious diseases: Part i," *Nature*, vol. 280, no. 5721, pp. 361–367, 1979.

[41] Z. Ghalmane, C. Cherifi, H. Cherifi, and M. El Hassouni, "Exploring hubs and overlapping nodes interactions in modular complex networks," *IEEE Access*, vol. 8, pp. 79 650–79 683, 2020.