



# Comparative evaluation of community-aware centrality measures

Stephany Rajeh<sup>1</sup> · Marinette Savonnet<sup>1</sup> · Eric Leclercq<sup>1</sup> · Hocine Cherifi<sup>1</sup>

Accepted: 21 April 2022 / Published online: 10 May 2022  
© The Author(s), under exclusive licence to Springer Nature B.V. 2022

## Abstract

Influential nodes play a critical role in boosting or curbing spreading phenomena in complex networks. Numerous centrality measures have been proposed for identifying and ranking the nodes according to their importance. Classical centrality measures rely on various local or global properties of the nodes. They do not take into account the network community structure. Recently, a growing number of researches have shifted to community-aware centrality measures. Indeed, it is a ubiquitous feature in a vast majority of real-world networks. In the literature, the focus is on designing community-aware centrality measures. However, up to now, there is no systematic evaluation of their effectiveness. This study fills this gap. It allows answering which community-aware centrality measure should be used in practical situations. We investigate seven influential community-aware centrality measures in an epidemic spreading process scenario using the Susceptible–Infected–Recovered model on a set of fifteen real-world networks. Results show that generally, the correlation between community-aware centrality measures is low. Furthermore, in a multiple-spreader problem, when resources are available, targeting distant hubs using Modularity Vitality is more effective. However, with limited resources, diffusion expands better through bridges, especially in networks with a medium or strong community structure.

**Keywords** Complex networks · Centrality · Influential nodes · Community structure · SIR model

## 1 Introduction

Complex networks can describe a wide range of real-world complex systems such as power grids, air transportation, social networks, and financial transactions. Influential nodes identification within these systems is of great interest. Indeed, one can use this knowledge to enhance the diffusion process in viral marketing applications and control epidemic spreading through appropriate immunization strategies. Numerous centrality measures have been proposed to identify influential nodes (Lü et al. 2016). Local centrality measures compute the centrality of a node based on its neighborhood. Global

---

✉ Stephany Rajeh  
stephany.rajeh@u-bourgogne.fr

<sup>1</sup> Laboratoire d'Informatique de Bourgogne, University of Burgundy, Dijon, France

centrality measures quantify the centrality of a node by inspecting its position in the entire network. Finally, mixed centrality measures combine both types of information (Berahmand et al. 2019). These centrality measures, which we call classical, exploit various topological properties of the nodes ignoring the network organization in communities. However, it is well-known that community structure is one of the main features characterizing real-world networks (Girvan and Newman 2002).

Current developments exploit the modular organization of networks to derive community-aware centrality measures (Ghalmane et al. 2019; Guimera and Amaral 2005; Tulu et al. 2018; Gupta et al. 2016; Magelinski et al. 2021; Zhao et al. 2015; Luo et al. 2016). Indeed, the structuring of networks in communities is ubiquitous in many real-world networks. Furthermore, it significantly affects the diffusion dynamics of a network Chérifi et al. (2019).

Consequently, nodes that may not be considered influential by a classical centrality measure (i.e., agnostic about the community structure) may be of ultimate influence when one considers the mesoscopic organization of the network. Community-aware centrality measures distinguish intra-community links from inter-community links. The former links join nodes in the same community while the latter join nodes in different communities. Intra-community links allow quantifying the node's local influence inside its community. In contrast, inter-community links account for the node's global effect on the various communities forming the network.

Community-aware centrality measures differ on how they combine the two types of links. Community Hub-Bridge (Ghalmane et al. 2019) picks up highly connected nodes (hubs) and bridges (connections between communities) simultaneously. It weights the intra-community links with the node's community size. The weight of inter-community links is the number of neighboring communities. With a slightly different approach, Comm Centrality (Gupta et al. 2016) gives more importance to bridges by raising the inter-community links to the power of 2. Community-based Centrality (Zhao et al. 2015) ponderates the intra-community links and the inter-community links depending on their respective community sizes. For example, suppose a node has an inter-community link connecting to a community made up of more than 80% of the total nodes in a network. This specific inter-community link will receive a larger weight than another inter-community link connecting to a smaller community. K-shell with Community (Luo et al. 2016) identifies important nodes based on their  $k$ -shell value by considering intra-community links and inter-community links separately. Participation Coefficient puts more emphasis on the heterogeneity of inter-community links (Guimera and Amaral 2005). If a node participates in several communities, it will receive a high participation coefficient. Similarly, Community-based Mediator (Tulu et al. 2018) targets influential nodes based on the heterogeneity of their intra-community and inter-community links. However, in this case, it combines the links using entropy. If intra-community and inter-community links of a node are equal, Community-based Mediator reduces to the normalized degree centrality. Modularity Vitality (Magelinski et al. 2021) is based on the modularity measure of the community structure strength. It computes the node centrality using the modularity variation induced when one removes it from the network. Modularity Vitality distinguishes bridges from hubs. The idea is that eliminating hubs decreases modularity while removing bridges increases it.

In previous works Rajeh et al. (2021a, c), we studied the interactions between classical and community-aware centrality measures and the influence of the network topological features. Results show that the community structure strength is the main feature driving the correlation between classical and community-aware centrality measures.

In contrast, this work focuses on the interactions between community-aware centrality measures. We perform a systematic investigation of seven popular community-aware centrality measures using fifteen real-world networks. These networks originate from various domains and cover a wide range of basic topological properties. First, we perform a correlation analysis of the multiple pairs of community-aware centrality measures. Second, we investigate the influence of the community structure strength. Finally, we analyze the effectiveness of the community-aware centrality measures. To this end, we consider their relative diffusive power in a scenario based on the Susceptible–Infected–Recovered (SIR) model using a multiple-spreader infection scheme. This extensive comparative analysis provides new insights to preliminary investigations conducted in online social networks in the SIR single-spreader infection scheme reported in Rajeh et al. (2021b).

The main contributions of the paper summarize as follows:

1. We investigate the correlation of several influential community-aware centrality measures.
2. We perform an extensive comparative analysis of their effectiveness in a SIR simulation scenario using real-world networks from different domains with a wide range of topological properties.
3. We study the interplay between the community structure characteristics, the availability of resources, and the performance of the various community-aware centrality measure under test.
4. We give clear indications about the most effective strategies to employ according to the community structure strength and the availability of resources at hand.

The paper is organized as follows. Section 2 briefly discusses the related work. The community-aware centrality measures are introduced in Sect. 3. The data and the tools used in the evaluation process are presented in Sect. 4. In Sect. 5, the experimental results are given. In Sect. 6, a discussion on the findings is developed. Section 7 concludes the article.

## 2 Related work

The relationship between centrality measures and topological network features is a fundamental issue. Initially focused on macroscopic topological properties, it has recently shifted to mesoscopic topological properties. One can categorize related works into three main topics. The most widespread research works deal with the interplay between classical centrality measures and networks' macroscopic topological features. Uncovering the relations between essential network features and community structure is also a well-studied issue. Finally, although there is a growing trend to design community-aware centrality measures, to our knowledge, there are few exhaustive comparative evaluations.

Studies that investigate the relationship between networks' macroscopic topological properties and centrality measures concern mainly classical centrality measures (Li et al. 2015; Ronqui and Travieso 2015; Rajeh et al. 2020; Schoch et al. 2017; Oldham et al. 2019). In Li et al. (2015), Li et al. study the correlations between the betweenness, the closeness, the components of the principal eigenvector, the degree, the first-order degree mass, and the second-order degree mass, in the Erdős-Rényi and the scale-free network models and 34 real-world networks. They show that classical centrality measures are generally positively correlated independently of the network's size. Ronqui and Travieso

(2015) show that classical centralities have stronger correlations in scale-free network models such as the Barabási-Albert model. Rajeh et al. (2020), observe that the higher the network's density and transitivity, the higher the correlation between centrality measures and hierarchy measures ( $k$ -core and  $k$ -truss). Schoch et al. (2017) show that the closer a graph to a threshold graph, the more correlated classical centrality measures, independent of their conceptual distinctness. Note that a threshold graph is a class of graphs that generalizes the star graph property and represents the purest form of a core-periphery structure. Indeed, one can partition the nodes into a clique and an independent set. Oldham et al. (2019) show that modularity plays a significant role in the correlation between classical centrality measures. Although these works give precious insights on how network topology affects the redundancy between centrality measures, they do not investigate community-aware centrality measures.

Uncovering the relations between the macroscopic topological properties and the community structure of networks has also attracted many researchers (Nematzadeh et al. 2014; Orman et al. 2013; Wang and Qin 2010; Wharrie et al. 2019; Lancichinetti et al. 2010). For example, Nematzadeh et al. (2014) show that a strong community structure enhances information diffusion under the linear threshold model through intra-community links. Orman et al. (2013) demonstrate the positive relationship between the community structure strength and transitivity. The work of Wang and Qin (2010) illustrates that communicability is more difficult in networks with a well-defined community structure. Wharrie et al. (2019) show that the number of communities naturally increases when networks exhibit high clustering. Lancichinetti et al. (2010) report that networks from the same domain often have several similar mesoscopic characteristics, such as the community size distribution. These studies demonstrate the interplay between the community structure and the macroscopic topological properties of the network. Nonetheless, they do not relate these outcomes with community-aware centrality measures.

Even though various community-aware centrality measures have been developed (Ghalmane et al. 2019; Guimera and Amaral 2005; Tulu et al. 2018; Gupta et al. 2016; Magelinski et al. 2021; Zhao et al. 2015; Luo et al. 2016; Rajeh et al. 2020), the literature reports few comparative studies. Authors generally introduce a new community-aware centrality measure and compare it with a few alternatives. Furthermore, there is no common framework to assess their performance. In recent works Rajeh et al. (2021a, c), we performed extensive comparative studies to evaluate the relationship between classical and community-aware centrality measures. Results show that the community structure strength is the top network feature influencing their correlation. More specifically, in networks with a strong community structure, the local influence of a community-aware centrality measure highly correlates with its classical counterpart. In contrast, correlation is low for the global impact of the community-aware centrality measures. One observes the opposite behavior in networks with a weak community structure.

In preliminary work, we investigated the diffusion dynamics of seven community-aware centrality measures on online social networks. This study uses the Susceptible-Infected-Recovered (SIR) model under a single-spreader activation scheme (Rajeh et al. 2021b). Results give precious indications about the effectiveness of each measure for identifying top spreaders. However, more experiments are necessary to consolidate these results. Here, we extend this work in three directions. First, we consider a multiple-spreader activation scheme for the diffusion process. Second, we use additional networks from various domains and different topologies and sizes. Third, we investigate the influence of the community structure strength in correlating the community-aware centrality measures.

### 3 Community-aware centrality measures

The definitions of the seven community-aware centrality measures are provided in this section. Assume that  $G(V, E)$  is a simple, undirected, and unweighted graph where  $V$  is the set of nodes of size  $N = |V|$  nodes and  $E \subseteq V \times V$  is the set of edges of size  $M = |E|$  edges. The graph  $G$  is partitioned into  $\zeta = \{c_1, c_2, \dots, c_q, \dots, c_{|\zeta|}\}$  communities, where  $C = |\zeta|$  is the total number of communities,  $c_q$  is the  $q$ -th community, and  $n_{c_q}$  is the number of nodes in community  $c_q$ . Each node  $i$  in  $G$  has a total degree of  $k_i^{tot} = k_i^{intra} + k_i^{inter}$  where  $k_i^{intra}$  and  $k_i^{inter}$  are the intra-community and inter-community links, respectively. Additionally,  $k_{i,c_q}$  is the number of links node  $i$  has in a given community  $c_q$ . Table 1 reports the list of symbols used in this paper and Table 2 states the main characteristics of the community-aware centrality measures under investigation.

**1. Community Hub-Bridge** Ghalmane et al. (2019) targets hubs and bridges simultaneously. It weights the node’s intra-community links by the size of its community and

**Table 1** Symbols used in this study

Symbol	Meaning
$V$	Set of nodes
$N$	Total number of nodes
$E$	Set of edges
$M$	Total number of edges
$\zeta$	Set of communities
$C$	Total number of communities
$c_q$	$q$ -th community
$n_{c_q}$	Total number of nodes in community $c_q$
$k_i^{intra}$	Intra-community links of a node
$k_i^{inter}$	Inter-community links of a node
$k_{i,c_q}$	Total number of links of a node in community $c_q$
$k_i^{tot}$	Total number of links of a node
$a_{ij}$	Connectivity between nodes $i$ and $j$
$H_i$	Entropy of the node’s intra/inter-community links
$\rho_i^{intra}$	Fraction of intra-community links of a node
$\rho_i^{inter}$	Fraction of inter-community links of a node
$\mu_{c_q}$	Fraction of inter-community links in community $c_q$
$\mu$	Fraction of inter-community links in graph $G$
$R$	A constant for standardization
$f_o$	Fraction of initially infected nodes
$Q(G)$	Newman’s modularity
$Q(G \setminus \{i\})$	Newman’s modularity after the removal of node $i$
$\alpha_{CHB}$	Community hub-bridge
$\alpha_{PC}$	Participation coefficient
$\alpha_{CBM}$	Community-based mediator
$\alpha_{Comm}$	Comm centrality
$\alpha_{MV}^+$	Modularity vitality
$\alpha_{CBC}$	Community-based centrality
$\alpha_{ks}$	K-shell with community

**Table 2** Main differences between the investigated community-aware centrality measures

Measure	Complexity	Description	Limitation(s)
Community hub-bridge	$O(N^2C)$	Detects hubs and bridges by giving preference to the local community size and direct belonging communities	May miss influential hubs if the network has a high number of communities
Participation coefficient	$O(N < k >)$	Finds nodes having a uniform link distribution across all the communities	It indirectly prioritizes bridges over hubs and it may result in non-unique of centrality values
Community-based mediator	$O(NM < k >)$	Discovers intermediate nodes capable of quickly disseminating information across the communities	Reduces to degree centrality if the internal and external link density of a node are equal
Comm centrality	$O(NC)$	Searches for hubs and bridges while taking into consideration the community structure strength	Directly prioritizes bridges and is undefined if at least one community is disconnected from the others
Modularity vitality	$O(NC + M)$	A signed measure that can be used to explicitly target hubs, bridges, or both	Susceptible to the limitations of Newman's modularity such as the resolution limit and local maxima
Community-based centrality	$O(N < k >)$	Detects influential nodes based on the communities they participate in and their sizes	The higher the number of communities, the more similar it is to normalized degree centrality
K-shell with community	$O(NM)$	Determines core nodes embedded in local and global communities	Redundancy of nodes existing at the same hierarchical level

the inter-community links by the number of communities a node can reach in one hop. It is defined as follows:

$$\alpha_{CHB}(i) = n_{c_q} \times k_i^{intra} + \sum_{c_l \subset \mathcal{C} \setminus c_q} \bigvee_{j \in c_l} a_{ij} \times k_i^{inter} \tag{1}$$

where  $n_{c_q}$  is node  $i$ 's community size and  $\bigvee_{j \in c_l} a_{ij} = 1$  if node  $i$  is connected to at least one node  $j$  in community  $c_l$ .

**2. Participation Coefficient** Guimera and Amaral (2005) emphasizes a node's influence based on the fraction of its inter-community links. If a node is linked only to nodes in its community, its Participation Coefficient is zero. It is defined as follows:

$$\alpha_{PC}(i) = 1 - \sum_{q=1}^C \left( \frac{k_{i,c_q}}{k_i^{tot}} \right)^2 \tag{2}$$

**3. Community-based Mediator** Tulu et al. (2018) is based on the entropy of the intra-community and inter-community links of a node. The more mixed the links of a node, the higher its centrality value. It is defined as follows:

$$\alpha_{CBM}(i) = H_i \times \frac{k_i^{tot}}{\sum_{i=1}^N k_i^{tot}} \tag{3}$$

where  $H_i = [-\sum \rho_i^{intra} \log(\rho_i^{intra})] + [-\sum \rho_i^{inter} \log(\rho_i^{inter})]$  is the entropy of node  $i$  based on its  $\rho_i^{intra}$  and  $\rho_i^{inter}$  which represent the node's ratio of intra-community and inter-community links over the total degree and  $\sum_{i=1}^N k_i^{tot}$  is the total sum of all degrees in the network.

**4. Comm Centrality** Gupta et al. (2016) preferentially targets bridges. However, hubs locally situated with their intra-community links are not discarded. It is defined as follows:

$$\alpha_{Comm}(i) = (1 + \mu_{c_q}) \times \chi + (1 - \mu_{c_q}) \times \varphi^2 \tag{4}$$

where  $\mu_{c_q}$  is the fraction of inter-community links over the total community links in community  $c_q$ ,  $\chi = \frac{k_i^{intra}}{\max_{j \in c_q} k_j^{intra}} \times R$ ,  $\varphi = \frac{k_i^{inter}}{\max_{j \in c_q} k_j^{inter}} \times R$ , and  $R$  is a constant to standardize intra-community and inter-community values.

**5. Modularity Vitality** Magelinski et al. (2021) is a signed community-aware centrality measure. It can differentiate a hub from a bridge based on Newman's modularity Newman (2006) variation after the node's removal. It is defined as follows:

$$\alpha_{MV}(i) = Q(G) - Q(G \setminus \{i\}) \tag{5}$$

where  $Q(G)$  is the network's modularity and  $Q(G \setminus \{i\})$  is the network's modularity after the removal of node  $i$ . Note that in this study, we use the version of Modularity Vitality which targets hubs first (i.e., nodes are ordered from positive to negative magnitude). We denote it as  $\alpha_{MV}^+$ . Results with other versions targeting bridges first are not reported because they show poor performance. Note that computation of the Modularity Vitality upon node removal involves limited information. Indeed, it only requires knowledge about the node's 1-hop neighborhood and the total degrees of communities.

**6. Community-based Centrality** Zhao et al. (2015) weights the node's intra-community and inter-community links by the size of their communities. It is defined as follows:

$$\alpha_{CBC}(i) = \sum_{q=1}^C k_{i,c_q} \left( \frac{n_{c_q}}{N} \right) \quad (6)$$

**7. K-shell with Community** Luo et al. (2016) splits the network  $G$  into two networks. One is made of the nodes and their intra-community links, and the other contains the nodes and the inter-community links. Then, a linear combination of the  $k$ -shell hierarchical decomposition of these networks assesses the node's influence Dorogovtsev et al. (2006). It is defined as follows:

$$\alpha_{ks}(i) = \delta \times \alpha^{intra}(i) + (1 - \delta) \times \alpha^{inter}(i) \quad (7)$$

where  $\alpha^{intra}(i)$  and  $\alpha^{inter}(i)$  refer to the  $k$ -shell value of node  $i$  on the graphs constituting intra-community links and inter-community links, respectively. In this study,  $\delta$  is set to 0.5 so that neither hubs nor bridges are preferentially selected over one another.

## 4 Data and tools

This section briefly presents the networks used to conduct the study. It also reports on the SIR evaluation process and the measures used in the comparative analysis.

### 4.1 Data

The experiments concern fifteen real-world networks spanning biological networks, infrastructural networks, and offline/online social networks. Table 3 reports their basic topological characteristics.

**Biological networks:** In Yeast Protein (Rossi and Ahmed 2015) and Yeast Collins (Peixoto 2020), the nodes represent proteins and are connected to each other if there's a physical interchange.

**Infrastructural networks:** In the EU Airlines network (Peixoto 2020) and U.S. Airports network (Kunegis 2014), nodes represent European and U.S. airports, respectively. Nodes are connected if there's a direct flight between two airports. In the U.S. Power Grid network (Kunegis 2014), nodes represent either a generator, transformer, or substation, and edges represent a power supply line.

**Offline social networks:** In the Adolescent Health network (Kunegis 2014), nodes represent students and edges represent friendships. Students have been asked to list 5 of their female and male friends.

**Online social networks:** Six Facebook networks are under study [Facebook Friends, Ego Facebook, Caltech, Princeton, Facebook Organizations, and Facebook Politician Pages (Peixoto 2020; Rossi and Ahmed 2015)]. In all networks, nodes represent Facebook users, and edges represent online friendships except for Facebook Politician Pages. In the latter, nodes represent politician pages from different countries, and edges represent mutual likes. In an online social pet platform called "Hamsterster" Kunegis (2014), nodes represent users of this platform, and edges represent online friendships. The DeezerEU network Rozemberczki and Sarkar (2020) is formed based on the Deezer platform for music streaming. In this network, nodes are users from European countries, and edges represent mutual follower relationships. Finally, in DNC Emails Peixoto (2020), nodes represent members of the Democrat National Committee, and edges represent email exchanges.



**Table 3** Topological features of the networks

Network	$N$	$M$	$\langle k \rangle$	$\kappa$	$Q$	$\lambda_{th}$
(A) Facebook friends*	329	1954	11.88	0.51	0.69	0.05
(B) EU airlines	417	2953	14.16	0.30	0.11	0.02
(C) U.S. Airports	500	2980	11.92	0.35	0.16	0.02
(D) Caltech*	762	16,651	43.70	0.29	0.39	0.05
(E) DNC emails*	849	10,384	24.46	0.55	0.42	0.01
(F) Yeast collins*	1004	8319	16.57	0.62	0.75	0.03
(G) Yeast protein*	1458	1993	2.73	0.05	0.75	0.16
(H) Hamsterster*	1788	12,476	13.49	0.09	0.39	0.02
(I) Adol. health	2539	10,455	8.23	0.04	0.57	0.11
(J) Ego Facebook	4039	88,234	43.69	0.52	0.81	0.01
(K) U.S. Power Grid	4941	6594	2.66	0.10	0.83	0.35
(L) Facebook organizations	5524	94,219	34.11	0.22	0.59	0.02
(M) Facebook politician Pages	5908	41,729	14.12	0.30	0.84	0.02
(N) Princeton*	6575	293,307	89.21	0.16	0.42	0.01
(O) DeezerEU	28,281	92,752	6.55	0.10	0.57	0.07

$N$  is the number of nodes.  $M$  is the number of edges.  $\langle k \rangle$  is the average degree.  $\kappa$  is the transitivity.  $Q$  is the modularity.  $\lambda_{th}$  is the epidemic threshold. \* means the largest connected component of the network is taken if it is disconnected

## 4.2 Infomap community detection algorithm

We use Infomap (Rosvall and Bergstrom 2008) to uncover the communities of the real-world networks. Indeed, this popular algorithm has proved to be quite efficient in artificial and real-world benchmarks (Yang et al. 2016; Orman et al. 2011, 2012; Jebabli et al. 2018). It minimizes the information flow used to characterize the connections in a network. Consider a random walker. It is more likely to stay longer in a densely connected cluster and rarely jumps to another cluster. Accordingly, Infomap uses Huffman coding to compress the encoded information of the random walker in each community. The optimal partitioning minimizes the map equation ( $L$ ). It assigns nodes to clusters to reduce the random walker's movement. First, each node belongs to a unique community. Then, nodes are considered to be in the same community if they result in the largest decrease in  $L$ . This procedure iterates until the map equation reaches a minimum. In the end, one associates each node with a code in two parts: the first part (i.e., prefix) determines its community. The second part (i.e., codeword) refers to the node within the community.

## 4.3 Kendall's Tau correlation

The Kendall's Tau correlation is used to measure the ordinal consistency of two sets of ranked nodes. Given two ranked sets  $X = (x_1, x_2, \dots, x_n)$  and  $Y = (y_1, y_2, \dots, y_n)$  of size  $n$ , a pair of ranks  $(x_i, y_i)$  and  $(x_j, y_j)$  is considered concordant if  $x_i > x_j$  and  $y_i > y_j$ , or if  $x_i < x_j$  and  $y_i < y_j$ . It is said discordant if  $x_i > x_j$  and  $y_i < y_j$ , or if  $x_i < x_j$  and  $y_i > y_j$ . In case of ties (if  $x_i = x_j$  or  $y_i = y_j$ ), the pair of ranks is neither concordant nor discordant. The Kendall's Tau correlation  $\tau_b$  between two ranking sets  $X$  and  $Y$  of size  $t$  is given by:

$$\tau_b(X, Y) = \frac{n_{co} - n_{di}}{\sqrt{(n_{co} + n_{di} + x)(n_{co} + n_{di} + y)}} \quad (8)$$

where  $n_{co}$  and  $n_{di}$  denote the number of concordant and discordant pairs, respectively.  $x$  and  $y$  denote the number of tied pairs in sets  $X$  and  $Y$ , respectively.  $\tau_b$  provides a value in the interval  $[-1, 1]$ . If  $\tau_b > 0$ , there is a positive monotonic association between the two sets. If  $\tau_b < 0$ , there is a negative monotonic association between the two sets.

#### 4.4 Pearson correlation

The Pearson correlation coefficient measures the linear association and direction between two variables. It is given by:

$$\rho(X, Y) = \frac{\sum_{i=1}^n (X_i - \bar{X})(Y_i - \bar{Y})}{\sqrt{\sum_{i=1}^n (X_i - \bar{X})^2 \sum_{i=1}^n (Y_i - \bar{Y})^2}} \quad (9)$$

where  $n$  is the sample size,  $\bar{X} = \frac{\sum_{i=1}^n X_i}{n}$  is the sample mean of variable  $X$ , and  $\bar{Y} = \frac{\sum_{i=1}^n Y_i}{n}$  is the sample mean of variable  $Y$ . Pearson's correlation values are in the range  $[-1, +1]$ . The greater the absolute value of the correlation coefficient, the stronger the relationship. Extreme values indicate a perfect linear relationship. The sign of the correlation coefficient gives the direction of the relationship. It is positive when the value of one variable increases, the value of the other variable also tends to increase. Inversely, negative coefficients represent cases when the value of one variable increases and the other variable's value tends to decrease.

#### 4.5 Susceptible–Infected–Recovered model

The Susceptible–Infected–Recovered (SIR) model (Anderson and May 1979) is a popular diffusion model for assessing the effectiveness of centrality measures. It can be used to model a wide range of real-world scenarios such as viral marketing and cyberattacks and has numerous implications (Ismail et al. 2017; Nguyen 2017; Buckee et al. 2021; Toda 2020). At first, one ranks the nodes in the decreasing order of their centrality value. A given proportion ( $f_o$ ) of the top-ranked nodes is set in the infectious state (I). All the other nodes are in the susceptible state (S). An infectious node infects its susceptible neighbors with a probability  $\lambda$ , as the propagation proceeds. At the same time, an infected node can recover at a rate  $\gamma$ . By the end of the propagation phenomenon, the disease propagation stops when all nodes are either in the susceptible or recovered state. Here, the outbreak size based on the number of nodes in the recovered state (R) is measured. The outbreak size determines the spreading power of  $f_o$ . Therefore, the goal is to maximize this value. All networks have an epidemic threshold that controls the epidemic spreading. In this study, we use the following definition to calculate it (Wang et al. 2016):

$$\lambda_{th} = \frac{\langle k \rangle}{\langle k^2 \rangle - \langle k \rangle} \quad (10)$$

where  $\langle k \rangle$  and  $\langle k^2 \rangle$  are the first and second moments of the network's degree distribution. The epidemic threshold value of each network is provided in Table 3. Note that we use three values of the infection rate in the experiments (the epidemic threshold  $\lambda_{th}$ ,  $\lambda_{th}/2$

and  $\lambda_{th} \times 2$ ). As we do not observe significant differences, we report only the results equal to the epidemic threshold. Moreover, the SIR simulations are averaged over 100 independent iterations for each network.

#### 4.6 The mixing parameter

The mixing parameter characterizes how strong or weak a community structure is. If the network has well-separated communities, its mixing parameter is low, indicating a strong community structure. If the network's communities are not clearly defined, the mixing parameter is high, indicating a weak community structure. It is quantified by the fraction of inter-community links over the total links in a network. Its value is in the range of [0,1]. The less the inter-community links (i.e., links from one community to another), the stronger the community structure is. It is given by:

$$\mu = \frac{\sum_{i=1}^N k_i^{inter}}{\sum_{i=1}^N k_i^{tot}} \quad (11)$$

where  $\sum_{i=1}^N k_i^{inter}$  represents the inter-community links of all the nodes in the network and  $\sum_{i=1}^N k_i^{tot}$  represents the total sum of all degrees in the network.

#### 4.7 Evaluation criterion

We use the degree centrality as a baseline to compare the spreading outbreak size of community-aware centrality measures. Indeed, all the community-aware centrality measures are extensions of degree centrality ( $k_i^{tot} = k_i^{intra} + k_i^{inter}$ ) which in turn is agnostic to community structure. The relative difference is defined as:

$$\Delta R = \frac{R_c - R_b}{R_b} \quad (12)$$

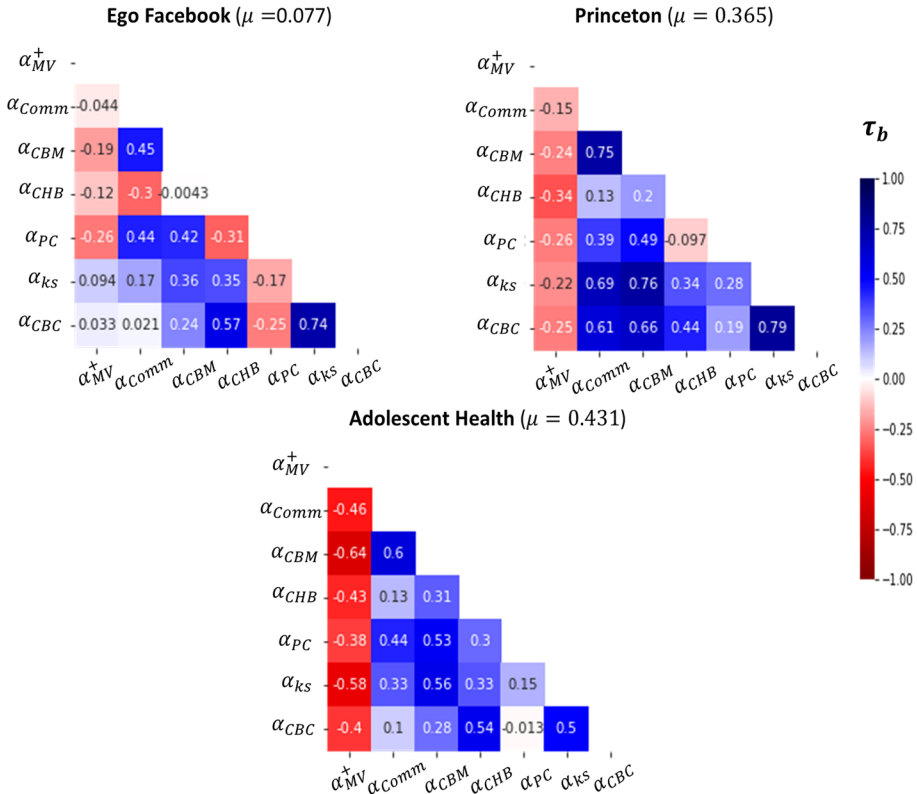
where  $R_c$  denotes the total number of recovered nodes using a specific community-aware centrality measure  $c$ .  $R_b$  represents the total number of recovered nodes with the baseline degree centrality. A positive value of  $\Delta R$  indicates that the community-aware centrality measure is more effective than the baseline.

## 5 Experimental results

This section reports the results of the experiments. First of all, we examine how the various community-aware centrality measures correlate. Then, we investigate the influence of the network's community structure strength on the correlation. Finally, we compare the spreading effectiveness of the centrality measures.

### 5.1 Correlation of the community-aware centrality measures

The first experiment investigates the correlation between the various couples of community-aware centrality measures in each network. Figure 1 illustrates the correlation



**Fig. 1** Kendall’s Tau correlation ( $\tau_b$ ) between all the community-aware centrality measures: Modularity Vitality =  $\alpha_{MV}^+$ , Comm Centrality =  $\alpha_{Comm}$ , Community-based Mediator =  $\alpha_{CBM}$ , Community Hub-Bridge =  $\alpha_{CHB}$ , Participation Coefficient =  $\alpha_{PC}$ , K-shell with Community =  $\alpha_{ks}$ , and Community-based Centrality =  $\alpha_{CBC}$

heatmaps for three typical networks. The heatmaps of the other networks are available in the supplementary material.

The heatmap of the first typical network, Ego Facebook, is very patchy. In other words, the correlation between the various community-aware centrality measure ranges from low negative correlation to high positive values. For example, Community-based Centrality has a negative correlation value (−0.25) with Participation Coefficient and a high positive value (+0.74) with K-shell with Community. Similar behavior characterizes EU Airlines, U.S. Airports, Facebook Politician Pages, Facebook Friends, Yeast Collins, and U.S. Power Grid.

The second typical network is Princeton. Modularity Vitality has a weak negative or positive correlation with its alternatives. The correlation values are more uniform between the other community-aware centrality measure. Note, however, that some low negative correlation values may appear. Networks that follow this behavior are DNC Emails, Yeast Protein, Hamsterster, and Facebook Organizations.

The third typical network is Adolescent Health. Correlation of Modularity Vitality exhibits weak to medium negative correlation values with the others. The remaining

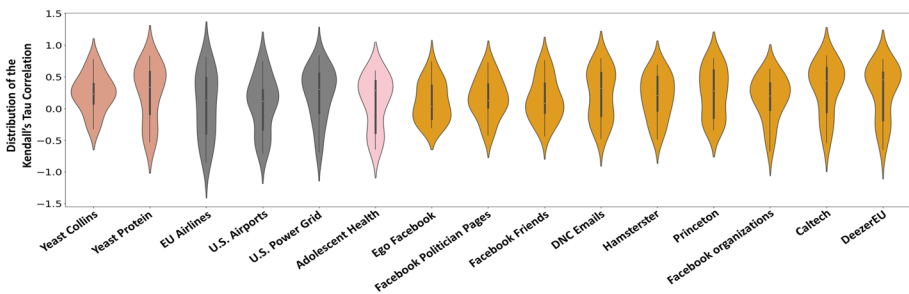
correlation values are positive, ranging from low to medium values. One observes similar behavior in Caltech and DeezerEU.

Figure 2 shows the distribution of Kendall's Tau correlation between the community-aware centrality measures for each network. Note that some distributions may pass the extremes  $[-1, +1]$  due to the smoothing of the kernel density function. One observes significant variation of the distributions' range for biological networks and online/offline networks. For example, the distribution of Yeast Collins is more compact as compared to Yeast Protein. Similar observations apply to Ego Facebook compared to DeezerEU. Additionally, three networks have bimodal distributions (EU Airlines, DNC Emails, and Princeton). The first modal value is positive, and the second is negative. The remaining network distributions are unimodal. That being said, the average mean of networks' correlation distributions is  $0.19 \pm 0.08$  and the average median is  $0.24 \pm 0.12$ . Hence, the correlation between the community-aware centrality measures is globally low. Additionally, there is an emerging pattern related to Modularity Vitality. A closer inspection of its correlation values shows that the negative tail of the distribution is generally related to the combinations including Modularity Vitality.

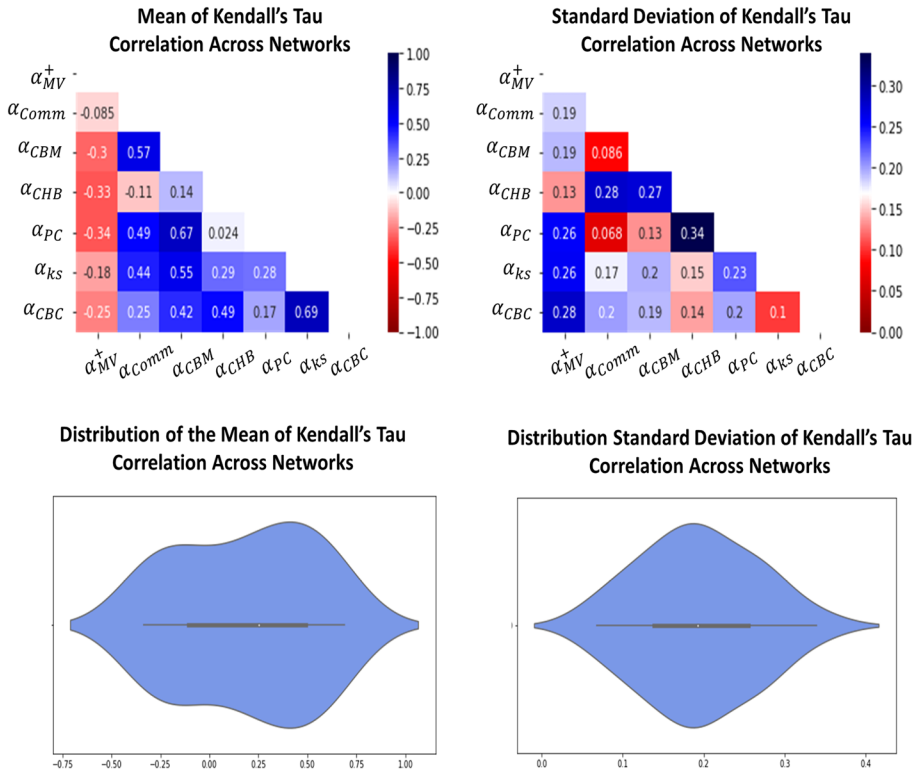
Hence, Modularity Vitality is extracting diverging information with the alternative community-aware centrality measures. The negative correlation values are not unexpected. Indeed, Modularity Vitality favors hubs while the other community-aware centrality measures mainly pinpoint bridges as the most influential nodes. Nonetheless, even if they aim at targeting bridges, it does not mean that they are highly correlated.

We calculate the mean and the standard deviation of the Kendall's Tau correlation between the community-aware centrality measures across the fifteen networks. This enables us to investigate if they generally tend to correlate or not independently of the network structure. The top left figure in Fig. 3 shows two emerging patterns. The first concerns Modularity Vitality. It has a weak negative correlation with the rest of the community-aware centrality measures. The couple consisting of Community Hub-Bridge with Comm Centrality also shows a weak negative correlation. The second concerns the other community-aware centrality measures. They present low to medium positive correlation values. Indeed, whatever the combination, correlation is always smaller than 0.70.

The bottom left figure shows the distribution of the top left figure in in Fig. 3. It is bimodal, with a positive mode and a negative one. This observation corroborates previous results on the distribution of each network reported in Fig. 2. It further explains the two emergent patterns encountered with the community-aware centrality measures under test.



**Fig. 2** Violin plots of Kendall's Tau correlation between the community-aware centrality measures for each network. Colors represent the network's domain. Biological networks are dark pink. Infrastructural networks are grey. Offline social networks are light pink. Online social networks are orange



**Fig. 3** Upper figures: Mean and standard deviation of Kendall's Tau correlation between the community-aware centrality measures across the fifteen real-world networks. Bottom figures: The distribution of the mean and standard deviation of Kendall's Tau correlation between the community-aware centrality measures across the fifteen real-world networks. The community-aware centrality measures are: Modularity Vitality targeting hubs =  $\alpha_{MV}^+$ , Comm Centrality =  $\alpha_{Comm}^+$ , Community-based Mediator =  $\alpha_{CBM}$ , Community Hub-Bridge =  $\alpha_{CHB}$ , Participation Coefficient =  $\alpha_{PC}$ , K-shell with Community =  $\alpha_{ks}$ , and Community-based Centrality =  $\alpha_{CBC}$

The top right figure in Fig. 3 shows the standard deviation of the means across networks. It ranges from 0.068 to 0.34. Its most frequent value is around 0.2. The violin plot reported in the bottom right figure presents its distribution. The mean value is  $0.19 \pm 0.07$ . Here, the distribution is symmetric and unimodal.

To summarize, these experiments show that one can consider three typical behaviors for the correlation heatmaps of the community-aware centrality measures. These cases span from low correlation to high correlation. However, generally, the correlation between community-aware centrality measures ranges from low to medium. The values can mainly be negative when Modularity Vitality is involved or positive for the other community-aware centrality measures, which target mainly bridges.

## 5.2 Influence of the community structure strength on correlation

The previous experiment shows that one can distinguish three groups of networks according to the visual similarity of their heatmaps. To further this analysis, we associate each network to a vector made of its set of Kendall's Tau correlation values. Afterward, we compute the Pearson correlation between the couples of vectors.

Figure 4 reports the heatmap of the Pearson correlation between the fifteen networks and its violin plot. Networks are ordered based on their mean correlation with other networks in increasing order. One can see that EU Airlines, U.S. Airports, and Ego Facebook are the least correlated with other networks, while Caltech, DeezerEU, and Yeast Protein are the most correlated. Inspecting the violin plot shows that predominantly the networks correlate well (with a mean of  $0.77 \pm 0.12$ ). Nevertheless, the left tail of the distribution indicates the presence of low correlation values.

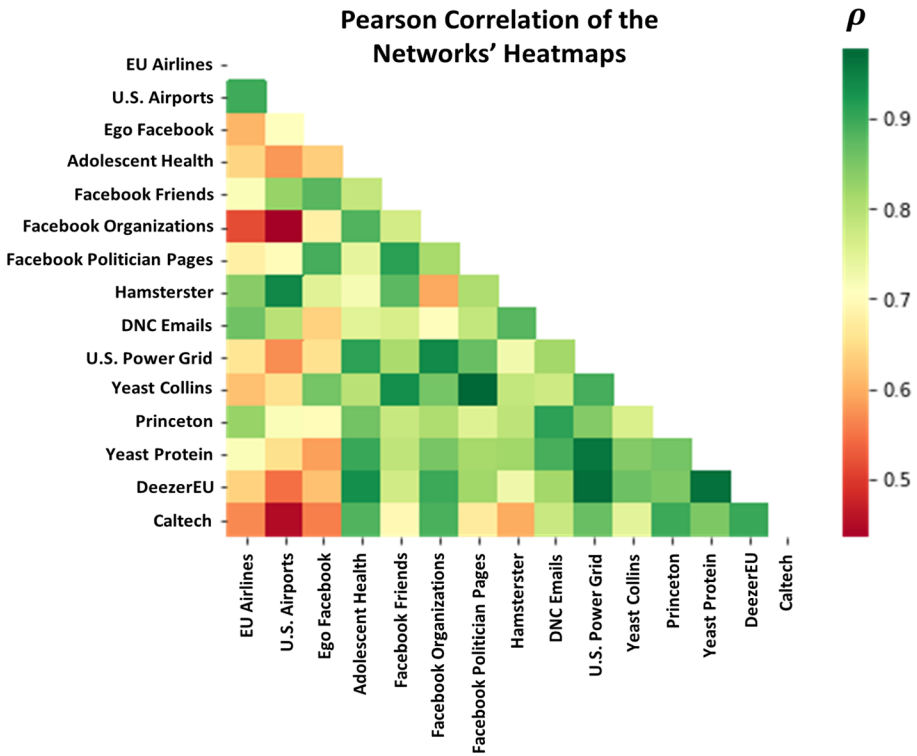
A closer inspection of the network topological properties shows that networks that do not correlate well with the others have a strong community structure. In contrast, the most correlated to their pairs have a medium to weak community structure. It suggests that the community structure strength is playing a significant role in this behavior.

We perform a simple linear regression to relate the community structure strength with the mean correlation value associated with each network. Note that we discard Modularity Vitality in the mean calculation. Indeed, it exhibits a negative correlation with the other community-aware centrality measures. Hence, it may cancel out the contributions of the other community-aware centrality measures to the final mean correlation. The linear regression process uses ordinary least squares estimators. The mean values are the dependent variables, and the mixing parameter quantifying the community structure strength is the independent variable. The relationship between the dependent and independent variables is statistically significant when the  $p$ -value is below a threshold value. Figure 5 reports the results of the linear regression. The linear relation hypothesis is statistically significant ( $p < 0.01$ ). The predicted slope is positive (+0.557). In other words, as the mixing parameter increases (i.e., as the network community structure gets weaker), the mean correlation tends to increase.

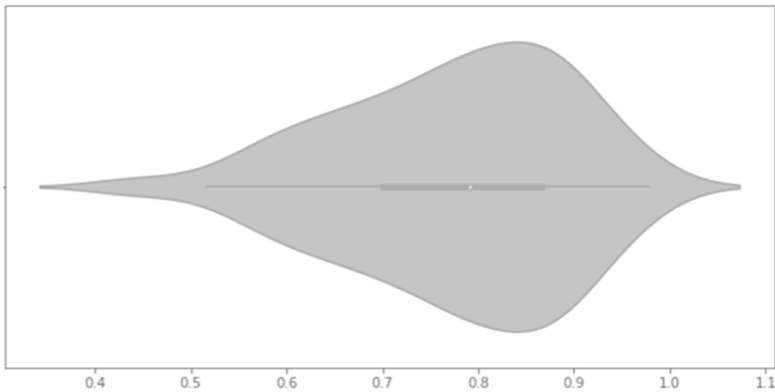
We also perform a linear regression analysis to study the relationship between the number of communities in each network and the correlation of community-aware centrality measures. As results show no significant relationship, they are not reported here.

## 5.3 Diffusion effectiveness of the community-aware centrality measures

The diffusion performance of the community-aware centrality measures is studied using the SIR simulation process. One can consider any centrality measure as a baseline. In this study, we decide to use the classical degree centrality. Indeed, all the community measures are correlated with degree centrality, which is the sum of the total intra-community and inter-community links of the node ( $k_i^{tot} = k_i^{intra} + k_i^{inter}$ ). Nonetheless, they are more complex as they exploit the mesoscopic information distinguishing hubs and bridges. Thus, a negative value indicates that the classical degree centrality is more efficient than the community-aware centrality measure under test to diffuse information in the network. In contrast, a positive value illustrates the benefits of community-aware centrality measures over the classical degree centrality measure.



**Distribution of the Pearson Correlation of the Networks' Heatmaps**

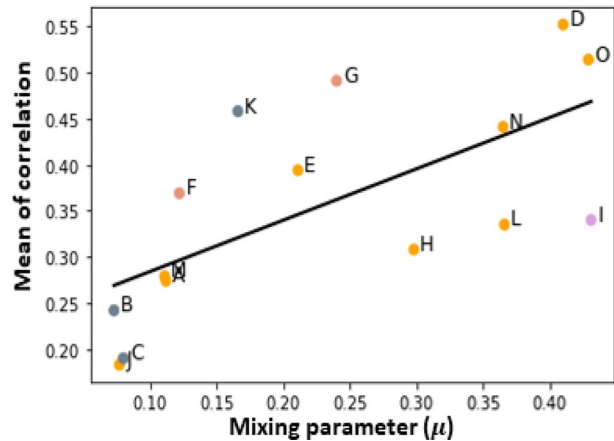


**Fig. 4** Top figure: Pearson Correlation ( $\rho$ ) between networks based on their Kendall's Tau correlation values where each network is associated to its Kendall's Tau correlation vector. Bottom figure: Violin plot of the Pearson correlation

To present the results of these experiments, we choose to consider three categories of networks according to the community structure strength. The first category regroups the networks with a strong community structure ( $\mu \leq 0.084$ ). It contains Ego Facebook, EU Airlines, and U.S. Airports. Networks with a medium community structure



**Fig. 5** Linear regression between the mean of correlation of the community-aware centrality measures for each network and their mixing parameter ( $\mu$ ). Ordinary least squares estimates are used



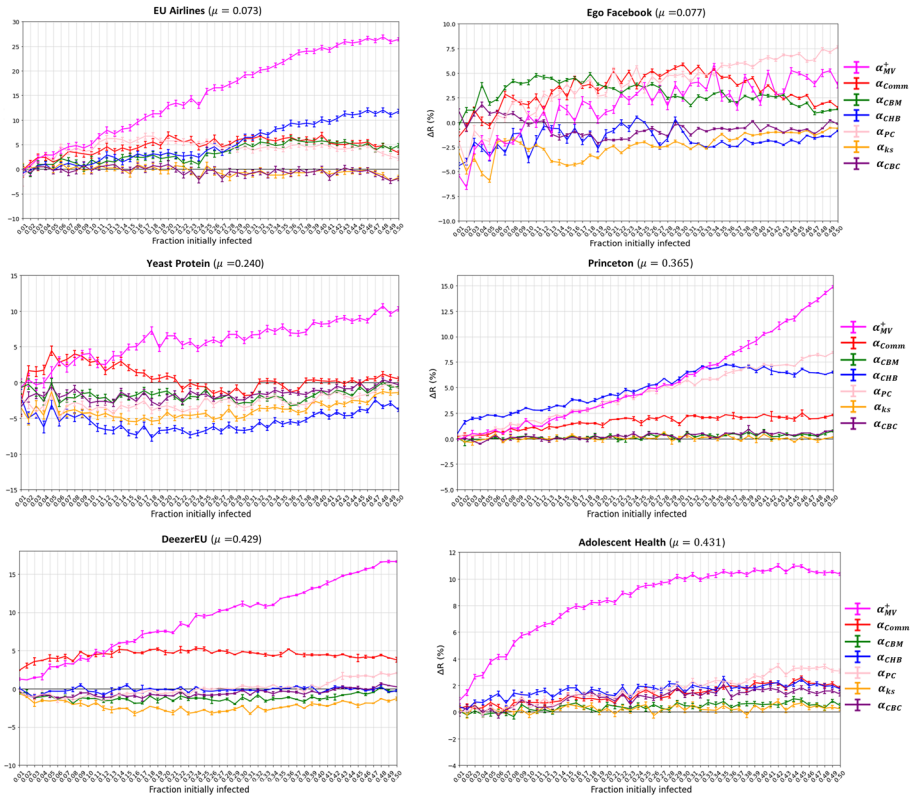
strength ( $0.084 < \mu \leq 0.366$ ) make the second category. It is the biggest one with Facebook Politician Pages, Facebook Friends, Yeast Collins, U.S. Power Grid, Princeton, DNC Emails, Yeast Protein, Hamsterster, and Facebook Organizations. Finally, the networks with a weak community structure ( $\mu \geq 0.410$ ) form the last category. It contains Caltech, Deezer EU, and Adolescent Health. Figure 6 shows the relative difference of the outbreak size ( $\Delta R$ ) as a function of the fraction of initially infected nodes for six networks belonging to the three categories. Results for the other networks are available in the supplementary material.

### 5.3.1 Networks with a strong community structure

Two typical behaviors emerge in networks with a strong community structure. The first is illustrated by EU Airlines and Ego Facebook represents the second. In EU Airlines, Modularity Vitality outperforms the rest of the centrality measures. Its gain over the baseline reaches 27%. It is 15% higher than Community Hub-Bridge, which ranks second. The U.S. Airports network shows similar behavior. Indeed, Modularity Vitality has the highest epidemic outbreak. Ego Facebook is a typical illustration of the second behavior. There is no single winner. The relative effectiveness of the community-aware centrality measures is tied to the proportion of initially infected nodes. In Ego Facebook, Community-based Mediator exhibits the highest outbreak difference for a low range of initially infected nodes. Then Comm Centrality takes the lead in the medium range. Finally, Participation Coefficient performs better in the last interval.

### 5.3.2 Networks with a medium community structure

One can distinguish two cases in networks with a medium community structure strength. Yeast Protein network illustrates the first. It shows that at a small fraction of initially infected nodes, Comm Centrality outperforms the other community-aware centrality measures. Then Modularity Vitality takes over. Facebook Politician Pages, Facebook Friends, U.S. Power Grid, and Yeast Collins follow this behavior. The gain of Modularity Vitality over the baseline can reach up to 19% (in Facebook Friends). Note that Comm Centrality stays the most performing at all fraction of initially infected nodes for the Yeast Collins



**Fig. 6** Relative difference of the outbreak size ( $\Delta R$ ) as a function of the fraction of initially infected nodes for three real-world networks. The initial spreaders set is build according to the ranks associated to a given community-aware centrality measure. The community-aware centrality measures are: Modularity Vitality targeting hubs =  $\alpha_{MV}^+$ , Comm Centrality =  $\alpha_{Comm}$ , Community-based Mediator =  $\alpha_{CBM}$ , Community Hub-Bridge =  $\alpha_{CHB}$ , Participation Coefficient =  $\alpha_{PC}$ , K-shell with Community =  $\alpha_{ks}$ , and Community-based Centrality =  $\alpha_{CBC}$

network. Princeton represents the second case. It shows that at first, Community Hub-Bridge has the highest epidemic outbreak, and then Modularity Vitality which targets hubs, takes over. One observes a similar behavior in DNC Emails and Hamsterster. It is worth noticing that Modularity Vitality is the most effective at all fraction of initially infected nodes in Facebook Organizations network.

### 5.3.3 Networks with a weak community structure

One can also notice two typical behaviors in networks with a weak community structure strength. The first one illustrated with DeezerEU shows that at a small fraction of initially infected nodes, Comm Centrality outperforms its alternatives. Then, Modularity Vitality takes over. Its gain reaches 17%. It is 13% higher than Comm Centrality that ranks second. Adolescent Health illustrates the second case. Modularity Vitality outperforms the other community-aware centrality measures at all fractions of initially infected nodes. Caltech also shows similar behavior.

## 6 Discussion

This study conducts a systematic evaluation of prominent community-aware centrality measures using a set of fifteen real-world networks spanning a wide variety of complex network structural patterns. First, we compute Kendall's Tau correlation between the community-aware centrality measures on each network. Results indicate that the correlation spans from low to medium. Moreover, one can arrange the community-aware centrality measures in two groups. The first group contains Modularity Vitality. A negative correlation with all the other community-aware centrality measures characterizes it. The remaining centrality measures form the second group. Correlation between these centrality measures ranges from low to medium positive values. These two groups seem to reflect the ability of community-aware centrality measures to prioritize either hubs or bridges. It all depends on how they integrate the intra-community and inter-community links.

To investigate the variation across networks of the correlation pattern between community-aware centrality measures, we calculate the Pearson correlation of the networks' heatmaps. It appears that networks with a strong community structure do not correlate well with the others. In addition, the ones correlating the most with the others have a medium or weak community structure. A linear regression analysis corroborates these findings.

We show that the community structure strength is a crucial feature driving the correlation variation between community-aware centrality measures. In networks with a strong community structure, community-aware centrality measures do not correlate well. Correlation between the centrality measures increases when the community structure gets weaker. Indeed, the weaker the community structure, the fewer differences between the various community-aware centrality measures. This result is quite intuitive. For example, consider Comm Centrality and Community-based Mediator. The first favors bridges against hubs by raising the inter-community links proportion to the power of 2. The second exploits the entropy of the intra-community and inter-community connections. If a network has a well-defined community structure, these two measures select different "bridges." Indeed, in Comm Centrality, intra-community links might not play a significant role in Community-based Mediator, which is affected by a simple change in the latter. Now consider a network with a weak community structure. In this case, a node's local and global influence are less diverse since many nodes from different communities are connected. Consequently, the Comm Centrality and Community-based Mediator extract on average more similar information than in a network with a strong community structure. Note that these results are in the same vein that findings reported in previous works (Rajeh et al. 2021a, c). Extensive investigation has shown that the correlation variation between classical and community-aware centrality measures is mainly related to the community structure strength.

We compare the diffusion ability of the various community-aware centrality measures using SIR simulations. Results show that Modularity Vitality obtains the most significant outbreak difference with the baseline in networks with strong, medium, and strong community structure. However, it may take not take the lead when a small fraction of nodes is initially infected. Note that Modularity Vitality used in this study targets hubs. The low correlation values observed with the other community-aware centrality measures suggest that they mainly focus on bridges.

To further compare the distribution of the top nodes extracted by Modularity Vitality to its alternatives, we visualize the six representative networks of Fig. 6 with their communities reported in different colors. The big nodes represent the fraction of top nodes initially infected based on their respective community-aware centrality measure.

Figures 7 and 8, represent respectively EU Airlines and Ego Facebook. These networks have a strong community structure. In EU Airlines, Modularity Vitality outperforms the other measures with a small fraction of initially infected nodes ( $f_o = 1\%$ ). One can see in Fig. 7 representing the EU Airlines network that Modularity Vitality neglects hubs at the core of the largest community (purple). Instead, it targets hubs in all the other small communities. This remark still holds when the fraction of initially infected nodes grows ( $f_o = 25\%$ ). Indeed, the few nodes selected in the purple community are in its periphery in this situation. Modularity Vitality does not favor hubs in the purple community since their removal does not severely impact the modularity of the network. In other words, since the purple community is very dense, removing a hub from another community has a higher effect on modularity.



(a) Nodes infected by K-shell With Community at  $f_o = 1\%$

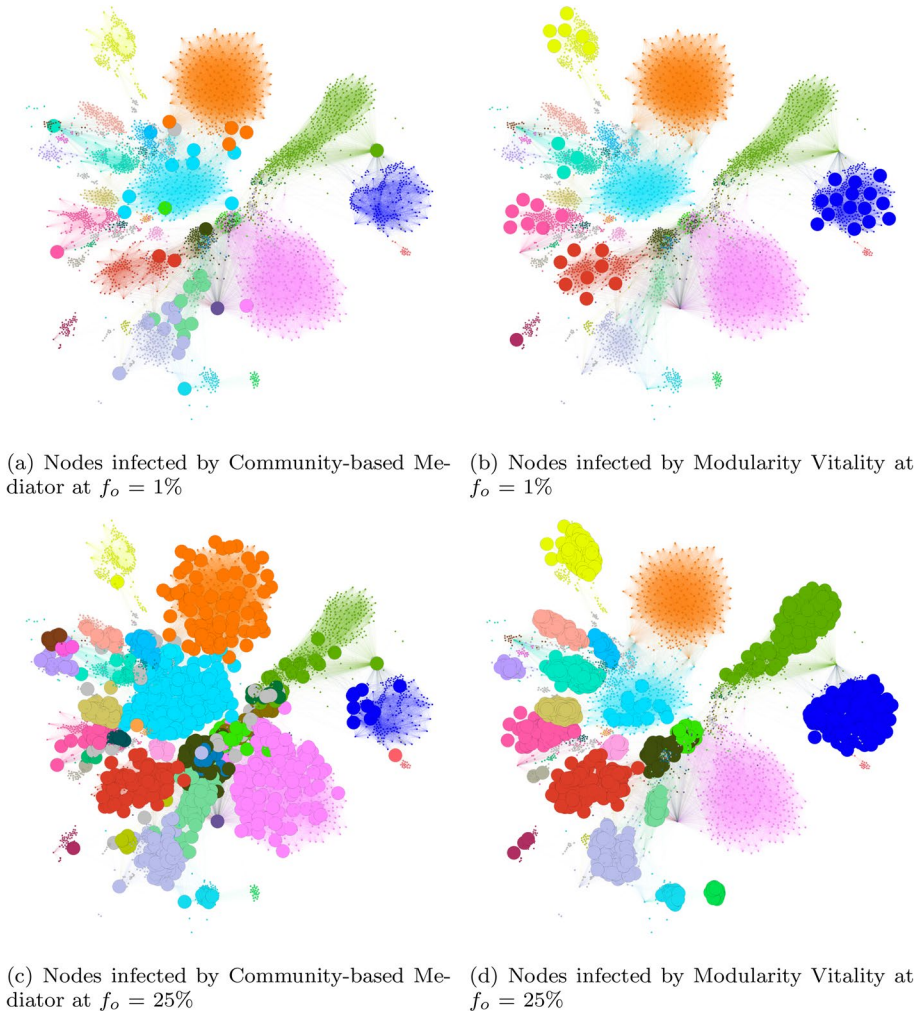
(b) Nodes infected by Modularity Vitality at  $f_o = 1\%$



(c) Nodes infected by K-shell with Community at  $f_o = 25\%$

(d) Nodes infected by Modularity Vitality at  $f_o = 25\%$

**Fig. 7** Comparing the position of the nodes infected at low availability of resources, at  $f_o = 1\%$  (top figures) and at high availability of resources, at  $f_o = 25\%$  (bottom figures) in the EU Airlines network. The bigger nodes in the left figure are the nodes picked by K-shell with Community and the bigger nodes in right figure are the nodes picked by Modularity Vitality



**Fig. 8** Comparing the position of the nodes infected at low availability of resources, at  $f_o = 1\%$  (top figures) and at high availability of resources, at  $f_o = 25\%$  (bottom figures) in the Ego Facebook network. The bigger nodes in the left figure are the nodes picked by Community-based Mediator and the bigger nodes in right figure are the nodes picked by Modularity Vitality

In contrast, the left figure shows nodes chosen by K-shell with Community. One can see that, whatever the value of the fraction of initially infected nodes, all the extracted nodes are at the core of the purple community. Consequently, K-shell with Community exhibits poor performance. Indeed, the epidemic does not reach the other communities and the nodes located at the network's periphery. Overall, the more scattered the top centrality nodes, the higher the probability that the infection reaches the various part of the network.

Although the Ego Facebook network is also a network with a strong community structure, its topology is quite different than EU Airlines. Indeed, the size of the communities is more balanced, while in EU Airlines, it is more nonhomogeneous. In this situation,

the hubs chosen by Modularity Vitality, even though they are from different communities, don't result in the highest epidemic outbreak. Indeed, there is no single winner in Ego Facebook. First, Community-based Mediator leads to the highest epidemic outbreak. It is followed by Comm Centrality and by Participation Coefficient. Consider the upper subfigures of Fig. 8 where 1% of the nodes are infected. One can see that Community-based Mediator selects nodes located at the borders of the communities. In contrast, Modularity Vitality chooses hubs in small communities at the expense of those in large communities. The epidemic spreading is more effective with Community-based Mediator in this case. The same behavior persists with more initially infected nodes (bottom subfigures). Modularity Vitality neglects the pink and orange communities, while Community-based Mediator infects all the communities. Overall, Modularity Vitality favors hubs in smaller communities since their removal has a higher impact on the network's modularity. Consequently, as it neglects larger communities with numerous hubs, an infection may not reach them.

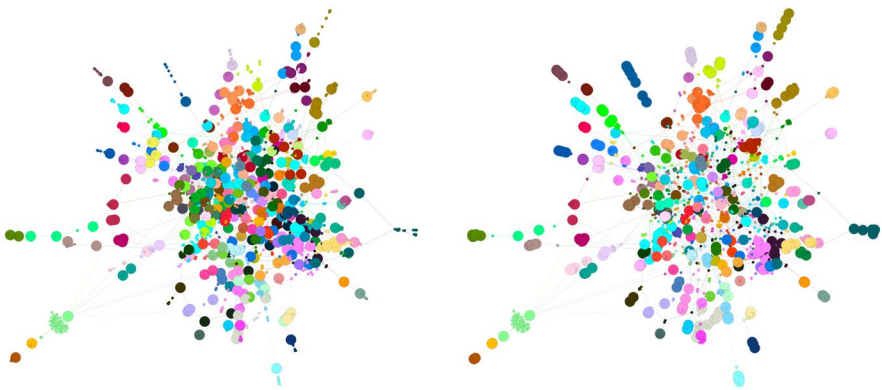
Figures 9 and 10 represent Yeast Protein and Princeton networks, respectively. These networks have a medium community structure strength. In Yeast Protein, Comm Centrality performs slightly better than Modularity Vitality with a low proportion of initially infected nodes ( $f_o = 1\%$ ). As this proportion increases, Modularity Vitality takes the lead. One can see that Modularity Vitality covers more communities at  $f_o = 25\%$ . Indeed, Modularity Vitality infects nodes located in the network periphery in numerous small communities. In return, the epidemic outbreak reaches parts of the network that are not easily reachable from the core communities. Concerning the Princeton network, Community Hub-Bridge has a slightly higher epidemic outbreak than Modularity Vitality and Participation Coefficient till  $f_o = 33\%$ . After that, Modularity Vitality takes over. We can see that Community Hub-Bridge selects nodes in the green community, while Modularity Vitality picks most nodes in the blue community. Nevertheless, even with a small fraction of infected nodes, it covers more communities. In the end, Modularity Vitality leads to a higher epidemic outbreak as the fraction of initially infected nodes grows. One can see that initially infected nodes surround the neglected green community. Thus, infections will most likely enter it through different nodes.

The limiting case of a network with a weak community structure is a network with no community structure. In this situation, the influence of the community structure is limited. The best strategy to reach the various zones of the network is to target the most distant nodes. Indeed, this is what Modularity Vitality shows in the networks DeezerEU and Adolescent Health, illustrated in Figures 11 and 12, respectively. In DeezerEU, at a small infection rate ( $f_o = 1\%$ ), the nodes selected by Comm Centrality are more distant compared to Modularity Vitality. It results in a better performance. At  $f_o = 25\%$ , nodes chosen by Modularity Vitality are more evenly distributed across the network, resulting in higher performance. One observes a similar behavior in the Adolescent Health network illustrated in Fig. 12. In this case, Modularity Vitality outperforms the other measures, whatever the proportion of initially infected nodes.

To summarize, Modularity Vitality achieves the highest epidemic outbreak when there is a fair proportion of initially infected nodes. If there are not enough resources, it tends to underestimate large communities densely connected. Its performance is more pronounced as the community structure gets weaker. It is due to its natural behavior of selecting hubs that are distant from each other. It tends to pick hubs in communities containing few hubs since their removal significantly affects the network's modularity. Nonetheless, targeting a small number of hubs is not enough below a specific fraction of infections (or resources). Here, it is better to target bridges, especially if the network has a strong or medium community structure strength. Indeed, there are enough external connections between



(a) Nodes infected by Comm Centrality at  $f_o = 1\%$  (b) Nodes infected by Modularity Vitality at  $f_o = 1\%$

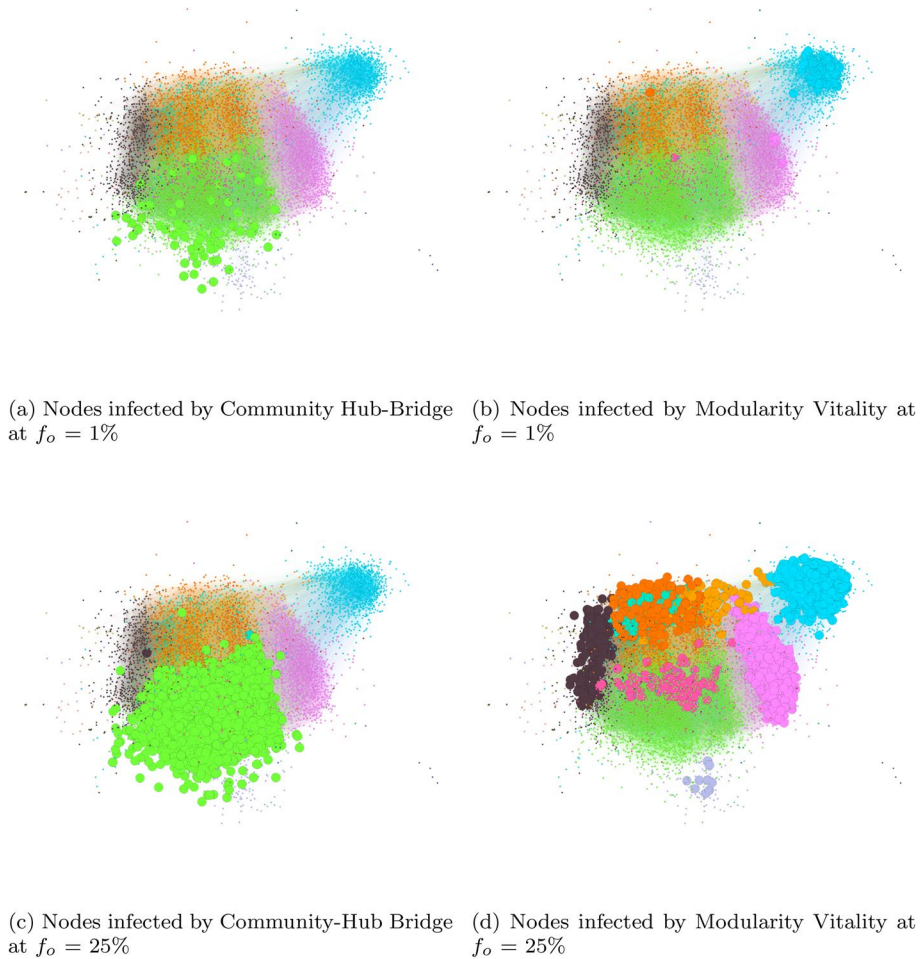


(c) Nodes infected by Comm Centrality at  $f_o = 25\%$  (d) Nodes infected by Modularity Vitality at  $f_o = 25\%$

**Fig. 9** Comparing the position of the nodes infected at low availability of resources, at  $f_o = 1\%$  (top figures) and at high availability of resources, at  $f_o = 25\%$  (bottom figures) in the Yeast Protein network. The bigger nodes in the left figure are the nodes picked by Comm Centrality and the bigger nodes in right figure are the nodes picked by Modularity Vitality

communities in these networks to diffuse in all the communities of the network. Hence, infected bridges spread the epidemic to communities more effectively than hubs.

It is worth noticing that a couple of community-aware centrality measures don't perform better than the baseline. It is evident with K-shell with Community in Ego Facebook, U.S. Airports, Facebook Politician Pages, Facebook Friends, U.S. Power Grid, Yeast Protein, Facebook Organizations, and DeezerEU. To a lesser extent, it applies to Community Hub-Bridge and Community-based Centrality. Sometimes, community-aware centrality measures perform poorly, even though they incorporate relevant information about the community structure. For example, in K-shell with Community, many nodes share the same  $k$ -shell (or  $k$ -core) value. This behavior is illustrated by K-shell with Community in the EU Airlines network in Fig. 7. Therefore, infecting one or the other on the same level may not

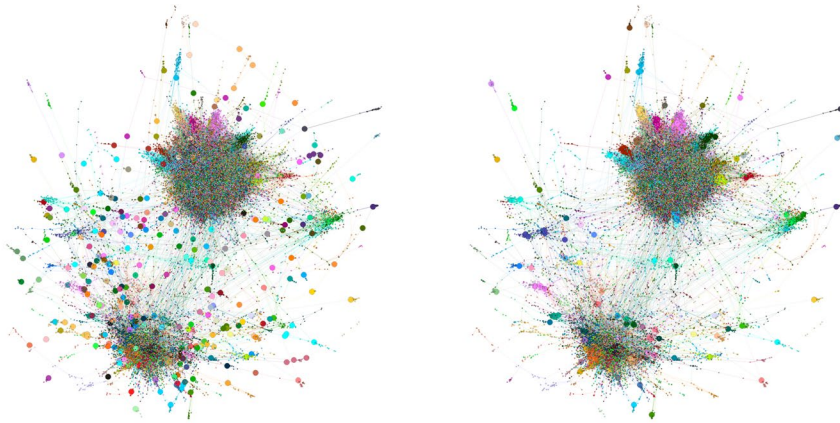


**Fig. 10** Comparing the position of the nodes infected at low availability of resources, at  $f_o = 1\%$  (top figures) and at high availability of resources, at  $f_o = 25\%$  (bottom figures) in the Princeton network. The bigger nodes in the left figure are the nodes picked by Community Hub-Bridge and the bigger nodes in right figure are the nodes picked by Modularity Vitality

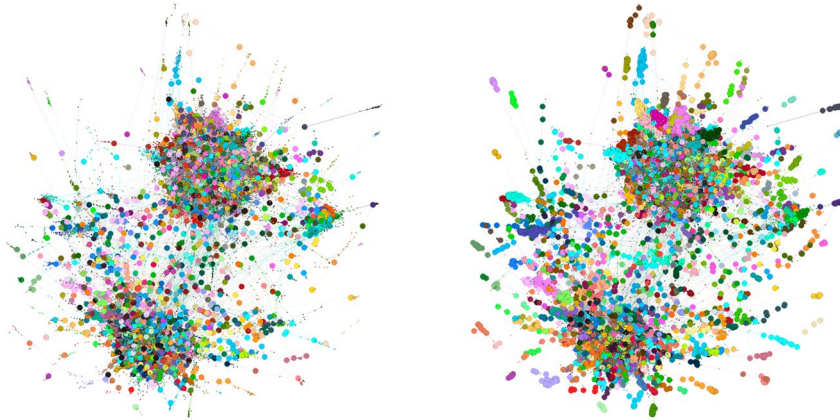
lead to a higher epidemic spreading than infecting more distant nodes (i.e., not on the same level). Similarly is the case for Community-based Centrality, which shrinks the inter-community and intra-community links based on their community sizes. So, many nodes receive very similar centrality values. Moreover, these nodes may be close to each other. Note that this is only true when one considers a multiple-spreader problem where several nodes are infected simultaneously. Nonetheless, if the infection starts from a single node, nodes that are the most embedded in the network may result in the highest epidemic outbreak (Rajeh et al. 2021b; Kitsak et al. 2010). For more insights on the single-spreader problem versus the multiple-spreader problem, one can consult the following literature Everett and Borgatti (1999, 2005).

Finally, we compute the evolution of the size of the largest connected component (LCC) when removing a given fraction of top-ranked nodes using the various





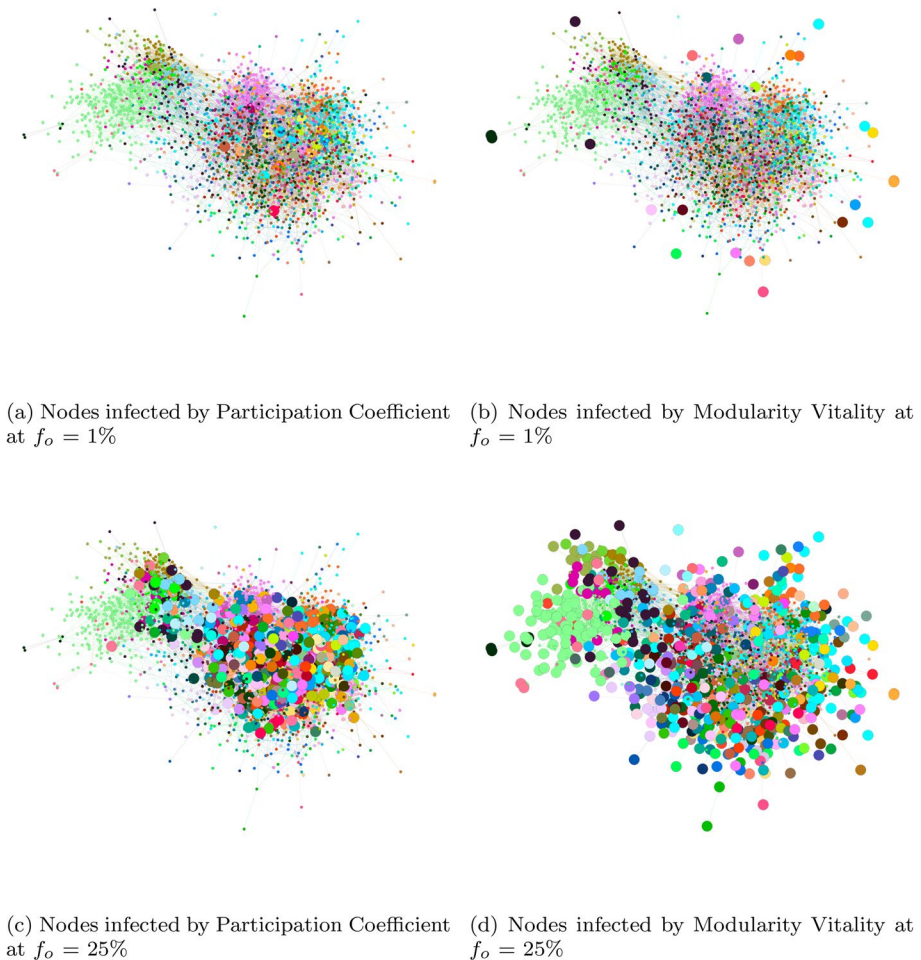
(a) Nodes infected by Comm Centrality at  $f_o = 1\%$  (b) Nodes infected by Modularity Vitality at  $f_o = 1\%$



(c) Nodes infected by Comm Centrality at  $f_o = 25\%$  (d) Nodes infected by Modularity Vitality at  $f_o = 25\%$

**Fig. 11** Comparing the position of the nodes infected at low availability of resources, at  $f_o = 1\%$  (top figures) and at high availability of resources, at  $f_o = 25\%$  (bottom figures) in the DeezerEU network. The bigger nodes in the left figure are the nodes picked by Comm Centrality and the bigger nodes in right figure are the nodes picked by Modularity Vitality

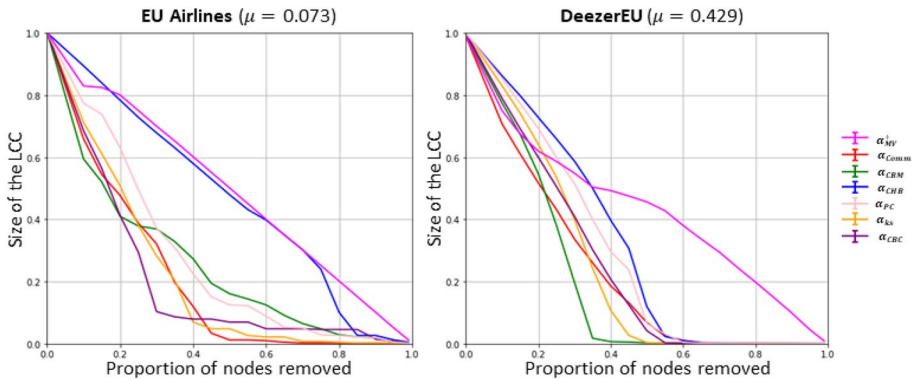
community-aware centrality measures. Figure 13 illustrates the general trend observed in two networks (EU Airlines and DeezerEU). According to the LCC, Modularity Vitality performs poorly in dismantling the network in both cases. In contrast, considering a diffusion process based on the SIR model (Fig. 6), Modularity Vitality results in the highest epidemic outbreak once a specific fraction of initially infected nodes is reached. One should be cautious when looking at these results, which may seem contradictory. Remember that the LCC is a conservative upper bound corresponding to the worst-case scenario with no information about the diffusion process. In this situation, strategies targeting bridges such as Community-based Mediator or nodes embedded in the large



**Fig. 12** Comparing the position of the nodes infected at low availability of resources, at  $f_o = 1\%$  (top figures) and at high availability of resources, at  $f_o = 25\%$  (bottom figures) in the Adolescent Health network. The bigger nodes in the left figure are the nodes picked by Participation Coefficient and the bigger nodes in right figure are the nodes picked by Modularity Vitality

communities using Community-based Centrality are more suitable strategies for efficient network dismantling.

Remember that Modularity Vitality targets hubs in small communities in priority because they induce the highest variation of modularity. The largest communities are affected in the SIR diffusion process when the fraction of initially infected nodes is high enough. However, the size of the LCC is barely affected compared to the other measures prioritizing bridges. Consequently, as the most prominent communities are not involved in the early stage of the infection process, the size of the LCC remains significant even if the epidemic diffusion is contained. Note that if Modularity Vitality uses the bridges-first ranking scheme, networks may disintegrate much quicker (Magelinski et al. 2021). These results demonstrate that one should be cautious in using the LCC to compare the



**Fig. 13** The size of largest connected component (LCC) of the EU Airlines and DeezerEU networks. The community-aware centrality measures are: Modularity Vitality targeting hubs =  $\alpha_{MV}^+$ , Comm Centrality =  $\alpha_{Comm}$ , Community-based Mediator =  $\alpha_{CBM}$ , Community Hub-Bridge =  $\alpha_{CHB}$ , Participation Coefficient =  $\alpha_{PC}$ , K-shell with Community =  $\alpha_{CS}$ , and Community-based Centrality =  $\alpha_{CBC}$

performance of centrality measures. If the diffusion process is known, one should use it for a more realistic investigation.

## 7 Conclusion

Identifying influential nodes in a network has always attracted scientists due to its broad theoretical and applicative dimensions. Several previous studies investigate the relations between classical centrality measures. In contrast, this work focuses on the interplay between popular community-aware centrality measures using real-world networks from various domains and diverse topological characteristics.

Results show that the correlation between the community-aware centrality measures ranges from negative to positive. Modularity Vitality exhibits low negative correlation values with the other community-aware centrality measures. It differs from the others mainly because it is a signed centrality. Furthermore, it explicitly targets hubs. The other community-aware centrality measures target bridges or hubs and bridges simultaneously. Their correlation is positive. It generally ranges from weak positive to medium positive values. The community structure strength affects the correlation pattern. The stronger the community structure strength, the lower the correlation. Conversely, networks with weak community structures exhibit higher correlation.

If enough resources are available, SIR simulations suggest targeting hubs using Modularity Vitality instead of targeting bridges. This strategy is more effective since it extracts from different distant communities. It is even more effective in networks with a weak community structure. Indeed, rather than selecting nodes in the network core, it allows covering the various modules of the network and leads to a higher epidemic outbreak. In networks with strong to medium community structures, the choice of a community-aware centrality measure is more dependent on the resources at hand. It is better to use Comm Centrality that targets bridges rather than hubs when resources are limited. Indeed, in these situations, the epidemic spreads into the communities through inter-community links between communities.

These findings have multiple implications. First, one can choose the adequate measure knowing the network's community structure strength and the availability resources at hand. Second, they open new perspectives in the design of community-aware centrality measures adapted to specific networks' mesoscopic characteristics. Third, they pave the way for future investigations on the interplay between various diffusion models and the networks' mesoscopic topological properties.

**Supplementary Information** The online version contains supplementary material available at <https://doi.org/10.1007/s11135-022-01416-7>.

**Author contributions** All authors have contributed to this work. SR implemented the models and the analyses. All authors participated in the formulation and writing of this paper. All authors approved the final manuscript.

**Funding** Not applicable.

**Data availability** The datasets used in this article are all freely accessible in the cited resources.

**Code availability** The code of the following study is accessible via GitHub: <https://github.com/StephanyRajeh/CompAnalysisCACM>

## Declarations

**Conflict of interest** Author HC has served on the editorial board of *Quality and Quantity*. The authors declare that they have no competing interests.

**Consent to participate** Not applicable.

**Consent for publication** Not applicable.

**Ethical approval** Not applicable.

## References

- Anderson, R.M., May, R.M.: Population biology of infectious diseases: part I. *Nature* **280**(5721), 361–367 (1979)
- Berahmand, K., Bouyer, A., Samadi, N.: A new local and multidimensional ranking measure to detect spreaders in social networks. *Computing* **101**(11), 1711–1733 (2019)
- Buckee, C., Noor, A., Sattenspiel, L.: Thinking clearly about social aspects of infectious disease transmission. *Nature* **595**(7866), 205–213 (2021)
- Cherifi, H., Palla, G., Szymanski, B.K., Lu, X.: On community structure in complex networks: challenges and opportunities. *Appl. Netw. Sci.* **4**(1), 1–35 (2019)
- Dorogovtsev, S.N., Goltsev, A.V., Mendes, J.F.F.: K-core organization of complex networks. *Phys. Rev. Lett.* **96**(4), 040601 (2006)
- Everett, M.G., Borgatti, S.P.: The centrality of groups and classes. *J. Math. Sociol.* **23**(3), 181–201 (1999)
- Everett, M.G., Borgatti, S.P.: Extending centrality. *Models Methods Soc. Netw. Anal.* **35**(1), 57–76 (2005)
- Ghalmane, Z., El Hassouni, M., Cherifi, H.: Immunization of networks with non-overlapping community structure. *SNAM* **9**(1), 1–22 (2019)
- Girvan, M., Newman, M.E.: Community structure in social and biological networks. *PNAS* **99**(12), 7821–7826 (2002)
- Guimera, R., Amaral, L.A.N.: Functional cartography of complex metabolic networks. *Nature* **433**(7028), 895–900 (2005)
- Gupta, N., Singh, A., Cherifi, H.: Centrality measures for networks with community structure. *Physica A* **452**, 46–59 (2016)

- Ismail, S.S., Akil, K.A.K., Chulan, M., Sharif, N.: The susceptible-infected-recovered (SIR) model for viral marketing. In: AIP Conference Proceedings, 1, p. 030018. AIP Publishing LLC (2017)
- Jebabli, M., Cherifi, H., Cherifi, C., Hamouda, A.: Community detection algorithm evaluation with ground-truth data. *Physica A* **492**, 651–706 (2018)
- Kitsak, M., Gallos, L.K., Havlin, S., Liljeros, F., Muchnik, L., Stanley, H.E., Makse, H.A.: Identification of influential spreaders in complex networks. *Nat. Phys.* **6**(11), 888–893 (2010)
- Kunegis, J.: Handbook of network analysis [konekt—the koblenz network collection] (2014). [arXiv:1402.5500](https://arxiv.org/abs/1402.5500)
- Lancichinetti, A., Kivela, M., Saramaki, J., Fortunato, S.: Characterizing the community structure of complex networks. *PLoS ONE* **5**(8), e11976 (2010)
- Li, C., Li, Q., Van Mieghem, P., Stanley, H.E., Wang, H.: Correlation between centrality metrics and their application to the opinion model. *Eur. Phys. J. B* **88**(3), 1–13 (2015)
- Lü, L., Chen, D., Ren, X.L., Zhang, Q.M., Zhang, Y.C., Zhou, T.: Vital nodes identification in complex networks. *Phys. Rep.* **650**, 1–63 (2016)
- Luo, S.L., Gong, K., Kang, L.: Identifying influential spreaders of epidemics on community networks. [arXiv:1601.07700](https://arxiv.org/abs/1601.07700) (2016)
- Magelinski, T., Bartulovic, M., M. Carley, K.: Measuring node contribution to community structure with modularity vitality. *IEEE Trans. Netw. Sci. Eng.* **8**(1), 707–723 (2021)
- Nematzadeh, A., Ferrara, E., Flammini, A., Ahn, Y.Y.: Optimal network modularity for information diffusion. *Phys. Rev. Lett.* **113**(8), 088701 (2014)
- Newman, M.E.: Modularity and community structure in networks. *Proc. Natl. Acad. Sci.* **103**(23), 8577–8582 (2006)
- Nguyen, B.: Modelling cyber vulnerability using epidemic models. In: SIMULTECH, pp. 232–239 (2017)
- Oldham, S., Fulcher, B., Parkes, L., Arnatkeviciūtė, A., Suo, C., Fornito, A.: Consistency and differences between centrality measures across distinct classes of networks. *PLoS ONE* **14**(7) (2019)
- Orman, G.K., Labatut, V., Cherifi, H.: On accuracy of community structure discovery algorithms (2011). [arXiv:1112.4134](https://arxiv.org/abs/1112.4134)
- Orman, G.K., Labatut, V., Cherifi, H.: Comparative evaluation of community detection algorithms: a topological approach. *J. Stat. Mech.* **2012**(08), P08001 (2012)
- Orman, K., Labatut, V., Cherifi, H.: An empirical study of the relation between community structure and transitivity. In: *Complex Networks*, pp. 99–110. Springer, Berlin (2013)
- Peixoto, T.P.: The netzschleuder network catalogue and repository (2020). <https://networks.skewed.de/>
- Rajeh, S., Savonnet, M., Leclercq, E., Cherifi, H.: Interplay between hierarchy and centrality in complex networks. *IEEE Access* **8**, 129717–129742 (2020a)
- Rajeh, S., Savonnet, M., Leclercq, E., Cherifi, H.: Investigating centrality measures in social networks with community structure. In: *International Conference on Complex Networks and Their Applications*, pp. 211–222. Springer, Berlin (2020b)
- Rajeh, S., Savonnet, M., Leclercq, E., Cherifi, H.: Characterizing the interactions between classical and community-aware centrality measures in complex networks. *Sci. Rep.* **11**(1), 1–15 (2021a)
- Rajeh, S., Savonnet, M., Leclercq, E., Cherifi, H.: Comparing community-aware centrality measures in online social networks. In: *International Conference on Computational Data and Social Networks*, pp. 279–290. Springer, Berlin (2021b)
- Rajeh, S., Savonnet, M., Leclercq, E., Cherifi, H.: How correlated are community-aware and classical centrality measures in complex networks? In: *International Conference on Complex Networks*, pp. 120–132. Springer, Berlin (2021c)
- Ronqui, J.R.F., Travieso, G.: Analyzing complex networks through correlations in centrality measurements. *J. Stat. Mech.* **2015**(5), P05030 (2015)
- Rossi, R.A., Ahmed, N.K.: The network data repository with interactive graph analytics and visualization. In: *AAAI* (2015)
- Rosvall, M., Bergstrom, C.T.: Maps of random walks on complex networks reveal community structure. *PNAS* **105**(4), 1118–1123 (2008)
- Rozenberczki, B., Sarkar, R.: Characteristic functions on graphs: birds of a feather, from statistical descriptors to parametric models (2020)
- Schoch, D., Valente, T.W., Brandes, U.: Correlations among centrality indices and a class of uniquely ranked graphs. *Soc. Netw.* **50**, 46–54 (2017)
- Toda, A.A.: Susceptible-infected-recovered (SIR) dynamics of covid-19 and economic impact (2020). [arXiv:2003.11221](https://arxiv.org/abs/2003.11221)
- Tulu, M.M., Hou, R., Younas, T.: Identifying influential nodes based on community structure to speed up the dissemination of information in complex network. *IEEE Access* **6**, 7390–7401 (2018)

- Wang, G.X., Qin, T.G.: Impact of community structure on network efficiency and communicability. In: 2010 International Conference on Intelligent Computation Technology and Automation, vol. 2, pp. 485–488. IEEE (2010)
- Wang, W., Liu, Q.H., Zhong, L.F., Tang, M., Gao, H., Stanley, H.E.: Predicting the epidemic threshold of the susceptible-infected-recovered model. *Sci. Rep.* **6**(1), 1–12 (2016)
- Wharrie, S., Azizi, L., Altmann, E.G.: Micro-, meso-, macroscales: the effect of triangles on communities in networks. *Phys. Rev. E* **100**(2), 022315 (2019)
- Yang, Z., Algesheimer, R., Tessone, C.J.: A comparative analysis of community detection algorithms on artificial networks. *Sci. Rep.* **6**(1), 30750 (2016). <https://doi.org/10.1038/srep30750>
- Zhao, Z., Wang, X., Zhang, W., Zhu, Z.: A community-based approach to identifying influential spreaders. *Entropy* **17**(4), 2228–2252 (2015)

**Publisher's Note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.