

Spreading Power of Key Nodes in Online Social Networks with Community Structure

STEPHANY RAJEH, MARINETTE SAVONNET, ERIC LECLERCQ, and HOCINE CHERIFI, University of Burgundy, France

Identifying key nodes is crucial for accelerating or impeding dynamic spreading in a network. Community-aware centrality measures tackle this problem by exploiting the community structure of a network. Although there is a growing trend to design new community-aware centrality measures, there is no systematic investigation of the proposed measures' effectiveness. This study performs an extensive comparative evaluation of prominent community-aware centrality measures using the Susceptible-Infected-Recovered (SIR) model on real-world online social networks. Overall, results show that K-shell with Community and Community-based Centrality measures are the most accurate in identifying influential nodes under a single spreader problem. Additionally, the epidemic transmission rate doesn't significantly affect the behavior of the community-aware centrality measures.

CCS Concepts: • **Mathematics of computing** → **Graph algorithms**; **Network flows**; • **Theory of computation** → **Dynamic graph algorithms**; • **General and reference** → **Empirical studies**; **Experimentation**; • **Computing methodologies** → **Network science**.

Additional Key Words and Phrases: Complex Networks, Centrality, Influential Nodes, Community Structure, SIR model

ACM Reference Format:

Stephany Rajeh, Marinette Savonnet, Eric Leclercq, and Hocine Cherifi. 2021. Spreading Power of Key Nodes in Online Social Networks with Community Structure. In *epiDAMIK 2021: 4th epiDAMIK ACM SIGKDD International Workshop on Epidemiology meets Data Mining and Knowledge Discovery*. ACM, New York, NY, USA, 6 pages. <https://doi.org/xx.xxxx/xxxxxxxxx.xxxxxxx>

1 INTRODUCTION

With the plethora of data flowing into online social networks, representing the main entities and their interactions is essential. Networks offer an ideal representation of such complex systems to investigate their structure and dynamics. Identifying influential nodes is crucial for many applications such as designing lucrative marketing campaigns, targeting terrorist attacks, controlling epidemic spreading, and detecting financial risks [Lü et al. 2016]. Centrality is one of the main approaches employed to do so. Classically, centrality measures exploit the topology and dynamics of networks [Lü et al. 2016]. They can be classified into two main groups, namely local and global. The former uses the node's neighborhood, while the latter incorporates all of the network's information to quantify a node's influence. They can also be combined [Sciarra et al. 2018].

Many real-world networks contain densely connected zones that are loosely linked to each other. This so-called community structure

is a ubiquitous feature in natural and artificial systems [Girvan and Newman 2002]. The network's structure and dynamics are significantly affected by communities [Nematzadeh et al. 2014]. Recently developed centrality measures exploit this information to identify influential nodes [Ghalmane et al. 2019; Guimera and Amaral 2005; Gupta et al. 2016; Luo et al. 2016; Magelinski et al. 2021; Tulu et al. 2018; Zhao et al. 2015]. We refer to them as "community-aware" centrality measures. Unlike classical centrality measures, community-aware centrality measures differentiate between the node's intra-community links (links between nodes in the same community) and inter-community links (links between nodes in different communities). Intra-community links exert influence at the community level, while inter-community links exert influence at the network level. The difference between community-aware measures is mainly based on how intra-community links and inter-community links are associated together. For example, Comm centrality [Gupta et al. 2016] preferentially selects bridges over hubs by prioritizing inter-community links over intra-community links. Community-aware Mediator [Tulu et al. 2018] favors nodes with unbalanced intra-community and inter-community links.

With limited resources, it is essential to identify top influential nodes either for maximizing or for minimizing the diffusion in online social networks. The Susceptible-Infected-Recovered (SIR) model [Anderson and May 1979] is commonly used to model disease and rumor spreading. Starting with a small set of initial spreaders defined by a specific centrality measure, the goal is to evaluate its ability to reach the maximum outbreak size.

The SIR model has been widely used to investigate the behavior of various classical centrality measures [Bucur 2020; Kitsak et al. 2010; Liu et al. 2015]. Studies on community-aware centrality measures examine either a small number of the proposed solutions in the literature or experiments are performed on a small sample of networks [Gupta et al. 2016; Luo et al. 2016; Magelinski et al. 2021; Tulu et al. 2018; Zhao et al. 2015]. Therefore, there is no consensus about the effectiveness of the most popular measures on online social networks, where communities are naturally prevalent [Gaisbauer et al. 2021; Labatut et al. 2014; Traud et al. 2011]. This paper aims to fill this gap. An extensive investigation of seven community-aware centrality measures is performed on ten real-world online social networks using the SIR diffusion model.

The paper is organized as follows. Section 2 introduces the community-aware centrality measures. Section 3 presents the networks, the tools, and the methodology applied. Experimental results are provided in section 4. The main findings are discussed in section 5. Finally, in section 6, the conclusion is given.

2 COMMUNITY-AWARE CENTRALITY MEASURES

In this section, we briefly recall the definitions of the seven community-aware centrality measures under test. Let $G(V, E)$ be an undirected

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than ACM must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from permissions@acm.org.

epiDAMIK 2021, Aug 15, 2021, Virtual

© 2021 Association for Computing Machinery.

ACM ISBN 978-1-xxxx-XXXX-X...\$15.00

<https://doi.org/xx.xxxx/xxxxxxxxx.xxxxxxx>

and unweighted graph where V is the set of nodes, E is the set of edges, and $N = |V|$ is the size of the network. It is partitioned into N_c non-overlapping communities where c_k is k -th community. A node i possess k_i^{intra} intra-community links and k_i^{inter} inter-community links such that $k_i^{tot} = k_i^{intra} + k_i^{inter}$ represents its degree. Note that if the community structure is unknown, a community detection algorithm is needed to uncover it.

1. Community Hub-Bridge [Ghalmane et al. 2019] weights the intra-community links of a node by their community size. The inter-community links are weighted by the number of communities reached by the node through its inter-community links. It is defined as follows:

$$\alpha_{CHB}(i) = |c_k| \times k_i^{intra} + NNC_i \times k_i^{inter} \quad (1)$$

where $|c_k|$ is the size of the community of node i and NNC_i is the number of communities linked to the node i .

2. Participation Coefficient [Guimera and Amaral 2005] gives more importance on the heterogeneity of the inter-community links of a node. If the node's links are uniformly distributed across the communities, its centrality value is one. It is defined as follows:

$$\alpha_{PC}(i) = 1 - \sum_{c=1}^{N_c} \left(\frac{k_{i,c}}{k_i^{tot}} \right)^2 \quad (2)$$

where $k_{i,c}$ is the number of links node i has in a given community c .

3. Community-based Mediator [Tulu et al. 2018] uses entropy to quantify the node's importance through its intra-community and inter-community links. It is defined as follows:

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

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 ρ_i^{intra} and ρ_i^{inter} which represent the node's ratio of intra-community and inter-community links and $\sum_{i=1}^N k_i^{tot}$ is the total degrees in the network.

4. Comm Centrality [Gupta et al. 2016] weights the intra-community links and inter-community links by the ratio of external links. It also prioritizes bridges over hubs. It is defined as follows:

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

where μ_{c_k} is the proportion of inter-community links over the total community links in community c_k , $\chi = \frac{k_i^{intra}}{\max_{(j \in c)} k_j^{intra}} \times R$, $\varphi = \frac{k_i^{inter}}{\max_{(j \in c)} k_j^{inter}} \times R$, and R is a constant to scale intra-community and inter-community values to the same range.

5. Modularity Vitality [Magelinski et al. 2021] is based on the modularity variation due to the node removal from the network. Removal of a bridge node increases the modularity, while removal of internal a hub decreases the modularity. It is defined as follows:

$$\alpha_{MV}(i) = M(G_i) - M(G) \quad (5)$$

where M is the modularity of a network and $M(G_i)$ is the network's modularity after the removal of node i . Note that Modularity Vitality is a signed centrality. In this study, we use its absolute value to rank the nodes.

6. Community-based Centrality [Zhao et al. 2015] is based on weighting the node's intra-community and inter-community links by the size of their communities. It is defined as follows:

$$\alpha_{CBC}(i) = \sum_{c=1}^{N_c} k_{i,c} \left(\frac{n_c}{N} \right) \quad (6)$$

where n_c is the number of nodes in community c and $k_{i,c}$ is the number of links node i has in a given community c .

7. K-shell with Community [Luo et al. 2016] is based on the k -shell (also called k -core) hierarchical decomposition of the network composed of intra-community links and the network composed of inter-community links, separately. A weighting parameter then combines the two values to prioritize the selection of hubs or bridges. 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 $\alpha^{inter}(i)$ stand for the k -shell value of node i by only considering intra-community links and inter-community links, respectively. δ is set to 0.5 in this study.

3 DATA, TOOLS, AND METHODS

3.1 Data

This study uses ten unweighted and undirected online social networks publicly available. They originate from various online platforms (Facebook, Twitter, Deezer, Hamsterster, and Pretty Good Privacy). Table 1 reports their basic topological characteristics. As their community structure is unknown, it is uncovered by Infomap [Rosvall and Bergstrom 2008].

1. Facebook Friends [Peixoto 2020]: Nodes are users from a Facebook ego network extracted in April 2014. Edges between two users mean they are "friends" on Facebook.

2. Retweets Copenhagen [Rossi and Ahmed 2015]: Nodes are Twitter users tweeting while the United Nations conference in Copenhagen about climate change was taking place. Edges represent retweets.

3. Caltech [Rossi and Ahmed 2015]: Nodes are users on Facebook enrolled at Caltech University. Edges between two users mean they are "friends" on Facebook.

4. Ego Facebook [Rossi and Ahmed 2015]: Nodes are users on Facebook participating in a survey conducted on Facebook. Edges between two users mean they are "friends" on Facebook.

5. Hamsterster [Kunegis 2014]: Nodes represent users from an online social pet network hamsterster.com. Edges represent friendships between the users.

6. Facebook Organizations [Peixoto 2020]: Nodes are users on Facebook who work in the same company. Edges between two users mean they are "friends" on Facebook.

7. Facebook Politician Pages [Rossi and Ahmed 2015]: Nodes are Facebook pages of politicians from different countries. Edges represent mutual likes of Facebook users among the given pages.

8. Princeton [Rossi and Ahmed 2015]: Nodes are users on Facebook enrolled at Princeton University. Edges between two users mean they are "friends" on Facebook.

Table 1. Topological features of the networks. N is the number of nodes. $|E|$ is the number of edges. $\langle k \rangle$ is the average degree. ζ is the transitivity. μ is the mixing parameter. λ_{th} is the epidemic threshold. * means the largest connected component of the network is taken if it is disconnected.

Network	N	$ E $	$\langle k \rangle$	ζ	μ	λ_{th}
Faceb. Frien.*	329	1,954	11.88	0.512	0.112	0.048
Retweets Co.	761	1,029	2.70	0.060	0.287	0.139
Caltech*	762	16,651	43.70	0.291	0.410	0.048
Ego Faceb.	4,039	88,234	43.69	0.519	0.077	0.009
Hamsterster*	1,788	12,476	13.49	0.090	0.298	0.022
Faceb. Org.	5,524	94,219	34.11	0.222	0.366	0.016
Faceb. Pol.	5,908	41,729	14.12	0.301	0.111	0.024
Princeton*	6,575	293,307	89.21	0.163	0.365	0.006
PGP	10,680	24,316	4.55	0.378	0.172	0.056
DeezerEU	28,281	92,752	6.55	0.095	0.429	0.066

9. PGP [Kunegis 2014]: Nodes are users from the web of trust, utilizing Pretty Good Privacy (PGP) encryption for sharing information online. Edges between users represent sharing data under secure connections.

10. DeezerEU [Rozemberczki and Sarkar 2020]: Nodes represent users from Deezer, a European platform for music streaming. Edges represent online friendships between users.

3.2 Susceptible-Infected-Removed Model

The Susceptible-Infected-Removed (SIR) model is one of the widely used diffusion models in networks. Initially, a set of nodes (f_0) is in the infectious state (I) while the remaining nodes are in the susceptible state (S). At each iteration, an infectious node infects its susceptible neighbors at a rate λ . Previously infected nodes recover and are removed from the network at a rate γ . The spreading continues until there are no infectious nodes. At this point, the number of nodes in the "Recovered" state indicates the spreading power of the initial set of nodes (f_0). Each network has an epidemic threshold (λ_{th}) controlling the epidemic spreading. It is defined as [Wang et al. 2016]:

$$\lambda_{th} = \frac{\langle k \rangle}{\langle k^2 \rangle - \langle k \rangle} \quad (8)$$

where $\langle k \rangle$ and $\langle k^2 \rangle$ are the first and second moments of the network's degree distribution. The epidemic threshold values are reported in table 1.

3.3 Imprecision function

The imprecision function [Kitsak et al. 2010] measures the performance of a centrality measure in predicting influential spreaders. It is based on the average number of infections due to an infected seed node. It is defined as follows:

$$\epsilon_c(p) = 1 - \frac{M_c(p)}{M_{eff}(p)} \quad (9)$$

where p is a value between $[0,1]$, $M_c(p)$ is the average spreading power of top pN nodes ranked according to a specific centrality measure c , and $M_{eff}(p)$ is the average spreading power of top pN

nodes ranked according to their influence in the SIR model (N is the number of nodes). The smaller the value of $\epsilon_c(p)$, the better the performance of the centrality measure c .

3.4 Methods

The SIR model runs on each network using different transmission rates around the epidemic threshold ($\frac{\lambda_{th}}{2}$, $\frac{\lambda_{th}}{1.5}$, $1.5 \times \lambda_{th}$, $2 \times \lambda_{th}$). The recovery rate γ is set to 1 to measure the spreading ability of the seed node initiating the spreading only. For each transmission rate, 1000 independent simulations of the SIR model are performed in networks with less than 6000 nodes and 100 independent simulations otherwise. The SIR spread size of each node in the network is computed after setting it as the seed of diffusion. The set ordered from highest to smallest SIR spread size is called the reference set. The community-aware centrality measures are computed, and nodes are ranked from higher to lower centrality value. For each transmission rate (λ), we calculate the imprecision function over the top fraction pN nodes.

4 EXPERIMENTAL RESULTS

4.1 Performance of the community-aware centrality measures within networks

Figure 1 illustrates the performance of the community-aware centrality measures for the ten networks under study. The transmission rate is set equal to the epidemic threshold (λ_{th}) for each network. Each figure reports the evolution of the seven community-aware centrality measure's imprecision function when the top spreading nodes' size ranges from $p=0.02$ to $p=0.2$ of the network size. Remember that the lower the value of the imprecision function, the more effective the centrality measure. One can observe that the performance generally increases with the proportion of top spreading nodes. Furthermore, no community-aware centrality measure outperforms the others in all the situations. Overall, there is a high variability of community-aware centrality measures performances within and across networks. For example, In Ego Facebook, at $p=0.02$, the imprecision value of K-shell with Community is 0.38, followed by Modularity Vitality at 0.6. Then all others have an imprecision value between 0.9 and 1. The variability among the community-aware centralities persists till $p=0.2$. K-shell with Community now has a value of 0.05, indicating its high accuracy at higher p . In the same vein comes Community-based Centrality, which has an imprecision value of 0.1. Its accuracy improves by almost 90% compared to its value at $p=0.02$. Modularity Vitality follows with an imprecision value of 0.25, improving in almost half value of $\epsilon(p)$. Community-based Mediator improves from 0.92 ($p=0.02$) to a value of 0.61 ($p=0.20$). Community Hub-Bridge also improves, but in a lower proportion. Finally, Participation Coefficient and Comm Centrality show a negligible improvement. There is also a high variability for the same community-aware centrality measures across networks. For example, in Facebook Politician Pages, the imprecision value of Community-based Mediator at $p=0.02$ is 0.81, while in Caltech, it amounts to 0.25. Another example is Community-based Centrality in Ego Facebook amounting to 0.9 at $p=0.02$ while it amounts to 0.11 in PGP.

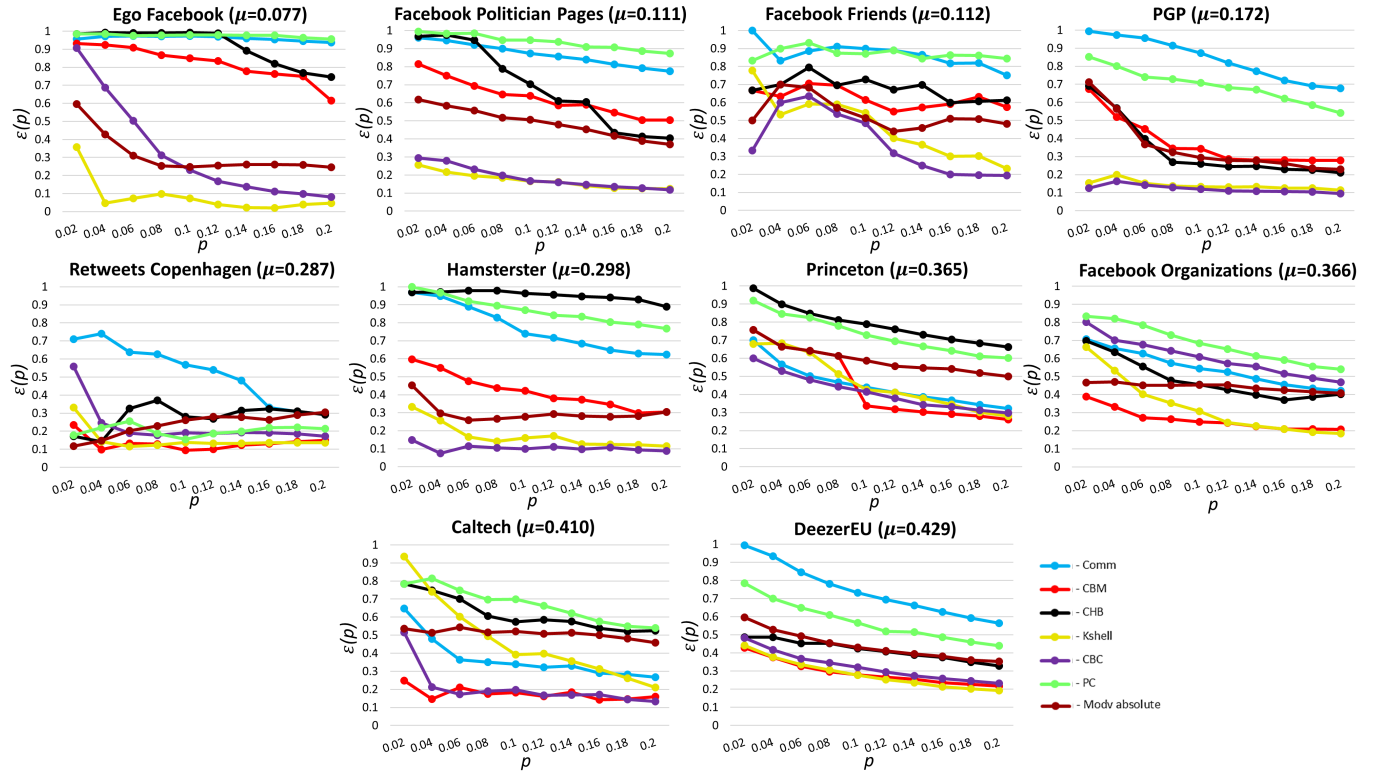


Fig. 1. The imprecision function $\epsilon(p)$ for the 7 community-aware centrality measures on each network. The transmission rate is set to λ_{th} and the recovery rate is set to 1. μ is the mixing parameter, the ratio of inter-community over total community links in a network.

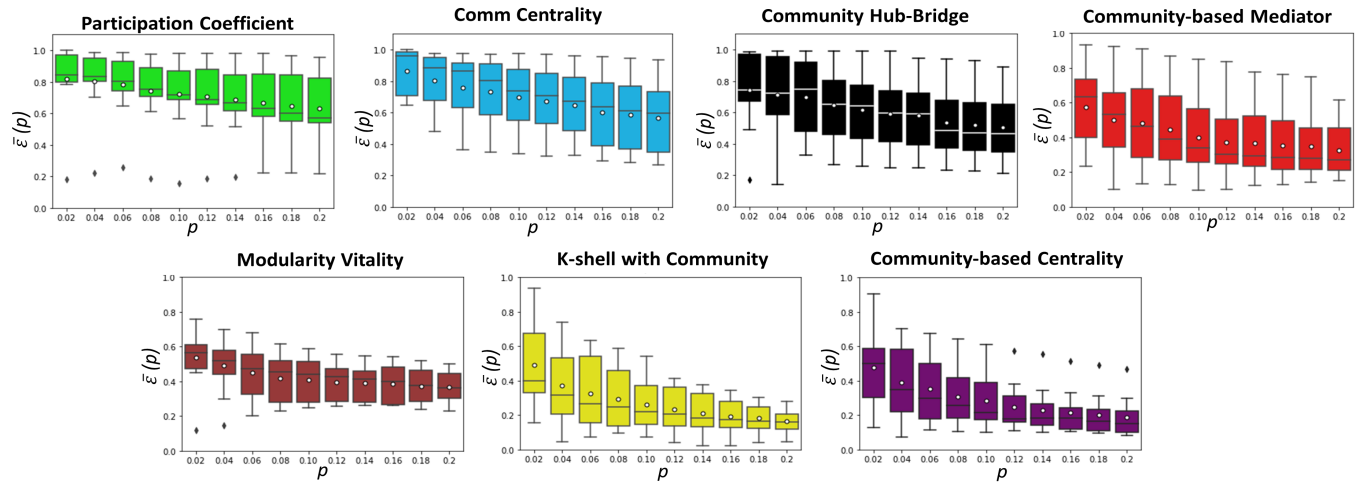


Fig. 2. The average of the imprecision function $\bar{\epsilon}(p)$ over the 10 online social networks. The transmission rate is set to λ_{th} and the recovery rate is set to 1.

4.2 Performance of the community-aware centrality measures across networks

Each community-aware centrality measure’s imprecision function is averaged over the ten networks for all p values. The goal is to better understand the performance consistency. Figure 2 illustrates

these results. The most stable (low variability) community-aware centrality measure is Modularity Vitality. Despite the change in p , the imprecision function values remain stable and condensed. Then comes K-shell with Community and Community-based Centrality. Even though they show high variability when $p \leq 0.08$,

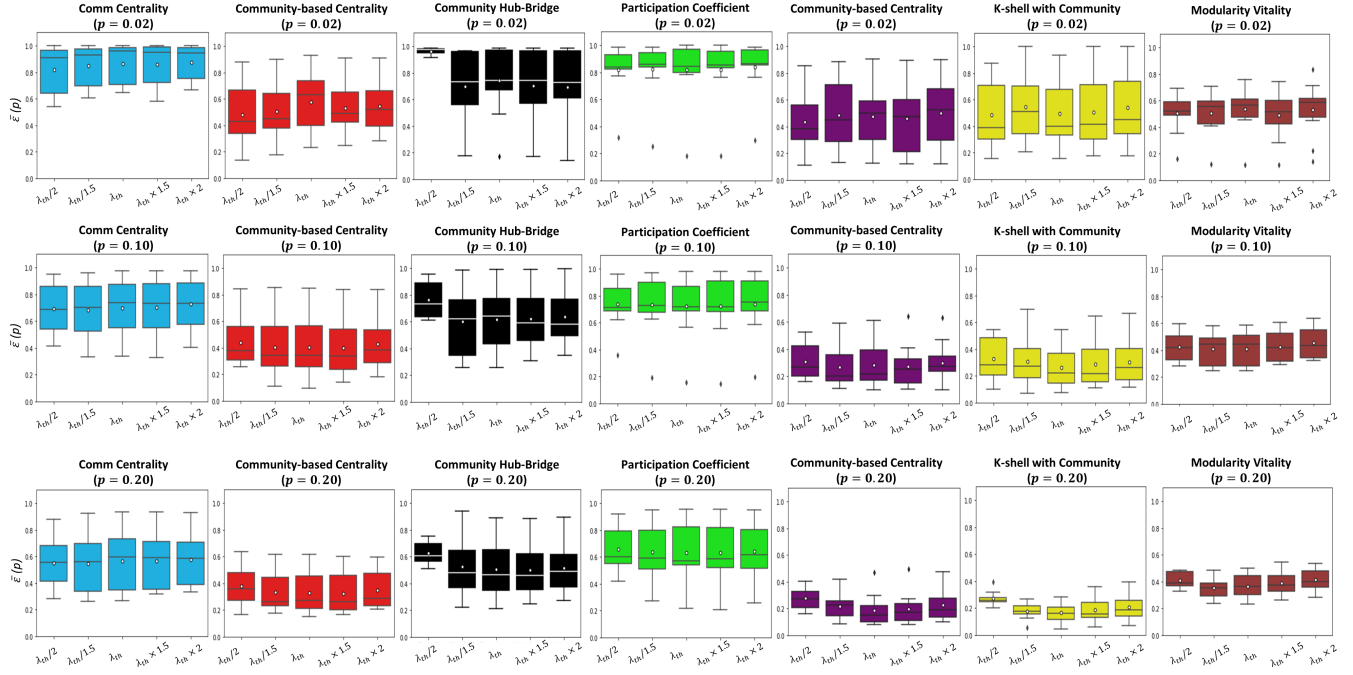


Fig. 3. The average of the imprecision function $\bar{\epsilon}(p)$ over the 10 online social networks as a function of five different transmission rates ($\frac{\lambda_{th}}{2}$, $\frac{\lambda_{th}}{1.5}$, λ_{th} , $1.5 \times \lambda_{th}$, $2 \times \lambda_{th}$). The recovery rate is set to 1. The upper, middle, and bottom figures show the results at $p=0.02$, $p=0.10$, and $p=0.20$, respectively.

both are very consistent afterward. On the opposite, the remaining community-aware centrality measures show higher variability as p increases. The average imprecision function $\bar{\epsilon}(p)$ illustrates the high accuracy of K-shell with Community and Community-based Centrality for all p values. It ranges from 0.5 for the lowest p value to 0.1 at the highest p value. Then comes Modularity Vitality, with $\bar{\epsilon}(p) = 0.55$ at $p=0.02$ and $\bar{\epsilon}(p) = 0.40$ at $p=0.20$. Community-based Mediator has similar $\bar{\epsilon}(p)$ values as Modularity Vitality, yet it has high variability. Community Hub-Bridge shows $\bar{\epsilon}(p)$ between 0.75 and 0.5 at $p=0.02$ and $p=0.20$, respectively. Participation Coefficient and Comm Centrality perform poorly. Their minimum for $\bar{\epsilon}(p)$ is around 0.6, and their maximum is around 0.8. These results confirm the results of figure 1.

4.3 Influence of the transmission rate

In this experiment, we study the effect of varying the transmission rate (λ) in the SIR model around the epidemic threshold (λ_{th}). Figure 3 shows the average imprecision function $\bar{\epsilon}(p)$ of the seven community-aware centrality measures at five different transmission rates. The average imprecision function $\bar{\epsilon}(p)$ is calculated considering a low portion of top nodes ($p=0.02$), a medium portion of top nodes ($p=0.10$), and a high portion of top nodes ($p=0.20$).

At low p values ($p=0.02$), results are generally comparable. For example, the mean of $\bar{\epsilon}(p)$ for Comm Centrality at the five different transmission rates is in the vicinity of 0.8. Also, the boxplots' interquartile range is quite similar, indicating that the values are consistent across λ . Participation Coefficient, Community-based centrality, K-shell with Community, and Modularity Vitality also

show consistent results. In contrast, Community Hub-Bridge is the most sensitive to the variation of the transmission rate. When $\lambda = \frac{\lambda_{th}}{2}$, Community Hub-Bridge cannot differentiate the nodes. Indeed, the mean $\bar{\epsilon}(p)$ is 0.98, and the interquartile range's height is very narrow. When the transmission rate is set to $\frac{\lambda_{th}}{1.5}$, λ_{th} , $1.5 \times \lambda_{th}$, and $2 \times \lambda_{th}$, $\bar{\epsilon}(p)$ becomes quite comparable. The consistency of the mean and the interquartile range of $\bar{\epsilon}(p)$ persists at $p=0.10$ and at $p=0.20$. Indeed, they share approximately the same values of $\bar{\epsilon}(p)$ for all community-aware centrality measures except for Community Hub-Bridge. Although now its interquartile range is wider compared to that of $p=0.02$ when $\lambda = \frac{\lambda_{th}}{2}$, the mean and interquartile range are much different than the others.

5 DISCUSSION

This study aims to investigate the behavior of popular community-aware centrality measures in online social networks. Community-aware centrality measures quantify a node's importance based on its local influence (inside its community using intra-community links) and its global impact (outside of its community using inter-community links). Yet, each community-aware centrality measure processes these two types of links distinctively.

A comparative evaluation of seven community-aware centrality measures is performed using the SIR diffusion model. The imprecision function quantifies the centrality measure's accuracy by comparing the spreading power of top nodes according to a centrality measure compared to their ground truth spreading efficiency. Results indicate that K-shell with Community and Community-based Mediator outperform the alternative community-aware centrality

measures. K-shell with Community exploits the hierarchical structure of the networks while taking into consideration its community structure. This result corroborates the study reported in [Kitsak et al. 2010]. Indeed, the authors show that under a single spreader setting, nodes identified using k -shell are the most accurate in predicting spreading outbreaks in networks. The performance of Community-based Centrality is also on the same line as the findings of the authors who proposed this measure [Zhao et al. 2015]. This study shows that this community-aware centrality measure is accurate in online social networks with communities of different sizes. The performance of these two measures can be attributed to their functioning in a single-spreader problem. K-shell with Community can pinpoint the nodes that are found in the core of the local and global network. The most efficient single-spreaders are then found by combining the node's coreness in these two networks, representing the influence of a node inside and outside its community. Note that in the case of a multiple-spreader problem, the distance between the nodes will then be the most discriminating feature in determining the spread of the epidemic [Kitsak et al. 2010]. This abates the power of k -shell and will also be the case with K-shell with Community since many nodes will exist on the same core level. For Community-based Centrality, by considering both the size of the communities that a node links to and its connections to these communities, the influential nodes are better identified. Indeed, this approach will prioritize nodes that are well connected to other communities, but the overall power is directly proportional to the communities' sizes.

Results also show that Community-based Mediator is somewhat sensitive to the community structure strength. Indeed, as shown in figure 1, when the network has a strong community structure ($\mu \leq 0.172$), it performs poorly. Yet, as the network has a weaker community structure, it becomes as accurate as K-shell with Community and Community-based Centrality. This centrality exploits the heterogeneity of links to assess the node's importance. Indeed, in a weak community structure, a node possesses a higher number of inter-community links than intra-community links. It explains why it performs better in a weak community structure. Modularity Vitality is the most consistent across networks, regardless of the strength of the community structure. The low accuracy of Participation Coefficient, Comm Centrality, and Community Hub-Bridge may be linked to the fact that they give a lot of importance to bridges. Besides bridges, online social networks also include hubs inside their communities that play a major role in information dissemination.

6 CONCLUSION

Identifying influential nodes in online social networks is fundamental for maximizing information diffusion and inhibiting fake news spreading. The community structure of a network plays a crucial role in the dynamics of these spreading processes. This work investigates the effectiveness of prominent community-aware centrality measures to target influential nodes under the SIR diffusion process context. Results show that K-shell with Community and Community-based Centrality are the most accurate community-aware centrality measures. Additionally, performances are pretty insensitive to variation of the transmission rate. Therefore, this work gives clear indications about which community-aware centrality

measure to use. Nevertheless, practitioners need to be conscious that the community aware-centrality measures accuracy depends on the seed node size. As in numerous situations, the community structure is unknown. Future work will investigate the results consistency using alternative community detection algorithms. Another direction of research is to study the influence of the propagation process. Finally, we are planning to link the performances to the network topological properties.

REFERENCES

- Roy M Anderson and Robert M May. 1979. Population biology of infectious diseases: Part I. *Nature* 280, 5721 (1979), 361–367.
- Doina Bucur. 2020. Top influencers can be identified universally by combining classical centralities. *Scientific reports* 10, 1 (2020), 1–14.
- Felix Gaisbauer, Armin Pournaki, Sven Banisch, and Eckehard Olbrich. 2021. Ideological differences in engagement in public debate on Twitter. *Plos one* 16, 3 (2021), e0249241.
- Zakariya Ghalmane, Mohammed El Hassouni, and Hocine Cherifi. 2019. Immunization of networks with non-overlapping community structure. *SNAM* 9, 1 (2019), 1–22.
- Michelle Girvan and Mark EJ Newman. 2002. Community structure in social and biological networks. *PNAS* 99, 12 (2002), 7821–7826.
- Roger Guimera and Luis A Nunes Amaral. 2005. Functional cartography of complex metabolic networks. *nature* 433, 7028 (2005), 895–900.
- Naveen Gupta, Anurag Singh, and Hocine Cherifi. 2016. Centrality measures for networks with community structure. *Physica A: Statistical Mechanics and its Applications* 452 (2016), 46–59.
- Maksim Kitsak, Lazaros K Gallos, Shlomo Havlin, Fredrik Liljeros, Lev Muchnik, H Eugene Stanley, and Hernán A Makse. 2010. Identification of influential spreaders in complex networks. *Nature physics* 6, 11 (2010), 888–893.
- Jérôme Kunegis. 2014. Handbook of Network Analysis [KONECT—the Koblenz Network Collection]. arXiv:1402.5500
- Vincent Labatut, Nicolas Dugué, and Anthony Perez. 2014. Identifying the community roles of social capitalists in the twitter network. In *2014 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM 2014)*. IEEE, Beijing, China, 371–374.
- Ying Liu, Ming Tang, Tao Zhou, and Younghae Do. 2015. Core-like groups result in invalidation of identifying super-spreader by k -shell decomposition. *Scientific reports* 5, 1 (2015), 1–8.
- Linyuan Lü, Duanbing Chen, Xiao-Long Ren, Qian-Ming Zhang, Yi-Cheng Zhang, and Tao Zhou. 2016. Vital nodes identification in complex networks. *Physics Reports* 650 (2016), 1–63.
- Shi-Long Luo, Kai Gong, and Li Kang. 2016. Identifying influential spreaders of epidemics on community networks.
- T. Magelinski, M. Bartulovic, and K. M. Carley. 2021. Measuring Node Contribution to Community Structure With Modularity Vitality. *IEEE Transactions on Network Science and Engineering* 8, 1 (2021), 707–723.
- Azadeh Nematzadeh, Emilio Ferrara, Alessandro Flammini, and Yong-Yeol Ahn. 2014. Optimal network modularity for information diffusion. *Physical review letters* 113, 8 (2014), 088701.
- Tiago P. Peixoto. 2020. The Netzschleuder network catalogue and repository. [Online]. Available: <https://networks.skewed.de/>.
- Ryan A. Rossi and Nesreen K. Ahmed. 2015. The Network Data Repository with Interactive Graph Analytics and Visualization. In *Proceedings of the Twenty-Ninth AAAI Conference on Artificial Intelligence (AAAI'15)*. AAAI Press, Austin, Texas, 4292–4293.
- Martin Rosvall and Carl T Bergstrom. 2008. Maps of random walks on complex networks reveal community structure. *PNAS* 105, 4 (2008), 1118–1123.
- Benedek Rozemberczki and Rik Sarkar. 2020. Characteristic Functions on Graphs: Birds of a Feather, from Statistical Descriptors to Parametric Models. arXiv:2005.07959 [cs.LG]
- Carla Sciarra, Guido Chiarotti, Francesco Laio, and Luca Ridolfi. 2018. A change of perspective in network centrality. *Scientific reports* 8, 1 (2018), 1–9.
- Amanda L Traud, Eric D Kelsic, Peter J Mucha, and Mason A Porter. 2011. Comparing community structure to characteristics in online collegiate social networks. *SLAM review* 53, 3 (2011), 526–543.
- Muluneh Mekonnen Tulu, Ronghui Hou, and Talha Younas. 2018. Identifying influential nodes based on community structure to speed up the dissemination of information in complex network. *IEEE Access* 6 (2018), 7390–7401.
- Wei Wang, Quan-Hui Liu, Lin-Feng Zhong, Ming Tang, Hui Gao, and H Eugene Stanley. 2016. Predicting the epidemic threshold of the susceptible-infected-recovered model. *Scientific reports* 6, 1 (2016), 1–12.
- Zhiying Zhao, Xiaofan Wang, Wei Zhang, and Zhiliang Zhu. 2015. A community-based approach to identifying influential spreaders. *Entropy* 17, 4 (2015), 2228–2252.