



Using Link Clustering to Detect Influential Spreaders

Simon Krukowski¹(✉) and Tobias Hecking²

¹ University of Duisburg-Essen, Forsthausweg 2, 47057 Duisburg, Germany
simon.krukowski@stud.uni-due.de

² German Aerospace Center, Linder Höhe, 51147 Cologne, Germany

Abstract. Spreading processes are increasingly analysed in the context of complex networks, for example in epidemics research, information dissemination or rumors. In these contexts, the effect of structural properties that facilitate or decelerate spreading processes are of high interest, since this allows insights into the extent to which those processes are controllable and predictable. In social networks, actors usually participate in different densely connected social groups that emerge from various social contexts, such as workplace, interests, etc. In this paper, it is examined if the number of groups an actor connects to can be used as a predictor for its capability to spread information effectively. The social contexts (i.e. groups) a node participates in are determined by the Link Communities approach by Ahn et al. (2010). The results are contrasted to previous findings of structural node properties based on the k -shell index of nodes (Kitsak et al. 2010) by applying both methods on artificially generated and real-world networks. They show that the approach is comparable to existing ones using structural node properties as a predictor, yet no clear evidence is found indicating that one or the other approach leads to better predictions in all investigated networks.

Keywords: Link clustering · Spreading processes · Information diffusion

1 Introduction

Spreading processes, originally examined in areas such as disease modelling [1, 2] and epidemiological mathematics [3], are increasingly examined to study social phenomena of information diffusion within complex networks, such as the spreading of rumours [4] or the communication during crisis events, for example [5]. They also gain special relevance considering the global COVID-19 pandemic, possibly yielding results to better understand and mitigate its spread. As a result of the way users connect and interact with each other, the social networks used for these analyses often exhibit properties of small-world and scale-free networks [6], making topological characteristics of the network an important aspect when analysing spreading processes. A common goal of the mentioned studies is to predict the efficiency of the spreading process, with the broader intention to acquire knowledge on how to control it. In this context, both local and global network properties related to spreading processes have to be explored. This paper focuses on local

properties (topological properties of the network attributed to nodes) but goes beyond their immediate neighbourhood. Using local properties to predict spreading efficiency, Kitsak et al. [7] already showed that the most efficient spreaders within a network are not necessarily the most connected nodes (i.e. nodes with the highest degree), but the ones that are located in densely connected cores of the network indicated by a high k -shell index. However, apart from being located within the core of a network or having a certain degree, structural properties of the network such as its tendency to form clusters might also yield an informative measure to predict the spreading efficiency. The general idea is, that someone who is member of many different overlapping social groups (workplace, sports club, friendship circles) is better capable of injecting information into various densely connected regions of the network where it further circulates. This notion of overlapping social groups is operationalised by applying the link-clustering approach by Ahn, Bagrow & Lehmann [8], where resulting clusters are highly interleaved and sometimes even nested. The approach clusters the links instead of the nodes, resulting in nodes possibly belonging to multiple clusters. It is reasonable to assume then, that the number of groups a node belongs to predicts its spreading capability as good as or even better than its k -shell index. Thus, we derive the following research questions:

- RQ1:** Is the community membership of nodes as determined by the Link Communities approach a good predictor for efficient spreaders within complex networks?
- RQ2:** Are the two approaches (Link Communities and k -Core) comparable for determining influential spreaders?

2 Background

Spreading processes and the analysis of potential spreaders have a long history in science and generally describe a flow of *information* between actors or members of a network [9]. For complex networks such as computer networks or networks of real individuals, *information* can refer to diseases and computer viruses [1, 10], whereas for other networks (i.e. created from Social Media data), it can refer to opinions, news articles or influence [11–13]. The spreading, i.e. the flow of *information* within these networks can result in diverse operationalisations, and obviously in contrary motifs regarding its analysis. For disease spreading, potential strategies to mitigate are sought-after, whereas for influence maximisation or opinion spreading, strategies to accelerate the flow of information are desired. For that reason, influencing factors are of great interest. Within complex networks where spreading can only happen between adjacent nodes, there is one aspect that affects the spreading, and likewise the efficiency of a single spreader – regardless of the motifs of analysis – to an equal degree: the topology of the network (see [9]). With regard to this topology, the origin of the diffusion process (i.e. the spreader) is of interest, as these so called “seeds” [14] and their properties yield important information from which inferences regarding the efficiency of the spreading process can be drawn.

2.1 Properties of the Network

As described above, the topology of a network results in certain characteristics of nodes, from which inferences regarding their spreading capability can be drawn. On an individual level, the degree centrality of a node is one such characteristic, as nodes with high degree centralities naturally have more possibilities to potentially spread information to other nodes [15]. Thus, so called “hubs” mark efficient spreaders (see [7]), which is also reflected by the fact that an uneven degree distribution (many hubs) results in more efficient spreading [7]. Apart from this, the community structure of a network can also influence spreading processes, as it is conceivable that information can spread more easily within highly interconnected sub-communities [16].

One notion to describe the community structure and the cohesiveness of subgraphs is the k -core of a network and the respective k -shell index of a node, which is the highest k such that the node is still part of the respective k -core. The index results in nodes which have at least k connections to other nodes within their core, resulting in highly interconnected nodes, whose spreading capability is high and where spreading is likely to occur. Furthermore, community detection techniques such as the Louvain method [17] can be used to examine the community structure of a network. However, as it assigns sub-communities to nodes based on high connections within a community and little connections between different sub-communities, each node is assigned a unique sub-community. To infer the spreading capability of a node however, its membership of a subcommunity yields little information: Information that originates from a spreader located within a highly interrelated sub-community might propagate quickly within the respective sub-community but is less likely to propagate to other sub-communities in the case of highly separated communities. In contrast, nodes with many different community memberships, indicating activities in various social contexts, are hypothesised to be capable spreaders. Such multiple community memberships can be found by clustering methods such as Clique Percolation [18] or Link Communities [8] that allow for overlapping sub-communities.

3 Approach

In this paper, it is hypothesised that actors who connect groups in different social contexts and thus are part of different overlapping and nested link communities are capable spreaders. The underlying assumption is that information items, diseases, etc. mainly circulate within densely connected groups. Actors in the overlap of such groups can be infected within one group and inject the spreading process into several other groups. To this end, we investigate whether the membership in multiple link communities (see Sect. 3.1) is another factor that determines spreading capability in addition to the node’s k -shell index [7]. It is further argued, that the number of link communities a node belongs to, also constituting a topological feature of the network organisation, suggests that the spreader has close connections to many other actors from different sub-communities within the network, and is thus able to spread between different highly interconnected communities more easily. Kitsak et al. [7] argue, that especially during the early stages of spreading processes, through the many pathways that exist for nodes located within the core of the network, the k -Shell index of a node predicts its spreading capability. However, nodes

within these cores also tend to exhibit multiple community memberships, which yields an additional inferential value in comparison to solely taking the node's k -Shell index into account (see Fig. 1). Additionally, when multiple outbreaks happen, information can spread more easily between different sub-communities, whereas for different cores, the distance between them needs to be taken into account [7]. The Link Communities approach by Ahn, Bagrow, & Lehmann [8] offers a method to determine overlapping communities, as it assigns multiple community-memberships to a single node. From this, inferences regarding the spreading capability of single nodes can be drawn.

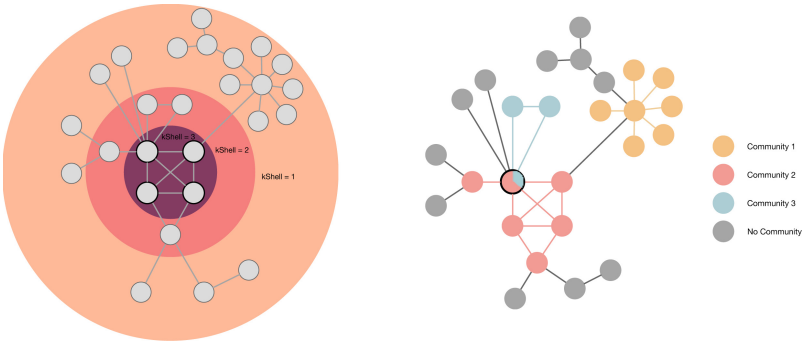


Fig. 1. On the left, an example from Kitsak et al. (2010) can be seen. Nodes within the core of the network, i.e. with a high k -Shell index, were found to be good spreaders. On the right, the same network is shown, but clustered with the Link Communities approach by Ahn et al. (2010). Nodes with black borders are nodes which are hypothesized to exhibit a high spreading capability. In this example, nodes with multiple community memberships also exhibit a high community centrality.

3.1 Link Communities

To determine the community-memberships of nodes, the Link Communities approach by Ahn et al. (2010) is used. In their approach, a sub-community is characterised as a set of closely interrelated links instead of closely interconnected nodes. As a result, sub-communities can overlap, and single nodes can be members of multiple sub-communities. The procedure to cluster the links by Ahn et al. (2010) is described as follows: Edges (e_{ik} and e_{jk}) with a common neighbour k are compared pairwise. Node k is called keystone node, while the other two nodes are called impost nodes. It should be noted, that

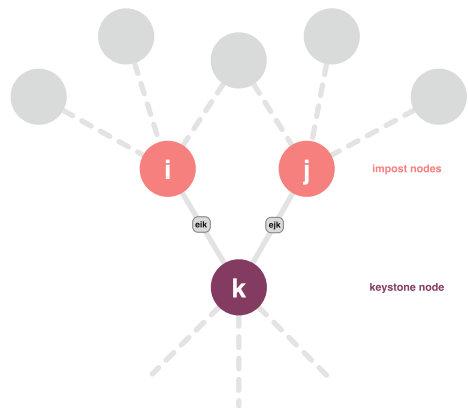


Fig. 2. Illustration from Ahn et al. (2010). As can be seen, only the neighbourhood of the impost nodes is taken into account.

only the neighbours of the impost nodes are taken into account for the calculation, as the neighbours of k (except the impost nodes) are of no interest. To calculate the similarity of the nodes, the similarity criterion S (Jaccard-index, [19]) is applied (see Eq. 1). The set of the node i and its neighbours is denoted as $n + i$.

$$S(e_{ik}, e_{jk}) = \frac{|n + (i) \cap n + (j)|}{|n + (i) \cup n + (j)|} \quad (1)$$

For the above example (Fig. 2), this would result in $S = \frac{1}{4}$. A dendrogram is then built through single-linkage hierarchical clustering and cut at a certain threshold according to the partition density, which then results in the link communities [8]. From these link communities, the community memberships of the nodes can be derived, and thus each node is assigned a vector of community memberships, from which the actual number of communities it belongs to can be calculated. The possibility to detect multiple community memberships differentiates our approach from other approaches analysing properties of influential spreaders [20], where nodes can only belong to one community each due to the community detection algorithm being used.

Community Centrality. Although the number of communities a node belongs to might be an important predictor for its spreading capability, there are possible limitations. Generally, due to the nested nature of link communities, a node can be a member of many communities. One can assume, that being a member of many communities goes along with a high spreading capability. However, because of this pervasive overlap, the set of directly reachable nodes can be limited, even for nodes belonging to multiple groups and the inferential value of the nodes' number of community-memberships might be limited. This is reflected by the community centrality by Newman [21], which assigns higher centrality values to nodes if they belong to many communities with little overlap. In this study, this concept is extended, additionally taking the size of the communities into account. For simplicity, it will also be denoted as community centrality. Formally, it is defined as the cardinality of the union of nodes in all communities a node belongs to. Consequently, it is high if a node belongs to many large communities with little overlap. Community centrality will be denoted as CC.

4 Evaluation

To evaluate the capability of nodes to spread information through the network, spreading processes are simulated according to well-known SIR models [3]. The process starts with one initially infected node. This node infects its neighbours at a given infection rate (denoted as β) and recovers. The resulting infected nodes then try to infect their neighbours themselves. The process terminates when no new infections occur. In this study, the spreading capability of a node x corresponds to the average fraction of infected population in 100 SIR runs starting at node x . For the community detection, it is decided to cut the resulting dendrogram at a smaller threshold to also detect smaller sub-communities.

Analysed Datasets. To evaluate our measure, we chose to use both real-world networks to increase the external validity, as well as artificially generated networks to maintain a high internal validity. The generated networks were created according to the forest-fire algorithm as shown by Leskovec, Kleinberg & Faloutsos [22], as it creates networks with properties typical of real-world graphs such as heavy-tailed degree distributions and communities. To this end, 8 undirected networks were created, each with 1000 nodes, with forward burning probabilities ($fwprob$) ranging from .05 to .40 (Table 1) and a fixed backward burning probability of 1 (see [22]). The $fwprob$ controls for the tendency to form densely connected and potentially nested clusters. Additionally, we analysed one ego-networks from the SNAP (Stanford Network Analysis Project) at Stanford University [23] to evaluate our metric on real-world data. The network represents the connections between all friends of the individual of which the ego network is derived from (thus *ego*), with all connections between the ego and friends removed. To increase informative value and evaluate our measure on a bigger network than our created ones, we chose the *107* network because of its size (at 1912 nodes and 53498 edges), a local average clustering coefficient of .534 and a mean spreading capability of .683.

Table 1. Analysed datasets. We used the average local clustering coefficient. The spreading capability is the mean spreading capability of all nodes.

Network	Edges	avg. Degree	Clust. Coeff.	Diameter	Spreading Cap.	FwProb
1	1,076	2.152	0.160	23	0.066	0.050
2	1,193	2.386	0.306	24	0.136	0.100
3	1,275	2.550	0.358	21	0.236	0.150
4	1,416	2.832	0.385	21	0.339	0.200
5	1,829	3.658	0.490	19	0.552	0.250
6	2,271	4.542	0.530	16	0.638	0.300
7	5,455	10.910	0.540	11	0.792	0.350
8	48,355	96.710	0.837	8	0.952	0.400

4.1 Metrics for Evaluation

In addition to descriptively comparing the proposed measure with established measures, certain metrics are applied to objectively evaluate it.

The Imprecision Function. To quantify the importance of nodes with a high community centrality during the spreading, an objective measure has to be calculated. The imprecision function serves just that purpose. Similar to Kitsak et al. (2010), these functions are calculated for each of the three relevant measures, and they are denoted as $\varepsilon_{k_S}(p)$, $\varepsilon_{CC}(p)$ and $\varepsilon_d(p)$, respectively. For each subset p of nodes (here, p refers to a specific percentage of the dataset) with the highest spreading capability (denoted as ϕ_{eff}) and the highest value according to one of the three measures (denoted as φ_{k_S} , φ_{CC} and φ_d), the average spreading is calculated. Then, the difference in spreading between the

p nodes with highest values in any of the measures and the most efficient spreaders is calculated. Formally, for ϵ_{CC} , the function is defined as follows:

$$\epsilon_{CC}(p) = 1 - \frac{\varphi_{CC}(p)}{\varphi_{eff}(p)} \tag{2}$$

Note that through subtracting the fraction from 1, higher values correspond to more imprecision, while smaller values for ϵ show less imprecision and thus a better measure.

4.2 Results

In the following, the results of the evaluation will be presented. All of the evaluations are calculated at a beta value (infection rate) of 7%.

Descriptive Results

General Observations. To assess how well the community memberships of a node allow inferences about its spreading capability, they are compared to other measures, more specifically to the k -Shell index of a node [7] and to its degree. As can be seen in Fig. 3, the measures generally correspond to each other, and higher $fwprob$ values result in higher values for CC, degree and k -Shell index. Figure 4 shows the results of a bivariate comparison for our generated networks with 1000 nodes, beta = 7 and $fwprob$ values of .05, .15 and .30. It can be seen that higher $fwprob$ values correspond to higher spreading capabilities, likely because of more community structure within the graphs. While the predictive value of all shown measures seems to be smaller for lower $fwprob$ values with regard to the spreading capability, it increases for higher $fwprob$ values. The figure shows, that generally, the CC corresponds with the degree and the k -Shell index of a node. Especially in comparison with the degree of a node, it can be seen that low degree values do not consistently correlate with low spreading capabilities, whereas for the CC, they do. For $fwprob = .30$, higher CC values also consistently reflect high spreading capabilities, whereas for degree, there are effective spreaders with small degrees. However, this effect shows less so for the k -Shell index of a node, where effective spreaders can be found for a greater variety of CC values, especially at high k -Shell values. Regarding **RQ1**, it can be said that the CC can also be used as a predictor for the spreading capability of the nodes: For high CC values, there are high spreading capabilities whereas for small CC values, spreading capabilities generally remain low. This discriminatory value increases with higher $fwprob$ values and thus higher community structures within the

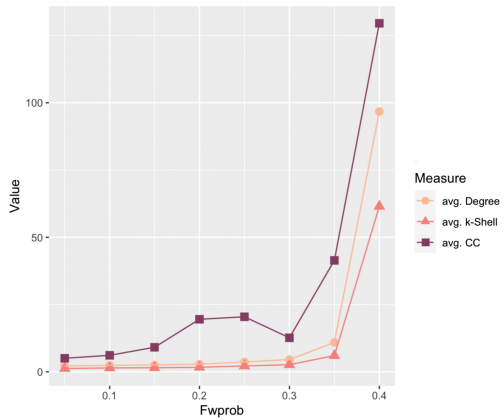


Fig. 3. Average values of the compared measures for the chosen $fwprob$ values.

graphs. To additionally evaluate this on a real network, we created the same plot for the *107* network (Fig. 5) at $\beta = 1$ (high β values would result in little variance of the spreading capability), where a similar pattern as above and in Kitsak et al. [7] emerges. Even more clearly than for the generated networks, it can be seen that high values in either community centrality, degree or k -Shell index, go along with a high spreading capability. Thus, and also considering what can be seen in Fig. 3, both the k -Shell index and the Link Communities approach are comparable in predicting the spreading capability (RQ2). However, this effect shows less so for our generated networks at high $fwprob$ values than for our analysed *107* network.

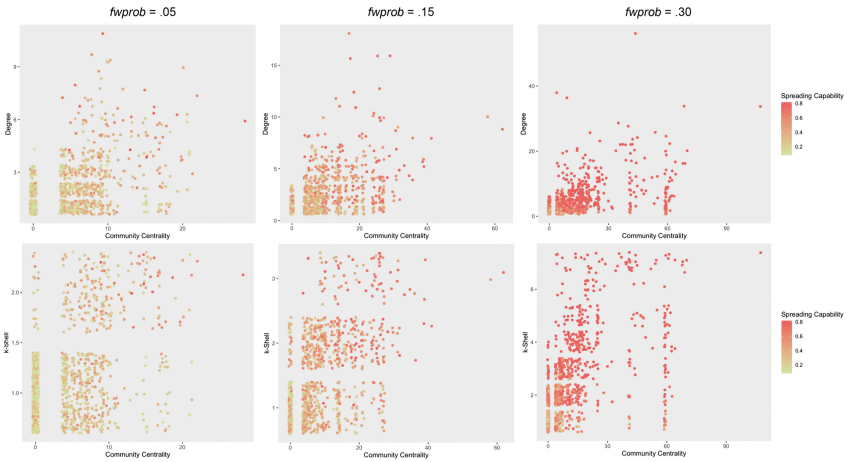


Fig. 4. Bivariate distribution for the examined measures on our generated networks. The colours indicate the spreading capability.

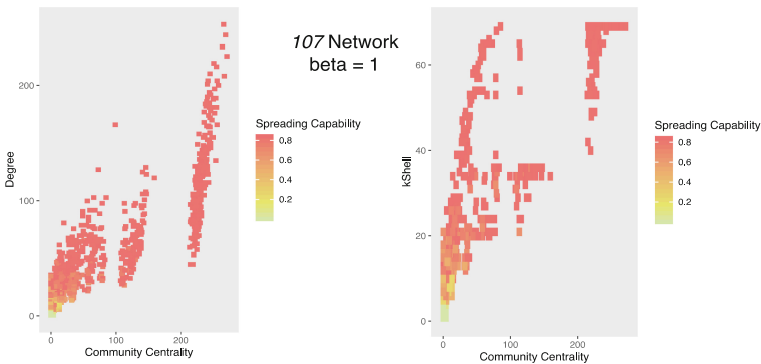


Fig. 5. Bivariate distribution for the 107 Facebook ego network

Correlational Measures. Correlations between the CC and the established measures are calculated. The mean correlation across all of our generated networks between the CC

and the k -Shell index of a node is $r = .56$ and $r = .54$ between CC and the degree. All measures correlate with the spreading capability of a node ($r = .40$ for CC, $r = .40$ for degree and $r = .47$ for k -Shell). In the *107* network, the CC also correlates with the other measure ($r = .90$ with k -Shell and $r = .88$ for degree) and with the spreading capability of a node (r values at $.50$ for CC, $.55$ for degree and $.63$ for k -Shell). Regarding **RQ2**, this makes the CC comparable to the k -Shell index of a node.

Imprecision Function. Apart from using correlational measures and looking at the distribution of data points, focusing on the top n nodes, the imprecision function is applied to all of the studied networks, in order to objectively evaluate the proposed measure in comparison to the other measures. It was calculated for all of our generated networks at $\beta = 7$. As can be seen in Fig. 6 very clearly, the errors decrease with higher $fwprob$ values, reflecting the observation described above and indicating that the predictive value of all measures (the CC included) seems to increase when there is a higher community structure. However, with $M = 0.18$ ($SD = 0.20$) for the k -Shell index, $M = 0.10$ ($SD = 0.12$) for degree and $M = 0.15$ ($SD = 0.20$) for community centrality, they are generally quite low, with the k -Shell index showing the highest average error. For the *107* network, the error values are especially low at $.01$ for degree, $.02$ for k -Shell and $.01$ for the CC. As there is no significant difference between the imprecision of our measure and the imprecision of other measures for both real and generated networks, the measures are comparable (**RQ2**), while the low error values additionally indicate the CC to be a good predictor for the spreading capability of a node (**RQ1**).

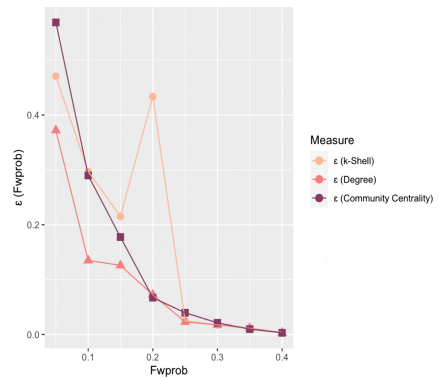


Fig. 6. Results of the imprecision function for our generated networks.

5 Discussion

The aim of this paper is to contribute to research on spreading processes within complex networks and identify properties which increase the efficiency of spreading. To this end, a novel measure is introduced, from which inferences regarding the spreading capability of single actors can potentially be drawn. Using link clustering [8] to infer multiple community memberships of nodes, we evaluate and compare the measure to other approaches. Generally, our results extend the understanding of the effect of structural properties of networks on information diffusion beyond centralities and k -shell and show that the community centrality of a node is comparable in predicting its spreading capability.

Descriptively, the average values of the measures indicate a comparability (**RQ2**): Higher $fwprob$ values and thus a higher community structure go along with higher k -Shell indices, degrees and community centrality. The bivariate plots (Fig. 4) extend this

with regard to predicting effective spreaders, additionally hinting at a broad range of degrees for a small range of community centralities. While the correlation shows less so in comparison with the k -Shell indices for our generated networks, it does show for the observed *107* network. There, high community centralities reflect less variance in spreading capability, implying a more robust prediction. This is especially true for the comparison between the CC and the k -Shell index, making it a good predictor in our analysed real network (**RQ1**). The calculated correlations imply the comparability of our measure, as it correlates with the other measures and also shows a medium-high correlation with the spreading capability of a node - for both generated and real networks. Judging from these correlations alone, regarding **RQ1**, this does also make the CC a good predictor for effective spreaders, yet slightly less than the k -Shell index. However, it should be noted that the expressiveness of correlations is limited in this case because of long-tailed and different distributions of the variables. The high correlations between the variables underline this further, as for example, our generated networks are scale free with many nodes showing a low degree, and consequently, also a low community centrality. Thus, in addition, the imprecision function focusing on the top n nodes is evaluated. The results clearly show that the errors in predicting the most efficient spreaders decrease when there are higher *fwprob* values and thus more community structure – for the CC along with the other analysed measures. This is especially relevant, as it shows that with regard to the results of the imprecision function, our measure is not only a good predictor for efficient spreaders (**RQ1**), yet this predictive value increases when the networks show more properties of real world networks with communities and heavy-tailed degree distributions. This is also reflected by the results for the *107* network and additionally undermined by the generally small error values. In our analysed networks, error values were even slightly higher for the k -Shell index than for the CC, exceeding our **RQ2** of comparability. In conclusion, the evaluations of the proposed measure show its comparability to other measures, specifically the degree and the k -Shell index of a node.

Certain things should be taken into account. For our generated networks with high *fwprob* values and the *107* network, the mean spreading capability of the sample is very high. This extreme right-skewness means, that many nodes show a high spreading capability, and takes away possibilities to examine properties that lead to such high spreading capability. Apart from that, due to computational constraints, the possibilities of simulating the spreading processed in the networks were limited, resulting in capping the size at 1000 nodes. Additionally, *fwprob* values greater than .40 resulted in networks with more than 100,000 edges, also making the simulations computationally intensive. It might therefore be possible that bigger networks or higher *fwprob* values show different results – calling for future analyses with bigger networks. For our chosen real network, its characteristics might have also influenced the metrics used for the conducted analyses. Thus, other networks (real-world, computer-networks or networks with ground-truth communities) should also be used for future analyses.

While we systematise the degree of real-world properties of the generated networks by varying the *fwprob* values with which they are created, we do not systematise all aspects of our analyses: There are aspects of our Julia implementation which we use to simulate the spreading that are fixed, specifically the steps (fixed to 30) and the iterations

(fixed to 10), and likewise the beta value at .7 (for the generated networks). Future studies should also vary these fixed parameters and try to systematise them like the systematised *fwprob* values in this paper. Additionally, the spreading capabilities obtained are also not deterministic, that means they rely heavily on chance. It is therefore possible that another run yields slightly different results.

Conclusion and Outlook. Our analyses and evaluations showed, that apart from the *k*-Shell index and centralities, structural properties of the network can affect spreading processes. To this end, community centrality, the examined measure, proved to be comparable in doing so. Along with the other analysed measures, the CC's inferential value increases, as the *fwprob* used to create the network with the forest fire algorithm [21] increases – meaning that for increasing real-world and scale free properties of a graph, the CC becomes better in predicting efficient spreaders. While the *k*-Shell index of a node seems to be a better predictor for the spreading capability under certain conditions, there might be applications in which the *k*-Shell index yields little inferential value or where there are multiple outbreaks simultaneously. In this case, the community centrality might be used instead of the *k*-Shell index coupled with the distance between cores [7]. This paper contributes to our understanding of the underlying processes through offering another measure that can be used to infer the spreading capability of nodes, and thus the efficiency of information diffusion due to structural properties of complex networks. Due to factors that could have influenced the evaluations in the present paper, future studies should further evaluate the measure, apply it to different contexts and networks and possibly apply new evaluation metrics.

References

1. Goltsev, A.V., Dorogovtsev, S.N., Oliveira, J.G., Mendes, J.F.F.: Localization and spreading of diseases in complex networks. *Phys. Rev. Lett.* **109**(12), 128702 (2012). <https://doi.org/10.1103/PhysRevLett.109.128702>
2. Kong, X., Qi, Y., Song, X., Shen, G.: Modeling disease spreading on complex networks. *Comput. Sci. Inf. Syst.* **8**(4), 1129–1141 (2011)
3. Keeling, M.J., Rohani, P.: *Modeling Infectious Diseases in Humans and Animals*. Princeton University Press (2011)
4. Qian, Z., Tang, S., Zhang, X., Zheng, Z.: The independent spreaders involved SIR rumor model in complex networks. *Phys. Stat. Mech. Its Appl.* **429**, 95–102 (2015). <https://doi.org/10.1016/j.physa.2015.02.022>
5. Stieglitz, S., Bunker, D., Mirbabaie, M., Ehnis, C.: Sense-making in social media during extreme events. *J. Contingencies Crisis Manag.* **26**(1), 4–15 (2018). <https://doi.org/10.1111/1468-5973.12193>
6. Wu, F., Huberman, B.A., Adamic, L.A., Tyler, J.R.: Information flow in social groups. *Phys. Stat. Mech. Appl.* **337**(1), 327–335 (2004). <https://doi.org/10.1016/j.physa.2004.01.030>
7. Kitsak, M., et al.: Identification of influential spreaders in complex networks. *Nat. Phys.* **6**(11), 888–893 (2010). <https://doi.org/10.1038/nphys1746>
8. Ahn, Y.-Y., Bagrow, J.P., Lehmann, S.: Link communities reveal multiscale complexity in networks. *Nature* **466**, 761 (2010)
9. Karunakaran, R.K., Manuel, S., Narayanan Satheesh, E.: Spreading information in complex networks: an overview and some modified methods. In: *Graph Theory - Advanced Algorithms Applications*, December 2017. <https://doi.org/10.5772/intechopen.69204>

10. Liu, Q., Han, L., Zou, M., Wei, O.: Spread dynamics with resistances of computer virus on complex networks. In: 2010 3rd International Conference on Advanced Computer Theory and Engineering (ICACTE), vol. 2, pp. V2-27–V2-31, August 2010. <https://doi.org/10.1109/icacte.2010.5579087>
11. Törnberg, P.: Echo chambers and viral misinformation: modeling fake news as complex contagion. *PLoS ONE* **13**(9), e0203958 (2018). <https://doi.org/10.1371/journal.pone.0203958>
12. Vosoughi, S., Roy, D., Aral, S.: The spread of true and false news online. *Science* **359**(6380), 1146–1151 (2018). <https://doi.org/10.1126/science.aap9559>
13. Alvarez-Galvez, J.: Network models of minority opinion spreading: using agent-based modeling to study possible scenarios of social contagion. *Soc. Sci. Comput. Rev.* **34**(5), 567–581 (2016). <https://doi.org/10.1177/0894439315605607>
14. Comin, C.H., da Fontoura Costa, L.: Identifying the starting point of a spreading process in complex networks. *Phys. Rev. E* **84**(5), 056105 (2011). <https://doi.org/10.1103/physreve.84.056105>
15. Albert, R., Jeong, H., Barabási, A.-L.: Error and attack tolerance of complex networks. *Nature* **406**(6794), 378–382 (2000). <https://doi.org/10.1038/35019019>
16. Stegehuis, C., van der Hofstad, R., van Leeuwen, J.S.H.: Epidemic spreading on complex networks with community structures. *Sci. Rep.* **6**, 29748 (2016). <https://doi.org/10.1038/srep29748>
17. Blondel, V.D., Guillaume, J.-L., Lambiotte, R., Lefebvre, E.: Fast unfolding of communities in large networks. *J. Stat. Mech: Theory Exp.* **2008**(10), P10008 (2008). <https://doi.org/10.1088/1742-5468/2008/10/P10008>
18. Palla, G., Derényi, I., Farkas, I., Vicsek, T.: Uncovering the overlapping community structure of complex networks in nature and society. *Nature* **435**(7043), 814–818 (2005). <https://doi.org/10.1038/nature03607>
19. Jaccard, P.: The distribution of the flora in the alpine zone. *New Phytol.* **11**(2), 37–50 (1912). <https://doi.org/10.1111/j.1469-8137.1912.tb05611.x>
20. Hu, Q., Gao, Y., Ma, P., Yin, Y., Zhang, Y., Xing, C.: A new approach to identify influential spreaders in complex networks. In: *Web-Age Information Management*, pp. 99–104, Berlin, Heidelberg, (2013). http://doi.org/10.1007/978-3-642-38562-9_10
21. Newman, M.E.J.: Finding community structure in networks using the eigenvectors of matrices. *Phys. Rev. E* **74**(3), 036104 (2006). <https://doi.org/10.1103/PhysRevE.74.036104>
22. Leskovec, J., Kleinberg, J., Faloutsos, C.: Graph evolution: densification and shrinking diameters. *ACM Trans. Knowl. Discov. Data*, **1**(1), 2–es (2007). <https://doi.org/10.1145/1217299.1217301>
23. Leskovec, J., McAuley, J.J.: Learning to discover social circles in ego networks. In: Pereira, F., Burges, C.J.C., Bottou, L., Weinberger, K.Q. (eds.) *Advances in Neural Information Processing Systems*, vol. 25, Curran Associates, Inc. pp. 539–547 (2012)