On the Percolation Centrality measure

Neta Singer   Francesco Salvi

I. Introduction

We investigate *Percolation Centrality: Quantifying Graph-Theoretic Impact of Nodes during Percolation in Networks* [1], a paper introducing a new measure for computing node centrality in the context of percolation. The measure is the first one to dynamically take into account the percolation status of each node, providing a useful tool to assess the influence of individual nodes on the transmission of some contagion and to efficiently allocate available resources. In §II we briefly summarize the paper, highlighting its main results. Then, we attempt to replicate such results by running our own simulations, discussed in §III. Finally, in §IV we consider some limitations and possible extensions of the original measure.

II. Summary of the paper

A. Background

Percolation theory has been successfully employed in a variety of applications to model the spread over a complex network of a "contagion" of some kind, propagating through the links of the network and altering the state of its nodes. A prominent example of this setting is epidemics, where we consider nodes as representing people, communities, or geographical units, and links as representing interactions between them [2, 3]. Each node is then equipped with a state that indicates its infection status, with labels such as "susceptible", "infected", and "recovered", and values corresponding to the percolation or infection percentage.

In such a scenario, it is crucial to assess the impact of each node on the contagion's transmission, i.e. how likely nodes are to disperse a disease throughout the network. In practical settings, this information is key to enabling an effective response in the early stages of an epidemic, through an efficient allocation plan of the available resources.

Over the years spanning prior to [1], several *centrality measures* have been introduced in the literature to estimate the importance of a single node v in a network, including:

- *Degree Centrality*, given simply by the degree of node v.
- *Betweenness Centrality* (BC), a family of measures that take into account the information flow through shortest paths of the network. The classical BC, introduced in [4], is defined as the fraction of shortest paths between pairs of nodes that pass through node v, averaged over the total number of pairs in the network:

$$BC(v) = \frac{1}{(N-2)(N-1)} \sum_{s \neq v \neq t} \frac{\sigma_{s,t}(v)}{\sigma_{s,t}} \quad (1)$$

Where N stands for the number of nodes, $\sigma_{s,t}$ the number of shortest paths $s \to t$ and $\sigma_{s,t}(v)$ the number of shortest paths $s \to t$ passing through v.

- *Closeness Centrality* [5], inversely proportional to the average shortest paths distance from v to any other node of the network:

$$CC(v) = \frac{1}{\sum_{i \neq v} d_g(v,i)}$$

Where $d_g(v,i)$ is the shortest path distance $v \to i$.

- Random-path centralities, expanding BC by relaxing the assumption that information flows through shortest paths, simulating instead random walks over the network. Measures of that kind include computing the expected number of times a walk pass through v (*power centrality*, [6]), the average speed at which a random message reaches v (*random-walk centrality*, [7]) or the harmonic mean of the length of paths ending in v (*information centrality*, [8]).

All those static metrics, however, are only based on the system's inherent structure and its potential evolution, but do not take into account how node importance varies over time, depending on the percolation states. At the time of [1], some early proposals of dynamical metrics had already been introduced by [9, 10] to model cascading failures in power grids, but with a topologically-driven time evolution, changing the weights of links over the network and recomputing shortest paths. Therefore, a time-dependent measure reflecting variations in percolation states was still missing.

B. Contribution

To address the problem outlined in §II-A, Piraveenan, Prokopenko and Hossain introduce *Percolation Centrality* (PC) [1], a new metric that depends on the nodes' percolation states at each time step of the spreading process. The measure is designed to be useful in the general settings previously outlined, dealing with a generic network and the spreading of some contagion. Therefore, the process of assigning and evolving percolation states is flexible and highly application-dependent, with states that in general can vary anywhere from 0 (non-percolated) to 1 (fully percolated).

Percolation Centrality builds on Betweenness Centrality by taking into account only the paths starting from a (partially) percolated node, and is thus defined as the proportion of shortest paths that pass through a given node weighted by the amount of percolation that can be spread through the path from the source:

$$PC^t(v) = \frac{1}{(N-2)} \sum_{s \neq v \neq r} \frac{\sigma_{s,r}(v)}{\sigma_{s,r}} \cdot \frac{x_s^t}{[\sum_i x_i^t] - x_v^t} \quad (2)$$

Where x_i^t represents percolation state of node i at time t and $x_s^t = 0 \implies$ no contribution for paths starting from s.

Comparing (2) and (1), it is evident that PC differs from BC by the presence of an extra weight term, measuring the ratio of the percolation that the path can send with respect to the total percolation in the network. For any given path, therefore, the higher the source node's current state is, the greater it weights the importance of the path. Notice that this weight factor is normalized to add up to 1 when considering all nodes except v.

A few interesting properties of Percolation Centrality include the following:

- If only a single node s is infected in the graph, then the percolation weight $\frac{x_s^t}{[\sum_i x_i^t] - x_v^t}$ for paths starting from s is 1, for any $v \neq s$.
- $PC(v)$ averaged over all starting nodes s that are infected gives exactly $BC(v)$, so that in expectation over the percolated sources it only reduces to the topological measure of betweenness.
- If all nodes in the graph are percolated, with the same percolation percentage, then $PC(v) = BC(v) \ \forall v$, so that the classical result is taken when the percolation states do not give preference to certain infection paths. Therefore, PC is most relevant for assessing a network at the start of a contagion, when only few nodes are infected and percolation states have high variability over the network.
- As with the BC measure, Percolation Centrality can be efficiently implemented with an algorithm running in $O(NM)$ time, with N the number of nodes and M the number of edges [11].

As an extension of (2), a version of PC taking into account the targets' infection states can also be considered. In that case, the weight factor is modified so that paths are weighted by the difference in percolation state between the source and the target, with the measure again reducing to BC if all the nodes are equally percolated.

C. Experiments and results

To illustrate how percolation centrality works and to advocate for its potential benefits, the authors showcase its usage with some simple simulations, first on toy models and then on larger random networks. In all these experiments, a very naive spread model is employed, to reflect the fact that PC is introduced as a *generic* resource allocation tool, with the model then needing to be tailored in concrete applications on a case-by-case basis. In particular, all nodes start from a percolation state $x^{(0)} = 0$, except one single infected node with $x^{(0)} = 1$. At each timestep, transmission happens with a fixed percolation probability $p = 0.2$ of changing the state of each newly infected node from 0 to 1. In this simple setting, no recovery is introduced, and all percolated nodes remain infected until the end of the simulation.

Firstly, using a small model with traceable nodes, [1] shows how, when the contagion starts from a peripheral node, PC is significantly different from BC in the early stages of the spreading, the difference remaining relevant until the number of percolated nodes undergoes a phase transition. Over the course of the process, the measure converges for each node to the classical BC. Again, this shows how Percolation Centrality

can be used for early intervention, with a suggested allocation strategy consisting of targeting the non-percolated nodes with the highest percolation centrality.

Subsequently, this strategy is put to the test with larger random graphs, both by using scale-free networks and Erdős–Rényi networks with a fixed edge probability. In this setting, "immunisation" is realised by permanently fixing the percolation state of a node to 0, which is equivalent to removing it from the network. The authors experiment with immunising, for differing values of α and β, the top $\beta\%$ of non-percolated nodes, ranked by their PC, after $\alpha\%$ of nodes have been infected. Then, they recorded the number of timesteps necessary to saturate the network, after which the number of percolated nodes passes a certain threshold. As a benchmark, the same experiment is repeated ranking the nodes by two other centrality measures: BC and the *Hop distance* (HD), i.e. the shortest-path distance to the closest infected node.

The results of their experiments are shown in Figure 1, highlighting in different colours which measure yielded the best results. When the ratio α/β is small, meaning that we are in the beginning stages of the percolation and/or we have many resources for immunising, hop distance comes out on top as a measure for choosing nodes to immunise. This is intuitive as in the case of low infection and high vaccination supply, the best strategy would be to ring-vaccinate all nodes around the few infected ones. Otherwise, when the ratio α/β is high, meaning that we are in the late stages of the percolation and/or we have only few immunisation resources, BC performs best, since in that case it is better to protect the topological core of the network. Finally, in the intermediate ranges of infections to resources, the best results are given by Percolation Centrality. In practice, intervention would likely start in these intermediate ranges of population infection and vaccination supply. The experiment therefore confirms the effectiveness of PC in aiding a society towards immunisation.

III. REPRODUCIBILITY AND SIMULATIONS

To better understand how percolation centrality works, we worked through reproducing the main results of the paper, using the same parameters and choices as the authors when possible. In particular, we used the same percolation model, giving each node a percolation probability $p = 0.2$.

For all the toy examples, we obtained the results as seen in the paper, with PC substantially different than BC in the beginning and converging to the latter over time.

For larger graphs, however, the outcome with respect to immunisation by measure was not so evident. Given that the paper did not indicate the explicit parameters used in their simulations, we experimented with Erdős–Rényi graphs with differing edge probabilities p, as well as with scale-free graphs generated in differing ways, given an expected degree sequence sampled from a power-law distribution [12, 13] or using the Barabási–Albert preferential attachment model [14].

Remarkably, all of our simulations, some of whose outcomes are shown in Figure 2, deviate considerably from the original results of Figure 1. For Erdős–Rényi networks, when the edge probability is sufficiently low, we observe a dominance of hop distance in the region where the ratio α/β is
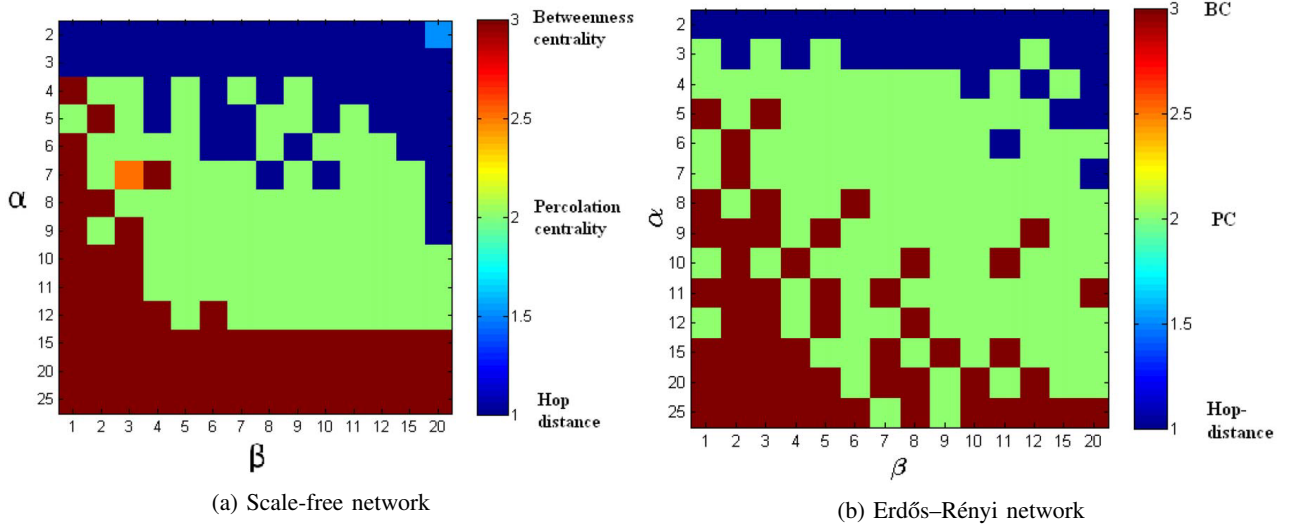
(a) Scale-free network

(b) Erdős–Rényi network

Fig. 1: Results of the immunisation experiments from the original paper, for different values of α and β [1]. The colours represent which measure gave the best results in terms of timesteps to reach saturation, averaged over 5 simulations.
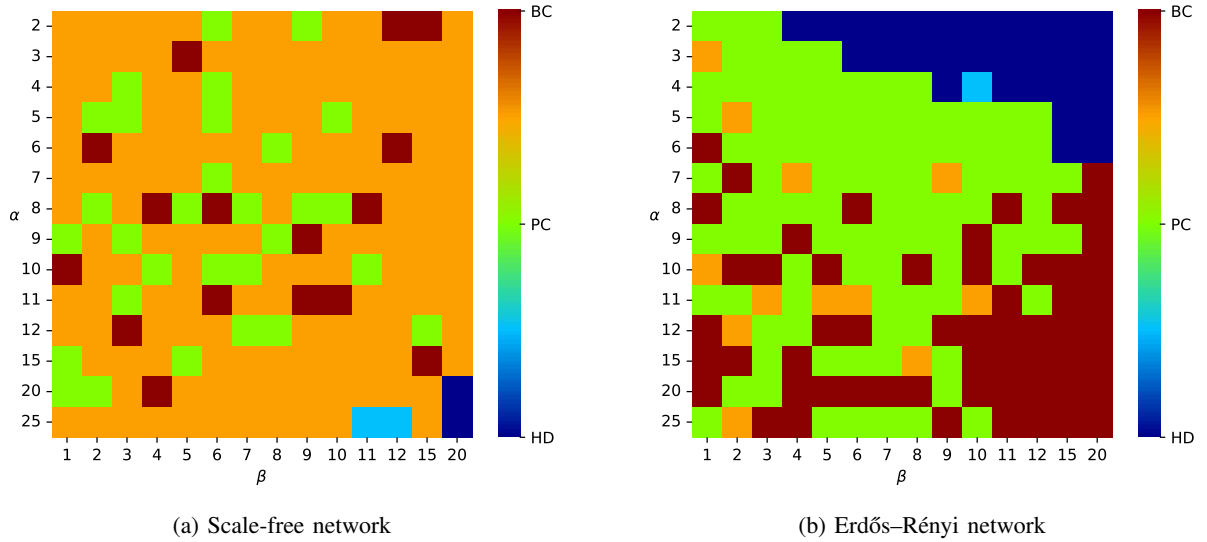

(a) Scale-free network

(b) Erdős–Rényi network

Fig. 2: Results of the immunisation experiments reproduced by us, for different values of α and β. The colours represent which measure gave the best results in terms of timesteps to reach saturation, averaged over 20 simulations.

small, as claimed in the paper, but in the intermediate to higher region of the ratio, PC and BC alternate in achieving best measure, with no observable pattern. For scale-free networks we do not obtain the results seen in the paper, and instead find that HD is almost never on top as a measure for immunisation and PC and BC almost equally split in effectiveness. Arguably, the finding that HD performs badly in scale-free networks agrees with intuition, as nodes are on average only a couple of hops away from any infected node given the scale free nature of the network. Therefore, ranking nodes by hop distance would result in a near random immunisation strategy.

The differences we find in our simulations outcomes to those seen in the paper could be explained in several ways:

1) The authors used for scale-free networks one specific graph based on a real dataset, instead of generating a random graph according to some model. This might be relevant because the assumption that real-networks are scale-free is debated [15] and log-normal distributions,

which might be a better fit than power laws, have fewer highly connected nodes.
2) The authors used a specific set of parameters, possibly tweaked or cherry-picked so that the results complied with the provided explanations.

It is also possible that other implementation choices, such as the choice of initial infection location or the manner of averaging over repetitions, have an impact on the results. In either case, the paper does not specify the specific setting for which their conclusions hold, and the results fail to generalize to broader classes of graphs or random models.

IV. LIMITATIONS AND EXTENSIONS

Following the original publication [1], Percolation Centrality has been successfully applied in a variety of scenarios, including the development of viral marketing campaigns [16], mimicking swarm intelligence in a bee colony [17], or tracking the evolution of beta-amyloid plaques in the brain [18].

However, the original measure exhibits some important limitations, that have fostered further research on the topic. In the following, we discuss some of those limitations and propose potential extensions to PC addressing such issues.

A. Computational efficiency

As for betweenness-centrality and other shortest-paths-based measures, percolation centrality is computationally very expensive to calculate, with the original algorithm running in $O(N^3)$ time. An efficient implementation based on Brandes' algorithm [11] can reach $O(NM)$, but for big-scale applications with large and well-connected graphs this is still too slow. Recently, randomized algorithms have been proposed for faster computation of a PC approximation running in about $O(M)$ or $O(N \log M \log \operatorname{diam}(G))$, accounting also for target node states [19, 20]. An effort is also spent in developing parallel implementations of Brandes' algorithm, speeding up computations by using larger hardware and GPUs [21].

B. Weighted centrality

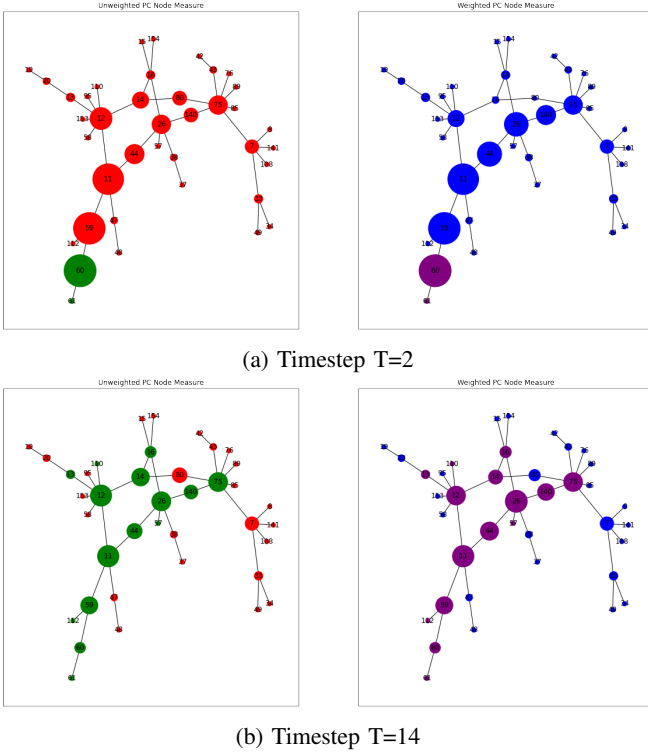(a) Timestep T=2

(b) Timestep T=14

Fig. 3: Network at time $T = 2$ and $T = 14$ of the percolation process. The infection begins at node 61, and nodes 140, 7, and 26 are given super-spreader status. Classical PC is shown in red for uninfected nodes and green for infected ones, while weighted PC is shown in blue and purple respectively. The size of each node is proportional to its centrality.

Though Percolation Centrality performs well on a variety of network settings, the measure falls short in capturing certain scenarios where a contagion evolves asymmetrically, such as in the case where percolation probabilities differ from node to node. This would be a likely situation in epidemic settings, where the spreading of an infection can depend on several factors other than just topological closeness, such as mask-wearing regulations, lifestyle choices, immunocompetence, and more. Since classical PC accounts only for the percolation states and the network topology, the measure would then fail to adequately represent the strength of percolation spread through each node. As an efficient alternative, *Weighted Percolation Centrality* could instead be used in such scenarios, introducing edge weights inversely proportional to percolation probabilities and hence computing with weighted shortest paths. Paths with higher likelihoods of infection are then captured as shorter. This weighted extension of PC is not discussed in [1], but follows closely from the analogous extension of BC.

To illustrate a situation relevant to this measure, we consider a specific scenario wherein certain nodes in a small network are designated as "super spreaders", and therefore have an increased probability (from 20% to 50%) of infecting a neighbor. In Figure 3, a network of this kind is shown, at two different timesteps of a simulated percolation process. We focus, in particular, on nodes 26 and 140, set as super-spreaders, and on nodes 80 and 14, similarly placed topologically. As can be seen in the image, at $T = 1$ PC treats these two pairs of nodes almost equally, giving them similar scores, while weighted PC assigns a much higher centrality to the first pair. In fact, being super-spreaders, nodes 26 and 140 are much more likely to receive and pass the virus than 80 and 14. Indeed, at $T = 14$, both the super spreaders and almost all of their immediate neighbors are infected, while node 80 has not been reached yet by the virus, despite being closer to its origin. In this case, the use of weighted PC would have allowed for a much more efficient early intervention than with PC, detecting from the beginning potential paths at high risk.

C. Random walks

Similarly to betweenness centrality, the definition of PC assumes that information, i.e. a contagion, flows in the network only through shortest paths. This is not the case in many real-world situations, where infections are likely to follow random paths. To compensate for this, a version of Percolation Centrality based on random walks was also proposed in the original paper [1], relaxing such assumption. However, its effect is not further investigated, leaving the discussion to potential future works. At present, random-walk percolation centrality has been studied for developing strategies for vaccines' allocation [22], but this still remains an open area for further research.

V. CONCLUSION

Percolation centrality [1], discussed in the present paper, has an important potential in a variety of scenarios involving the spreading of a contagion, helping to devise efficient strategies of resource allocation for early intervention. However, the simple recipe proposed by the original paper, consisting of choosing one measure between PC, BC and HD according to the ratio α/β (cf §II-C), does not seem to generalize well to broader settings, as shown in §III. Despite that, the measure has been successfully employed in practice, and several extensions have been proposed to overcome some of its initial pitfalls (cf §IV). Further efforts are still needed though, in particular concerning the computational complexity of the algorithm and on better understanding random-walk PC.

REFERENCES

[1] Mahendra Piraveenan, Mikhail Prokopenko, and Liaquat Hossain. "Percolation Centrality: Quantifying Graph-Theoretic Impact of Nodes during Percolation in Networks". In: *PLoS ONE* 8.1 (Jan. 2013). Ed. by Petter Holme, e53095. DOI: 10.1371/journal.pone.0053095.

[2] L.M. Sander et al. "Percolation on heterogeneous networks as a model for epidemics". In: *Mathematical Biosciences* 180.1-2 (Nov. 2002), pp. 293–305. DOI: 10.1016/s0025-5564(02)00117-7.

[3] M. E. J. Newman. "Spread of epidemic disease on networks". In: *Physical Review E* 66.1 (July 2002). DOI: 10.1103/physreve.66.016128.

[4] Linton C. Freeman. "A Set of Measures of Centrality Based on Betweenness". In: *Sociometry* 40.1 (Mar. 1977), p. 35. DOI: 10.2307/3033543.

[5] Linton C. Freeman. "Centrality in social networks conceptual clarification". In: *Social Networks* 1.3 (Jan. 1978), pp. 215–239. DOI: 10.1016/0378-8733(78)90021-7.

[6] M.E. J. Newman. "A measure of betweenness centrality based on random walks". In: *Social Networks* 27.1 (Jan. 2005), pp. 39–54. DOI: 10.1016/j.socnet.2004.11.009.

[7] Jae Dong Noh and Heiko Rieger. "Random Walks on Complex Networks". In: *Physical Review Letters* 92.11 (Mar. 2004). DOI: 10.1103/physrevlett.92.118701.

[8] Karen Stephenson and Marvin Zelen. "Rethinking centrality: Methods and examples". In: *Social Networks* 11.1 (Mar. 1989), pp. 1–37. DOI: 10.1016/0378-8733(89)90016-6.

[9] R. Kinney et al. "Modeling cascading failures in the North American power grid". In: *The European Physical Journal B* 46.1 (July 2005), pp. 101–107. DOI: 10.1140/epjb/e2005-00237-9.

[10] Valentina Cupac, Joseph T. Lizier, and Mikhail Prokopenko. "Comparing dynamics of cascading failures between network-centric and power flow models". In: *International Journal of Electrical Power & Energy Systems* 49 (July 2013), pp. 369–379. DOI: 10.1016/j.ijepes.2013.01.017.

[11] Ulrik Brandes. "A faster algorithm for betweenness centrality". In: *The Journal of Mathematical Sociology* 25.2 (June 2001), pp. 163–177. DOI: 10.1080/0022250x.2001.9990249.

[12] Fan Chung and Linyuan Lu. "Connected Components in Random Graphs with Given Expected Degree Sequences". In: *Annals of Combinatorics* 6.2 (Nov. 2002), pp. 125–145. DOI: 10.1007/pl00012580.

[13] Joel C. Miller and Aric Hagberg. "Efficient Generation of Networks with Given Expected Degrees". In: *Lecture Notes in Computer Science*. Springer Berlin Heidelberg, 2011, pp. 115–126. DOI: 10.1007/978-3-642-21286-4_10.

[14] Albert-Laszlo Barabasi and Reka Albert. "Emergence of Scaling in Random Networks". In: *Science* 286.5439 (Oct. 1999), pp. 509–512. DOI: 10.1126/science.286.5439.509.

[15] Anna D. Broido and Aaron Clauset. "Scale-free networks are rare". In: *Nature Communications* 10.1 (Mar. 2019). DOI: 10.1038/s41467-019-08746-5.

[16] Anastasia Mochalova and Alexandros Nanopoulos. *Non-intrusive Viral Marketing Based on Percolation Centrality*. 2015. DOI: 10.18151/7217433.

[17] Jiho Park et al. "A forwarding scheme based on swarm intelligence and percolation centrality in opportunistic networks". In: *Wireless Networks* 22.8 (Nov. 2015), pp. 2511–2521. DOI: 10.1007/s11276-015-1113-y.

[18] Gautam Kumar Baboo et al. "Tracking the Progression Influence of Beta-Amyloid Plaques Using Percolation Centrality and Collective Influence Algorithm: A Study using PET Images". In: (Oct. 2020). DOI: 10.1101/2020.10.12.20211607.

[19] Alane M. de Lima, Murilo V. G. da Silva, and André L. Vignatti. "Estimating the Percolation Centrality of Large Networks through Pseudo-dimension Theory". In: *Proceedings of the 26th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*. ACM, Aug. 2020. DOI: 10.1145/3394486.3403235.

[20] Alane M. de Lima, Murilo V.G. da Silva, and Andre L. Vignatti. "Percolation centrality via Rademacher Complexity". In: *Discrete Applied Mathematics* (2021). DOI: 10.1016/j.dam.2021.07.023.

[21] Athreya Chandramouli, Sayantan Jana, and Kishore Kothapalli. "Efficient Parallel Algorithms for Computing Percolation Centrality". In: *2021 IEEE 28th International Conference on High Performance Computing, Data, and Analytics (HiPC)*. IEEE, Dec. 2021. DOI: 10.1109/hipc53243.2021.00025.

[22] Benjamin J Singer, Robin N Thompson, and Michael B Bonsall. "Evaluating strategies for spatial allocation of vaccines based on risk and centrality". In: (Sept. 2021). DOI: 10.1101/2021.09.07.21263209.