9^{th} International Conference on Computational Social Science IC^2S^2
July 17-20, 2023, Copenhagen, Denmark

Towards Fairer Access to Information: Addressing Fundemental Biases in Influence Maximization

Keywords: Social networks, Influence maximization, Fairness, Equitable policies, Information diffusion

Extended Abstract

Social connections are a conduit through which individuals communicate, information propagates, and diseases spread [1]. Identifying individuals that are more likely to adopt ideas or technologies and spread them to others is essential in order to develop effective information campaigns [2], fight epidemics [3], and to maximize the reach of limited resources [4]. Consequently a lot of work has focused on identifying sets of influencers. Here we show that seeding information using traditional influence maximization methods, only benefits connected and central individuals, consistently leaving the most vulnerable behind. Our results highlight troublesome outcomes of influence maximization algorithms: they do not disseminate information equally, threatening to create an increasingly unequal society. To overcome this issue we devise a simple, multi-objective algorithm, which maximises both influence and information equity. Our work demonstrates how to find fairer influencer sets, highlighting that in our search for maximizing information, we do not need to compromise on information equality.

We define information equality in two ways. 1) In terms of frequency: how often does an individual receive information. 2) In terms of recency: how old is the received information. Using Independent Cascade Model (ICM) we simulate information diffusion processes on networks. ICMs are often used to study influence maximization in social networks and are a special case of susceptible-infected-recovered (SIR) models, where the recovery probability is fixed to 1. As such, an individual has one attempt to convince their neighbors to adopt a behavior; the neighbors, if convinced, will then try to convince their neighbors, and so on. Many methods have been proposed to identify "*the set*" of influential nodes in a network [5] through which information should be disseminated. We focus on 3 state-of-the art methods for finding influencers (Fig. 1a): highest degree [6] (HD), coreHD [7] (CHD), and degree discount [8] (DD).

We quantify information equality by comparing how nodes get information from influencers identified by the 3 heuristics, and from influencer sets selected at random (Fig 1b). To understand the implications of information inequalities we look at real-world networks (communication, interaction, and social networks varying in size from hundreds to tens of thousands of individuals). We measure fairness as the fraction of nodes that receive information at a higher frequency and speed than what is expected from a benchmark model, where influencers are selected at random. Looking across the different influencer heuristics we find that if a node is under-informed by one method, it will most likely not be better reached by any other method. Overall, we find that up to 59.5% of nodes receive information less frequently compared to if it was input at random, and information reaches 77.5% of individuals slower.

To bridge the information gap we propose a multi-objective formulation of fair influence maximization, where both spread and fairness are taken into account in the fitness of a candidate solution (influencers nodes). Traditionally influence maximization has only focused on maximizing a single objective, information spread (measured as the cascade-size). However, as literature from the field of Machine Learning and Artificial Intelligence shows, focusing solely on optimizing one parameter can lead to troubling and unfair outcomes. As such, the more fair an influencer set is, the fewer nodes will be vulnerable, so we also maximize the number of non-vulnerable nodes. To find fairer influencer seeds we then use a genetic algorithm to solve the optimization problem.

Fig. 1c shows the theoretical Pareto front (in multi objective optimization a pareto front denotes a line of optimal solutions) for a village household network. Our method demonstrates that there are seed sets which are more fair and, at the same time, equally effective at maximizing influence as the influence maximization heuristics (HD, CHD, DD). For this network, our algorithm identifies nine possible influencer sets, undiscovered by the traditional heuristics, with different trade-offs between maximizing information reach (cascade size) and the number of non-vulnerable nodes (fairness). Fig. 1d, shows our theoretical predictions are consistent with results found by numerically simulating information spread. As such, for a negligible reduction in cascade size we can choose fairer seeds that result in 10 to 17 less vulnerable households, this roughly corresponds to 6-10% fewer vulnerable nodes in the network. Figs. 1e-f illustrate the difference between nodes identified by using a state-of-the-art influencer heuristics and our approach. Figs. 1c-f shows an example for one specific network, however, our fair influence maximization method works equally well for other networks we investigated.

We believe this can act as a starting point towards a more systematic solution towards fair information access, which does not only arise in the network influence maximization context but in many other network- and computational science problems. Our algorithm is not perfect. It is a first approximation at solving this important problem. However, it shows that there exist more fair seed sets which current state-of-the-art algorithms are unable to find.

References

[1] D. Centola, "The spread of behavior in an online social network experiment," *science*, vol. 329, no. 5996, pp. 1194–1197, 2010.

[2] A. Banerjee, A. G. Chandrasekhar, E. Duflo, and M. O. Jackson, "The diffusion of microfinance," *Science*, vol. 341, no. 6144, p. 1236498, 2013.

[3] G. F. Chami, S. E. Ahnert, N. B. Kabatereine, and E. M. Tukahebwa, "Social network fragmentation and community health," *Proceedings of the National Academy of Sciences*, vol. 114, no. 36, pp. E7425–E7431, 2017.

[4] L. Beaman, A. BenYishay, J. Magruder, and A. M. Mobarak, "Can network theory based targeting increase technology adoption," *Working Paper*, 2015.

[5] F. Radicchi and C. Castellano, "Fundamental difference between superblockers and superspreaders in networks," *Physical Review E*, vol. 95, no. 1, p. 012318, 2017.

[6] R. Albert, H. Jeong, and A.-L. Barabási, "Error and attack tolerance of complex networks," *nature*, vol. 406, no. 6794, p. 378, 2000.

[7] L. Zdeborová, P. Zhang, and H.-J. Zhou, "Fast and simple decycling and dismantling of networks," *Scientific reports*, vol. 6, 2016.

[8] W. Chen, Y. Wang, and S. Yang, "Efficient influence maximization in social networks," in *Proceedings of the 15th ACM SIGKDD international conference on Knowledge discovery and data mining*, pp. 199–208, ACM, 2009.

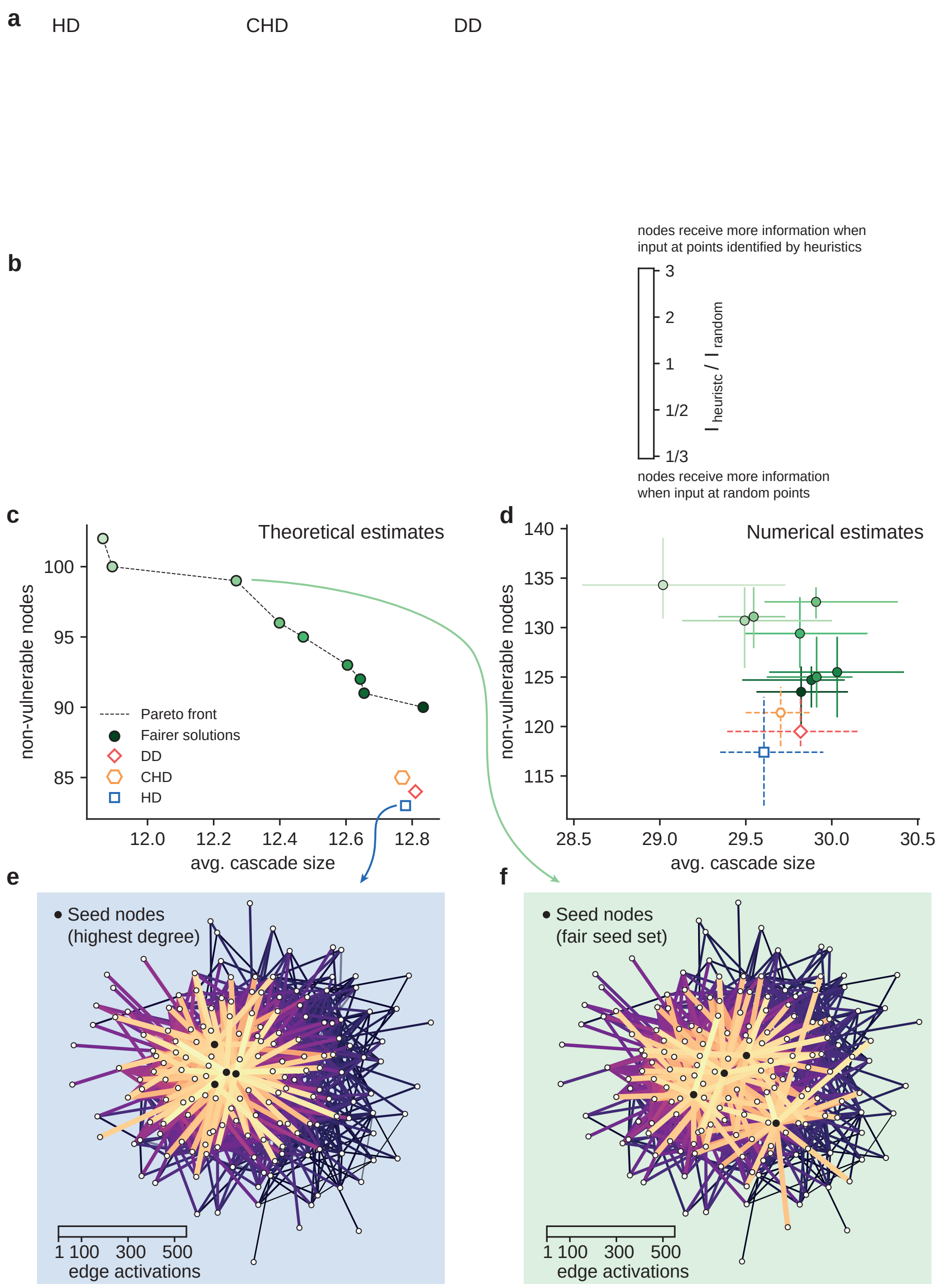

Figure 1: **Inequalities and diffusion of information in empirical networks.** **a,** Initial seed sets selected according to HD, CHD, and DD, illustrating variations in how the four methods select influencers. In this example 5% of nodes are selected as information insertion points. **b,** The effect of initial seed set on the information speed (how fast information reaches individual nodes), relative to a reference model (random seed sets). **c,** Fair influence maximization for a small social network between households in an Indian village. Theoretical Pareto front of optimal influencer sets found by our algorithm, compared to the the ones found by the influencer heuristics. Higher values of non-vulnerable nodes indicate higher values of fairness. **d,** Numerical evaluation of influencer sets using ICM. Error bars are the standard deviation of 10 realization of 10000 ICM simulations. **e,** Edge activations for the set of influencers identified by HD. Edges are colored and sized according to how often they were activated. Black nodes are original seed nodes. **f,** Edge activations for one of the fairer seed sets identified by our algorithm. Our method identifies nodes which are more evenly distributed in the network, resulting in larger parts of the network being more easily reached.