# Affect of network structure on influence maximization in dynamic networks

Habiba
hhabib3@uic.edu
Department of Computer Science,

Tanya Berger-Wolf
tanyabw@uic.edu
University of Illinois Chicago

In this work we examine how density and modularity of *dynamic networks* affect the choice of influence maximization heuristics and the optimality of the resulting solutions.

In recent years, online social networks, such as Facebook and Twitter, have provided real examples for understanding diffusion of information, ideas, and adoption of products in fast evolving networks [1, 8]. Other than social networks, applications like contamination detection in water distribution networks or spread of diseases also inherently rely on the dynamic process of diffusion [5, 6]. In the past, almost all efforts focus on "static" networks [4] in which the ties among nodes remain fixed over time. In recent years, there has been some work done in studying diffusion in "evolutionary" networks, such as co–authorship networks. However, this is also a limiting example of social networks as the assumption here is that links are never lost.

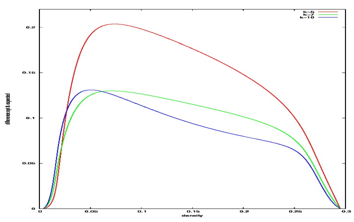
Yet, most networks are highly dynamic. In social networks, individuals maintain different connections in different contexts, such as friends or work colleagues and these relationships change with time. Similarly, the animal networks change with age, migration, and risk factors in the environment. Despite all the real world examples, there has been little work done in incorporating these relationship dynamics while analyzing the extent of influence in such networks. And most of the work that has been done focused on designing algorithms or heuristics that work for any general dynamic networks.

We analyze the global structure of explicitly dynamic networks for understanding when to employ a sophisticated algorithm vs a simple heuristic. Moreover, we propose simple structural measures that can guide this choice. Specifically, we analyze networks in term of their changing density and community structure to investigate the trends in extent of diffusion of influence. It has long been shown that maximizing the influence in a static network is an NP–hard problem [3]. Even the best approximation is infeasible for very large networks [3]. A number of heuristics have been shown to work very well in practice [7]. However, they do not provide any guarantees of performance.

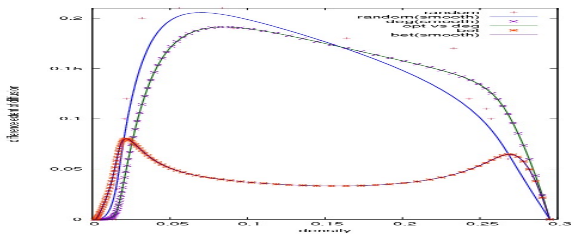
In this work, we use density of a dynamic network as an indicator of influence maximization: (a) when it is necessary to employ a sophisticated yet computationally expensive method? or (b) when even a random set of spread initiators perform as well as the best in expectation for maximizing the spread in the network? and (c) why certain heuristics like high degree as indicator of high spread work for certain networks and not for others. We show that for network densities above and below a certain threshold, the difference between the optimal and expected spread is negligible. In between the two extremes, the difference between the two approaches is markedly large. This region, rich with non-trivial and complicated structures, requires further work to devise efficient techniques for finding optimal spreaders.

We use a statistical dynamic network generative model introduced in [2] to sample dynamic networks with a wide range of properties. This generating model provides us with an extensive set of parameters through which networks of varying densities and community structures can be generated. For this analysis, we sample networks with skewed community structures that evolve over time, using this generative process.

We use the Independent cascade diffusion model [3] to simulate diffusion in networks. We estimate three main types of extent of diffusion. *i*) Optimal(or the greedy approximation) extent of diffusion. *ii*) Expected extent of diffusion by a uniformly random set of diffusion initiators. *iii*) Extent of diffusion by degree, betweenness, and Eigen values heuristics. We compare the optimal results with the expected extent of diffusion in the networks. The difference in these



**(a)** Difference in opt. & random



**(b)** Opt vs Random, Degree, Betweenness

extents highlight the disparity in the network structure. We find that networks with very low density have highly skewed structure. Which renders heuristics like highest degree very effective for diffusion maximization. On the other hand, networks that are very dense, the best methods for estimating the extent of diffusion are comparable to the baseline, that is random spreaders. However, there is an in–between pocket of density for which the difference between the optimal and random is very significant. This is meaningful since we are able to determine with only the basic knowledge about the underlying network structure, when we need to employ a sophisticated yet computationally expensive method and when something as simple as random set of spread initiators will work for diffusion maximization. We also compare the optimal spreaders to the heuristic methods in that intermediate density region to evaluate their effectiveness. We find that although degree heurisitic gives result closer to the optimal, it is the nodes that connect communities that consistently perform well in this region. For dynamic networks we use an extended definition of betweenness to find nodes that not just connect disparate communities spatially but temporally as well.

We performed extensive diffusion simulation on large set of networks sampled from the network generative process mentioned above. Figure 1a shows results of difference in optimal and expected for various seed set = 5, 10, and 15. Clearly, there are two phase shifts for when the expected performs as well as optimal. Figure 1b shows the difference in optimal and random, degree, and betweenness. Degree performs well at low density. Betweenness in the intermediate density region. And random performs comparable to optimal in high density networks.

In conclusion, we showed that dynamic networks with densities above and below a certain threshold are amenable to simple heuristics. In dense networks, in fact, there is no differentiation between optimal influence of individuals and any random set of individuals is good, given a high enough rate of diffusion. In sparse networks simple heuristics like highest degree nodes perform well. In between the two extremes, the difference between the best and a random seed set is non–trivial in expecation. That difference, in fact, depends on how modular (clumped, clustered, non uniform) the network is. Networks with such densities have complex structures that require sophisticated methods for influence maximiation.

## References


[1] S. Asur and B. A. Huberman. Predicting the future with social media. *CoRR*, abs/1003.5699, 2010.

[2] Habiba, C. Tantipananandh, and T. Y. Berger-Wolf. Dynamic networks generative model for skewed component distribution. "`http://compbio.cs.uic.edu/habiba/NetworkGenerativeModel.pdf`", 2013.

[3] D. Kempe, J. Kleinberg, and E. Tardos. Maximizing the spread of influence through a social network. In *Proc. of the Ninth ACM SIGKDD Int. Conf. on Knowledge Discovery and Data Mining*, 2003.

[4] J. Leskovec, A. Krause, C. Guestrin, C. Faloutsos, and J. VanBriesen. Cost-effective outbreak detection in networks. In *Proc. of the Thirteenth ACM SIGKDD Int. Conf. on Knowledge Discovery and Data Mining*, 2007.

[5] C. Nicolaides, L. Cueto-Felgueroso, M. C. González, and R. Juanes. A metric of influential spreading during contagion dynamics through the air transportation network. *PLoS ONE*, 7(7):e40961, 07 2012.

[6] J. paul Watson, W. E. Hart, and H. J. Greenberg. An analysis of multiple contaminant warning system design objectives for sensor placement optimization in water distribution networks.

[7] B. A. Prakash, H. Tong, N. Valler, M. Faloutsos, and C. Faloutsos. Virus propagation on time-varying networks: Theory and immunization algorithms. In *ECML/PKDD (3)*, pages 99–114, 2010.

[8] T. R. Zaman, R. Herbrich, J. Van Gael, and D. Stern. Predicting information spreading in twitter. 2010.