2021 International Conference on COMmunication Systems & NETworkS (COMSNETS) | 978-1-7281-9127-0/20/$31.00 ©2021 IEEE | DOI: 10.1109/COMSNETS51098.2021.9352737

Cost Effective Influence Maximisation

Somyadeep Shrivastava*, Dheeraj Chaudhary†
Department of Computer Science & Engineering,
Indian Institute of Information Technology Dharwad
Hubli, India
*somyadeep99@gmail.com,
†dheeraj12000@gmail.com

Yayati Gupta‡, Sanatan Sukhija§
École Centrale School of Engineering,
Mahindra University
Hyderabad, India
‡yayati.gupta@mahindrauniversity.edu.in,
§sanatan.sukhija@mahindrauniversity.edu.in

***Abstract*—In the context of virality prediction, many researchers have leveraged the existence of a core-periphery structure in a network to identify the super-spreaders of information. Topologically, the nodes in the core of a network are the most efficient spreaders. However, these nodes are less susceptible, i.e., unlikely to be influenced by the periphery nodes. Consequently, large payoffs are required to market information (ideas, products, memes, etc.) via them. In this paper, we show the presence of several non-core nodes whose spreading power is close to that of the core nodes. In the 4 real-world datasets under consideration, the number of such nodes is 7 times more than the number of core nodes on average. Given a limited budget, digital marketers can target such non-core nodes to advertise their products with lesser payoffs. Moreover, from a sociological perspective, interpersonal closeness can help in reducing the payoff further. With the help of friendship connections, we propose a cost-effective strategy to reach the influential nodes. The proposed hill-climbing based strategy can be effectively used with both, global as well as local characteristics of the nodes in a network. In terms of the cost metric, it outperforms the conventional independent cascade model by more than 5 times for the core and 2 times for the non-core super-spreaders.**

***Index Terms*—Influence maximisation, Information diffusion, Virality, Core-periphery structure**

I. Introduction

Over the last two decades, researchers have investigated the reasons behind “What makes a meme go viral on social networks?”. In the Ted talk “How Videos go Viral?” [1], YouTube’s trends manager Kevin Allocca emphasised the importance of tastemakers (highly followed celebrities or influential identities) in viralising the YouTube videos. The research problem of identifying such super-spreaders [16] has attracted a lot of attention lately. Formally known as the “influence maximisation problem”, it aims at identifying $k \in \{1, 2, 3, \dots\}$ most efficient spreaders in a network. The large body of research literature in this domain includes optimisation algorithms [12], analysis of network topology [13], and studies based on real-world meme diffusion data [18] etc.

Predominantly, in addition to the content of a meme, the connections of its spreaders play an important role in deciding its virality [7]. To make a meme go viral, the super-spreaders in a network need to be influenced. An analysis of spreading patterns on Twitter has shown that the super-spreaders tend to be the most influential people [4] such as popular celebrities, prominent political leaders etc. However, in comparison to the common mass, these super-spreaders cannot be influenced easily [2]. In this paper, we show that, in addition to the scarce influential super-spreaders, there exist numerous other efficient spreaders in real-world networks. We term such super-spreaders as “pseudo-influentials”. These nodes have a spreading power equivalent to that of the most influential super-spreaders. We propose algorithms to find a chain of connections (“friends of friends of friends...”) to make a meme reach the most efficient spreaders. The experiments on several real-world datasets suggest that the proposed algorithm is more effective as compared to the conventional information spreading strategies.

The major contributions of the paper are listed below.

1) Proving the presence of numerous non-core super-spreaders in real-world social networks.
2) Proposed a cost-effective strategy to make a meme reach the super-spreaders without the need for global information.

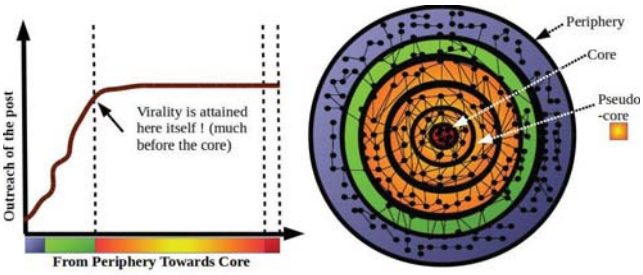
Outreach of the post
Virality is attained here itself ! (much before the core)
From Periphery Towards Core
Periphery
Core
Pseudo -core

Fig. 1. Left: correlation of shell number (on the X-axis) with spreading power (on the Y-axis), Right: An example social network showing the presence of core, pseudo-core, and periphery.

II. Related Work

Understanding the spreading of information [8; 9; 11] in online social networks is a problem of high significance. Information spreading models play a prominent role in influence maximisation [12; 13]. A number of approaches

978-1-7281-9127-0/21/$31.00 ©2021 IEEE

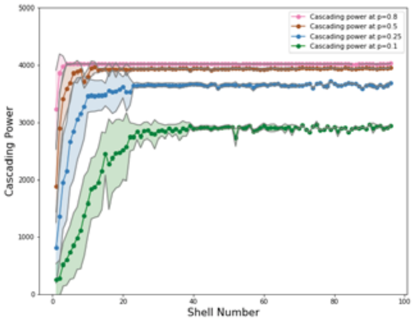

Fig. 2. Facebook Network

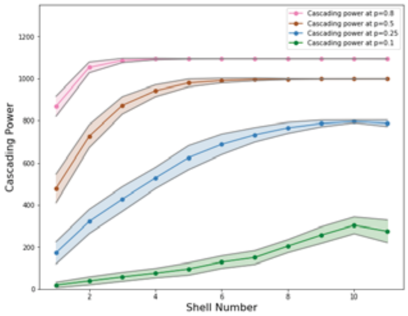

Fig. 3. Email Interaction Network

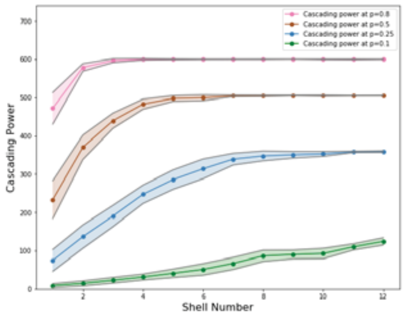

Fig. 4. Weblog Network

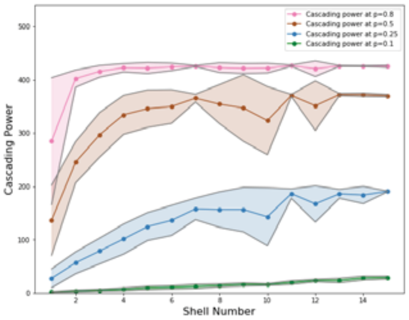

Fig. 5. Animal Interaction Network

have been used to solve the influence maximisation problem like approximation algorithms, network structure, budget minimisation [6], centrality measures and specifically coreness [3] (k-shell decomposition [13]). In many real-world scenarios, people are eager to achieve the desired influence maximization with limited time and low budget, where a minimum seed set is taken with constrained time and influence, known as budget minimizing problem. For a long time, the scientific community has considered the core nodes in a network as the super-spreaders and the periphery nodes are considered to be the seeds as they are easy to approach. Pei et al. [18] proposed the sum of degrees of neighbors (degree sum) of a node as a proxy for the coreness of a node in a network. Many other measures like leader rank [17], cluster rank [5] etc. have been proposed for calculating the spreading power of nodes in a network. Although the k-shell decomposition algorithm and its local variants identify the core nodes, this alone is not sufficient for super-spreading the information. This is because the core nodes are non-susceptible and do not readily spread a piece of information. In this regard, Gupta et al. [10] used a hill-climbing based approach to reach the core nodes in the network which creates a path from the periphery nodes to core nodes in the network. This approach is similar to the Milgram's small world experiment aimed at passing a letter from a source person to the target person through a chain of acquaintances [19]. However, unlike the Milgram experiment, the target in their approach is a set of nodes (core) instead of just one person. As Kleinfield specified that the success rate of Milgram's experiment was very low [14] when the targets were low-status people and vice versa. Being based on the shell number of nodes, the hill-climbing approach proposed by Gupta et al. required the global information of the network. Further, they did not consider the possibility of a person in the chain not forwarding the meme further. We address both of these limitations in our work.

III. Pseudo-Influentials

The current study revisits the idea of identifying super-spreaders using the k-shell decomposition algorithm. A k-core of a graph G is its maximal connected subgraph in which all vertices have degree at least k. We consider a node to have coreness/shell-number c, if it belongs to c-core but not $(c+1)$-core. We show that the spreading power does not increase at every successive shell in the network. There are several consecutive shells whose diffusion capacity (spreading power) is equivalent to that of the core. A schematic plot for this behaviour is shown in Figure 1. The diffusion capacity increases with an increasing shell number until a certain shell q, thereafter stabilising. This proves the existence of a set of non-core super-spreading nodes that are capable of making a post viral. Hereafter, we call such super-spreaders as pseudo-core (or pseudo-influentials).

A. Cascading Power of Pseudo-cores

For finding the spreading power of a shell k, we calculate the spreading power of the nodes in this shell using Independent Cascade [8] model. Given a network G, the spreading power $\delta_p(u)$ of a node u is calculated using 100 simulations of independent cascade model with u as the seed node and the infection probability p. The spreading power of node u, $\delta_p(u)$, is defined as the average of number of infected nodes over 100 iterations. The spreading power of shell k, δ_p^k is defined as the average spreading power of all nodes in this shell.

$$\delta_p^k = \frac{\sum_{v \in \tau(k)} \delta_p(v)}{|\tau(k)|}$$

Coreness proxies	**Influential shell**	**Facebook**			**Email university**			**Weblog**			**Animal interaction**		
		avg	*min*	*max*	*avg*	*min*	*max*	*avg*	*min*	*max*	*avg*	*min*	*max*
Baseline	Core	956.12	1	1836.1	156.88	4.71	299.04	94.80	1	231.73	121.62	32.35	185.20
	Pseudo-core	85.91	1	337.5	1.62	1	6.47	1.69	1	7	9.57	1	22.94
Proposed-shell	Core	65.55	1	702.52	6.53	2.12	9.625	4.10	1	7.18	64.27	25.88	104.32
	pseudo-core	61.21	1	711	1.45	1	3.5	1.45	1	4.29	5.41	1	15.55
Proposed-degree	Core	213.29	1	765.56	31.92	12.86	38.15	8.18	1	15.4	101.04	36.71	142.22
	pseudo-core	89.78	1	811	1.46	1	3.51	1.43	1	1.44	4.11	1	14.83
Proposed-h2_index	Core	188.46	1	897.7	31.33	3.31	38.76	3.86	1	7	60.20	6.24	94.17
	pseudo-core	51.02	1	631.26	1.46	1	3.62	1.45	1	4.19	5.02	1	16.51
Proposed-degree_sum	Core	200.06	1	827.87	32.81	4.81	40.39	4.73	1	8.67	66.33	9.81	97.36
	pseudo-core	49.29	1	640	1.45	1	3.48	1.43	1	4.23	3.98	1	13.4

TABLE I
Number of edges explored for the baseline and the proposed strategy (with various proxies)

where $\tau(k)$ is the set of nodes having shell number k.

The spreading (cascading) power of each shell for 4 real-world networks[15] (Facebook, Email interaction, Weblog and Animal Network) is shown in Figure [2, 3, 4, 5] respectively, where X-axis corresponds to the shell number (k) and Y-axis represents (δ_p^k) for varying values of p. It can be observed that the spreading power saturates much before the core is encountered. Furthermore, the standard deviation of the cascading power of a shell decreases with increasing shell number. This indicates that the nodes in the core/pseudo-core, unlike the periphery nodes, have equivalent spreading power.

B. Cascading power of individual pseudo-influential nodes

Figure 6 depicts the spreading power of the periphery, core, pseudo-core, and non-pseudo-core shell nodes.

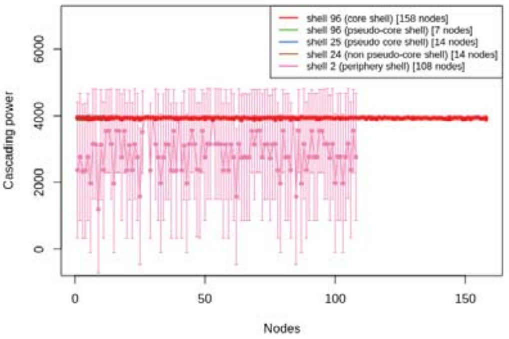

Fig. 6. Spreading power of nodes in various shells. The X-axis corresponds to the nodes. Y-axis represents the average and standard deviation of the spreading power (100 experiments).

The X-axis in the plots represent the nodes in a particular shell (Please note that this number will vary across different shells). Y-axis shows the average $\delta_p(u)$ and the standard deviation $\sigma_p(u)$ of the spreading power of nodes. It is observed that $\sigma_p(u)$ for the nodes in non-core and non-pseudo-core shells is higher as compared to the core and pseudo-core shells. Further, it is seen that the nodes of core and pseudo-core shells have similar spreading power as opposed to the other nodes.

IV. Cost Effective Influence Maximisation

Assume a company wants to advertise its product through the super-spreaders of a network. These super-spreaders such as celebrities will demand a payoff for the endorsement. As Interpersonal closeness (IC) plays an important role in influencing an individual,we propose an alternate cost-efficient strategy for super-spreading that does not involve directly approaching the super-spreaders.

Instead of targeting the core/pseudo-core nodes directly in the network, the proposed strategy chooses an individual in the lower-most shell to start the information spread. Thereafter, the company pays a reward r to the person u for each infected edge E_{uv} in the network. The rewarding process continues until the product/meme reaches a core/pseudo-core node. Hence, the overall cost incurred by the company is $r \times (the\ number\ of\ infected\ edges)$. We propose an approach that greatly reduces the advertising cost as compared to the conventional information diffusion strategy (referred to as the baseline model hereafter).

1) Baseline Model: We use a variant of the conventional independent cascade model as the baseline model. In the experiment, information diffusion starts from a seed node, say $u \in V(G)$ where $V(G)$ is the vertex set of the underlying graph G. For each edge (u, v) there is a probability, $p_{u,v} \in [0, 1]$ of information propagation across the edge. At each time step $t \in N$, each node in the G is in one of the states; infected (the node infected in the current iteration which attempts to infect its neighbors for one iteration), susceptible (a node which has not been infected before) or inactive (a node which has been infected before and has tried infecting its neighbors, thereafter achieving an inactive state). In each iteration, every infected node tries infecting its susceptible neighbors with a probability p, thereafter becoming inactive. The experiment stops when either a) there are no infected nodes in the network or b) the information cascade reaches a super-spreader.

2) Proposed Model: In the baseline model, an infected node infects all of its neighbors. This increases the number of edges explored leading to an increase in the payoff. The proposed model introduces a controlled behavior in the experiment. To reach a core or a pseudo-core node, instead of infecting all the neighbors, each node tries infecting its neighbor which has the highest spreading power (shell number), with a success probability p. If it fails to infect the most influential neighbor, it tries infecting the neighbor having the second-highest spreading power and so on. The experiment continues until a core/pseudo-core node is infected. We use shell numbers of the nodes as a measure of the spreading power.

Using Proxies (Local Measures): In real-world practice, calculating the shell number of neighbors is impractical because of the unavailability of the entire network. However, many proxies can be used as the approximation for the shell number of a node. We perform our experiments with these proxies, namely, degree, degree sum of neighbors and h-index. The cost for all experiments is calculated in the same manner as that of the baseline model.

From **Table I**, it can be observed that in comparison to the baseline, the proposed strategy takes significantly less number of edges to reach a core and a pseudo-core shell. The other important observations from the conducted experiments are listed below.

- For both, the conventional as well as the proposed model, the number of edges explored to reach a pseudo-core node is significantly lesser than that of a core node.
- The proposed hill-climbing based strategy explores a lesser number of edges to reach a core/pseudo-core node as compared to the conventional approach.
- For most experiments; when the target is a core node, the hill-climbing approach when used with shell number outperforms all other proxies.
- For most experiments; when the target is a pseudo-core node, the hill-climbing approach when used with proxies, namely, h2-index and degree-sum, outperforms the shell-number based climbing.

From our experiments, we conclude that; despite not knowing the path to reach the super-spreaders, one can still influence them with the proposed hill-climbing based strategy with the discussed proxies. The information will go viral as soon as a super-spreader shares it upon being influenced.

V. Conclusion

The study leverages the core-periphery structure in real-world social networks to show the presence of a large number of non-core nodes that can spread the information as effectively as a core node. A cost-effective information diffusion strategy has been proposed that only requires the neighborhood information (friendship connections) of a node to make a meme go viral. Digital marketing agencies with a limited advertising budget can use the proposed strategy to popularise their product. One limitation of the proposed approach is the assumption of an equal cost for every infected edge. However, for each node, an impact of both, the susceptibility and influence-ability can be considered.

References

[1] Allocca, K. Why videos go viral. *Ted talk* (2011).

[2] Aral, S., and Walker, D. Identifying influential and susceptible members of social networks. *Science 337*, 6092 (2012), 337–341.

[3] Borgatti, S. P., and Everett, M. G. Models of core/periphery structures. *Social networks 21*, 4 (2000), 375–395.

[4] Cha, M., Haddadi, H., Benevenuto, F., and Gummadi, K. P. Measuring user influence in twitter: The million follower fallacy. In *fourth international AAAI conference on weblogs and social media* (2010).

[5] Chen, D.-B., Gao, H., Lü, L., and Zhou, T. Identifying influential nodes in large-scale directed networks: the role of clustering. *PloS one 8*, 10 (2013), e77455.

[6] Dou, P., Du, S., and Song, G. Budget minimization with time and influence constraints in social network. In *Web Technologies and Applications* (2016), pp. 304–315.

[7] Fu, Y.-H., Huang, C.-Y., and Sun, C.-T. Identifying super-spreader nodes in complex networks. *Mathematical Problems in Engineering 2015* (2015).

[8] Goldenberg, J., Libai, B., and Muller, E. Talk of the network: A complex systems look at the underlying process of word-of-mouth. *Marketing letters 12*, 3 (2001), 211–223.

[9] Granovetter, M. Threshold models of collective behavior. *American journal of sociology 83*, 6 (1978), 1420–1443.

[10] Gupta, Y., Das, D., and Iyengar, S. Pseudo-cores: the terminus of an intelligent viral meme's trajectory. In *Complex Networks VII*. Springer, 2016, pp. 213–226.

[11] Gupta, Y., Iyengar, S., Saxena, A., and Das, D. Modeling memetics using edge diversity. *Social Network Analysis and Mining 9*, 1 (2019), 2.

[12] Kempe, D., Kleinberg, J., and Tardos, É. Maximizing the spread of influence through a social network. In *Proceedings of the ninth ACM SIGKDD international conference on Knowledge discovery and data mining* (2003), pp. 137–146.

[13] Kitsak, M., Gallos, L. K., Havlin, S., Liljeros, F., Muchnik, L., Stanley, H. E., and Makse, H. A. Identification of influential spreaders in complex networks. *Nature physics 6*, 11 (2010), 888.

[14] Kleinfeld, J. S. The small world problem. *Society 39*, 2 (2002), 61–66.

[15] Leskovec, J., and Krevl, A. SNAP Datasets: Stanford large network dataset collection. http://snap.stanford.edu/data, June 2014.

[16] Li, Y., Fan, J., Wang, Y., and Tan, K.-L. Influence maximization on social graphs: A survey. *IEEE Transactions on Knowledge and Data Engineering 30*, 10 (2018), 1852–1872.

[17] Lü, L., Zhang, Y.-C., Yeung, C. H., and Zhou, T. Leaders in social networks, the delicious case. *PloS one 6*, 6 (2011), e21202.

[18] Pei, S., Muchnik, L., Andrade Jr, J. S., Zheng, Z., and Makse, H. A. Searching for superspreaders of information in real-world social media. *Scientific reports 4* (2014), 5547.

[19] Travers, J., and Milgram, S. An experimental study of the small world problem. In *Social Networks*. Elsevier, 1977, pp. 179–197.