

CKS: A Community-Based K-shell Decomposition Approach Using Community Bridge Nodes for Influence Maximization (Student Abstract)

Inder Khatri^{*1}, Aaryan Gupta^{*2}, Arjun Choudhry^{*1}, Aryan Tyagi^{*2}, Dinesh Kumar Vishwakarma¹, Mukesh Prasad³

¹ Biometric Research Laboratory, Delhi Technological University, New Delhi, India

² Delhi Technological University, New Delhi, India

³ School of Computer Science, FEIT, University of Technology Sydney, Sydney, Australia

{inderkhatri999, aryan227227, choudhry.arjun, tyagiaryan82}@gmail.com, dinesh@dtu.ac.in, mukesh.prasad@uts.edu.au

Abstract

Social networks have enabled user-specific advertisements and recommendations on their platforms, which puts a significant focus on Influence Maximisation (IM) for target advertising and related tasks. The aim is to identify nodes in the network which can maximize the spread of information through a diffusion cascade. We propose a community structures-based approach that employs K-Shell algorithm with community structures to generate a score for the connections between seed nodes and communities. Further, our approach employs entropy within communities to ensure the proper spread of information within the communities. We validate our approach on four publicly available networks and show its superiority to four state-of-the-art approaches while still being relatively efficient.

Introduction

Online Social Networks are platforms for people to publicize their ideas and products. This approach, called Viral Marketing, raises the problem of Influence Maximisation (IM), which requires selecting k seed nodes to maximize the information spread in the network. Several approaches have been proposed for IM in recent years, based on local, semi-local, and global structures. Global structure-based approaches are efficient due to their consideration of whole network. They find out core nodes with maximum connectivity to the remaining network. However, in real-world situations, where number of seed nodes is very small, influence gets restricted to only a few sub-groups (or communities) in the network containing the selected core seed nodes. To overcome this, we instead consider community bridge nodes as influential seed nodes due to their connections to a larger number of communities, leading to simultaneous information propagation to these communities. We propose CKS centrality measure, which incorporates community structures and K-shell Decomposition to identify influential spreaders in a network. We define three novel measures: Community K-Shells, Community K-Shell Entropy, and CKS-Score, which qualitatively and quantitatively evaluate connections of a node to various communities.

^{*}These authors contributed equally.

Proposed Methodology

Obtaining Community K-Shell (CKS): Community K-Shell concept incorporates knowledge of information flow in a network. Obtaining CKSs comprises the following steps: identifying community structures using Louvain’s algorithm (Blondel et al. 2008); isolating communities by removing connections between different communities; and passing isolated communities through K-Shell algorithm (Kitsak et al. 2010) to obtain K-Shell scores particular to the community of each node (i.e., Community K-shell score). The higher the Community K-shell score of a node, the closer it is to the core of the given community.

Computing K-Shell Entropy (KSE): KSE evaluates connectivity of a node to different regions of a community. We formulate KSE for a node v corresponding to each community c as follows:

$$KSE_{v,c} = - \sum_{s=1}^{shells_c} K_s * \frac{\eta_{v,s}}{\eta_v} * \log\left(\frac{\eta_{v,s}}{\eta_v}\right) \quad (1)$$

$$\eta_v = \sum_{s'=1}^{shells_c} \eta_{v,s'} \quad (2)$$

where, $\eta_{v,s}$ is number of connections of node v with shell s for a given community, $shells_c$ represents the unique CKS in community c , and K_s represents K-value of given shell.

Influence of an activated node lasts only till 2-3 hops. Thus, for maximum influence over a given community, connectivity of a node to the community should be well-distributed across all its regions, which also reduces the chances of overlapping influence in a community. This requires a higher KSE score. Equation 1 represents the entropy submission over shells of the respective community weighted by the K-Value of the respective shell.

CKS-Score: CKS-Score evaluates the overall connectivity of a node to all adjacent communities to which it is connected. It is defined as:

$$CKS-Score(v) = \sum_{c=1}^{comm} NN_c * KSE_{v,c} * \eta_v \quad (3)$$

where, η_v is number of connections of node v with the respective community as shown in Equation 2, $KSE_{v,c}$ represents K-Shell Entropy for node v and community c , and NN_c represents number of nodes in community c . We compute CKS-Score for each node by considering its KSE values corresponding to each community weighted by the community’s size. This ensures a higher score for community bridge nodes connected to the most significant communities.

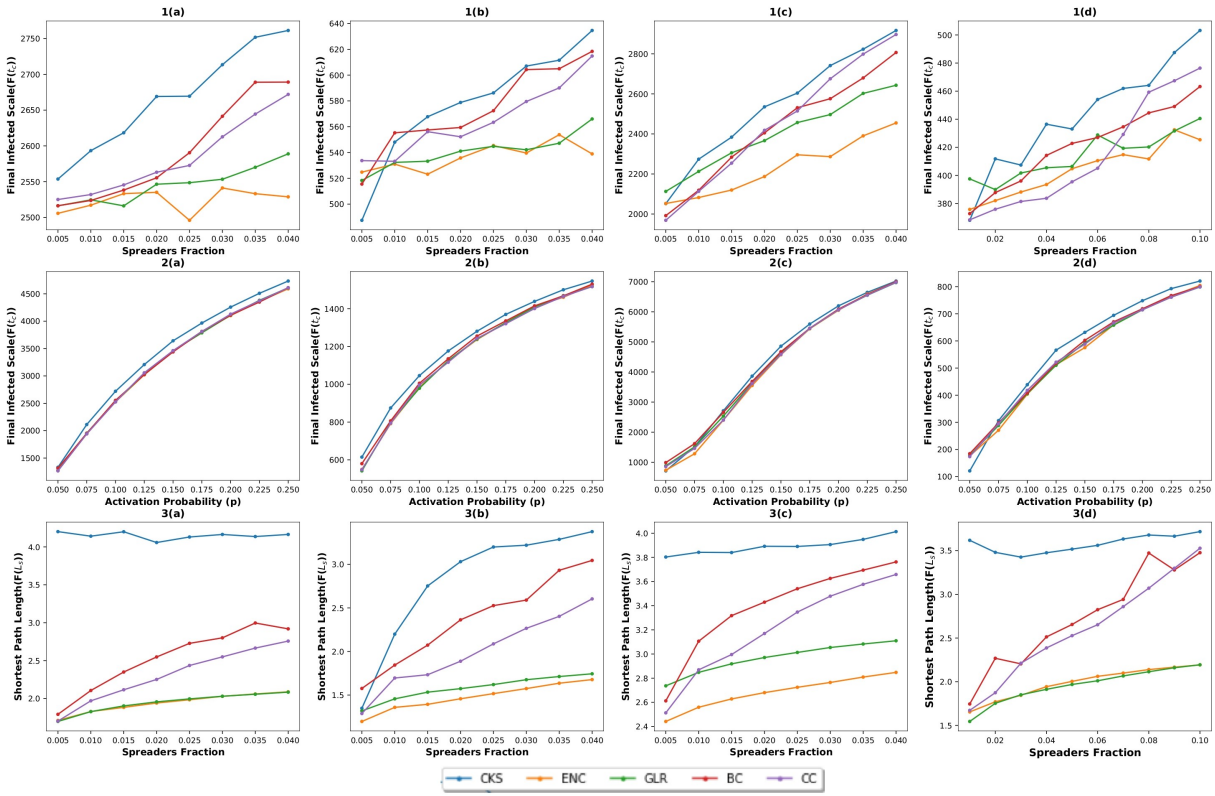


Figure 1: Results of our experiments on datasets (a) Twitch, (b) soc-Hamsterster (c) p2p-Gnutella04 (d) email-univ. (1) FIS after IC model simulation with activation probability 0.5. (2) FIS-v-p after IC model simulation with initial spreaders fraction 0.2. (3) ASPL among initial selected seed nodes vs spreaders fraction.

Dataset	CKS	ENC	GLR	BC	CC
Twitch	28.97	0.176	37.72	533.63	159.07
soc-Hamsterster	49.21	0.04	3.44	40.12	11.84
p2p-Gnutella04	83.21	0.19	54.41	1205.83	352.65
email-univ	0.83	0.01	0.71	10.09	1593.18

Table 1: Execution time for CKS and other approaches.

Experimental Results and Discussion

We evaluated CKS on four metrics: Final Infected Scale (FIS), Average Shortest Path Length (ASPL), Final infected scale vs Activation Probability (FIS-v-p), and Execution Time. We compared CKS with BC (Freeman 1977), CC (Okamoto, Chen, and Li 2008), ENC (Bae and Kim 2014), and GLR (Salavati, Abdollahpouri, and Manbari 2018) on four real-world datasets¹: Twitch, soc-Hamsterster, p2p-Gnutella04, and email-univ. We simulated each model 100 times using Independent Cascade (Goldenberg, Libai, and Muller 2001). Infection probability was set to 0.1. Table 1 and Figure 1 show our findings. We observed:

CKS consistently outperformed other core nodes-based approaches on FIS, ASPL, and FIS-v-p, while being reasonably efficient. This confirms our hypothesis that bridge nodes are more influential core nodes.

¹<https://networkrepository.com>, <https://snap.stanford.edu/snap>

CKS showed a higher inter-spreader distance than competing approaches, while also having a higher infection rate, leading to much lesser overlap between the influence of spreaders (or *rich club effect*).

References

- Bae, J.; and Kim, S. 2014. Identifying and ranking influential spreaders in complex networks by neighborhood coreness. *Physica A Statistical Mechanics and its Applications*, 395: 549–559.
- Blondel, V.; Guillaume, J.-L.; Lambiotte, R.; and Lefebvre, E. 2008. Fast Unfolding of Communities in Large Networks. *Journal of Statistical Mechanics Theory and Experiment*, 2008.
- Freeman, L. 1977. A Set of Measures of Centrality Based on Betweenness. *Sociometry*, 40: 35–41.
- Goldenberg, J.; Libai, B.; and Muller, E. 2001. Talk of the Network: A Complex Systems Look at the Underlying Process of Word-of-Mouth. *Marketing Letters*, 12: 211–223.
- Kitsak, M.; Gallos, L. K.; Havlin, S.; Liljeros, F.; Muchnik, L.; Stanley, H. E.; and Makse, H. A. 2010. Identification of influential spreaders in complex networks. *Nature Physics*, 6(11): 888–893.
- Okamoto, K.; Chen, W.; and Li, X.-Y. 2008. Ranking of Closeness Centrality for Large-Scale Social Networks. In *Frontiers in Algorithmics*, volume 5059, 186–195. ISBN 978-3-540-69310-9.
- Salavati, C.; Abdollahpouri, A.; and Manbari, Z. 2018. Ranking nodes in complex networks based on local structure and improving closeness centrality. *Neurocomputing*, 336.