Effectiveness of the K-core Nodes as Seeds for Influence Maximisation in Dynamic Cascades

SARAH ELSHARKAWY
R&D Department
ITWORX
EGYPT
sarah.elsharkawy@itworx.com

GHADA HASSAN
Faculty of
Computer & Information Sciences
The British University
and
Ain shams University
EGYPT
ghada.hassan@bue.edu.eg

TAREK NABHAN
R&D Department
ITWORX
EGYPT
tarek.nabhan@itworx.com

MOHAMED ROUSHDY
Faculty of
Computer and Information Sciences
Ain shams University
EGYPT
mroushdy@cis.asu.edu.eg

Abstract: Influence maximisation is the problem of finding a small subset of nodes in a social network, known as seed nodes, which could maximise the spread of influence. Identifying seed nodes is of interest for marketing and information dissemination purposes. One of the algorithms used to rank a single node's influence in the spreading process is the k-core decomposition. Most of the work done in this area had two main limitations: 1) worked with a static graph representation of the social network, and 2) did not tackle the possibility of the k-core being formed as a consequence of the domination of other sources such as the mass media or electronic militias and hence the limited contribution of the k-core in the influence dissemination.

In this paper, we investigate the evolution of the k-core influence in dynamic meme cascades and scrutinize the effectiveness of using the k-core nodes for the influence maximisation. We propose a measure to estimate the ability of the k-core to disseminate its influence in dynamic cascades, and another to measure the relative strength of the k-core as an influence source among other sources of influence contributing to the cascade development. Finally, we propose combining the two measures to determine the influence domination of the k-core nodes, and hence the effectiveness of targeting these specific nodes for influence maximization.

Using four real-life Twitter datasets, we evaluate the proposed measures. Due to the lack of ground truth about the various influence sources affecting the cascade, we examine the datasets for the existence of spam accounts where spam and electronic militias are one potential type of influence sources. We demonstrate how the spam existence indeed affects the correlation between the inner-most k-core size and the cascade size, and in effect, distorts the traditional evaluation of the k-core nodes as influence maximisers.

Key–Words: Influence Maximisation, K-core, Influential Spreaders

1 Introduction

It is of vital importance to identify which users act as influential spreaders that can propagate information to a large portion of the network, in order to ensure efficient information diffusion, optimise available resources, and control the spreading. The ability of influencers to initiate a large-scale spreading is attributed to their privileged locations in the underlying social networks [26, 31, 41]. A lot of research [6, 19, 32, 39, 44] focused on identifying those super-spreaders; however social cascades are dynamic by nature, and studying the impact and influence reach of those spreaders over the lifetime of a social cascade has not been considered.

A line of research [2, 19, 26, 28] showed how the most influential spreaders are consistently located at the k-core. The k-core is a maximal subgraph, where all nodes are connected to some number *(k)* of other nodes in the same subgraph. The k-core can be found by iteratively deleting nodes with degree less than k, where the degree of a node is the number of edges connected to the node. However, the propagation of a meme cascade cannot be attributed only to the influence of the super-spreaders located at the k-core, as other sources of influence, either internal such as nodes with a high degree or high betweenness centrality, or external such as mass media and social spammers, may affect the propagation. Where the betweenness centrality is an indicator of a node's centrality in a network. Hence, the effectiveness of the k-core nodes as the seed users for influence maximisation is directly related to the consistent superiority of the internal influence of the k-core nodes over the other factors.

In dynamic social media cascades, there is possibly more than one source of influence guiding the cascade growth. In addition, the dynamics of these sources and their relative strength is possibly also changing. For instance, a cascade could be initiated by some influential nodes in the social network and then the mass media becomes the influencing sources at a later point in time. This happens, for instance, when some users create a blog or video and it becomes viral and then, mass media such as popular TV shows talk about it. At this point in time, any user who adopts this video or blog, by sharing it, is either influenced by the internal influential spreaders or the external mass media. In this scenario, the influence source was internal, at the beginning, and then became internal and external at the same time. A variety of influence contribution scenarios could occur in online social networks and they differ from one cascade to another.

In this paper, we address the following limitations of the previous research:

1. Most of the work on the k-core decomposition has been conducted on static graphs. The influence-ability of the k-core in a dynamic cascade has not been studied before.

2. The implication of the multiplicity of the sources of influence was not taken into consideration. Where, despite the k-core nodes being influential spreaders, yet the majority of the cascade nodes, including the k-core nodes, may be impacted by other sources.

Specifically, we make the following contributions:

1. We propose a new measure to estimate the consistent influence-ability of the k-core nodes in dynamic cascades.

2. We propose another measure to estimate the multiplicity of the sources of influence that affect the cascade.

3. We combine the previous two measures to conclude the effectiveness of the k-core in the influence dissemination in the given cascade and hence, its candidacy to become a seed for influence maximisation.

The paper is organised as follows: Section 2 gives an overview of the related research. Section 3 defines the methodology of problem-solving. Section 4 describes the Twitter datasets used in our experiments. Section 5 explains the conducted experiments and the results. Finally, section 6 discusses and concludes this work.

2 Related Work

Two prior research directions form the basis of our paper: (1) Information spread and virality, and (2) influential spreaders and influence maximisation. Researchers have tackled various aspects of these fields, but in this section, we focus on the work related to our study.

2.1 Information Spread and Virality

The information spread investigates how information (news, rumours, etc.) propagates among people. Most of the fundamental research on the flow of information and influence through networks have been implemented in the context of epidemiology and the spread of diseases [15, 21, 38].

The virality of memes has been examined from various perspectives. A meme may become viral because it appeals to many [4, 9], but virality of a meme also depends on other factors, such as network structure, randomness, adoption patterns of influential users, timing, and many others [10, 34, 42].

In social networks, large information cascades are often driven by extrinsic events, including political campaigns [7, 22] and natural disasters [29]. External influence in networks was studied on YouTube videos [14]. The authors argued that since some videos became popular quicker than their predictions, the additional popularity must have been a result of an external influence.

2.2 Influential Spreaders and Influence maximisation

A line of research focused on identifying influential spreaders using different measurement techniques. Examples of such measurement techniques are the number of retweets, the number of followers, the number of mentions, betweenness centrality [18], and existence in k-core [3, 8, 17, 26, 43]. The authors in [33] found that the best spreaders are consistently located in the k-core across different social platforms and that the other measurement techniques are less effective in ranking the user influence. The k-core is commonly used as a measure of importance and well-connectedness for nodes in diverse applications in social networks [1].

In [5], the authors studied the idea of anchoring some nodes to the graph to remain engaged in the social network. The natural equilibrium of this model corresponds to the k-core of the social network. In [20], the authors evaluated communities based on the k-core concept, as means of evaluating their collaborative nature.

Another interesting study was conducted in [23], the authors showed that the k-cores have an important role in counter-contagions in online social networks. They stated that to start a counter-contagion to an existing contagion, one needs to search for the most influential nodes to start with. The k-core was one of the methods they proposed to identify those influential nodes. The research in [26] shows that the most efficient spreaders are not necessarily the most connected people in the network, but rather are those located within the core of the network as identified by the k-core decomposition analysis [37].

Influence maximisation is one of the fundamental problems in studying social influence. The influence maximisation problem was studied in [16, 36] as a probabilistic model of interaction. The selection of the most influential seeds was based on individual's overall effect on the network. In other studies [12, 24, 25, 27, 40] researchers dealt with this seeding selection as a discrete optimization problem. In [11–13] efforts were made to improve the efficiency of influence maximisation. The authors of [45] studied the interplay between users social roles and their influence on information diffusion.

3 Methodology and Proposed Measures

3.1 Modelling Dynamic Cascades

In previous studies, Identifying influential spreaders and influence maximisers was conducted mainly on static graphs. To model the dynamic cascades, our study is performed on periodic snapshots of the meme cascade to evaluate the progression of the cascade with respect to the k-core evolution. For a given dataset, we build the cascade that represents the information dissemination of the meme under study. In our model, a node represents a user, and an edge from a node A to another node B is created under the two conditions that: user B follows user A, and B tweeted or retweeted in the context of the same dataset *after* A has tweeted or retweeted in the same context. A node that remains on the graph without tweeting or retweeting for a 24-hour period of time is removed from the graph, as well as, all the edges linked to it. A node remains on the graph as long as it is active by tweeting or retweeting relative content to the topic of the dataset under investigation. We then take a periodic snapshot of the cascade every 24-hours. In each snapshot, the k-core decomposition is computed, in addition to the two measurements defined later in this section. We used the k-core decomposition approach proposed in [30] that enables the runtime computation of the k-cores in live distributed systems. We selected the time span to be 24-hours in order to cover the whole cycle of human change in behaviour from day to night. For the rest of this paper, we name the inner-most k-core as the k_d-core.

3.2 The Effectiveness of the k_d-core as an Influence Source

In this paper, we argue that the effectiveness of the k_d-core as an influential source is evaluated by the percentage of the other cascade nodes that it is able to disseminate the information to, and also by its consistent ability to remain the dominating influence source in the presence of other competing sources of influence especially the external sources such as mass media and electronic militias.

In the absence of ground-truth about the different sources of influence guiding the cascade, a value that asserts the k_d-core contribution in the cascade formation is crucial. For that, we propose calculating the percentage of the k_d-core descendants. There is a direct relationship between the ratio of the k_d-core successors and the strength of the k_d-core influence. This value defines whether the k_d-core is a contributing source of influence, but it does not reveal whether it is the only contributing source. A given cascade could possibly be driven by more than one source of influence either internal such as news agency accounts or accounts of famous public figures, or external such as mass media or electronic militias.

The association between the k_d-core size and the cascade size reveals the multiplicity of influencing sources controlling the cascade growth. When a cascade encounter various influencing sources, the bond between the k_d-core size and the cascade size breaks, because each portion of the cascade grows with different dynamics based on the strength of the source of influence that it follows. Moreover, in some cases, the strength of another source, other than the k_d-core, becomes very powerful to the extent that it influences the k_d-core itself. Hence, the k_d-core becomes a subsidiary of the dominating source, and the size of the k_d-core and its successors should be attributed to the influence of that dominating source.

Below, we propose two measures to estimate the contribution of the k_d-core in the cascade formation and the multiplicity of the influence sources acting on the cascade.

1. **Influence-ability of the K_d-core in Dynamic Cascades:** We measure the influence-ability of the k_d-core nodes on the propagation of a dynamic meme cascade by (S): the average percentage of the k_d-core successors with respect to the cascade size over the observed lifetime of the cascade.

$$S = \frac{\sum_{i=1}^{n} s_i}{n} \quad (1)$$

   The k_d-core successors (S) are the distinct nodes descendant from any of the k_d-core nodes. Where s_i is the total number of successors of the k_d-core nodes in a single snapshot i, divided by the i^{th} cascade size and n is the number of snapshots. The value of S directly indicates the percentage of the cascade size that is being influenced by the k_d-core. In order to avoid over-estimating the k_d-core successors, if a node C is descendant from two other nodes A and B, where A is in the k_d-core and B is not, node C would not be counted in the number of k_d-core's successors.

2. **Multiplicity of Influencing Sources in Dynamic Cascades:** To measure the consistent ability of the k_d-core to influence other nodes in the cascade, we measure the correlation (C) between the k_d-core size and the cascade size over the observed lifetime of the given cascade. The C is calculated using the Spearman's rank coefficient:

$$C = 1 - \frac{\sum d_i^2}{n(n^2-1)} \quad (2)$$

   Where d_i is the difference between the ranks of the k_d-core size and the cascade size of each snapshot i, and n is the number of snapshots. A high value of C generally implies the domination of a single source of influence which may be the k_d-core or another source either internal or external. Moreover, a low C may indicate the existence of multiple sources affecting the spread of influence.

Assuming positive values for C, we categorize the values into three levels: low [0 - 0.3), moderate [0.3 - 0.7), and high [0.7 -1). The values of S are grouped into two levels: low [0-0.5) and high [0.5 - 1]. In this section, we analyse the indication of the following four extreme cases:

- **Low S and low C:** A low C reveals the existence of various influence sources controlling the cascade growth. In addition, the low S shows that the k_d-core has a weak influence strength on the other cascade nodes. Hence, the k_d-core is not a good seed for influence maximisation.

- **High S and high C:** This is the case where all the credit of the meme spread is attributed to the k_d-core influence. The k_d-core has the favour of introducing a large number of nodes to the cascade which is revealed by the high S. Moreover, as the number of super-spreaders in the k_d-core increases, the cascade size consequently grows as well. Hence, the k_d-core nodes are the best seed candidates for influence maximisation.

- **Low S and high C:** The high C, as explained earlier, is an indication of the existence of a single influencing source that is guiding the whole cascade growth. However, in this case, the low value of S gives the insight that the k_d-core is not the source of influence.

- **High S and low C:** The high value of S indicates that the k_d-core nodes are the origins of a large portion of other nodes in the cascade. This is an obvious indication that the k_d-core nodes are in fact the super-spreaders. However, the low value of C indicates the possibility of the existence of other sources of influence albeit with a weaker strength than the k_d-core.

4 Datasets Description

In this work, we are interested in cascades that model a meme and the evolution of these cascades over a period of time. In the absence of datasets that represent such a dynamic social graph, we collected our own Twitter datasets using the free Twitter search API. We selected two general interest datasets: *Tsunami* and *Royal baby*. *Tsunami* is a natural phenomenon that led to a disaster, and *Royal baby* is the set of tweets discussing the birth of the United Kingdom's Prince's son. The other two datasets, P1 and P2, represent two competing political campaigns that were active concurrently during a constrained period of time in Egypt. The campaigns were rallying for opposite causes with a noticeable difference in societal popularity. Table 1 describes the time span, the total number of tweets collected and the count of users in each dataset.

The gathered data in each dataset represent tweets and users discussing a single meme/topic. The collected data is described by the following four groups: user profile, user's followers list, tweet information, and retweeters list.

5 Experiments and Results

In this section, we present results of two sets of experiments. In the first, we study the effectiveness of the k_d-core in dynamic cascades through analysing the strength and consistency of the k_d-core as an influence source in the observed lifetime of the meme cascade. In the second experiment, we validate our assumption that the higher C values imply the domination of a single source of influence as opposed to the multiplicity of influence sources at lower C values.

5.1 Evaluating the k_d-core as an Influence Source

Table 2 shows the results of measuring the values of S and C on each of the four datasets over the observed lifetime of the cascades. The S values in Royal baby, P1, and P2 reveal that the k_d-core derives more than a half of the cascade size. Its influence is responsible for approximately an average of 60%, 65%, and 71% of the cascade nodes respectively. While, in Tsunami, the k_d-core influence is less, with the k_d-core nodes being responsible for approximately an average of 45% of the whole cascade over its lifetime.

The results also show that a positive correlation C exists in all the datasets, yet, with values higher in Tsunami and Royal baby than in the other two political datasets. The high C in Tsunami implies that the k_d-core nodes and the other cascade nodes are being influenced similarly by a single source. And according to the low/moderate value of S, this single source is not the k_d-core. Thus, the k_d-core nodes are not the most influential sources, and the community size of this dataset cannot be attributed to the k_d-core influence. In the Royal Baby dataset, we find a very high value for C. In this dataset, the k-core nodes are the most influential nodes responsible for the information dissemination and hence the cascade size growth as well. This conclusion is enforced by the high measured value of S for this dataset.

On the other hand, in P1 and P2, we find lower values of C (0.54% and 0.47% in P1 and P2 respectively), in spite of the high values of S (65% in P1 and 75% in P2). We believe that these values reveal the existence of multiple sources of influence that causes the cascade growth to lose its ties with the k_d-core growth. The two datasets have high values of S, yet the k_d-core is only a single source of influence affecting the cascade growth. In the next section, we aim to investigate this case further.

5.2 The Spam Effect as an External Source of Influence

In the previous experiment, we showed that in P1 and P2 datasets, the S value implies the influence strength of the k_d-core nodes, yet only a moderate correlation between the k_d-core size and the cascade size is evident. We were intrigued by these results and argued that this case is an indication that more than one influence source is affecting the cascades. This section aims to further study the existence of other sources of influence in these datasets.

The spam is one type of the potential sources of influence that could affect the dissemination of influence in the cascade. In this experiment, we locate the spam accounts in our datasets and study the relationship between the number of spam accounts in each dataset and its impact on the correlation between the k_d-core size and the cascade size. We use Truthy BotOrNot [35], which is a spam detection tool that measures the likelihood of a Twitter user account being spam, to detect fake accounts in the datasets. We also counted the number of cascaded successor nodes of these spam accounts in our datasets in order to measure the portion of the cascade that is affected by the spam.

Table 3 shows the ratio of the cascade size that is either spam or a descendant from a spam account in each dataset. We notice that in P1 and P2, where the C value is moderate, the spam size is 13% and 10% respectively. Despite the high influence of the k_d-core in these two datasets, the spam size is significantly high and it negatively impacts the correlation between the k_d-core size and the cascade size. This enforces our assumption that the low value of C implies the multiplicity of influence sources acting on the cascade.

On the other hand, in Tsunami and Royal Baby the spam size is 2% and 5% respectively. The spam in these two datasets is an insignificant source of influence, hence does not contradict our assumption that the high value of C is an indication of the domination of a single source of influence. And specifically in these two datasets, the dominant source is not the k_d-core as indicated by the low to moderate values of S.

6 Discussion and Conclusion

In this paper, we studied the question of whether the k-core is an effective seed for influence maximisation

Table 1. Datasets description

Dataset name	Time span	Tweets count	Users count
Tsunami	10 March 2011 - 10 April 2011	770,083	415,642
Royal baby	2 May 2015 - 2 June 2015	137,036	130,788
P1	1 June 2013 - 1 July 2013	110,782	42,760
P2	1 June 2013 - 1 July 2013	13,446	7,821

Table 2. Spearman's correlation coefficients (C) between the k_d-core size and the cascade size, and the average percentage ratio (S) of k_d-core successors of each dataset.

Dataset name	S	C
Tsunami	0.448	0.767 (P-value = 1.01 e^{-6})
Royal baby	0.597	0.803 (P-value = 9.8 e^{-10})
P1	0.652	0.541 (P-value = 4.58 e^{-10})
P2	0.709	0.471 (P-value = 8.42 e^{-10})

Table 3. The percentage of spam accounts and their successors present in the datasets.

Dataset name	Percentage of spam and their successors
Tsunami	0.0279
Royal baby	0.0561
P1	0.1357
P2	0.1087

in a dynamic meme cascade. We presented two new measures to estimate the effectiveness of the k-core as seeds for influence maximisation by observing the average percentage of k-core successors in the cascade, as well as, the correlation between the k-core size and the cascade size on a set of consecutive snapshots of the meme cascade. We explained the implications of these two measurements and the usefulness of combining them to deduce the impact of the k-core on the influence propagation.

We conducted a case study on four real-life Twitter datasets and used the two measurements to evaluate the effectiveness of the k-core as an influential source for the given meme cascades. Furthermore, we investigated the existence of spam accounts in the datasets and their effect on the influence propagation.

Our findings indicate that, while some of the cascades are propagated due to the influence of the super-spreaders, others propagate due to an external influence such as mass media. In addition, the cascade size could be inflated by social spammers, crowd-turfers, and shallow users, who might adopt the idea momentarily but are not helpful in spreading it further. The selection of the k-core nodes, on their own, as seeds for influence maximisation in the later types of cascades is less effective.

We conclude that marketers and politicians, seeking influence maximization, would gain the ultimate influence dissemination by targeting the nodes located at the k-core in dynamic cascades having a high number of k-core descendants and a high association between the k-core size and the cascade size. A lower value of any of these two measures would consequently imply that the influence propagation would reach a smaller portion of the cascade, hence a less influence propagation.

Future work is planned to extend this research by looking into a suitable measure that augments the two proposed measurements, as a way to score the influence effect of the nodes in the cascade. In addition, it is of interest to monitor the fluctuations of correlation between the k-core size and the cascade size as a possible early sign of spam introduction in the cascade.

References:

[1] A. Adiga and A. K. Vullikanti. How robust is the core of a network? *Proceedings of the European Conference on Machine Learning and Knowledge Discovery in Databases*, 8188:541–556, 2013.

[2] M. A. Al-garadi, K. D. Varathan, and S. D. Ravana. Identification of influential spreaders in online social networks using interaction weighted k-core decomposition method. *Physica A: Statistical Mechanics and its Applications*, 468(C):278–288, 2017.

[3] J. I. Alvarez-Hamelin, L. Dall'Asta, A. Barrat, and A. Vespignani. How the k-core decomposition helps in understanding the internet topology. *ISMA Workshop on the Internet Topology*, 2006.

[4] J. Berger and K. L. Milkman. What makes online content viral? *Journal of Marketing Research*, 49(2):192–205, 2012.

[5] K. Bhawalkar, J. Kleinberg, K. L. andTim Roughgarden, and A. Sharma. Preventing unraveling in social networks: The anchored k-core problem. *Proceedings of the 39th International Colloquium, Part II*, pages 440–451, 2012.

[6] J. Borge-Holthoefer, S. Meloni, B. Gonalves, and Y. Moreno. Emergence of influential spreaders

in modified rumor models. *Journal of Statistical Physics*, 151(1-2):383–393, 1 2013.

[7] J. Borge-Holthoefer, A. Rivero, I. Garca, E. Cauh, A. Ferrer, D. Ferrer, D. Francos, D. Iiguez, M. P. Prez, G. Ruiz, F. Sanz, F. Serrano, C. Vias, A. Tarancn, and Y. Moreno. Structural and dynamical patterns on online social networks: the spanish may 15th movement as a case study. *PloS one*, 6(8):e23883, 2011.

[8] S. Carmi, S. Havlin, S. Kirkpatrick, Y. Shavitt, and E. Shir. A model of internet topology using k-shell decomposition. *Proceedings of the National Academy of Sciences of the United States of America*, 104(27):11150–11154, 2007.

[9] M. Cataldi, L. Di Caro, and C. Schifanella. Emerging topic detection on twitter based on temporal and social terms evaluation. In *Proceedings of the Tenth International Workshop on Multimedia Data Mining*, MDMKDD '10, pages 4:1–4:10, New York, NY, USA, 2010. ACM.

[10] D. Centola and M. Macy. Complex contagions and the weakness of long ties. *American Journal of Sociology*, 113(3):702–734, 2007.

[11] N. Chen. On the approximability of influence in social networks. In *Proceedings of the Nineteenth Annual ACM-SIAM Symposium on Discrete Algorithms*, SODA '08, pages 1029–1037, Philadelphia, PA, USA, 2008. Society for Industrial and Applied Mathematics.

[12] W. Chen, C. Wang, and Y. Wang. Scalable influence maximization for prevalent viral marketing in large-scale social networks. In *Proceedings of the 16th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, KDD '10, pages 1029–1038, New York, NY, USA, 2010. ACM.

[13] W. Chen, Y. Yuan, and L. Zhang. Scalable influence maximization in social networks under the linear threshold model. In *Proceedings of the 2010 IEEE International Conference on Data Mining*, ICDM '10, pages 88–97, Washington, DC, USA, 2010. IEEE Computer Society.

[14] R. Crane and D. Sornette. Robust dynamic classes revealed by measuring the response function of a social system. *Proceedings of the National Academy of Sciences*, 105(41):15649–15653, 2008.

[15] D. J. Daley and D. G. Kendall. Epidemics and rumors. *Nature*, 204:225–228, 1964.

[16] P. Domingos and M. Richardson. Mining the network value of customers. In *Proceedings of the Seventh ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, KDD '01, pages 57–66, New York, NY, USA, 2001. ACM.

[17] S. Dorogovtsev, A. Goltsev, and J. Mendes. k-core organization of complex networks. *Physical Review Letters*, 96:4, 2006.

[18] L. C. Freeman. Centrality in social networks conceptual clarification. *Social Networks*, 1(3):215–239, 1979.

[19] C. Giatsidis, D. Thilikos, and M. Vazirgiannis. Evaluating cooperation in communities with the k-core structure. In *Proc. Int. Conf. Advances in Social Networks Analysis and Mining*, pages 87–93, 2011.

[20] C. Giatsidis, D. M. Thilikos, and M. Vazirgiannis. Evaluating cooperation in communities with the k-core structure. *Proceedings of the 2011 International Conference on Advances in Social Networks Analysis and Mining*, pages 87–93, 2011.

[21] W. GOFFMAN and V. A. NEWILL. Generalization of epidemic theory: An application to the transmission of ideas. *Nature*, 204:225 – 228, 1964.

[22] S. González-Bailón, J. Borge-Holthoefer, A. Rivero, and Y. Moreno. The Dynamics of Protest Recruitment through an Online Network. *Scientific Reports*, 1, 2011.

[23] A. Guille, H. Hacid, C. Favre, and D. A. Zighed. Information diffusion in online social networks: a survey. *ACM SIGMOD Record*, 42(2):17–28, 2013.

[24] D. Kempe, J. Kleinberg, and E. Tardos. Maximizing the spread of influence through a social network. In *Proceedings of the Ninth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, KDD '03, pages 137–146, New York, NY, USA, 2003. ACM.

[25] D. Kempe, J. Kleinberg, and E. Tardos. Influential nodes in a diffusion model for social networks. In *Proceedings of the 32Nd International Conference on Automata, Languages and Programming*, ICALP'05, pages 1127–1138, Berlin, Heidelberg, 2005. Springer-Verlag.

[26] M. Kitsak, L. K. Gallos, S. Havlin, F. Liljeros, L. Muchnik, H. E. Stanley, and H. A. Makse. Identification of influential spreaders in complex networks. *Nature Physics*, 6:888–893, 2010.

[27] J. Leskovec, A. Krause, C. Guestrin, C. Faloutsos, J. VanBriesen, and N. Glance. Cost-effective outbreak detection in networks. In *Proceedings of the 13th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, KDD '07, pages 420–429, New York, NY, USA, 2007. ACM.

[28] J.-H. Lin, Q. Guo, W.-Z. Dong, L.-Y. Tang, and J.-G. Liu. Identifying the node spreading influence with largest k-core values. *Physics Letters A*, 378(45):3279 – 3284, 2014.

[29] X. Lu and C. Brelsford. Network structure and community evolution on twitter: Human behavior change

in response to the 2011 japanese earthquake and tsunami. 4:6773 EP –, Oct 2014. Article.

[30] A. Montresor, F. De Pellegrini, and D. Miorandi. Distributed k-core decomposition. In *Proceedings of the 30th Annual ACM SIGACT-SIGOPS Symposium on Principles of Distributed Computing*, PODC '11, pages 207–208, New York, NY, USA, 2011. ACM.

[31] S. Pei and H. A. Makse. Spreading dynamics in complex networks. *Journal of Statistical Mechanics: Theory and Experiment*, 2013(12):P12002, 2013.

[32] S. Pei, L. Muchnik, J. S. Andrade, Z. Zheng, and H. A. Makse. Searching for superspreaders of information in real-world social media. *Scientific Reports*, 4, July 2014.

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

[34] H. Pinto, J. M. Almeida, and M. A. Gonçalves. Using early view patterns to predict the popularity of youtube videos. In *Proceedings of the Sixth ACM International Conference on Web Search and Data Mining*, WSDM '13, pages 365–374, New York, NY, USA, 2013. ACM.

[35] J. Ratkiewicz, M. Conover, M. Meiss, B. Goncalves, S. Patil, A. Flammini, and F. Menczer. Truthy: mapping the spread of astroturf in microblog streams. *Proceedings of the 19th ACM International Conference on Information and Knowledge Management, ACM*, pages 249–252, 2011.

[36] M. Richardson and P. Domingos. Mining knowledge-sharing sites for viral marketing. In *Proceedings of the Eighth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, KDD '02, pages 61–70, New York, NY, USA, 2002. ACM.

[37] S. B. Seidman. Network structure and minimum degree. *Social Networks*, 5(3):269–287, 1983.

[38] E. Stattner and N. Vidot. Social network analysis in epidemiology: Current trends and perspectives. *International Conference on Research Challenges in Information Science*, pages 1–11, May 2011.

[39] M. Trusov, A. V. Bodapati, and R. E. Bucklin. Determining Influential Users in Internet Social Networks. *Journal of Marketing Research*, 47(4):643–658, Aug. 2010.

[40] Y. Wang, G. Cong, G. Song, and K. Xie. Community-based greedy algorithm for mining top-k influential nodes in mobile social networks. In *Proceedings of the 16th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, KDD '10, pages 1039–1048, New York, NY, USA, 2010. ACM.

[41] D. Watts, P. Dodds, J. D. served as editor, and T. E. served as associate editor for this article. Influentials, networks, and public opinion formation. *Journal of Consumer Research*, 34(4):441–458, 2007.

[42] L. Weng, A. Flammini, A. Vespignani, and F. Menczer. Competition among memes in a world with limited attention. *Scientific Reports*, 2(335), 2012.

[43] S. Wuchty and E. Almaas. Evolutionary cores of domain co-occurrence networks. *BMC Evolutionary Biology*, 5(24), 2005.

[44] W. Xu, W. Liang, X. Lin, and J. X. Yu. Finding top-k influential users in social networks under the structural diversity model. *Information Sciences*, 355356:110 – 126, 2016.

[45] Y. Yang, J. Tang, C. W. ki Leung, Y. Sun, Q. Chen, J. Li, and Q. Yang. Rain: Social role-aware information diffusion. In *AAAI'15*, AAAI'15, 2015.