6-12-2017

Viral Marketing for Smart Cities: Influencers in Social Network Communities

Madhura Kaple
San Jose State University

Ketki Kulkarni
San Jose State University

Katerina Potika
San Jose State University, katerina.potika@sjsu.edu

Recommended Citation

Madhura Kaple, Ketki Kulkarni, and Katerina Potika. "Viral Marketing for Smart Cities: Influencers in Social Network Communities" *2017 IEEE Third International Conference on Big Data Computing Service and Applications (BigDataService)* (2017): 106-111. https://doi.org/10.1109/BigDataService.2017.46

Viral Marketing for Smart Cities: Influencers in Social Network Communities

Madhura Kaple
Computer Science Department
San Jose State University
San Jose, CA, USA
madhuramukesh.kaple@sjsu.edu

Ketki Kulkarni
Computer Science Department
San Jose State University
San Jose, CA, USA
ketki.kulkarni@sjsu.edu

Katerina Potika
Computer Science Department
San Jose State University
San Jose, CA, USA
katerina.potika@sjsu.edu

***Abstract*—Social networks are used by cities primarily for announcing local-area events, but also for increasing engagement of citizens in votes and elections. Given the current plethora of heterogeneous social networks, city administrators can benefit from social networks to promote initiatives, which are important to a current smart city as well use them to discover future needs in order to manage resources more efficiently. Our focus in this paper is how we can adapt commercial and viral marketing techniques to smart city systems to influence the behavior, opinion and choices of citizens in order to improve their well being and that of the whole society as well as predicting future trends and events.**

***Index Terms*—smart city; social network; community detection; influential**

I. INTRODUCTION

Today, social networks form an important communication platform. People located at geographically distinct locations can communicate with each other via various social networks. These networks can be modeled as graphs where people form the vertices and the relation between the people form the edges. In these networks, people sharing common interests form communities. Members of these communities can effectively share information and communicate with each other. Humans have a natural tendency to get influenced by the decisions of others. In a social network too, some people act as key influencers such that they extend their influence to others. Thus, by targeting such influencer seed sets we can do better marketing of a product. A possible scenario of using social networks in city planning is to increase public participation by conducting polls to seek majority opinion of citizens. In addition, social networks are also helpful for the efficient resource management of cities, e.g., estimating the scale and size of events.

A. Motivation

Smart cities use modern information technology measures to build a well planned and a resource efficient city. Building such a smart city is a gradual process, where identifying current problems that a city faces is the primary step. Examining the different factors and players of each local government can be the first step towards proposing smart city initiatives [1]. The basis of an integrative framework, that will cope with the top challenges of a modern smart city, should be operational cost-wise and time-wise efficient and improve city management [2].

Citizens play an important role in identifying such challenges. Involvement of the public is important to build a smart city. People must voice their opinion on various issues faced as well as share ideas for the betterment of the city with others. Social networks is widely used as a communication platform to share information. Various social networking sites like Facebook, Twitter etc provide a central location for sharing information among remote users. The like, comment and forming of groups features of Facebook or re-tweet, follow features of Twitter provide an effective way of sharing information. Micro blogging is another popular social medium where users express their opinions about any issue with others. Use of these social networks will benefit in building the smart cities in the planning and managing of resources.

People express different opinions about different issues in social networks. We can group these people in the network, based on the common issues as well as common sentiments about a particular issue to form different communities. Communities formation will lead to better sharing and in depth analysis of an issue as all people in a community are facing or sharing interest in that common issue. The communities, can propose solutions and ideas for overcoming the problems or challenges faced using technology. Polls and surveys can be conducted via social medias, to finalize such proposals. People can also keep track of the status of the implementation of plan via social networks. Thus, use of social networks increases the public participation in smart cities and enhance local government accountability.

In each community, some people are more active in voicing their opinions. For example, some micro bloggers are more popular and have large number of followers, some journalists are more effective in spreading the information. They act as influencers in their network. The opinion and behavior of the rest of the people in participating in that network is influenced by such influencers. Their positive comments and writing motivates people. Finding such influential individuals, called influencers, in a community will influence the community at large. This will benefit in campaigns of spreading awareness, increasing public participation in the planning process etc. Thus, finding influencers in communities can lead to effective and organized planning to build smart cities.

B. Problem Definition

A social network is represented as a graph $G = (V, E)$, where V is the set of vertices (people, users or citizens) and E is the set of their edges describing their connections.

In the first phase we identify the communities in the social network. We can construct these communities based on every issue or opinion posted on the social media. Identifying such communities is the community detection problem (see [3], [4]) and this will allows us to evaluate each vote or comment, relation with the current topic and predict missing information. This gives a clear segregation of people in city by the common views shared by the people.

In the second phase, we will determine the k-influencers set in each community. Identifying such a set of k total influencers is known as the influence maximization problem. The influence maximization problem was formulated as a discrete optimization problem by Kempe in his paper [5]. Moreover, the potential influence between the users determines the edge weights. In our approach the edges are unweighted. Given is a graph $G = (V, E)$ and a budget k; the goal is to find a set, called seed set, of k-users such that on activating these k vertices as influencers they can spread the maximum influence to the other users of the social graph. The influence spread occurs in discrete steps. Finding influencers in each community can play a vital role in spreading the awareness among people of each community.

II. Framework

We propose a two phase framework that uses social networks to enhance everyday operations of smart cities. More specific, in the first phase we form communities based on the structure of the underlined graph. Each edge can be formed using some issue that is an active topic, an opinion or/and some sentiments that a pair of people share and can be extracted from keywords search in comments. We use an agglomerative hierarchical clustering approach to discover communities. In the second phase, we identify the k-influencers set of each community that was created in the first phase.

A. Community Detection

The problem of community detection is to cluster people into one group based on a common opinion and comment. This can be achieved in two steps. In the first step, we construct a network graph where each vertex represents an individual user and edges between vertices are based on topic similarity such as common comment, same opinion or sentiment for a particular topic. In the second step we use the generated graph and apply a community detection algorithm. For our experiments we have used an agglomerative hierarchical clustering approach known. In this approach each vertex is assigned to a community iteratively, based on some common property they possess. One of the algorithm that uses this approach is the Louvain algorithm [4], which is a fast greedy optimization algorithm. It iteratively combines two communities in order to maximize the network modularity Q. The network modularity is calculated as follows:

$$Q = \frac{1}{2m} \cdot \sum_{i,j} \left[A_{ij} - \frac{k_i k_j}{2m} \right] \cdot \delta(c_i, c_j) \qquad (1)$$

Here $m =$ the number of edges, A_{ij} represents the weight of an edge between vertices i and j, k_i is the degree of vertex i, c_i represents the community to which i belongs to and δ is a function such that $\delta(u, v) = 1$, if $u = v$ else 0.

The Louvain algorithm has two parts. In part 1, all the vertices of the graph are assigned to different communities, i.e. the number of communities is equal to the number of vertices in that graph. For each vertex i, the gain in modularity is calculated by considering each neighbor of i. A vertex i is placed in a community for which the gain in modularity is the maximum, provided that the gain modularity is positive. If the gain in modularity for each of its neighbors is negative, then i remains in its original community. This process continues till no vertex in the graph can further be assigned to a different community or improve the modularity.

Part 2 consists of creating a new graph where each vertix corresponds to a community as obtained in part 1. Edges between the same communities correspond to self-loops. Once this is done, a new pass starts and part 1 of the algorithm is repeated to the new graph etc. This process is continued till no changes can be done and the maximum modularity is attained. At the end of this algorithm, we get the hierarchy of communities, i.e., communities within communities. The height of the hierarchy is equal to the number of passes.

This phase will iteratively cluster together people that share common views and/or sentiments while keeping track of the quality of the formed communities, i.e., Q. Once the people are segregated into different communities, most influential or active individuals are found in Phase II as described next.

B. Finding Influencers

Recall that in the influence maximization problem we seek to identify a seed set of k users in a social network that will cause a maximum spread of information, regarded a specific topic or issue, to the remaining users of the social network. This problem is NP-hard and there exists a greedy approximation algorithm, with a 63% approximation ratio [5], that solves the problem. This greedy algorithm inserts vertices into the seed set iteratively, one by one, by adding in each iteration the vertex that manages to maximize the influence spreading or has the maximum centrality value. In graphs the centrality of a vertex defines it's importance in the network. Some of the most common centrality are degree centrality, eigenvector centrality, betweeness centrality, PageRank centrality. In the paper, we will be using the PageRank centrality based approach for the seed set of influencers detection. PageRank centrality accounts for the importance of vertices that a vertex is connected to [6].

A vertex once activated as influencer or influenced remains in the active state. If a graph consists of cycles it causes influence feedback of a vertex to itself. This will deviate the information diffusion measurement. Thus, we construct an acyclic spanning graph starting with the most central vertex in the original graph. The advantage of constructing an acyclic spanning graph to identify the seed set is that the acyclic spanning graph removes the cycles present in the original social graph and thus make the running time better [7]. We can then detect the seed set of influencers by finding the k-top vertices based on their PageRank centrality in the acyclic spanning graph. Hence, we observe that the seed set constructed using acyclic spanning graph is done faster as compared to the seed set constructed from the original social graph.

C. Algorithm

The steps of our Algorithm is summarized in the following.

Algorithm 1 Finding influencers inside social network communities.

Input: A graph $G = (V, E)$ built from a dataset.

Output: Partition of vertices in disjoint communities and find a subset $K \subseteq V$ such that remaining vertices V/K are influenced.

1) Phase I: Finding communities
    a) Run Louvain algorithm on the input graph G.
    b) Divide detected communities into subgraphs.
2) Phase II: Finding influencers
    a) Find the most central node u in every subgraph using PageRank centralities.
    b) Construct an acyclic spanning graph using node u as a starting point.
    c) Find k-top central nodes on the acyclic spanning graph using PageRank centralities.

III. Implementation Details and Results

The objective of the experiments is to find influencers present inside communities in synthetic and real-world data sets. To conduct the experiments, we have used python [8] and igraph library [9] along with some synthetic and real-world social networks.

We are generating a random graph, which is based on the Erdos-Rényi model [10]. Given the number of vertices n and a parameter p $(0 \leq p \leq 1)$, this model constructs a graph $G_(n, p)$ such that for each pair of vertices (i, j) it generates an edge joining i and j with probability p. Specifically, for our experiments we have used the igraph implementation of this model were it allows to specify the number of edges m instead of the probability p.

The random graph that was generated is an undirected graph with 70 vertices connected by 120 edges. After applying Phase I of Algorithm 1, the 6 communities were detected as shown in Figure 1. Vertices belonging to the same community are colored with the same color. Within each of these communities we have found two influencers ($k = 2$). The influencers are shown in Figure 2 as red enlarged vertices.

Although these computer generated random models give controlled and appropriate test cases, it is very important to test our algorithm against real-world dataset. For this purpose

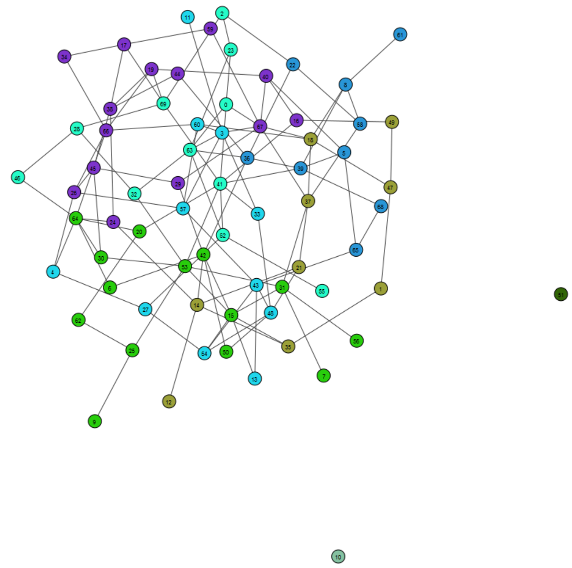

Fig. 1. Phase I: Six Communities in a Random Graph.

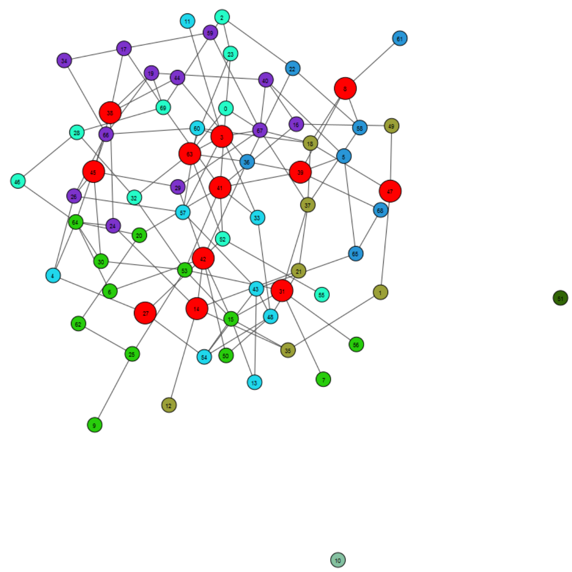

Fig. 2. Phase II: Two Influencers in each community of a Random Graph. Influencers are shown with red color enlarged.

we have selected two real-world datasets. The first one is the famous Zachary's karate club [11]. Zachary observed one karate club and built a friendship network based on the members of this club. This is an undirected network with 34 vertices and 78 edges. When we used this dataset with our algorithm we found four communities as shown in Figure 3. Although the original karate club was split into two groups, the Louvain algorithm splits this network in 4 communities. Within each these communities we found one influencer in each community ($k = 1$). The karate club was split in real life into two groups because of the disagreement between two members, which are labeled by 0 and 33. Our algorithm correctly identifies these as influencers as shown in Figure 4 along with two additionally ones.

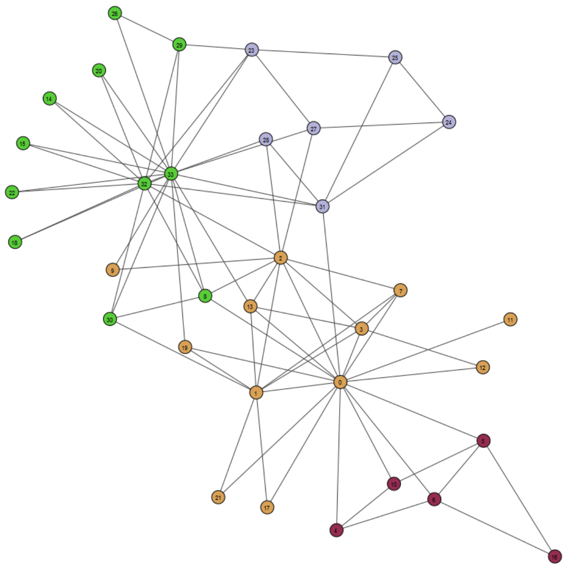

Fig. 3. Phase I: Four Communities in the Karate Club data set.

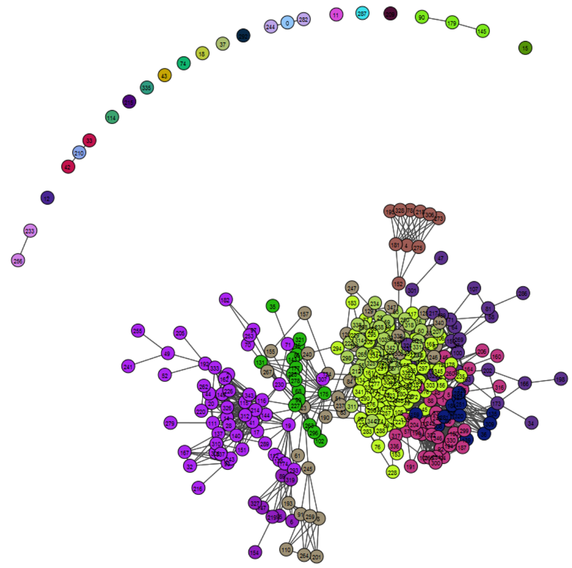

Fig. 5. Phase I: Ten Communities in the Facebook dataset. Communities are shown with different colors.

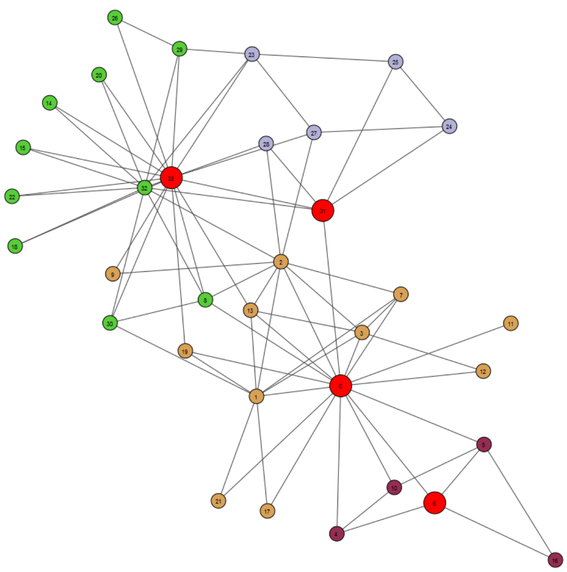

Fig. 4. Phase II: One Influencer in each community in the Karate Club data set. Influencers are shown with red color enlarged.

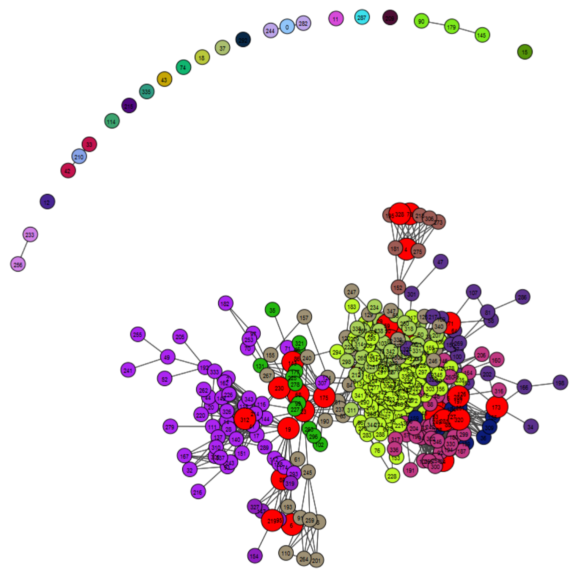

Fig. 6. Phase II: Three influencers in each community of Facebook dataset. Influencers are shown with red color enlarged.

The second real-world data set we are using is from Facebook [12]. This network is made up of 4039 individuals and their relationships and is constructed by conducting a survey amongst Facebook users. Thus, it has 4039 vertices and 88234 edges. We used a sample of this data set for our simulations which consists of 348 vertices and 5038 edges. Phase I of Algorithm 1 produces ten communities as shown in Figure 5. Different communities are colored with different colors. After Phase 2 of Algorithm 1 we get the influencers shown with red color and enlarged in Figure 6. Here, each community has three influencers ($k = 3$).

IV. Related Work

Social networks have large number of applications in domains like education, entertainment, public awareness, marketing, health-care etc. Due to such wide range of applications large amount of research has been carried out in social networks. Community Detection and Influence Propagation are two such areas which extensively studied and connected to various social network applications.

A. Community Detection

The problem of community detection requires breaking large networks into small chunks, clusters or groups of vertices which possess similar properties [4]. The vertices belonging to the same cluster form a strong network. Finding such clusters in large networks is usually computationally expensive and hard. Therefore, many heuristic algorithms have been proposed to efficiently find communities within these networks.

There are four methods for graph clustering: graph partitioning, spectral clustering, partitional clustering and hierarchical clustering (see [13]). The problem of graph partitioning is explained as a division of vertices into partitions. The size of each partition is fixed and the number of edges connecting two different partitions should be minimal. These partitions are called cuts. Finding such cuts is a NP-hard problem. One of the algorithms that uses that approach is proposed in [14]. It is based on an evaluation function that is defined as the difference of intra-community edges and inter-community edges. During the local search process, this algorithm only accepts better neighbor solutions and rejects all worse solutions. Hence this approach gives mostly local optimal solutions. Spectral clustering approaches include all the methods that group the set of vertices into a community using eigenvectors of matrices. Most of the algorithms based on this method use Laplacian matrices. Most of the social networks have hierarchical structures, i.e., clusters within clusters and so on. This approach aims to find such multilevel structures in social networks. This approach divides the set of vertices into communities based on vertex similarity. This method is further classified as Agglomerative and Divisive.

One of the most famous agglomerative algorithm is the Louvain algorithm [4], which we are using in our framework and our experiments. Another algorithm is proposed by Newman et al. [15]. This algorithm finds the largest change in the value of modularity Q for two adjoining communities and then combines those two communities. Finding the change in the value of Q is a time-consuming process and hence this algorithm instead of maintaining an adjacency matrix and then calculating ΔQ, maintains and updated matrix of ΔQ. ΔQ_{ij}, which is calculated for only those communities which are joined by at least one edge. This process is continued until only one community remains. The running time of this algorithm is $\mathcal{O}(md \log n)$, where m is the number of edges, d is the depth of the dendrogram and n is the number of vertices. The divisive approach is a top-down approach in which a large network graph is iteratively broken down into different chunks based on centrality measures. One of the most popular algorithms is defined by Girvan et al. [3]. It uses edge betweenness instead of using traditional vertex betweenness. It removes edges in decreasing order of their betweenness ratio. The algorithm starts by calculating edge betweenness for all edges and then removing an edge one-by-one with betweenness ratio. After the removal of the edge, betweenness is re-calculated for all the edges. The removal of an edge might re-route some of the shortest paths and thus affect the betweenness of other edges. This process continues till there are no edges left in the network.

B. Influence Maximization

Given a fixed budget, finding a set of influencers such that they will extend their influence to the maximum number of the remaininf vertices in the network, was first studied as part of viral marketing. Later, Kempe in [5] stated that the influence maximization problem is NP-hard. They describe an approximation greedy algorithm for finding the most influential vertices in social networks. This greedy algorithm based on submodularity functions that has an $(1 - \frac{1}{e})$-approximation ratio. It also describes two mainly used information diffusion models, the Linear Threshold and the Independent Cascade models.

CELF proposed in [16] is another popular approximation algorithm based on submodularity and the lazy computation property. The algorithm considers unit as well as non-constant costs of vertices in a network. The algorithm explores lazy evaluation of marginal increments to reduce the computations and attain better efficiency. An enhancement to this algorithm is CELF++ algorithm [17]. This algorithm uses the principle that if the previous best vertex is selected as an influencer in the current iteration, then we do not recompute the marginal gain. SIMPATH presented in [18] is another efficient algorithm for maximizing the influence spread under the Linear Threshold model. Vertex Cover optimization and Look Ahead optimization are the two techniques used by this algorithm to reduce the number of spread estimation calls.

In addition to graph theory based techniques, machine learning and data mining based algorithms can be used for influence detection and information spread estimation.

C. Smart cities

In a similar line of research [19], they consider the problem of finding microblogger influencers, which they call key microbloggers. In their experiments they collect data from Sina Weibo, and evaluate the ranking accuracy of their proposed model. To identify key microbloggers they consider specific domains which can also change. The ranking is based on three primary scores of influence and three secondary scores of activity assigned to each microblogger. In our model we use topic specified edges and then use the structure of the constructed graph to determine communities and the influencers within each community.

Paper [20] proposes a framework on how cities can prevent crime occurrences by listening on tweets of Twitter. Additionally, applies the framework to a case study by performing real-time acquisition of crime detection information from social media messages.

Another example of how predictive models, which use social data in the public sector, can assist cities is described in [21]. In that article a pilot project is described that supplements San Francisco's food-safety analytics with a predictive tool derived from Yelp reviews. By isolating Yelp posts with keywords related to foodborne illness, they created a model that identified health code violations in restaurants with high accuracy.

The paper [22] proposes an approach to deploy the small cell, which is a 5G internet technology, in an energy efficient way in order to support wireless access for numerous devices anywhere and anytime in smart cities. This deployment framework takes into account the dynamic traffic patterns. The simulation results obtained suggest that this achieves better energy efficiency with the same quality of service.

V. Conclusion

In this paper, we propose an approach to utilize social networks to increase public participation and build smart cities optimally. In our two phase framework we form communities based on the common issues faced, opinion of people and detect influencers in each community formed. We carry out experiments with the synthetic and real world data sets. As a part of future work, we plan to improvise our community and influencers detection algorithms to attain better performance. We also plan to maximize the influence using different influence diffusion techniques. In addition, experiments with more real world social network data sets will be carried out followed by analysis of results obtained.

References

[1] H. Chourabi, T. Nam, S. Walker, J. R. Gil-García, S. Mellouli, K. Nahon, T. A. Pardo, and H. J. Scholl, "Understanding smart cities: An integrative framework," in *45th Hawaii International International Conference on Systems Science (HICSS-45 2012), Proceedings, 4-7 January 2012, Grand Wailea, Maui, HI, USA*, 2012, pp. 2289–2297. [Online]. Available: http://dx.doi.org/10.1109/HICSS.2012.615

[2] A. Meola, "How smart cities & IoT will change our communities," http://www.businessinsider.com/internet-of-things-smart-cities-2016-10, published: 2016-12-20.

[3] M. Girvan and M. E. Newman, "Community structure in social and biological networks," *Proceedings of the National Academy of Sciences*, vol. 99, no. 12, pp. 7821–7826, 2002. [Online]. Available: http://www.pnas.org/content/99/12/7821.full

[4] V. D. Blondel, J.-L. Guillaume, R. Lambiotte, and E. Lefebvre, "Fast unfolding of communities in large networks," *Journal of Statistical Mechanics: Theory and Experiment*, vol. 2008, no. 10, p. 10008, 2008. [Online]. Available: http://iopscience.iop.org/article/10.1088/1742-5468/2008/10/P10008/meta

[5] D. Kempe, J. M. Kleinberg, and É. Tardos, "Maximizing the spread of influence through a social network," *Theory of Computing*, vol. 11, pp. 105–147, 2015. [Online]. Available: http://dx.doi.org/10.4086/toc.2015.v011a004

[6] R. Zafarani, M. A. Abbasi, and H. Liu, *Social media mining: an introduction*. Cambridge University Press, 2014.

[7] I. Gaye, G. Mendy, S. Ouya, and D. Seck, "Spanning graph for maximizing the influence spread in social networks," in *Proceedings of the 2015 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining, ASONAM 2015, Paris, France, August 25 - 28, 2015*, 2015, pp. 1389–1394. [Online]. Available: http://doi.acm.org/10.1145/2808797.2809309

[8] "Python 3.5.3." [Online]. Available: https://www.python.org/

[9] "Python-igraph." [Online]. Available: http://igraph.org/python/

[10] P. Erdos and A. Rényi, "On the evolution of random graphs," *Publ. Math. Inst. Hung. Acad. Sci*, vol. 5, no. 1, pp. 17–60, 1960. [Online]. Available: http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.348.530&rep=rep1&type=pdf

[11] "Zachary's karate club." [Online]. Available: http://nexus.igraph.org/api/dataset_info?id=1&format=html

[12] "Snap dataset (facebook)." [Online]. Available: http://snap.stanford.edu/data/egonets-Facebook.html

[13] S. Fortunato, "Community detection in graphs," *Physics Reports*, vol. 486, no. 3, pp. 75–174, 2010. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0370157309002841

[14] B. W. Kernighan and S. Lin, "An efficient heuristic procedure for partitioning graphs," *Bell System Technical Journal*, vol. 49, no. 2, pp. 291–307, 1970.

[15] A. Clauset, M. E. Newman, and C. Moore, "Finding community structure in very large networks," *Physical Review E*, vol. 70, no. 6, p. 066111, 2004. [Online]. Available: http://journals.aps.org/pre/abstract/10.1103/PhysRevE.70.066111

[16] J. Leskovec, A. Krause, C. Guestrin, C. Faloutsos, J. M. VanBriesen, and N. S. Glance, "Cost-effective outbreak detection in networks," in *Proceedings of the 13th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, San Jose, California, USA, August 12-15, 2007*, 2007, pp. 420–429. [Online]. Available: http://doi.acm.org/10.1145/1281192.1281239

[17] A. Goyal, W. Lu, and L. V. Lakshmanan, "Celf++: optimizing the greedy algorithm for influence maximization in social networks," in *Proceedings of the 20th international conference companion on World wide web*. ACM, 2011, pp. 47–48.

[18] A. Goyal, W. Lu, and L. V. S. Lakshmanan, "SIMPATH: an efficient algorithm for influence maximization under the linear threshold model," in *11th IEEE International Conference on Data Mining, ICDM 2011, Vancouver, BC, Canada, December 11-14, 2011*, 2011, pp. 211–220. [Online]. Available: http://dx.doi.org/10.1109/ICDM.2011.132

[19] L. Huang, W. Wang, X. Chen, and J. Chen, "A heuristic method of identifying key microbloggers," in *Inclusive Smart Cities and Digital Health - 14th International Conference on Smart Homes and Health Telematics, ICOST 2016, Wuhan, China, May 25-27, 2016. Proceedings*, 2016, pp. 417–426. [Online]. Available: http://dx.doi.org/10.1007/978-3-319-39601-9_37

[20] A. Souza, M. Figueredo, N. Cacho, D. Araújo, J. Coelho, and C. A. Prolo, "Social smart city: A platform to analyze social streams in smart city initiatives," in *IEEE International Smart Cities Conference, ISC2 2016, Trento, Italy, September 12-15, 2016*, 2016, pp. 1–6. [Online]. Available: http://dx.doi.org/10.1109/ISC2.2016.7580848

[21] J. P. Schomberg, O. L. Haimson, G. R. Hayes, and H. Anton-Culver, "Supplementing public health inspection via social media," *PloS one*, vol. 11, no. 3, p. e0152117, 2016.

[22] L. Zhou, Z. Sheng, L. Wei, X. Hu, H. Zhao, J. Wei, and V. C. M. Leung, "Green cell planning and deployment for small cell networks in smart cities," *Ad Hoc Networks*, vol. 43, pp. 30–42, 2016. [Online]. Available: http://dx.doi.org/10.1016/j.adhoc.2016.02.008

[23] S. Goldsmith, "Social media's place in data-smart governance," published: 2016-05-19. [Online]. Available: http://datasmart.ash.harvard.edu/news/article/social-medias-place-in-data-smart-governance-837

[24] 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*. ACM, 2009, pp. 199–208.

[25] M. Eirinaki, N. Moniz, and K. Potika, "Threshold-bounded influence dominating sets for recommendations in social networks," in *2016 IEEE International Conferences on Big Data and Cloud Computing (BDCloud), Social Computing and Networking (SocialCom), Sustainable Computing and Communications (SustainCom), BDCloud-SocialCom-SustainCom*, 2016, pp. 408–415. [Online]. Available: http://dx.doi.org/10.1109/BDCloud-SocialCom-SustainCom.2016.67

[26] D. Easley and J. Kleinberg, *Networks, crowds, and markets: Reasoning about a highly connected world*. Cambridge University Press, 2010.

[27] N. Meghanathan, "A greedy algorithm for neighborhood overlap-based community detection," *Algorithms*, vol. 9, no. 1, 2016.

[28] M. E. Newman, "Fast algorithm for detecting community structure in networks," *Physical Review E*, vol. 69, no. 6, p. 066133, 2004. [Online]. Available: https://arxiv.org/pdf/cond-mat/0309508.pdf%22

[29] P. De Meo, E. Ferrara, G. Fiumara, and A. Provetti, "Generalized louvain method for community detection in large networks," in *Intelligent Systems Design and Applications (ISDA), 2011 11th International Conference on*. IEEE, 2011, pp. 88–93. [Online]. Available: https://arxiv.org/pdf/1108.1502.pdf

[30] M. E. Newman, "Modularity and community structure in networks," *Proceedings of the National Academy of Sciences*, vol. 103, no. 23, pp. 8577–8582, 2006. [Online]. Available: http://www.pnas.org/content/103/23/8577.short

[31] L. Tang and H. Liu, *Community Detection and Mining in Social Media*. Morgan & Claypool Publishers, 2010, vol. 3.

[32] M. E. Newman and M. Girvan, "Finding and evaluating community structure in networks," *Physical Review E*, vol. 69, no. 2, p. 026113, 2004. [Online]. Available: https://arxiv.org/pdf/cond-mat/0308217.pdf

[33] J. MacQueen *et al.*, "Some methods for classification and analysis of multivariate observations," in *Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability*, vol. 1, no. 14. Oakland, CA, USA., 1967, pp. 281–297. [Online]. Available: http://www-m9.ma.tum.de/foswiki/pub/WS2010/CombOptSem/kMeans.pdf