# Identifying most prominent nodes in social network for Information Diffusion by Weak Nodes approach*

Deepak Khirey†
*Indiana University, Bloomington*
(Dated: May 1, 2019)

In Social network analysis , it is most frequent question to identify prominent nodes from the network. Prominent node has various definitions based on its applications. And there are different centrality measures to rank and identify prominent nodes such as Degree centrality to identify most connected node, closeness centrality to identify closeness with hub for viral epidemic studies.

Similarly there is betweenness centrality measure to identify prominent nodes for Information diffusion. Betweenness centrality is measured by the amount of shortest paths passing through the given node. If the node is present on many shortest paths then it is most likely to play vital role in passing on the information through it.

We also know that weak nodes play important role in information diffusion. As Granovetter study [1] points out, information diffusion through weak nodes is much more effective than through strong nodes. Most of the times weak nodes are the ones with higher betweenness centrality value. But sometimes, in scale free network hubs have high betweenness centrality value due to ultra small world phenomenon. [2] In those cases, weak nodes and high value betweenness centrality nodes will be different.

In this study, I am going to evaluate method to detect community structure in underlying network and then identifying weak nodes from it as top nodes playing important role for information diffusion. I am also going to compare this method with betweenness centrality measure to establish statistical significance. Goal of this study is to assess weak nodes method based on community structure vis-a-vis betweenness centrality method to identify top nodes contributing to information diffusion.

## I. INTRODUCTION

Online Marketing on Social networks is spreading very fast with the rise of social platforms like Facebook, Instagram, LinkedIn etc. It provides a very effective platform of product promotion through users, groups and brand specific pages.

Consider a use case where a new webseries is launched on online entertainment site like Netflix or Youtube and an online social marketing campaign is started on Facebook to promote that webseries. Now, we have to place an advertisement on Facebook Pages and obviously we cannot place our advertisement on millions of available pages. So we have to figure out top pages which will reach to maximum number of subscribers and achieve maximum spread of word. Suppose, we know the Facebook Page network for all TVShows on Facebook which is our input for this campaign. There are lot of such outreach campaigns in online social marketing space and they have significant commercial value. Clearly in such scenarios Information Diffusion plays an important role.

Hence finding top nodes from information diffusion aspect in social network is very critical and thats motivation for current study. I am evaluating weak nodes method to in context of

**Research Question-** How to get maximum reach on Facebook network by targeting specific pages ?

**Statistical Hypothesis**

Betweenness centrality is most popular method to identify nodes contributing to Information Diffusion. However weak nodes in community structure also contribute to Information spread. I am going to consider this centrality as benchmark to compare performance of weak nodes method. I will evaluate whether both approaches are same in terms of identifying top nodes, whether both methods indicate to same set of nodes in the network for maximum information diffusion and what are factors affecting performance of weak nodes method in terms of reach. To carry out Statistical Significance test, Let me define my hypothesis as below-

Null Hypothesis H0 : K2weaknodes = K2betweenness
There is no difference in reach if top nodes are selected by betweenness method or weak node method.
Alternate Hypothesis H1 : K2weaknodes  K2betweenness
There is difference in reach if top nodes are selected by betweenness method or weak node method.

## II. RELATED WORK

Rise of Online Social Networks are relatively new in the area of Network Science. These networks are huge datasets which may be leveraged for large scale network analysis and study of social behavior. This helps to overcome some of the statistical limitations of earlier studies as we have more accurate and most updated data

* keyworkds: network science, weak nodes, facebook marketing, information diffusion
† Network Science, Spring 2019; dkhirey@iu.edu

which helps us to generalize patterns observed in sample dataset. Although there was limited data and restrictive means to carry out experiment in early years, there has been extensive research in Social Network Analysis with empirical methods.

In his early study of 'The Strength of Weak Ties (1973)', [1] Granovetter studied Friendship Network, to establish role of weak ties in information diffusion. His reasearch was one of the pioneering work in network science. He suggested that weak links work as bridge between two friendship communities rather than strong ties. He mentions "Intuitively speaking, this means that whatever is to be diffused can reach a larger number of people, and traverse greater social distance (i.e., path length), when passed through weak ties rather than strong". [1] He refers to previous studies like Stanley Milgram Parcel study, Michigan Junior High school study and Hysterical Contagion in Southern Textile Plant to support his argument. He also gives ineteresting example of American Blue collar workers and their job serach. He observes that there is more possibility to get a job through a remote contact rather than immediate contact.

Ferrara et el (2012) [3] in their 'On Facebook Most Ties are weak',take this notion one step ahead and try to apply Granovetter's weak link approach to Online Social Network such as Facebook in modern days. This study shows that empirical study about weak links is equally applicable to most sophisticated modern social networks even today. Granovetter's study was on friendship network, but in large Online Social netwroks, Ferrara sugegsts that "weak ties as those edges that, after dividing up the network into communities (thus obtaining the so-called community structure), connect vertices belonging to different communities". [3] In current study, I am going to use this definition of weak links to identify weak nodes associated with it.

Further, Nouh and Nurse [4] in their 'Identifying Key-Players in Online Activist Groups on the Facebook Social Network' provide more insights on ways to analyse social network to detect most influential people in the communities. With example of a UK based Facebook Online Activist group, they use Degree centrality to target nodes which are at the center of destructive activities and then betweenness centrality to prevent further spread of negative sentiments. This study tries to combine findings based on multiple centrality measures such as Degree Centrality, Betweenness Centrality and Eigenvector Centrality. It argues that these centralities are overlapping in real world networks and most active members can not be considered as most influential members. This is very important reasoning because different centrality measures indicate different use cases of information flow within network.

Facebook Marketing is a most significant use case when it comes to study information flow in Online Social Network because it has commercial value to reach to maximum number of users at lowest possible cost. Maurer(2011) [5] in study 'Effectiveness of Advertising on Social Network Sites: A Case Study on Facebook', evaluated various avenues on Facebook such as Groups, Events, Fanpages as marketing tool and found that an advertisement on Facebook Page is most likely to influence customers' purchasing decision. As there are restrictions on placing advertisement on friend's timeline or group network due to privacy settings, its not very efficient way for marketing. Considering this fact, it becomes necessary to identify most prominent pages on Facebook network which can amplify marketing campaign and thats main motivation behind my current study.

In 'GEMSEC: Graph Embedding with Self Clustering' study, Rozemberczki et el(2018) [6] have collected Facebook Page Network dataset from various areas of interest to propose new clustering and graph embedding algorithm. I think, this is an ideal dataset to study weak nodes for Facebook Page network. Most of the studies on effects of weak nodes have been in the context of egocentric networks. So if we want to apply theory of information diffusion through weak nodes to Facebook as marketing tool through Page advertisement, then we need large Facebook page network as diverse as possible and which is not ego-centric network. This is satisfied by GEMSEC dataset, hence I have chosen this for my study.

## III. METHOD

I am using eight step method [FIG. 1] for repetitive samples of network to establish statistical test results in this study. The code is written in python using networkx package.[7] Network Visualization is done in Gephi.[8] Below are details of method of study-

**Data Source**

This study aims to identify top pages for information diffusion for Facebook Marketing. So I have selected a network dataset which is Facebook Page Network from GEMSEC research. It is publicly available on snap.standford.edu. [9] This dataset was used for developing Clustering algorithm by original researchers but I am going to use this dataset only to detect community structure and identify top nodes. I am not using GEMSEC clustering algorithm for this study purpose.

This dataset consists of multiple networks such as tvshows, politician, company, publicfigures, government, atheletes, newsites, artists. Each of the network consist of information about Facebook Pages in respected area of interest. So Facebook pages are Nodes of these networks.

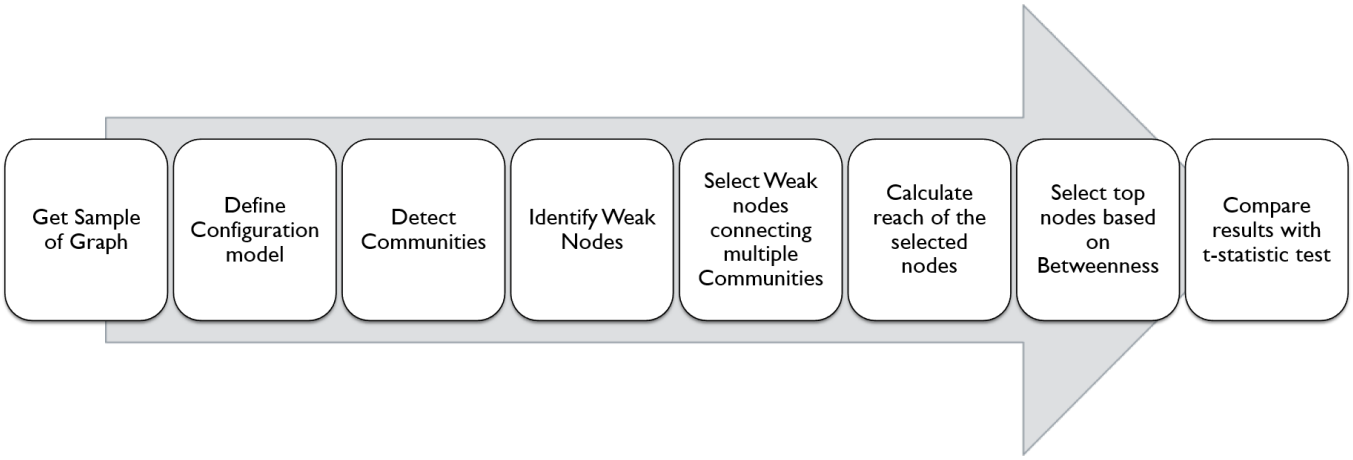


FIG. 1. Method of study
This method is executed for each sample size repetitively to perform statistical test for comparison of Weak nodes method and Betweenness methods on TVshows network

[FIG.2] If there are any mutual subscribers between two pages then that is considered as link between those two pages. If there are no mutual subscribers then there is no link between those two pages. All networks are anatomized meaning real names of pages are replaced by random numbers. All networks are undirected and unweighted. They are available in CSV format as list of source and target nodes.

TVShows network is smallest network in GEMSEC dataset.[TABLE. I] Hence I have chosen to experiment on this network due to computational limitation. Although I have used TVShows network extensively in current study, my method can be applied to any social network of similar nature.

As most of the social network are, this network also shows scale free properties. It is evident from CCDF plot.[FIG. 3] We can see power law traits as there are large number of nodes with smaller degree and very few hubs with large degree values.

### Data Sampling

Due to computational limitations, it is recommended to take sample from network and perform multiple statistical trials on it. [10] However, It should be ensured that network characteristics are preserved while taking sample from the graph. The degree sequence is not distorted drastically while scaling down the graph. I have used Snowball Sampling [11] as preferred sampling technique. It explores the graph by starting with random k nodes and then explores the graph by random walk method for those nodes. It adds k neighbors for selected nodes and keeps repeating this step for every selected neighbor until sample size is reached. This ensures that we collect nodes from significant clusters of the graph and also it tries to reduce bias I selected initial nodes k=5 randomly. Also I am using multiple sample sizes and repeated tests to make sure that sampling does not affect the outcome of the study due to bias.

TABLE I. TVShows Network Characteristics

| | |
|---|---|
| Nodes | Facebook Pages |
| Links | Mutual subscribers |
| Type | Undirected, unweighted |
| Number of Nodes | 3892 |
| Number of Links | 17239 |
| Average Degree | 8.859 |
| Max Degree | 126 |
| Average Clustering coefficient | 0.443 |
| Average Path length | 6.276 |
| Network Diameter | 20 |
| Connected Components | 1 |

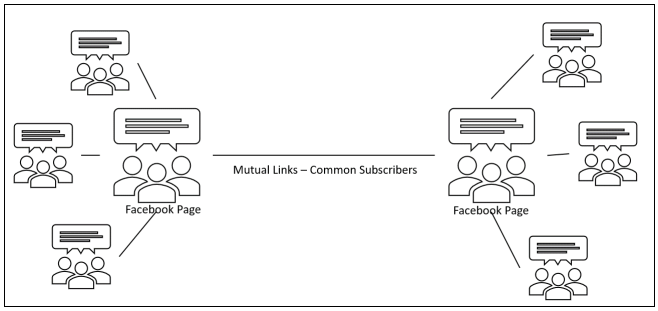


FIG. 2. Nodes and Links in TVShows network.
Nodes are Facebook pages and Links are Mutual subscribers between pages.

### Configuration Model

Next steps is to generate null model from the degree sequence of sample graph. Since input graph shows scale free properties, I have not used Erdos-Renyi Random graph method. Instead, I have used configuration model

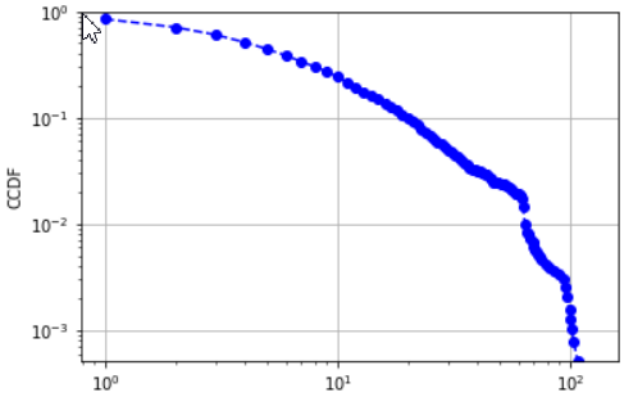


FIG. 3. CCDF plot for TVShows network on log-log scale. This plot shows power law distribution indicating scale-free network property.

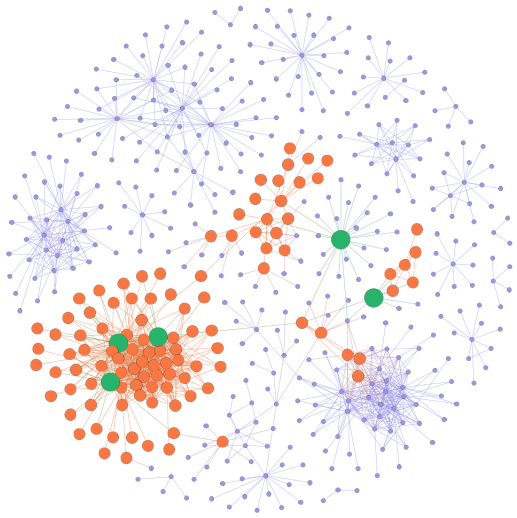

FIG. 4. Top nodes and their reach using Weak Nodes method. This is Fructerman Reingold [8] layout visualization for 500 nodes sample from TVShows network and top 5 nodes selected. The coverage is 172 nodes i.e. 34.4 percent of sample network.

method using networkx API. [7] This method uses degree sequence as input and then randomly rewires stubs to generate a simulated network which is different than input network but preserves network characteristics. Using such simulated model for repeated statistic tests ensures that experiment is not biased and it can be generalized.

**Detect Community Structure**

Once I simulate configuration model, then most crucial step of this method is to detect underlying Community structure of the network. I have used Modularity approach to arrive at optimum community structure. It is difference between actual and expected number of links for each node of the parameter. If actual links are more, then the node is considered part of same cluster. It is observed that community detection depends on Resolution parameter, hence I used multiple trials on original network to arrive at maximum modularity value. Then I used that resolution value for all tests of that network for different sample sizes. This is manual effort involved to ensure that community structure is optimum for network. There is opportunity to automate optimization of community detection by deploying machine learning methods.

**Identify Top Weak Nodes**

As per Farrera et al. [3] , weak links are the ones which join nodes from different communities. Hence using this definition I identified all links which connect nodes from different modular class. There are number of nodes found this way. So next question was to identify top weak nodes from this list. Intuitively, the nodes which are connecting multiple communities are most important for information diffusion. So I counted how many times each weak node appears in the list of weak links and then ranked all weak nodes based on this count in descending order. Required number of top weak nodes are picked from this list during statistical test.

**Identify Top Betweenness Nodes**

To compare weak nodes method I have chosen Betweenness Centrality measure as standard parameter. Hence I calculated Betweenness Centrality for input network and assigned it as parameter to the node using networkx API [7]. Then sorted list of nodes based on betweenness value in descending order. Required number of top nodes are selected from this list for statistical test.

**Count Unique Nodes Reached**

Once top nodes are identified by both methods, it is required to compare the information diffusion achieved by these nodes in the network. We have to remember here that this is not an egocentric friendship network so it does not directly give number of people reached if we select a node. Rather, a node is Facebook page here. So when we select a page we are actually referring to a group of subscribers associated with that page. Considering this fact, intuitively we can say that if we display advertisement content on one Facebook page then fraction of its subscribers will visit it. These subscribers are, in turn, mutual subscribers of other Facebook page in the link. So, the information diffusion achieved by placing advertisement on one page reaches to neighboring page and neighbours of neighbouring page through mutual subscribers. This way, the information diffusion parameter to consider for this type of network is the "Reach" which is unique nodes in network from neighbors (geodesic distance d=1) and neighbors of neighbors (geodesic distance d=2) of selected nodes. For each experiment in statistical test, reach is counted for

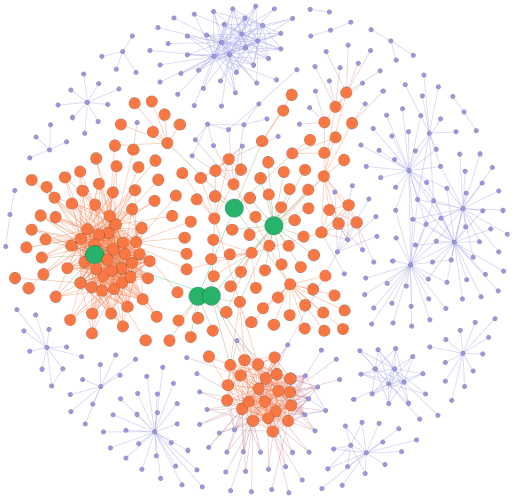

FIG. 5. Top nodes and their reach using Betweenness method. This is Fructerman Reingold [8] layout visualization for 500 nodes sample from TVShows network and top 5 nodes selected. The coverage is 239 nodes i.e. 47.8 percent of sample network.

both methods and used for comparing results.

**Statistical Test**

To establish statistical proof for hypothesis, the experiment of selecting top nodes by weak nodes method and by betweenness method is repeated 100 times on different simulated configuration models of size ranging from 100 to 1000. So thats totally 100 * 10 = 1,000 tests for selecting top 5 nodes. Similarly it is repeated for multiple top node selection like top 10, 15, 20 nodes. Hence for TVShows network, there were 1000 * 4 = 4000 tests carried out and the average count of reach i.e. unique nodes up-to geodesic distance =2 were compared using 2-sided t-test. This test compares given 2 input arrays and results describe whether the input arrays are same or different based on p-value. I am testing for confidence interval of 95%, hence if p-value is smaller than (1 - 0.95)/2 = 0.0025 then we can reject the null hypothesis and say that two methods are different methods. The t-stats score tells us which method is better of the two. If t-stats is positive then first method (weak nodes method) is better and if t-stats is negative then second method (betweenness method ) is better.

## IV. RESULTS

**Different sample sizes**

When I selected Top 5 nodes from TVShows network for different sample size ranging from 100 to 1000, it is observed that p-value is very much smaller than 0.0025, [TABLE. II] which shows that weak nodes method and

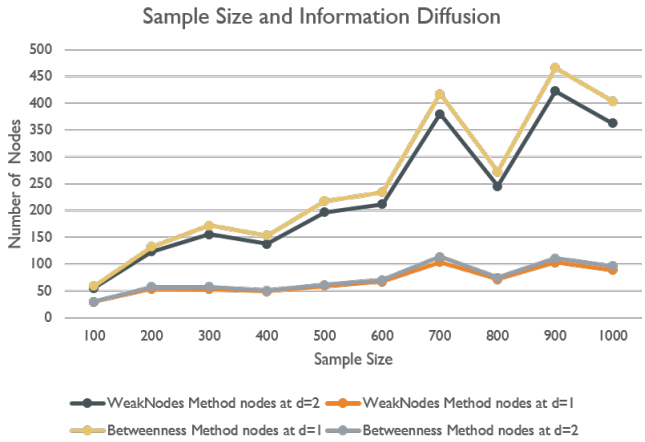


FIG. 6. Plot of Reach for Different Sample size for Top 5 nodes
This plot clearly shows that for larger sample size, the reach of both methods diverges.

betweenness method are different and they produce different set of top nodes and have different reach within the network. This trend is consistent for selecting top 5 node for all sample sizes. [TABLE. III] Here Betweenness method performs better than weak nodes method. As we can see from the graph, there is small gap for number of nodes reached at smaller sample size, but this gap diverges as the sample size grow. [FIG. 6] This indicates that for larger network, betweenness method performs much better than weak nodes method.

**Different top nodes selected**

Then I repeated tests for different top node selection for same sample size range. It is observed that the reach for weak nodes method and for betweenness method converges in this case. [FIG. 7] We can see that for selection of Top 5 nodes, betweenness method is significantly better but as we select Top 15 nodes

TABLE II. Statistical Test Results
Test conducted on multiple sample sizes of TVShows network and top 5 nodes selected for both methods. Results clearly show that Betweenness method is performing better than weak nodes method in terms of network coverage.

| Sample Size | t-stats | p-value | Better Method |
|---|---|---|---|
| 100 | -5.43 | 0.000000408 | Betweenness |
| 200 | -5.14 | 0.000001377 | Betweenness |
| 300 | -4.13 | 0.000075218 | Betweenness |
| 400 | -6.1 | 0.000000020 | Betweenness |
| 500 | -4.32 | 0.000037719 | Betweenness |
| 600 | -8.21 | 0.000000000 | Betweenness |
| 700 | -6.96 | 0.000000000 | Betweenness |
| 800 | -13.88 | 0.000000000 | Betweenness |
| 900 | -10.69 | 0.000000000 | Betweenness |
| 1000 | -10.08 | 0.000000000 | Betweenness |

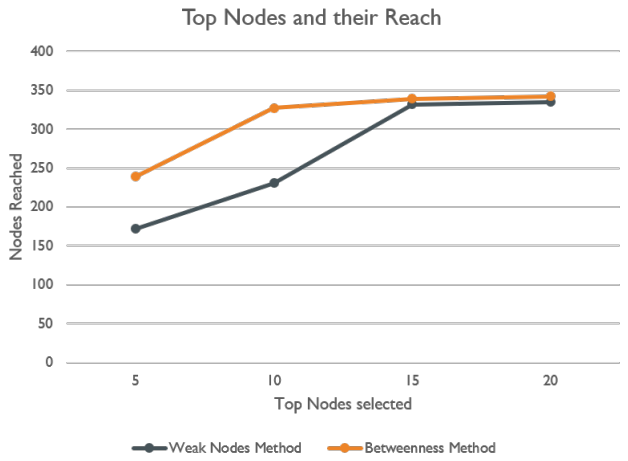


FIG. 7. Plot for selecting more Top Nodes
It shows that as we select more top nodes the reach converges for both methods.

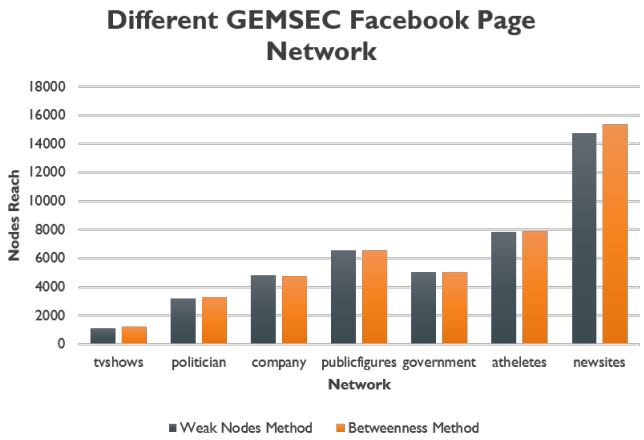


FIG. 8. Plot for Different Networks
This plot shows that weak nodes method and betweenness method are comparable for different networks irrespective of network size.

then the reach for both methods is almost the same. [TABLE. IV] This indicates that both methods are equally effective if we choose greater number of seeds for information diffusion.

### Different networks

I carried out statistical tests on TVShows network as it is smallest of the GEMSEC dataset and it was computationally feasible for me to use this test for repetitive simulation. But to validate my observations, I applied same method to all other networks in the GEMSEC dataset and compared outputs. This also ensured that the results are not biased due to sampling strategy used on TVShows network because I did not use any sampling or simulation method on other networks. Rather I subjected entire networks to the same set of logical steps to arrive at the results. Here we can see that for all networks, the weak nodes method and betweenness method have produced comparable outputs. [FIG. 8] There is marginal difference in the number of nodes reached. In fact, for company, publicfigures and government networks, weak nodes methods outperformed betweenness method. [TABLE. V]

TABLE III. Reach for Different Sample size for Top 5 nodes
This table shows count of neighbor nodes (geodesic distance d=1) and neighbors of neighbors nodes (geodesic distance d=2)

| | Weak Nodes Method | | Betweenness Method | |
|---|---|---|---|---|
| Sample Size | Nodes d=2 | Nodes d=1 | Nodes d=2 | Nodes d=1 |
| 100 | 55 | 30 | 59 | 30 |
| 200 | 123 | 54 | 132 | 57 |
| 300 | 156 | 53 | 172 | 57 |
| 400 | 137 | 49 | 153 | 50 |
| 500 | 197 | 59 | 218 | 61 |
| 600 | 212 | 67 | 234 | 70 |
| 700 | 380 | 104 | 418 | 114 |
| 800 | 245 | 72 | 272 | 74 |
| 900 | 422 | 103 | 466 | 110 |
| 1000 | 363 | 89 | 404 | 96 |

## V. DISCUSSION

As we see from the results, weak nodes method and betweenness method provide comparable reach for selected top nodes. And we also know that Betweenness centrality measure is quite proven to identify top nodes for Information diffusion. [12] In that case what is value of Weak nodes method over betweenness method? In my view, the key value add is knowledge underlying of Community structure. Betweenness method considers network as whole and calculates centrality values based on shortest path approach. However, in weak nodes method, when we identify communities, we consider network as collection of small clusters which are inter linked. For the use case of Facebook marketing, this brings immense advantage to the campaign as it will be input for which user categories to target. For TVShows network, let's say if we are promoting Reality show

TABLE IV. Selecting More Top Nodes
This result is for sample size of 500 nodes from TVshows network.

| Top Nodes | Weak Nodes Method | Betweenness Method |
|---|---|---|
| 5 | 172 | 239 |
| 10 | 231 | 327 |
| 15 | 332 | 339 |
| 20 | 335 | 342 |

then it is more logical to woo viewers of Adventure shows rather than Fictional shows. This distinction helps to select weak nodes which are joining Reality shows community and Adventure shows community and promote shows accordingly. This level of details cannot be identified by Betweenness method which considers all nodes similarly. Although I have given example of TVShows, similar logic can be applied to any canvasing exercise based on social networks.

We can see that for all sample sizes in TVShows network, betweenness centrality method is performing better than weak nodes method. Betweenness centrality is a very consistent method and it is already an optimized approach with out-of-the-box API from networkx package [8]. On the other hand Weak nodes method is totally dependent on quality of community detection method. If the underlying communities are effectively detected then top weak nodes selected can have grater reach than betweenness method. In current study, community detection is based on modularity method which depends heavily on Resolution parameter. Resolution value is different for different methods to achieve maximum modularity. This is one of the weakness of community based method. Another observation is that during simulation based on graph sample, the results do not repeat consistently. This can be attributed to Snowball sampling method. This method starts with random k nodes and is dependent on initial choice. Since for each simulation, seed nodes are different, we get different configuration model for testing. Although I have tried to address this issue by increasing number of tests and sampling size, there is still some room to evaluate more effective sampling technique. [3]

Above issues can be addressed by using alternate techniques which will make weak nodes method more effective and reliable. Sampling strategies line MHRW are known to overcome sampling bias [10] and traversing through network to provide an even sample but that might distort degree distribution of the sample which is key characteristic of scale-free network. Hence to avoid bias during sampling, respondent driven sampling may be evaluated.

TABLE V. Results with Different Networks
Applying study to different networks in GEMSEC dataset. This result shows that weak node method performs better than betweenness method for company, publicfigures and goverment networks.

| Network | Weak Nodes Method | Betweenness Method |
|---|---|---|
| tvshows | 1094 | 1212 |
| politician | 3197 | 3277 |
| company | 4830 | 4773 |
| publicfigures | 6583 | 6583 |
| government | 5040 | 5039 |
| atheletes | 7871 | 7896 |
| newsites | 14735 | 15374 |

Community detection can be achieved better by deploying alternate algorithms like Infomap, Louvain Method or Link communities. It would be more reliable if same network is subjected to multiple community detection algorithms and then weak nodes are identified which are common in all methods. This way we can reduce the dependency on Modularity method and repeatability issues therein.

Also, for optimizing the community detection method, Machine Learning approach can be deployed which can identify maximum modularity for given network. This will overcome issue of plateau in modularity optimization. Similarly machine learning methods can be provided data from multiple networks under study and it can identify parameters affecting optimized community detection.

## VI. CONCLUSION

We can observe from current study that detecting community structure in the network and then selecting top weak nodes is very effective method to identify most prominent nodes in the network for information diffusion. This method provides very much comparable results to betweenness method where we can select top node by highest betweenness centrality measure. Weak nodes method may provide totally different set of prominent nodes which can achieve same amount of reach within the network. On top of this, since we already know the community structure underlying the network, weak nodes method caters to additional advantage of choosing specific community for targeted information flow. This can be of immense commercial value in the field of Facebook as marketing tool as it provides capability to focus or avoid specific user population based on campaign requirements.

However, effectiveness of weak nodes method entirely depends on effectiveness of community detection. If communities are detected in most optimized way, then top weak nodes selection can provide best possible reach in the network. This area needs to be explored further by using optimization principles to community detection.

Also, current study uses Facebook page network, but similar study may be carried out in other Online Social Network platforms such as Twitter, Instagram, LinkedIn etc. This will help to understand whether similar results can be obtained on different types of network and generalize method to detect prominent nodes for information diffusion.

## ACKNOWLEDGMENTS

I would like to thank Prof. Yong-Yeol Ahn (Indiana University, Bloomington) for providing his valuable Suggestions and Guidance for this project and throughout the course of Network Science.



[1] M. S. Granovetter, The strength of weak ties, in *American Journal of Sociology*, No. 6, Vol. Vol. 78 (1973) pp. 1360–1380.

[2] . M. M. T. Wohlgemuth, J., Small-world properties of facebook group networks, Complex Systems **23**, 197 (2014).

[3] P. . F. G. . P. A. Ferrara, Emilio & De Meo, The role of strong and weak ties in facebook: a community structure perspective, in *Communications of the ACM*, 10.1145/2629438, Vol. 57 (2012).

[4] . N. J. R. Nouh, M., Identifying key-players in online activist groups on the facebook social network, in *IEEE International Conference on Data Mining Workshop (ICDMW)* (IEEE, Atlantic City, NJ, USA, 2015).

[5] W. R. Maurer C., Effectiveness of advertising on social network sites: A case study on facebook, in *Information and Communication Technologies in Tourism 2011*, edited by R. F. Law R., Fuchs M. (Springer, Vienna, 2011).

[6] R. . S. R. . S. C. Rozemberczki, Benedek & Davies, Gemsec: Graph embedding with self clustering (2018).

[7] D. A. S. Aric A. Hagbergand P. J. Swart, Exploring network structure, dynamics, and function using networkx, in *7th Python in Science Conference (SciPy2008)*, edited by T. V. Gel Varoquauxand J. Millman (Pasadena, CA USA, 2008) pp. 11–15.

[8] M. Bastian, S. Heymann, and M. Jacomy, Gephi: An open source software for exploring and manipulating networks (2009).

[9] J. Leskovecand A. Krevl, SNAP Datasets: Stanford large network dataset collection, `http://snap.stanford.edu/data` (2014).

[10] C. T. B. M. Gjoka, M. Kurantand A. Markopoulou, Walking in facebook: A case study of unbiased sampling of osns, in *Proceedings IEEE INFOCOM* (San Diego, CA, 2010) pp. 1–9.

[11] Ashish7129, Graph sampling package (2019).

[12] M. Barthlemya, Betweenness centrality in large complex networks, Eur. Phys. J. B **38**, 163 (2004).

[13] E. F. Pasquale De Meo, On facebook, most ties are weak, in *Communications of the ACM*, No. 11, Vol. Vol. 57 (2014) pp. 78–84.

[14] A. C. Dan Frank, Zhiheng Huang, Sampling a large network: How small can my sample be? (2012).

[15] A.-L. Barabsi, *Network Science* (Cambridge University Press, 2016).