

Similarities of Influencers across Different Social Media Platforms by Using Four Centrality Measures

Tingyu Shi

Department of Computing and Software, Faculty of Engineering, McMaster University, Hamilton, ON, L8S 4L7

tingyushi0803@163.com

Abstract. Searching for influencers among a social network is important because marketers can then use this information to conduct word-of-mouth (WOM) advertisement, which is an important marketing technique. Literature Review provides detailed information about WOM advertisement. There are many ways to search influencers and often they are network centrality measurements. This paper aims to investigate whether each centrality measurement could produce similar results across different social media platforms (eg. Facebook, Twitter, Instagram). The social network data used in this research is from Huawei Company. This research uses four centrality measurements and three set similarity methods to analysis the data. As a result, this paper draws a conclusion about the binary question "Does it provide similar results or not?". Since various companies and applications may have different standards and definitions about being similar, please also check similarity data provided in this paper.

Keywords: social network, centrality measurements, marketing strategy.

1. Introduction

With the development of science and technology, social media(network) is becoming essential to people's daily lives. According to the data from Pew Research Center, about 81% of Americans use YouTube and 69% of Americans use Facebook, which means the social media usage is very high [1]. Social media does not only impact people's lives, it also brings challenges and opportunities to all walks of life. One of them is the marketing strategy. When a salesperson wants to promote a product, he/she always wants to reach the maximum number of people by using the minimum amount of resources. As a result, it is essential to find all the influencers in a social network since they can help with the advertisement by using their influence through social networks. This marketing technique is called word-of-mouth (WOM).

Then the problem comes with searching for influencers in a social network. Various methods and models are there to help with searching for influencers, and they will be introduced in the Literature Review section. Usually, centrality measures are used to find the social media influencers.

The following four centrality measures are investigated in this paper:

- Degree Centrality
- Closeness Centrality
- Betweenness Centrality

- Eigenvector Centrality

Some researchers have compared and contrasted the above four centrality measures in many aspects. For example, they have investigated the performance, consistency and correlations among four centrality measures. However, this paper examines the similarity of influencers when a specific centrality measurement is applied to different social media platforms. The significance of this research is that the research results may save time for market analysts. Suppose we can prove that a centrality measure can produce similar influencers across different platforms. Then, marketers can just analyze one platform instead of different platforms since all platforms will contain similar influencers. This research is quantitative.

2. Literature review

2.1. Importance of social network influencers

WOM marketing means that customers can promote products to their friends through dialogue or conversation [2]. In other words, customers can advertise products to other potential customers. According to a Harvard Business Review, the true valuable customers are those people who can bring new customers [3]. Thus, WOM is an important technique for marketing. Therefore, marketers should try to find influencers on a social network. Social network influencers are the proper individuals to start WOM marketing. Since they have a massive influence on a social network, they can better advertise the products and reach a greater range of potential customers. The following is a direct quote from a paper:

“To succeed today, you need to connect with people who are at the center of the conversation... Specifically, you should make sure you are reaching the decision makers who are influential in others' decisions. Influentials are well connected, they have ties to a significantly larger number of groups than the average American.” [4].

2.2. How to find social network influencers

There are many methods to discover social influencers among a social network, the following are some of them found in a paper [5]:

1. Centrality Measures (Specific information will be introduced below in section 2.3)
2. Link Topology Ranking Measures

Some centrality measures (except eigenvector centrality measure) neglect the influence nodes. For example, if a node ($node_a$) connects to a really influential node ($node_b$), then the importance of $node_a$ should be increased as well. Link topology ranking measures can solve this kind of problem. Some popular algorithms in this category are Hyper-Induced Topic Search algorithm and PageRank algorithm for web search.

Another useful model for social influencers is diffusion model [5]. This kind of model is used for modelling the transmission of information among different nodes within a social network. For instance, they can model virus spreading in a pandemic or message spreading within a social network. This paper concentrates on the first measure of influences in a network --- Centrality Measures.

2.3. Centrality measurement

Many problems can be modelled as networks, such as human brains, city traffic, relationships among people, etc. A typical network is formed with two essential elements. The first one is nodes. A node can represent a biological neuron in human brains, a bus station in a city traffic network or an individual in a social network. The second one is edges which connect different nodes. An edge can represent the signal transmission between biological neurons, roads connecting different bus stations or relationships between different individuals in a social network. Since the network structure can vary in various ways, each node has a distinct influence on the whole network [6]. Centrality measure is a quantitative measurement of the importance of a node in a network. The following four sections give brief introductions to four types of centrality measures.

2.3.1. *Degree centrality.* Degree Centrality is measured by counting the number of nodes connected to a specific node [7]. If the graph is directed, then the concept of in-degree centrality and out-degree centrality can be defined. In-degree centrality of a node means the number of edges coming into that node. The node's out-degree centrality implies the number of edges coming out of that node.

If the graph is undirected, then the Degree Centrality of node k is the following:

$$DC(k) = \sum_{j=1}^N A_{jk} \quad (1)$$

$$N = |V| \quad (2)$$

$$A_{ab} = \begin{cases} 1 & \text{(if edge between node}_a \text{ and node}_b \text{ exists)} \\ 0 & \text{(else)} \end{cases} \quad (3)$$

If the graph is directed, then in-degree centrality is the following:

$$DC(k) = \sum_{j=1}^N A_{jk} \quad (4)$$

$$N = |V| \quad (5)$$

$$A_{ab} = \begin{cases} 1 & \text{(if edge from node}_a \text{ to node}_b \text{ exists)} \\ 0 & \text{(else)} \end{cases} \quad (6)$$

If the graph is directed, then out-degree centrality is the following:

$$DC(k) = \sum_{j=1}^N A_{kj} \quad (7)$$

$$N = |V| \quad (8)$$

$$A_{ab} = \begin{cases} 1 & \text{(if edge from node}_a \text{ to node}_b \text{ exists)} \\ 0 & \text{(else)} \end{cases} \quad (9)$$

2.3.2. *Closeness centrality.* Closeness centrality measures the sum distances from one node to the other nodes. Specifically, the distance means the shortest path from one node to other nodes. After calculation, if a node has a smaller sum distance to other nodes, then it is easier for this node to communicate with other nodes since the information does not have to travel far to reach other nodes [7]. Therefore, smaller sum distances mean higher closeness centrality.

The Closeness Centrality of node k is the following:

$$CC(k) = \frac{N}{\sum_{j=1 \text{ and } j \neq k}^N d_{kj}} \quad (10)$$

$$N = |V| \quad (11)$$

d_{kj} means the shortest distance between node_k and node_j

2.3.3. *Betweenness centrality* Betweenness centrality measures the importance of a node by checking the importance of this node as a bridge [7]. If a graph is connected, then the shortest path between any two nodes must exist. The betweenness centrality of a node is calculated by checking how many times this node acts as a bridge for communication between two other nodes. If a node is needed many times when the other two nodes want to communicate via the shortest path, then this node has higher betweenness centrality.

The Betweenness Centrality of node k is the following:

$$BC(k) = \sum_{s \neq t \neq k \text{ and } s, t, k \in V} \frac{\sigma_{st}(k)}{\sigma_{st}} \quad (12)$$

$\sigma_{st}(k)$ means the number of shortest path between node_s and node_t that passes node_k

σ_{st} means the number of shortest path between node_s and node_t

2.3.4. Eigenvector centrality. Eigenvector centrality is an improvement of degree centrality. For degree centrality, it only considers the number of neighbour nodes connected to a particular node instead of considering the importance of neighbour nodes [8]. For example, if node_i is only connected to node_j and node_j is connected to other 1000 nodes. Then from the point of degree centrality, node_i has a very low centrality. However, since node_i has a valuable neighbour node, it should be considered an important node. With eigenvector centrality, we can better evaluate the importance of a node.

The Eigenvector Centrality of node k is the following:

$$EC(k) = \frac{1}{\lambda} \sum_{j=1}^N A_{kj} * EC(j) \quad (13)$$

$$N = |V| \quad (14)$$

$$A_{ab} = \begin{cases} 1 & \text{(if edge between node}_a \text{ and node}_b \text{ exists)} \\ 0 & \text{(else)} \end{cases} \quad (15)$$

2.4. Previous research about the above four centrality measurements

2.4.1. Performance analysis of centrality measures. Kiss and Bichler measured the performance of different centrality measures based on the call data from a telecom company [5]. They used centrality measures to select initial sets of individuals (influencers) and then monitored the messages diffusion level of different centrality measures. They found out that the performances of centrality measures depend on the structure of networks.

2.4.2. Correlation analysis of centrality measures. Koschutzki and Schreiber investigated the correlation of different centrality measures based on protein-protein-interaction(PPI) network and transcriptional regulation (TR) network [9]. They are two biological networks. They found that the correlation between degree centrality and eigenvector centrality is higher compared with other centralities.

3. Methodology

3.1. Research type

This research is a quantitative research. The research problem is that for each centrality measurement, does it produce similar influencers across different social media platforms? Different social media platforms can be Facebook, Twitter and Instagram. Since this paper investigated across different social media platforms, data about user relationships from different platforms were required. The controlled variable here was the individuals in the database, which means databases representing different social media platforms should contain the same group of people. For example, if D_T and D_F are two databases from two platforms, then D_T and D_F should contain identical individuals.

3.2. Data collection

An existing dataset from Kaggle was used.

This dataset was collected by crawling social media platforms i.e. Facebook, Twitter and Instagram Huawei pages. The web crawler mainly focused on Facebook posts, Twitter tweets and Instagram posts. This dataset is used by Huawei Company to enhance their business positions. Social media analytics helps Huawei organizations understand their targeted audience. I can use this dataset to create the following three networks:

- Huawei Facebook Communication Network
- Huawei Twitter Communication Network
- Huawei Instagram Communication Network

All above three communication networks were directed. For Facebook Network, it contained 1000 nodes and 250315 edges. For Twitter Network, it contained 1000 nodes and 50153 edges. For Instagram Network, it contained 1000 nodes and 4933 edges.

3.3. Data analysis

3.3.1. Social network construction. The data is represented as three big matrices from the collected dataset. These three matrices were named as M_T , M_F and M_I . The subscripts here mean Twitter, Facebook and Instagram. Then, three social networks representing three social media platforms were constructed from M_T , M_F and M_I . These three social networks were named as G_T , G_F and G_I .

How to construct G_T :

1. Add 1000 nodes to G_T first.
2. If $M_T[i][j]=1$, this means that person_{*i*} and person_{*j*} know each other. As a result, an edge was added between node_{*i*} and node_{*j*} in G_T .

How to construct G_F :

3. Add 1000 nodes to G_F first.
4. If $M_F[i][j]=1$, this means that person_{*i*} and person_{*j*} know each other. As a result, an edge was added between node_{*i*} and node_{*j*} in G_F .

How to construct G_I :

5. Add 1000 nodes to G_I first.
6. If $M_I[i][j]=1$, this means that person_{*i*} and person_{*j*} know each other. As a result, an edge was added between node_{*i*} and node_{*j*} in G_I .

3.3.2. Calculate centrality Measurements and Find Influencers. Four types of centrality measures were all calculated for each node from each social network. The following example uses degree centrality to illustrate:

1. Calculate degree centrality for each node from three social networks.
2. For each social network, we selected top 100 nodes which have the highest degree centrality values. After that, three sets of nodes were formed, and these three sets of nodes were named as S_{DT} , S_{DF} and S_{DI} . The first letter in subscript represents degree centrality and the second letter represents different social media platforms.

Repeat the above two steps for three other centrality measures. After that, the following sets existed:

- Degree Centrality: S_{DT} , S_{DF} , S_{DI}
- Closeness Centrality: S_{CT} , S_{CF} , S_{CI}
- Betweenness Centrality: S_{BT} , S_{BF} , S_{BI}
- Eigenvector Centrality: S_{ET} , S_{EF} , S_{EI}

For the second step mentioned above, the experiment selected top 100 nodes and considered them as influencers. The experiment also selected top 200 and top 300 as influencers to do analysis.

3.3.3. Calculate similarities of influencers. For each centrality measure, similarity of influencers from different social media was calculated. For example, similarities between (S_{DT}, S_{DF}) , (S_{DT}, S_{DI}) and (S_{DF}, S_{DI}) were calculated for degree centrality. Then, the experiment repeated the above process for

other three centrality measures. The following are three ways to calculate the similarity between two sets:

- Comment elements method:

For two sets S_1 and S_2 such that $|S_1|=|S_2|$, count the number of elements e such that $(e \in S_1 \wedge e \in S_2)$. More common elements means that two sets are more similar.

- Jaccard Index: For two sets A and B, Jaccard Index is the following:

$$J(A,B)=\frac{|A\cap B|}{|A\cup B|} \quad (16)$$

- Sørensen-Dice coefficient: For two sets A and B, Sørensen-Dice coefficient is the following:

$$DSC(A,B)=\frac{2|A\cap B|}{|A|+|B|} \quad (17)$$

3.4. Methodology justification

The experiment also selected the top 200 and top 300 nodes with the highest centrality measures to do the similarity analysis. The experiment was designed in this way because various companies or organizations may have different standards for social media influencers. As a result, in this way, the experiment can provide more data for companies to evaluate.

4. Results

4.1. Degree centrality

Table 1 presents similarity data by choosing the top 100 nodes with the highest Degree centrality measures as influencers. Three pairs of comparisons were made between three social media platforms.

Table 1. Similarity data of top 100 influencers by using Degree centrality.

	# Of common elements	Jaccard Index	Sørensen-Dice coefficient
S_{DF} VS S_{DT}	3	1.52%	3%
S_{DF} VS S_{DI}	11	5.82%	11%
S_{DT} VS S_{DI}	12	6.38%	12%

Table 2 presents similarity data by choosing the top 200 nodes with the highest Degree centrality measures as influencers. Three pairs of comparisons were made between three social media platforms.

Table 2. Similarity data of top 200 influencers by using Degree centrality.

	# Of common elements	Jaccard Index	Sørensen-Dice coefficient
S_{DF} VS S_{DT}	27	7.24%	13.5%
S_{DF} VS S_{DI}	43	12.04%	21.5%
S_{DT} VS S_{DI}	42	11.73%	21%

Table 3 presents similarity data by choosing the top 300 nodes with the highest Degree centrality measures as influencers. Three pairs of comparisons were made between three social media platforms.

Table 3. Similarity data of top 300 influencers by using Degree centrality.

	# Of common elements	Jaccard Index	Sørensen-Dice coefficient
S_{DF} VS S_{DT}	82	15.83%	27.33%
S_{DF} VS S_{DI}	89	17.42%	26.97%

Table 3. (continued).

S_{DT} VS S_{DI}	87	16.95%	29%
----------------------	----	--------	-----

4.2. Betweenness centrality

Table 4 presents similarity data by choosing the top 100 nodes with the highest Betweenness centrality measures as influencers. Three pairs of comparisons were made between three social media platforms.

Table 4. Similarity data of top 100 influencers by using Betweenness centrality.

	# Of common elements	Jaccard Index	Sørensen-Dice coefficient
S_{BF} VS S_{BT}	4	2.04%	4%
S_{BF} VS S_{BI}	9	4.71%	9%
S_{BT} VS S_{BI}	10	5.26%	10%

Table 5 presents similarity data by choosing the top 200 nodes with the highest Betweenness centrality measures as influencers. Three pairs of comparisons were made between three social media platforms.

Table 5. Similarity data of top 200 influencers by using Betweenness centrality.

	# Of common elements	Jaccard Index	Sørensen-Dice coefficient
S_{BF} VS S_{BT}	28	7.53%	14%
S_{BF} VS S_{BI}	43	12.04%	21.5%
S_{BT} VS S_{BI}	45	12.68%	21.5%

Table 6 presents similarity data by choosing the top 300 nodes with the highest Betweenness centrality measures as influencers. Three pairs of comparisons were made between three social media platforms.

Table 6. Similarity data of top 300 influencers by using Betweenness centrality.

	# Of common elements	Jaccard Index	Sørensen-Dice coefficient
S_{BF} VS S_{BT}	79	15.16%	26.33%
S_{BF} VS S_{BI}	87	16.96%	29%
S_{BT} VS S_{BI}	89	17.42%	29.67%

4.3. Closeness centrality

Table 7 presents similarity data by choosing the top 100 nodes with the highest Closeness centrality measures as influencers. Three pairs of comparisons were made between three social media platforms.

Table 7. Similarity data of top 100 influencers by using Closeness centrality.

	# Of common elements	Jaccard Index	Sørensen-Dice coefficient
S_{CF} VS S_{CT}	3	1.52%	3%
S_{CF} VS S_{CI}	12	6.38%	12%
S_{CT} VS S_{CI}	13	6.95%	13%

Table 8 presents similarity data by choosing the top 200 nodes with the highest Closeness centrality measures as influencers. Three pairs of comparisons were made between three social media platforms.

Table 8. Similarity data of top 200 influencers by using Closeness centrality.

	# Of common elements	Jaccard Index	Sørensen-Dice coefficient
S_{CF} VS S_{CT}	27	7.24%	13.5%
S_{CF} VS S_{CI}	42	11.73%	21%
S_{CT} VS S_{CI}	44	12.36%	22%

Table 9 presents similarity data by choosing the top 300 nodes with the highest Closeness centrality measures as influencers. Three pairs of comparisons were made between three social media platforms.

Table 9. Similarity data of top 300 influencers by using Closeness centrality.

	# Of common elements	Jaccard Index	Sørensen-Dice coefficient
S_{CF} VS S_{CT}	82	15.83%	27.33%
S_{CF} VS S_{CI}	90	17.65%	30%
S_{CT} VS S_{CI}	85	16.50%	28.33%

4.4. Eigenvector centrality

Table 10 presents similarity data by choosing the top 100 nodes with the highest Eigenvector centrality measures as influencers. Three pairs of comparisons were made between three social media platforms.

Table 10. Similarity data of top 100 influencers by using Eigenvector centrality.

	# Of common elements	Jaccard Index	Sørensen-Dice coefficient
S_{EF} VS S_{ET}	3	1.52%	3%
S_{EF} VS S_{EI}	13	6.95%	13%
S_{ET} VS S_{EI}	11	5.82%	11%

Table 11 presents similarity data by choosing the top 200 nodes with the highest Eigenvector centrality measures as influencers. Three pairs of comparisons were made between three social media platforms.

Table 11. Similarity data of top 200 influencers by using Eigenvector centrality.

	# Of common elements	Jaccard Index	Sørensen-Dice coefficient
S_{EF} VS S_{ET}	31	8.40%	15.5%
S_{EF} VS S_{EI}	37	10.19%	18.5%
S_{ET} VS S_{EI}	44	12.36%	22%

Table 12 presents similarity data by choosing the top 300 nodes with the highest Eigenvector centrality measures as influencers. Three pairs of comparisons were made between three social media platforms.

Table 12. Similarity data of top 300 influencers by using Eigenvector centrality.

	# Of common elements	Jaccard Index	Sørensen-Dice coefficient
S_{EF} VS S_{ET}	85	16.50%	28.33%

Table 12. (continued).

S_{EF} VS S_{EI}	94	18.58%	31.33%
S_{ET} VS S_{EI}	90	17.65%	30%

5. Discussion

The results can indicate two key findings. The first finding is that for each type of centrality measure, it cannot produce similar influencers across different social media platforms. No matter how many individuals were chosen as influencers, the Jaccard indices are consistently below 20%, therefore, all centrality measures cannot produce similar influencers across different social media platforms. The second key finding is that as the number of selected influencers increases, similarity also increases. You can verify this from section 4.

Other researchers have studied performance, consistency and correlation. This research results showed some data about the influencers similarity when applying a certain centrality measure to different social media.

The experiment results showed that all four types of centrality measures can not produce similar influencers across different social media platforms. This result may be caused by two reasons. The first one is that people may have different preferences for different social media. For example, an individual may spend more time on Facebook than on Twitter. As a result, he/she presents more influence on Facebook. The second reason is that the data size may not be big enough. In this research, the dataset only has relationships among 1000 people. The results may differ if the same methodology can be applied to a more extensive dataset.

For marketers, this research indicates that it is essential to analyze different social media platforms since different platforms do not contain similar influencers. However, as mentioned before, various companies and organizations may have different definitions of being similar. Therefore, it is also recommended to check out the results presented in the previous section.

For future researchers, the first improvement can be collecting more social media data and applying the same methodology to the more extensive dataset. Also, there are some other solutions to find social media influencers. Future researchers can apply the same methodology to those solutions to check if they can generate similar influencers across different social media platforms.

6. Conclusion

This research provides some experimental results about the similarity aspects when comparing four types of centrality measures and presents constructive advice for markers. Based on the experiment results, for all four centrality measures (degree centrality, closeness centrality, betweenness centrality and eigenvector centrality), they can not generate similar influencers across different social media platforms. The significance of this research is to show markers that it is necessary to analyze different social media platforms individually since different social media platforms contain different influencers.

References

- [1] Auxuer B and Anderson M 2021 Social media use in 2021 *Pew Research Center* <https://www.pewresearch.org/internet/2021/04/07/social-media-use-in-2021/>
- [2] Akbari M, Foroudi P, Fashami RZ, Nasrin M and Khodayari 2022 Let us talk about something: The evolution of e-WOM from the past to the future *J.Bus.Res* **149** 663-89
- [3] Kumar V, Petersen J, and P.Leone R 2007 How valuable is word of mouth? *Har.Bus.Review* **85(10)** 139
- [4] Keller E and Berry J 2003 *The Influentials: One American in ten tells the other nine how to vote, where to eat, what to buy* (New York: The Free Press) p 279
- [5] Kiss C and Bichler M 2008 Identification of influencers — Measuring influence in customer networks *Decision.Support.System* **46** 233-53
- [6] Rodrigues FA 2019 Network Centrality: An Introduction *A Mathematical Modeling Approach*

- from Nonlinear Dynamics to Complex Systems* Macau E vol 22 (Springer, Cham) p 177-196
- [7] Zhang J and Yu L 2017 Degree centrality, betweenness centrality, and close-ness centrality in social network *MSAM2017* **132** 300-3
 - [8] Ruhnau B 2000 Eigenvector-centrality—a node-centrality *Social.Networks* **22** 357-65
 - [9] Koschuetzki D and Schreiber F 2004 Comparison of centralities for biological networks *GermanConf.Bioinformatics(GCB'04)* 199-206