Customer Analysis in the Marketing Strategy of Huawei: A Social Network Analysis Approach

Peijin Li
05/10/2023

1 Introduction

One of the major success behind Huawei Company[1] is that they promote their products through Social Media. Social media analysis helps Huawei better understand their targeted audience, evaluate the impact of marketing campaigns and enhance their business. In this report, I will use social network analysis techniques to study the Huawei's customer connecting pattern. This report is based on the data from Huawei Social Network Data on Kaggle platform[2].

The dataset includes 3,000 nodes from three social media platform—Facebook, Twitter, and Instragram, each of the platform has 1,000 nodes. The data is extracted from the API of these three platforms, and is collected from the Facebook posts, twitter tweets and Instagram posts and comments under the Huawei official account. The edge of the nodes refers to a user leaving a comment under another user, a node with high degree of edges means that the user is very active on such platform. If users are active under Huawei's posts on the platform, this may mean that the platform is a good platform to promote the newly released product/service. Each of the platform has the same subset of the 1,000 users, so the dataset employed in this study has three layers.

2 Methodology

2.1 Centrality Analysis

The centrality metric is used to identify the most important vertices in a graph. Its application includes identifying the most important and influential individuals in a social network. There are various metrics for measuring centrality, such as degree centrality, closeness centrality, betweenness centrality, eigenvector centrality, PageRank centrality., etc. This report employs degree centrality, closeness centrality and betweenness centrality to identify the important users in the network.

- ✓ Degree centrality is a measure of the number of direct connections that a node has in a network. Nodes with a high degree centrality are those that have many connections to other nodes in the network.
- ✓ Closeness centrality measures the proximity of a node to all other nodes in the network. A node with a high closeness centrality is one that can reach all other nodes in the network quickly through its direct connections.

[1] Huawei Technologies Co.Ltd. is a Chinese multinational networking, telecommunications equipment, and services company headquartered in Shenzhen, Guangdong. It is the largest telecommunications equipment manufacturer in the world, having overtaken Ericsson in 2012.

[2] Data Link: Huawei Social Network Data | Kaggle

✓ Betweenness centrality measures the extent to which a node lies on the shortest paths between all pairs of other nodes in the network. Nodes with high betweenness centrality act as important bridges or intermediaries between other nodes in the network.

2.2 Exponential Random Graph Models

Exponential Random Graph Models (ERGMs) provide a framework for studying how network properties, such as the presence of certain patterns or the degree to which nodes are connected, influence the likelihood of other network properties.

ERGMs assume that the probability of a particular network configuration (e.g., the presence or absence of ties between certain nodes) can be modeled as a function of the underlying social mechanisms that govern the formation of those ties. These mechanisms can include factors such as homophily (the tendency for individuals to form ties with others who are similar to them), transitivity (the tendency for friends of a person to be friends with each other), and popularity (the tendency for highly connected individuals to attract more ties)[3]. ERGMs estimate the parameters that describe the effects of these mechanisms on the probability of observing a particular network configuration. By doing so, ERGMs can provide insight into the social processes that shape the structure of social networks.

3 Findings

3.1 Exploratory Data Analysis

I initialized the analysis by exploring the properties of these three platform, including diameter, density, clustering coefficient and average degree.

Chart 1 Network Fundamental Features

Metric	Facebook	Instagram	Twitter
Number of Vertices	1000	1000	1000
Number of Edges	50153	4933	250315
Diameter	3	5	2
Density	0.1004	0.0098	0.5011
Clustering Coefficient	0.1002	0.0083	0.5012
Average Degree	100.306	9.866	500.63

Diameter measures the longest shortest path between any two nodes in a network. Twitter has the shortest diameter of 2 and the largest density of 0.511, indicating that the nodes in the network are closely connected; while Instagram has the largest diameter of 5 and the lowest density of 0.0098, suggesting the network is less connected to each other.

With a high clustering coefficient of 0.5012 and high average degree of 500.63, **the Twitter network is highly interconnected and there are many closely-knit communities within the network.** The Instagram has the lowest clustering coefficient and the lowest average degree, indicating that **it is a sparse and disconnected network**. There are few connections

[3] Wikipedia contributors. (2023). Exponential family random graph models. Wikipedia. https://en.wikipedia.org/wiki/Exponential_family_random_graph_models

between nodes, and the nodes tend to be relatively isolated. This can result in low overall network efficiency and communication between nodes may be limited.

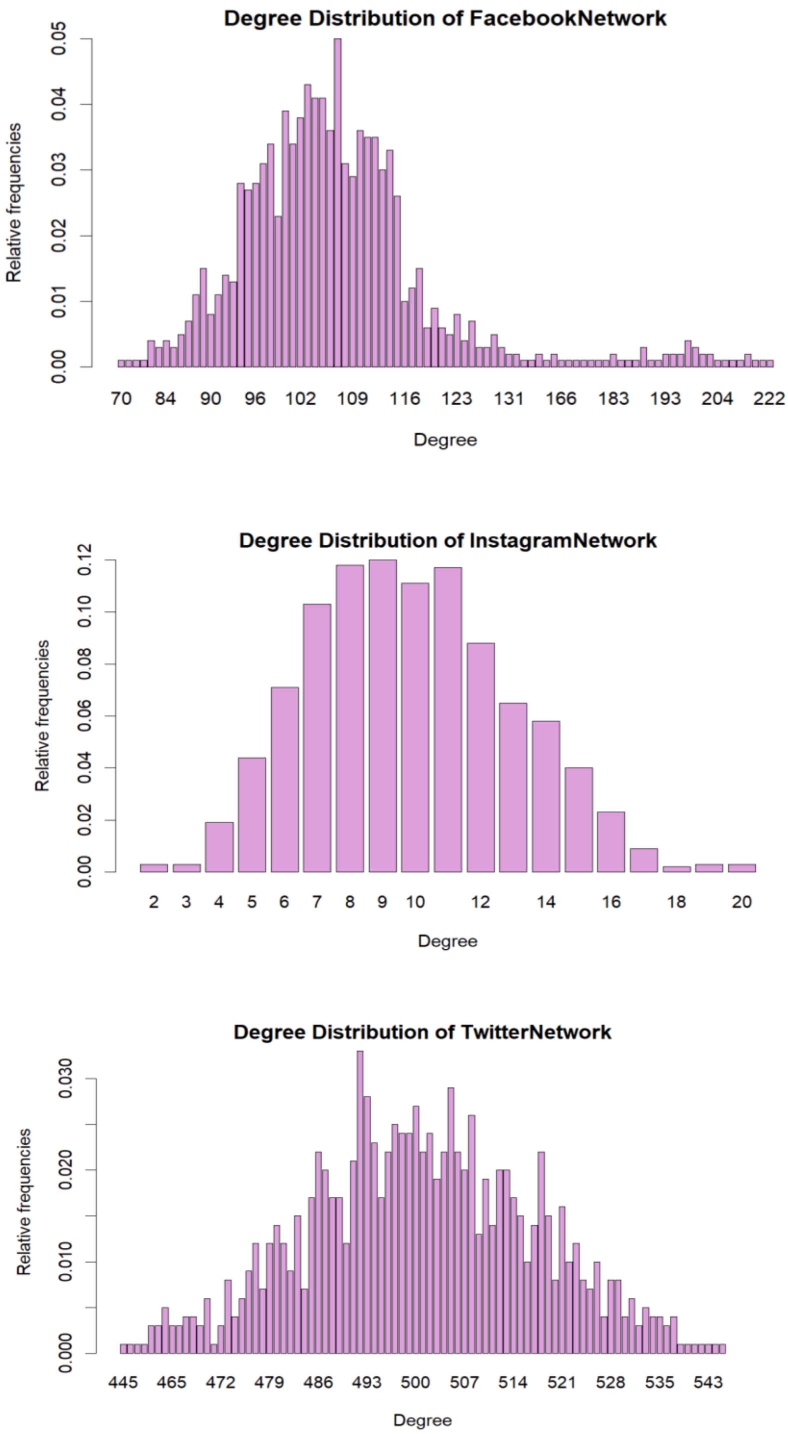


Figure 1 Degree Distribution of Networks

By comparing the x-axis and y-axis scales of the three networks' degree-distribution, it can be seen that there are fewer connections between the nodes of the Instagram network and a high concentration of connections between the nodes of the Twitter network (with y-axis scale to be 0- 0.03). In Facebook network, there are fewer nodes with a large number of connections and there is a normal distribution in the 70-130 degree interval.

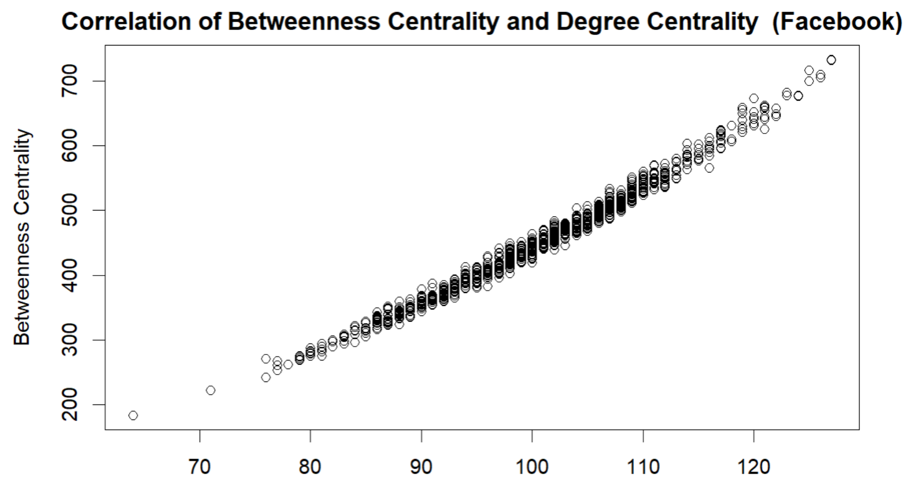
3.2 Centrality Analysis

Centrality Analysis aims to answer the question "What characterizes an important vertex?"[4] Between centrality measures the extent to which a node acts as a bridge or intermediary between other nodes in the network. Nodes with high betweenness centrality are those that lie on many of the shortest paths between pairs of other nodes. This can make them critical nodes in terms of network structure and function, as their removal can lead to the fragmentation or breakdown of the network.

By finding the top 10 highest level of between centrality users on the three platforms, it can be found that based on the same subgroup of users, the highest between centrality users on the three platforms are different, which is intuitive because in general each user will choose to post mainly on one platform.

Chart 2 Top 10 Highest Level of Between Centrality Users

Facebook	Instagram	Twitter
Ernie	Alveena	Homer
Engkos Kosasih	Ishku Ishku	Ellis
Sylvia	KH Hassan	Dililah
Zack	Alexis	Kattie
Fahad Rehman	Farah Samad	Jasmine
Umtiti	Hallie	Azhar
Noor	Betsy	Zak
Ahmed	Misno	Adnan
Asghar	Danish Saifullah	Bruce Wix
Abdullah Khan	Waseem	Benjamin



[4] Wikipedia contributors. (2023b). Centrality. Wikipedia. https://en.wikipedia.org/wiki/Centrality

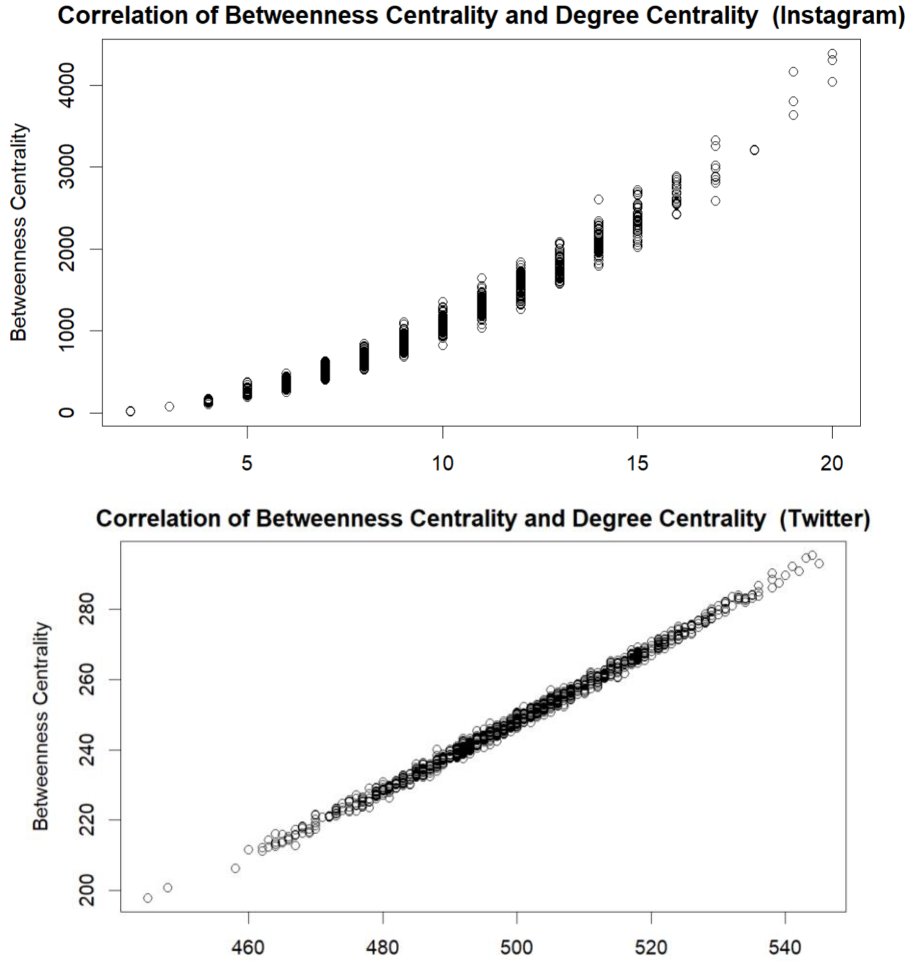


Figure 2 Correlation of Between Centrality and Degree Centrality

Degree centrality measures the number of edges (or connections) that a node has in a network. It is a local measure of a node's importance within its immediate neighborhood; while betweenness centrality measures how often a node acts as a bridge along the shortest path between two other nodes. It is a global measure of a node's importance in the entire network. It shows that between centrality and degree centrality for all three platforms are positively correlated, suggesting that nodes with high degrees also tend to have high betweenness centrality. This indicates that **highly connected nodes also tend to be important bridges in the network**, connecting different parts of the network and facilitating communication between nodes.

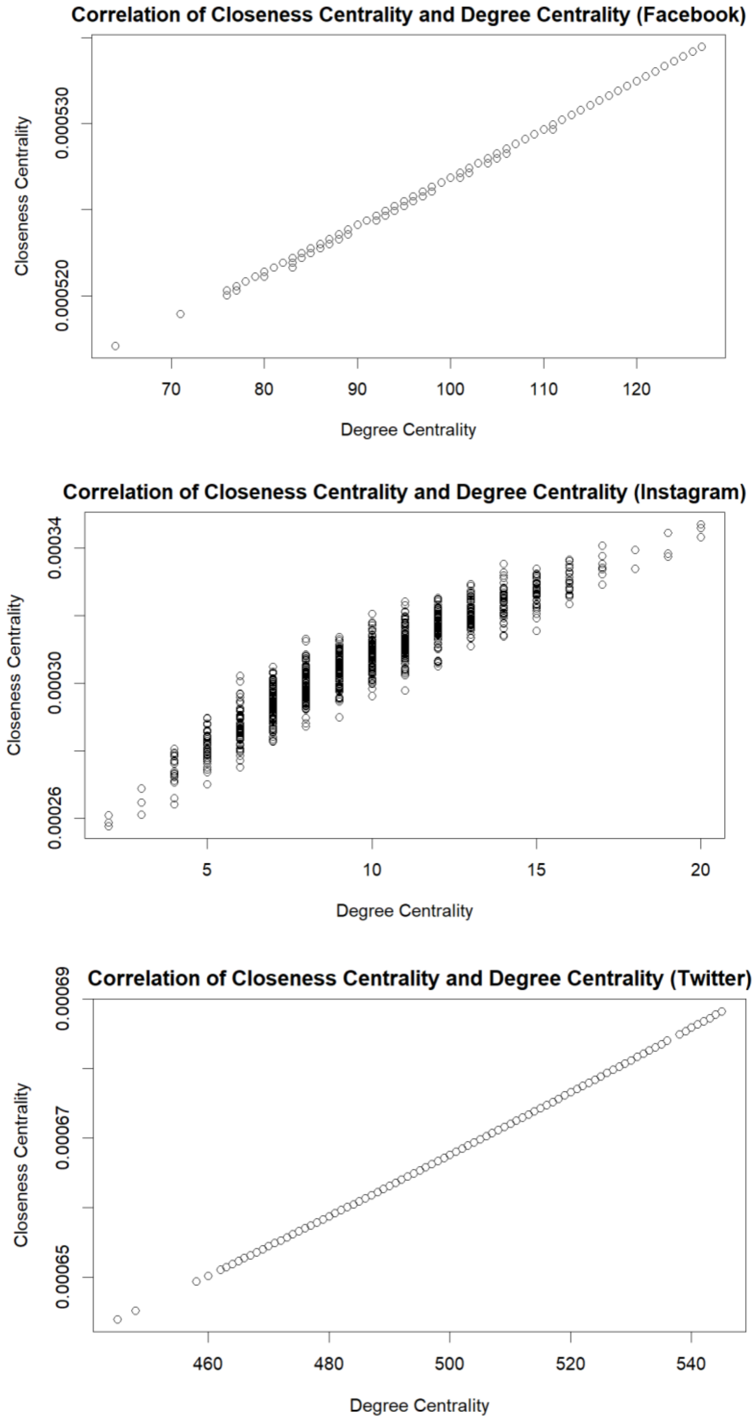


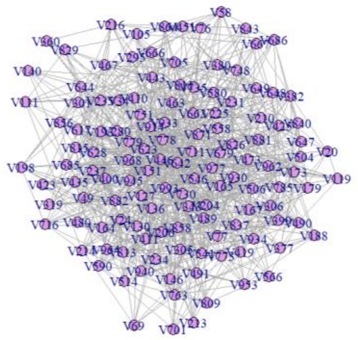
Figure 3 Correlation of Closeness Centrality and Degree Centrality

Degree centrality measures the number of direct connections a node has to other nodes in the network, while closeness centrality measures the average distance of a node to all other nodes in the network. Due to the large sample size, the closeness centrality and degree centrality for all three platforms are positively correlated. Since nodes with high degree centrality are often more central in the network and have shorter average distances to other nodes, which contributes to higher closeness centrality scores. Such correlation is almost linear in Facebook and Twitter networks. For Instagram network, there are nodes with the same degree centrality have different closeness centrality, it means that some users may have a similar number of connections but are positioned differently in the network. **The users with a higher closeness centrality are located in areas of the network where the average shortest path to all other users is shorter.** This could be due to various reasons such as a user being located in a central location with many direct connections to other important nodes, or the node being part of a densely connected group within the network.

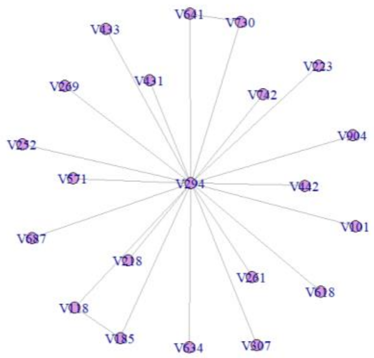
In this step, I created the ego network for the highest degree node of each graph. In the Facebook network, Engkos Kosasih has the highest degree. The node count of the largest ego network (excluding the ego node) is 127, and the density of the ego network is 0.1139272. In the Instagram network, Alexis has the most connections. The node count of the largest ego network (excluding the ego node) is 20, and the density of the ego network is 0.1047619. In the Twitter network, Dililah has the highest degree. The node count of the largest ego network (excluding the ego node) is 545, and the density of the ego network is 0.5047081.

Largest ego network is representative for the patterning of the social network in which it is located. As revealed in the previous exploratory data analysis, it can be seen that **the Twitter and Facebook network are highly interconnected and there are many closely-knit communities within the network. The Instagram network is a sparse and disconnected network.**

Largest Ego Network in FacebookNetwork



Largest Ego Network in InstagramNetwork



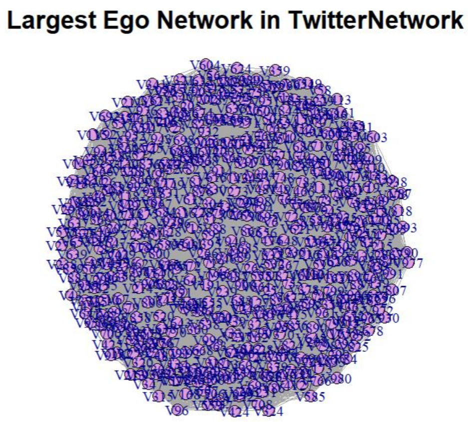


Figure 3 Largest Ego Network

3.3 Exponential Random Graph Models

Before conducting the ERGM analysis, I used the "rethnicity" package[5] to predict the race of 1,000 users by their last names. According to the prediction results, out of 1,000 users, there are 502 Asian people, 166 Black people, 113 Hispanic people and 219 White people. Huawei, as a Chinese brand, has a large proportion of Asian users, so there are more comments on social media on topics related to Huawei products/services come from Asian consumers.

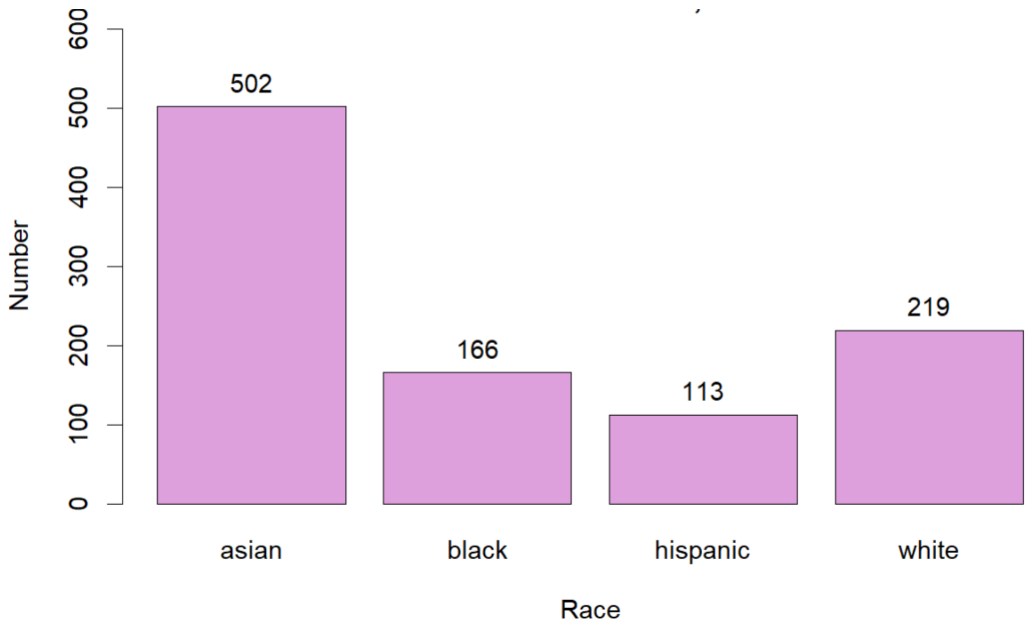


Figure 4 Race Prediction for the Users

Then, using race as the attributes of the network, I used ERGM to analyze the three platforms separately.

[5] Rethnicity Github repo for more information: https://github.com/fangzhou-xie/rethnicity

```
Maximum Likelihood Results:

                         Estimate Std. Error MCMC %  z value Pr(>|z|)    
edges                   -2.192311   0.011119      0 -197.176   <1e-04 ***
nodematch.Race           0.004001   0.011421      0    0.350    0.726    
nodefactor.Race.black   -0.004720   0.010195      0   -0.463    0.643    
nodefactor.Race.hispanic -0.012455   0.011880      0   -1.048    0.294    
nodefactor.Race.white    0.005915   0.009101      0    0.650    0.516    
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

     Null Deviance: 692454  on 499500  degrees of freedom
 Residual Deviance: 325646  on 499495  degrees of freedom
 
AIC: 325656  BIC: 325712  (Smaller is better. MC Std. Err. = 0)
```

Figure 5 ERGM Results for FacebookNetwork

In these three results, we can see that the edges are statistically significant(p-value < 0.001) in the ERGM result, it means that **the presence or absence of those edges is not random and has a significant effect on the network structure**. In other words, the model suggests that the observed network is more likely to have those edges compared to a random network with the same number of nodes and edges. Nodematch.Race is not statistically significant in all the ERGM result, indicating that **the formation of a tie is not based on whether two people are from the same race**.

Based on the results of Instagram network, it is observed that nodefactor.Race.white is statistically significant, suggesting that the race of White has an impact on the number of his/her connections on Instagram. With an estimate value of -0.07436, it indicates that the presence of that this attribute decreases the likelihood of an edge forming between two nodes on the platform. In other words, White people are less likely to be connected to other nodes in the network.

```
Maximum Likelihood Results:

                         Estimate Std. Error MCMC %  z value Pr(>|z|)    
edges                    -4.54746    0.03418      0 -133.028  < 1e-04 ***
nodematch.Race           -0.03154    0.03527      0   -0.894  0.37121    
nodefactor.Race.black    -0.04592    0.03143      0   -1.461  0.14399    
nodefactor.Race.hispanic -0.01125    0.03597      0   -0.313  0.75438    
nodefactor.Race.white    -0.07436    0.02852      0   -2.608  0.00911 ** 
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

     Null Deviance: 692454  on 499500  degrees of freedom
 Residual Deviance:  55367  on 499495  degrees of freedom
 
AIC: 55377  BIC: 55433  (Smaller is better. MC Std. Err. = 0)
```

Figure 6 ERGM Results for InstagramNetwork

```
Maximum Likelihood Results:

                          Estimate Std. Error MCMC % z value Pr(>|z|)
edges                     0.012373   0.006686      0   1.850  0.06424 .
nodematch.Race            0.005372   0.006870      0   0.782  0.43429
nodefactor.Race.black    -0.018520   0.006122      0  -3.025  0.00249 **
nodefactor.Race.hispanic -0.005606   0.007115      0  -0.788  0.43076
nodefactor.Race.white    -0.005153   0.005485      0  -0.940  0.34746
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

     Null Deviance: 692454  on 499500  degrees of freedom
 Residual Deviance: 692438  on 499495  degrees of freedom

AIC: 692448  BIC: 692503  (Smaller is better. MC Std. Err. = 0)
```

Figure 7 ERGM Results for TwitterNetwork

The result of Twitter network ERGM shows that the race of Black has an impact on the number of his/her connections, statistically significant with p-value < 0.01. With an estimate value of -0.01852, it indicates that the presence of that Black people are less likely to be connected to other nodes in the Twitter network.

4 Conclusion

Based on the analysis of network data from the three platforms, we could learn that the Twitter network and Facebook network are highly interconnected and there are many closely-knit communities within the Twitter network. Information spreads quickly in both communities, and because of the presence of close-knit communities, discussions about new products/services are very lively. Therefore, Twitter and Facebook are good platforms to promote information about the latest product/service launches, as well as events that have recently taken place. **Twitter could be the best platform to launch a marketing campaign.**

Compared to users on Twitter and Facebook, the Instagram has the lowest clustering coefficient and the lowest average degree, indicating that it is a sparse and disconnected network. Information is less mobile on Instagram.

According to the results of Exponential Random Graph Models, we found that race of an individual has an impact on the formation of edges between nodes. Specifically, on Instagram, the race of White has a negative impact on the number of his/her connections. On Twitter, the race of Black has a negative impact on the number of his/her connections. However, the formation of a tie is not based on whether two people are from the same race. Such results do not necessarily mean that the Huawei community on Twitter and Instagram is racially biased, because, as a Chinese brand with a strong national spirit, Huawei is also influenced by countries' relations and the international situation. Moreover, the results are also influenced by the number of different ethnic users on different platforms. Future research could further validate the study's findings by controlling for these variables.