벤처창업연구 제14권 제4호 (통권64호)
Asia-Pacific Journal of Business Venturing and Entrepreneurship Vol.14 No.4 pp.199-209

Exploring Twitter Follower-Networks of Startup Companies Employing Social Network Analysis and Cluster Analysis

Yu, Seunghee (Associate Professor, School of Business, Sejong University)*

Abstract

The importance of business strategy for successful social media engagement has quickly increased as more businesses engage in social media. The importance is even greater for startup companies because startup companies are genuinely new to business, and they need to increase their presence in the market, and quickly access future customers. The objective of this paper lies in exploring key indicators of social media engagements by selected startup companies. The key indicators include two aspects of social media usages by the companies: i) overall social media activities, and ii) properties of network structure of the information flow platform provided by social media service. To better assess and evaluate the key indicators of social media usages by startup companies, the indicators will be compared with those of selected large established companies.

Twitter is selected as a social media service for the analysis of this paper, and using Twitter REST API, data regarding the key indicators of overall Twitter activities and the Twitter follower-network of each company in the sample are collected. Then, the data are analyzed using social network analysis and hierarchical clustering analysis to examine the characteristics of the follower-network structures and to compare the characteristics between startup companies and established companies. The results show that most indicators are significantly different across startup companies and established companies. One key interesting finding is that the startup companies have proportionally more influencers in their follower-networks than the established companies have. Another interesting finding is that the follower-networks of startup companies in the sample have higher modularity and higher transitivity, suggesting that the startup companies tend to have a proportionally larger number of communities of users in their follower-networks, and the users in the networks are more tightly connected and cohesive internally. The key business implication for the future social media engagement efforts by startup companies in general is that startup companies may need to focus on getting more attention from influencers and promoting more cohesive communities in their follower-networks to appreciate the potential benefits of social media in the early stage of business of startup companies.

KeyWords: Social Network Services, Twitter Follower Networks, Social Network Analysis, Cluster Analysis

Ⅰ. Introduction

Social media has proven beneficial to business for the past decades, and businesses have started recognizing and realizing the potential benefits of social media for the success of their business. The most fundamental advantage of social media is that it provides an effective and efficient platform for information flow among businesses and customers. Some of the key business benefits of social media enabled by the key advantage include faster and easier communication with customers, better understanding about customers and increased brand awareness. Social media helps business to understand and access their target audience better and allows customers to provide instant feedback, enabling the business to have valuable insight on how customers perceive the products and services provided by the business and to improve its market opportunities.

As businesses become more aware of these benefits, social media has become a huge part of everyday business activities nowadays. For example, 91% of 2018 Fortune 500 companies have active corporate Twitter accounts, and top leading Fortune companies continue to post on their Twitter accounts. Also, at least one company in 73 industries represented in the 2018 Fortune 500 has a corporate Twitter account, and 100% of the Fortune 500 companies in some industries(e.g., utilities, commercial banks and specialty retailers) have Twitter

* Associate Professor, School of Business, Sejong University, shyu@sejong.edu
· 투고일: 2019-08-20. · 수정일: 2019-08-27. · 게재확정일: 2019-08-28.

accounts(Barnes et al., 2019). This trend is also similar among small businesses, and 70% of them are on Twitter(Bansal & Annika, 2019).

The importance of business strategy for successful social media engagement has quickly increased as more businesses engage in social media. The importance is even greater for startup companies because startup companies are genuinely new to business, and they need to increase their presence in the market, and quickly access future customers. In fact, startup companies may have more potential to take advantages of social media. There are quite a few industry suggestions about how business can take advantage of social media for business success, especially for startup companies. However, we may first need to investigate and improve our understanding about the way that social media is used in business. We especially need to have a better understanding about the way that social media provides a new platform for effective and efficient information flow among businesses and customers. The better understanding will greatly facilitate the development of a viable social media strategy for business.

The objective of this paper lies in exploring key indicators of social media engagements by selected startup companies. The key indicators cover two aspects of social media usages by the companies: i) overall social media activities, and ii) properties of network structure of the information flow platform provided by social media service. To better assess and evaluate the key indicators of social media usages by startup companies, the indicators will be compared with those of selected large established companies.

To serve the objective of this paper, first, Twitter is selected as a social media service for the analysis of this paper. Twitter is one of the most popular social media services being actively used by 91% of the Fortune 500 companies, compared with Facebook being used by 89% of the Fortune 500 companies(Barnes et al., 2019). Barnes et al.(2019) also claims that it is proven that companies should focus their social media efforts on Twitter. Second, using Twitter REST API, data regarding the key indicators of social media activities are collected, and to examine the properties of network structure of the information flow provided by Twitter, the follower-network of each company is collected. The follower-network is selected because the number of followers is the most representative indicator of social media engagement, and therefore, follower-networks seem best suited for investigating how well social media plays its role as an effective and efficient platform for information flow. Then, the data are analyzed using social network analysis and hierarchical clustering analysis to examine the characteristics of structure of the follower-networks and to compare the characteristics between two sample groups: startup companies and established companies. The findings will help us to gain a better understanding and useful insights of the key aspects of social media engagements by startup companies. The findings are also expected to have valuable implications for developing viable social media engagement strategies for startup companies.

Ⅱ. Social Media Activities and Network Structure

This paper investigates two types of indicators of social media engagements in the context of Twitter as follows: i) indicators of overall Twitter activities, and ii) properties of social network structure of the Twitter follower-network. Indicators of the overall Twitter activities are related to how popular a company is and how actively a company engages in Twitter. The indicators for popularity include the number of followers, the number of Twitter lists, and the number of influential followers in the follower-network. The indicator for active engagement include the number of friends and the number of status update. The properties of social network structure encompass key metrics from social network analysis perspective. Key metrics include density, average path length, diameter, transitivity, centralization and modularity. In this paper, all these metrics are investigated in the context of Twitter follower-network, and they are compared between startup companies and established companies. Since this paper is an exploratory study in nature, these indicators and metrics will be examined without any specific hypotheses being proposed up front. All the indicators and metrics will be further explained below.

2.1 Indicators of Overall Twitter Activities

The foremost important indicator of engagement on Twitter is the number of followers. Having a large number of followers on Twitter means that many Twitter users are subscribing to the tweets or messages from the company that they are following, and therefore, the company can reach a broad range of potential customers, possibly leading to increased sales for the company in the future(Barnes et al., 2019). The number of followers may improve the company’s brand popularity, and therefore, an increase in the number of followers is considered one of the key indicators of social media effectiveness.

Another indicator of popularity is the number of Twitter lists a company is on. Twitter users are allowed to create lists of their interests and add accounts that they find interesting to the lists. A company may be considered more popular if the company is added to more lists. It's very likely that a high correlation between the number of followers of a company and the number of lists the company is on. However, this may not be always the case, and therefore, the number of lists may be a complementary indicator of Twitter popularity.

The number of friends is also important measure of Twitter engagement of a company. Having Twitter friends means that the company is following other users. Having many Twitter friends may indicate that a company is very active on Twitter and eager to participate in the Twitter activities. Also, having many friends may help the company get involved with other Twitter users and increase the chance to get useful information from them. Another indicator is the number of Twitter status update. Through the status update, Twitter users can simply indicate when they are on Twitter. A more detailed status update could allow a Twitter user to let other users know what he or she is up to at certain time. The number of status update is certainly a good indicator of how active a Twitter user is.

Finally, the number of influential followers or influencers in the follower-network will be a good indicator of a company's Twitter activities. Influencers are individual Twitter users who have authority, knowledge or relationship with other users, and therefore may have the influential power to affect other Twitter users. Attracting more influencers in the follower-network of a company means that the company is already popular, and more importantly is very likely to attract more followers in the future.

2.2 Metrics of Social Network Structure of the Twitter Follower-network

A social network structure is created connections(i.e., "links," or "edges") that are formed among actors(i.e., "nodes" or "vertices"), such as Twitter followers(Wasserman & Faust, 1994). Research on social media from a social network perspective focuses on relational ties between social entities. Bruns & Stieglitz(2013) discuss various Twitter-specific metrics. On Twitter, social networks are composed of users and the connections they form with other users when they follow one another(Hansen et al., 2011). Understanding the overall structure of a network is essential for assessing how information flows among its users. Examining Twitter-follower network in the perspective of social network analysis will help us understand the key social network metrics of Twitter-follower networks and identify the key characteristics of the Twitter follower-networks of startup companies especially in comparison to those of established companies. The major social network metrics examined in this paper are concerned about two key aspects of social networks, interconnectedness and centralization. The specific metrics are briefly explained below.

2.2.1 Interconnectedness Metrics

Density is the key measure of the interconnectivity of actors in a network. The density of network affects the rate of information flow within the network. Density is low when a group of actors are loosely connected and high when actors are highly interlinked(Hansen et al., 2011). Interaction between actors causes shared knowledge and leads to even more interaction, enhancing the stability of a group(Carley, 1991). Lerman & Ghosh(2010) examined the role of social network structures in the spread of information on Digg and Twitter and found that the rate of information flow through a network depends on its density.

Another key metric of network connectivity is modularity. Modularity is a measure of the quality of clustering(Newman & Givran, 2004, Clauset et al., 2004) and is designed to measure the degree of a network divided into modules(also called clusters or communities). High modularity means that there exist dense connections between the nodes within modules, but sparse connections between nodes in different modules. Measuring the characteristics of clusters and the embedded network can enhance our understanding of information flow in social media. A related metric is transitivity(or global clustering coefficient), which is a measure of the degree to which nodes in a graph tend to cluster together. While modularity looks at edge densities in given clusters compared to edge densities between clusters, transitivity looks at density of triangles in comparison to induced density of triplets(Wasserman & Faust, 1994). These two metrics employ two somewhat different concepts, but both may work very well to measure the quality of clustering in a network. Other measures of network interconnectivity include average degree, average path length, and network diameter. Average degree is the average number of links per nodes and is closely related to the density of a network. Average path length is defined as the average number of steps along the shortest paths for all possible pairs of network nodes. It is a measure of the efficiency of information flow on a network. Network diameter is the shortest distance between the two most distant nodes in the network and represents the linear size of a network(Wasserman & Faust, 1994).

2.2.2 Centralization Metrics

Centralization captures another key characteristic of network structure. The centralization of a network is a measure of the extent to how central its most central node is in relation to how central all the other nodes are(Freeman, 1979). Centralization is an extension of the definition of centrality on the node level to the whole network. Centralization can be calculated on node-level centrality measures including degree centrality, closeness centrality, betweenness centrality and eigenvector centrality, at the network level. For example, in a network with high degree centralization, one or a few nodes bring many other connections, and the information flow in this network depends highly on these central nodes with high degree centrality. Two common structures are tree-like network and star-shaped network(Kumar et al., 2006). Tree-like network has lower level of degree centralization than star-shaped network. Ediger et al.(2010) found that news dissemination on Twitter fell into tree-like broadcast patterns, and Himelboim & Han(2013) found that star-shaped clusters disappeared as soon as their core actors stopped tweeting about the topic, exhibiting the vulnerability of star-shaped network despite its higher level of degree centralization. In this paper, degree centralization, closeness centralization and eigenvector centralization metrics are examined to explore the centralization characteristics of Twitter-follower networks.

Ⅲ. Research Methods

3.1 Data Collection

Data on the overall Twitter activities were collected for two sample groups: selected startup companies and selected established companies. 20 startup companies and another 20 Fortune 100 companies were selected from the lists provided by Sajid(2019) and Ranker(2019), respectively. The data collection was done through Twitter REST API using R packages, twitteR(Gentry, 2015) and rtweet(Kearney, 2019), in July 2019.

The data on the indicators of overall Twitter activity are presented in <Table 1> and <Table 2> below. Also, the summary statistics and the results of tests for sample means are shown in <Table 3>. The screen name in the table is the user name of a Twitter account, which is also known as ‘Twitter handle.’ OrgID is an ID assigned to each company in the sample for the convenience of identifying each company in the series of analyses of this study. The ‘Listed Ratio’ in the tables means the ratio of ‘Listed’ and ‘Followers.’ The ‘Influencers Ratio’ is calculated as the proportion of influencers among all 1st-degree followers. A 1st-degree follower is categorized as an influencer when the follower has more than or equal to 3,000 followers.

Then, data on Twitter follower-networks of all sampled companies were collected. Because of Twitter’s API rate limit, the follower-network data were mined for a sample of 1st-degree followers of each company and all followers of the sampled 1st-degree followers(or 2nd-degree followers). Therefore, the follower-network of each company has only two degrees of separation, and yet it leaves us with a network of up to 90,623 edges. The sample of the 1st-degree followers was randomly selected from all the followers of each company and the sample size was determined employing a simplified formula proposed by Yamane(1967). The social networks metrics of the follower-network of each company were calculated using iGraph package(Csardi & Nepusz, 2006). The summary statistics of the network metrics and the results of tests for comparing the means of the metrics over two sample groups are presented in <Table 4> and <Table 5>.

<Table 1> Overall Twitter Activities: Startup Companies (ST)

Screen Name	Org.ID	Followers	Friends	Listed	List Ratio	Status Updates	Influencers Ratio
FrontApp	ST01	6,587	4,597	249	0.037802	6,847	0.080808
_FiveAI	ST02	3,052	552	127	0.041612	883	0.054945
DandelionEnergy	ST03	1,053	48	18	0.017094	254	0.072289
bundilapp	ST04	282	217	5	0.017730	111	0.176471
sano_int	ST05	1,462	33	103	0.070451	226	0.047619
molekuleair	ST06	2,798	408	47	0.016798	1,721	0.121951
AuraHealthHQ	ST07	1,058	541	27	0.025520	193	0.064935
Aaptiv	ST08	13,421	504	69	0.005141	5,360	0.078947
possiblefinance	ST09	113	112	0	0.000000	14	0.161290
inamo	ST10	274	45	101	0.368613	1,042	0.152778
TransferGo	ST11	2,636	743	204	0.077390	1,667	0.083333
AlpacaHQ	ST12	1,544	1,256	48	0.031088	1,205	0.123596
tributi_online	ST13	97	130	0	0.000000	60	0.108108
imbellus	ST14	594	68	25	0.042088	90	0.088235
Fluentify	ST15	1,585	502	27	0.017035	923	0.037975
aceable	ST16	3,192	837	78	0.024436	10,807	0.082353
carserv_io	ST17	264	357	10	0.037879	364	0.112903

KobitonMobile	ST18	335	214	42	0.125373	740	0.111111
SoloStove	ST19	3,444	44	44	0.006388	811	0.043478
squartwolf	ST20	119	173	1	0.008403	114	0.040816

<Table 2> Overall Twitter Activities: Fortune 100 Companies (FT)

Screen Name	Org.ID	Followers	Friends	Listed	List Ratio	Status Updates	Influencers Ratio
Alcoa	FT01	27,158	5,262	685	0.025223	6,571	0.011628
allstate	FT02	84,830	4,943	1,274	0.015018	25,976	0.039474
BankofAmerica	FT01	528,696	660	2,011	0.003804	9,228	0.037975
Citi	FT04	913,848	342	3,164	0.003462	17,096	0.050000
CocaCola	FT05	3,339,538	62,057	12,917	0.003868	267,530	0.028169
comcast	FT06	176,374	0	1,705	0.009667	56,907	0.039474
exxonmobil	FT07	305,081	270	2,482	0.008136	6,905	0.022989
GE_Reports	FT08	60,845	2,354	897	0.014742	6,369	0.023529
HoneywellNow	FT09	2,389	2	68	0.028464	1	0.065934
IngramMicroInc	FT10	21,614	280	547	0.025308	5,771	0.088889
libertymutual	FT11	97,653	2,553	500	0.005120	15,524	0.097561
northropgrumman	FT12	194,209	671	2,446	0.012595	12,566	0.000000
pfizer_news	FT13	16,182	891	379	0.023421	785	0.048193
Safeway	FT14	77,521	3,390	983	0.012680	62,364	0.064935
sprintnews	FT15	107,731	84	1,224	0.011362	3,969	0.027027
StateFarm	FT16	100,587	9,087	1,272	0.012646	48,627	0.025641
Target	FT17	1956,767	2,626	7,909	0.004042	753,48	0.012500
verizon	FT18	1,658777	356	5,757	0.003471	123,426	0.042254
Walgreens	FT19	880,928	1,587	2,677	0.003039	60,294	0.033333
WellsFargo	FT20	309,323	330	2,526	0.008166	17,257	0.040000

3.2 Preliminary Analysis

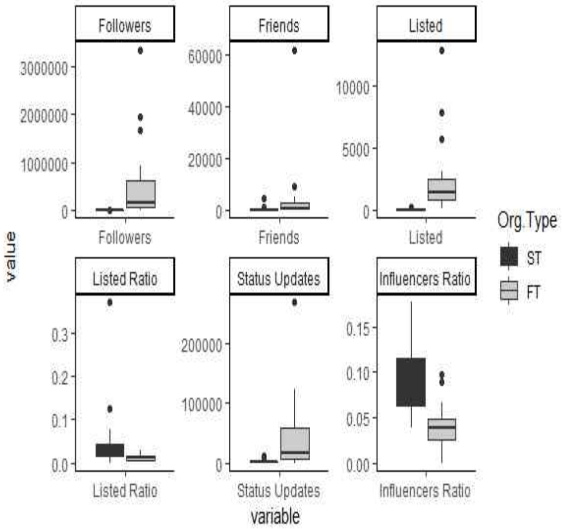
As stated previously, this paper examines the key Twitter activity indicators and the key metrics of social network structure of Twitter follower-networks. In particular, employing hierarchical cluster analysis method, further analysis will examine whether the two sample groups differ each other on the key network metrics by investigating if companies in one group are more similar to each other than companies in the other group. In the next section, a cluster analysis will be done to see if the startup companies and the established companies are categorized into different clusters, and the clusters will be further examined to see how the companies assigned to each cluster are different from each other on the key network metrics. In this section, as a preliminary analysis, sample means of the indicators and metrics of the two sample groups are simply compared to see if there exist any meaningful differences between the two groups before the companies are categorized into different clusters based on a cluster analysis. The test results along with summary statistics of the key Twitter activity indicators and the network structure metrics are presented in the tables below(<Table 3>, <Table 4> and <Table 5>). Given that the data is non-normal, testing for comparing sample means was conducted using Kruskal-Wallis rank sum test(Campbell & Swinscow, 2009).

<Table 3> Summary Statistics and Tests for Sample Means: Overall Twitter Activities

	ST (N=20)	FT (N=20)	Total (N=40)	p value
Followers				< 0.001
Mean (SD)	2,195.5 (3,105.5)	543,002.6 (859,432.1)	272,599.0 (659,423.9)	
Range	97.0 - 1,3421.0	2,389.0 - 3,339,538.0	97.0 - 3,339,538.0	
Friends				0.030
Mean (SD)	569.0 (1001.3)	4,887.2 (13,656.0)	2,728.2 (9,804.2)	
Range	33.0 - 4,597.0	0.0 - 62,057.0	0.0 - 62,057.0	
Listed				< 0.001
Mean (SD)	60.1 (68.2)	2571.2 (3,083.5)	1315.7 (2,500.2)	
Range	0.0 - 249.0	68.0 - 12,917.0	0.0 - 12,917.0	
Listed Ratio				0.005
Mean (SD)	0.0485 (0.0812)	0.0117 (0.0082)	0.0301 (0.0599)	
Range	0.0000 - 0.3686	0.0030 - 0.0285	0.0000 - 0.3686	
Status Updates				< 0.001
Mean (SD)	1,671.6 (2,790.5)	41,125.7 (62,287.4)	21,398.7 (47,885.8)	
Range	14.0 - 10,807.0	1.0 - 267,530.0	1.0 - 267,530.0	
Influencers Ratio				< 0.001
Mean (SD)	0.0922 (0.0407)	0.0400 (0.0246)	0.0661 (0.0424)	
Range	0.0380 - 0.1765	0.0000 - 0.0976	0.0000 - 0.1765	

Regarding the indicators of overall Twitter activities, the results show that all the indicators are statistically different across two sample groups: startup companies and Fortune companies. Startup companies have higher Listed Ratio and Influencers Ratio than Fortune companies in the sample. Fortune companies have more followers, friends and status updates as shown in <Table 3> and <Fig. 1.>.

With regard to the metrics of social network sturucture of the follower-networks, all metrics are statistically different across two sample groups except for closeness centralization as shown in <Table 4> and <Table 5>. To make it easier to compare the differences, boxplots are prepared and presented in <Fig. 2>. Startup companies have higher on most of the metrics except on two metrics, density and degree centralization. These results from the preliminary analysis will be discussed further along with results from the cluster analysis in the later section.



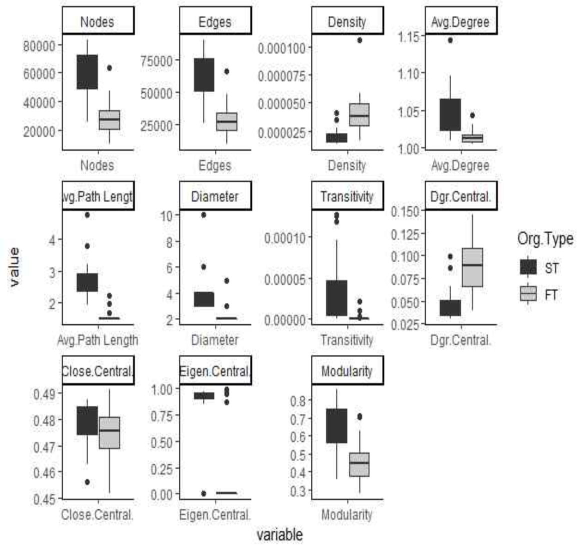
<Fig. 1> Boxplots: Overall Twitter Activities of Startup (ST) and Fortune 100 (FT) Companies

<Table 4> Summary Statistics and Tests for Sample Means: Social Network Structure Part 1

	ST (N=20)	FT (N=20)	Total (N=40)	p value
Nodes				< 0.001
Mean (SD)	56,895.9 (16702.4)	29,287.5 (128,90.4)	43,091.7 (20,305.2)	
Range	25180.0 - 83,062.0	9,543.0 - 63,525.0	9,543.0 - 83,062.0	
Edges				< 0.001
Mean (SD)	59,820.4 (17,923.4)	29,826.8 (13,446.3)	4,823.6 (21,800.5)	
Range	25,793.0 - 90,623.0	9,675.0 - 66,255.0	44,823.6 (21,800.5)	
Density				< 0.001
Mean (SD)	0.000020 (0.000007)	0.000041 (0.000020)	0.000031 (0.000018)	
Range	0.000013 - 0.000041	0.000016 - 0.000106	0.000013 - 0.000106	
Avg. Degree				< 0.001
Mean (SD)	1.0494 (0.0346)	1.0147 (0.0097)	1.0321 (0.0306)	
Range	1.0100 - 1.1440	1.0040 - 1.0430	1.0040 - 1.1440	
Avg. Path Length				< 0.001
Mean (SD)	2.7642 (0.6461)	1.5720 (0.1931)	2.1681 (0.7655)	
Range	1.9439 - 4.7735	1.4939 - 2.2396	1.4939 - 4.7735	
Diameter				< 0.001
Mean (SD)	4.2 (1.7)	2.2 (0.7)	3.2 (1.6)	
Range	3.0 - 10.0	2.0 - 5.0	2.0 - 10.0	

<Table 5> Summary Statistics and Tests for Sample Means: Social Network Structure Part 2

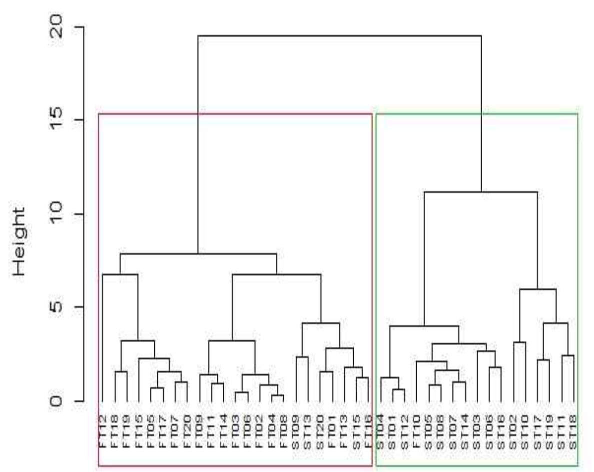
	ST (N=20)	FT (N=20)	Total (N=40)	p value
Transitivity				< 0.001
Mean (SD)	0.000036 (0.000044)	0.000002 (0.000005)	0.000019 (0.000035)	
Range	0.000001 - 0.000126	0.000000 - 0.000021	0.000000 - 0.000126	
Dgr. Centralization				< 0.001
Mean (SD)	0.0487 (0.0183)	0.0886 (0.0302)	0.0686 (0.0319)	
Range	0.0301 - 0.0993	0.0394 - 0.1449	0.0301 - 0.1449	
Close. Centralization				0.123
Mean (SD)	0.4782 (0.0090)	0.4749 (0.0090)	0.4766 (0.0091)	
Range	0.4562 - 0.4872	0.4517 - 0.4913	0.4517 - 0.4913	
Eigen. Centralization				< 0.001
Mean (SD)	0.839217 (0.288737)	0.188572 (0.387489)	0.513894 (0.471501)	
Range	0 - 0.967413	0 - 0.988862	0 - 0.988862	
Modularity				
Mean (SD)	0.6491 (0.1336)	0.4506 (0.1267)	0.5499 (0.1632)	
Range	0.3537 - 0.8598	0.2754 - 0.7134	0.2754 - 0.8598	



<Fig. 2> Boxplots: Social Network Structures of Startup (ST) and Fortune 100 (FT) Companies

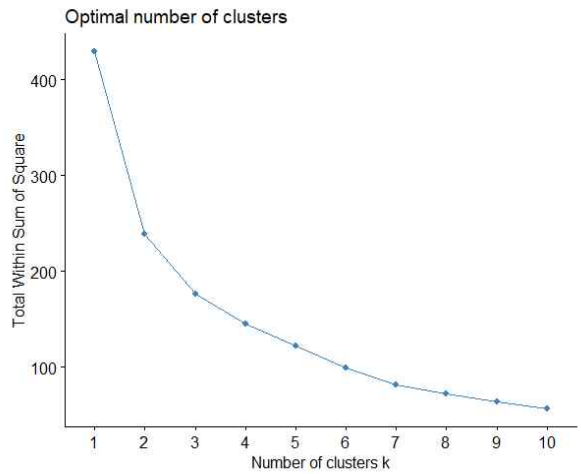
IV. Cluster Analysis and Key Findings

Cluster analysis is intended to categorize a set of objects in such a way that objects in the same group(called a cluster) are more similar to each other than to those in other groups. It is one of the popular unsupervised machine learning techniques, which is suitable for an exploratory data mining. Considering that the main objective of this study is to explore the key indicators of Twitter activities and metrics of Twitter follower-networks of startup companies and compare them with those of established companies as a reference group, cluster analysis seems to serve the objective quite well.



<Fig. 3> Cluster Dendogram on Social Network Structure Metrics of the Twitter Follower-networks

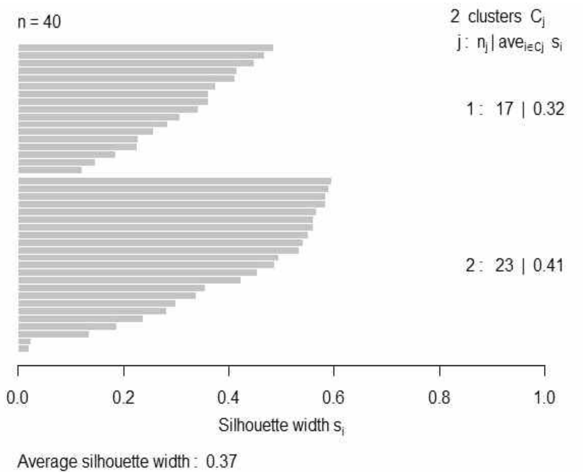
Hierarchical clustering analysis was conducted especially on the metrics of social network structure of the Twitter follower-networks to see if the companies in the same group are similar to each other than the companies in the other group. For the hierarchical clustering analysis in this study, Euclidean distance matrix and ward's linkage method were adopted(Ward, 1963). A dendogram is known as the main graphical tool for looking at the hierarchical cluster solution. The dendogram of the network structure metrics is shown in <Fig. 3>.



<Fig. 4> Elbow Plot of Social Network Structure Metrics of the Twitter follower-networks

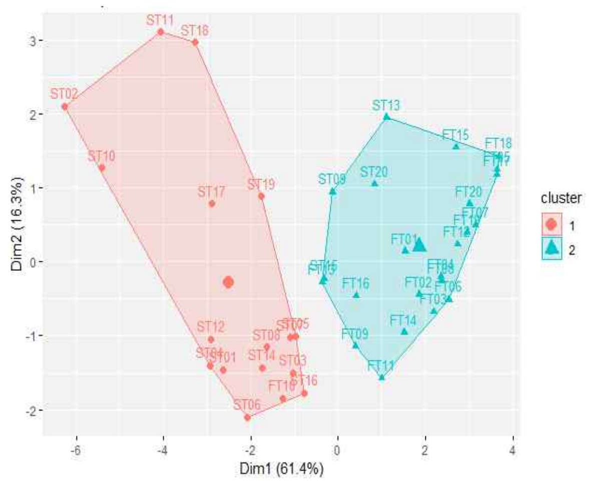
To determine the optimal number of clusters, a elbow method(Kassambara & Mundt, 2017) is employed and the result is presented in <Fig. 4>. The graph plots the percentage of variance explained by the clusters against the number of clusters. The number of clusters is chosen at some point where the marginal gain from adding a cluster drops, giving an angle(just like an elbow) in the graph(Ketchen & Shook, 1996). The elbow plot in <Fig. 4> suggests that the optimal number of clusters is 2.

As a complementary and an alternative method for selecting the number of clusters, a silhouette plot provides very useful information for locating observations in a cluster analysis that may not be categorized properly. A silhouette plot is intended to show how well each observation fits into the cluster of which each observation is a member(Rousseeuw, 1987). The information provided by the silhouette plot can be used to choose the proper number of clusters. A silhouette plot for the cluster analysis solution was prepared and presented in <Fig. 5>. The plot shows that most observations seem to belong to the cluster assigned, supporting that there exists a good structure in the cluster solution.



<Fig. 5> Silhouette Plot for the Cluster Solution

To visualize the result of the cluster analysis of this study, a cluster plot is also prepared(Kassambara & Mundt, 2017) and presented in <Fig. 6>. The cluster plot illustrates the clusters with the two principal components that are identified based on the results from principal components analysis, exhibiting the majority of the variance in the model. The plot depicts two clusters and shows how the observations assigned in the two clusters are separated from each other. The scatter plot shows that the two clusters are nicely separated from each other, supporting that the cluster analysis result shows a reasonable cluster structure in the data.



<Fig. 6> Cluster Plot of Social Network Structure Metrics of the Twitter Follower-networks

Finally, agglomerative coefficient is also calculated(Maechler et al., 2019) to evaluate the cluster analysis solution. The agglomerative coefficient measures the amount of clustering structure, and the value being closer to 1 suggests a strong clustering structure. The value of coefficient of the cluster solution of this study is 0.9175, suggesting that the solution provides a reasonable cluster structure in the data.

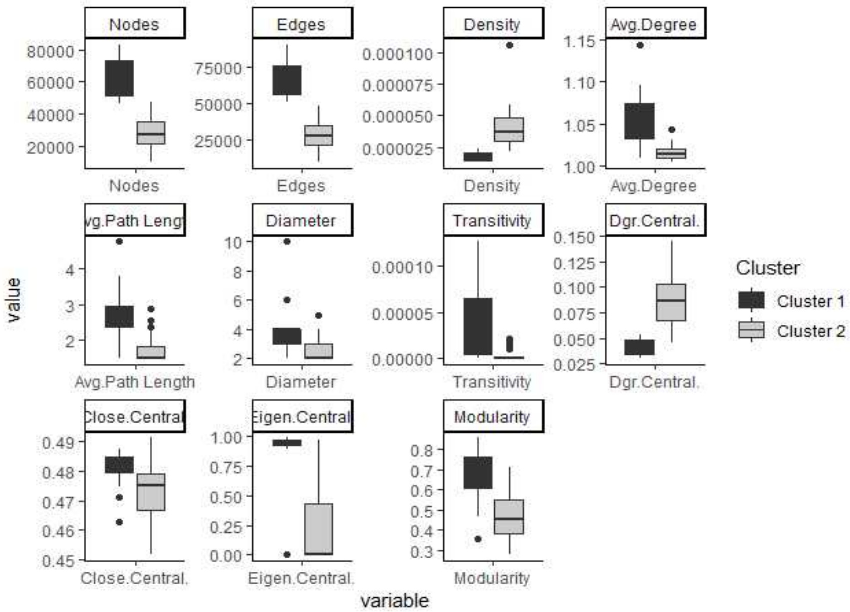
Combined the results of the series of analyses, the cluster analysis result presented in the dendogram(<Fig. 3>) shows that categorizing the sample companies into two clusters seems a reasonable solution. 17 companies and 23 companies are assigned into two clusters, respectively. The results show that most of the companies in the two sample groups(i.e., startup companies and established companies) gather together reasonably well except a small number of cases. The 17 companies(Cluster 1) are all startup companies with one exception. 'FT10' is a Fortune 100 company, and yet it hangs together with startup companies assigned to the Cluster 1. The 23 companies(Cluster 2) are mostly established companies except 4 startup companies ('ST09,' 'ST13,' 'ST15' and 'ST20). Despite these exceptional cases, the two clusters seem to represent the two respective sample groups reasonably well, and therefore, further discussions will be provided in the next section, especially focusing on the sources of the differences of the two clusters in terms of the two key aspects of Twitter engagement of the companies assigned to each cluster: i) the indicators of overall Twitter activities and ii) the metrics of social network structure of the Twitter follower-networks.

V. Discussion and Conclusion

With regard to the indicators of overall Twitter activities, as shown in <Table 3> and <Fig. 1.>, the Fortune companies have more followers, friends and status updates than the startup companies. This result seems quite obvious because the Fortune companies have a considerably longer history in business, and should have more resources for social media engagement. In contrast, the startup companies have higher Listed Ratio and Influencers Ratio than Fortune companies in the sample. These results may suggest that the startup companies tend to have proportionally tighter and more intimate relationships with their followers. The startup companies tend to have a proportionally larger number of followers who are more eager to get information from the companies that they are following. Also, in the sampled follower-networks, the startup companies tend to have proportionally a larger number of influencers. This may arise from the tendency for influencers to be more interested in the startup companies because of potential influences that the influencers may be able to exercise with the startup companies. Also, the startup companies are likely to be more eager to attract influencers and have made active efforts to attract them, hoping that influencers will facilitate improving and expanding Twitter engagement by the companies. Regardless of the reasons behind the results of higher Listed Ratio and Influencers Ratio, considering that attracting more influencers is one of key effective social media strategies especially for small and new startup companies, the sampled startup companies seem to have been relatively successful in their social media engagement efforts.

<Table 6> Tests for Sample Means: Social Network Structure of Two Clusters

	Cluster	Nodes	Edges	Density	Avg. Degree	Avg.Path. Length	Diameter	Transitivity	Dgr. Central.	Close. Central.	Eigen. Central.	Modularity
Mean (Scaled)	1 (N=17)	0.979308	0.988186	-0.74889	0.729234	0.754251	0.621748	0.578328	-0.86594	0.503907	0.796298	0.686092
	2 (N=23)	-0.72384	-0.7304	0.553526	-0.539	-0.55749	-0.45955	-0.42746	0.640045	-0.37245	-0.58857	-0.50711
Mean (Original)	1 (N=17)	62976.8	66366.5	0.000017	1.054412	2.745481	4.235294	0.000039	0.041052	0.481164	0.889349	0.661811
	2 (N=23)	28394	28900.6	0.000041	1.015565	1.741333	2.478261	0.000004	0.089038	0.473219	0.236384	0.467128
p value	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001	0.002	<0.001	<0.001



<Fig. 7> Boxplots: Social Network Structures of Two Clusters

To further examine the differences in the metrics of the social network structure of the Twitter follower-networks, the sample means of the two clusters are tested and the results are presented in <Table 6>. Given that the data is non-normal, the sample means across two clusters were tested using Kruskal-Wallis rank sum test(Campbell & Swinscow, 2009). Also, to make it easier to compare the results, a series of boxplots are prepared and presented in <Fig. 7>.

The cluster 1 and 2 are mostly represented by startup companies and established companies, respectively. One key finding from the analysis is that the cluster 1 has higher average path length and diameter, suggesting that the companies assigned to the cluster 1 tend to have wider networks in their sampled follower-networks. The cluster 1 also has higher transitivity, eigenvector centralization and modularity, showing that the companies in the cluster 1 tend to have several clusters and have dense and tight connections between followers within clusters. This finding looks consistent with the earlier finding that the startup companies have a proportionally larger number of influencers and have more intimate connection among the followers around influencers.

In contrast, the cluster 2 has higher density and higher degree centralization than the cluster 1, suggesting that the companies in the cluster 2 have users highly connected but connected around only a few users with high degree centrality.

The characteristics of the cluster 1 are expected to promote more effective and efficient information flow with potential customers and business partners, which startup companies definitely need for their success in their early stage of business. This finding has a useful implication for the future social media engagement efforts by startup companies in general. To better appreciate the potential benefits of social media for business success, startup companies may need to focus on getting more attention from influencers and promoting more cohesive communities in their follower-networks.

This study may contribute to both practice and research. This study examined the key indicators of Twitter engagements by startup companies and compare the indicators with those of established companies to facilitate more sensible interpretation of the indicators. The key findings and implications of this study may help startup companies to make more effective and efficient social media engagement efforts. In addition, although this study is an exploratory study, the findings of this study may also provide a useful framework for more rigorous confirmatory studies investigating Twitter engagement efforts by startup companies in the future.

One main limitation of this study stems from the access to the data for analysis. It has become increasingly difficult for academics to access social media data. Most social media platforms either block access data or limit data provided through their API. Twitter still provides its data via a number of API, and remains the most popular platform for academic research. However, working with Twitter data requires extensive resources, including time, computing power and money. This study was limited by such resources and had to compromise its approach, resulting in weaknesses in the validities of the key findings.

REFERENCE

Bansal, & Annika(2019). *10 Insightful Twitter Statistics for Small Business,* SmallBusinessSense, Retrived from https://small-bizsense.com/10-insightful-twitter-statistics-for-small-business/

Barnes, N. G., Allison, K., & Kyile, M.(2019). *The 2018 Fortune 500 Target Millennials and Seek Uncensored Expression*, Center for Marketing Research-UMass Dartmouth, Retrieved(2019/7/8) from https://www.umassd.edu/cmr/research/2018-fortune-500.html

Bruns, A., & Stieglitz, S.(2013). Towards more systematic Twitter analysis: Metrics for tweeting activities. *International Journal of Social Research Methodology*, 16, 91-108.

Campbell, M. J., & Swinscow, T. D. V.(2009). *Statistics at Square One, 11th ed.* Wiley-Blackwell: BMJ Books

Carley, K.(1991). A theory of group stability. *American Sociological Review,* 56, 331-354.

Clauset, A., Newman, M. E. J., & Moore, C.(2004), Finding community structure in very large networks, *Phyisical Review,* 70, 066111

Csardi, G., & Nepusz, T.(2006). The igraph software package for complex network research, I*nterJourmal, Complex Systems* 1695(5). http://igraph.org

Ediger, D., Jiang, K., Riedy, J., Bader, D. A., & Corley, C. (2010). *Massive social network analysis: Mining twitter for social good, In 2010 39th International Conference on Parallel Processing(583-593).* New York, NY: IEEE.

Freeman, L. C.(1979). Centrality in social networks conceptual clarification, *Social networks* 1(3), 215-239.

Gentry, Jeff, G.(2015). *twitteR: R Based Twitter Client. R package version 1.1.9. Retrieved(2019/3/15)* from https://CRAN.R-project.org/package=twitteR

Himelboim, I., & Han, J. Y.(2013). Cancer talk on Twitter: Patterns of information seeking in breast and prostate cancer networks. *Journal of Health Communication*, 19, 210-225.

Hansen, D. L., Shneiderman, B., & Smith, M. A.(2011). *Analyzing social media networks with NodeXL: Insights from a connected world,* Burlington, MA: Morgan Kaufmann.

Kassambara, A., & Mundt, F.(2017). *factoextra: Extract and Visualize the Results of Multivariate Data Analyses.* R package version 1.0.5., Retrieved(2019/7/25) from https://CRAN.R-project.org/package=factoextra

Kearney, M. W.(2019). *rtweet: Collecting Twitter Data. R package version 0.6.9,* Retrieved (2019/3/15) from https://cran.r-project.org/package=rtweet

Ketchen, Jr, D. J., & Shook, C. L.(1996). The application of cluster analysis in Strategic Management Research: An analysis and critique, *Strategic Management Journal,* 17(6), 441-458.

Kumar, R., Novak, J., & Tomkins, A.(2006). *Structure and evolution of online social networks. Proceedings of the 12th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining,* Philadelphia, PA.

Lerman, K., & Ghosh, R.(2010). Information contagion: An empirical study of the spread of news on Digg and Twitter social networks. I*CWSM,* 10, 90-97.

Maechler, M., Rousseeuw, P., Struyf, A., Hubert, M., & Hornik, K.(2019). c*luster: Cluster Analysis Basics and Extensions,* R package version 2.1.0.

Newman, M. E., & Girvan, M.(2004). Finding and evaluating community structure in networks. *Physical Review E*, 69(2), 026113.

Ranker(2019). *Fortune 100 Companies on Twitter: Corporate Twitter Accounts, Retrieved (2019/6/15)* from https://www.ranker.com/list/fortune-100-companies-on-twitter-corporate-twitter-accounts/business-and-company-info

Rousseeuw, P. J.(1987). Silhouettes: A graphical aid to the interpretation and validation of cluster analysis, *J. Comput. Appl. Math.*, 20, 53-65.

Sajid, Arsalan(2019). *60 Best Startups You Need to Watch Out for in 2019, Retrieved (2019/6/10)* from https://www.cloudways.com/blog/best-startups-watch-out/

Ward, J. H., Jr.(1963). Hierarchical Grouping to Optimize an Objective Function, *Journal of the American Statistical Association*, 58, 236-244.

Wasserman, S., & Faust, K.(1994). *Social network analysis: Methods and applications,* New York, NY: Cambridge University Press.

Yamane, T.(1967). *Statistics: An Introductory Analysis, 2nd Edition*, New York: Harper and Row.

소셜네트워크 분석과 클러스터 분석 방법을 활용한 스타트업 회사의 트위터 팔로워 네트워크에 대한 탐색적 연구

류승희 (세종대학교 경영대학 부교수)*

국 문 요 약

기업의 소셜미디어 활용이 빠른 속도로 증가함에 따라 성공적인 소셜미디어 활용전략의 중요성이 커지고 있다. 이러한 중요성은 새로이 시장에 진입하여 신속하게 시장에서의 인지도를 확대하고 미래고객을 확보해야 할 필요성이 큰 스타트업 회사에게 더욱 절실하다고 할 수 있다. 본 연구의 목적은 스타트업 회사의 소셜미디어 활용의 특징을 보여주는 지표를 탐색적으로 조사, 분석하는데 두고 있다. 주요 지표는 전반적인 소셜미디어 관련 활동을 보여주는 지표와 소셜미디어 서비스을 통해 형성된 소셜네트워크 구조의 특성과 관련 지표를 포함한다. 스타트업 회사의 이러한 지표를 좀 더 객관적으로 평가하기 위하여 잘 갖춰진 기존 회사의 지표와 비교, 분석 하였다.

본 연구를 위해 여러 소셜미디어 서비스 중 트위터를 선정하고, 트위터 REST API를 통해 측정지표와 관련된 데이터와 팔로워네트워크(follower-network)에 대한 데이터를 수집하였다. 주요 분석방법으로 각 회사의 소셜네트워크 구조의 특성을 분석하기 위해 소셜네트워크분석 기법이 활용되었으며, 클러스터분석 기법을 이용하여 스타트업 회사와 기존 회사의 측정지표를 비교, 분석하였다. 분석결과에 따르면 대부분의 측정지표에서 스타트업 회사와 기존 회사 간에 유의미한 차이를 보여주고 있다. 특징적인 분석결과의 하나로 스타트업 회사들이 상대적으로 많은 수의 인플루언서 (influencer)를 팔로워네트워크에 가지고 있다는 점이다. 또한, 스타트업 회사를 포함하는 클러스터의 네트워크 모듈성(modularity)과 추이성(transitivity)이 기존 회사에 비해 상대적으로 높은 것으로 나타났다. 이러한 결과는 스타트업 회사의 소셜네트워크 안에 기존 회사에 비해 내부결속력이 높은 상대적으로 많은 수의 커뮤니티가 존재한다는 점을 시사한다고 할 수 있다. 스타트업 회사의 이러한 특징은 잠재고객 및 비즈니스 파트너와의 효과적인 정보교환을 촉진할 수 있으며, 따라서 향후 일반적인 스타트업 회사의 소셜미디어 노력은 어떻게 인플루언서를 확보할 것인지, 또한 어떻게 내부결속력이 높은 긴밀한 네트워크를 구축할 것인지에 초점을 두어야 할 필요성이 있음을 시사하고 있다.

핵심주제어: 소셜네트워크서비스, 트위터 팔로워네트워크, 소셜네트워크분석, 클러스터분석

* 세종대학교 경영대학 부교수, shyu@sejong.edu