

Social Network Analysis and Visualization with R

Kim Gye-Soo*

Professor, Department of Business Administration And Big Data Center, Semyung University, Korea
215 Social Science Building, 65 Semyung-Ro Jecheon, Chungbuk 390-711, Korea
E-mail: gskim@semyung.ac.kr

Abstract: In industrial revolution 4.0 period, social network Service (SNS) usage is rising. In this circumstance, Social network analysis has emerged as a key technique in modern sociology. It has also gained a significant following in anthropology, biology, communication studies, business management, economics, geography, history, information science, organizational studies, political science, social psychology, development studies, sociolinguistics, and computer science and is now commonly available as a consumer tool.

SNA is a branch of Network Science, which is an attempt to understand networks emerging in nature, technology and society using R. Researcher can find hub and node for practical application. And word cloud is the representative tool for visualization of communication. The resulting networks, which can contain thousands of nodes, are then analyzed by using R from network theory to identify the key actors, the key communities or parties, and general properties such as robustness or structural stability of the overall network, or centrality of certain nodes.

Key words.: *Social Network Service, R, Word Cloud, Hub, Node, Link.*

I. INTRODUCTION

Businesses are collecting more data than they know what to do with. To turn all this information into competitive gold, they'll need new skills and a new management style. Networks are everywhere, from the Internet, to social networks, and the genetic networks that determine our biological existence. Barabási, Albert-László (2002) showed that a Web-based view of nature, society, and business, a new framework for understanding issues ranging from democracy on the Web to the vulnerability of the Internet and the spread of deadly viruses. Networks are present everywhere. All we need is an eye for them. Network theory can have some useful actuarial and business applications. Network is related to health care service providers and evaluating how failures at critical companies can propagate through the economy and potentially impact your firm's investment returns. Other applications might include the determination of key personnel at your organization or attempting to target a marketing campaign preferentially at the social hubs of a community.

Organization needs data scientist for analysis data and developing strategy. Data scientists want to build things, not just give advice. Data scientists are the people who understand how to fish out answers to important business questions from today's tsunami of unstructured information (Davenport and Patil, 2012). As data become cheaper, the complements to data become more valuable. Some of the most crucial of these are data scientists and other professionals skilled at working with large quantities of information. Statistics are important, but many of the key techniques for using big data are rarely taught in traditional statistics courses. Perhaps even more important are skills in cleaning and organizing large data sets; the new kinds of data rarely come in structured formats. Visualization tools and techniques are also increasing in value.

Data collection for such efforts can be problematic, but in the absence of a formal evaluation a network approach can still be of use qualitatively. All are networks, and all are a part of a surprising scientific revolution. Networks are everywhere. Networks exhibit interesting phenomenon. Networks analysis is useful for decision making. Networks will often have few hubs with a lot of connections and a much larger number of nodes with few connections. This type of structure provides resiliency against indiscriminate attacks on networks, as the bulk of the nodes targeted in such an attack have few connections, and the rest of the network can continue to function without them. Attacks targeted at hubs, however, can cripple a network by removing a small number of nodes from the network. This effect has been seen in denial-of-service attacks on websites.

This paper has three goals. The first goal of our research was to introduce basic concept and word cloud of social network analysis and word cloud method using R to readers. A second goal of our work was to develop generalized insight into analyses that wore informative and actionable for practitioners.

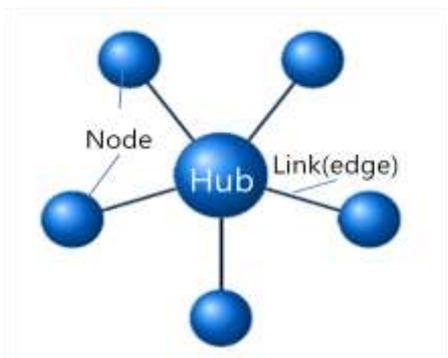
II. Social Network Analysis and Word Cloud Social Network Analysis and R package

All over the world is linked each other. A Social network analysis (SNA) can be yield sufficient benefit to justify the investment of time and energy on part of the organization and individuals. Cultural anthropologists independently invented the notion of social network to provide a new way to think about social structure and the concepts of role and position an approach that culminated in rigorous algebraic treatment of kinship system (White, 1963).

It is about how networks emerge, what they look like, and

how they evolve. It shows you a Web-based view of nature, society, and business, a new framework for understanding issues ranging from democracy on the Web to the vulnerability of the Internet and the spread of deadly viruses. Networks are present everywhere (Barabási, Albert-László, 2002). What Barabasi and his colleagues found were that there were particular points in networks that had high levels of connections—they seemed to be the “connectors” for smaller points in the web. These connectors they came to see as hubs in their networks with the less connected points being the nodes in the network, and Barabasi uses examples from the Web and popular culture to help describe and clarify this concept.

Visualizations can be hard to interpret for large networks. Important features will often be impossible to see. Common features (Small worlds, Hubs, and Clustering) include. Small worlds are small-world networks grow the path length between vertices grows very slowly. Hubs are Networks with hubs have a small number of vertices with a large number of connections. Clustering is friends are often also friends with each other. To measure and interpret these structural properties along with some more complex centrality measures for finding important vertices. The following figure is the basic structure. Social network composed node, link (edge) and Hub.



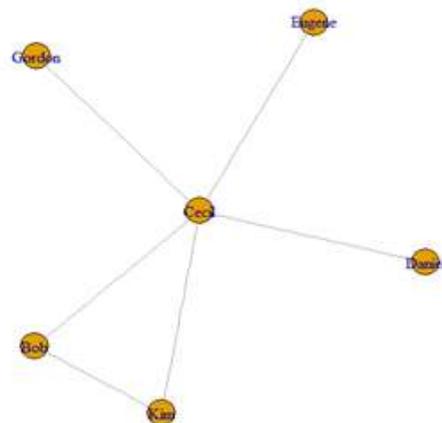
[Figure 1] Node, Hub, Link

Analyst of SNA can have unique effects. The first is promoting effective collaboration within a strategically important group. The second is supporting junctures in networks that cross function, and hierarchical boundaries. The third is finding the target market and point for marketing. In fact, SNA can be a very useful means of assessing the impact of strategic restructuring initiatives on the informal structure of an organization (Cross, 2002).

Managers need transparent methods for using the new models and algorithms on a daily basis. By necessity, terabytes of data and sophisticated modeling are required to sharpen marketing, risk management, and operations (Barton & Court, 2012).

R is a language and environment for statistical computing and graphics. It is a GNU project which is similar to the S

language and environment which was developed at Bell Laboratories (formerly AT&T, now Lucent Technologies) by John Chambers and colleagues. R can be considered as a different implementation of S. There are some important differences, but much code written for S runs unaltered under R. R and R studio provides a wide variety of statistical (linear and nonlinear modeling, classical statistical tests, time-series analysis, classification, clustering, ...) and graphical techniques, and is highly extensible. The S language is often the vehicle of choice for research in statistical methodology, and R provides an Open Source route to participation in that activity (<https://www.r-project.org/about.html>). In order to work r package, Analyst should down and load r and rstudio program. Script for Social Network Analysis is the following. After insert the script, analyst can push the Run button. And analyst can get the following figures.



[Figure 2] Script and result for SNA

Analyst can find hub and node in figure. And he/she can find strategic point for accomplish the task or goal. In [Figure 2] we can find that Cecil is hub and centrality of community. In this circumstance, marketer can determine how contact to

decision making hub and get effective result.

Word cloud

A word cloud (tag cloud, or weighted list in visual design) is a visual representation of text data. Tags are usually single words, and the importance of each tag is shown with font size or color. This format is useful for quickly perceiving the most prominent terms and for locating a term alphabetically to determine its relative prominence. Word clouds add simplicity and clarity. The most used keywords stand out better in a word cloud. Word clouds are a potent communication tool. They are easy to understand, to be shared and are impactful Word clouds are visually engaging than a table data.

The following example is the forecasting reducing job of the 4.0 industry revolution effect from Korea Press Foundation (2017). The employment impact of disruptive change have often been polarized between those who foresee limitless opportunities in newly emerging job categories and prospects that improve workers’ productivity and liberate them from routine work, and those that foresee massive labor substitution and displacement of jobs. Academics, chief executives and labor leaders hold strong and diverse views on the debate, as do policymakers.

Now, analyst can make syntax script for World Crowd for visualization like the following.

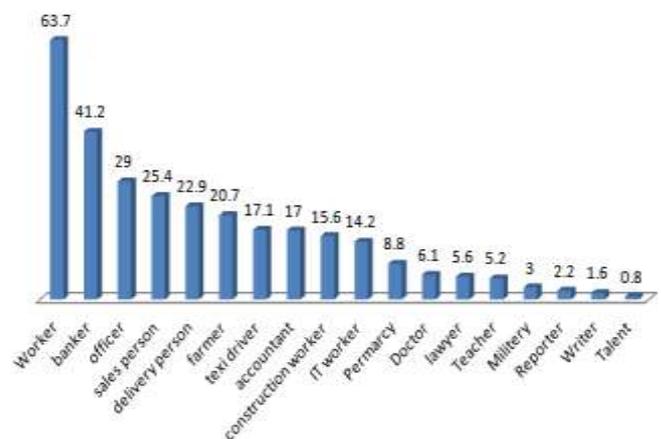
<Table2> Script for word cloud

```
library("wordcloud")
library("RColorBrewer")
d <-read.csv("D:/data/words1.csv", header=T)
head(d, 18)
set.seed(1234)
wordcloud(words = d$word, freq = d$freq, min.freq = 1,
           max.words=200, random.order=FALSE,
           rot.per=0.35,
           colors=brewer.pal(8, "Dark2"))
head(d, 18)
barplot(d[1:18,]$freq, las = 2, names.arg =
d[1:18,]$word,
        col = "lightblue", main = "Most frequent
words",
        ylab = "Word frequencies")
```

And we can find the result of word crowd in Rstudio package.



[Figure 3] Script for word cloud



[Figure 4] Barchart result

<Table 1> Reducing job for the Fourth Industrial Revolution

	word	freq
Worker	Worker	63.7
banker	banker	41.2
officer	officer	29
sales person	sales person	25.4
delivery person	delivery person	22.9
farmer	farmer	20.7
taxi driver	taxi driver	17.1
accountant	accountant	17
construction worker	construction worker	15.6
IT worker	IT worker	14.2
Permarcy	Permarcy	8.8
Doctor	Doctor	6.1
lawyer	lawyer	5.6
Teacher	Teacher	5.2
Military	Military	3
Reporter	Reporter	2.2
Writer	Writer	1.6
Talent	Talent	0.8

III. Conclusion

In today's fast-paced and turbulent industry revolution 4.0, work of importance is increasingly accomplished collaboratively through informal networks. As a result, assessing and supporting strategically important network in organizations can yield substantial performance benefits. Exploiting vast new flows of information can radically improve your company's performance. But first you'll have to change your decision-making culture (McAfee and Brynjolfsson, 2012).

The procedure of creating SNA and word cloud is very simple in R if you know the different steps to execute. Analyst can get strategic insight for strategy and plan. And he/she can present research result with impactful effect.

Data-driven decisions tend to be better decisions. In Korea NC Dinos Professional baseball team recruits foreign players based on big data analysis and multi source feedback. These methods have showed good result comparing competitors. Leaders will either embrace this fact or be replaced by others who do. In sector after sector, companies that figure out how to combine domain expertise with data science will pull away from their rivals. We can't say that all the winners will be harnessing big data to transform decision making. SNA and Word cloud methods allow us to highlight the most frequently used key person and keywords in a paragraph of texts. One can create social network and a word cloud, also referred as text cloud or tag cloud, which is a visual representation of text data and organization. The era of big data is evolving rapidly, and our experience suggests that most companies should act now. Executives should concentrate target efforts to source data, build model, and transform the organization culture.

REFERENCES

- [1] Barabási, Albert-László (2002), *Linked: The New Science of Networks*. Perseus Books Group.
- [2] Barton, D., Court, D. (2012), *Making Advanced Analytics Work for You*, Harvard Business Review, October, pp. 3-7.
- [3] Cross, R., Borgatti, S. P., Parker, A. (2002), *Making Invisible Work*.
- [4] Davenport T. H., Pati, D.J. (2012), *Data Scientist: The Sexiest Job of the 21st Century*, Harvard Business Review, October, pp. 70-76.
- [5] Lusseau, D. K. Schneider, O. J. Boisseau, P. Haase, E. Sloaten, and S. M. Dawson (2003), *Behavioral Ecology and Sociobiology* 54, 396-405.
- [6] <https://www.r-project.org/about.html>
- [7] Korea Press Foundation (2017).
- [8] McAfee, A., Brynjolfsson, E. (2012), *Big Data: The Management Revolution*, Harvard Business Review, October, pp. 59-68.
- [9] White, H. C. (1963), *An Anatomy of Kinship*, Englewood Cliffs, NJ: Prentice-Hall.