The Eighteenth International Symposium on Artificial Life and Robotics 2013 (AROB 18th '13),
Daejeon Convention Center, Daejeon, Korea, January 30-February 1, 2013

A Procedure for Constructing Social Network Using Web Search Engines: The Case for Japanese Automotive Industry

Yuichi Kubota[1], Reiji Suzuki[1] and Takaya Arita[1]

[1] Graduate School of Information Science, Nagoya University, Japan
kubota@alife.cs.is.nagoya-u.ac.jp, {reiji, arita}@nagoya-u.jp

Abstract: Recently, search engines have enabled us to access immense quantities of useful information in an instant. In this paper, we propose a procedure for analyzing the social relationship and structure using Web search engines, which includes novel ways to create a search query and to use the number of hits. This allows us to construct various networks that reflect directed and undirected relationships among actors under arbitrary contexts. As a case study for evaluations of the proposed procedure, we focus on 50 companies belonging to automotive industry in Japan. We constructed several directed and undirected networks under different temporal and geographical contexts. We show that we can obtain more general knowledge about this industrial community from the analyses of these created networks and their centrality measures.

Keywords: Social Network, Web Mining, Visualization

1 INTRODUCTION

Characteristics of social networks are a valuable piece of information for understanding social activities and communications. However, it had been very costly, or even impossible for non-specialists, to obtain large-scale data for measuring the strength of the relationship among many actors (e.g., persons, companies) to create their social network, which had made such an approach only applicable to limited and well-organized data.

On the other hand, the progress in the field of Web technologies have enabled us to access immense quantities of useful information for creating social networks in an instant. Especially, various methodologies for extracting social networks using search engines have been proposed [1].

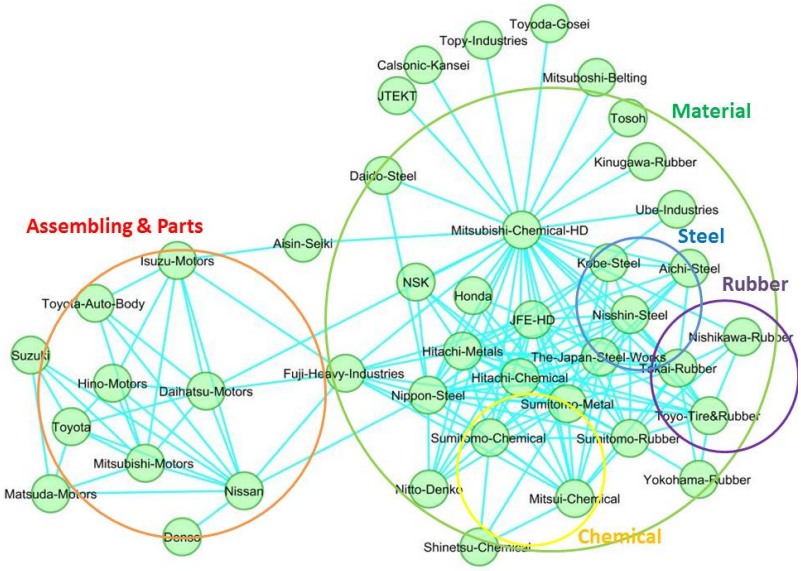
Lee et al. proposed a method for social network analyses based on the statistical physics using the number of Google search hits in order to estimate the relatedness between two actors [2]. They introduced a general framework for measuring the disparity or heterogeneity of weights a node bears. They constructed a social network of the 109th US Senate members, and successfully showed the division or community structure, reasonably consistent with the senators' political parties, and so on. Recently, Akaishi et al. proposed a method to analyze the temporal change in a social network every year by adding the enumeration of years to a query [3]. They observed the temporal dynamics of the social network composed of the 93 people who have played major roles in the US economy, showing its relational changes surrounding the economic crisis in 2008. They also analyzed the interdisciplinarity of researchers and research topics by constructing a bipartite graph using queries composed of topics and researchers, using a new quantity named *visibility boost* for the calculation of relatedness [4]. These studies clearly show that we can observe various aspects of a social relationship by constructing networks using different ways to use search queries and the search results, although each previous study mainly focused on a specific topic in detail. We believe that such a multifaceted approach gives us valuable insights into understanding of essential properties of social networks.

From this viewpoint, in this paper, we propose a simple procedure for analyzing social relationships and structures using Web search engines, which includes novel ways to create a search query and to use the number of hits. This allows us to construct various networks that reflect directed and undirected relationships among actors under arbitrary contexts. As a case study for evaluations of this procedure, we focus on 50 companies belonging to automotive industry in Japan. We constructed and analyzed several networks using indices for measuring directed and undirected relatedness under different temporal and geographical contexts.

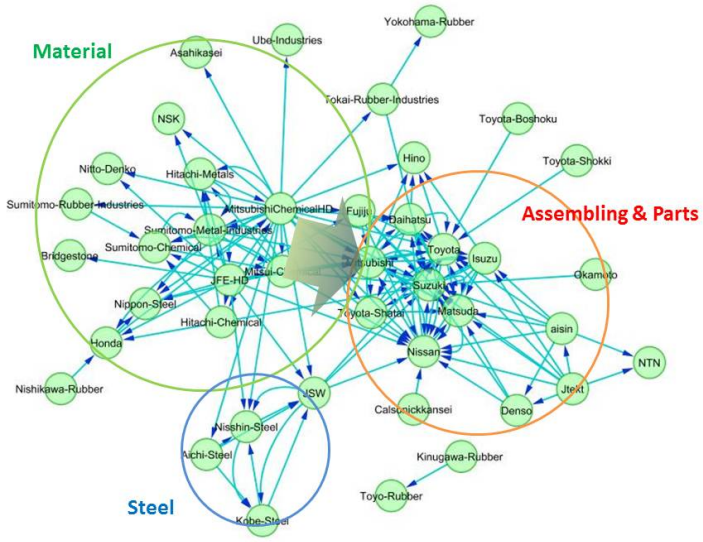
2 METHOD

Generally, the number of hits obtained by a search query consisting two keywords reflects the co-occurrence frequency of these words in the Web. We can approximate the strength of the relationship between actors (e.g. companies, researchers) by using this property of search engines.

If X and Y are the sets of the web pages obtained by search queries for "x y", respectively, we can regard statistical metrics for the relationship between these two sets as the strength of the relatedness between the actors x and y. As for undirected relationships between actors, we use a standard Jaccard coefficient: $\frac{h(X \cap Y)}{h(X \cup Y)}$, where $h(X \cap Y)$ and $h(X \cup Y)$ are the number of hits obtained by the search query for "x y" and "x OR y" (using the Google search engine), respectively. In order to measure the asymmetric relationship between two



(a) The undirected network based on the Jaccard coefficient.

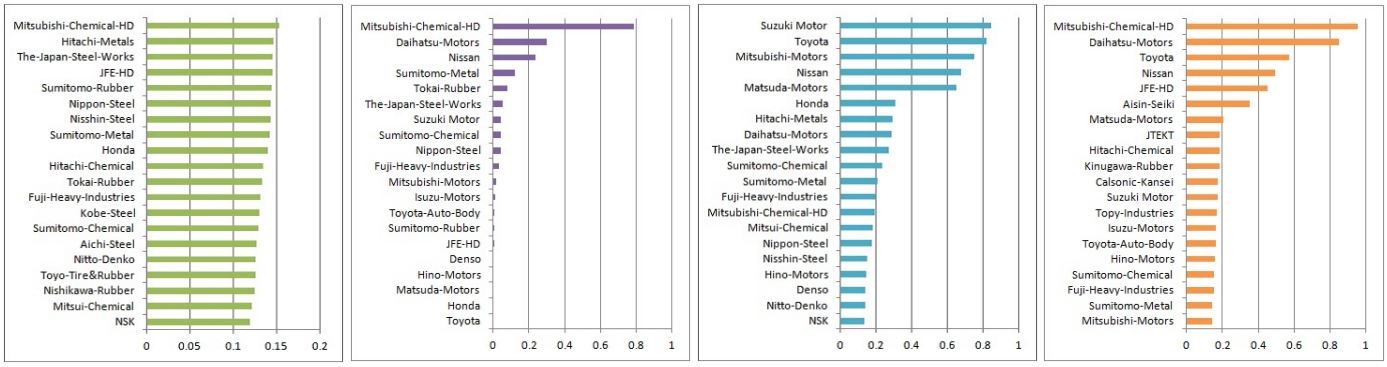


(b) The directed network based on the directed Simpson coefficient.

Fig. 1. The social networks of automotive industry in Japan, constructed by using (a) the Jaccard coefficient and (b) the Simpson coefficient.

Table 1. The 50 companies of automotive industry in Japan.

Toyota, Honda, Nissan, Matsuda Motors, Suzuki Motor Corporation, Daihatsu Motors, Mitsubishi Motors, Fuji Heavy Industries, Isuzu Motors, Hino Motors, Denso, Aisin Seiki, Toyota Auto Body, Toyota Industries, Toyota Boshoku, JTEKT, Calsonic Kansei, NSK Ltd., Toyoda Gosei, NTN Corporation, Bridgestone, Sumitomo Rubber, Yokohama Rubber, Toyo Tire and Rubber, Tokai Rubber, Bando Chemical, Okamoto, Kinugawa Rubber, Nishikawa Rubber, Mitsuboshi Belting, Nippon Steel, JFE Holdings, Kobe Steel, Sumitomo Metal, Nisshin Steel, Hitachi Metals, Daido Steel, The Japan Steel Works, Topy Industries, Aichi Steel, Mitsubishi Chemical HD, Sumitomo Chemical, Mitsui Chemical, Shinetsu Chemical, DIC Corporation, Tosoh, Asahikasei, Nitto Denko, Ube Industries, Hitachi Chemical



(a) Closeness (b) Betweenness (c) In-degree (d) Out-degree

Fig. 2. Network centralities of the top 20 companies. (a) and (b) are the closeness and betweenness in the undirected network. (c) and (d) are in-degree and out-degree in the directed network.

actors, we introduce a new index named *directed Simpson coefficient*: $\frac{h(X \cap Y)}{h(X)}$, where h(X) is the number of hits obtained by the search query for "x", which reflects the relatedness of the actor y for the actor x. Thus, this coefficient allows us to measure the asymmetric relationship between two actors. As far as we know, the directed Simpson coefficient has not been used in the previous research despite of its simplicity.

In addition, we can calculate the relatedness under an arbitrary context if we measure these indices by adding a specific word related to the context to the all queries for the calculation of relatedness. For example, we use the query "x y z" instead of "x y" in order to obtain $h(X \cap Y)$ under the context of "z". In this paper, we use the name of places in order to analyze the geographical variation of the networks as well as the years to analyze the temporal variation, which has also not been used in the previous research for this purpose, as far as we know.

We construct a network that reflects their social relationship using these indices. Each actor is represented as a node, and the strength of relationship between two actors is represented as the width of the link between their corresponding nodes. Note that two nodes are connected only when the strength between them is equal to or higher than a threshold T. We visualized a network using Cytoscape[1] and adopted a spring model in which the higher relatedness between nodes pull them together, in order to arrange the nodes in a two-dimensional plane. We can examine the topological properties of the network by measuring various metrics used in complex network studies such as the closeness (the average path length from the focal node to the other nodes), the betweenness (the proportion of shortest paths from all nodes to all others that pass through the focal node), and the degree (the number of edges which a node has).

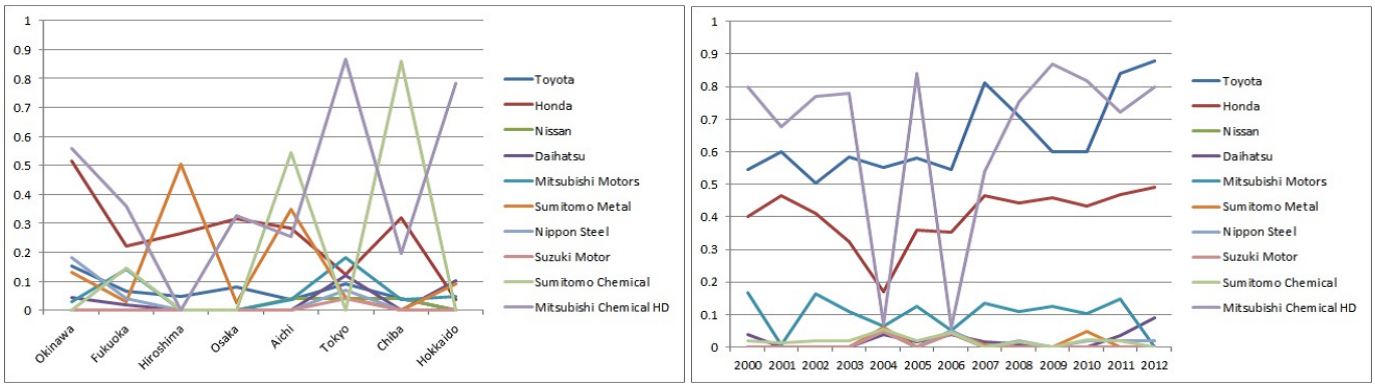
[1]Cytoscape: http://www.cytoscape.org/

3 MULTIFACETED ANALYSES

As a case study for evaluations of the proposed procedure, we constructed social networks of 50 companies, whose names are listed in Table 1, belonging to automotive industry (assembling, material and part makers) in Japan. We used Google Search API for web searches. We also used the setting T = 0.1(Jaccard) and 0.4(directed Simpson).

Fig. 1 (a) and (b) show the constructed networks using the Jaccard coefficient and the directed Simpson coefficient, respectively. In the undirected network (Fig. 1 (a)), we see the two major clusters composed of material companies (including steel, chemical and rubber companies) and parts & assembling companies, respectively. The Jaccard coefficient tends to become large when h(X) and h(Y) are close if h(X OR Y) is large. Because the number of hits obtained by a search query for a company is expected to reflect the scale of the company's business, we can expect that mutually related companies with the more similar scale of business tended to be clustered in the network in Fig. (a). When we used the directed Simpson coefficient (Fig. 2 (b)), we observed the "flows" of the relationships from the former cluster to the latter, implying the significance of the latter companies. We also see that the assembling makers (e.g. Toyota, Nissan) have more input links than output links. This implies that it is likely that the assembling maker is socially depended on by other suppliers in automotive industry.

We calculated the several centrality measures of each node in these networks, and listed the top nodes, as shown in Fig. 2. In this paper, we focused on the closeness and the betweenness in the undirected network (Fig. 2 (a) and (b)). It should be noticed that both centrality measures of a major chemical company (Mitsubishi chemical holdings) tended to be higher than others'. This seems to reflect that this company has wide relationship with other companies in this field. As for the directed network, we focused on the in-degree and



(a) The geographical variation (b) The temporal variation

Fig. 3. The betweenness of each node in the directed networks under the different (a) geographical contexts and (b) temporal contexts.

out-degree (Fig. 2 (c) and (d)). We found that major car manufacturers (e.g. Toyota, Nissan) tended to have both high in-degree and high out-degree, which is expected to be due to their high importance in this industry.

We also constructed different directed networks using the different temporal (from 2000 to 2012) and geographical (from Okinawa to Hokkaido) contexts. Note that these years and names of places themselves were used for additional keywords for a query. The keywords except the focal context are also used to limit search results to a particular context (e.g., "Toyota Nissan Tokyo -Osaka -Aichi, -···" for calculating the relatedness between Toyota and Nissan under the geographical context of Tokyo). Then, we calculated the betweenness of each node in the all networks as shown in Fig. 3. We have chosen some major companies in order to observe the general tendency of the relational variation under these contexts.

We see from Fig 3 (a) that there were significant differences in the betweenness of each company among the networks in different geographical regions. This might to be due to the large difference in the industrial structure or the size of the market in these regions. Fig 3 (b) also shows the temporal changes in the betweenness of each company through the years, implying that some major companies played important roles in turns through this period.

4 CONCLUSION

We have proposed a simple procedure for analyzing social relationships and structures using Web search engines, which includes novel ways to create a search query and to use the number of hits, in order to understand them from a multifaceted view. As a case study, we constructed several networks of companies in automotive industry in Japan, using different indices for measuring directed and undirected relatedness under different temporal and geographical contexts. The analyses clarified not only the overall topological properties of the social network, but also the existence of its geographical and temporal variations. We believe that the proposed procedure enables many people including non-specialists to understand social relationships and structure between individuals, companies or countries through constructing and analyzing the networks by making use of search engines.

ACKNOWLEDGEMENTS

The authors thank Dr. Jin Akaishi for technical advices for web searches using the Google search engine.

REFERENCES

[1] Matsuo, Y., Mori, J., Hamasaki, M., Nishimura, T., Takeda H., Hashida, K. and Ishizuka, M.: POLYPHONET: An advanced social network extraction system from the Web, *Web Semantics*, 5: 262-278 (2007).

[2] Lee, S. H., Kim P. J., Ahn Y. Y. and Jeong, H., Googling social interactions: Web search engine based social network construction, *PLoS ONE*, 5(7): e11233.

[3] Akaishi, J., Sayama, H., Dionne, S. D., Chen, X., Gupta, A., Hao, C., Serban, A., Bush, B. J., Head, H. J. and Yammarino, F. J.: Reconstructing history of social network evolution using web search engines, *Proceedings of BIONETICS 2010*, 155-162 (2010).

[4] Sayama, H. and Akaishi, J.: Characterizing interdisciplinarity of researchers and research topics using web search engines, *PLoS ONE*, 7(6): e38747 (2012).