Semiconductor Optoelectronics
ISSN: 1001-5868

SOUTH KOREA'S ECONOMY: A BIG DATA ANALYSIS

Namkil Kang
Far East University, South Korea

The ultimate goal of this paper is to provide an in-depth analysis of 35 articles of Google written from 2020 to 2023 concerning South Korea's economy. A point to note is that the frequency of common nouns is 1,474, whereas that of proper nouns is 634. A further point to note is that in the word cloud of 35 articles, the keyword *South Korea* is expressed as the biggest. It is deemed to be the most central one. It is worthwhile pointing out, on the other hand, that the word *economy* is expressed as the second biggest, which is considered to be the second most central one. When it comes to 16 topics, it is significant to note that topic 13 was the most occurred one, followed by topic 8, topic 4, and topic 9 (topic 10 and topic 16), in that order. With respect to the use of key words, it is worth noting that the key word *South Korea* was the most occurred one, followed by the key word *economy*, the key word *country*, the key word *growth*, the key word *year*, and the key word *export*, in descending order. This paper argues that the key word *South Korea* obtains the highest degree centrality. Simply put, the centrality of *South Korea* is the highest since it has the most linked neighbors. This paper further argues that the key word *South Korea* obtains the highest closeness centrality since it has the shortest links. Finally, this paper further shows that the key word *South Korea* obtains the highest betweenness centrality since this word acts a bridge among many words that are linked.
Keywords: NetMiner, topic, keyword, word cloud, map, centrality

1. Introduction

The main goal of this paper is to provide an in-depth analysis of 35 articles of Google written from 2020 to 2023 concerning South Korea's economy. For this goal, we used the software package NetMiner. First, we aim to look into the proportion and cumulative proportion of common and proper nouns. Second, we aim at considering the properties of networks consisting of 35 articles. More specifically, we attempt to investigate the number of links, their density, and average degree. Also, we try to investigate inclusiveness and the number of weak components and strong components. Also, attention is paid to reciprocity (Arc and Dyad) and transitivity. We also inquire into the properties of networks such as clustering coefficient, the mean distance of nodes, and the number of isolated nodes. Third, we aim at investigating the word cloud of 35 articles. In the word cloud, important nouns in 35 articles are provided as bigger words. Fourth, we try to consider 16 topics and their keywords that form 35 articles of Google. Also, we provide the use and frequency of 16 topics in 35 articles of Google. Fifth, we aim at considering the use and the frequency of important and key words. Sixth, we are chiefly concerned with degree centrality and provide its map. The degree centrality refers to the so-called local centrality. Put differently, it indicates the number of directly linked neighbors.

Also, we aim to consider closeness centrality and provide its map. The closeness centrality indicates that if the distance between nodes is close, the nodes count as important. Finally, we aim to investigate betweenness centrality and provide its map. The betweenness centrality indicates acting as a bridge. If nodes that act as a bridge disappear, the networks will be heavily influenced. In the three notions, centrality refers to the so-called importance and influence and nodes that have high centrality count as important and prominent.

2. Results

2.1. Network properties

This section is devoted to showing the proportion and cumulative proportion of common and proper nouns. Also, this section centers on examining the properties of networks. Table 1 shows the use and frequency of common and proper nouns in 35 articles:

Table 1 Common and proper nouns

Value	Frequency	Proportion	Cumulative Proportion
Common Noun	1474	0.699	0.699
Proper Noun	634	0.301	1
Total	2108	1	

When it comes to common nouns, their frequency is 1,474 and their proportion is 0.699 (They account for 69.9%). It is worth mentioning that the frequency of proper nouns is 634 and that their proportion is 0.301 (They account for 30.1%). Now we turn our query to network properties:

Table 2 Network properties

	# of Links : O(m)	Density : O(m)	Average Degree : O(m)
Word Network (Sentences)	3,455	0.002	3.727

Note that NetMiner produces networks by linking nouns to the relevant nouns. It is worth noting that the number of links that were used in 35 articles is 3,455. Notice that the density of networks is 0.002. With respect to average degree, it is interesting to note that it is 3.727. The term degree indicates the number of directly linked neighbors. Thus, that average degree is 3.727 means that a node is associated with 3.7 neighbors.

Now let us consider the number of components and inclusiveness:

Table 3 Network properties

	# of Components(Weak) : O(m)	# of Components(Strong) : O(m)	Inclusiveness : O(m)
Word Network (Sentences)	53	53	1

The term inclusiveness indicates the proportion of linked nodes. As can be seen from Table 3, the proportion of linked nodes is 1, which in turn indicates 100%. On the other hand, the term component indicates a set of nodes that are linked. In this case, nodes are linked with one another (links are not stopped). The weak components mean that there is only one link between two nodes. On the other hand, the strong components mean that there is an exchange of links between two nodes. More specifically, node X gives a link to node Y and the former receives a link from the latter. Quite interestingly, the number of strong and weak components is the same, namely 53. Simply put, 53 strong components and weak components do in fact exist in the whole network. A component refers to a group. Thus, in each case, 53 groups do in fact exist in the whole network.

Now let us consider reciprocity and transitivity:

Table 4 Network properties

	Reciprocity(Arc) : O(m)	Reciprocity(Dyad) : O(m)	Transitivity : O(nm)
Word Network (Sentences)	1	1	0.042

The term reciprocity (Dyad) indicates the proportion of mutually linked pairs. As indicated in Table 4, it accounts for 100%. The term reciprocity (Arc) indicates the proportion of reciprocal links. As illustrated in Table 4, it accounts for 100%. When it comes to transitivity, things are different. It happens among 3 nodes. If there is a link between X and Y and there is a link between Y and Z, then there is a link between X and Z. As exemplified in Table 4, transitivity in the whole network accounts for 4.2%.

Now let us observe the following network properties:

Table 5 Network properties

	Clustering Coefficient : O(n^3)	Mean Distance : O(nm)	Diameter : O(nm)	# of Isolated Nodes : O(m)

Word Network (Sentences)	0.058	4.544	14	0

The term clustering coefficient indicates the probability value of associating the neighbors of a node with one another. The clustering coefficient value is between 0 and 1. As indicated in Table 5, the clustering coefficient value is 0.058. This in turn indicates a low level of cohesion. The term mean distance indicates Average Path Length. It indicates the average distance among nodes. As exemplified in Table 5, the mean distance is 4.544. On the other hand, the term diameter indicates the furthest distance between two nodes. As shown in Table 5, the diameter of two nodes is 14. Finally, it is interesting to observe that there are no isolated nodes in the whole network.

2.2. A word cloud

In this section, we explore the word cloud of 35 articles. In this word cloud, the so-called keywords are expressed as bigger words. Figure 1 shows the word cloud of 35 articles with respect to South Korea's economy:

Figure 1 Word cloud

It is significant to note that the keyword *South Korea* is expressed as the biggest, which is deemed to be the most central one. It is also worth pointing out that the word *economy* is expressed as the second biggest, which is considered to be the second most central one. This also indicates that this word is the second most occurred one. It must also be stressed that the keyword *country* is expressed as the third biggest, which counts as the third most widely used one. Noteworthy is that the word *growth* is expressed as the fourth biggest, which is regarded as the fourth most central one. It should be pointed out, on the other hand, that the word *export* is expressed as the sixth biggest, which in turn suggests that it is the sixth most pivotal one. We thus conclude that the keyword *South Korea* is the most widely used one in 35 articles of Google.

2.3. Topics

In this section, we discuss 16 topics and their keywords that constitute 35 articles. Those keywords frequently occurred in 35 articles of Google:

Table 6 Topics

	1st Keyword	2nd Keyword	3rd Keyword	4th Keyword	5th Keyword
Topic-1	export	product	semiconductor	number	trade
Topic-2	percent	South Koreas	US	level	inflation
Topic-3	rate	Korea	OECD	average	interest
Topic-4	country	facility	capital	resource	Korean
Topic-5	South Korea	market	economy	time	impact
Topic-6	South Korea	world	China	US	demand
Topic-7	economy	crisis	growth	IMF	datum
Topic-8	Koreas	South	USD	government	enterprise
Topic-9	growth	GDP	quarter	economy	yy
Topic-10	sector	development	service	manufacturing	firm
Topic-11	Korea	China	cent	Japan	growth
Topic-12	industry	South Korea	technology	company	import

Topic-13	South Korea	trade	United States	relationship	partner
Topic-14	cooperative	economy	government	year	enterprise
Topic-15	population	Seoul	Korea	person	US
Topic-16	year	economy	growth	recession	business

It is worthwhile noting that the keywords *export*, *product*, *semiconductor*, *number*, and *trade* consist of topic 1. As indicated in Table 6, the word *semiconductor* is the third keyword in topic 1. This keyword is assumed to be much used since one of South Korea's major exporting items is semiconductors. It is particularly noteworthy that the keywords *South Korea*, *market*, *economy*, *time*, and *impact* are made up of topic 5. It must be noted, on the other hand, that the keywords *Korea*, *China*, *cent*, *Japan*, and *growth* consist of topic 11. In this topic, the word *China* is the second keyword since it is one of South Korea's key trading and economic partners. It must be emphasized, on the other hand, that the keywords *industry*, *South Korea, technology*, *company*, and *import* form topic 12. It is interesting to observe that the keywords *year*, *economy*, *growth*, *recession*, and *business* include topic 16. Note that these five keywords are used frequently in 35 articles of Google.

Now let us consider the use of each topic in 35 articles of Google:

Table 7 Use of each topic

	# of each topic
Topic-1	60
Topic-2	49
Topic-3	68
Topic-4	80
Topic-5	62
Topic-6	31
Topic-7	60

Topic-8	82
Topic-9	74
Topic-10	74
Topic-11	45
Topic-12	52
Topic-13	111
Topic-14	47
Topic-15	71
Topic-16	74

It is interesting to point out that topic 13 was the most occurred one. More specifically, it occurred 111 times in 35 articles. As exemplified in Table 6, the keywords *South Korea*, *trade*, *United States*, *relationship*, and *partner* consist of topic 13. It is also worth mentioning that topic 8 appeared 82 times in 35 articles. As shown in Table 6, the keywords *Koreas*, *South*, *USD*, *government*, and *enterprise* constitute topic 8. From all of this, it seems evident that these five keywords were the second most widely used ones. It is also interesting to consider topic 4. It turned up 80 times in 35 articles. The keywords *country*, *facility*, *capital*, *resource*, and *Korean* are made up of topic 4. It must also be said that topic 9 (topic 10) occurred 74 times in 35 articles. Exactly the same can be said of topic 16. It also appeared 74 times in 35 articles. It can thus be inferred that topic 13 was the most occurred one, followed by topic 8, topic 4, and topic 9 (topic 10 and topic 16), in that order.

2.4. Words

This section centers on examining the use of key nouns in 35 articles of Google. Table 8 shows the use and frequency of key nouns in 35 articles:

Number	Word	Frequency
1	South Korea	231
2	economy	199
3	country	165
4	growth	119
5	year	112
6	export	107
7	Korea	102
8	market	78

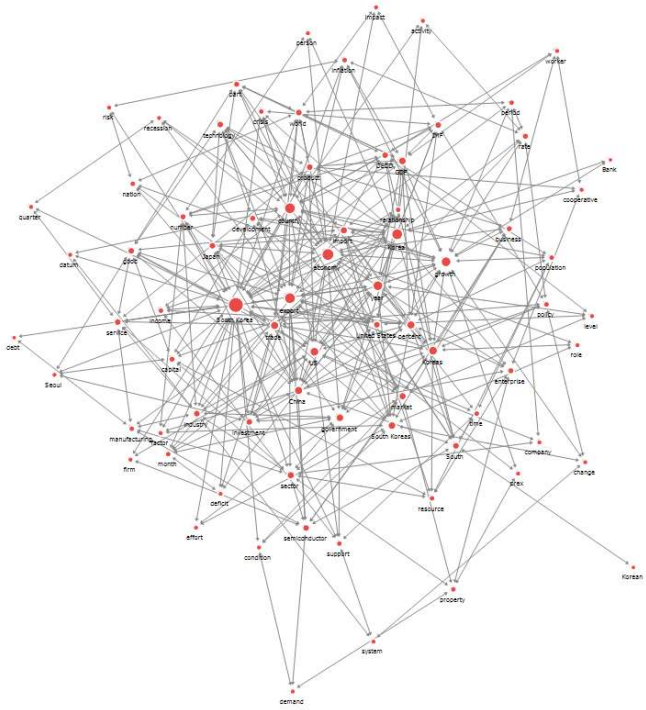
9	trade	70
10	Koreas	67
11	South	66
12	government	64
13	US	62
14	China	58
15	sector	57
16	development	53
17	crisis	52
18	rate	50
19	world	48
20	product	45
21	United States	45
22	GDP	45
23	industry	44
24	percent	41
25	service	39

It is worth noting that the word *South Korea* occurred 231 times in 35 articles. It has the highest frequency (231 tokens). It must also be noted that the word *economy* appeared 199 times in 35 articles (the second highest). It is also interesting to observe the key word *country*. It turned up 165 times in 35 articles. When it comes to the key word *growth*, it occurred 119 times. It is worth observing, on the other hand, that the key word *year* is the fifth highest. More specifically, it appeared 112 times in 35 articles. With respect to the use of the key word *export*, it is worth pointing out that it was the sixth highest. To be more specific, it turned up 107 times in 35 articles. It can thus be concluded that the key word *South Korea* was the most occurred one, followed by the key word *economy*, the key word *country*, the key word *growth*, the key word *year*, and the key word *export*, in descending order. It is also interesting to note that the key word *China* occurred 58 times in 35 articles. It should also be noted that the key word *market* appeared 78 times in 35 articles. We thus conclude that the key word *South Korea* was the most occurred one in 35 articles.

2.5. Degree centrality

This section is devoted to providing the map of degree centrality. It indicates the number of directly linked neighbors. Note that the more there are many neighbors that are directly linked to a word, the more its centrality is high. In this map, we can see core words with many neighbors that are directly linked:

Figure 2 Map of Degree centrality

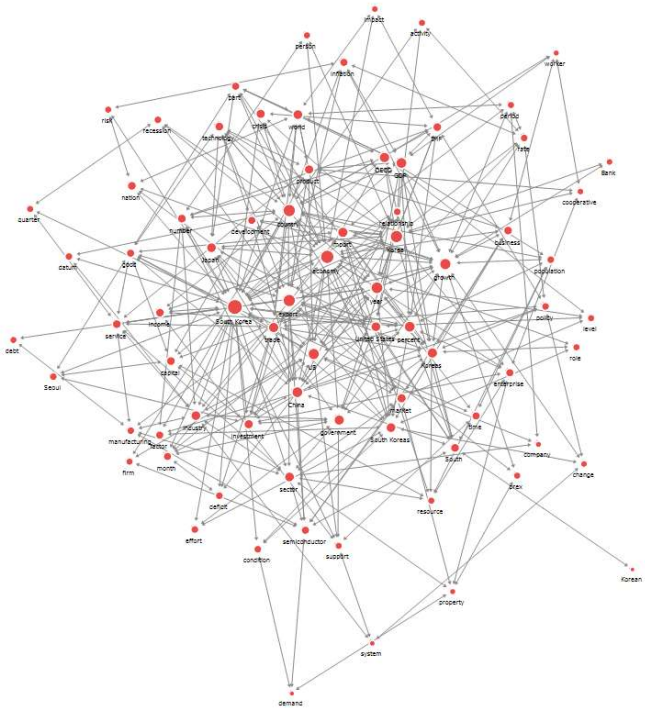


As exemplified in Figure 2, the key word *South Korea* obtains the highest degree centrality. Simply put, the centrality of *South Korea* is the highest since it has the most linked neighbors. More specifically, the key words *trade*, *export*, *country*, *development*, *capital*, *income*, *China*, *industry*, etc. are linked to the key word *South Korea*. It is also worth mentioning that the centrality of the key word *economy* is the second highest. The linked neighbors of the key word *economy* are by far smaller than those of the key word *South Korea*. The key words *import*, *country year*, *growth*, *GDP*, etc. are linked to the key word *economy*. When it comes to the key word *export*, it has smaller neighbors than the key words *South Korea* and *economy*. The key words *technology*, *recession*, *Japan*, *China*, *trade*, etc. are linked to the key word *export*. We thus conclude that the key word *South Korea* obtains the highest degree centrality, thus being a core word.

2.6. Closeness centrality

This section is devoted to providing the map of closeness centrality. The closeness centrality indicates that if the distance between nodes is close, the nodes count as important:

Figure 3 Map of closeness centrality

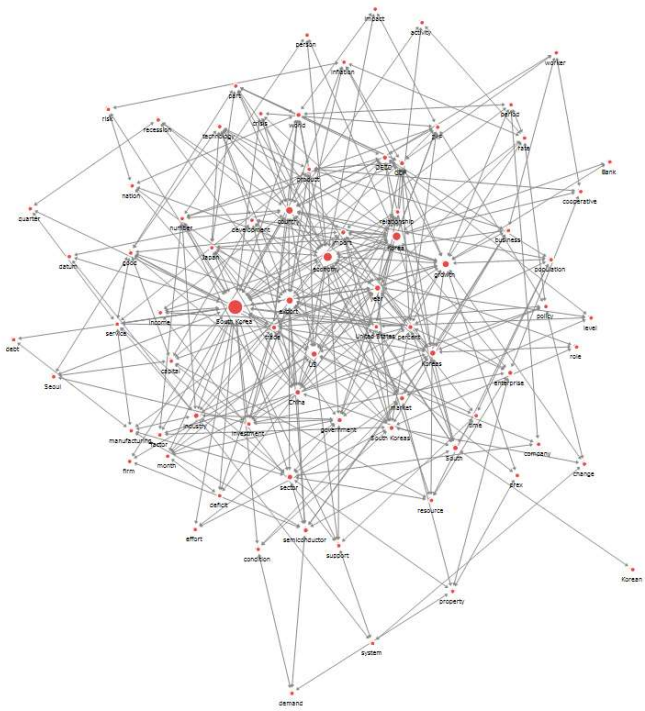


As shown in Figure 3, the key word *South Korea* has the biggest node, which indicates that it has the shortest links. Even though there are many neighbors that are linked, their links are the shortest since the key word *South Korea* occurred around the center of the map, thus being a core word. Note that the key word *OECD* has longer links than the key word *South Korea* since it appeared around an edge. It seems thus appropriate to assume that core words (e.g. *South Korea*, *economy*, *export*, *country*) appear around the center of the map. We thus conclude that the key word *South Korea* obtains the highest closeness centrality.

2.7 Betwenness centrality

This section focuses on providing the map of betweenness centrality. It indicates acting as a bridge. If nodes that act as a bridge disappear, the network will be heavily influenced:

Figure 4 Map of Betweenness centrality



Notice that nodes acting as a bridge count as important. As indicated in Figure 4, the key word *South Korea* obtains the highest betweenness centrality. This word acts a bridge among many words that are linked. As exemplified in Figure 4, many words give their links to the key word *South Korea* and it gives its links to many words, thus acting as a bridge. Note that the key word *South Korea* is followed by the key word *economy*. The latter has the second highest centrality. The number of the links of the key word *economy* is by far smaller than that of the links of the key word *South Korea*, hence the second highest. We thus conclude that the key word *South Korea* obtains the highest betweenness centrality.

3. Conclusion

To sum up, we have provided an in-depth analysis of 35 articles of Google written from 2020 to 2023 concerning South Korea's economy. In section 2.1, we have shown that the frequency of common nouns is 1,474, whereas that of proper nouns is 634. We have also provided the properties of networks consisting of 35 articles. More specifically, we have investigated the number of links, their density, and average degree. Also, we have provided inclusiveness and the number of weak components and strong components. In section 2.2, we have shown that the keyword *South Korea* is expressed as the biggest. It is deemed to be the most central one. We have also shown that the word *economy* is expressed as the second biggest, which is considered to be the second most central one. In section 2.3, we have argued that topic 13 was the most occurred one, followed by topic 8, topic 4, and topic 9 (topic 10 and topic 16), in that order. In section 2.4, we have maintained that the key word *South Korea* was the most occurred one, followed by the key word *economy*, the key word *country*, the key word *growth*, the key word *year*, and the key word *export*, in descending order. In section 2.5, we have contended that the key word *South Korea* obtains the highest degree centrality. Simply put, the centrality

of *South Korea* is the highest since it has the most linked neighbors. In section 2.6, we have argued that the key word *South Korea* obtains the highest closeness centrality since it has the shortest links. In section 2.7, we have further argued that the key word *South Korea* obtains the highest betweenness centrality since this word acts a bridge among many words that are linked.

References

[1] Kang, N. (2022a). A Comparative Analysis of Search for and Look for in Four Corpora. *Advances in Social Sciences Research Journal*, *9*(3), 168-178.

[2] Kang, N. (2022b). A Comparative Analysis of Impressed by and Impressed with in Two Corpora. *Theory and Practice in Language Studies*, *12*(5), 819-827.

[3] Kang, N. (2022c). On Speak to and Talk to: A Corpora-based Analysis. *Theory and Practice in Language Studies*, *12*(7), 1262-1270.

[4] Kang, N. (2022d). On Speak with and Talk with: A Corpora-based Analysis. *International Journal of Social Science and Human Research*, *5*(8), 3354-3360.

[5] Kang, N. (2023a). K-Pop in BBC News: A Big Data Analysis. *Advances in Social Sciences Research Journal*, *10*(2), 156-169.

[6] Kang, N. (2023b). K-Dramas in Google: A NetMiner Analysis. *Transaction on Engineering and Computing Sciences*, *11*(1), 193-216.