**Transactions on Engineering and Computing Sciences - Vol. 11, No. 2**
**Publication Date: April 25, 2023**
**DOI**:10.14738/tecs.112.14252.
Kang, N. (2023). The Korean Movie Minari: A Big Data Analysis. Transactions on Engineering and Computing Sciences, 11(2). 42-52.

SERVICES FOR SCIENCE AND EDUCATION

# The Korean Movie Minari: A Big Data Analysis

**Namkil Kang**
Far East University, South Korea

## ABSTRACT

**The ultimate goal of this paper is to analyze 42 articles of Google regarding the Korean movie *Minari*. As a research tool for our goal, we used the software package NetMiner. A major point to note is that one word has the highest frequency (1,247 tokens) and the highest proportion (57.8%). More specifically, it occurs 1,247 times in 42 articles of Google. A further point to note is that topic 11 is the most occurred one in 40 articles, followed by topic 3, topic 4, and topic 10, in that order. When it comes to the use of keywords in 42 articles, the word *film* is the most widely used one, followed by the word *Minari*, the word *family*, the name *Jacob*, and the name *Monica*, in that order. Talking about degree centrality, the word *family* has the highest degree centrality, followed by the name *Monica*, the word *Minari*, and the word *film*, in that order. It is worthwhile noting, on the other hand, that the word *family* has the highest closeness centrality since the distance among the word *family* and the other words is the closest. Finally, this paper argues that the name *Monica* has the highest Eigenvector centrality since it has the most neighbors. Thus, the name *Monica* counts as the most influential and important.**

**Keywords:** Minari, NetMiner, degree centrality, closeness centrality, Eigenvector centrality

## INTRODUCTION

The main purpose of this paper is to analyze 42 articles of Google written from December 2020 to May 2022 regarding the Korean movie *Minari*. We analyzed 42 articles of Google in terms of the software package NetMiner. First, we aim to investigate the frequency, proportion, and cumulative proportion of nouns such as common nouns and proper nouns. Also, we inquire into the frequency, proportion, and cumulative proportion of nouns occurred in 42 articles of Google. Second, we aim at observing 14 topics occurred in 42 articles of Google and 5 keywords consisting of each topic. Also, we observe how often each topic turns up in 42 articles of Google. Third, we attempt to consider how often major keywords constituting 42 articles of Google occur. Fourth, we aim to examine degree centrality, closeness centrality, and Eigenvector centrality and provide their maps. The term centrality refers to the so-called importance, reputation, prominence, prestige, and influence. Through degree centrality, we can see core words directly linked with their neighbors. The term degree refers to the number of directly linked neighbors. On the other hand, closeness centrality indicates that the more the distance between nodes is close, the more centrality is high. Finally, the term Eigenvector centrality indicates that centrality is high if nodes are linked to nodes with high centrality. Put differently, if my neighbors have many neighbors, my influence grows.

## RESULTS

### Word Information

This section centers on investigating the frequency, proportion, and cumulative proportion of nouns occurred in 42 articles of Google. Table 1 shows information on the frequency, proportion, and cumulative proportion of nouns such as common nouns and proper nouns occurred in 42 articles:

**Table 1 Information on nouns**

| Value | Frequency | Proportion | Cumulative Proportion |
|---|---|---|---|
| **Common Noun** | 1607 | 0.745 | 0.745 |
| **Proper Noun** | 549 | 0.255 | 1 |
| **Total** | 2156 | 1 | |

It is interesting to note that in 42 articles of Google, common nouns occur 1,607 times, whereas proper nouns appear 549 times. It should also be pointed out that the use of common nouns is almost three times higher than that of proper nouns. The proportion of common nouns is 0.745 (74.5%), whereas that of proper nouns is 0.255 (25.5%). As exemplified in Table 1, the total number of common nouns and proper nouns is 2,156.

Now Table 2 shows information on the frequency, proportion, and cumulative proportion of nouns occurred in 40 articles of Google:

**Table 2 Word information**

| Value | Frequency | Proportion | Cumulative Proportion |
|---|---|---|---|
| 1.0 | 1247 | 0.578 | 0.578 |
| 2.0 | 344 | 0.16 | 0.738 |
| 3.0 | 158 | 0.073 | 0.811 |
| 4.0 | 104 | 0.048 | 0.859 |
| 5.0 | 43 | 0.02 | 0.879 |
| 6.0 | 50 | 0.023 | 0.903 |
| 7.0 | 25 | 0.012 | 0.914 |
| 8.0 | 17 | 0.008 | 0.922 |
| 9.0 | 21 | 0.01 | 0.932 |
| 10.0 | 20 | 0.009 | 0.941 |
| 11.0 | 10 | 0.005 | 0.946 |
| 12.0 | 5 | 0.002 | 0.948 |
| 13.0 | 11 | 0.005 | 0.953 |
| 14.0 | 10 | 0.005 | 0.958 |
| 15.0 | 6 | 0.003 | 0.961 |
| 16.0 | 8 | 0.004 | 0.964 |
| 17.0 | 4 | 0.002 | 0.966 |
| 18.0 | 2 | 0.001 | 0.967 |
| 19.0 | 4 | 0.002 | 0.969 |
| 20.0 | 2 | 0.001 | 0.97 |

| | | | |
|---|---|---|---|
| 21.0 | 5 | 0.002 | 0.972 |
| 22.0 | 6 | 0.003 | 0.975 |
| 23.0 | 4 | 0.002 | 0.977 |
| 25.0 | 6 | 0.003 | 0.98 |
| 27.0 | 1 | 0 | 0.98 |
| 28.0 | 1 | 0 | 0.981 |
| 29.0 | 3 | 0.001 | 0.982 |
| 30.0 | 2 | 0.001 | 0.983 |
| 31.0 | 3 | 0.001 | 0.984 |
| 32.0 | 2 | 0.001 | 0.985 |
| 33.0 | 3 | 0.001 | 0.987 |
| 35.0 | 3 | 0.001 | 0.988 |
| 36.0 | 2 | 0.001 | 0.989 |
| 37.0 | 1 | 0 | 0.989 |
| 39.0 | 1 | 0 | 0.99 |
| 40.0 | 1 | 0 | 0.99 |
| 42.0 | 1 | 0 | 0.991 |
| 43.0 | 1 | 0 | 0.991 |
| 45.0 | 2 | 0.001 | 0.992 |
| 48.0 | 1 | 0 | 0.993 |
| 50.0 | 1 | 0 | 0.993 |
| 53.0 | 1 | 0 | 0.994 |
| 59.0 | 2 | 0.001 | 0.994 |
| 63.0 | 1 | 0 | 0.995 |
| 68.0 | 1 | 0 | 0.995 |
| 71.0 | 1 | 0 | 0.996 |
| 72.0 | 1 | 0 | 0.996 |
| 78.0 | 1 | 0 | 0.997 |
| 86.0 | 1 | 0 | 0.997 |
| 89.0 | 1 | 0 | 0.998 |
| 93.0 | 1 | 0 | 0.998 |
| 110.0 | 1 | 0 | 0.999 |
| 167.0 | 1 | 0 | 0.999 |
| 191.0 | 1 | 0 | 1 |
| 232.0 | 1 | 0 | 1 |
| Total | 2156 | 1 | |

Perhaps it is worthwhile noting that one word has the highest frequency (1,247 tokens) and the highest proportion (57.8%). Simply put, it occurs 1,247 times in 42 articles of Google. It is interesting to observe two words. They occur 344 times (the second highest) in 42 articles. Note that their proportion and cumulative proportion are 0.16 and 0.738, respectively. It is also worth mentioning that three words appear 158 times (the third highest). Notice that their proportion and cumulative proportion are 0.073 and 0.811, respectively. It must be stressed, on the other hand, that four words occur 104 times in 42 articles and that their proportion is 0.048 (4.8%). We thus conclude that one word appears 1,247 times and that it ranks first.

## Topics

This section is focused on looking into 14 topics and their keywords occurred in 42 articles of Google. Table 3 shows information on 5 keywords constituting each topic. Notice that 5 keywords consist of each topic, which in turn constitutes 42 articles:

**Table 3 Information on 14 topics**

| | 1st Keyword | 2nd Keyword | 3rd Keyword | 4th Keyword | 5th Keyword |
|---|---|---|---|---|---|
| **Topic-1** | Minari | story | water | film | award |
| **Topic-2** | Minari | character | film | life | award |
| **Topic-3** | Chung | film | movie | Minari | dream |
| **Topic-4** | film | Lee | story | director | Isaac Chung |
| **Topic-5** | film | Korea | year | Minari | US |
| **Topic-6** | country | film | home | experience | church |
| **Topic-7** | American | family | language | Minari | Korea |
| **Topic-8** | Minari | life | immigrant | family | actor |
| **Topic-9** | time | scene | story | life | person |
| **Topic-10** | David | Anne | Alan | Noel | Cho |
| **Topic-11** | Jacob | Monica | farm | Han | Steven Yeun |
| **Topic-12** | family | life | Yeun | person | story |
| **Topic-13** | family | Soonja | Arkansas | story | California |
| **Topic-14** | Youn | movie | actress | home | Soonja |

It is interesting to observe topic 1. Topic 1 includes the keywords *Minari*, *story*, *water*, *film*, and *award*. As expected, the word *Minari* is the first keyword, which in turn suggests that it is the most widely used one in topic 1. As illustrated in Table 3, the keywords *film*, *Korea*, *year*, and *Minari* constitute topic 5. In this topic, the keyword *film* is the most occurred one, thus leading to the first topic. It is worth noting, on the other hand, that topic 8 includes the keywords *Minari*, *life*, *immigrant*, *family*, and *actor*. Again, the word *Minari* is the most frequently used one, hence becoming the first keyword. Finally, it must be noted that the five keywords *family*, *Soonja*, *Arkansas*, *story*, and *California* consist of topic 13. It is worth observing that the word *family* is the most occurred one, thereby resulting in the first topic. Now attention is paid to the information of each topic:

**Table 4 Information on each topic**

| | # of sentences |
|---|---|
| **Topic-1** | 78 |
| **Topic-2** | 36 |
| **Topic-3** | 114 |
| **Topic-4** | 113 |
| **Topic-5** | 80 |
| **Topic-6** | 89 |
| **Topic-7** | 81 |
| **Topic-8** | 44 |
| **Topic-9** | 87 |
| **Topic-10** | 98 |

| | |
|---|---|
| **Topic-11** | 168 |
| **Topic-12** | 56 |
| **Topic-13** | 88 |
| **Topic-14** | 87 |

It is particularly noteworthy that topic 11 occurs 168 times in 42 articles of Google and that it ranks first. Note that the keywords *Jacob*, *Monica*, *farm*, *Han*, and *Steven Yeun* form topic 11. Quite interestingly, topic 11 is followed by topic 3. When it comes to topic 3, it appears 114 times (the second highest). As indicated in Table 3, the keywords *Chung*, *film*, *movie*, *Minari*, and *dream* constitute topic 3. Perhaps it is worthwhile pointing out that topic 4 occurs 113 times in 42 articles of Google (the third highest). As exemplified in Table 3, topic 4 includes the keywords *film*, *Lee*, *story*, *director*, and *Isaac Chung*. It is worth pointing out, on the other hand, that topic 10 appears 98 times and that it ranks fourth. It is thus reasonable to conclude that topic 11 is the most occurred one in 40 articles, followed by topic 3, topic 4, and topic 10, in that order.

## Word Frequency

In what follows, we aim at examining how often a particular word occurs in 42 articles of Google. Table 5 shows how frequently a particular word appears in 42 articles:

**Table 5 Word frequency**

| Number | Word | Frequency |
|---|---|---|
| 1 | film | 212 |
| 2 | Minari | 185 |
| 3 | family | 158 |
| 4 | Jacob | 106 |
| 5 | Monica | 90 |
| 6 | David | 83 |
| 7 | story | 80 |
| 8 | movie | 75 |
| 9 | Youn | 71 |
| 10 | Arkansas | 67 |
| 11 | American | 67 |
| 12 | Chung | 63 |
| 13 | life | 55 |
| 14 | farm | 55 |
| 15 | Soonja | 52 |
| 16 | home | 48 |
| 17 | time | 45 |
| 18 | Lee | 45 |
| 19 | dream | 43 |
| 20 | Han | 42 |
| 21 | child | 40 |
| 22 | award | 37 |
| 23 | America | 36 |

| | | |
|---|---|---|
| 24 | grandmother | 35 |
| 25 | Steven Yeun | 35 |
| 26 | director | 27 |
| 27 | language | 33 |
| 28 | Korean | 33 |
| 29 | immigrant | 32 |
| 30 | character | 32 |

It must be stressed that the word *film* occurs 212 times in 42 articles of Google. Simply put, it has the highest frequency (212 tokens). It is interesting to consider the word *Minari*. It appears 185 times in 42 articles and it ranks second. It is important to mention, on the other hand, that the word *family* occurs 158 times and that it ranks third. It is also important to note that *Jacob* appears 106 times and that it ranks fourth. It is interesting to note that *Monica* occurs 90 times in 42 articles and that it ranks fifth. From all of this, it seems clear that the word *film* is the most widely used one in 42 articles of Google, followed by the word *Minari*, the word *family*, the name *Jacob*, and the name *Monica*, in that order. Perhaps it is worthwhile noting that the word *farm* appears 55 times in 42 articles. Finally, it should also be pointed out that the word *dream* occurs 43 times and that it ranks nineteenth. It can thus be concluded that the word *film* has the highest frequency (212 tokens).

## Network Properties

In the following, we aim to consider the network properties of 42 articles. To begin with, let us consider the number of links and components (weak and strong) and their density:

**Table 6 Network properties**

| | # of Links: O(m) | Density: O(m) | Average Degree: O(m) | # of Components (Weak): O(m) | # of Components (Strong): O(m) |
|---|---|---|---|---|---|
| **Word Network (Sentences)** | 2,874 | 0.001 | 1.591 | 95 | 1,161 |

It is worth observing that the total number of links is 2,874. Note that the software package NetMiner provides networks by linking the relevant words to other words. It seems thus reasonable to assume that 2,874 links form the whole network. The term weak components indicate that there is a link between node A and node B. On the other hand, the term strong components indicate that node A gives a link to node B and node A receives a link from node B. Put differently, strong components happen in two directions. As exemplified in Table 6, the number of strong components (1,161) is twelve times higher than that of weak components (95). Now attention is paid to network properties such as inclusiveness, reciprocity (Arc and Dyad), and transitivity:

**Table 7 Network properties**

| | Inclusiveness: O(m) | Reciprocity (Arc): O(m) | Reciprocity (Dyad): O(m) | Transitivity: O(nm) |
|---|---|---|---|---|
| **Word Network (Sentences)** | 1 | 0.045 | 0.023 | 0.024 |

The term inclusiveness refers to the proportion of linked nodes. Note that as illustrated in Table 7, the proportion of linked nodes is 100%. The term reciprocity (Arc) refers to the proportion of reciprocal links. On the other hand, the term reciprocity (Dyad) refers to the proportion of node pairs. As shown in Table 7, the proportion of reciprocal links (reciprocity Arc) is 45%, whereas the proportion of node pairs is 23%. The term transitivity takes place among three nodes. That is to say, if node A gives a link to node B and node B gives a link to node C, then node A gives a link to node C. Finally, let us turn our attention to network properties such as clustering coefficient and the number of isolated nodes:

**Table 8 Network properties**

| | Clustering Coefficient: O(n^3) | Mean Distance: O(nm) | Diameter: O(nm) | # of Isolated Nodes: O(m) |
|---|---|---|---|---|
| **Word Network (Sentences)** | 0.073 | 5.711 | 23 | 0 |

The term clustering coefficient refers to the probability value of neighboring nodes associated with a node. It takes place between 0 and 1. The term mean distance called Average Path Length indicates the number and stage of links. That is to say, the mean distance is 5.711 links. On the other hand, the term diameter refers to the furthest distance between two nodes. Finally, it is worth noting that the number of isolated nodes is 0.

## CENTRALITY

### Degree Centrality

The so-called degree centrality is concerned with the number of directly linked neighbors. Therefore, it is called local centrality. Eigenvector centrality indicates, on the other hand, that if my neighbors have many neighbors, then my influence grows, thus being called global centrality. This section centers on providing the map of degree centrality and analyzing the map. We got rid of many nodes and included core nodes only for the reason of space:

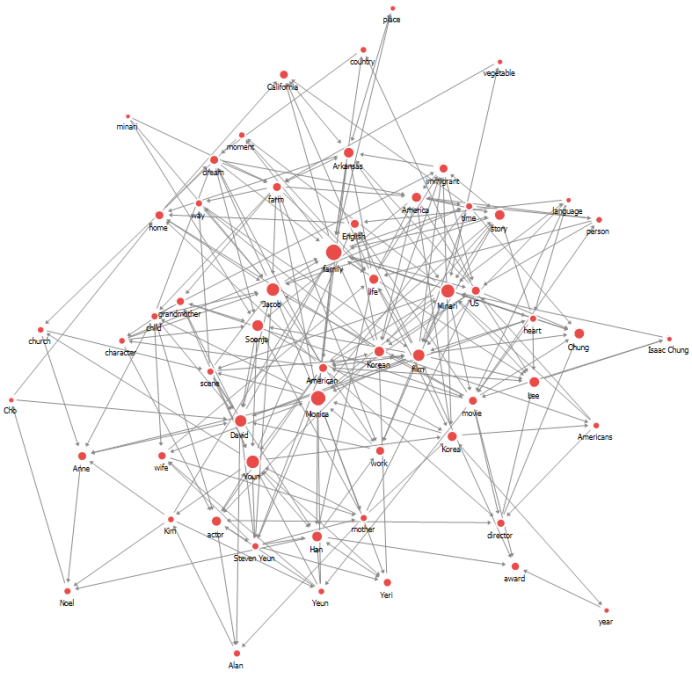



**Figure 1 map of degree centrality**

It is worth mentioning that the word *family* has the highest degree centrality. This in turn implies that this word has the most neighbors that are directly linked. More specifically, the word *family* shares the direct links with the words *farm*, *Jacob*, *American*, *Korean*, *life*, *time*, *immigrant*, *scene*, *Arkansas*, etc. It is important to mention that we can see core words in terms of degree centrality since many words are directly linked to them. It should also be noted that the name *Monica* has the second highest degree centrality. Note that the words *farm*, *Jacob*, *family*, *film*, *work*, *David*, etc. are linked to it. It is also noteworthy that the word *Minari* has the third highest degree centrality. It is interesting to consider the word *film*. When it comes to the word *film*, it has the fourth highest degree centrality. This in turn indicates that the word *film* is the fourth highest in the number of directly linked nodes. We thus conclude that the word *family* has the highest degree centrality, followed by the name *Monica*, the word *Minari*, and the word *film*, in that order. To sum up, the words *family*, *Monica*, *Minari*, *film*, etc. are assumed to be core words since many words (neighbors) are directly linked to these words.

## Closeness Centrality

This section is focused on providing the map of closeness centrality and analyzing the map. The term closeness centrality indicates that the more the distance among nodes is close, the more they are important. Thus, the more the distance among nodes is close, the more the value of closeness centrality is large. Now let us observe the map of closeness centrality:

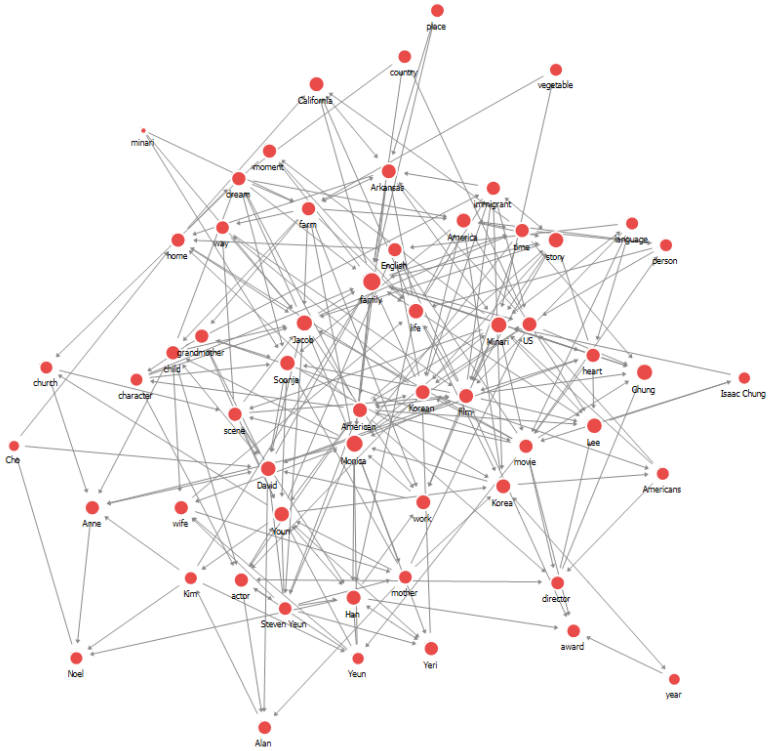



**Figure 2 map of closeness centrality**

Note, to begin with, that the central place of this map has high closeness centrality since the distance among nodes is close. For instance, the word *family* occurs around the central place, whereas the word *person* appears around an edge. Thus, closeness centrality of the word *family* is by far higher than that of the word *person*. Put differently, the word *family* rather than the word *person* counts as more important. As can be seen from Figure 2, the distance among the word *family* and the other words is the closest, compared to other words. It is also worth noting that the name *Monica* has the second highest closeness centrality. Similarly, the name *Monica* also occurs around the central place of the map, thereby leading to the second highest closeness centrality. It must also be stressed that the word *Minari* has the third highest closeness centrality. Its place is near to the central place of the map. It seems thus reasonable to assume that the word *Minari* counts as the third most important. To sum up, the word *family* has the highest closeness centrality since the distance among the word *family* and the other words is the closest, as compared with other words, thus counting as the most important.

**Eigenvector Centrality**

In this section, we aim at providing the map of Eigenvector centrality and analyzing the map. The term Eigenvector centrality refers to links with many neighbors. To put it briefly, if my neighbors have many neighbors, then my influence grows. Now let us observe the map of Eigenvector centrality:

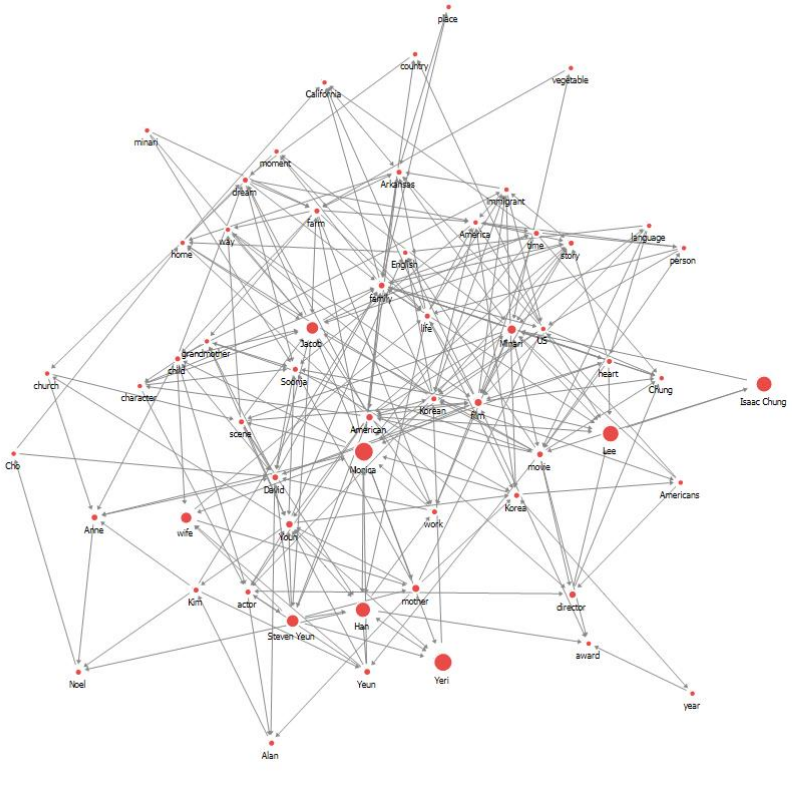



**Figure 3 Map of Eigenvector centrality**

Perhaps it is worthwhile pointing out that the name *Monica* is the most influential and important since it has the most neighbors. More specifically, *Monica* is linked to the words *film*, *Jacob*, *American*, etc., which are also linked to many other words (many neighbors). In a word, the name *Monica* has the highest Eigenvector centrality. It is also worth mentioning that the name *Yeri* has the second highest Eigenvector centrality. To be more specific, *Yeri* is linked to the words *Monica*, *work*, etc., which are also linked to many other neighbors. Note, however, that *Monica* has more neighbors than *Yeri*, thus implying that *Monica* is more influential and important than *Yeri*. It should also be pointed out that *Lee* has the third highest Eigenvector centrality. More specifically, *Lee* is linked to the words *heart*, *film*, etc., which are also linked to many other neighbors (*film*, *Minari*, etc.). Notice, however, that *Yeri* has more neighbors than *Lee*, thereby indicating that *Yeri* is more influential and important than *Lee*. It seems thus reasonable to conclude that *Monica* has the most neighbors, so she is regarded as the most influential and important. For the map of synonyms and keywords, see Kang (2022a, 2022b, 2022c, 2022d, 2023a, 2023b).

## CONCLUSION

To sum up, we have analyzed 42 articles of Google regarding the Korean movie *Minari*. In section 2.1, we have shown that one word has the highest frequency (1,247 tokens) and the highest proportion (57.8%). More specifically, it occurs 1,247 times in 42 articles of Google. In section 2.2, we have argued that topic 11 is the most occurred one in 40 articles, followed by

topic 3, topic 4, and topic 10, in that order. In section 2.3, we have further argued that the word *film* is the most widely used one in 42 articles of Google, followed by the word *Minari*, the word *family*, the name *Jacob*, and the name *Monica*, in that order. In section 2.3, we have provided network properties such as inclusiveness, reciprocity (Arc and Dyad), transitivity, and the number of links and components (weak and strong). In section 3.1, we have contended that the word *family* has the highest degree centrality, followed by the name *Monica*, the word *Minari*, and the word *film*, in that order. In section 3.2, we have maintained that the word *family* has the highest closeness centrality since the distance among the word *family* and the other words is the closest. In section 3.3, we have argued that the name *Monica* has the highest Eigenvector centrality since it has the most neighbors. Therefore, she counts as the most influential and important.

## References

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.

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.

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

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.

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

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