

# 한국 영화배우 소셜 네트워크 데이터 분석을 통한 중심성 변화 연구

최준영<sup>1</sup> · 이오준<sup>1</sup> · 정재은<sup>1</sup> · 용환성<sup>2</sup>

<sup>1</sup>중앙대학교 · <sup>2</sup>리얼리티랩

## Understanding Temporal Change of Centrality by Analyzing Social Network among Korean actors

Joonyoung Choi<sup>1</sup> · O-Jun Lee<sup>1</sup> · Jason J. Jung<sup>1</sup> · Hwan-Sung Yong<sup>2</sup>

<sup>1</sup>Chung-Ang University · <sup>2</sup>Reality Lab

E-mail :{mydream757, concerto9203}@gmail.com, j3ung@cau.ac.kr, zeus@realitylab.co.kr

### 요 약

이 논문에서는 한국 영화의 출연 배우들을 노드로 하여, 동반 출연의 경우에 엣지를 설정하여 그래프를 구축하는 방법을 보인다. 이렇게 구축한 그래프를 통하여, 논문에서 정의한 ‘중심 지수’를 각각의 배우에 대해 구하고, 이를 바탕으로 중심성 수치에 따른 순위를 매겨 그 분포를 조사하고, 또한 연도별 중심 배우의 변화를 살펴본다. 마지막으로 6단계 분리 이론과 접목하여 사회적 집단의 범위를 수치적으로 한정하는 방안을 제시한다.

### ABSTRACT

On this paper, we show the way of forming graph data structure via setting an edge between Korean actors if they appeared in the same movie. From this graph, we calculate the ‘centralities’ (which declared on this paper) for each actor, then examine distribution by ranking the actors of the centralities and analyze the change of the actor who is/was center on the graph by years. Finally, we suggest the way that sets the numerically Range limits on social group.

### 키워드

Graph analytics, Korean movies, Korean actors, Centrality

### 1. 서론

그래프(graph) 자료 구조란, 연결되어있는 노드(node)간의 관계를 표현하는 자료 구조[1]이다. 그래프의 성격에 따라 엣지(edge)의 방향성, 엣지의 가중치가 존재할 수 있다. 그래프 자료 구조는 자료들을 저장함에 있어 단순히 인덱싱(indexing)하는 것이 아니라, 연관 관계를 파악하는 데에 유용하다 [1]. 때문에, 인물 지도 혹은 도시 지도 등의 자료를 구조화하는 데에 많이 쓰이며, 주목받고 있는 자료 구조의 대표적인 예시이다[2, 3]. 이 논문에서

는 이러한 그래프 자료 구조의 사용에 대한 예로서, ‘한국 영화에 출연한 배우들의 관계’를 그래프로 구조화할 것이다. 또한 ‘중심 지수(centrality)를 정의하고, 연도별로 중심 지수가 가장 높은 배우를 중심 배우로 선정하여, 중심 배우의 변화가 어떻게 이루어지는지 알아볼 것이다.

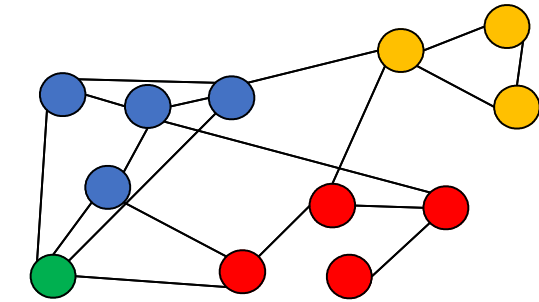


그림 1. 그래프 자료 구조 시각화

## II. 제약 사항

### 2.1 영화 및 배우 자료 수집원

자료는 한국영화진흥위원회에서 기재한 자료만을 유일하게 사용한다.

### 2.2 자료 수집 형식

- 1) 대표 국적: 한국
- 2) 개봉 년도: 1959~2019년
- 3) 데이터 파일: 엑셀(excel) 파일
- 4) 결과 파일: 씨에스브이(.csv) 파일

### 2.3 자료 유효성 판단

- 1) 동일한 영화가 중복 노드로서 기록되지 않도록 한다.
- 2) 영화의 대표 국적 이외에 출연진의 국적은 무시한다.

## III. 그래프 형성

영화진흥위원회에서 얻은 영화와 출연 배우에 대한 자료를 이용해서, 영화배우 각각을 노드로 하는 그래프를 형성한다. 자료 수집 기간(1959~2019년)을 통틀어서 그래프를 구축하지 않고, 10년 단위의 영화 데이터를 이용하여 그래프를 동적으로 구축하였다. 예를 들어, 개봉년도 1959~1968년에 해당하는 자료를 통해 그래프를 구축하고, 알고리즘을 수행한 후, 시작년도와 종료년도를 1년씩 증가시켜, 마지막으로 2010~2019년에 대한 그래프를 형성하여 알고리즘을 수행한다. 이 알고리즘은, 각각의 배우들의 중심 지수를 구하고, 이를 정렬하여 가장 높은 중심도를 가진 배우를 도출한다.

그래프의 모든 엣지는 가중치 1을 갖고, 방향을 갖지 않는다. 이 가중치는 각각의 배우가 서로 얼마만큼의 거리를 가지고 있는지 계산하기 위한 상대적 수치이다. 예를 들어, 하나의 엣지를 만드는 두 노드는 1의 거리를 갖고, 2개의 엣지를 통해 연결되어 있는(사이에 하나의 경유 노드를 가진) 두 노드는 2의 거리를 갖는다.

## IV. 중심 지수와 중심도

중심 지수는 특정 배우가 자신 이외의 다른 모든 배우와 갖는 거리의 평균으로 정의한다. 이는 6단계 분리 이론(Six degrees of separation)[4, 5]에 대한 놀이인 케빈 베이컨의 6단계(Six degrees of Kevin Bacon)[6]에서 제시한 방법을 참고하였다.

중심 지수를 구하기 위해서는 몇 가지 예외에 대한 정의가 필요하다. 예를 들어, 모든 배우들이 하나의 그래프로 연결되어있을 수도 있지만, 하나의 큰 그래프가 형성되고, 나머지 작은 그래프들이 각각 존재하게 될 수도 있다. 따라서 이러한 상황에 공통적인 다음의 조건을 설정하였다. 먼저, 임의의 노드  $N_a$ 와  $N_b$ 가 인접한 노드라면 거리를 1로 계산한다. 두 번째로, 임의의 노드  $N_a$ 와  $N_b$ 의 거리는 두 노드의 최단 경로에 대한 거리이다. 만약  $N_a$ 와  $N_b$  사이에 경로가 존재하지 않는다면,  $N_a$ 와  $N_b$ 의 거리는, 그래프 상에 존재하는 거리 중 가장 먼 거리인  $D_{MAX}$ 에 1을 더한 값으로 한다.

위의 조건을 토대로 중심 지수는 식(1)에 의해 산출된다.

$$C_a = \sum_{i=0}^{K-1} d_{a,i} / (k-1) \quad (1)$$

중심 지수는 결국, 다른 배우들과 평균적으로 얼마나 거리가 떨어져 있는지를 나타낸다. 따라서, 수치적으로 낮은 중심 지수는, 다른 배우들과 평균적으로 가까이에 있다는 것을 나타내므로, 이를 중심도가 높다고 칭할 것이다. 이를 토대로, 가장 높은 중심도를 가진 배우를 중심 배우라 한다.

## V. 결론

중심도가  $n$ 이라는 것은, 그래프에서(다시 말해, 한국 영화 배우라는 집단에서) 다른 배우들과 평균적으로  $n$ 만큼 떨어져있다고 할 수 있다. 연결되지 않은 노드를 제외하고, 연결되어있는 노드들 간의 거리의 최대치는 6으로, 이는 6단계 분리 이론을 부분적으로 만족한다. 단, 영화 출연이라는 특성상, 단 하나의 영화에만 출연한 데다가, 해당 영화의 출연진들이 모두 그 영화에만 출연했다면, 다시 말해 한국 영화 배우라는 집단 내에서 분리된 집단이라면 그래프에 연결되지 않는 문제가 생긴다. 이를 통해 생각할 수 있는 바는, 사회적 집단에 대한 수치적 정의이다. 우리는, 만약 6단계를 초과하는, 다시 말해 해당 논문에서  $D_{MAX}(=6)$ 으로 제시한 거리를 초과하는 집단을 다른 사회적 집단으로

규정할 수 있는가라는 의문을 남긴다. 반대로,  $D_{MAX}$  이내에 있으면 반드시 동일한 사회적 집단에 속해있는가도 중요한 의문점이다. 다음의 그림2를 보면, 특정 거리(= 중심도)를 기준으로, 매우 급격한 변화가 생기는 것을 확인할 수 있다. 이를 순위로 생각해보았을 때, 사실상 연결되어있지 않은 노드들과 순위상 차이가 미약하다.

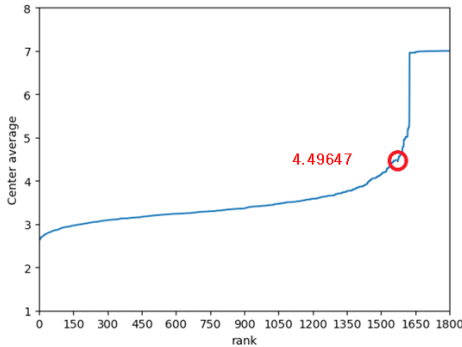


그림 2. 2013~2018년 한국 영화 출연 배우들의 중심도 순위

한편, 1959년부터 시작해 1년씩 증가시키면서, 10년 단위의 데이터들로 그래프를 구성하여 결과적으로 총 52개의 데이터를 얻었다. 각각의 데이터는 10년 동안의 영화와 출연 배우들로 이루어진 그래프에서, 각 출연 배우의 중심도를 계산해 얻어졌다.

표 1. 중심 배우의 변화

수집 년도	배우	중심도
1959-1968	신영균	1.6
1960-1969	신영균	2.03571
1961-1970	신영균	2.08333
1962-1971	신영균	2
1963-1972	신영균	2.84615
1964-1973	신영균	3.06897
1965-1974	신영균	3.69697
1966-1975	신영균	3.95
1967-1976	신성일	3.93182
1968-1977	신성일	3.02174
1969-1978	신성일	2.97959
1970-1979	신성일	3.38356
1971-1980	신성일	3.07407
1972-1981	신성일	2.47525
1973-1982	신성일	2.47619
1974-1983	김지영	2.38889
1975-1984	김지영	2.41228
1976-1985	김지영	2.28302
1977-1986	김지영	1.97222
1978-1987	김지영	1.95238
1979-1988	김지영	1.90826
1980-1989	김지영	1.83516
1981-1990	김지영	1.75962
1982-1991	김지영	1.81395

1983-1992	김지영	2.02326
1984-1993	김지영	2.16049
1985-1994	박중훈	2.38356
1986-1995	박중훈	2.12195
1987-1996	박중훈	2.11236
1988-1997	박중훈	2.21905
1989-1998	최종원	2.12605
1990-1999	박길수	2.18056
1991-2000	박길수	2.34591
1992-2001	기주봉	2.17972
1993-2002	기주봉	2.49603
1994-2003	송강호	2.20061
1995-2004	송강호	2.46194
1996-2005	송강호	2.31422
1997-2006	송강호	2.36059
1998-2007	송강호	2.90615
1999-2008	송강호	2.81712
2000-2009	송강호	2.65472
2001-2010	송강호	2.77307
2002-2011	조덕제	2.56892
2003-2012	조덕제	2.60313
2004-2013	송강호	2.65531
2005-2014	황정민	2.49401
2006-2015	조덕제	2.54673
2007-2016	황정민	2.49821
2008-2017	이경영	2.57282
2009-2018	이경영	2.69384
2010-2019	이경영	2.6972

이를 통해, 시간의 흐름에 따라 중심 배우의 변화를 파악할 수 있다. 중심 배우들은 1.75~4 정도 범위의 중심도를 나타냈다. 수집 년도가 2019년에 가까워질수록 중심도의 차이는 줄어들어, 2~3의 수치를 보였다.

수집 년도가 2019년에 가까워진다는 것은, 그래프의 구성이 더 복잡해짐을 의미한다. 왜냐하면 영화의 작품작과 그 규모가 증가하면서, 배우들의 수와 그 배우들의 출연 작품들도 다양해지기 때문이다. 그래프의 구성이 복잡해질 수록, 현실에 존재하는 사회적 집단과 유사한 형태를 보여줄 것이라 생각할 수 있다. 그러므로, 우리는 특정 집단에 대한 규정을 위에서 제시한 수치로서 규정할 수 있다. 해당 결과는 수집 자료의 조건에 따라 달라질 수 있다.

### Acknowledgement

본 연구는과학기술정보통신부 및 정보통신기술진흥센터의 SW중심대학지원사업의 연구결과로 수행되었음 (20170001000031001).

### References

[1] Neo4j, Why graph database. <https://neo4j.org>

- [2] Wikipedia, Graph (abstract data type),  
[https://en.wikipedia.org/wiki/Graph\\_\(abstract\\_data\\_type\)](https://en.wikipedia.org/wiki/Graph_(abstract_data_type)).
- [3] [Hae-Sung Lee, Joon-Hee Kwon], “The Study of Graph Database Model-based Social Search Techniques”, Proceedings of KIIT Summer Conference, pp 123-126, May. 2012.
- [4] S Milgram, “The Small World Problem”, Psychology today, Vol. 1, No.1, pp. 61-67, May 1967.
- [5] Wikipedia, Six degrees of separation.  
[https://en.wikipedia.org/wiki/Six\\_degrees\\_of\\_separation](https://en.wikipedia.org/wiki/Six_degrees_of_separation)
- [6] Wikipedia. Six degrees of Kevin Bacon.  
[https://en.wikipedia.org/wiki/Six\\_degrees\\_of\\_Kevin\\_Bacon](https://en.wikipedia.org/wiki/Six_degrees_of_Kevin_Bacon)