**네트워크 및 텍스트 분석법을 이용한 한국뮤지션의 특징 분석**

---------------------------------------------------------------------------------------------------------

통계학과 20190779 정다인

**1. 한국드라마 네트워크**

**1-1. 데이터 수집**

1. 위키피디아에서 한국뮤지션 목록 가져오기

* Urlib와 json 모듈을 불러온 후, 위키피디아 API를 사용
* json 데이터 형태로 페이지 본문만 추출

**1-2. 전처리 방법**

1. 첫번째 타입(link\_type1) : [[(실제페이지명)|(표시페이지명)]] 처리

* 처음 2개는 한국 뮤지션이 아니므로 제외



그림 - 'link\_type1' 중 제거부분

1. 두번째 타입(link\_type2) :  [[페이지명]] 처리

* 처음 2개와 마지막 5개는 한국뮤지션이 아니므로 제외함

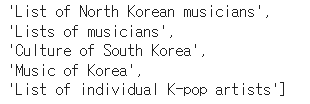
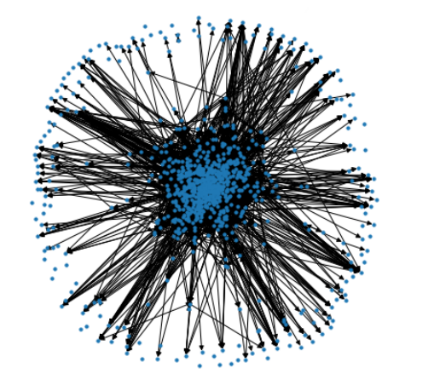


그림 2 - 'link\_type2' 중 제거부분

1. 첫번째 타입과 두번째 타입으로부터 노드(한국뮤지션) 목록 생성

* 첫번째 타입은 두 부분 중 앞 부분만 추출하고, 두번째 타입은 노드 이름이므로 그대로 추출 후 실행
* 집합으로 생성 후, 리스트로 변환 : 집합은 중복을 허락하지 않으므로 중복 제거

 **1-3. 네트워크생성**

* 노드 수는 855, 엣지 수는 5892
* 노드는 네트워크의 구성원이고, 엣지는 네트워크를 형성하는 관계로 노드들 간의 연결을 나타낸다.
* 다른 노드들과 연결되어 있는 노드가 많다.

고립노드가 적은 것을 보아, 한국뮤지션은 상대적으로 노드들 간 높은 네트워크를 갖고 있는 분야다.

그림 - 유향네트워크



그림 – ‘IU’ 유향 네트워크 일부

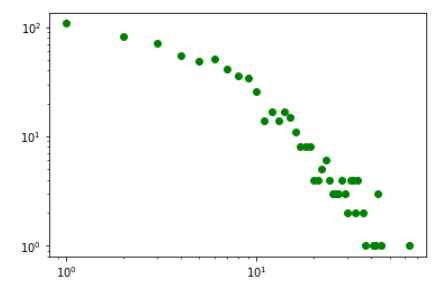
* IU는 작곡/작가사, 가수이면서 배우다. 가수 이외에도 박명수, 김수현, 현빈 등 다양한 직업을 가진 사람들(entertainer)과 관계가 있다.

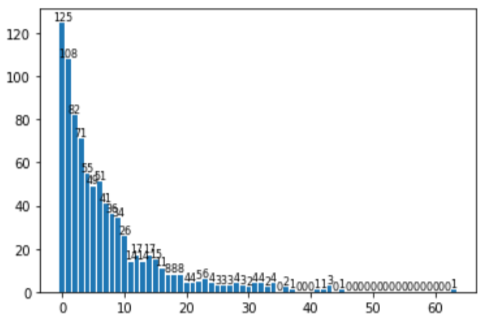
**2. 드라마의 영향력 및 중요도 분석**

**2-1. 차수 분포**

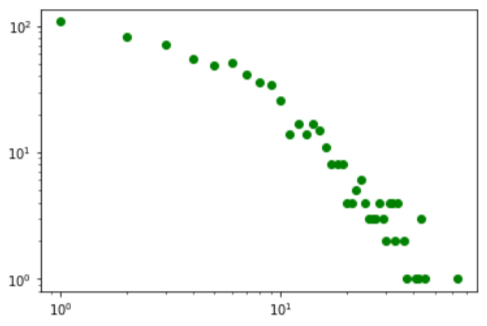
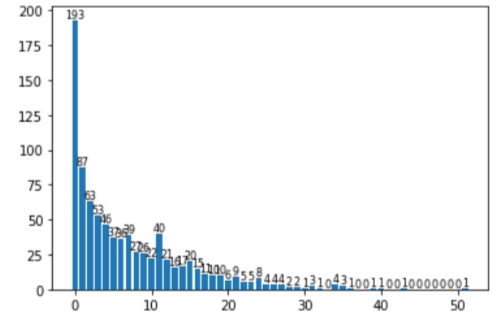
**:** 노드 차수 분포는 노드가 갖고 있는 링크의 개수(차수)를 나타낸다.

1. **내차수 분포**

* 내차수 값들의 top3은 63, 45, 43
* 차수 0을 갖는 노드의 개수가 123개로, 전체의 14.6%에 해당
* 내차수는 인기의 척도로, 차수가 63인 노드가 존재한다. 즉, 특히 인기가 많은 노드가 존재한다고 볼 수 있음.



1. **외차수 분포**

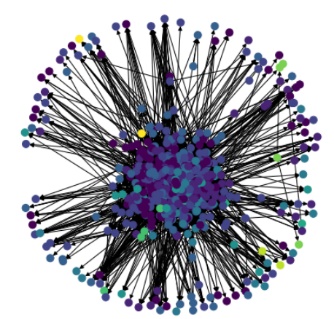
* 외차수 값들의 top3은 51, 43, 40
* 193개의 노드가 외차수 값 0을 갖는다. 즉, 내차수 값 0을 갖는 노드(125개)보다 많다.
* 내차수, 외차수 모두 Log-LogScale 결과, 기울기가 -1과 -3사이 직선식을 따르므로 각 네트워크의 노드 분포가 Power-Law 분포를 따른다.

**2-2. 노드 중심성 분석**

* 내향 연결 중심성, 외향 연결 중심성, 매개 중심성

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  | Indegree Centrality | | Outdegree Centrality | | Betweenness Centrality | |
| **1** | Super Junior | **0.074** | Brave Brothers | **0.060** | Zico | **0.031** |
| **2** | Jay Park | **0.053** | Verbal Jint | **0.050** | BoA | **0.030** |
| **3** | TVXQ | **0.050** | Crush | **0.047** | IU | **0.029** |
| **4** | Shinee | **0.050** | Mino | **0.046** | Brave Brothers | **0.025** |
| **5** | BoA | **0.050** | Zico | **0.042** | I.O.I | **0.023** |

* 내향 연결 중심성을 보면 슈퍼주니어, 박재범, 동방신기, 샤이니, 보아 등 인기가 많은 한류스타이며 박재범을 제외한 4명은 대기업인 sm소속이다.
* 외향 연결 중심성을 TOP5는 용감한 형제들, 버벌진트, 크러쉬, 송민호, 지코이며 작곡가 겸 가수다. 즉, 많은 다른 가수들에게 곡을 제공한다.
* 매개 중심성을 통해 zico, boa, IU, 용감한 형제들, IOI 등이 소수의 엣지만으로도 다른 노드들 사이를 매개하는 중개자임을 알 수 있다. 정보가 지나갈 때 정보를 거르고 왜곡할 수 있는 위치이므로 활동은 방해하거나 중단시킬 수 있는 잠재력을 가졌다. 실제로 Boa는 SM의 이사이며, zico와 용감한 형제는 연예 기획사 대표다.



**4. 한국뮤지션 커뮤니티 분석**

**4-1. 커뮤니티 분석**

* 커뮤니티 개수는 11개다.

**4-2. 워드클라우드 분석**

|  |  |  |
| --- | --- | --- |
| 0번 커뮤니티 | 4번 커뮤니티 | 10번 커뮤니티 |
|  |  |  |
| 싱글(single)이 특징 | 싱글이면서 그룹(group) | ost 뮤지션 |

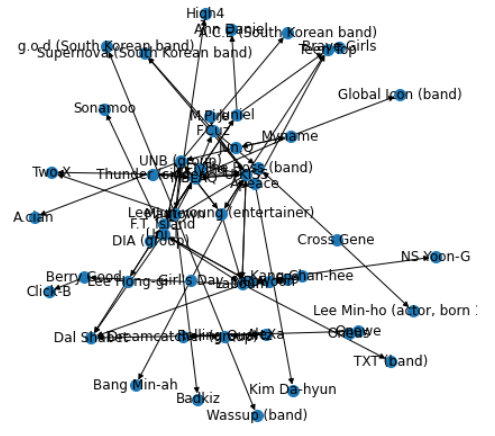


그림 - 4번 커뮤니티에 해당하는 부분 네트워크

* Ft-아일랜드와 이홍기, 걸스데이와 방민아, sf9과 로운 등 그룹과 솔로 활동을 같이 하는 사람들이 구성되어 있음.