

# Graph Centrality를 이용한 음원차트에 따른 프로듀서와 아티스트의 영향력 분석 및 신곡 추천

김건희(2017103958)

## 요약

Graph Centrality를 이용해 음원차트에 따른 국내 음악시장에서의 프로듀서와 아티스트 간의 사회연결망을 확인하고, 프로듀서와 아티스트의 각 관계가 어떤 영향을 끼치며, 어떠한 상호 관계가 있는지 분석한다. 이것을 토대로 신곡들 중 높은 순위를 차지할 가능성이 있는 음악들을 선별하는 것을 목표로 한다.

## 1. 서론

### 1.1 연구배경

최근 음원차트를 봤을 때 발견할 수 있는 점은 유명한 아티스트의 노래라고 하더라도 무조건 높은 순위를 차지하지는 않는다는 것이다. 많은 사람들이 알고 있는 발매 후 한 달이 지난 '트와이스'의 신곡 'Talk that Talk'은 9월 20일 기준 인기차트에서 32위를 차지하고 있다. 그에 반하여 상대적으로 인지도가 적은 '김승민'의 '내 기쁨은 너가 벤틀리를 끄는 거야'는 발매된 지 1년이 넘었지만 인기 차트 27위를 차지하고 있다. 이외에도 곡의 발매 시기와 관계없이 'sokodomo', '시온' 등 유명세가 낮은 아티스트의 곡들을 인기차트에서 꽤나 찾아볼 수 있다.

일반적으로 우리는 아티스트가 곡의 흥행을 좌우한다고 생각하는 경향이 있다. 그러나 위에서도 알 수 있듯이 꼭 그런 것만은 아니다. 프로듀서도 아티스트에 못지 않은 영향력을 끼친다. 음악은 기본적으로 아티스트와 프로듀서의 합작으로 이루어진다. 아티스트는 노래나 랩을 하고, 프로듀서는 비트를 제작하며 음원 제작을 총괄하는 역할을 맡는다.

이에 따라 본 프로젝트에서는 국내 음악의 다양한 분야에서 아티스트뿐만 아니라 프로듀서의 영향력이 음원 성적과 얼마나 긴밀한 연관이 있는지 Graph Centrality를 이용해 분석하고, 아티스트와 프로듀서 간의 상호 관계에 대해서도 연구한다. 이를 기반으로 신곡을 추천하는 것까지 나아갈 예정이다.

### 1.2 연구목표

국내에서 가장 이용자수가 많은 멜론 차트를 이용하여 댄스, 힙합, 인디 음악 장르 차트에서 지난 3년간의 월별 데이터를 수집하고 정제한다. 그 과정에서 Graph Centrality를 이용하여 가장 적합한 중심성을 찾는다. 그렇게 나온 중심성 수치를 통해 각 분야별 아티스트와 프로듀서의 영향력을 확인해보고, 그것이 실제 차트와 연관성이 있는지 확인한다. 이후 분석을 토대로 영향력이 높은 아티스트와 프로듀서를 소개하고, 신곡들 중 높은 음원 순위를 차지할 가능성이 있는 곡들을 추천한다.

## 2. 관련연구

### 2.1 Graph Centrality

그래프에서 중심성이란 그래프 혹은 사회 연결망에서 꼭짓점 혹은 노드의 상대적 중요성을 나타내는 척도이다. 이 중심성은 지수로 계산되는데 기준에 따라서 다양하다. 노드들 간의 최단 경로를 가지고 계산하는 매개 중심성(Betweenness centrality), 그래프의 노드와 다른 모든 노드들 사이의 최단 경로 길이의 합의 역수를 이용해서 구하는 근접 중심성(Closeness Centrality), 한 노드에 연결된 모든 엣지의 개수로 중심성을 평가하는 연결 중심성(Degree Centrality) 등 이 밖에도 다른 계산 방법들이 많이 존재한다[1].

### 2.2 Graph Centrality를 이용한 국내 힙합 프로듀서와 아티스트 간 영향력과 음원 성적 분석

지난 학기 캡스톤 디자인1에서 진행하였던 해당 연구에서는 Graph Centrality 중 가중 연결 중심성 (Weighted Degree Centrality)을 이용해 힙합 장르에서 각 아티스트와 프로듀서의 영향력을 파악했고, 이를 토대로 음원 성적 사이의 연관성을 분석했다.

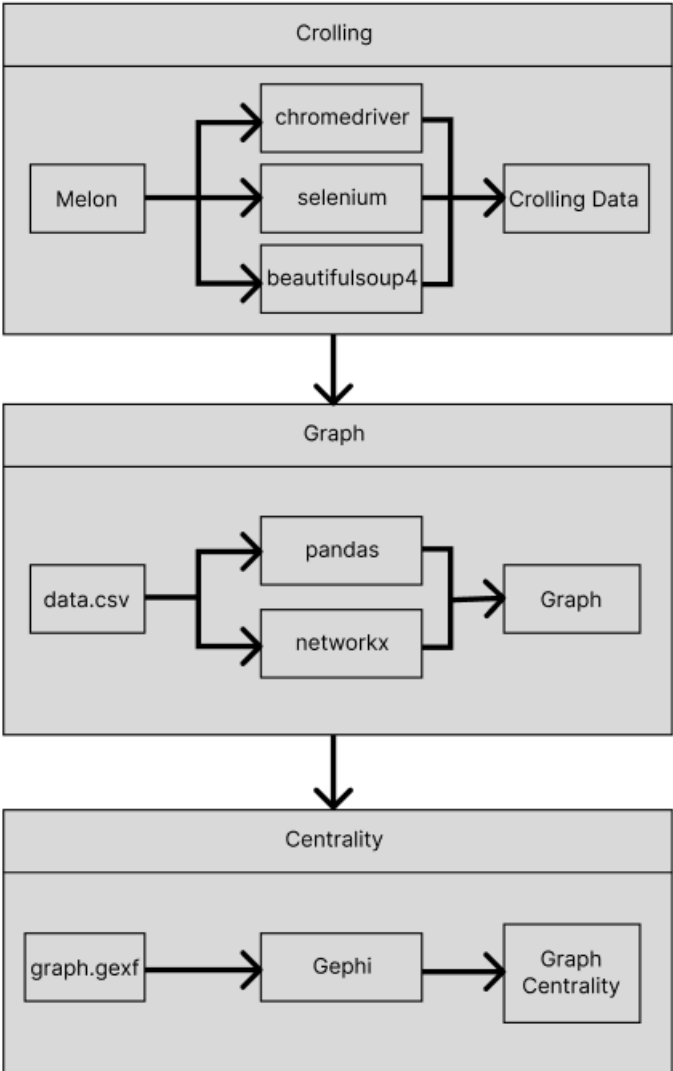
각 엣지에는 음원성적을 기반으로 가중치를 부여해 중심성을 도출해 시각화하였다. 그 결과, 해당 연도 장르에서 어떤 아티스트와 프로듀서가 높은 영향을 끼쳤는지 파악할 수 있었다. 아티스트는 프로듀싱까지 직접 하는 사람이 높은 영향력을 보였고, 프로듀서는 같은 회사 소속의 아티스트와 작업할 때 높은 영향력을 보였다[2].

### 2.3 기존 연구의 문제점과 해결 방안

기존 연구는 2020년 1월부터 12월까지의 데이터만 이용하였으므로 신뢰성이 상대적으로 떨어진다고 볼 수 있다. 그러므로 2018, 2019년도의 데이터도 이용할 것이다. 또한 기존 연구는 힙합 장르에만 한정하여 진행했기 때문에 이번 연구에서는 댄스 및 인디음악 분야로 확장하여 다른 분야에서는 아티스트와 프로듀서가 어떠한 특징을 보이는지 확인할 것이다.

### 3. 프로젝트 내용

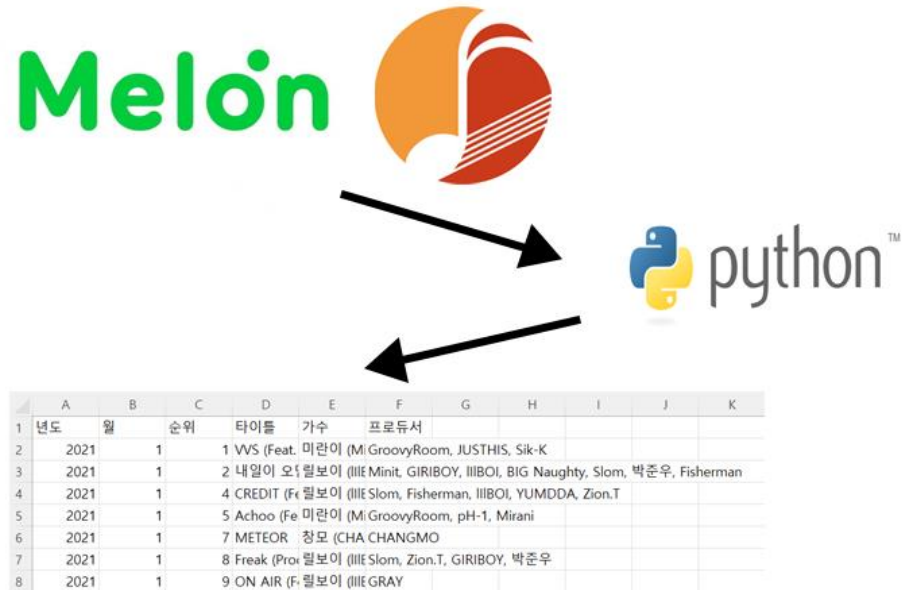
#### 3.1 시스템 설계



[그림 1] 프로젝트 모듈 구조

본 연구의 그래프 중심성을 구하는 과정은 [그림 1]과 같이 진행된다. Selenium과 BeautifulSoup4, 그리고 Chrome Driver를 이용하여 데이터 추출 과정을 거치고 Pandas와 NetworkX를 이용하여 노드와 엣지를 만들고 가중치를 설정한 그래프를 생성한다. 이후 Gephi를 통하여 Graph Centrality를 계산하는 과정을 거친다.

### 3.2 데이터 추출 및 정제



[그림 2] 데이터 추출 과정

데이터 분석에 필요한 데이터는 국내 최대 음원 사이트인 멜론에서 이용하였다. 월간 차트를 이용하여 2019년부터 2021년까지의 댄스, 힙합, 인디 분야의 차트 데이터를 크롤링하였다. 이후 각 장르와 적합하지 않은 곡들을 제외하는 정제 작업을 진행하였으며, 프로듀서 정보는 저작권협회에서 수집을 진행하였다.

### 3.3 그래프 생성

수집한 데이터를 분석하기 위해 ‘그래프’를 이용했다. 아티스트와 프로듀서를 노드로 설정하고, 같은 곡을 작업한 아티스트와 프로듀서를 엣지로 연결하였다. 그리고 엣지에는 음원차트를 기반으로 한 가중치를 부여하였다.

1. (가중치1) =  $[(1 - (\text{해당 곡의 순위}/101)^2) * 100]$
2. 전체 아티스트와 프로듀서의 엣지 가중치 중 최대, 최솟값 도출
3. (가중치2) =  $(\text{가중치1} - \text{최솟값}) / (\text{최댓값} - \text{최솟값})$
4. 최종 가중치 = 두 노드 간 (가중치2) 값의 총합

가중치 공식은 먼저 (가중치1) =  $[(1 - (\text{해당 곡의 순위}/101)^2) * 100]$ 로 설정하였다. 이를 바탕으로 작성된 전체 아티스트와 프로듀서의 엣지의 가중치 중 최대, 최솟값을 구하고 나서 (가중치2) =  $(\text{가중치1} - \text{최솟값}) / (\text{최댓값} - \text{최솟값})$ 의 식으로 가중치를 계산한다. 마지막으로 각 두 노드 간 협업된 곡들의 가중치2 값을 총합하여 최종 가중치를 구한다[2]. 가중연결 중심성(Weighted Degree Centrality), 고유벡터 중심성(Eigenvector Centrality)을 이용할 때는 해당 가중치를 그대로 사용하고, 매개 중심성(Betweenness Centrality), 근접 중심성(Closeness Centrality)을 사용할 때는 해당 가중치의 역수를 이용한다.

### 3.4 중심성 선정

#### 3.4.1 중심성 선정 과정

적합한 중심성 선정을 위하여 수집한 데이터에 대하여 매개 중심성(Betweenness Centrality), 근접 중심성(Closeness Centrality), 가중연결 중심성(Weighted Degree Centrality), 고유벡터 중심성(Eigenvector Centrality) 네 가지 중심성을 이용한 분석을 진행하였다.

#### 3.4.2 힙합 분야

힙합 장르는 '한국 힙합 어워즈'를 이용하여 올해의 아티스트, 올해의 프로듀서 두 가지 부분의 후보들의 중심성 순위 평균을 계산하여 적합한 중심성을 확인한 결과 아티스트는 가중연결 중심성(Weighted Degree Centrality)을, 프로듀서는 매개 중심성(Betweenness Centrality)을 사용하였을 때 가장 높은 정확도를 보였다.

#### 3.4.3 댄스 분야

댄스 장르는 'Mnet 아시안 뮤직 어워드'를 이용하였는데 3 아티스트는 남자 그룹상, 여자 그룹상, 남자 가수상, 여자 가수상 총 4가지 부분의 후보들의 중심성 순위 평균을 비교해보았고, 프로듀서는 베스트 프로듀서 부분을 이용하였다. 그 결과 아티스트와 프로듀서 모두 가중연결 중심성(Weighted Degree Centrality)을 이용하는 것이 더 높은 정확성을 보였다.

#### 3.4.4 인디 분야

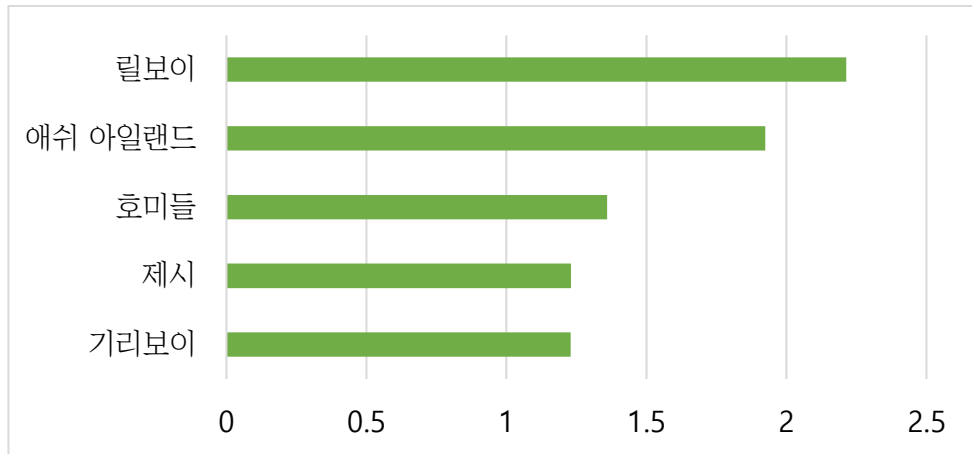
인디 장르는 '한국대중음악상'의 최우수 모던록-노래와 음반 부문, 그리고 '멜론 뮤직 어워드'의 TOP10과 인디 부문 장르 상을 이용하여 적합한 중심성을 선정하였다. 이를 통하여 인디 부분에서도 가중연결 중심성(Weighted Degree Centrality)이 더 높은 정확도를 보여줌을 확인할 수 있었다.

#### 3.4.5 최종 중심성

신곡은 다음과 같은 기준으로 선별한다. 연도가 가까울수록 현재 영향력이 더 높다고 생각하기 때문에 공식을  $(\text{영향력}) = [(2019\text{년의 중심성}) \times 0.2 + (2020\text{년의 중심성}) \times 0.3 + (2021\text{년의 중심성}) \times 0.5]$ 로 세웠다. 해당 식을 이용하여 아티스트나 프로듀서의 최종 중심성 수치를 계산한다.

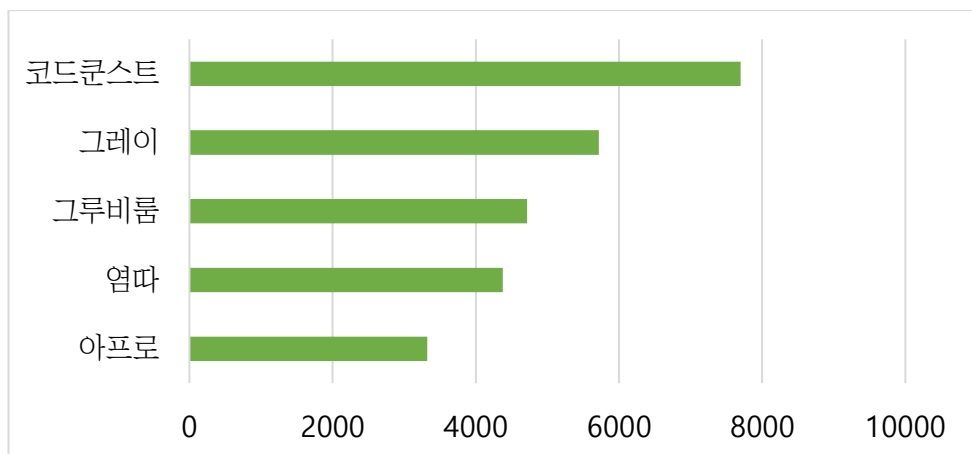
### 3.5 중심성 분석

#### 3.5.1 힙합 분야



[표 1] 힙합 아티스트 가중연결 중심성(Weighted Degree Centrality) 상위 5명

가중연결 중심성(Weighted Degree Centrality)을 이용하여 연도별로 중심성 상위권에 있는 힙합 아티스트들을 확인해보면 2가지 특징을 발견할 수 있었다. 첫 번째 특징은 해당 아티스트들의 차트인한 곡들은 셀프 프로듀싱하는 경우가 많다는 점이다. 그리고 또 다른 특징으로는 셀프 프로듀싱을 하지 않는 아티스트의 경우 작업을 자주 같이 진행하는 특정 프로듀서가 존재한다는 것이다.

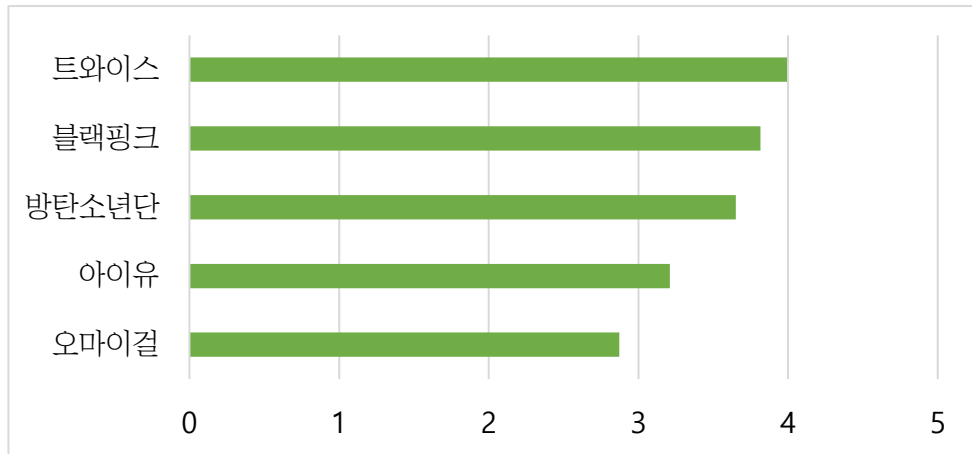


[표 2] 힙합 프로듀서 매개 중심성(Betweenness Centrality) 상위 5명

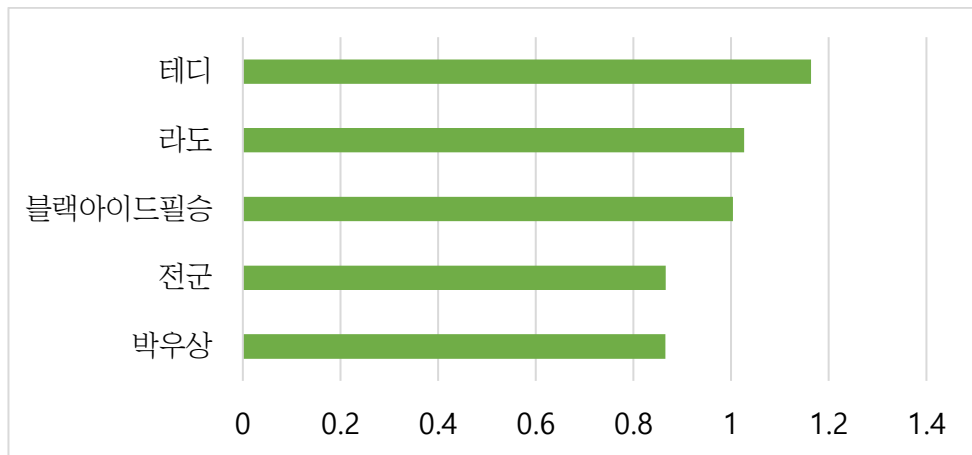
매개 중심성(Betweenness Centrality)을 이용하여 연도별로 중심성 상위권인 힙합 프로듀서들을 확인해보았을 때도 마찬가지로 두 가지 특징을 확인할 수 있었다. 하나는 프로듀서들의 차트인 한 곡들은 같은 회사 소속의 아티스트와 작업한 경우가 많다는 것이고, 다른 하나는 아티스트 역할도 하는 프로듀서들은 셀프 프로듀싱을 한 곡들이 많이 차트인 했다는 점이다. 그러나 매개 중심성(Betweenness Centrality)을 이용한 만큼 셀프 프로듀싱을 하는 프로듀서보다 그렇지 않은 프로듀서의 등수가 상대적으로 더 높았다.

아티스트와 프로듀서가 서로 적합한 중심성이 다른 이유를 생각해보자면 아티스트는 일차적으로 프로듀서와 연결되는데, 하나의 아티스트가 많은 프로듀서와 연결되기는 어렵다. 그렇기 때문에 인기가 많은 아티스트라고 하더라도 노드 간의 최단 거리에 포함되기는 어려우며, 한 노드에 연결된 모든 엣지의 개수로 중심성을 평가하는 가중연결 중심성(Weighted Degree Centrality)이 더 적합하였던 것으로 예상된다. 반대로 프로듀서는 상대적으로 많은 아티스트와 작업하기 수월하며 이에 따라 노드 간의 최단 거리에 포함되기 상대적으로 쉬울 것으로 예상된다. 그러므로 매개 중심성(Betweenness Centrality)이 적합하였다고 생각된다. 하지만 프로듀서 부문에서 매개 중심성(Betweenness Centrality)과 가중연결 중심성(Weighted Degree Centrality)이 큰 차이를 보이지는 않았다.

### 3.5.2 댄스 분야



[표 3] 댄스 아티스트 가중연결 중심성(Weighted Degree Centrality) 상위 5명

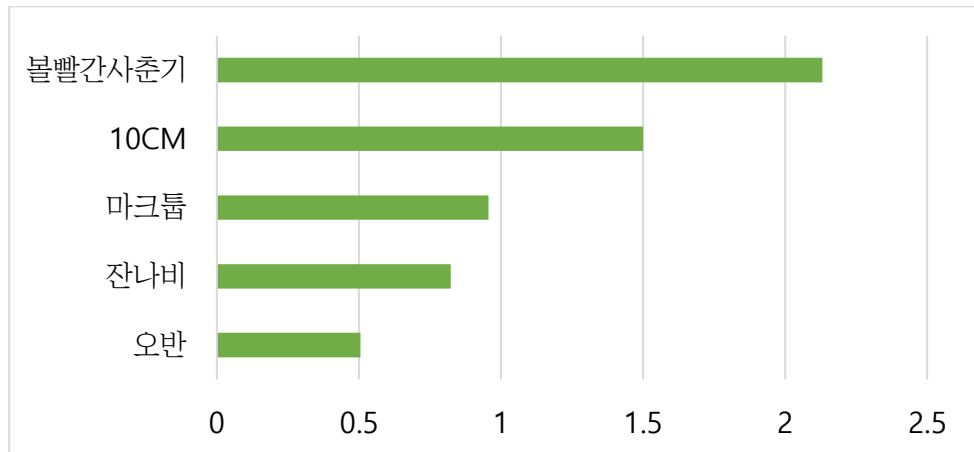


[표 4] 댄스 프로듀서 가중연결 중심성(Weighted Degree Centrality) 상위 5명

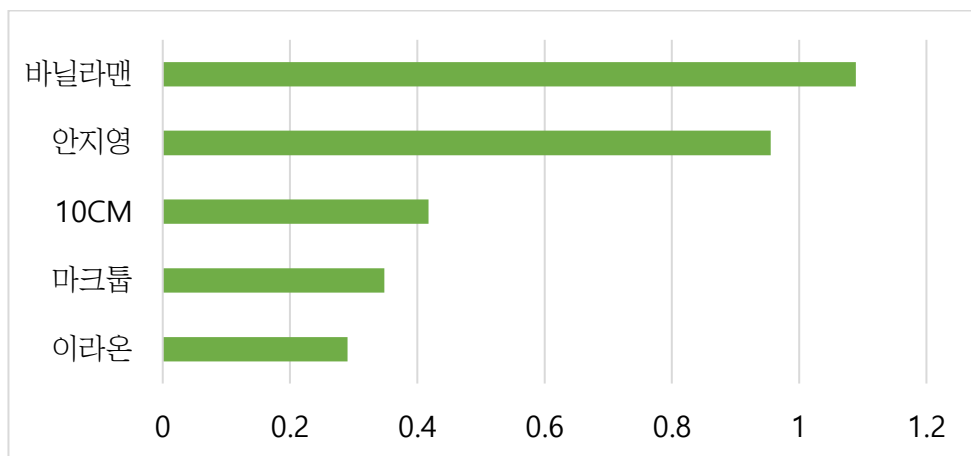
다음으로는 가중연결 중심성(Weighted Degree Centrality)을 이용하여 댄스 분야의 아티스트와 프로듀서를 연도별로 상위권을 뽑아서 분석해보았다. 발견한 점은 아티스트와 프로듀서 모두 차트인 곡들을 보았을 때 주로 함께 작업하는 특정 프로듀서나 아티스트가 존재한다는 점이다.

댄스 분야에서 유명한 프로듀서는 같은 회사 소속의 아티스트가 아닌, 다른 회사 소속의 아티스트와 작업하는 경우도 더러 있기 때문에 한 노드에 연결된 모든 엣지의 개수로 중심성을 평가하는 가중연결 중심성(Weighted Degree Centrality)이 더 적합하였다고 판단된다.

### 3.5.3 인디 분야



[표 5] 인디 프로듀서 가중연결 중심성(Weighted Degree Centrality) 상위 5명



[표 6] 인디 프로듀서 가중연결 중심성(Weighted Degree Centrality) 상위 5명

마지막으로 가중연결 중심성(Weighted Degree Centrality)을 이용하여 인디 분야 아티스트와 프로듀서를 분석해보았다. 이전 힙합 분야와 마찬가지로 상위권의 아티스트들은 셀프 프로듀싱하는 경우가 많고, 작업을 함께 하는 특정 프로듀서가 존재한다. 프로듀서 역시 작업을 같이하는 특정 아티스트가 존재하거나, 셀프 프로듀싱을 한 경우가 다수였다.

인디 분야에서는 함께 작업하는 특정 아티스트와 특정 프로듀서가 지속해서 곡을 만드는 경향이 다른 분야에 비해 더 심하였다. 이에 따라 노드 간의 최단 거리에 포함되기 어렵고, 가중연결 중심성(Weighted Degree Centrality)이 더 영향력 높은 모습을 보인다고 생각된다.



## 3.6 신곡 추천

### 3.6.1 추천 신곡 선정

추천 신곡을 선정하는 과정은 다음과 같다. 먼저 해당 주차의 신곡 데이터를 벅스에서 수집한다. 모든 음원 차트에서 상세 페이지에 들어가지 않는다면 곡의 프로듀서 정보는 제공하지 않는다. 그렇기 때문에 사람들은 제목과 아티스트만 보고 신곡을 듣는 경우가 대부분이라고 생각하였다. 이러한 생각을 바탕으로 일차적으로 아티스트가 최종 중심성 리스트에 들어가 있는지 확인하여 따로 정제한다. 이후 정제된 곡에 대한 프로듀서 정보를 수집한 뒤 아티스트와 프로듀서 중심성을 합쳐 순위를 매긴다.

### 3.6.2 신곡 추천 검증 과정

추천 신곡에 대한 검증은 다음과 같이 진행하였다. 먼저 해당 곡이 발매된 주차와 그다음 주차의 순위, 그리고 발매 4주 후 주간 순위 데이터를 수집하였다. 이후 주차 별로 예측 순위와 실제 순위를 비교 및 분석해보았다.

### 3.6.3 신곡 추천 분석

#### 3.6.3.1 힙합 분야

먼저 힙합 분야의 예측 결과에 대해서 분석해보았다. 힙합 분야에서는 아티스트는 가중연결 중심성(Weighted Degree Centrality), 프로듀서는 매개 중심성(Betweenness Centrality)이 더 적합하였다. 그러나 서로 점수 차이가 너무 많이 나기 때문에 보정하기에도 어려움이 있었다. 프로듀서 부문 분석에서 가중연결 중심성(Weighted Degree Centrality)과 매개 중심성(Betweenness Centrality)이 약간의 차이를 보였기 때문에 신곡 추천에는 프로듀서 역시 가중연결 중심성(Weighted Degree Centrality)을 사용하기로 하였다.

가중연결 중심성(Weighted Degree Centrality)을 이용한 분석 결과 실제로 해당 주차의 신곡 추천 순위가 높았던 곡들이 발매 1, 2주 차에 차트인 하는 경우가 많았다. 차트인 한 곡 중 차트인을 예측하지 못한 곡은 77곡 중 6곡 밖에 없었다. 발매 이후 4주 차의 주간 차트 순위도 분석해보았는데 총 41곡이었고, 그중 예측하지 못한 곡은 단 4곡밖에 없었다. 특히 예측한 37곡 중 30곡이 모두 중심성 수치가 1이 넘었다.

#### 3.6.3.2 댄스 분야

댄스 분야를 가중연결 중심성(Weighted Degree Centrality)을 이용하여 분석 결과 총 107개의 차트인한 곡 중 84개의 곡을 선별하였으나 점수대는 오락가락하였다. 84곡 중 중심성 수치가 1점 아래인 곡들이 49곡이나 되었다. 발매 4주 뒤 차트에서도 순위를 유지한 곡은 66곡이었으나 실제 예측은 48곡만 성공하였다. 그중 절반인 24곡이 1점 아래였다.

힙합 분야와 상대적으로 예측 정확도도 떨어지고 중심성 수치가 제대로 반영되지 않은 현상의 원인으로는 먼저 최근 ‘아이브’, ‘르세라핌’, ‘뉴진스’

등과 같은 대형 신인 아이돌이 많이 나왔기 때문으로 보인다. 신인 아이돌은 최근 인기가 많더라도 지난 3년간의 데이터에 없을 가능성이 높고, 있더라도 중심성 수치가 낮을 것이기 때문에 위와 같은 현상이 나타났을 것으로 예상된다. 두 번째로는 솔로로 활동을 시작한 사람도 많기 때문이다. 기존의 그룹 활동이 아닌 솔로 활동한다면 당연히 지난 3년간의 데이터에 잡히지 않을 것이기 때문이다.

### 3.6.3.3 인디 분야

인디 분야를 가중연결 중심성(Weighted Degree Centrality)을 이용하여 분석 결과 총 61개의 차트인한 곡 중 45개의 곡을 추천한 것으로 확인된다. 그러나 추천곡들의 중심성 수치를 본다면 점수가 엄청 낮은 것을 알 수 있다. 45개의 곡 중 1점을 넘는 곡은 단 13곡밖에 되지 않는다. 발매 4주 이후의 주간 차트에서는 차트인 한 33곡 중 29곡을 예측하였다. 그러나 이 역시 1점을 넘는 곡은 10곡밖에 되지 않는다.

이러한 결과가 나온 이유로는 특정 사람들과 작업하는 경향이 인디 분야에서 특히 자주 보이기 때문이라고 예상한다. 특정 사람들이랑만 작업을 한다면 연결되는 노드가 제한적이기 때문에 가중연결 중심성(Weighted Degree Centrality)을 분석에 부정적인 요소가 된다.

## 4. 결론 및 향후연구

### 4.1 결론

Graph Centrality 중 네 가지 중심성을 이용하여 적합한 중심성을 판단하였고 힙합 분야의 프로듀서는 매개 중심성(Betweenness Centrality)을 이용하는 것이, 나머지는 모두 가중연결 중심성(Weighted Degree Centrality)을 사용하여 분석하는 것이 가장 적합하다는 것을 알 수 있었다.

이후 해당 중심성을 바탕으로 분석을 진행한 결과 분야와 관계없이 공통적인 특징 하나를 발견할 수 있었다. 차트인을 자주 하는 영향력 높은 아티스트나 프로듀서들은 함께 작업하는 특정 프로듀서나 아티스트가 존재한다는 점이다. 또 힙합 분야와 인디 분야에서 높은 영향력을 보이는 아티스트와 프로듀서는 셀프 프로듀싱을 하는 경우에 차트인하는 경우가 많다는 점 또한 알 수 있었다. 그리고 분야별 선택된 중심성이 적합한 이유에 대해서도 분석했다.

신곡 추천 과정에서는 모두 가중연결 중심성(Weighted Degree Centrality)을 이용하여 진행하였으며, 힙합 분야에서는 차트인할 곡을 매우 잘 예측하였으며 중심성 수치 역시 곡마다 적절하였다. 다른 분야에서는 중심성 수치는 부정확한 부분이 있었으나 힙합 분야만큼은 아니더라도 어느 정도 높은 예측률을 보여 주었다.

## 4.2 향후 연구

향후 연구로는 새롭게 데뷔하는 아티스트나 프로듀서들에 대한 정보를 빠르게 반영하기 위하여 1년 단위로 데이터를 수집하는 것에서 반년이나 분기 단위로 데이터를 수집하는 것으로 변경하여 연구를 진행하는 것이다. 그렇게 한다면 댄스 분야에서 문제가 되었던 새롭게 데뷔하거나, 솔로 데뷔하는 사람들에 대한 부분들을 놓치지 않을 수 있을 것이다.

이 외에도 매개 중심성(Betweenness Centrality), 근접 중심성(Closeness Centrality), 가중연결 중심성(Weighted Degree Centrality), 고유벡터 중심성(Eigenvector Centrality) 네 가지 중심성뿐만 아니라 다른 중심성까지 확인해보아 댄스 분야나 인디 분야에 더욱 적합한 중심성을 찾는 연구도 진행할 수 있을 것이다.

## 5. 참고문헌

[1] 조태수, 한치근, 이상훈, “그래프 중심성들을 이용한 그래프 유사도 측정”, 한국컴퓨터정보학회논문지, 2018.

[2] 김건희, 송근영, 한치근, “Graph Centrality를 이용한 국내 힙합 프로듀서와 아티스트 간 영향력과 음원 성적 분석”, 한국정보과학회 한국컴퓨터종합학술대회, 2022.

[3] <https://gephi.org/>