**Graph Centrality 이용한 음원차트에 따른**

**국내 힙합 프로듀서와 아티스트의 영향력 분석**

김건희(2017103958)

송근영(2017103997)

**요약**

최근 세계적인 추세에 따라 우리 나라에도 힙합 장르의 영향력이 점점 커지고 있다. 이러한 힙합 분야는 아티스트와 프로듀서의 역할이 모두 중요하다. 그러므로 Graph Centrality를 이용해 국내 랩/힙합 분야의 프로듀서와 아티스트의 영향력을 분석한다. 그 중 가중연결 중심성(Weighted Degree Centrality)을 이용하며 그에 따른 가중치는 2차원 함수를 사용한다. 분석 결과 아티스트의 경우 셀프 프로듀싱을 하는 아티스트들이 높은 영향력을 보였고, 프로듀서의 경우 같은 회사 소속이거나 친분이 있는 아티스트와 작업을 많이 하는 프로듀서가 높은 영향력을 보였다.

**1. 서론**

**1.1 연구배경**

최근 음원차트를 보면 눈에 띄는 장르가 있다. 세계적인 음원차트 빌보드에서는 힙합 장르가 주를 이룬 지 꽤 되었고, 국내 음원차트 역시 몇 년 사이 힙합 음악의 영향력이 커졌다. 국내 최대 음원 스트리밍 서비스 ‘멜론’에서 집계한 통계를 보면 2020년도 1, 2위를 힙합 장르가 차지했으며, 2021년 또한 4위를 차지했다. 이 밖에도 랩이 들어가는 노래는 상당히 많다.

일반적으로 아티스트가 곡의 흥행을 좌우한다고 생각하지만, 프로듀서도 그에 못지 않은 영향력을 끼친다. 힙합 음악은 기본적으로 아티스트와 프로듀서의 합작으로 이루어진다. 아티스트는 랩을 하고, 프로듀서는 비트를 제작하며 음원 제작을 총괄한다.

이에 따라 본 프로젝트에서는 국내 힙합 장르에서 아티스트뿐 아니라 프로듀서의 영향력이 음원 성적과 얼마나 긴밀한 연관이 있는지 Graph Centrality를 이용해 분석하고, 아티스트와 프로듀서 간의 상호 관계에 대해서도 연구한다.

**1.2 연구목표**

연구를 통해서 얻고자 하는 결론은 크게 두 가지다. 첫 번째, Graph Centrality를 이용해 힙합 장르에서 프로듀서와 아티스트의 영향력을 파악한다. 두 번째, 영향력 높은 아티스트와 프로듀서들이 어떤 특색을 지니며, 어떠한 연관성을 보이는지 파악한다

**2. 관련연구**

**2.1 Graph Centrality**

Graph Centrality이란 그래프 혹은 사회 연결망에서 꼭짓점 혹은 노드의 상대적 중요성을 나타내는 척도이다. 이 중심성은 지수로 계산되는데 중점으로 두는 기준에 따라 출력되는 결과값이 다르다.

노드들 사이의 최단 경로를 가지고 계산하는 매개 중심성(Betweenness Centrality), 그래프의 노드와 다른 모든 노드들 사이의 최단 경로 길이 합의 역수를 이용해서 구하는 근접 중심성(Closeness Centrality), 한 노드에 연결된 모든 엣지의 개수로 중심성을 평가하는 연결 중심성(Degree Centrality) 등 이 밖에도 다른 계산방법들이 다양하게 존재한다[1].

**2.2 기존 연구의 문제점과 해결 방안**

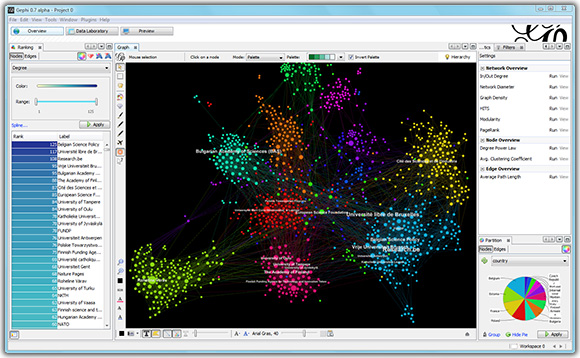
기존 연구는 데이터를 수집하여 기사의 개수와 음원차트 사이의 상관관계를 분석하였다.[2] 빅데이터 분석 기술 중 하나인 오피니언 마이닝을 사용해 연구를 진행했다. 결론은 기사 수와 음원차트 순위 간의 연관성은 존재하지만 높은 수준은 아니라고 밝혔다.

본 프로젝트는 오피니언 마이닝 기법만을 사용해 음원차트를 분석한 이전 프로젝트에 머물지 않고, Graph Centrality를 이용해 음원차트를 분석하기로 했다. 이에 따라 아티스트뿐 아니라 프로듀서 또한 음원차트 순위에 영향을 준다 보았고, 힙합 장르에서 아티스트와 프로듀서의 영향력이 음원차트와 얼마나 긴밀한 연관이 있는지 분석할 것이다.

**2.3 Gephi**

대규모 그래프를 시각화하고 분석하기 위한 오픈 소프트웨어이다.[3] 3D 렌더링 엔진을 사용해 실시간으로 그래프를 표시하고 탐색 속도를 향상해준다. 이를 사용해 음원차트 기반으로 제작한 아티스트, 프로듀서 그래프를 탐색, 분석, 조작이 가능하다.

**[그림 1] Gephi 화면 예시**



**3. 프로젝트 내용**

**3.1 중심성 선정**

**3.1.1 아티스트 측면**

**[표 1] 아티스트 매개 중심성(Betweenness Centrality) 상위 5명**

위는 매개 중심성(Betweenness Centrality) 분석 결과 중 아티스트 top 5를 나타낸 것이다.

**[표 2] 아티스트 가중연결 중심성(Weighted Degree Centrality) 상위 5명**

위는 가중연결 중심성(Weighted Degree Centrality) 분석 결과 중 아티스트 top 5를 나타낸 것이다. 매개 중심성(Betweenness Centrality)을 이용했을 때와 비교해보면 기리보이를 제외하면 top 5 중 4명이 바뀐 것을 확인할 수 있다.

새로 들어온 창모, 염따, pH-1, 더 콰이엇 모두 차트에 들어간 곡 중 셀프 프로듀싱을 한 곡들이 다수이며 곡들의 순위도 높은 것을 확인할 수 있다. 노드들 사이의 최단 경로를 가지고 계산하는 매개 중심성(Betweenness Centrality) 분석은 셀프 프로듀싱을 거의 반영하지 못하기 때문에 이러한 차이가 나는 것으로 보인다.

**3.1.2 프로듀서 측면**

**[표 3] 프로듀서 매개 중심성(Betweenness Centrality) 상위 5명**

위는 매개 중심성(Betweenness Centrality) 분석을 결과 중 프로듀서 top 5를 나타낸 것이다.

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

위는 가중연결 중심성(Weighted Degree Centrality) 분석 결과 중 프로듀서 top 5를 나타낸 것이다. 프로듀서에는 아프로와 그레이가 빠지고 NOISEMASTERMINSU, 창모가 top5에 들어온 것을 확인할 수 있다. 창모는 프로듀서로 참여해서 차트에 들어온 곡들 모두 셀프 프로듀싱을 하였고, NOISEMASTERMINSU는 차트인 한 곡들이 특정 소수의 아티스트들과 작업을 진행하였다. 특히 창모는 곡들의 순위도 높은 편이다. 그러나 매개 중심성(Betweenness Centrality) 분석에서는 이것들이 제대로 반영되지 못하였음을 알 수 있다. 위의 결과로부터 알 수 있듯이 매개 중심성(Betweenness Centrality)은 셀프 프로듀싱하는 경우, 소수의 사람과 작업하는 경우에 대하여 제대로 영향력을 반영하지 못한다. 그렇지만 가중연결 중심성(Weighted Degree Centrality)은 이러한 사항들을 적절하게 반영하기 때문에 가중연결 중심성(Weighted Degree Centrality)을 채택했다.

**3.2 가중치 선정**

**3.2.1 아티스트 측면**

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

위는 1차원 가중치 식을 사용해 진행한 가중연결 중심성 분석 중 아티스트 top 5를 나타낸 것이다.

**[표 6] 2차원 아티스트 가중연결 중심성(Weighted Degree Centrality) 상위 5명**

위는 2차원 가중치 식을 사용하여 진행한 가중연결 중심성 분석 중 아티스트 top 5를 나타낸 것이다. 염따와 창모의 순위가 바뀐 것을 알 수 있는데, 순위 측면에서 보면 창모는 매월 10위 안에 곡이 들어가는데 성공하였으나, 염따는 1, 2, 3월에만 10위 안에 곡이 들었고 나머지 기간에는 점점 곡들의 순위가 줄어들었다. 곡 순위의 중요성이 높아진 것을 확인할 수 있다.

**3.2.2 프로듀서 측면**

**[표 7] 1차원 프로듀서 가중연결 중심성(Weighted Degree Centrality) 상위 5명**

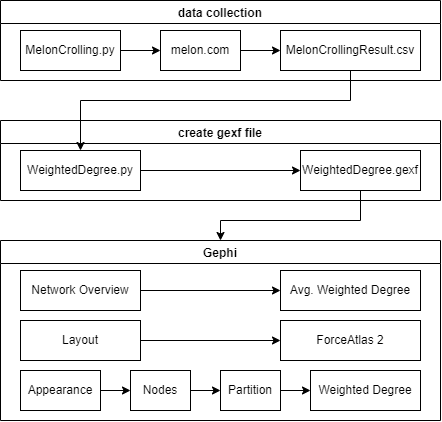
위는 1차원 가중치 식을 사용해 진행한 가중연결 중심성 분석 중 프로듀서 top 5를 나타낸 것이다.

**[표 8] 2차원 프로듀서 가중연결 중심성(Weighted Degree Centrality) 상위 5명**

위는 2차원 가중치 식을 사용하여 진행한 가중연결 중심성 분석 중 프로듀서 top 5를 나타낸 것이다. 여기서 발견할 수 있는 것은 김효은이 순위에서 밀려나고 코드쿤스트가 순위에 든 것이다. 김효은은 해당 기간에 본인이 프로듀서로 참여한 곡이 20위 안에 들었던 적은 5번뿐이었다. 그러나 코드쿤스트는 20위 안에 9번을 랭크시켰으며, 이를 통하여 2차원 가중치 식을 사용함으로써 곡 순위의 중요성이 높아졌다는 것을 확인할 수 있다.

1차원 가중치 식과 2차원 가중치 식을 비교해보면서 아티스트 측면으로 보나 프로듀서 측면으로 보나 곡의 중요성이 더 높아진 것을 확인할 수 있었다. 또한 1차원 식을 사용한다면 1위를 한 번 한 사람과 50위를 2번 한 사람의 가중치가 동일하다는 문제가 생긴다. 위와 같은 이유들 때문에 높은 순위의 곡의 영향력을 늘리기 위하여 가중치 식으로 2차원 함수를 사용하였다

**3.3 시스템 설계**



**[그림 2] 프로젝트 모듈 구조**

본 연구의 프로젝트 모듈 구조는 [그림 1]과 같다. 모듈의 처리 순서는 데이터 수집 단계, 데이터 정제 단계, 그래프 생성 및 시각화 단계 순으로 진행된다. 웹 크롤링을 통해 음원 서비스에서 데이터를 수집한다. 이후, 수집된 데이터를 시각화 프로그램 Gephi에 적합한 형태로 정제 및 가공하는 과정을 거친다. 그렇게 그래프 정보를 포함해 생성된 파일은 Gephi에서 가중연결 중심성(Weighted Degree Centrality)을 이용해 분석하고 원하는 결과를 도출한다.

**3.4 데이터 추출 및 정제**

음원 차트 제공 서비스로는 국내 최대 음원 사이트 멜론을 사용했다. 현재 음원 스트리밍 서비스 중 최대 사용자를 보유하고 있으며, 집계하는 방식을 명백히 밝히고 있기 때문에 신뢰성이 있다. 멜론에서는 1개월 단위로 장르별 차트를 제공한다. 본 연구에서 그래프의 데이터 대규모 수집과 정확도 향상을 위하여, 2020년 1월부터 12월까지 총 12개월의 데이터를 수집하여 정제 작업을 실행했다.

테이블이(가) 표시된 사진

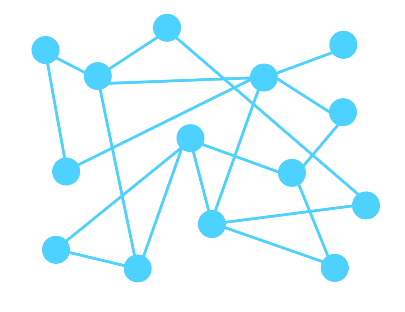
자동 생성된 설명

**[그림 3] 추출한 데이터의 일부**

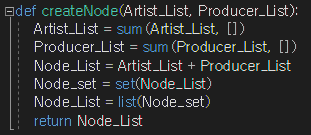
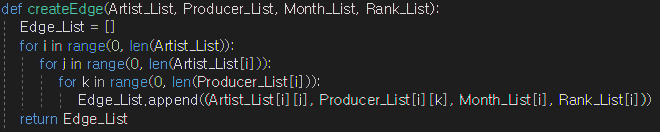
프로그래밍을 통해 웹사이트를 크롤링해 데이터를 수집했다. 멜론에서는 달, 음원차트, 제목, 아티스트 정보를 추출했고, 한국저작권협회에서 프로듀서의 데이터를 추출 후 정제하였다. 정제 후 노드의 개수는 276개, 엣지의 개수는 450개가 되었다. 추출된 데이터는 위와 같이 csv파일로 저장돼 있다.

**3.5 그래프 생성**

앞서 웹 크롤링 프로그램을 통해 월, 순위, 아티스트, 프로듀서의 정보를 포함한 csv파일을 생성했다. 이러한 데이터를 시각화하고, 보다 정확하게 분석하기 위해 가장 적합한 자료구조 ‘그래프’를 채택했다. 그래프는 사회 연결망 분석을 위해 사용된다. 국내 힙합 장르의 아티스트, 프로듀서를 그래프를 이용해 하나의 사회 연결망을 구축한다.



[그림 4] 그래프 생성 과정

파이썬을 이용해 호출된 csv파일에서 음원 차트 정보를 가져온다. 아티스트와 프로듀서를 중복되지 않게 각각 노드로 생성해주었고, 이들이 같은 곡을 작업했을 경우 두 노드를 엣지로 이어준다.

[그림 6] 엣지 생성 코드

[그림 5] 노드 생성 코드

또한, 이들의 영향력을 파악하고 음원 성적과의 상관 관계를 분석하는 것이 목적이기 때문에 엣지에 음원차트를 기반으로 한 가중치를 부여해준다. 앞선 비교를 통해 선택된 가중연결 중심성(Weighted Degree Centrality) 가중치를 도출하는 과정은 이와 같다.

1. (가중치1)=[(1-(해당 곡의 순위/101)^2)\*100]

2. 전체 아티스트와 프로듀서의 엣지 가중치 중 최대, 최소값 도출

3. (가중치2)=(가중치1-최소값)/(최대값-최소값)

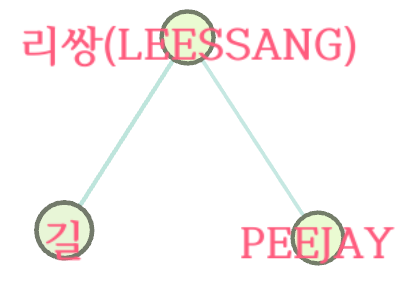
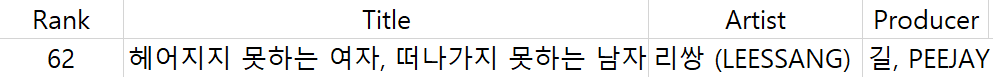
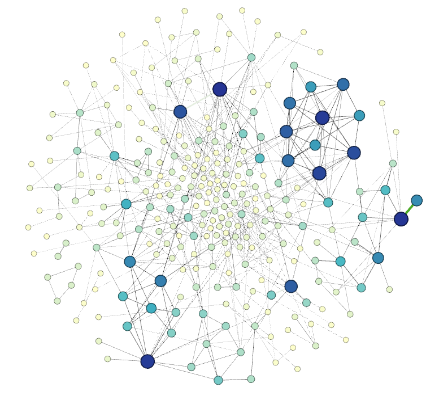
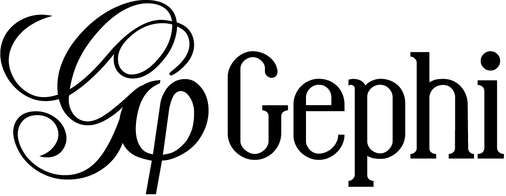
4. 최종 가중치 = 두 노드 간 (가중치2) 값의 총합``

(가중치1)=[(1-(해당 곡의 순위/101)^2)\*100]로 설정하여 임시 가중치1을 도출한다. 그리고 아티스트 및 프로듀서 각각 (가중치1)을 모두 합친다. 이를 바탕으로 작성된 전체 아티스트와 프로듀서의 엣지의 가중치 중 최대, 최소값을 구하고 나서 (가중치2)=(가중치1-최소값)/(최대값-최소값)의 식으로 (가중치2)를 구한다. 마지막으로 각 두 노드 간 협업된 곡들의 가중치2 값을 총합하여 최종 가중치를 구한다. 이후, 해당 노드와 엣지의 가중치 정보를 포함한 리스트를 ‘Gephi’에 적합한 확장자인 ‘.gexf’ 파일로 변형한다. 생성된 ‘WeightedDegree.gexf’ 파일은 아티스트, 프로듀서, 가중치를 포함한 엣지 데이터를 포함한다.

**3.6 그래프 시각화**

생성된 .gexf파일은 Gephi를 통해 시각화된다.

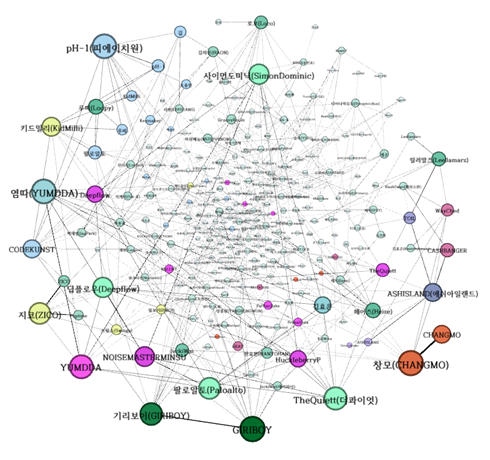
[그림 7] 그래프 시각화 과정



[그림 8] 그래프 생성 예시

예를 들어, ‘헤어지지 못하는 여자, 떠나가지 못하는 남자’라는 음원을 발매한 리쌍이라는 가수가 있다. 이 곡을 길과 PEEJAY가 프로듀싱했다면, 생성된 그래프는 이와 같다.

노드의 개수는 총 276개, 엣지의 개수는 총 450개가 생성됐다. 생성된 그래프는 아래와 같다.



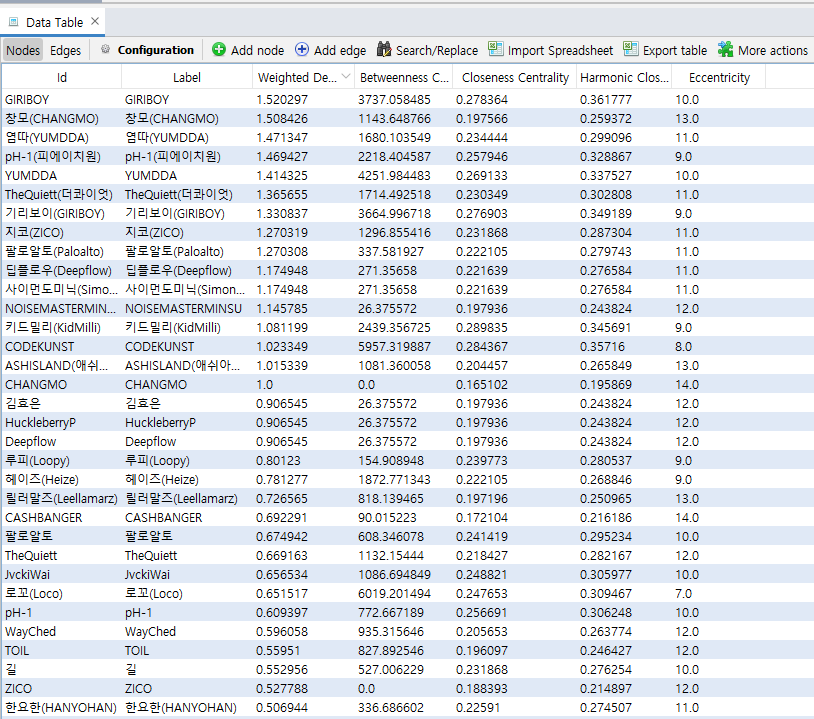
[그림 9] 그래프를 Gephi를 통해 시각화한 결과

앞서 분석했던 가중연결 중심성(Weighted Degree Centrality)을 보다 손쉽게 파악하기 위해 시각화하였다. 노드의 크기는 각 프로듀서의 영향력을, 엣지의 굵기는 음원 차트의 성적을 의미한다. 프로듀서의 영향력이 높을수록 노드의 크기가 커지며, 아티스트와 프로듀서 간 높은 성적을 거뒀을 경우 굵은 엣지가 나타난다.

**3.7 그래프 중심성 분석**

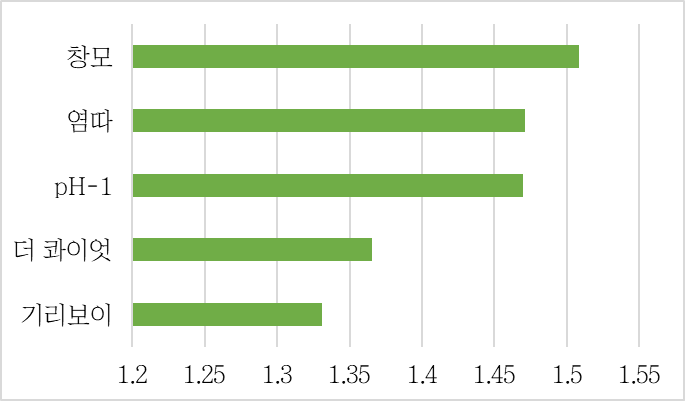
다음은 그래프의 중심성을 분석한 결과이다. Gephi에서는 그래프 시각화를 진행해주는 것뿐 아니라, 생성된 그래프의 가중치를 기반으로 각종 중심성 수치를 도출해준다. 해당 연구에서는 가중연결 중심성(Weighted Degree Centrality) 수치를 분석한다.

[그림 10] Gephi를 통해 도출된 중심성 표



도출된 가중연결 중심성(Weighted Degree Centrality) 수치를 아티스트와 프로듀서로 나누어 나열하였다. 먼저, 아티스트를 분석한 결과값이다.

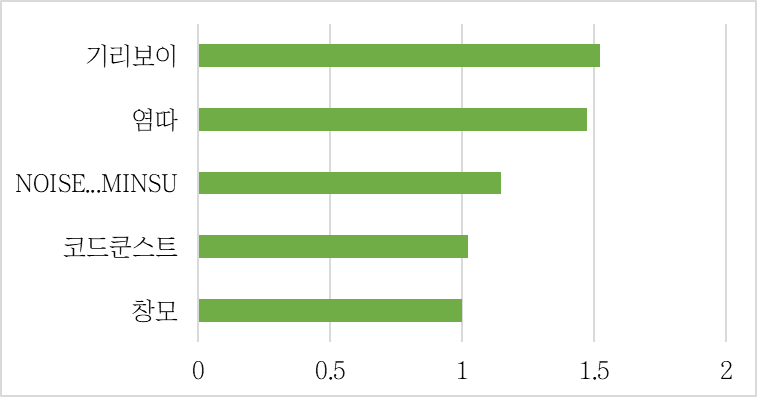
**[표 9] 아티스트 가중연결 중심성(Weighted Degree Centrality) 상위 5명**



해당 년도에서 아티스트 ‘창모’가 가장 높은 중심성을 지닌 것을 파악했다. 해당 아티스트의 ‘METEOR’라는 곡은 매달 국내 랩/힙합 부문 차트 10위 안에 들었다. 그 밖의 다른 많은 곡들 역시 꾸준히 차트인을 하며, 음원 차트에서 좋은 모습을 보여주었다. 또한, 염따, pH-1, 더 콰이엇, 기리보이가 그 뒤를 이었다.

프로듀서 부분에서는 기리보이가 가장 높은 중심성을 보여주었다. 기리보이가 프로듀싱한 곡 중 차트인 한 곡은 20곡이 넘으며 2위, 4위 등 좋은 순위를 거둔 곡들도 존재하기 때문에 높은 가중치를 보인다. 그 외에도 도출된 결과값을 통해 많은 프로듀서의 영향력을 파악했다.

**[표 10] 프로듀서 가중연결 중심성 (Weighted Degree Centrality) 상위 5명**



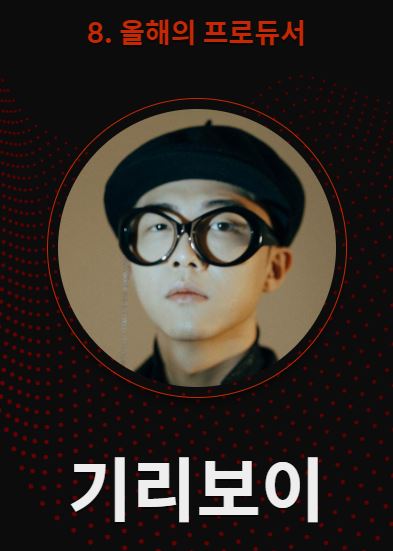
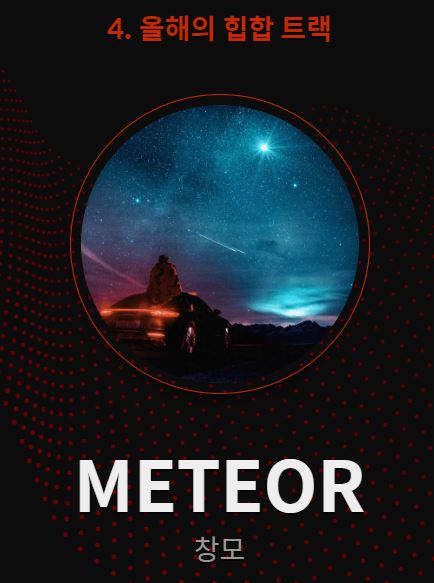
**4. 프로젝트 결과**

이를 토대로 영향력이 높은 아티스트와 프로듀서의 특색과 연관성을 분석했다. 먼저 아티스트 측면으로 분석했다. 가중연결 중심성(Weighted Degree Centrality)이 높은 아티스트 상위 10명을 분석한 결과, 차트인한 그들의 곡 중 셀프 프로듀싱의 비중이 높은 것을 파악할 수 있었다. 스스로 본인의 특색을 잘 알기 때문에 직접 프로듀싱을 한 것이 순위에 도움이 된 것으로 보인다.

다음으로 프로듀서 측면으로 분석했다. 가중연결 중심성(Weighted Degree Centrality)이 높은 프로듀서 상위 10명을 분석한 결과, 차트인 한 곡 대부분이 같은 소속사의 아티스트와 작업한 곡임을 확인할 수 있었다. 같은 회사일 경우, 서로 비슷한 특색을 지닌 경우가 많은데, 이것이 높은 음원 순위를 기록하는데 긍정적인 영향을 끼쳤다고 해석된다.

실제로 한국대중음악상처럼 한 해의 힙합 음악을 평가하는 시상식인 ‘한국힙합어워즈’에서 2020년 수상한 기록을 찾아볼 수 있다. 대중음악평론가 30인 투표 50%와 네티즌 투표 50%로 진행되는 이 수상식에서 올해의 인물을 선정하는데, 올해의 아티스트로 염따, 올해의 힙합 트랙으로 창모의 ‘METEOR’, 올해의 프로듀서로 ‘기리보이’가 수상했다. 실제로 중심성을 통해 도출해낸 결과에서도 아티스트 부문 염따가 2위, 프로듀서 부문 기리보이가 1위를 차지했다. 또한, 아티스트 부문 1위, 프로듀서 부문 5위한 창모의 METEOR는 올해의 힙합 트랙에 선정됐다. 이는 그래프 중심성을 통해 도출해낸 영향력이 유의미한 결과를 지닌다고 볼 수 있다.

[그림 11] 한국힙합어워즈 올해의 수상작



**5. 결론 및 향후연구**

**5.1 결론**

본 연구를 통하여, 힙합 분야에서 어떤 특성을 지닌 아티스트와 프로듀서가 작업했을 때 음원 성적이 좋았었는지를 파악할 수 있었다. 그렇기 때문에 높은 음원 성적을 위해서는 무작정 다른 사람들과 음원 작업을 진행하는 것이 아닌, 셀프 프로듀싱을 할 수 있는 능력을 키우거나, 여건상 힘들다면 같은 회사 사람들과 같이 서로 특색이 비슷한 사람들과 함께 작업하는 것이 음원 순위에 도움을 줄 수 있을 것이다.

**5.2 향후연구**

향후에는 다른 년도의 차트에도 본 연구를 통한 분석을 진행하여 본 연구에서 얻은 결론의 신뢰성을 높일 것이다. 또한, 더 나아가 힙합 장르에만 국한하는 것이 아니라 다양한 음악 차트에서의 아티스트와 프로듀서의 연관성을 분석하여 전체적인 음악 차트에서 프로듀서와 아티스트의 관계가 어떠한 영향이 있는지 알아보면서 범용성을 넓힐 생각이다.

**6. 참고문헌**

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

[2] 하정철, 강동훈, 박재모, 장으뜸, 이은영, 임성현, 길준민, “뉴스 기사와 음원차트 간의 상관관계 추출을 위한 R언어 기반 분석기의 설계 및 구현”, 한국정보과학회 학술발표논문집, 122-124, 2016.

[3] https://gephi.org/