**Graph Centrality 이용한 음원차트에 따른**

**국내 힙합 프로듀서와 아티스트의 영향력 분석**

김건희(2017103958)

송근영(2017103997)

**요약**

Graph Centrality를 이용해 음원차트에 따른 국내 힙합 프로듀서와 아티스트 간의 사회연결망을 확인하고, 프로듀서와 아티스트의 각 관계가 힙합 장르에서 어떤 영향을 끼치며, 어떠한 상호 관계가 있는지 분석한다.

1. **서론**
   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. **연구목표**

국내에서 가장 이용자수가 많은 멜론 차트를 이용할 것이다. 지난 힙합 장르 차트에서 데이터를 수집하고 정제한다. 이후 그 데이터를 통하여 아티스트와 프로듀서의 영향력을 알아본다. 그 과정에서 Graph Centrality를 이용하며, 그 중에서 Betweenness Centrality를 주로 이용할 계획이다. 그렇게 나온 중심성 수치를 통해 각 영향력을 확인해보고, 그것이 실제 차트와 연관성이 있는지 확인한다. 만약 연관성이 있다면 그것을 통하여 아티스트와 프로듀서 사이의 관계에 대해서도 알아볼 예정이다.

1. **관련연구**
   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.

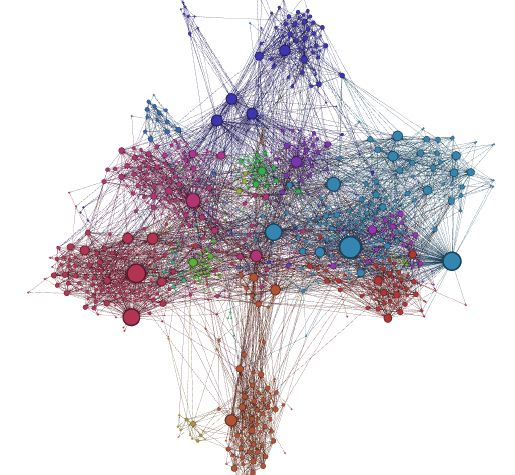
* 1. **기존 연구의 문제점과 해결 방안**

기존 연구는 데이터를 수집하여 기사의 개수와 음원차트 사이의 상관관계를 분석하였다. 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.

1. **프로젝트**
   1. **프로젝트 내용**

본 프로젝트는 공신력 있는 음원 서비스를 통해 수집한 데이터를 기반으로 제작된다. 아티스트와 프로듀서는 노드로 표현되고, 서로를 엣지로 연결한다. 이때, 음원차트 순위에 따라 각 엣지에게 가중치를 부여한다. 이때, 상위권의 음원 석차 횟수가 많을수록 엣지가 굵어지며, 이에 따라 각 노드의 크기가 커지게 된다. 이렇게 제작된 그래프를 Graph Centrality를 이용해 시각화해 국내 힙합 내에서 아티스트와 프로듀서의 영향력을 분석한다. 또한 이를 바탕으로 아티스트와 프로듀서의 영향력이 음원차트와 어떤 긴밀한 연관이 있는지 분석하고, 현재 국내 힙합 시장의 흐름을 분석한다. 최종 제작된 그래프는 아래와 유사한 형태를 띤다.



**그림 1 [시각화된 그래프]**

* 1. **요구사항**
     1. **정제된 데이터**

위와 같은 프로젝트를 위해 데이터 수집 및 정제가 필요하다. 본 프로젝트는 두 가지 경로로 데이터를 수집한다. 첫째, 음원 서비스 멜론에서 제공하는 음원차트를 사용한다. 멜론은 현재 카카오 엔터테인먼트에서 운영하며, 국내에서 최대 사용자를 보유하고 있다. 둘째, 한국음악저작권협회에서 제공하는 음원 저작권자 데이터를 이용한다. 이를 통해 곡의 정확한 아티스트와 프로듀서를 확인한다. 이후 시각화 도구 ‘Gephi’와 호환될 수 있도록 데이터를 교정 및 통합해 정제한다.

* + 1. **Graph Centrality 요구사항**

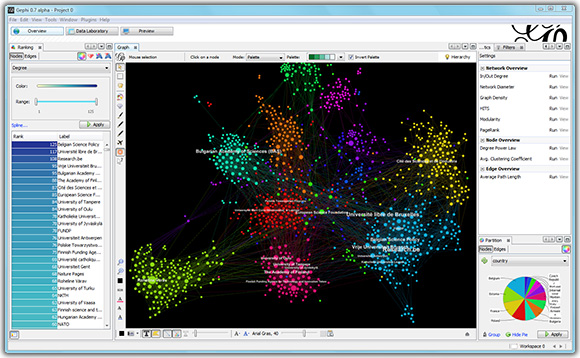
본 프로젝트를 수행하기 위한 핵심 알고리즘이다. 노드와 엣지로 표현된 그래프에서 노드들의 상대적 중요성을 계산하는 기법이다. 계산하는 방법에 따라 크게 매개 중심성(Betweenness Centrality), 연결 중심성(Degree Centrality), 근접 중심성(Closeness centrality), 고유벡터 중심성(Eigenvector Centrality)로 분류된다. 현재 프로젝트는 매개 중심성을 주로 사용하되, 상황에 맞게 다른 중심성 기법을 사용해 각자 다르게 출력된 결과를 대조해 분석한다.

매개 중심성(Betweenness Centrality)은 서로 다른 두 노드()의 최단거리 상에 자신()이 포함되는 비율이다.

여기서 는 에서 로의 최단 경로의 수이고, 는 에서 로의 최단 경로들 중에 를 거쳐 가는 경로의 수이다.

* + 1. **Gephi**

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



**그림 2 [Gephi 화면 예시]**

* 1. **개발환경**

운영체제: 윈도우 10

개발언어: Python

개발도구: Gephi, Igraph

1. **향후 일정 및 역할 분담**

|  |  |  |
| --- | --- | --- |
| **진행 일정** | **담당** | **내용** |
| ~ 4/10 | 팀 전원 | 프로젝트 데이터 수집 |
| 4/10 ~ 4/25 | 팀 전원 | 데이터 정제 및 알고리즘 작성 |
| 4/25 ~ 5/10 | 팀 전원 | 그래프 중심성 측정 |
| 5/10 ~ 5/28 | 팀 전원 | 중심성 기반 분석 |
| 5/28 ~ 6/5 | 팀 전원 | 새로운 데이터 적용 및 반복 |
| 6/5 ~ 6/9 | 팀 전원 | 발표 준비 및 발표 |
| 6/9 ~ 6/15 | 팀 전원 | 최종보고서 작성 및 제출 |

1. **결론 및 기대효과**

최근 각종 프로그램과 대중 매체의 발달로 힙합은 완전히 대중적인 음악 장르로서 자리를 잡았다. 이에 따라 국내 힙합 시장 규모도 대거 확대됐다. 본 프로젝트는 이러한 흐름에 맞춰 아티스트와 프로듀서의 관계를 그래프로 구축하고, Gephi를 이용해 시각화하여 각 아티스트와 프로듀서의 영향력을 확인할 것이다. 또한, 이러한 영향력이 음원차트와 어떤 상호연관을 지니고 있는지 분석할 것이다. 이를 통해 현재 국내 힙합 시장의 흐름을 파악하고, 더 나아가 향후 한국 힙합 장르의 동향을 예측할 수 있다.

1. **참고문헌**

하정철, 강동훈, 박재모, 장으뜸, 이은영, 엄성현, 길준민 R 기반의 빅데이터 기술을 활용한 뉴스기사와 음원차트의 상관관계 분석 한국정보처리학회 학술대회논문집, 2016.12 636-639(3p).