Problem Chosen	2021	Team Control Number
D	MCM/ICM Summary Sheet	2108926

Influence and similarity of music

‘Music is a higher revelation than all wisdom and philosophy.’ Music is an important part of human civilization and plays an important role in human collective experience. We were hired by the ICM to study the evolution and revolutionary trends of artists and genres. We mainly establish three types of models: basic model, graph model, and time model.

In the basic model, we construct the music influence parameters in the directed graph and measure the musical similarity. For the music influence parameter, we use the weight of the edge from the influencer to the follower to represent it, which depends on two sub-indicators: the node influence of each artist calculated by the Page Rank algorithm, and time influence related to the distribution of active starts.We compare the influence within and between genres and find that Pop/Rock has the strongest influence. In this genre, The Beatle, Bob Dylan and The Rolling Stone have the strongest influence. For music similarity, we use the Euclidean distance to measure artist’s music similarity. Based on this, we calculate the similarity of music of the same genre and music of different genres, and find that the average similarity of different genres was 1.55672,The similarity of music of the same genre is less than this value, artists of the same genre are more similar, and the similarity of music of different genres is smaller.

In the graph model, we first construct a sub-network of directed graphs, and use the music influence parameters in the network to reveal the influence of artists in the sub-network, the influence of music genres, the connection of influence age and the direction of influence. Subsequently, we use BFS to analyze whether the similarity indicates that influencer does have an impact. Using this algorithm to search for the average music distance between influencer and follower on the path, we find that when the depth is less than 7, the average music distance at different depths changes little, so it is believed that influencer does affect followers and that the influence is passed from generation to generation. In the same way, we find that the average Euclidean distance of the two musical features, danceability and valance, changes little at different search depths, so they are cotagious features.

In the time model, time becomes an important dimension. We first analyze the characteristics that may mark the evolution of music. By comparing the similarity of the standard characteristics with the average similarity of the genre, we believe that acousticness and energy can represent the revolutionary characteristics of music. We selecte the 1960s for analysis and find that The Beatles and The Rolling Stones are revolutionaries. Then we add the time dimension to the directed network and build an influence model with time dimension. Taking Country as the object, we find that in other genres, Country was mainly influenced by Folk in the 1930s, and was mainly influenced by Pop/Rock after the 1950s, the indicator of dynamic influencers is music similarity. After considering the influence of artists, we use the time difference of the average music distance in each year as an indicator to indicate the impact of cultural factors on music. The sharp fluctuation of this indicator indicates the impact of exogenous influence, we take rock music as an example and analyze it from three perspectives: political, society and technology.

Keywords: **PageRank, Euclidean distance, Depht-limited Breadth First Search**

Contents

1 Introduction

1.1Background and Problem Restatement

‘*Music is a higher revelation than all wisdom and philosophy*.’Beethoven once said.Music is the treasure of human civilization and plays an important role in the process of human evolution.Evolution of music has been an integral part of our lifestyle ever since the inception of human consciousness,it is accompanied by the evolution of human civilization. The evolution of music may come from the influence of other artists, or from the reaction to external events.In order to clarify the role of music evolution in human collective experience, people need to understand how music evolves through societies over time.

Based on the above background, we have been identified by the Integrative Collective Music (ICM) Society to develop models that measures musical influence. We will reveal evolutionary and revolutionary trends of artists and genres through models.Specifically, we need to complete the following tasks:

- Creat a directed network reflecting music influence, capture the parameters reflecting music influence and use sub-networks to explain; use the given data to find a measure of music similarity, and analyze whether artists in the same genre are more similar.
- Analyze similarity and influence, find the characteristics that distinguish genres and describe the evolution of genres; analyze whether certain genres are similar.
- Find out the characteristics that reveal music revolution and the main influencers of the revolution; analyze the influence process and the evolution process of one genre; find indicators that reveal the dynamic influencers.
- Use the similarity to analyze whether the influencer does have an influence, whether the influencer’s musical characteristics play a different role in the influence process. Analyze how to express the cultural influence of music (social, political or technical)
- Write a document to ICM.

1.2 Our Work

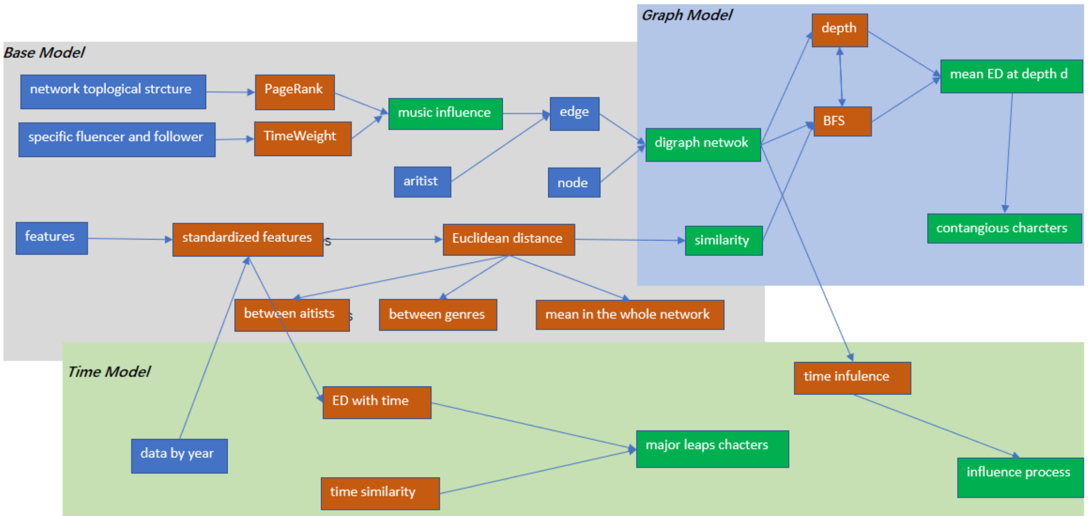


Figure1:Our work

2 Assumptions and Justifications

In order to solve the above problems, we make the following assumption to simplify the problem reasonably:

When analyzing the influence of a certain music genre from different active starting points, only the network relationship at the active starting point of the music genre is calculated, without considering the influence of accumulation over the years.

Although old music will also have an impact on current music, we believe that this is not a common situation. In order to simplify the model, we ignore the influence of aging accumulation.

3 Music influence model

3.1A network-based music influence model

3.1.1Problem analysis

Task 1 requires us to build a directed network of influencers and followers based on the' influence_data' data set, and construct parameters to quantify music influence. In this directed graph,we define artists are the vertexes, the direction of the edge is from the influencer to the follower, and edge's weight represents the influencer's music influence on the follower. Moreover, we use two sub-indicators to quantify music influence. From a macro perspective, the first sub-indicator is for the topological structure of this directed graph. We define it as the PR value of every node, and use the classic PageRank algorithm to quantify the artist's music influence in this network.; from a micro point of view, the second indicator is for a specific follower in the directed graph. It is defined as the difference of active_start between the influencer and the follower. We use this variable and the time distributions of music artists to construct every edge's time influence. Finally, by combining these sub-indicators, the final music influence, that is, the side weight is obtained.

3.1.2Notations

Important notations used in this model are listed in Table 1.

Table 1:Noations of music influence model

Notation	Definition
p	Network's node(artist)
E	Directed edge set in the network
M	A set of music artists pointing to a certain music artist
L	The number of outdegree music artists of a certain music artist
N	The number of nodes(srtists)
α	Damping coefficient
a	The indicator vector (confirm whether it is a hanging page)
e	Unit vector
d	The active_start of a srtist
Δd	The difference of active_start between two artists
wat	The proportion of the specified amount of Δd to the total
PR	Artist's PR value
$bound$	The time weight of edge
A	Adjacency matrix of directed graph
S	Transition matrix
w	Total weight of an edge
ε	iteration precision

3.1.3Model establishment

1.Data preprocessing

First, we preprocess the 'influence_data' data set, because there maybe the same names. We use each different id as the node representation in the network, and connect each row of data with the direction from the influencer to the follower. The weight of the edge is to be determined.

We calculate the number and proportion of different genres of music in this network,then we calculate the outdegrees of music artists in this network, the outdegree means the number of other artists influenced, taking the Top 10 shown in Table 2.

Table 2 : Top 10 artists

Rank	Influencer(genre)	Outdegree	Rank	Influencer(genre)	Outdegree
1	The Beatles	615	6	Jimi Hendrix	201
2	Bob Dylan	389	7	The Kinks	192
3	The Rolling Stones	319	8	The Beach Boys	186
4	David Bowie	238	9	Hank Williams	184
5	Led Zeppelin	221	10	The Velvet Underground	181

2.Model establishment and analysis

(1)PageRank algorithm

The PageRank algorithm is based on two basic assumptions:

- If a webpage is linked to by many other webpages, it means that this webpage is more important,and the PR value will be relatively high.
- If a webpage with a higher PR value links to other webpages, it means that the PR value of the linked webpage will increase accordingly.

Based on these two assumptions, it is unreasonable for us to directly use the PageRank algorithm in this network,since its edge direction is from influencers to followers. Therefore, when we use the PageRank algorithm, we keep the network node unchanged, and change the direction of the directed edge, that is, from the follower to the influencer. At this time, when a music artist is linked, it means that it inluences others, and its PR value will be relatively high, which is in line with the first assumption. The artist's out-link points to the person who influenced the artist,and its PR value is also relatively high,this is in line with the second assumption. Then we establish the PR attributes of this network node to quantify music influence.

At the same time, after changing the direction of the directed edge, in order for the PageRank algorithm to work normally, we need to introduce a directed edge to each node in the network (including itself) for those vertexes that has no outdegree ,this helps to solve the problem of hanging web pages. Based on this, the PageRank algorithm can be described as follows ：

Let p_i be a music artist in the network, and the calculation equation for its PR value $PR(p_i)$ is as follows:

$$PR(p_i) = \alpha \sum_{p_j \in M(p_i)} \frac{PR(p_j)}{L(p_j)} + \frac{1-\alpha}{N} \tag{1}$$

Among them, $p_i, i \in [1, N]$is the music artist, $M(p_i)$is the set of followers of the music artist p_i, $L(p_j)$ is the number of influencers of the music artist p_j, α is the damping coefficient, and its meaning is the probability that the user will continue to visit the next page after visiting a certain page at any time.

Then we will simply prove that the PR value is convergent and has nothing to do with the initial value PR_0.

We first define the adjacency matrix A, whose calculation equation is shown as:

$$A_{p_i p_j} = \begin{cases} \frac{1}{L(p_i)} & \text{if}(p_i, p_j) \in E \\ 0 & \text{others} \end{cases} \tag{2}$$

According to our previous solution to the hanging page, we introduce a column vector A containing N elements as the index vector of the hanging page, and its calculation formula is as follows:

$$a_{p_i} = \begin{cases} 1 & \text{if} L(p_i) == 0 \\ 0 & \text{others} \end{cases} \tag{3}$$

Then the calculation formula of the transition matrix S can be obtained as the formula x:

$$S \leftarrow A^T + \frac{ae^T}{N} \tag{4}$$

Where e is a unit column vector containing N elements.

Apply the PR calculation formula and update S as follows:

$$S \leftarrow \alpha S + (1-\alpha)\frac{ee^T}{N} \tag{5}$$

Therefore, we get the matrix iterative formula about PR value:

$$\begin{aligned} PR_{k+1} &= S \cdot PR_k \\ &= [\alpha(A^T + \frac{ae^T}{N}) + (1-\alpha)\frac{ee^T}{N}]PR_k \end{aligned} \tag{6}$$

The iterative process of PR value is a Markov process, and the transition matrix is S. In order to converge and have nothing to do with the selection of the initial value R_0, that is:

$$\lim_{k\to\infty} PR_k \text{ exists} \tag{7}$$

$$\text{s.t.} \begin{cases} \forall PR_0 \in \mathbb{R}^+ \\ \sum_{p_i} PR_0(p_i) = 1 \\ PR_{k+1} = S \cdot PR_k \end{cases}$$

The above feature requires the transition matrix S to meet three conditions:

1)S must be a random matrix

2)S must be irreducible

3)S must be aperiodic

- Random matrix: also called probability matrix, assume that $s_{i,j}$ is the element of random matrix S in the i-th row and j-th column, then the following conditions need to be met:

$$a_{i,j} \geq 0, \forall i,j = 1,2,\dots,N \tag{8}$$

$$\sum_{j=1}^{N} a_{i,j} = 1, \forall i = 1,2,\dots,N$$

  Obviously, in the above derivation, the matrix A is a random matrix, and S is derived from the change of A via the formula x,x, and the way of change is to use probability as a measure, so S is also a random matrix.

- Irreducible matrix: The square matrix S is irreducible if and only if the directed graph corresponding to the matrix S is strongly connected. Obviously, after the correction of the formula x, the matrix S is irreducible.

- Aperiodic matrix: If S is a primitive matrix, then S is aperiodic. The primitive matrix needs to satisfy a certain power of itself to be a positive matrix, namely:

$$S^k > 0,\ \exists k > 0 \tag{9}$$

Obviously, for the matrix S derived above, k=1 is sufficient, so S is a primitive matrix, S is aperiodic.

In summary, we briefly proved that the PR value is convergent and has nothing to do with the initial value PR_0. Therefore, we can take $PR_0 = \frac{e}{N}$ when calculating. We get the flow chart of PageRank algorithm as follows:

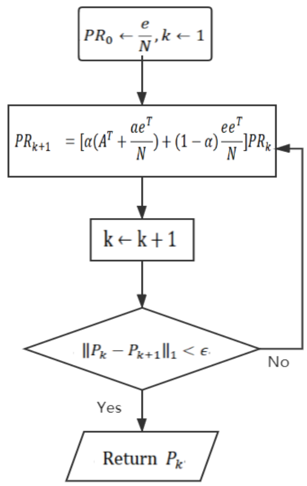


Figure 2: Flow chart of PageRank algorithm

(2)Calculation of the time weight of music influence

Our understanding of time weight is: the musical influence a follower receives is related to the concentrated age of the influencer. Specifically, among his influencers, if the number of influencers in the same period increases, then the overall influence value of influencers in the same period increases.

In order to calculate the time weight, first, we calculate the age difference Δd of different musicians in the directed edge of the network,the formula is as follows:

$$\Delta d_{p_i,p_j} = \frac{|d(p_i) - d(p_j)|}{10} \tag{10}$$

Then, we calculate the proportion of the active start difference wat：

$$\text{wat}(\Delta d_i) = \frac{\text{count}(\Delta d_i)}{\sum_j \text{count}(\Delta d_j)} \tag{11}$$

Among them, count represents the total number of occurrences of a certain age difference.

For a specific directed edge from influencer$p_i$to followerp_j, its time weight is:

$$bound(p_i,p_j)=\frac{wat(\Delta d_{p_i,p_j})}{\sum_{p_k\in M(p_j)} wat(\Delta d_{p_k,p_j})} \tag{12}$$

(3) Calculate the final directed edge weight that is music influence

First, we standardize PR and adopt min-max standardization, namely:

$$PR^*=\frac{PR-\min}{\max-\min} \tag{13}$$

Then, we get the calculation formula of the edge weight as follows:

$$w_{p_i,p_j}=\frac{PR^*(p_i)+bound(p_i,p_j)}{2} \tag{14}$$

3.1.4 Model solving process and results

First of all, from the previous analysis, we need to change the direction of the edge , then we use the PageRank algorithm to calculate the PR value of each music artist, and the iteration accuracy $\varepsilon=1\times10^{-6}$, the top 20 of the final iteration results are shown in Table 3 :

Table 3:Top 20 of PR value

Rank	Influencer-genre-active_start	PR value	Rank	Influencer-genre-active_start	PR value
1	Cab Calloway-Jazz- 1930	0.0205	11	Mississippi Sheiks-Blues-1930	0.0068
2	Billie Holiday-Vocal- 1930	0.0196	12	Mississippi Fred McDowell-Blues-1950	0.0066
3	Lester Young-Jazz- 1930	0.0169	13	Muddy Waters-Blues-1940	0.0066
4	Louis Jordan-Jazz- 1930	0.0134	14	Bob Dylan-Pop/Rock-1960	0.0057
5	T-Bone Walker-Blues-1930	0.0099	15	Woody Guthrie-Folk-1930	0.0056
6	Sister Rosetta Tharpe-Religious- 1930	0.0091	16	Nat King Cole-Jazz-1930	0.0054
7	The Beatles-Pop/Rock-1960	0.0091	17	Billy Eckstine-Vocal-1930	0.0051
8	The Mills Brothers-Vocal-1930	0.0076	18	Hank Williams-Country-1930	0.0049
9	Roy Acuff-Country-1930	0.0068	19	Chuck Berry-Pop/Rock-1950	0.0046
10	Charlie Christian-Jazz-1930	0.0067	20	Ray Charles-R&B-1940	0.0044

As can be seen from the above table: According to the PR value calculated using the Pagerank algorithm, the artists with the greatest musical influence are Cab Calloway, Billie Holiday, and Lester Young. At the same time, the genres with greater musical influence are of great variety, covering Jazz, Vocal, Blues, Country, Rock/Pop, etc. From a time point of view, artists began to be active in a wide range of times, including the 1930s to the 1950s.

Then, we calculate the time weight of each edge, and normalize the PR value to get the weight of the final directed edge, that is, the music influence, as shown in Table 4：

Table 4: Top 10 edge weight

Influencer(genre)	Follower(genre)	PR value	Time weight	Edge weight
Cab Calloway（Jazz）	Louis Prima（Jazz）	1	1	1
Cab Calloway（Jazz）	Erskine Hawkins（Jazz）	1	1	1
Billie Holiday（Vocal）	Lester Young（Jazz）	1	0.9536	0.9768
Lester Young（Jazz）	Charlie Christian（Jazz）	1	0.8242	0.9121
Lester Young（Jazz）	Paul Quinichette（Jazz）	1	0.8242	0.9121
Louis Jordan（Jazz）	Nat King Cole（Jazz）	1	0.6531	0.8266
Louis Jordan（Jazz）	Wynonie Harris（R&B）	1	0.6531	0.8266
Louis Jordan（Jazz）	Bill Doggett（R&B）	1	0.6531	0.8266
Louis Jordan（Jazz）	Big Maybelle（Blues）	1	0.6531	0.8266
Louis Jordan（Jazz）	Joe Liggins（Blues）	1	0.6531	0.8266
Louis Jordan（Jazz）	Louis Jordan（Jazz）	0.5	1.0000	0.7500

3.2 Sub-network—A specific analysis of music influence

In order to explore a subset of musical influence, we have constructed a sub-network whose nodes are composed of artists with high musical influence in various genres, such as: The Beatles, Bob Dylan belonging to Pop/Rock;Ray Charles, Roy Brown belonging to R&B, and Cab Calloway, Lester Young of Jazz genre, etc. The specific sub-networks are shown in the figures below. The horizontal axis in the figure represents the time when the influencer began to be active, reflecting the dimension of time, and the vertical axis represents the artist’s PR value, this indicator reflects the size of the artist's musical influence to a certain extent. Figure3 shows the relationship between the musical influence of artists within each genre, while Figure4 considers the influence of artists between different genres, comprehensively showing the relationship between the genres and the musical influence of artists within each genre.

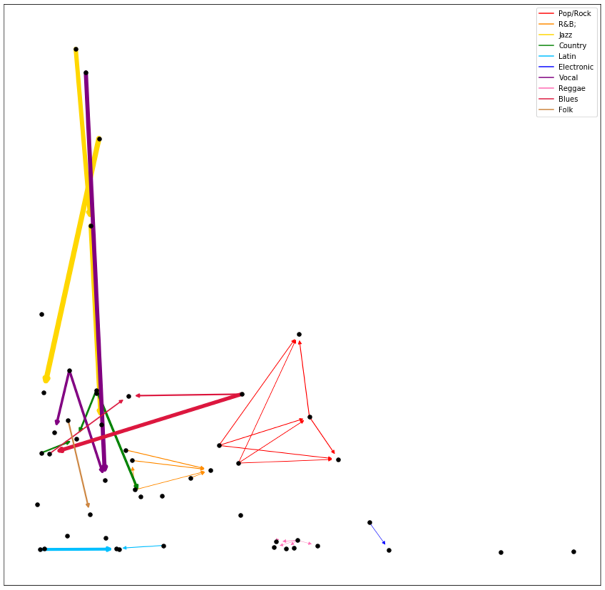


Figure 3:A sub-network(uncross genres)

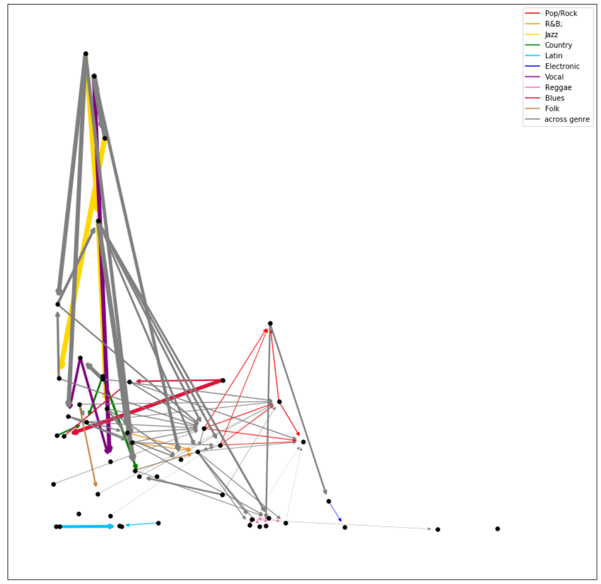


Figure 4:A sub-network(cross genres)

In the subnet we built, the indicator of music influence reveals very rich connotations. We will analyze from the following aspects:

- **It reveals the magnitude of the music influence of artists：**

  Since the edge power represents the magnitude of music influence, we can directly compare the music influence between artists through the edge power. Take Pop/Rock as an example: in this subnet, The Beatles' average edge weight is 0.5768, and Bob Dylan's average edge weight is 0.2978. Therefore, in this sub-network, The Beatles has a greater music influence than Bob Dylan.

- **It reveals the relative magnitude of the overall musical influence of various genres:**

  We can analyze the relative influence of different genres according to the distribution of the corresponding nodes of each genre in the graph. If the overall center of gravity of the corresponding nodes in a genre is far from the horizontal axis, then the corresponding music influence of the genre will be bigger. Through this subnet, we can compare the relative influence of different genres more conveniently. For example, the Figure 3shows that the center of gravity of the corresponding nodse of Jazz is farther from the horizontal axis than that of Pop/Rock, so the music influence of Jazz is stronger than that of Pop/Rock.

- **It reveals the concentration of the years affected by each artist:**

  Our subnet comprehensively considers the size of the musical influence of each artist and the dimensions of time. Since the horizontal axis represents the first decade when an artist becomes active, the closer the horizontal distance between the nodes that affect each other, it indicates their age concentration is higher. As shown in the figure 3: compared with Pop/Rock, Reggae has a closer horizontal distance between nodes , indicating that the inter-nodes in Reggae have a higher degree of chronological concentration,and music influences are mainly caused by artists in similar periods.

- **It reveasl the mutual influence between the artists:**

  First, the direction of the edge between two nodes in the subnet indicates the fact that who influences whom. The arrow in the figure points from a influencer to a follower, indicating the influence of a influencer on a follower. Secondly, the thickness of the edge shows the size of the influence. The thicker the edge indicates the greater influence . For example: it can be seen from the figure 4that in the genre of Jazz, Cab Calloway influenced Louis Jordan, Lester Young influenced Charlie Christian, and Lester Young's influence was greater than that of Cab Calloway; Cab Calloway not only has an impact on artists of the same genre, but also has an impact on artists of the Vocal genre.

4 Music similarity model

4.1Problem analyse

The title requires us to establish a measure of music similarity. Similarity refers to the relationship between two things. The higher the similarity, the more similar between the two. For the feature vector of each artist, we use the classic Euclidean distance to measure the similarity between artists. The larger the Euclidean distance, the lower the similarity. Then, we also defined the similarity of different types of music, which is the average value of the Euclidean distance between all artists of a certain type of music and all artists of another type of music.

4.2Model establishment

4.2.1Notations

Table 5:Notations of this model

Notation	Definition
X	Feature matrix
Z	Standard feature matrix
ED	Euclidean distance between standard features
S_{in}	Similarity within genre
$S_{between}$	Similarity between genres

4.2.2 Model building and analysis

Normalize the music feature matrix X to the standardized feature matrix Z.
The row vector Z_i of the standardized feature matrix Z represents the standard feature of an artist p_i, and the Euclidean distance between it and the row vector of another artistp_j is calculated, namely：

$$ED(p_i, p_j) = \sqrt{\sum_k Z_{ik} \cdot Z_{jk}} \tag{15}$$

For the internal similarity S_{in} of the same type of music, the calculation formula is as follows:

$$S_{in}(genre_u) = \frac{\sum_{p(i)}^{g(p_i)=genre_u} \sum_{p_j}^{g(p_j)=genre_u} E\,D(p_i, p_j)}{\binom{N(genre_u)}{2}} \tag{16}$$

Among them, $N(genre_i)$ represents the number of artists of the i music genre.
For the similarity $S_{between}$ between different types of music, the calculation formula is as x

$$S_{beetween}(genre_u, genre_v) = \frac{\sum_{p(i)}^{g(p_i)=genre_u} \sum_{p_j}^{g(p_j)=genre_v} E\,D(p_i, p_j)}{N(genre_u) \cdot N(genre_v)} \tag{17}$$

Further, we find the average of the similarity between different types of the entire network:

$$S = \frac{\sum_{genre_u} \sum_{genre_v \neq genre_v} S_{between}\,(genre_u, genre_v)}{(len(genre) - 1)^2} \tag{18}$$

Where $len(genre)$ represents the number of genres in the network

4.2.3Results

According to the formula(16), we get the similarity within the same genre:

Table 6:The similarity within the same genre

Genre	Electronic	Reggae	Country	Stage & Screen	Easy Listening	Folk
Similarity	1.2318	1.1273	0.831	1.295	1.255	1.0299
Genre	Classical	R&B	Avant-Garde	Vocal	Jazz	Blues
Similarity	1.3095	1.0120	1.71512	0.966	1.110	1.100
Genre	Comedy/Spoken	New Age	Latin	Children's	International	Pop/Rock
Similarity	1.334	1.350	1.0700	2.2459	1.1951	0.91074

According to the formula(18),similarity of different genres of music is 1.55672.
After comparison, we find that for most genres, the similarity of artists within genres is higher than that of artists between genres.

5 Distinguishments and changes of genres

5.1Comparison of music similarity

According to formula(16) ,we get the similarity within the same genre,as shown in Figure5:

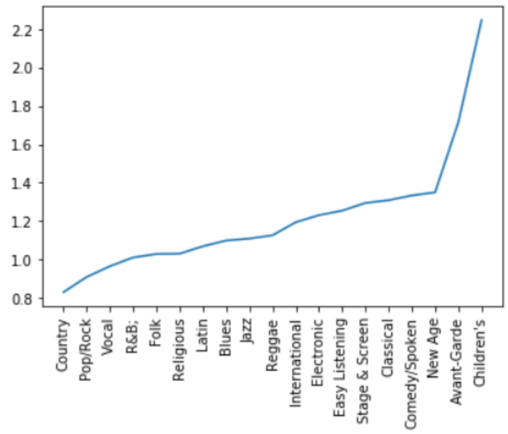


Figure 5:The similarity within the same genre

It can be seen that the similarity of the same type of music is very high, especially country and pop music music, but there are also low similarities of the same music, such as Avant-Garde, Children's.

According to formula(17) ,we get the similarity between genres,then we draw the heat map,we can get the conclusion that most genres have low similarity with other genres, which shows that most genres have distinct personalities.But there are still some genres related,such as Country and Children’s.

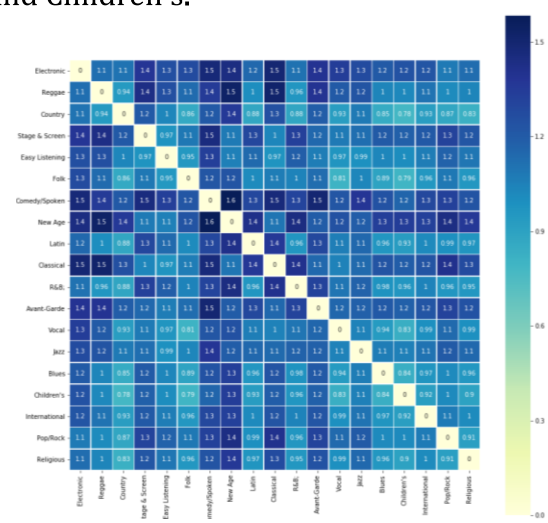


Figure 6 :The heat map of similarity between genres

5.2 Comparison of music influence

5.2.1Notations

Important notations used in this model are listed in Table x.

Table 7: Noations of this model

Notation	Definition
$genre$	Music genre
g	Artist’s music genre

F_{in}	Artist's music influence within a genre
$F_{between}$	The music influence between different genres

5.2.2Measurement of music influence

For the same music genre, we calculate the musical influence of a certain music artist of that music genre on all other music artists of the music genre, denoted as F_{in}, then we get the following calculation formula:

$$F_{in}(P_i) = \sum_{p_j}^{g(P_j)=g(P_i)} w\,(pi, pj) \tag{19}$$

As for the influence between different music genres, we calculate the musical influence of a certain music genre on other music genres, which is defined as the edge weight of all music artists of this music genre on all music artists of a certain music genre,the fomular is as follows:

$$F_{between}(genre_u, genre_v) = \sum_{p_i}^{g(p_i)=genre_u} \sum_{p_j}^{g(p_j)=genre_v} w\,(p_i, p_j) \tag{20}$$

5.2.3Comparation of music influence

In order to compare the internal influence of the same music genre, we selected some of the most influential music genres for analysis, and we selected the top five artists in each genre. When there are less than five artists, we select all music artists. The figure below reflects the influence of each artist within the main genre.

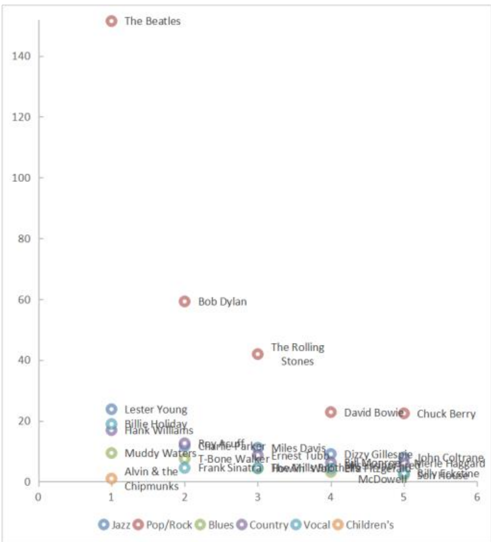


Figure 7: The influence of each artist within the main genre

From Figure 7 we can get these conclusions:

- The influence of **Pop/Rock** artists is significantly higher than that of artists of other genres . And **The Beatles** has the strongest influence. For artists with higher influence in Pop/Rock, their influence is significantly higher than that of other genres, such as Jazz, Blues, etc., which shows that Pop/Rock has played an important role in the evolution of music. Among them, **The Beatles** has the strongest influence, followed by **Bob Dylan**, **The Rolling Stones**, **David Bowie**, and **Chunk Berry.**

- Among the genres with higher influence except Pop/Rank, the difference in influence of artists is not big. The most influential artists in each genre are: **Lester Young (Jazz), Billie Holiday (Vocal) , Muddy Waters (Blues), Rank Williams (Country)**

In order to analyze the influence between music genres, we standardize $F_{between}$ and make a heat map as follows:

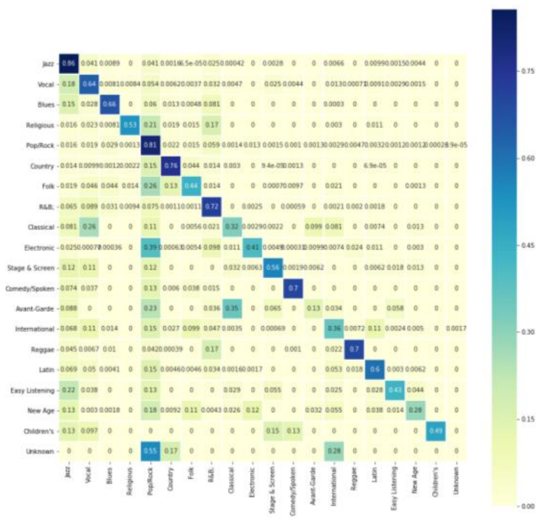


Figure 8:The heat map of music influence

It can be seen that different genres have different influences between genres.

- **Certain genres have influence internally and externally:**

  The four categories of Jazz, Vocal, Pop/Rock, and R&B not only have a strong influence on artists of the same genre, but also have a higher influence on artists of other genres.

- **Some genres are only influential inside:**

  Blues, Religious, and Electronic have little influence on other genres, and basically only exert influence within the genre;

- **Certain genres have little influence internally or externally:**

  There are also genres that have less influence within and between genres, such as Avant-Garde and New-Age.

6Model for real influencers and contagious features

6.1 Do influencers really influence—based on BFS

6.1.1Prblem analyse

We use the Breadth-first search algorithm (BFS)to calculate the average value of the music distance between a influencer and his followers on the search path(under different search depths).Then use the distribution of the average value to identify whether a influencer in fact has influence on followers. Subsequently, we use the standard feature Euclidean distance of the sub-features one by one to check the distribution change of sub-feature Euclidean distance's average value. The features with small changes in the average are contagious features.

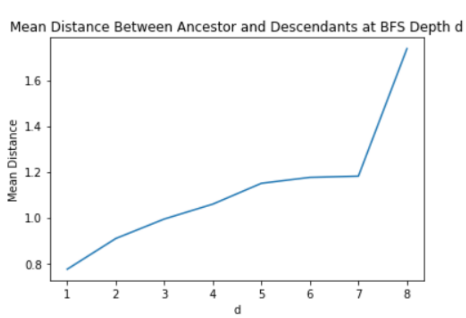
6.1.2Model establishment

Breadth-first search (BFS) is an algorithm for traversing or searching tree or graph data structures. It starts at the tree root (or some arbitrary node of a graph, sometimes referred to as a 'search key', and explores all of the neighbor nodes at the present depth prior to moving on to the nodes at the next depth level.

On the basis of BFS, we add the search depth k to limit the search depth, and by adjusting the search depth, we can view the changes in the music Euclidean distance or the similarity in the network.

6.1.3 Model solution and results

We first get the change of music distance and BFS depth,as shown in Figure 9 and Figure 10:



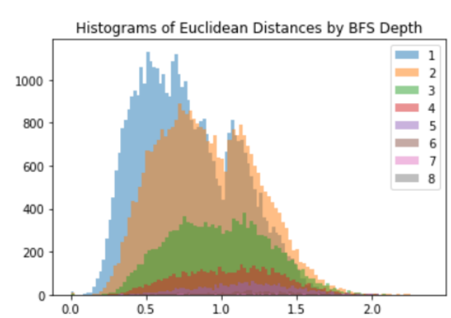


Figure 9:Mean distance between influence and followers at BFS Depth d

Figure 10:Histograms of Euclidean Distances by BFS Depth

It can be seen from the above two figures that when the BFS depth is less than 7, the average music distance between influencer and follower does not change too much. Although it is in a rising state, the rising speed gradually slows down. This shows that under the influence path of influencer,the influencer does have an impact on the follower, and the impact is transmitted in the generation ; when the BFS depth is 8, due to the small number of samples, the error is large. Although the average music distance becomes larger, we no longer consider this situation.

6.2 Are there more contagious features?—based on BFS

Also through BFS, we get the changes between the distances of different music features and the BFS depth, as shown in the following figures(For the following figures, we list all contagious features, but only one feature that is not contagious)

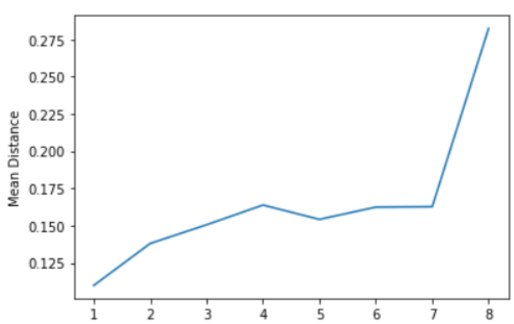


Figure 11:Danceability’s mean distance

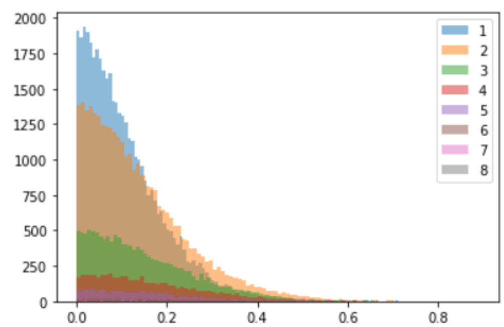


Figure 12: Danceability’s histograms

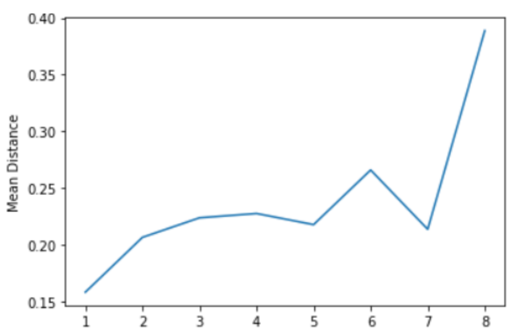


Figure 13:Valance’s mean distance

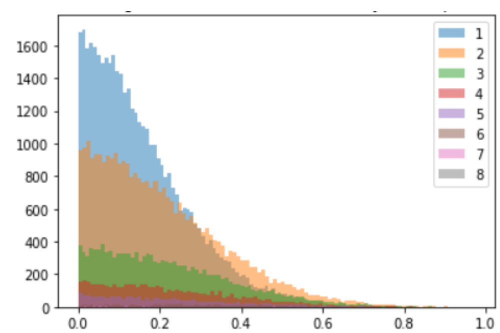


Figure 14: Valance’s histograms

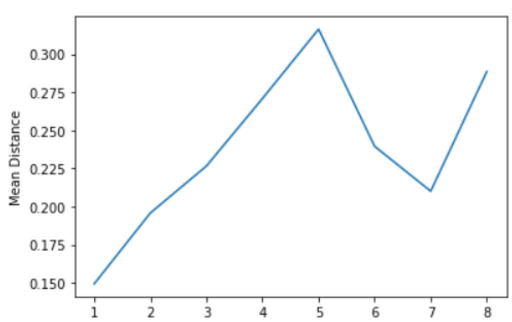


Figure 15:Energy’s mean distance

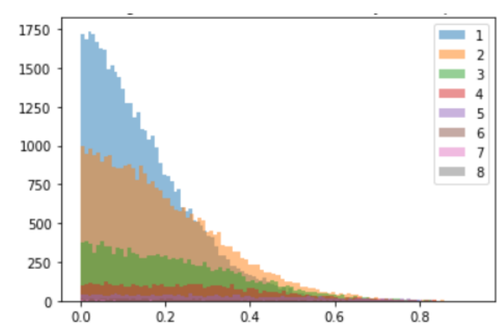


Figure 16:Energy’s histograms

As can be seen from the above figures, the average music distance of danceability and valace changes relatively smoothly, while the average music distance of music features such as energy fluctuates drastically. Therefore, we believe that danceability and valace have an actual impact and are contagious features.

7 Revolutionary characteristics and revolutionaries

There are a total of 13 features, we draw 13 curves that change with the year, as shown in Figurea,and then analyze which feature is the most capable of driving the music revolution.

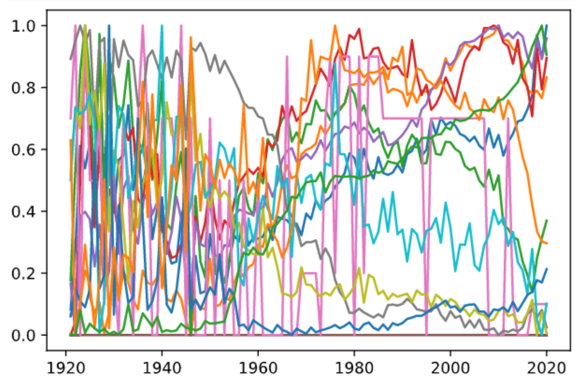


Figure 17: Changes in music characteristics

It can be seen from the figure that when the time is far away, the curve fluctuates more intensively, which shows that the genres experienced many great changes before the 1970s. We calculated the variance of each feature separately, and found that the variance of acousticness and energy is the largest, 0.37545807870367687 and 0.3346536050524824, respectively. This shows that acousticness and energy can best represent the revolution of music evolution. The most drastic changes in acousticness and energy were in the 1860s. Rock music was the most popular music in this era. According to the results of the Music Influence Model, the most influential rock genres were the Beatles and the Rolling Stones.So the Beatles and the Rolling Stones represent revolutionaries.

8 Genre evolution and dynamic influencers

8.1Problem analyse

We need to analyze the influence process of a music genre in the time of evolution. We established an influence model in the time dimension, and calculate the influence of different music genres on one music genre at different time points,finally we characterize the influence of different types at different time points.

8.2Model establishment

We define W as the influence of different genres in the time dimension,its formula is:

$$W(t, genre_u,\ \ genre_v) = \sum_{p_i}^{g(p_i)=genre_u, t(p_i)=t} \sum_{p_j}^{g(p_j)=genre_v} w\,(p_j, p_i) \tag{21}$$

$W(t, genre_u,\ \ genre_v)$ represents the influence of the v-th genre on the u-th genre at time t, and $t(p_i) = t$ indicates that the artist's active starting point is t.

8.3Results

We choose $genre_u$=Country to study the influence of other music in the evolution of Country music over time. For time t=1930, 1940, 1950, 1960, 1970, 1980, 1990, 2000, 2010, we selected the three music genres that have the greatest impact on Country music. The results are shown in Table 8.

Table 8: The most influential genres in different periods

Time	Genre	W
1930	Country	4.369624143326065
	Folk	1.248921333573762
	Jazz	0.2922404231241814
……		
1950	Country	34.85243977776871
	Pop/Rock	3.3854196057496773
	R&B	1.5869537572810617
……		
1970	Country	28.352614160487914
	Pop/Rock	7.347747658568013
	Folk	1.7869836954640865
……		
1990	Country	7.53752727283597

	Pop/Rock	3.1977391506368975
	Folk	0.6203760036036229

9 Cultural influence-an exogenous shock

We believe that the impact of music mainly includes the gradual influence from other artists, the sudden change caused by the impact of exogenous factors, and the cultural influence (or political, social, and technological influence) is the exogenous sudden change. We reflect this influence through ΔS_t, the corresponding formula is:

$$W\Delta S_t = S_t - S_{t-1} \tag{22}$$

Where S_t represents the average music distance of different genres in year t,ΔS_t reflects the musical distance between genres in different years. The larger the indicator, the greater the gap between genres. The drastic changes in this indicator indirectly indicate that some genres have undergone tremendous changes in the short term. We will illustrate with three examples:

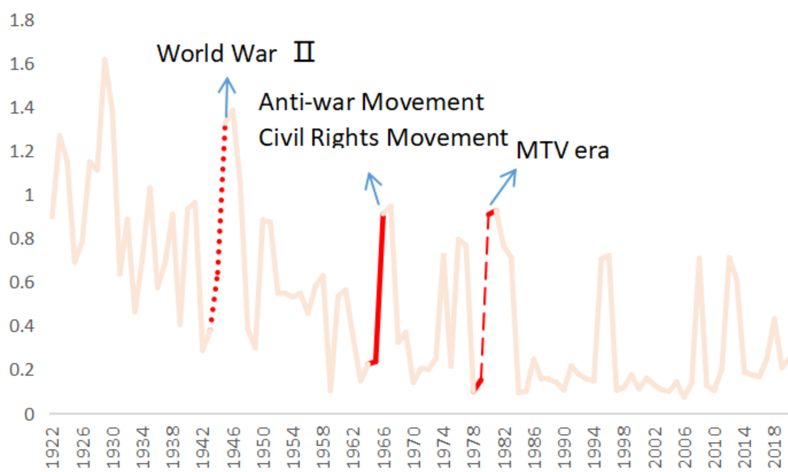


Figure 18:The difference of Euclidean distance

- Politics: World War II broke out in the 1940s, and jazz and swing music that reflected the wishes of the people and inspired people to keep hope gradually rose and became the mainstream.
- Society: In the 1960s, the anti-war movement and civil rights movement in the United States rose, and the sense of social responsibility of rock music was constantly awakened, arousing people's actions to save society, humans, and nature.
- Technology: In the 1980s, the arrival of music television (commonly known as MTV) led to the recording of more song videos by artists and bands. Michael Jackson's

creative music videos and skilled dance helped set popular music at that time standard. Hip-hop is accepted by the mainstream as a legitimate music genre.

10 Model Evaluation

10.1Strengths

- In the first question, we combine the network topology from a macro perspective and a micro perspective(the time relationship between influencers and followers) to construct the edge power

- In the second question, we use the Euclidean distance to measure the similarity and visually display the distribution of artists in the high-dimensional feature plane.

- In the fourth question, we use BFS of different depths to measure the changes in the musical distance between influencers and direct followers and profile followers.

- In the fifth and sixth questions, we establish influence and characteristic changes in the time dimension.

10.2 Weaknesses

- Although the Euclidean distance is intuitive, it may not perform as well as other measures.
- The establishment of influence and characteristic changes in the time dimension are not comprehensive enough.

A musical influence model

We are honored to that our team can help you. We use the... model to finally build a model to measure the influence of music. This model has strong use value and significant significance. model that measures musical influenc

The value of our method is reflected in the following aspects:

- Through the directed network, the artist's music influence, genre influence, influence direction and the age concentration of the artist are accurately measured, which will provide a solid data foundation for the evolution of music.
- Euclidean distance is used to measure the similarity of music, which is an indispensable information for studying the evolution of music.
- We propose a method to determine the genre, which helps us to judge the direction of music evolution
- Using our model, we can find truly infectious music characteristics, which helps us analyze the endogenous factors that affect music evolution
- We have proposed a method to find the hallmarks of music evolution, and at the same time analyze the characteristics of music revolutionaries, which will help us analyze and predict the evolution of music.
- We have identified indicators that reveal dynamic influencers, which helps us discover and analyze influencers who have influenced the evolution of music.
- In addition to the internal influence brought by the artist, we also propose indicators to identify external influences (politics, society, technology, etc.). At this time, we can comprehensively consider the influence of exogenous shocks and endogenous factors on the evolution of music.

Considering that the model we proposed this time is limited to certain genres and certain artists, if we are faced with richer data, we will make improvements in the following aspects:

- When the data becomes more abundant, we first need to preprocess and clean the data;
- Secondly, we will pay more attention to cover data, because cover songs are more popular and have important value for analyzing music evolution. We consider covering cover data as a separate data set, including cover singer, cover age and other data.
- Finally, we need to consider the overlap between the cover data set and the similar main data set given by the title and take further measures to solve it. .

After our research, we found that the evolution of music plays an important role in the development of human beings and the progress of human civilization. Take rock music as an example. Bob Dylan’s anti-war song *Blowin'in the Win* emphasizes social responsibility. Rock’s social responsibility consciousness awakens people's actions to save society, humans, and nature. Therefore, music becomes a political and cultural action based on the more precious wealth of mankind. We sincerely suggest that ICM can further study music and the influence of music on culture, which will be an asset to all mankind.

Appendixes

PageRank

```
#!/usr/bin/env python
# coding: utf-8
import pandas as pd
import networkx as nx
df = pd.read_csv("influence_data.csv", encod
ing=&apos;utf-8&apos;)
df_ve = df.to_numpy()
node = []
name = {}
delta = {&apos;0&apos;:0,&apos;10&apos;:0
,&apos;20&apos;:0,&apos;30&apos;:0,&apos;
;40&apos;:0,&apos;50&apos;:0,&apos;60&a
pos;:0,&apos;70&apos;:0,&apos;80&apos;:0
}
for i in range(len(df_ve)):
    n_f = str(df_ve[i,4])
    n_t = str(df_ve[i,0])
    if n_f not in node:
        node.append(n_f)
        name[n_f] = df_ve[i,5],df_ve[i,7],df_ve[i,6
],df_ve[i,4]
    if n_t not in node:
        node.append(n_t)
        name[n_t] = df_ve[i,1],df_ve[i,3],df_ve[i,
2],df_ve[i,0]
    delta_t = abs(df_ve[i,3]-df_ve[i,7])
    delta[str(delta_t)]+=1
order = [i for i in range(len(node))]
node_dic = dict(zip(node,order))
g = nx.DiGraph()
g.add_nodes_from(range(5603))
for i in range(len(df_ve)):
    n_t = str(df_ve[i,4])
    n_f = str(df_ve[i,0])
    g.add_edge(node_dic[n_f], node_dic[n_t])
name_byk = list(enumerate(name))
pr_dict = nx.pagerank(nx.reverse(g))
sorted_pagerank = sorted(pr_dict.items(), ke
y=lambda x: x[1], reverse=True)
for i, (id, pr) in enumerate(sorted_pagerank[
:30]):
    print (str(i+1) + &apos;.&apos;, name[na
me_byk[id][1]], pr)
```

depth k BFS

```
def mean_distance_at_k(k, influence_graph=g,
 ids=ids, dist_dict=dict_id):
    distances = []
    for u in ids:
        explored = set([u])
        last_fringe = [u]
        # Recursively explore successors until de
pth k is reached in BFS
        for i in range(k):
            current_fringe = set()
            for node in last_fringe:
                successors = influence_graph.success
ors(node)
            for successor in successors:
                if successor not in explored:
                    current_fringe.add(successor)
            last_fringe = current_fringe
            explored.update(last_fringe)
        for v in ids:
            if v in current_fringe:
                distances.append(dist_dict[str(u)][st
r(v)])
    return distances, np.mean(distances)
```