

---

There's always some evolution on music behind the auditory fest, and so are the cooperative effort of musicians across the generations behind the evolution of this great component of the society. Many artists have claimed that they got inspired by the artists of the same or different genre(s), making their works more popular and refreshing the overall atmosphere of music composing globally.

To explore the influence of the music, our team established a **Network Model** and analyzed the influence processes of musical evolution that occurred over time in one genre, making full use of the concerto provided by **PageRank Algorithm**. Moreover, this network model stood out when analyzing how our work express information about cultural influences of music in time or circumstances.

To find out the influence of music on the albums of the youngsters, our team established a **Principal Component Analysis Model**, having succeeded in finding the principal indicators for measuring the similarity between the artists or songs both within and beyond the genre and discovering whether the objects in the same genre possess more similarities. The model contains **Correlation Analysis**, **Hierarchical Analysis and Factor Analysis**, analyzing the similarity on the angle of artists and their works respectively.

To make the result more accurate, our team employed **Cosine Similarity** to calculate the similarity among genres separately. The cosine similarity, meanwhile, is also made full use of to make comparisons between similarities and influences among and within genres. our team also discovered what distinguishes a genre, and some potential relationship among different genres.

**Additional networks** was established to give the identification of whether there are characteristics that might signify revolutions (major leaps) in musical evolution from these data and what artists represent revolutionaries (influencers of major change) in our network. By analyzing the key phases of musical reforms which transformed into turning points on the line chart, our team found networks of at the decades matching the turning points.

**A bonus network** was employed for identification of whether the similarity data suggest that the identified influencers actually affect the music created by the followers. Our team have found that the network can also indicate the most contagious factors among all music characteristics when taking influencers and followers of one outstanding artist as an example.

Our team succeeded in revealing the dynamic influencers and how the genres and artists changed over time by establishing another **rich-in-chart model** conducted by the solo of analysis on provided stats.

The analysis on strengths and weaknesses of the models our team have built and the accuracy of them came up, analyzing the sensitivity of what our team have constructed, proving the stability of all above.

Finally our team issued a one-page document to the ICM society, our supervisor, about the value of using our approach to understanding the influence of music through networks. The document will also contain how our work would change with more or richer data.

**Keywords:** Influencing Network, In(ter)-genre Similarity, Musical Evolutions

## Contents

<b>1</b>	<b>Introduction</b>	<b>2</b>
1.1	Background . . . . .	2
1.2	Restatements of the Problem . . . . .	2
1.3	Our Work . . . . .	3
<b>2</b>	<b>Assumptions</b>	<b>3</b>
<b>3</b>	<b>Model 1: Network Model for Musical Influence</b>	<b>4</b>
3.1	Network for Indicators of Musical Influences . . . . .	4
3.2	Analysis on Impacts Conducted By Revolutions . . . . .	5
3.3	Identification on Impacts Brought By External Factors . . . . .	6
3.3.1	Model for Spread of the Influence on the Network Basing on scc . . . . .	6
3.3.2	Model for Examining External Disturbance on the Network . . . . .	6
<b>4</b>	<b>Model 2: Model for Analyzing Musical Characteristics</b>	<b>7</b>
4.1	Sub-Model for Developing Measures of Music Similarity . . . . .	7
4.1.1	Correlation Analysis . . . . .	7
4.1.2	Hierarchical Cluster Analysis . . . . .	8
4.1.3	Factor Analysis . . . . .	8
4.1.4	Result of This Sub-model . . . . .	10
4.2	Sub-model for Comparing Similarities/Influences Between/Within Genres . . . . .	11
4.2.1	Cosine Similarity . . . . .	11
4.2.2	Similarities Between and Within Genres . . . . .	11
4.2.3	Discussion of Songs of Different Genres . . . . .	12
4.3	Analyzing Turning Points in the Development of Music . . . . .	13
4.4	Analyzing Development of Music Over Time . . . . .	14
4.4.1	Sub-Model for Analyzing the Development of Artists Over Time . . . . .	14
4.4.2	Sub-Model for Analyzing Development in Genres Over Time . . . . .	16
<b>5</b>	<b>Discussions</b>	<b>18</b>
5.1	Sensitivity Analysis . . . . .	18
5.2	Strengths and weaknesses . . . . .	18
5.2.1	Strengths . . . . .	18
5.2.2	Weaknesses . . . . .	19
<b>6</b>	<b>One-page Document to ICM Society</b>	<b>20</b>
<b>References</b>		<b>21</b>
<b>Appendix</b>		<b>22</b>

**Notice:** All figures in the solution play critical roles in our analysis. Please enlarge them for details if necessary.

# 1 Introduction

## 1.1 Background

The past centuries worldwide have enjoyed an auditory fest of evolution of music. Many artists compose their tracks according to the heritage of the past generations, some novel instruments, and some external circumstances, exploring their preferred characteristics of their works. All these shifts and revolutions can be regarded as a concert of efforts conducted by both influencers and the followers. By studying the networks of songs and their musical characteristics, our team can get hold of the effects brought by the influencers of the songs' origins.

Our team was identified by the supervisor to develop a set of models which is able to measure musical influence. Our team should employ the provided data and later claim the value of our approaches.

## 1.2 Restatements of the Problem

The problem should be translated into the following tasks which requires to be finished:

- Create network(s) for the musical influences. Figure out the parameters of the influence and what they reveal.
- Use the statistics provided to explore the indicators to measure the similarity of music, analyzing the similarities between artists of the same and various genres simultaneously.
- Compare similarities and influences between and within genres and exploring the distinctions among genres.
- Indicate whether the similarity data, as reported in the data set *data\_influence*, suggest that the identified influencers in fact influence the respective artists.
- Identify whether there are characteristics that might signify revolutions (major leaps) in musical evolution from these data and what artists represent revolutionaries (influencers of major change) in your network.
- Analyze the influence processes of musical evolution that occurred over time in one genre and find the indicators which can reveal the influencers dynamically.
- Exploring for how our work express the information about influence of music in time or circumstances.

## 1.3 Our Work

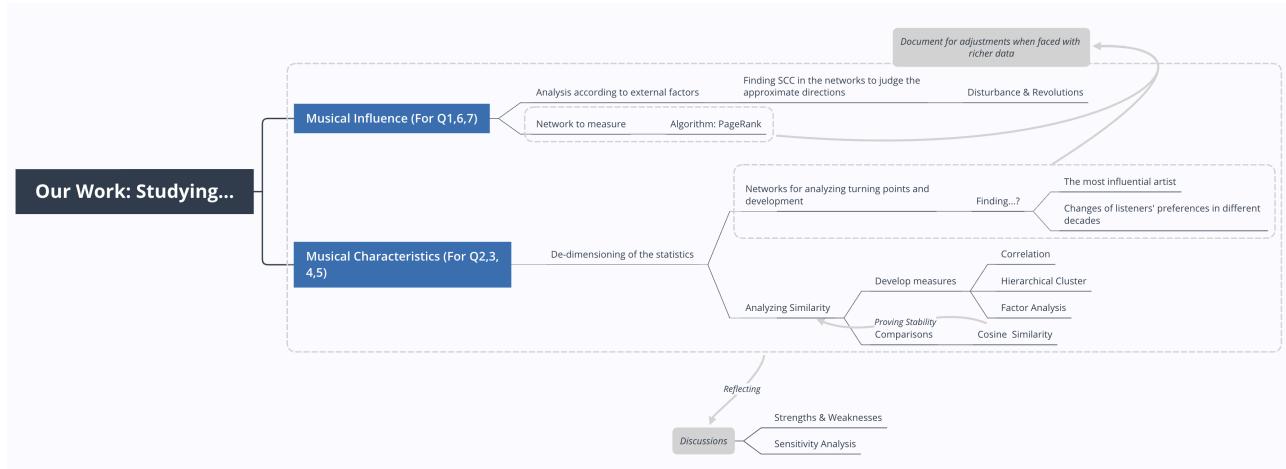


Figure 1: Sketch Map of The Works of Our Team

## 2 Assumptions

- The preferences of people with regard to music will not change greatly in a short time. In other words, popularity during other phases still stands the test of references.
- The overall trend of the musical market will remain stable.
- The sudden factors will leave no strike on the operation of models.
- Influencers of one artist feature different impacts on others. The more influential the influencer is, the more impact it conveys to his followers.
- Artist may take an active role in making a shift towards their own musical characteristics.

### 3 Model 1: Network Model for Musical Influence

Parameters	Definitions	Parameters	Definitions
$V$	Sets of Dots	$deg_{out}(u)$	Out-degree of the Node $u$
$E$	Sets of Directed-Edges	$deg_{in}(u)$	In-degree of the Node $u$
$G(V, E)$	Network Composed by $V$ and $E$	$PR(u)$	PageRank Value of the Node $u$
$(u, v)$	Directed-Edge Connecting $u$ and $v$	$q$	Damping Factor
$ V $	Size of $V$	$ E $	Size of $E$
$t$	Times of Running PageRank Algorithm	$G'(V', E')$	Network By Shrinking $G(V, E)$
$V'$	Sets of Dots Got By Shrinking $V$	$ V' $	Size of $V$
$E'$	Sets of Edges Got By Shrinking $E$	$ E' $	Size of $E$

The model is aimed at solving **Problem 1, 6 and 7** in the restatements, basing on the **Assumption 4 and 5**.

#### 3.1 Network for Indicators of Musical Influences

The algorithm of PageRank, initially, will grant the nodes with the same value of PageRank. During each phase of the calculation, the equation

$$PR(u) = \sum_{(v,u) \in E} \frac{PR(v)}{deg_{out}(v)}$$

will be the rule when the node  $u$  distributes  $PR(u)$  evenly to all his followers and makes a sum on the PageRank Value inherited from his influencers, which in turn triggers a new round of PageRank. The terminal PageRank occur when there's no change on PageRank of every node within the permission of the inaccuracy.

Existence of nodes of which the  $deg_{out}$  is 0 is inevitable. They may be affected by other factors which have nothing to do with the artists, so our team have to introduce the dumping factor to minimize the effect.

$$PR(u) = (1 - q) + q \cdot \sum_{(v,u) \in E} \frac{PR(v)}{deg_{out}(v)}$$

Generally, state-transition matrix is required for the calculation of the new value of  $PR$ .<sup>[1]</sup>

This network is purposed for **Problem 1** in the restatements. Our team established a directed network  $G(V, E)$ , where every node represents one artist and each of the directed-edge  $(u, v)$  symbolizes one relationship of influence ( $u$  for influencer, while  $v$  for follower). The network will display the influential network among the post-1960s artists.

The indicators of musical influences are as follows:<sup>[2]</sup>

- Degree Centrality (DC): The more the size of one node's out-degree in the network is, the more influential the artist at this node is.

- Betweenness Centrality (BC): The more the frequency of the node which exists on the shortest route between 2 other nodes is, the more influential the artist at this node is.

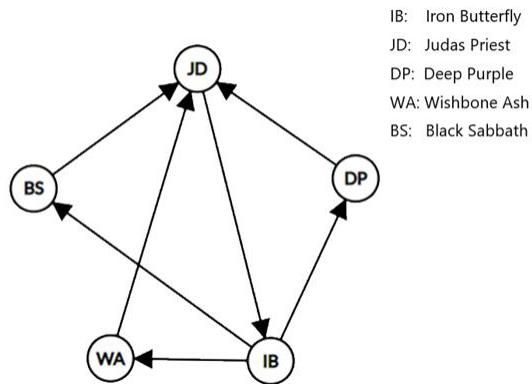


Figure 2: Sub-Network of the Active Artists in the 1960s

Our team consider the following sub-network where all nodes represent artists who became active in the 1960s, making the use of the PageRank Algorithm to find the artists who, in the whole network, are playing the most crucial role.

In the sub-network, *Iron Butterfly* represents the node which are quite critical according to BC, while those got from BC are named *Iron Butterfly* and *Judas Priest* respectively.

Our team then run the Algorithm of PageRank and discovered that the nodes follow the relationship *Iron Butterfly* > *Black Sabbath* = *Deep Purple* = *Wishbone Ash* > *Judas Priest* in the order of PageRank. The larger the value of PageRank the artist possesses, the more influential he is on the nodes.

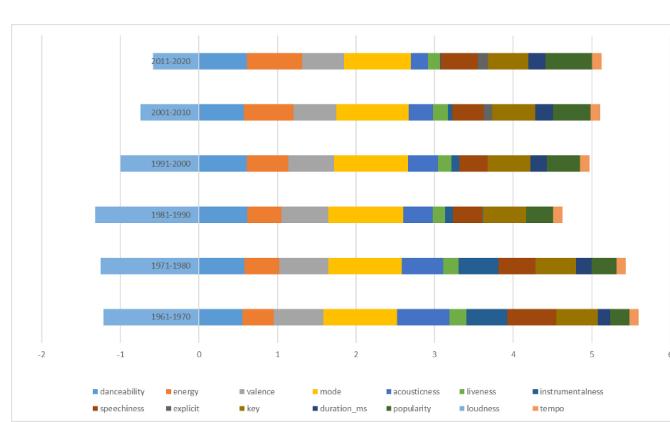
As a result, it's obvious to say that **the most influential node is *Iron Butterfly***.

### 3.2 Analysis on Impacts Conducted By Revolutions

The model is aimed at solving **Problem 6** in the restatements.

Our team measure the musical characteristics by making out the average of the music characteristics of one artist. Take the genre *Country* as an example. Our team depicted the characteristics in the same chart and succeeded in supervising the trend of musical characteristics from the 1960s to the 2010s.

The stacked column of musical characteristics are as follows.



It implies that **instrumentalness has dropped and so has acousticness since the 1980s**. In other words, instruments and acoustics have been increasingly absorbed into the genre *Country*. This may be attributed to the widespread development of computer. Similarly, indicators related to technology, such as *instrumentalness* and *acousticness*, may mirror the influence the technology has on music. The stable increase happening on *energy* can be originated from the 1960s thanks to the rising level of living citizens after WWII. The diversification of description also give rise to the increase of it.

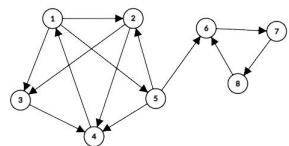
### 3.3 Identification on Impacts Brought By External Factors

#### 3.3.1 Model for Spread of the Influence on the Network Basing on scc

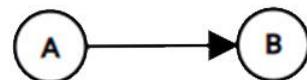
scc is the abbreviation of Strongly Connected Components.

Our team established scc in the directed map so that every node in the components can be influenced by each other. As is shown in figure (a) below, node 1, 2, 3, 4 and 5 are components of one scc which is marked as A, while node 6, 7 and 8 makes up scc B.

Our team found all scc by Tarjan Algorithm and depicted a directed acyclic graph on the basis of the initial chart by regarding every scc as a node.<sup>[3]</sup>



(a) Sketch Map of scc



(b) Sketch Map of Shrink Point

Figure 3: Sketch Maps With scc

In the acyclic graph, music characteristics of every node are the average value of those of the nodes in the scc.

#### 3.3.2 Model for Examining External Disturbance on the Network

Our team can discover the approximate directions of spread and the origin of the influence by sorting the acyclic graphs topologically<sup>1</sup>.

Our team added directed-edges to the network dynamically in the order of time according to the time when the follower turn active in every relationship of influences. Simultaneously, our team maintain the scc in the network and shrink them to DAG according to Incremental Strongly Connected Components Algorithm.<sup>[4]</sup> The moment when changes in the stats at the node whose in-degree is 0 in DAG turn obvious should indicate the possibility that external factors have left influences on the network.

---

<sup>1</sup>Topological sorting<sup>[5]</sup> runs on the following basis:

- Every node is included in the pattern and should appear for only once.
- Route from B to A should not exist if the node A stands in the front of the node B.

## 4 Model 2: Model for Analyzing Musical Characteristics

Parameters	Definitions	Parameters	Definitions
$D_{i,1}$	Danceability	$I_i$	Instrumentalness
$E_{i,1}$	Energy	$L_{i,2}$	Liveness
$V_i$	Valence	$S_i$	Speechiness
$T_i$	Tempo	$E_{i,2}$	Explicit
$L_{i,1}$	Loudness	$D_{i,2}$	Duration ms
$M_i$	Mode	$P_i$	Popularity
$s_n$	Comprehensive Score of Artist	$Y_i$	Year
$A_i$	Acoustics	$K_i$	Key
$R$	Constant for Correlation		

Table 1: Parameters and Their Definitions

The model is purposed at solving **Problem 2, 3, 4 and 5** issued in the restatement, basing on the **Assumption 1,2 and 3**.our team take the characteristics of music as significant indicators to measure the similarity among different works of music. We analyzed influences of artists from a different angle.

Our team discover the major influencers by employing the **Analytic Hierarchy Process, Factor Analysis and Entropy weight Method, etc.**

### 4.1 Sub-Model for Developing Measures of Music Similarity

Our team use this sub-model to solve **Problem 2** in the restatements. By using the measure, our team can find it easy to find out whether the artists within genre are more similar than artists between genres.

#### 4.1.1 Correlation Analysis

Our team initially introduced the Spearman Constant

$$r_s = 1 - \frac{6 \sum_{i=1}^n d_i^2}{n(n^2 - 1)}. \quad (1)$$

In the equation above,  $d_i$  represents the rank differences of the two data  $X_i$  and  $Y_i$ , which originates from the data sets  $X$  and  $Y$ . The rank of one datum should be the position of it after putting them in min-to-max order. It can also be proved that  $r_s$  can vary in the set [-1,1].

Figure 3(a) displays the value of the Spearman Constant with the help of the lightness of the color in each block. The heat map also illustrated the fact that **Energy, Loudness, Acoustics and Year are the strongest factors backing the stage of popularity.**

Our team then introduced Kendall's tau-b Constant of Correlation,  $\kappa_{TCC}$ , to prove the accuracy of the previous discoveries under the music of Spearman Constant, making the heat map in a similar way and can be seen on Figure 3(b).

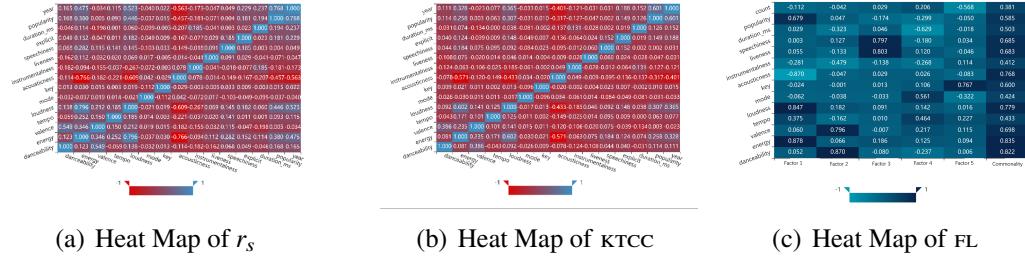
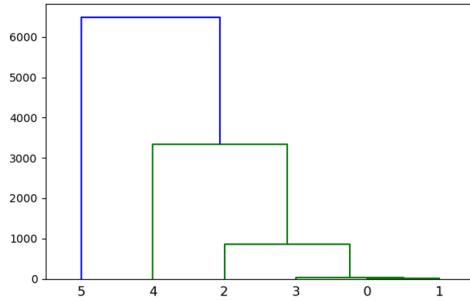


Figure 4: Heat Maps of Different Measures

We've discovered that the heat maps, (a) and (b), have succeeded in striking a chord with each other.

#### 4.1.2 Hierarchical Cluster Analysis



To simplify the influencing factors of the problem, we will only discuss the principal components figured out in the subsection 4.1.

Hierarchical cluster analysis features 3 types, Single Linkage, Complete Linkage and Average Linkage included. Our team chose the last one due to the fact that it can fit our model more perfectly. [6]

Our team then got the following chart and table with the help of our calculation.

Types	1			2		3		4	
	Danceability	Energy	Acoustics	Loudness	Popularity	Count			
Peak	0.072	-0.599	-0.909	2.934	0.523	78.903			
Skewness	-0.230	-0.267	0.592	-1.106	-0.507	7.333			
cv	0.248	0.369	0.858	-0.420	0.334	2.293			

cv:Coefficient of Variation

Table 2: Explanation of Some Components in the Tree Chart Above

#### 4.1.3 Factor Analysis

Our team calculated, making full use of the duet conducted by KMO Test and Bartlett Test, to discover that the parameters shown in the problems meet the demands of the conduction of Principal Component Analysis. Our team then can get Total Variance explained as follows.

It's universally acknowledged that the higher the PV is, the more important role the factor is playing.

Factor Number	Eigenvalue			Percentage of Variance After Rotation		
	Eigenvalue	PV	Total	Eigenvalue	PV	Total
1	3.713	24.757%	24.757%	2.987	19.916%	19.916%
2	1.646	10.971%	35.727%	1.844	12.293%	32.209%
3	1.336	8.904%	44.631%	1.715	11.436%	43.645%
4	1.200	8.000%	52.631%	1.309	8.727%	52.372%
5	1.119	7.459%	60.090%	1.158	7.718%	60.090%
6	0.992	6.616%	66.706%			
7	0.938	6.250%	72.956%			
8	0.830	5.532%	78.489%			
9	0.798	5.319%	83.807%			
10	0.741	4.937%	88.744%			
11	0.663	4.417%	93.161%			
12	0.373	2.487%	95.648%			
13	0.329	6.616%	97.843%			
14	0.204	1.357%	99.201%			
15	0.120	0.799%	100.000%			

PV: Percentage of Variance

Table 3: Total Variance Explanation Table

The statistics after rotation can also work when solving the formula of principal components.

Our team take the top 5 in the list as principal components because the eigenvalue has dropped below 1 after the Factor Number turns more than 6.

Our team then draw the Nightingale Rose-Chart and the Scree Chart to describe the extent to which the influences the principal components have on statistics. The amount of principal components which require to be chosen can be reflected by the slope of the drop on eigenvalue and can be adjusted according to Table 3.

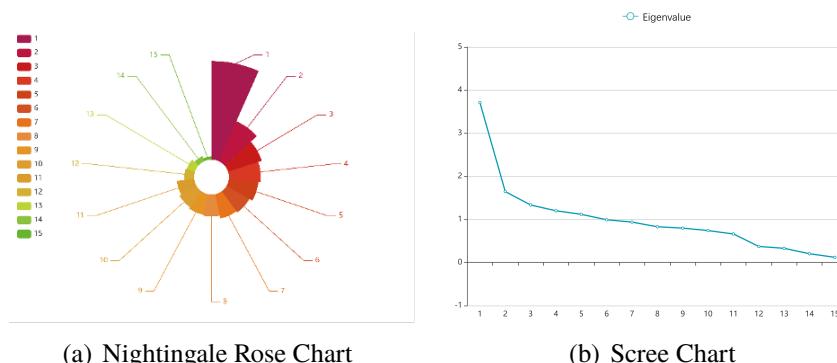


Figure 5: Charts to Find the Principal Components

Next, our team analyzed the significance of the potential variables with the heat map of factor loading shown in Figure 3(c). The nearer the value is to the 1 or -1, the more important the potential factors are.

Our team then are able to induce the formula of the model as follows.

$$F_i = \alpha_i D_{i,1} + \beta_i E_i + \gamma_i V_i + \delta_i T_i + \epsilon_i L_{i,1} + \zeta_i M_i + \nu_i K_i + \xi_i A_i + \kappa_i I_i + \pi_i L_{i,2} + \rho_i S_i + \sigma_i E_{i,2} + \eta_i D_{i,2} + \theta_i P_i + \iota_i Y_i \quad (2)$$

The value of  $i$  and its matching parameters in the equation (2) are shown below. All parameters but for  $i$  have multiplied by 1000.<sup>[7]</sup>

$i$	$\alpha_i$	$\beta_i$	$\gamma_i$	$\delta_i$	$\epsilon_i$	$\zeta_i$	$\nu_i$	$\xi_i$	$\kappa_i$	$\pi_i$	$\rho_i$	$\sigma_i$	$\eta_i$	$\theta_i$	$\iota_i$
1	38	77	-145	-89	128	-1	-44	-169	-74	-155	-58	182	58	358	359
2	469	20	449	-78	58	9	36	26	-200	-124	71	-8	-279	-58	-91
3	-216	339	170	420	224	58	17	-214	29	303	-50	-290	95	-177	-138
4	-11	34	-70	-74	2	-30	-2	67	-89	413	621	438	-77	-69	-17
5	134	63	100	-51	-34	-649	517	-69	232	-38	45	-26	283	-42	-31

Meanwhile, our team can safely reach an equation

$$F = \frac{199F_1 + 123F_2 + 114F_3 + 87F_4 + 77F_5}{601} \quad (3)$$

to get the comprehensive score of one artist,  $F$  in the equation (3), as is displayed in Table 3.

**The lower the value  $\delta = |s_1 - s_2|$  is, the stronger the link between these 2 scores' owners should be.**

Artist ID	Score
674417	1.68778
131866	1.53988
644216	1.43469
...	...
943649	-1.21662
1768491	-1.25053
100426	-1.26898

Table 4: Rank of Top 3 and Bottom 3 Artists' Comprehensive Score

#### 4.1.4 Result of This Sub-model

Having combined all discoveries of explorations above, we succeeded in reaching the conclusion that **Correlativity among artists of the same genre stands far more obvious than that among artists of various genres. In other words, artists of the same genre feature more similarities than those of inter-genre.**

## 4.2 Sub-model for Comparing Similarities/Influences Between/Within Genres

This model is aimed at solving **Problem 3** in the restatements.

### 4.2.1 Cosine Similarity

Our team mentioned the index  $F$  to find out  $\delta$ , the similarity between the works of music, previously. The  $\delta$  between different genres, however, will bring quite significant inaccuracy despite the fact that the absolute difference between their individual values. Thus, for better comparison between similarities and influences between and within genres, we're now introducing the **Cosine Similarity**, cs, to solve the problem of the unification among the measurements of the parameters.

cs employs the cosine of the angle of 2 vectors in one certain space to measure the differences between the 2 individuals they represent. cs focus on the differences on directions rather than lengths or anything else. It can be calculated in the following way:

$$\text{sim}(X, Y) = \cos \theta = \frac{\mathbf{x} \cdot \mathbf{y}}{|\mathbf{x}| |\mathbf{y}|} \quad (4)$$

Our team have to discuss the  $\cos \theta$  on the multi-dimension level thanks to the concert conducted by various factors. It can be calculated in the following equation if the n-dimensional vectors  $\mathbf{x}$  and  $\mathbf{y}$  possess the coordination  $(x_1, x_2, \dots, x_n)$  and  $(y_1, y_2, \dots, y_n)$  respectively.

$$\cos \theta = \frac{\sum_{i=1}^n x_i y_i}{\sqrt{\sum_{i=1}^n x_i^2} \sqrt{\sum_{i=1}^n y_i^2}} \quad (5)$$

Having calculated the index  $\text{sim}$  among genres, our team have discovered that the most significant indicators, which shoulder the most weight, are  $D_{i,1}$ ,  $Y_i$  and  $A_i$ .<sup>[8]</sup>

### 4.2.2 Similarities Between and Within Genres

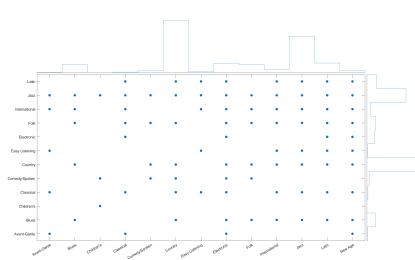


Figure 6: Chart of Link Between Genres

After processing the provided data by wrapping the genre *unknown* out, putting the index of various factors of musical work into calculation of cs among genres and making out the calculation the average of cs of the song of the same genre, we can gain access to the following chart, where similarity stay high in the red blocks and low in the green ones.



The figure shows that:

- The genres *pop/rock*, *Latin*, *R&B*, *reggae*, *religious* and *vocal* feature high similarity with each other, making the fact clear that the role they played during the development of music have been magnificent.
- It's not necessary that the works get the most of the influences from those of the same genres.
- Works of music may leave rare influence on others if they follow different paths of era or targets; But these influences can't be ignored.

#### 4.2.3 Discussion of Songs of Different Genres

Our team solved the average of all the information of the works of the same genre and got the following bar chart after the process of standardization of all stats.

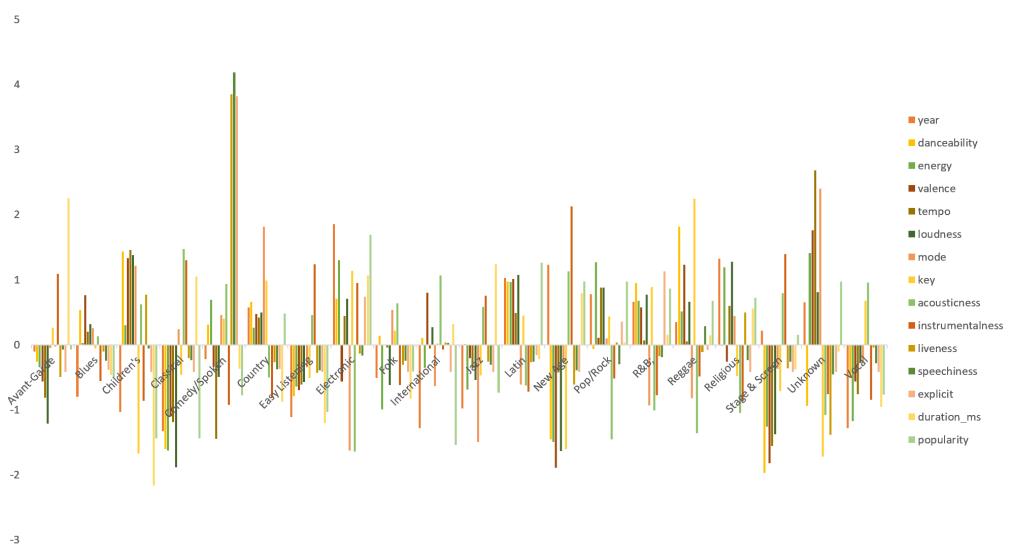


Figure 7: Bar Chart of Characteristics' Info

The figure above indicates that:

- *Classical international, Electronic and Religious* come into existence much later than others. The first one can be categorized as a classics, while the latter 2 can be regarded as modern music.
- *Reggae* and *Children's* stand out when it comes to the best dance-suited tracks. They possess optimistic tracks, the ones which *New Age, Stage & Screen* and *Comedy/Spoken* do not feature, as well.
- *Electronic* and *Pop/Rock* feature the most energetic tracks while the *New Age* and *Classical* stand the opposite.
- *Pop/Rock* and *Children's* feature the most strong tempo which *Stage & Screen* don't possess.
- *Children's, Religious* and *Latin* feature loudness.
- *Reggae* and *Electronic* are characterized with high pitch while the genre *Children's* feature low pitch.
- When it comes to the percentage of instrumentality, *New Age* and *Stage & Screen* stand out to be the highest, while *Classical New Age* turn out the lowest. More elements of electronics, namely, take place in *Electronics*.
- *Electronics, Latin* and *New Age* enjoy top popularity, while the scene of *Comedy/Spoken* and *Children's* may be hot when covered live.

### 4.3 Analyzing Turning Points in the Development of Music

The analysis is purposed at solving **Problem 5** in the restatements.

After the factors analysis conducted in **4.1.3** with the help of using the parameters in the file *data\_by\_year.csv*, we can get the value of  $F$  in the equation (3). Our team have to conduct a process of standardization for  $F$  due to the large amount of it, having succeeded in getting the new parameter  $F'$ . The line chart of  $F'$  in terms of year is as follows.

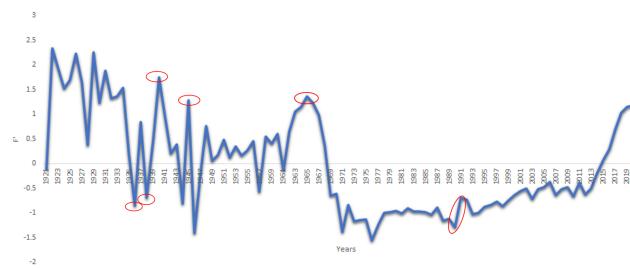
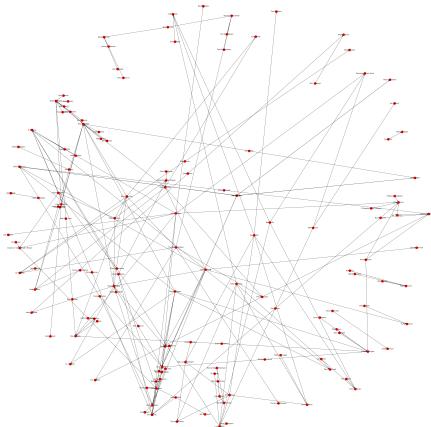


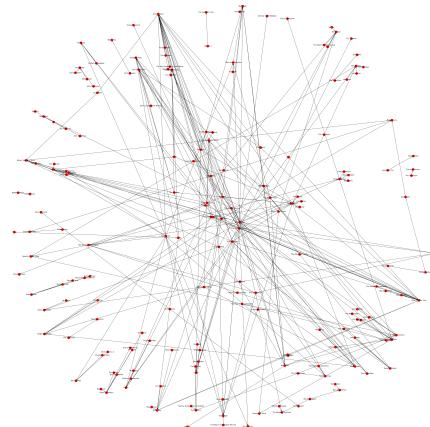
Figure 8: Line Chart Featuring  $Y_i$  and  $F'$

The figure indicates that the development of music, as is illustrated in the circle-marked point, have experienced great shifts in 1936, 1938, 1945 and 1965.

Having taken the years into account, we found out the revolutionary musicians in the 1940s and 1960s by depicting the networks respectively.



(a) The 1940s



(b) The 1960s

Figure 9: Networks in the 1940s &amp; the 1960s

Using the figure above, we can safely say that *vocal* and *jazz* enjoyed popularity in the 1940s, the era when Charlie Parker, John Coltrane and Lester Young leave the deepest influence on the atmosphere. *Pop/Rock*, as is reported in the network, were given birth to by the music of the mainstream. *Blue*, *R&B* also got influenced by *jazz and country*.

Similarly, *Pop/Rock* turn mainstream in the 1960s with over a half of the artists enjoying it. *Blue*, *jazz* and *R&B* also gained widespread applause in the society, making great impact on *Stage & Screen* *Reggae* and resulting in the birth of *Reggae*. The network indicates that The Beatles, Alice Cooper, Paul weller and Soundgarden turned the most popular band or musical individuals.

In short, the charts above displays the fact that **great shifts can be partly attributed to the existence of Pop/Rock in the 1940s and Reggae in the 1960s**.

## 4.4 Analyzing Development of Music Over Time

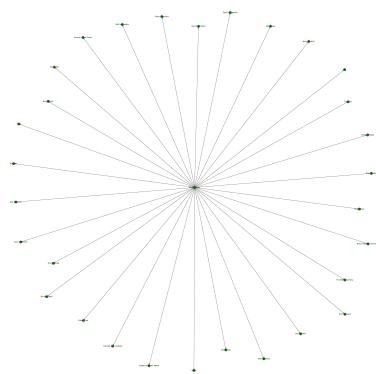
This model is a solution to **Problem 4** in the restatements. Our team solve this problem by analyzing with angles of artists and genres respectively.

### 4.4.1 Sub-Model for Analyzing the Development of Artists Over Time

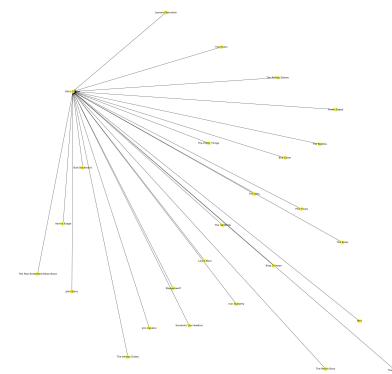
According to the analysis of stats, we succeeded in finding that when it comes to followers of one artist, over half of the followers will create music of the same genre with the influencer's. For the exceptions of the musicians above, however, their compositions of other genres may feature high similarity with the works of their influencers. *Rehab*, an *R&B* work, and *Amateur (Rock Me Amadeus)*, a *Latin* work, feature the incredible similarity of 99% with each other. Other compositions of these exceptions play

a great role in the phase of innovation. The Beatles once got affected by Slim Whitman, a *Country* musician, King Curtis and Barrett Strong, *R&B* musicians. It was influences from those artists that give rise to the band's widespread influence and its high popularity. As a result, **the influencers actually affect the music created by the followers.**

The researches conducted in **4.1** have told us that the indicators bringing the strong relationship with popularity are **energy, loudness and acoustics**. As a result, we're firmly convinced that these 3 parameters **plays the strongest pushing role for the influence on the followers** as well. To prove the conclusion we made, we studied the influencers and followers of Alice Cooper, the origin of *Hard Rock*, *Heavy Metal* and *Pop/Rock*, of which the result is shown below.



(a) Followers of Alice Cooper



(b) Influencers of Alice Cooper

Our team then analyzed key indicators of random 10 artists in the network and standardized the stats and depicted the radar charts below.

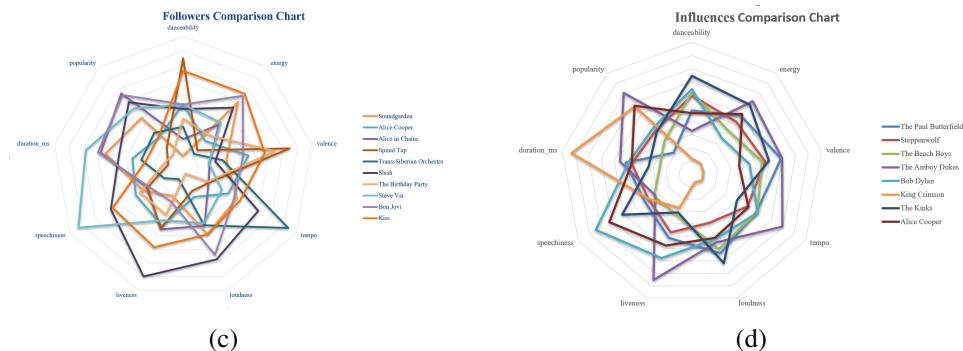


Figure 10: Network and Radar Charts of Followers and Influencers

As is illustrated, loudness and danceability are quite similar. In another word, Alice Cooper succeeded in absorbing these 2 factors into his work, thus gaining higher popularity than the former giants on their shoulders. Followers of Alice Cooper feature high similarities in valence, danceability and energy, but only 2 of them gained more popularity than their origins, indicating A.C.'s high musical achievements.

In short, **danceability and energy are more contagious than others. Meanwhile, the cliché better navigators, better destinations is still making sense.**

#### 4.4.2 Sub-Model for Analyzing Development in Genres Over Time

The rock carved with the name of the genres will stand the tests of the scrubbing of the history. The characteristics of works of one genre may change according to the influences from other genres. Influences may occur among artists of the same genre; even some contagious factors of works from different genres may get inherited by their followers and the influences, if piled up, may bring up the storm of reforms. The increment may also increase as the number of artists within the genre increases. This effect may bring shifts on characteristics of some genres and the birth of new genres. The coming-up research is focused on the development of genres over time.

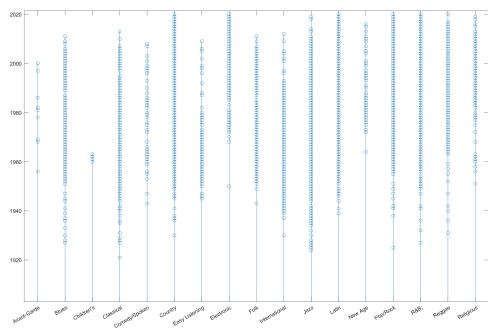
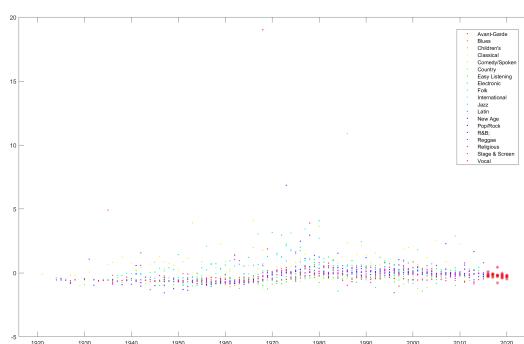


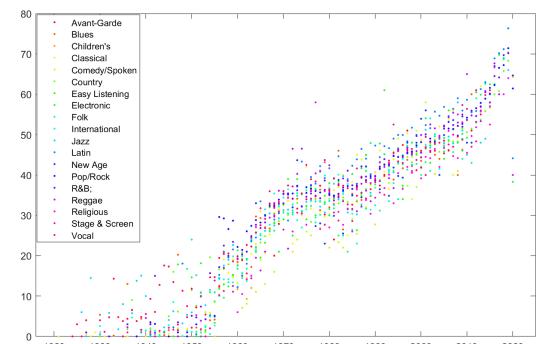
Figure 11: Genre-Year Stem Chart

The stem chart was depicted according to the performance of works of each genre every year. Our team can find from the chart that *classical*, *jazz*, *pop/rock*, *R & B*, *country* and *blue* came into being much earlier than others, leaving great and continuous impact on the novel genres from the mid to the terminal decade of 20th century. Many pieces music which are newly-born, including *avant-garde*, *children's*, *electronic*, *new age* and *reggae*, are still giving out their light and heat in this well-digitalized society, despite the weak prospects possessed by *avant-garde*, *children's* and *folk*.

To analyze the musical characteristics of different genres, the indicators can be de-dimensioned with the factor analysis method shown in 4.1.3 and shown in the scatterplot chart in Figure 11(a).



(a) Year-Characteristics Scatterplot Chart



(b) Year-Preferences Scatterplot Chart

Figure 12: Scatterplot Charts for Development on Genres

The chart shows that **some genres may play the leading role during a particular period**, with *jazz*, *country*, *pop/rock* and *Latin* stands out in the modern society. **Characteristics may make slight changes to follow the rhythm of the era.**

The preferences of people in different decades are illustrated in Figure 11(b). The scatterplot chart can dynamically mirror the genres which enjoy popularity according to the time. Having taken historic, social, political and technological changes into consideration, we discovered that:

- **Jazz began to grow in the 1950s.** The birth of *jazz* can be attributed to the flow-in and the development of black-composed music and the mainstream of *pop/rock* and *blues*. The combination of movies and music redefined the existence of these special kinds of art. Making the whole thing better, the unique musical system *jazz* constructed give birth to the development of modern pop music, *blues*, *R & B* and *pop/rock* being the branches of it.
- **The booming of knowledge era took place in the 1960s, inspiring the musicians of the significance of combining technology and their works.** Branches of *Electronics* and *jazz* stood out, backing the implementation of the inspiration on to everyone's mind.
- **The voice for liberty and equality swept the air worldwide in the 1970s, the decade when music of all genres were enjoying their booming era, pushing musicians to release the albums which call for freedom.** Speech delivered by Martin Luther King, Jr. and Mao Zedong succeeded in igniting the souls of the revolutionists worldwide, encouraging them to compose tracks of radical feelings. The works expressing the true emotions of their authors soon gained great popularity worldwide, criticizing the cruel ruling system, the unfairness they're faced with and catching the ears of listeners. Bob Dylan, for example, began to get the key to hall of fame with the metal *Blowing in the Wind*, which was born in that decade as well.
- The technology of Internet has made the spread of masterpieces easier for fans worldwide since the wide application of computers and other network-services-equipped devices. Existence of engineers for music and DJs has begun to turn more frequent, making the composition of tracks no longer limited to the inventions of inspiration. In other words, **A giant leap ahead in music has taken place in the late 1990s and the 21th century due to the small step forward in the development of computer.**

To conclude, **development in genres can be accelerated by both external and internal factors**. The overall trend of the development is to cater to the demands of people and to make full benefits of the cutting-edge technology at that age, absorbing the critical recipe for the popularity from works of other genres.

## 5 Discussions

### 5.1 Sensitivity Analysis

The results of PageRank Algorithm feature close links with the parameter  $q$ . In the second equation in **3.1**,  $(1 - q)$  represents possibility of self-changes under the condition that one node is spared from the influence of others. The PageRank Value in **3.1** is calculated under the condition  $q = 0.85$ . The varying of PageRank Value, which follows the varying  $q$ , is as follows:

q	Black Sabbath	Wishbone Ash	Iron Butterfly	Judas Priest	Deep Purple
0.05	0.1958	0.1958	0.1959	0.1920	0.1958
0.15	0.1865	0.1865	0.1868	0.1756	0.1865
0.25	0.1755	0.1775	0.1763	0.1588	0.1755
0.35	0.1627	0.1627	0.1642	0.1415	0.1627
0.45	0.1477	0.1477	0.1499	0.1235	0.1477
0.55	0.1301	0.1301	0.1329	0.1046	0.1301
0.65	0.1095	0.1095	0.1127	0.0847	0.1095
0.75	0.0850	0.0850	0.0883	0.0632	0.0850
0.85	0.0558	0.0558	0.0584	0.0399	0.0558
0.95	0.0205	0.0205	0.0217	0.0141	0.0205

It can be founded that the min-to-max order of PageRank Values remain the same when  $q$  varies. The algorithm, thus, shows great stability in the sub-network.

### 5.2 Strengths and weaknesses

#### 5.2.1 Strengths

- The network of influence can reveal the spread of the influence in it dynamically.
- The criticality of every node can be evaluated from different perspectives thanks to employing BC, DC and PageRank Value.
- High efficiency and excellent quality make it easier for analysis on ring-free map to be finished within a quite short time.
- Our team have simplified the description of similarities among works in various parameters by de-dimensioning the stats and maintaining high describability.
- Our team introduced cs at the condition of high accuracy of similarities so that the errors can hardly occur even if the data sets get larger.
- Researches on a single artist can be transplanted for that on other artists and can reach safe conclusions as well.

### 5.2.2 Weaknesses

- Complexity when applying BC may deny our this indicator when faced with piles of statistics.
- Extreme stats' existence may bring obvious fluctuation to the average.
- What we've founded are just limited to the statistics of reforms and influences among genres, which is much simpler than the actual situation of musical development.
- Inaccuracy may occur if time keeps going due to the assumption that value of parameters should not change with the time (included in **Assumption 1**).

## 6 One-page Document to ICM Society

The fence of these provided data sets should not restrict our models' usage; The genres in the entire world of music is not completely covered in these sets. The uncertainty, thus, stands out when the number of genres increases.

Networks, revealing the link of influence among artists, may be increasingly complex thanks to the network of addition as the time keeps going. This in turn provides the networks opportunities to mirror the process of the spread existing in them. Our team are then able to explore the significance of every node in the network, finding out the composer of the musical revolution and the clubs which are leaving influences on each other. The link of influence among genres and the style of one genre can be also discovered with the help of the concerto of network.

Faced with richer data, the network can still maintain the scc in the linear time complexity, employing Inc-scc, turning the network to DAG and explore the spread of the influence in the networks. DC can be calculated within a quite high rate, while the calculation of BC requires the time of  $O(|V||E|)$ <sup>[9]</sup>. In other words, when the number of nodes and edges gets higher, the digits displaying on the timer should follow its path when it comes to calculation of BC. Common PageRank algorithm requires the time of  $O(|V|^2t)$ ; but fortunately, plenty of algorithms have come to existence to optimize the solution of the PageRank Values<sup>[10]</sup>, which in turn gives us the chance to process statistics in larger scales, including the algorithm MMSI.

Our team can take historic events and social reform into account and analyze distribution of artists of different genres and shifts on music characteristics for exploration of the reasons behind the rise and fall, the changes and the spread of the genres worldwide. Meanwhile, we can get a good command of the influences among culture, music and society by getting hold of the directions of spread of music shown in the network of influence and the related profiles. Our team can then step into the analysis of the reasons behind the great influences of some particular artists under the particular era, along with features of music characteristics of the musical masters in different decades.

The ability of simulation for factors of similarities stands the test of richer data due to the de-dimensioning methods, including factors and cluster analysis. The analysis based on CS is suitable for spaces of any dimensions of vectors when it comes to researching on genres. Alternatively, we can still measure internal cohesion, figure out accurate range, power and trend of influences and even predict the coming-up revolutionary events in the near future when faced with data sets in a larger scale.

The analysis of the network isn't only limited to the accordance to the changes of music characteristics, but also combined with the background of events at that time, including shifts in the social structure, scientific discoveries and revolutions. Taking all above into account, the distribution of artists of different genres can be explored in search of the reason behind the rise and fall of different genres, the styles of the masterpieces and the spread among regions. The richer the data is, the more history-related profiles we can get access to and the more evidences about analysis on the correlation between cultural and musical spread, thus proving the discoveries regarding the reasons behind the strong influences of certain artists in one certain decade and the features of these musicians' masterpieces.

All about the value of using our approach to understanding the influence of music through networks have been on display above. Our team is hoping that the report is able to compose a wonderful track of sonata which combines mathematics and music, and that the report can be adopted for further studies.

## References

- [1] Page.L. Brin.S. *The anatomy of a large-scale hypertextual web search engine*. Computer Networks and ISDN Systems. 1998, 30: 107–117. ISSN 0169-7552. doi:10.1016/S0169-7552(98)00110-X.
- [2] <https://en.wikipedia.org/wiki/Centrality/>
- [3] Tarjan, Robert (1972-06-01). *Depth-first search and linear graph algorithms*". SIAM Journal on Computing. 1(2): 146–160. doi:10.1137/0201010. ISSN 0097-5397.
- [4] X.Liao, Y.Chen, Y.Zhang, et al. *An efficient incremental strongly connected components algorithm for evolving directed graphs* (in Chinese). Sci Sin Inform, 2019, 49: 988–1004, doi: 10.1360/N112018-00125
- [5] Kahn, Arthur B.(1962) *Topological sorting of large networks*. Communications of the ACM, 5 (11): 558–562, doi:10.1145/368996.369025, S2CID 16728233.
- [6] X.Shi, Y.Qian, F.Li. *MS<sup>2</sup>BC-a multi-view space structure-based clustering algorithm*. Journal of Southwest University, 2020, 42(11): 59-67. doi: 10.13718/j.cnki.xdzk.2020.11.007
- [7] M.R.Silva. O.A.Carvalho. R.F.Guimarães. R.A.T.Gomes. C.R.Silva. *Wheat planted area detection from the MODIS NDVI time series classification using the Nearest Neighbor Method calculated by the Euclidean Distance and Cosine Similarity Measures*. Geocarto International,2020,35(13).
- [8] Y.Shao. *Data similarity weight adjustment algorithm based on mathematical graph theory analysis*. Journal of Hunan University of Arts and Science (Science and Technology), 2021, 33(1): 20-24. doi: 10.3969/j.issn.1672-6146.2021.01.005
- [9] U.Brandes. *A faster algorithm for betweenness centrality*. Journal of Mathematical Sociology 25(2):163-177, 2001.
- [10] Z.Yang. *An effective accelerating iterative algorithm for computing PageRank*. Lanzhou University. 2019.

## Appendix

### I Codes for Solving PageRank Values

```

1 #include<bits/stdc++.h>
2 using namespace std;
3 const int maxn = 17;
4 int g[10][10], n = 5, cnt;
5 double PR[10], S[10][10], deg[10], a = 0.7, dv;
6 bool mul(){
7     cnt++;
8     if (cnt > 100) return 1;
9     double res[10], change = 0;
10    for (int i = 1; i <= 5; i++) {res[i] = 0;}
11    for (int i = 1; i <= 5; i++) {
12        for (int j = 1; j <= 5; j++) {
13            if (S[j][i] == 1) {res[i] += a * PR[j] / deg[j];}
14        }
15        res[i] += dv;
16        change += abs(PR[i] - res[i]);
17    }
18    for (int i = 1; i <= n; i++) {PR[i] = res[i];}
19    if (change < 0.0001) {return true;}
20    return false;
21 int main() {
22     g[1][4]=g[2][4]=g[3][5]=g[3][2]=g[3][1]=g[4][3]=g[5][4]=1;
23     deg[1] = 1; deg[2] = 1; deg[3] = 3; deg[4] = 1; deg[5] = 1;
24     for (int i = 1; i <= 5; i++) {
25         for (int j = 1; j <= 5; j++) {
26             if (i != j) {
27                 if (deg[i] != 0 && g[i][j]) {S[i][j] = 1.0 / deg[i];}
28                 else {S[i][j] = 0.2;
29                     deg[i] = 5;
30                     for (int k = 1; k <= 5; k++) S[i][j] = 1;}
31             }
32         }
33     PR[i] = 0.2; dv = (1.0 - a) / 5;
34 }
35 while (!mul());
36 for (int i = 1; i <= 5; i++) {printf("%.5lf\n", PR[i]);}
37 }
```

## II Codes for Factor Analysis (Key Part)

```
1 clc
2 clear all
3 A=xlswread(' C:\Users\..\Desktop\ influence_data.xlsx ');
4 a=size(A,1);
5 b=size(A,2);
6 for i=1:b
7     SA(:,i)=(A(:,i)-mean(A(:,i)))/std(A(:,i));
8 end
9 CM=corrcoef(SA);
10 [V,D]=eig(CM);
11 for j=1:b
12     DS(j,1)=D(b+1-j,b+1-j);
13 end
14 for i=1:b
15     DS(i,2)=DS(i,1)/sum(DS(:,1));
16     DS(i,3)=sum(DS(1:i,1))/sum(DS(:,1));
17 end
18 T=0.9;
19 for k=1:b
20     if DS(k,3) >= T
21         com_num=k;
22         break;
23     end
24 end
25 for j=1:com_num
26     PV(:,j)=V(:,b+1-j);
27 end
28 new_score=SA*PV;
29 for i=1:a
30     total_score(i,1)=sum(new_score(i,:));
31     total_score(i,2)=i;
32 end
33 result_report=[new_score,total_score];
34 result_report=sortrows(result_report,-4);
```

FactorAnalysis.m (Partial)

### III Codes for Shrinking Points (Key Part)

```
1 void tarjan(int u){  
2     dfn[u] = low[u] = ++tot;  
3     instack[u] = 1; stack[++top] = u;  
4     for (int i = head[u]; ~i; i = e[i].nxt) {  
5         int v = e[i].v;  
6         if (!dfn[v]) {  
7             tarjan(v);  
8             low[u] = min(low[u], low[v]);  
9         } else if (instack[v]) {  
10             low[u] = min(low[u], dfn[v]);  
11         }  
12     }  
13     if (dfn[u] == low[u]) {  
14         cnt++; int v = 0;  
15         do {  
16             v = stack[top--]; instack[v] = 0;  
17             col[v] = cnt; scc[cnt].push_back(v);  
18         } while (v != u);  
19     }  
20 }
```

ShrinkingPoints.cpp (Partial)