CS224W Final Report: Influence Networks in Popular Music

Marco Alban, Vivek Choksi, Stephanie Tsai

Stanford CS Department

Abstract

In our project, we address the overarching question: can we use audio data to identify and characterize musical influence relationships? Using human-annotated data about popular musicians from allmusic.com, we construct a musical influence graph wherein a directed edge between two artist nodes signifies a musical influence relationship. We perform link prediction on this graph by training a classifier with input features based on graph structure as well as song audio content. The results demonstrate that features based on song audio content have predictive power, and our highest-scoring feature combination achieves an area under the receiver operating characteristic (ROC) curve of 0.867. These findings show that audio features of songs can, to some extent, identify musical influence relationships between artists, and this finding affirms the promise of data-driven analyses of the progression of musical creativity.

---

1. Introduction

The study of music has not traditionally involved quantitative analysis. However, with advances in learning algorithms and data availability, it is becoming more feasible to apply data to musicology in a meaningful way.

In this paper, we study music influence relationships between popular musicians in recent history (1960 - 2010). We do this by using features based on graph structure and song audio content as inputs to a link prediction model applied to a music influence graph. We first scraped human-annotated musical influence relationships from allmusic.com to construct a graph of musical influence, where each directed edge in the graph denotes a musical influence relationship between the artists at the source and destination nodes. In Section 4, we present an analysis of the graph's structure and properties. Next, we pose the problem of link prediction on this graph, and we perform link prediction by training a classifier using features based on graph structure as well as features

Email addresses: `marcoal@stanford.edu` (Marco Alban), `vhchoksi@stanford.edu` (Vivek Choksi), `stsai612@stanford.edu` (Stephanie Tsai)

based on song audio. Finally, we discuss strengths and weaknesses of our results and methodology.

At a high level, the question we intend to answer is: can we predict artist-to-artist influence using audio features describing songs' musical content? Our findings are positive, affirming the promise of applying quantitative analysis to the study of music and its evolution.

2. Literature Survey

2.1. Prior Work in Link Prediction in Social Graphs

In order to better understand the difficulty of link prediction in our model, we first perform a literature survey on link prediction in social graphs. We focus on link prediction schemes that use the network structure. The following methods propose a score function for the edge (u, v) and then output k link predictions as the top-k highest-scoring edges.

Link prediction for edge (u, v) can be formalized as a process in a network growth model. The authors in [4] show that using the number of common neighbors for the two nodes has a correlation with network growth prediction in an author collaboration network. Note however that if two nodes have high degree in the graph, this measure is skewed. The Jaccard coefficient and Adamic-Adar score (proposed in [1]) both address this skew by dividing the number of common neighbors by some scaling factor.

The authors in [3] propose a method inspired by the resource allocation process; this is similar to the method in [1] but without the log scaling. Other scoring/similarity measures are proposed by the authors in [9] and [10].

Using preferential attachment as the model for network growth, we can score each edge by using a product of the magnitude of their degrees. Barabasi et al., [5] use this model of network growth in evaluating link prediction in author-author collaboration networks.

The authors in [6] formalize how the concept of random walks can be used to compute the relevance of a node in a graph. By using a bounded teleport set (Personalized Page Rank score), this can be considered a distance metric between nodes. As noted in [2], the personalized page rank score can be informative in link prediction.

2.2. Prior Work in Audio Signal Level Topic Modeling

We also present a literature review into prior work that quantifies songs' musical characteristics to predict or analyze connections between artists.

In [13], Bryan and Wang look at influence between artists and genres at a high level. Specifically, they examine a music sampling graph, in which a directed edge denotes an instance of one artist sampling from another artist's work. They use this graph to study artist-to-artist and genre-to-genre influence. We wish to predict influence edges as well, but instead of looking at specific music sampling between songs, we look at similarity in musical attributes of songs. Thus, we aim to create graph structures similar to [13], but using a lower-level data set instead.

We look to [11] and [12] for two methods of applying audio signal level topic modeling to study musical influence between artists.

The authors in [11] apply topic modeling to song audio data to analyze the spread of musical influence. Specifically, for each song, they consider the dominance of each of the 12 notes in the chromatic scale, the timbre of these notes, and other more general musical qualities such as tempo, key, and time signature. After creating a complete vocabulary for each of these topics, they extract one word (in this case, a single quantitative measure) per topic for each song. Each of these topics is then normalized in relation to the others, and a probability model is constructed, giving the topic-influence score (i.e. how much weight a particular song should have for a given topic). In addition, a separate probability model is constructed for each epoch in music history (each song belongs to a discrete epoch), thus allowing an analysis of the influence of songs in linear time.

Unlike Shalit et al., who use topics mostly comprised of weights of exact musical notes, Mauch et al., [12] define sixteen more general audio feature topics (described in more detail in Section 3.2) in order to study the evolution of musical creativity and diversity. Mauch et al. track the rise and decline of each topic over time and contextualize this analysis with a discussion of the evolution of musical styles in recent history.

3. Dataset

We used two primary datasets: (1) an influence graph scraped from allmusic.com, and (2) audio features from songs on US Billboard Hot 100 between 1960 and 2010.

3.1. Influence graph

Our first dataset, the influence graph, describes which artists influenced which other artists as annotated by allmusic.com's musicologists. We obtained this influence graph using the API of allmusic.com's data provider, Rovi Corporation. We scraped the influencer and follower relationships between artists in our second dataset, and our resulting graph contains 3228 artist nodes and 4734 influence edges. Figure 1 shows the influence graph, filtered to only display artists with a degree of 20 or higher. This filter captures the most popular artists whose songs reached the US Billboard Hot 100 between 1960 and 2010, ranging from classic jazz and swing musicians (e.g. Louis Armstrong, Ella Fitzgerald) to modern-era hip-hop artists (e.g. Dr. Dre, Snoop Dogg).

3.2. Song data

Our second dataset, consisting of audio features from popular songs, was made publicly available in CSV format in Mauch et al., [12]. It contains features extracted from 30-second audio clips from 17094 songs on the US Billboard Hot

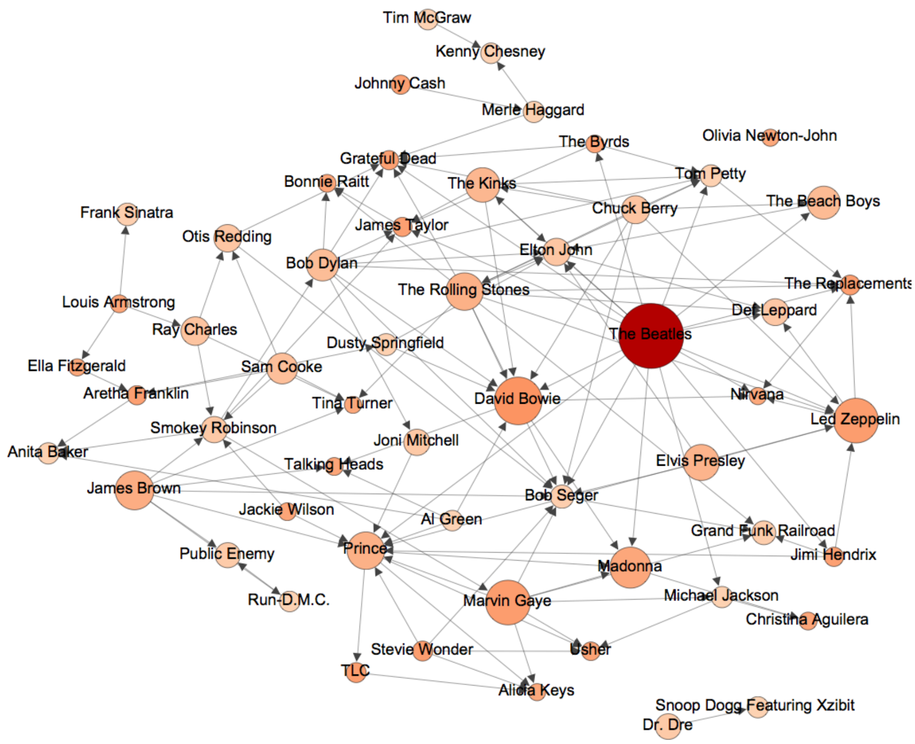

Figure 1: The influence graph, filtered to only include the artists with an out-degree of 15 or higher. Node size and color intensity are scaled with node degree.

100 between 1960 and 2010, along with other metadata such as songs' release years. We filtered out songs by artists for which allmusic.com has no influencer and follower data, resulting in a dataset of 13800 songs. For each song, the dataset provides 8 harmonic and 8 timbral features derived from topic modeling on audio content, chord change counts, song release dates, and other song-level features.

In [12], Mauch et al., summarize their 16 audio features using the following qualitative descriptions, annotated by experts.

Harmonic features

1. changes involving dominant 7th chords
2. natural minor key changes
3. changes involving minor 7th chords
4. simple diatonic changes used in major keys
5. unrecognized changes or no chordal content
6. stepwise changes indicating modal harmony
7. ambiguous major/minor attribution
8. sustained major chords

Timbral features

1. drums, aggressive, percussive

2. calm, quiet, mellow
3. energetic, speech, bright
4. piano, orchestra, harmonic
5. guitar, loud, energetic
6. /ay/, male voice, vocal
7. /oh/, rounded, mellow
8. female voice, melodic, vocal

The above 16 features resulted from topic modeling of audio of popular songs in Mauch et al., [12]. Each song in the dataset is represented by a combination of weights for each topic. These topics are also used to measure the evolution of characteristics of popular music between 1960 and 2010, as shown in Figure 2, taken from Mauch et al. [12].

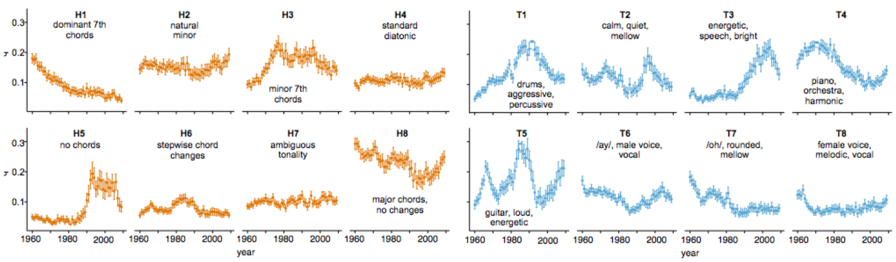

Figure 2: Evolution of musical topics in the Billboard Top 100

4. Dataset Summary Statistics

The portion of the influence graph that we obtained from allmusic.com is a directed graph consisting of 1097 weakly connected components and a giant component consisting of 2121 nodes. The graph contains 3228 nodes and 4734 edges. The average shortest undirected path length between two nodes is 4.63, and the average local clustering coefficient is 0.123.

Given the graph's low average shortest path length and high clustering coefficient (by comparison, the clustering coefficient of an Erdos-Renyi random graph with the same number of nodes and the same edge probability is 0.00091), the influence graph can be classified as a small world network.

The graph exhibits strong preferential attachment, as shown by the skewed in- and out-degree distributions of the graph (plotted in Figure 3).

5. Models

We performed link prediction on the influence graph by training a classifier with input features based on graph structure as well as song audio content.

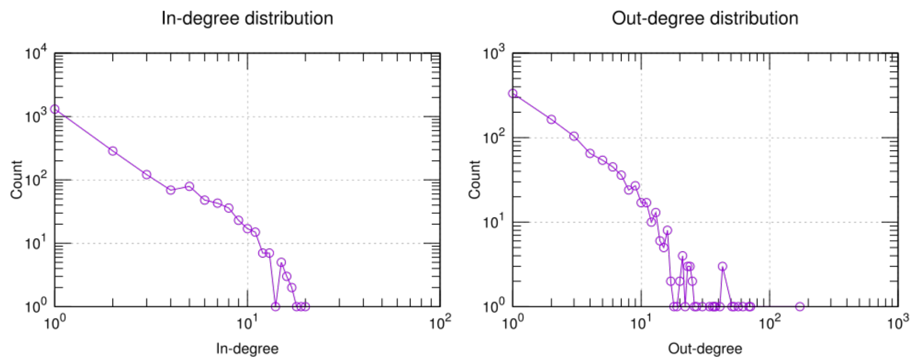

Figure 3: In- and out-degree distributions of the influence graph, plotted on a log-log scale

5.1. Pipeline and Evaluation Procedure

Given a graph G with V vertices and E edges (call the non-existing edges $\bar{E}$; note that there are $n \times (n-1) - \|E\|$ of them), we model existing edges as positive training examples and non-existing edges $\bar{E}$ as negative training examples.

We use k-fold bootstrap sampling cross validation (also called random sub-sampling cross validation) with $k = 5$ to evaluate the model, and we report Area Under the Receiver Operating Characteristic Curve (ROC AUC) as our evaluation metric on the class-balanced test set. For each iteration of cross validation, we randomly partition E into *positive train*, and *positive test* sets; similarly we sub-sample $\bar{E}$ into *negative train* and *negative test* sets so that the overall distribution is class-balanced. Subsequently, we remove all edges in G that are now in the partitioned positive test set, creating a modified graph G'.

In the training phase, we fit a classifier using the examples in the train sets. In the evaluation phase, for each example (node pair u, v) in the test sets, we compute the features of u, v based on the graph G'. Lastly, we report the ROC AUC score averaged across all runs of cross validation.

5.2. Addressing Class Imbalance

In our formulation of link prediction (with directionality), the class imbalance is severe. In our influence graph of size $n = 3228$ artists, there are $n \times (n-1)$ possible directed edges. Our dataset contains $w = 4734$ positive edges, which translates to a class imbalance ratio of $\frac{w}{n(n-1)-w} = 0.00045$ expected positive edges for every negative (i.e. non-existent) edge.

To address this class imbalance problem, we under-sample the majority class in training, testing, and validation.

5.3. Learning Algorithm

We use the Extremely Randomized Trees Classifier model, a learning algorithm that fits decision trees (with Gini impurity as the split measure) on randomized sub-samples of the dataset and averages across models to report a final classifier (commonly referred to as ensemble learning). We chose this model for two main reasons. First, the most direct application for this project is to provide useful information to musicologists that are labeling artist influence. Decision trees are highly interpretable and visualizable, which favors them for applications for musicologists who may have little machine learning expertise. Secondly, since we, the authors, do not have specific domain knowledge about music harmonics and timbre features, a decision tree model is appropriate because it allows us to avoid making assumptions about the data.

5.4. Notation

$\Gamma(u)$ the set of neighbors of node u
$k(u)$ the degree of node u
(u, v) an directed edge between node u and node v
$\{u, v\}$ an undirected edge between node u and node v
$u \to v$ a directed path between node u and node v
$u \sim v$ an undirected path between node u and node v
$dist(u, v)$ the shortest path length between node u and node v
$songs(u)$ the set of all of artist node u's songs in our dataset
$years(u)$ the set containing the year of release of all songs in $songs(u)$
$aud(s)$ the vector of 16 audio features for song s (see Section 3.2)
$avg\_aud(u)$ the vector of 16 audio features taken as the mean of $aud(s)$ for all $s \in songs(u)$

5.5. Baseline features based on graph structure

Our baseline features are based on existing techniques described in Section 2.1. However, instead of using the approach of scoring each edge and outputting the top-k as the authors did in 2.1, we use the scoring functions as features to a learning algorithm. Formally, the scoring functions evaluated are:

Common Neighbors: $|\Gamma(u) \cap \Gamma(v)|$
Jaccard Coefficient: $\frac{|\Gamma(u) \cap \Gamma(v)|}{|\Gamma(u) \cup \Gamma(v)|}$ if $|\Gamma(u) \cup \Gamma(v)| \geq 1$, and 0 otherwise
Adamic/Adar [1]: $\sum_z \frac{1}{log(k(z))}$ where $z \in \Gamma(u) \cap \Gamma(v)$
Resource Allocation [3] : $\sum_z \frac{1}{k(z)}$ where $z \in \Gamma(u) \cap \Gamma(v)$
Sørensen Index [10]: $\frac{|\Gamma(u) \cup \Gamma(v)|}{|k(u)+k(v)|}$
Leicht-Holme-Newman [9]: $\frac{|\Gamma(u) \cup \Gamma(v)|}{k(u) \times k(v)}$
Preferential Attachment [5]: $|\Gamma(u)||\Gamma(v)|$

5.6. Features based on graph metadata

We use song audio data to define the feature Joint Audio and the feature set Disjoint Audio, which both measure musical similarity between artists along the 16 audio features described in Section 3.2.

Joint Audio: $|avg\_aud(u)(u) - avg\_aud(u)(v)|$, where $|x|$ denotes the l^2 norm of vector x
Disjoint Audio: $avg\_aud(u)(u) - avg\_aud(u)(v)$, where each element of the resulting vector is a separate feature

All the features based on graph structure treat the graph as undirected. That is, for each feature function f evaluated on edge (u, v), $f(u, v) = f(v, u)$. Since the influence graph is directed, we need some feature to determine the direction of an edge. Since a music influence relationship implies chronological precedence, we define the following simple time-based feature to capture the extent of precedence:

Year Difference: $\frac{\sum_{y_v} y_v}{|years(v)|} - \frac{\sum_{y_u} y_u}{|years(u)|}$ for $y_v \in years(v)$ and $y_u \in years(u)$

5.7. Difficulties and Weaknesses in our Methodology

Our work encountered three main difficulties. Firstly, we rely on the influence graph data from allmusic.com as our ground truth. However, this dataset is imperfect and subject to the biases and constraints facing the people who defined its influence relationships. This difficulty in part justifies why a good link prediction model could be useful to musicologists – namely, that it could propose new edges to add to an incomplete graph.

Secondly, the definition of musical influence encoded in the influence graph is not entirely formalized, and it could mean different things in different contexts. In our work, we wish to study influence related to musical content, but influence relationships may exist for other reasons – for example, one could say that David Bowie influenced Queen not just with his music, but also with his brand of fashion and stage performance. Our link prediction model cannot capture such complex and various influence relationships since it is limited by its input features.

Thirdly, we face the technical problem of joining our two main datasets, the audio features dataset and the influence graph. The difficulty is in using song-level features to predict artist-level relationships. Here, we resolve this difficulty by representing artists' music as the mean of the audio features in each of their songs. However, this scheme is imperfect because it ignores that artists may channel different influences and musical styles in different songs.

6. Results and analysis

6.1. Results

In Table 1, we present the ROC AUC generated when training and evaluating our classifier using different features.

Next, we evaluated using combinations of features. In Table 2, we report the ROC AUC of our model using Year Difference combined with all pairs of the other features.

Scoring Function	ROC AUC
Random Guessing	0.490
Common Neighbors	0.543
Jaccard Coefficient	0.542
Adamic/Adar	0.536
Preferential Attachment	0.700
Resource Allocation	0.544
Sørensen Index	0.541
Leicht-Holme-Newman	0.543
Joint Audio	0.597
Disjoint Audio	0.789
Year Difference	0.740

Table 1: The ROC AUC metric when the model is run with 5-fold cross-validation on each feature.

	Random Guessing	Common Neighbors	Jaccard Coefficient	Adamic / Adar	Preferential Attachment	Resource Allocation	Sørensen Index	Leicht-Holme-Newman	Joint Audio	Disjoint Audio
Random Guessing	0.717	0.734	0.730	0.748	0.800	0.740	0.731	0.734	0.763	0.852
Common Neighbors		0.760	0.760	0.758	0.771	0.751	0.746	0.767	0.774	0.853
Jaccard Coefficient			0.751	0.759	0.760	0.754	0.753	0.754	0.780	0.857
Adamic / Adar				0.750	0.777	0.759	0.757	0.749	0.777	0.860
Preferential Attachment					0.771	0.777	0.763	0.774	0.811	0.867
Resource Allocation						0.750	0.751	0.755	0.782	0.856
Sørensen Index							0.759	0.757	0.777	0.866
Leicht-Holme-Newman								0.749	0.770	0.857
Joint Audio									0.766	0.855
Disjoint Audio										0.859

Table 2: Results using different feature sets. Each cell value represents the ROC AUC metric when the model is run with 5-fold cross-validation, using the pair of features given by the cell's row and column along with theYear Difference feature. Cell values along the table's diagonal were generated by running the model with only feature given by the cell's row and the Year Difference feature.

As Table 2 shows, the highest-scoring feature combination tested consisted of Preferential Attachment, Disjoint Audio, and Year Difference, and achieves ROC AUC = 0.867. Of the features based on graph structure, Preferential Attachment performed the best by far; given that the influence graph exhibits strong preferential attachment (as shown in Figure 3), it is unsurprising that favoring edges between nodes with high degree performs well.

These results demonstrate that audio features have discriminative power in predicting musical influence relationships. As reported in Table 1, Disjoint Audio was the feature with highest score.

6.2. Analysis of the model

Using the highest-scoring feature combination tested (Preferential Attachment, Disjoint Audio, and Year Difference), we analyze the precision at top k as well as the top features as ranked by importance in order to better understand the model's performance.

Figure 4 shows a plot of precision at top k. For some k, the plot shows the precision of the k examples predicted positively with the highest probability. The shape of the plot trends downward with k, confirming that examples predicted positively with high probability are more likely to be true positive examples.

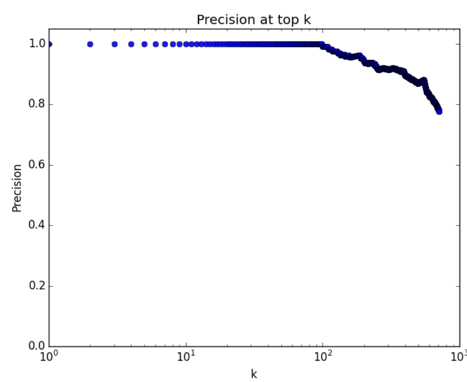

Figure 4: Precision at top k using the features Preferential Attachment, Disjoint Audio, and Year Difference

Below, we report the top 5 features ranked by feature importance as determined by discriminative power in our Extremely Randomized Trees Classifier model.

1. Year Difference
2. Preferential Attachment
3. Timbral feature 3 (energetic, speech, bright)
4. Timbral feature 5 (guitar, loud, energetic)

5. Harmonic feature 5 (unrecognized changes or no chordal content)

Of the 16 audio features included in the Disjoint Audio feature set, the most important features appear to also have been among the most volatile over time, as depicted in Figure 2. Mauch et al., [12] contextualize these features in music history: timbral feature 3 and harmonic feature 5 are associated with hip-hop and related genres, which only reached the mainstream in the 1990s; and timbral feature 5 peaked with the rises of rock and heavy metal music in the mid-1960s and mid-1980s, respectively. Perhaps these "volatile" audio features are most highly predictive because they provide a more meaningful proxy for genre than other features, such as harmonic feature 2 (natural minor key changes), which appears not to be uniquely associated with any particular genre.

6.3. Error analysis

For error analysis, we examine selected false positives misclassifications in order to better understand the model's strengths and weaknesses.

Below are three of the top false positives – that is, the false positive influence relationships predicted with the highest confidence.

1. George Jones → Tina Turner
2. The Beatnuts → Hurricane Chris
3. Marv Johnson → The Marvelettes

The first prediction does not seem plausible, since George Jones, a country musician, and Tina Turner, a pop and R&B artist, represented entirely different musical traditions. However, the pair (George Jones, Tina Turner) scores highly for Year Difference (feature value = 26.0) as well as for Preferential Attachment (feature value = 162), the model's two most important features. While these two features have discriminative power in classifying positive and negative examples, they do not capture any nuance in artists' musical content.

The second and third predictions give pairs of artists who do not score highly for Year Difference (feature values = 8.0 and 5.0, respectively) or Preferential Attachment (feature values = 2.0 and 3.0, respectively); however, the artists in both pairs share the same genre: hip-hop and Motown, respectively. In these cases, a musical influence relationship seems more plausible; however, without musicological expertise, this is difficult to verify.

7. Conclusion

We presented an analysis of musical influence based on audio features of songs and graph structure features. From our results, we see that preferential attachment and several audio features are good indicators of musical influence, which suggests that musical evolution may be studied through a data-driven approach. Further work can be done in investigating the propagation of the significant musical features along the influence graph to map the development of musical creativity.

[1] Lada Adamic, Eytan Adar, *Friends and Neighbors on the Web*. Social Networks, 25(3):211230, July 2003.

[2] Jon Kleinberg, David Liben-Nowell *The Link Prediction Problem for Social Networks*. Deparment of CS Cornell University, Jan, 2004.

[3] T. Zhou, L. Lu, Y.-C. Zhang. *Predicting missing links via local information.* Eur. Phys. J. B 71 (2009) 623.

[4] J. Newman. *Clustering and preferential attachment in growing networks.* Physical Review Letters E, 64(025102), 2001.

[5] L. Barabasi, H. Jeong, Z. Neda, E. Ravasz, A. Schubert, and T. Vicsek. *Evolution of the social network of scientific collaboration.* Physica A, 311(34):590614, 2002.

[6] Sergey Brin and Lawrence Page. *The anatomy of a large-scale hypertextual Web search engine.* Computer Networks and ISDN Systems, 30(17):107117, 1998.

[7] S. Soundarajan, John Hopcroft. *Using Community Information to Improve the Precision of Link Prediction Methods*. Proceedings of the 21st International Conference on World Wide Web. 2012.

[8] L. Barros, M. Finger et al., *Link Prediction in Complex Networks Based on Cluster Information.* Advances in Artificial Intelligence SBIA 2012.

[9] E. A. Leicht, Petter Holme, and M. E. J. Newman. *Vertex Similarity in Networks*, Department of Physics, University of Michigan

[10] T. Sorensen. *A method of establishing groups of equal amplitude in plant sociology based on similarity of species content.* 1948.

[11] Uri Shalit, Daphna Weinshall, Gal Chechik. *Modeling Musical Influence with Topic Models.* Journal of Machine Learning Research. 2013.

[12] Matthias Mauch, Robert M. MacCallum, Mark Levy, Armand M. Leroi. *The evolution of popular music: USA 19602010.* R. Soc. open sci. 2015.

[13] Nicholas Bryan, Ge Wang. *Musical Influence Network Analysis and Rank of Sample Based Music.* Proceedings of the 12th International Society for Music Information Retrieval Conference. 2011.