The Interconnectivity of Music: A Cluster Analysis

Ted Jesus Chua, Sophie Kaplan,
Jonathan Valyou, Mai Phuong Pham Huynh

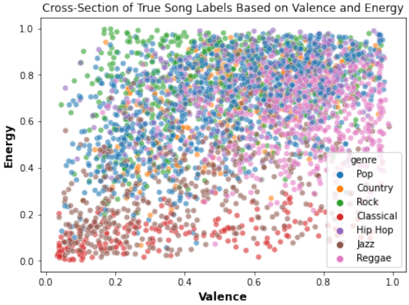
Emory University
MATH 347: Nonlinear Optimization
Spring 2021

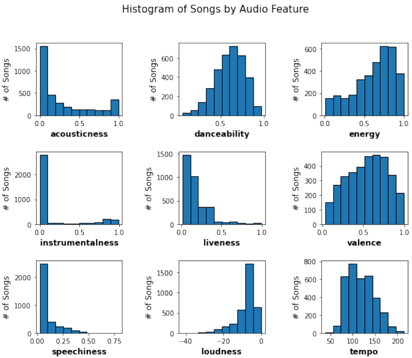
Introduction

- **Motivation:** Music is not independent across genres.
  - Spotify's *audio features* are proprietary metrics that classify relevant attributes of a song.
  - Audio features can potentially capture how music genres are linked.
- Clustering is the process of partitioning a data set into groups of data points with similar characteristics.
- **Objective:** Explore the inter-connectivity of music and the impact of clustering on a listener's discovery of new musical interests.

[1] Spotify API

Spotify Data Set





The data consists of nine audio features for 3455 songs across seven genres.

---

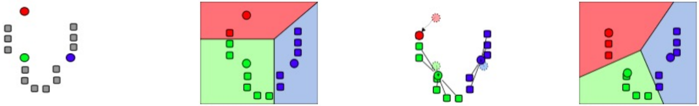
[1] Spotify API

K-means Clustering

- Description:
  Given a set of observations $(x_1, x_2, \cdots, x_n)$, where each observation is a d-dimensional real vector, k-means clustering aims to partition the n observations into k clusters with $k \leq n$ and $\mathbf{S} = \{S_1, S_2, \cdots, S_k\}$ so as to minimize the within-cluster sum of squares distance

$$\underset{S}{argmin} \sum_{i=1}^{k} \sum_{x_j \in S_i} ||x_j - u_i||^2$$

- Visualization for $k = 3$



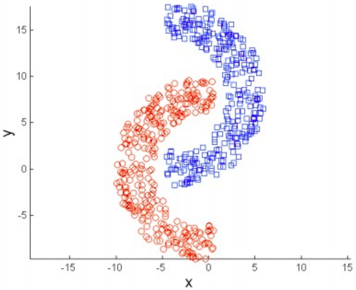
[2] University of Texas - "Spectral Clustering"

K-Means Clustering - Advantages and Disadvantages

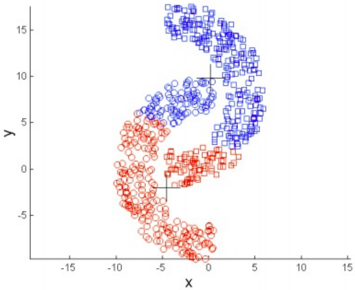
Advantages	Disadvantages
Simple to implement	Choose k manually
Scale to large data sets	Dependent on initial values
Guarantee convergence	Vary clusters' sizes and density
Flexible to new examples	Clustering outliers
Generalized to imbalanced clusters	Scales with dimensions

[3] Google Developers - "k-Means Advantages and Disadvantages"

K-Means Clustering - Data Manifold



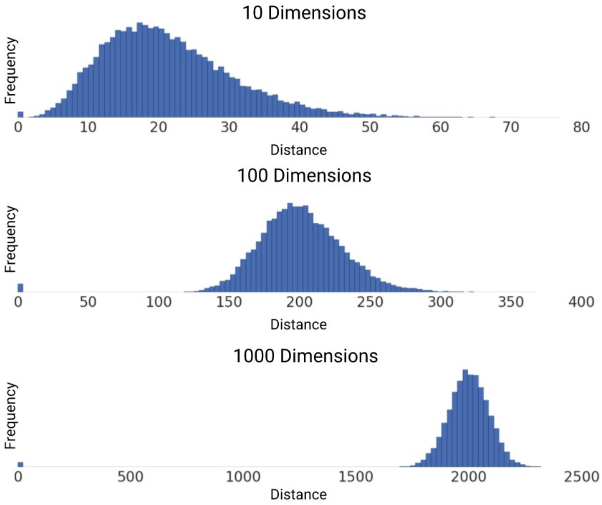
True Labels



K-Means Clustering

[2] University of Texas - "Spectral Clustering"

K-Means Clustering - Curse of Dimensionality

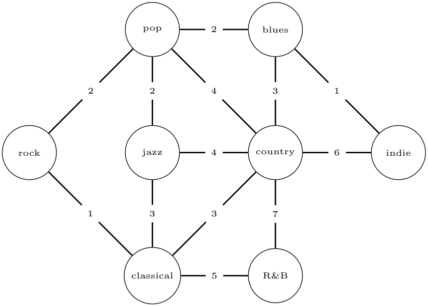


[3] Google Developers - "k-Means Advantages and Disadvantages"

What Are Graphs?

Let $G(\gamma, \theta)$ denote an undirected graph:

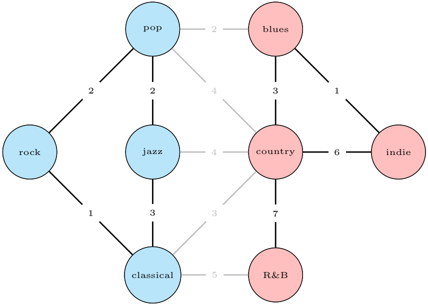
- $\gamma = \{\gamma_i | i = 1, \cdots, n\}$ are the vertices that represent data points
- $\theta = \{(\gamma_i, \gamma_j)\}$ are pairs of data points that are similar in some way with weight $w_{ij} \geq 0$



[4] Ulrike von Luxburg (2007). "A Tutorial on Spectral Clustering"

How Can Graphs Be Clustered?

Suppose we want to cluster our songs into two groups V and W. Let $cut(V, W)$ be the sum of the weights of the edges that connect the two clusters. **The smaller the graph cut, the more dissimilar are the resulting clusters.**



$$\min \, cut(V, W)\left(\frac{1}{|V|} + \frac{1}{|W|}\right) = \min \, \mathrm{RatioCut}(V, W)$$

[4] Ulrike von Luxburg (2007). "A Tutorial on Spectral Clustering"

An NP-Hard Problem

The Complexity of min RatioCut$(C_1, C_2, \ldots, C_k)$

As the number of vertices and desired clusters grow, the more difficult it becomes to find a solution to the optimization problem.

An Alternative Approach to Minimizing the Ratio Cut

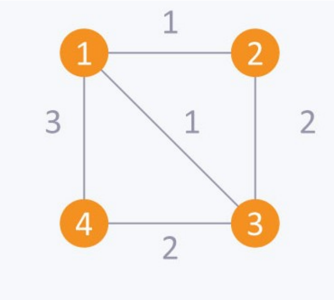
Spectral clustering circumvents the NP-hard problem by reformulating the ratio cut using the Laplacian—a matrix representation of a graph—and solving a relaxed optimization problem.

Derivation of the Laplacian

Let S denote the similarity matrix of a complete graph $G(\gamma, \theta)$. Let D be a diagonal matrix where the diagonals d_i are the degrees of γ_i.

$$S_{ij} = w_{ij}, (\gamma_i, \gamma_j) \in \theta \qquad D_{ii} = d_i = \sum_{j=1}^{n} w_{ij}$$

The Laplacian matrix L is given by $L = D - S$. Example:



$$L = \begin{pmatrix} 5 & 0 & 0 & 0 \\ 0 & 3 & 0 & 0 \\ 0 & 0 & 5 & 0 \\ 0 & 0 & 0 & 5 \end{pmatrix} - \begin{pmatrix} 0 & 1 & 1 & 3 \\ 1 & 0 & 2 & 0 \\ 1 & 2 & 0 & 2 \\ 3 & 0 & 2 & 0 \end{pmatrix}$$

Properties of the n by n Laplacian Matrix

- $L = L^T$
- L is diagonally dominant and the diagonal entries are bounded below by 0, so it is **positive semi-definite**. Therefore, **all the eigenvalues of L are non-negative.**
- 0 is always an eigenvalue of L. **The algebraic multiplicity of** 0 **is the number of connected components in $G(\gamma, \theta)$.**

[4] Ulrike von Luxburg (2007). "A Tutorial on Spectral Clustering"

A Reformulation of the Optimization Problem

- We arrive at a relaxed minimization problem:

$$\min_{f \in \mathbb{R}^n} f^\top L f \text{ subject to } e^\top f = 0, \ \|f\|^2 = n$$

- Solution: $f^\top L f = \lambda$, the smallest nonzero eigenvalue of L.
- How to Solve: Rayleigh Quotient OR Lagrange Multipliers/KKT Conditions.

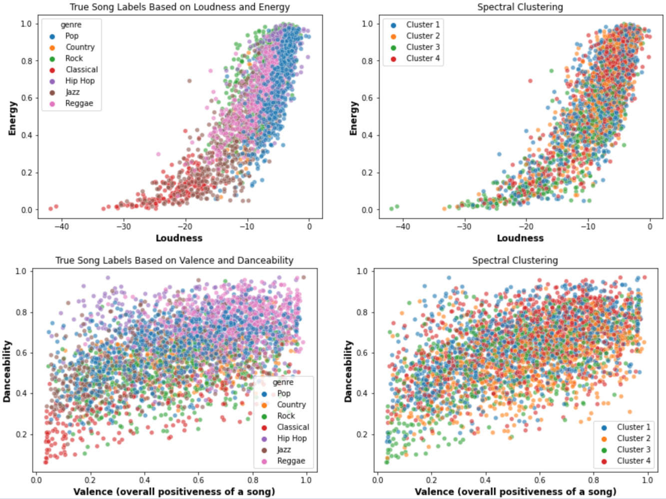
[4] Ulrike von Luxburg (2007). "A Tutorial on Spectral Clustering"

Spectral Clustering Algorithm

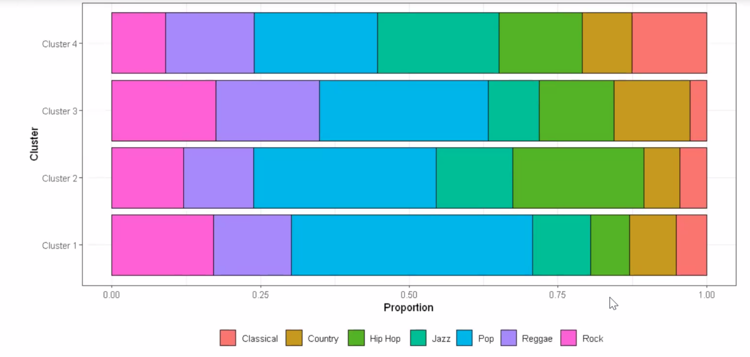
1. Specify the undirected graph $G(\gamma, \theta)$, where $|\gamma| = n$ and $w_{ij} \geq 0 \forall (\gamma_i, \gamma_j) \in \theta$.
2. Form the n by n Laplacian matrix L.
3. Let k be the desired number of clusters. Compute the k eigenvectors corresponding to the k smallest nonzero eigenvalues of L and store it in an n by k matrix K.
4. The cluster assignments of $\{\gamma_1, \ldots, \gamma_n\}$ are obtained from clustering the n rows of K using k-means clustering.

Song Clusters

All nine audio features were used to form clusters. Out of 36 possible 2D cross-sections, two are shown below:



Proportion of Musical Genres in Each Cluster



	acousticness	loudness	speechiness	tempo
cluster 1	0.27	-7.99	0.07	121.61
cluster 2	0.26	-7.53	0.11	99.57
cluster 3	0.23	-7.25	0.12	160.85
cluster 4	0.41	-10.16	0.11	79.68

Implications and Next Steps

- Music is subjective with many different interpretations
- Apply our algorithm to a music recommendation engine
- Form clusters using other optimization problems

References

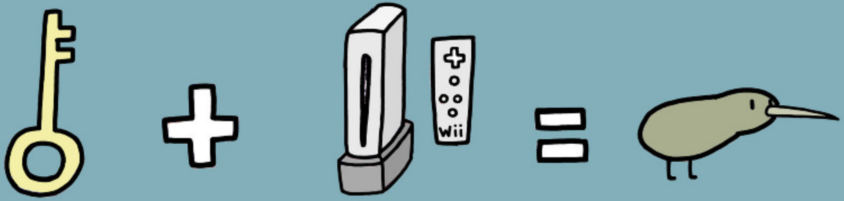
Spotify API
https://developer.spotify.com/documentation/web-api/

University of Texas
"Spectral Clustering"
http://ranger.uta.edu/~heng/CSE6363_slides/spectralclustering.pdf

Google Developers
"k-Means Advantages and Disadvantages"
https://developers.google.com/machine-learning/clustering/algorithm/advantages-disadvantages

Ulrike von Luxburg (2007)
"A Tutorial on Spectral Clustering"
Statistics and Computing Journal 17(4), 1 – 13.
https://people.csail.mit.edu/dsontag/courses/ml14/notes/Luxburg07_tutorial_spectral_clustering.pdf

THANK YOU!



Math. It explains everything.

KKT Conditions- Stationarity/Complementary Slackness

$$L(f, \mu_1, \mu_2) = f^\top L f + \mu_1 e^\top f + \mu_2(\|f\|^2 - n)$$

$$\nabla L(f, \mu_1, \mu_2) = \begin{pmatrix} Lf + \mu_1 e + 2\mu_2 f \\ e^\top f \\ \|f\|^2 - n \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}$$

Solving we will find:

$$\mu_1 n = 0 \Rightarrow \mu_1 = 0$$

Simplifying our stationarity condition, we can show that f is an eigenvector of L and it is any of $n-1$ eigenvectors ($n-1$ stationary points are present) given that we are constrained to have e as an eigenvector:

$$Lf = -2\mu_2 f$$

Since f is an eigenvector,

$$f^\top L f = \lambda$$