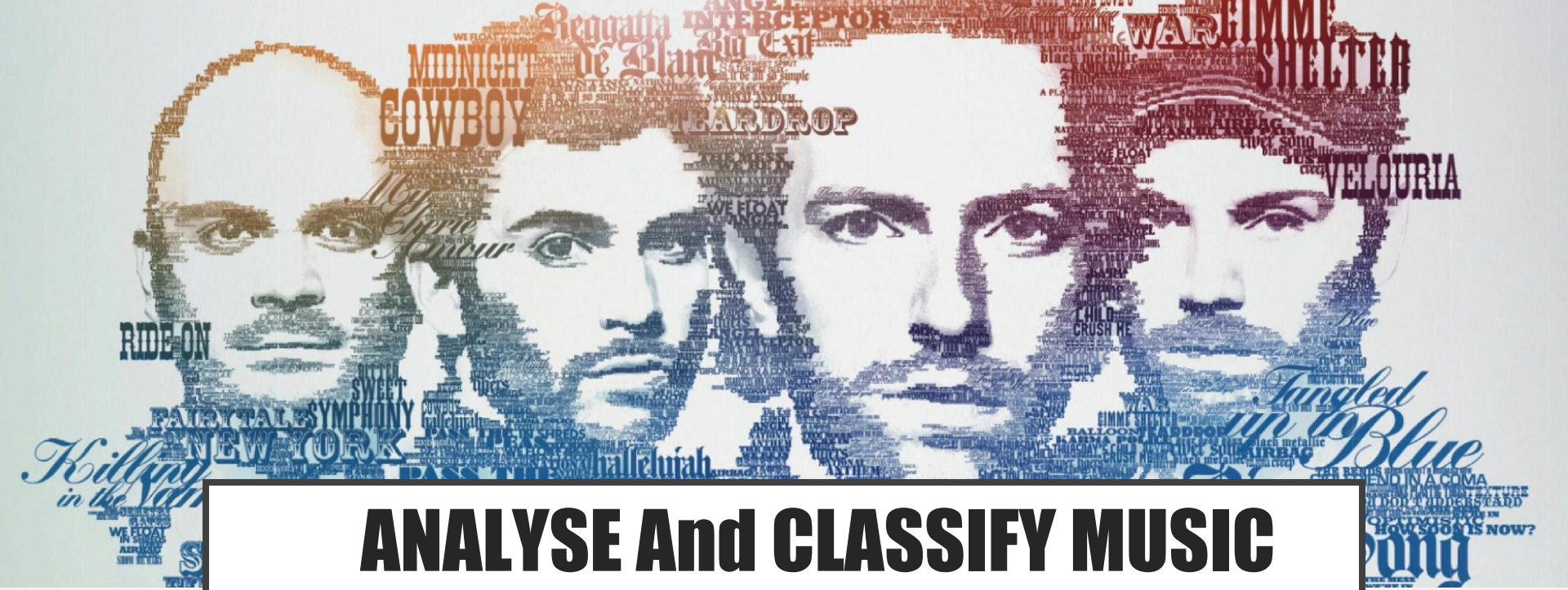


ANALYSE And CLASSIFY MUSIC BY LYRICS

Group 3: Danlei Qian, Siqi Jiang, Xi Zhang

The George Washington University

AGENDA

- Problem Statement
- Literature Review
- Motivation - Project Scope
- Methodology
  - Dataset
  - Preprocess
  - Modeling
- Results & Analysis
- Conclusion
- Limitation
- Further Research

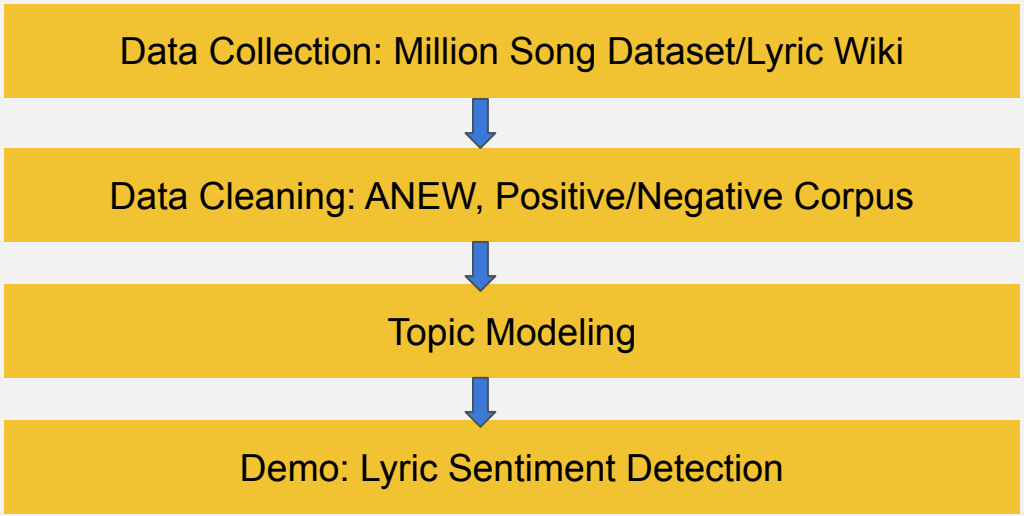
PROBLEM STATEMENT

- People find music by different preferences
- Lyrics of music can convey emotions
- Classify the music by lyrics can target different preferences

Literature Review

- **Natural Language Processing for Music Knowledge Discovery**
- Applying Natural Language Processing approaches to harness the potential of the text collections for automatic music knowledge discovery
- https://arxiv.org/abs/1807.02200
- **Natural Language Processing for Music Information Retrieval**
- Providing the audience with potential applications of Natural Language Processing (NLP) to MIR and Computational Musicology.
- https://www.upf.edu/web/mdm-dtic/tutorial-natural-language-processing-for-music-information-retrieval
- **Moody Lyrics: A Sentiment Annotated Lyrics Dataset**
- Provides a research direction for how to annotate the sentiment of lyrics.
- **Several researches about applying NLP models on sentiment analysis.**

MOTIVATION & SCOPE



Data Collection

- Data Resource:

  Million song dataset: http://millionsongdataset.com/pages/getting-dataset/

  Randomly Selected 1000 songs from it.

file	artist	title	lyrics
TRAZSEY128F424C12B.h5	Ivan Parker	Close To The Well (Live)	
TRASTIS128F92FA998.h5	Eri Esittäjiä	Dawn	
TRAFKDG12903CAB85C.h5	Webb Wilder	Too Cool for Love	

Lyric Wiki: https://lyrics.fandom.com/wiki/LyricWiki

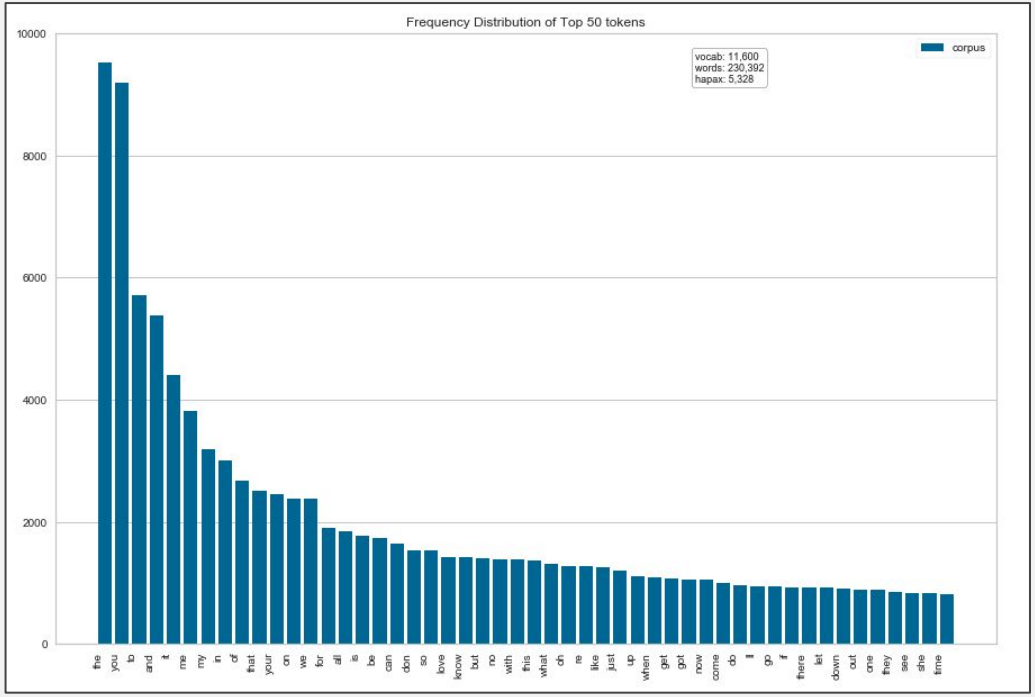
Write script to download lyrics and append to pandas dataframe.

file	artist	title	lyrics
TRAAAAW128F429D538.h5	Casual	I Didn't Mean To	Verse One:\n\nAlright I might\nHave had a litt...
TRAAAEF128F4273421.h5	Adam Ant	Something Girls	Adam Ant/Marco Pirroni\nEvery girl is a someth...
TRAAAFD128F92F423A.h5	Gob	Face the Ashes	I've just erased it's been a while, I've got a...
TRAABJV128F1460C49.h5	Lionel Richie	Tonight Will Be Alright	Little darling \nWhere you've been so long \nI...
TRAABLR128F423B7E3.h5	Blue Rodeo	Floating	Lead Vocal by Greg\n\nWell, these late night c...
...	...	...	...

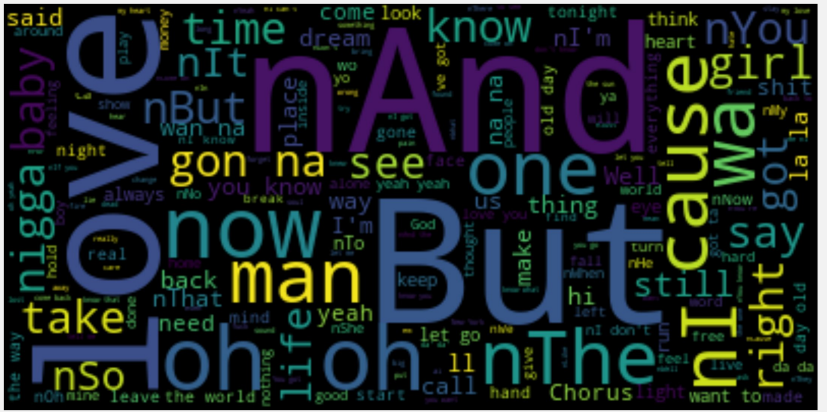
Data Description

```
Alright I might
Have had a little glare when I stared at ya ho
But I didn't know she was like that
She stared right back
My niggas warnin me that she was comin on to me
I react like a mack do I act cool
Just to test her cause I ain't no jester
I suggest her and her friend be outtie
Cause I don't want to make my pals get rowdy
And doubt me our friendship
But when lips touch
I go crazy in the clutch
Sorta like schitzo I forgets my bros and pals over gals
I didn't mean to but when you fiend you do
Strange things for the denim no matter who's in em
Grab a flooze then I'm in traffic
Don't laugh it might be your girl that I'm talkin about
I didn't mean to
```

Data Description

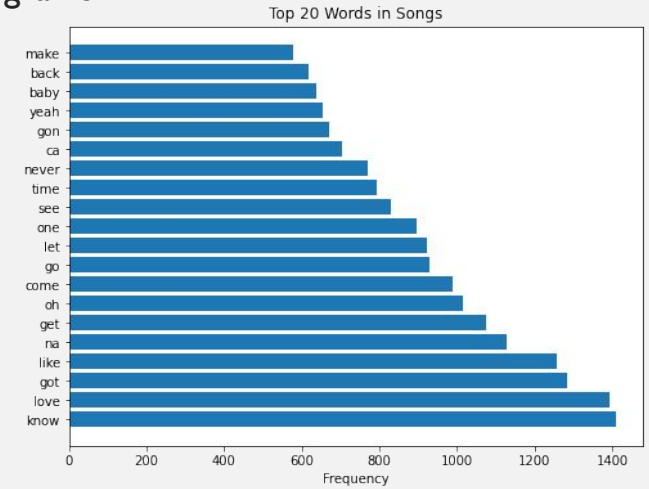


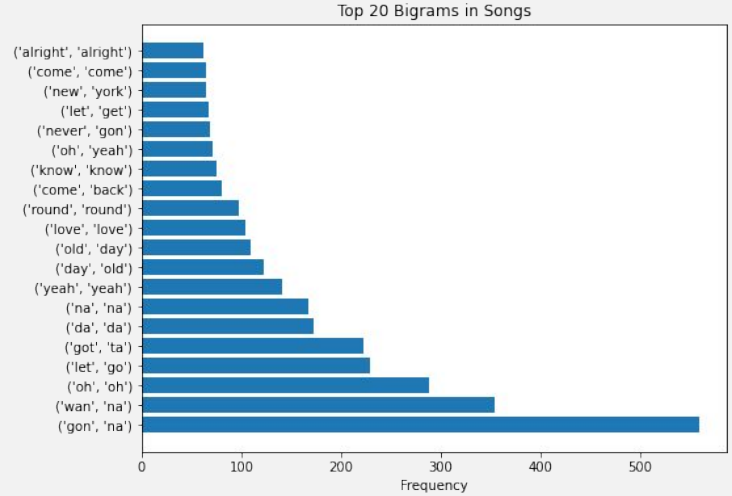
Articles, Pronouns, Conjunctions, musical terms(Verse, Chorus).

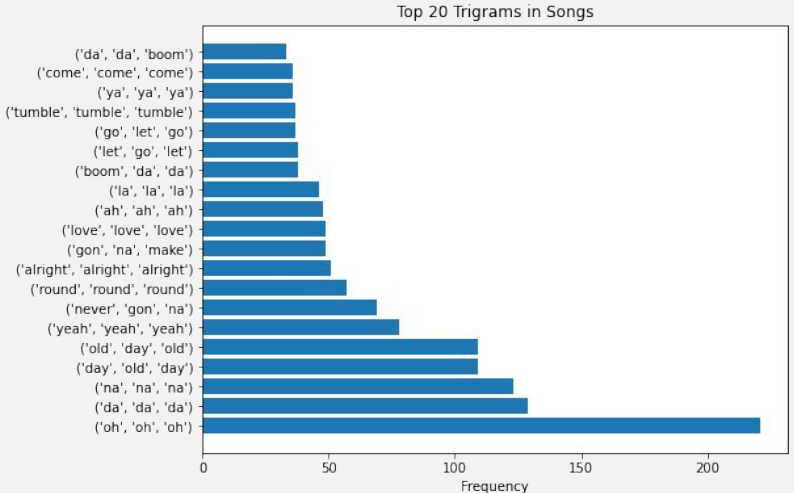


EDA

N-grams







Raw Data:
Too many words without moods.

Sentiment Lexicons

Filter useful data

Affective Norms for English Words(ANEW)

provides a set of normative emotional ratings for a large number of words in the English language.

Reference:
https://csea.phhp.ufl.edu/Media.html

anew

Description	Word No.	Valence Mean	Valence SD	Arousal Mean	Arousal SD	Dominance Mean	Dominance SD	Word Frequency
grin	773	7.4	1.87	5.27	2.64	6	1.86	13
honest	210	7.7	1.43	5.32	1.92	6.24	2.13	47
gripe	774	3.14	1.56	5	2.19	4.67	1.79	.
honey	792	6.73	1.7	4.51	2.25	5.44	1.47	25
guillotine	196	2.48	2.11	6.56	2.54	4.64	2.63	.
honor	211	7.66	1.24	5.9	1.83	6.7	2.04	66
guilty	197	2.63	1.98	6.04	2.76	3.09	2.22	29
hooker	793	3.34	2.31	4.93	2.82	4.73	2.48	.
gun	593	3.47	2.48	7.02	1.84	3.53	2.72	118

Positive and Negative

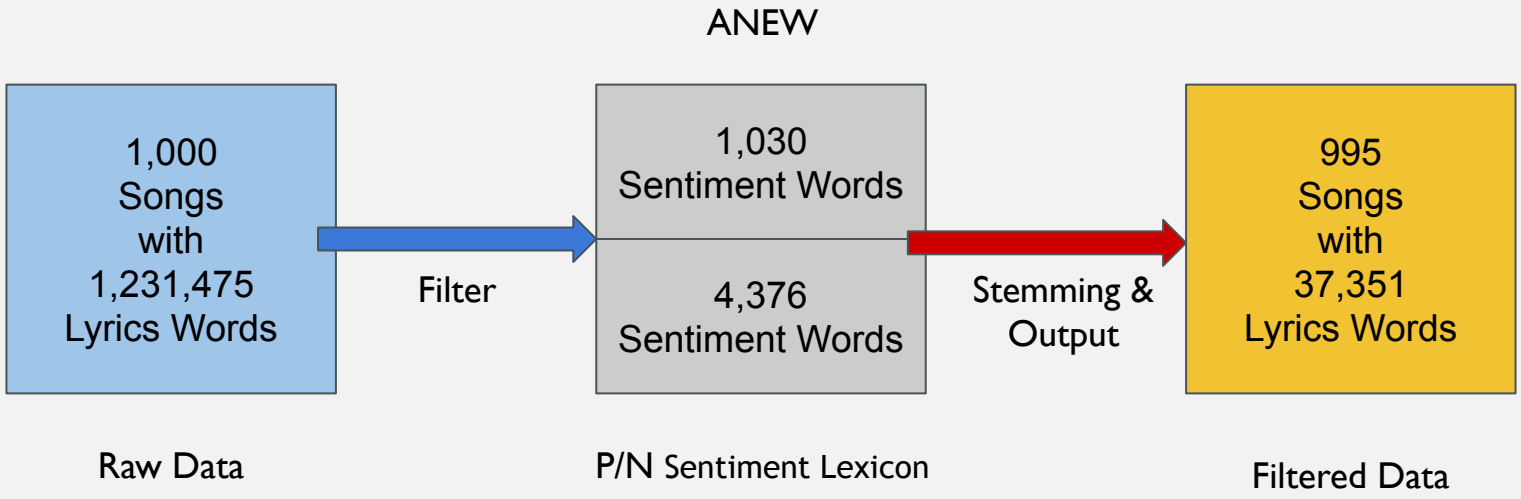
generated sentiment lexicon

Reference:
https://sites.google.com/site/datasciencesslab/projects/multilingualsentiment

Positive: well love like won right best top beautiful support important good popular famous peace

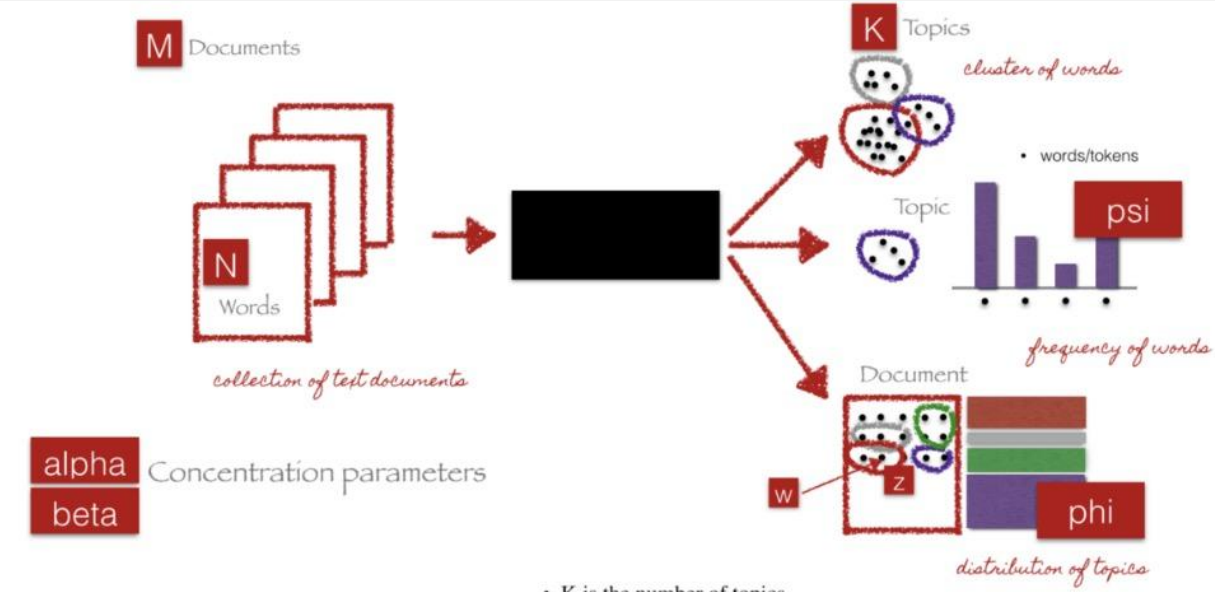
Negative: dead lost attack killed problems issues conflict lies failed hard difficult damage break suffered

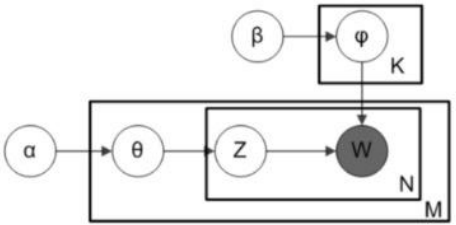
Data Cleaning and Preprocessing - Filter useful data



Modeling

LDA -Theoretical Overview





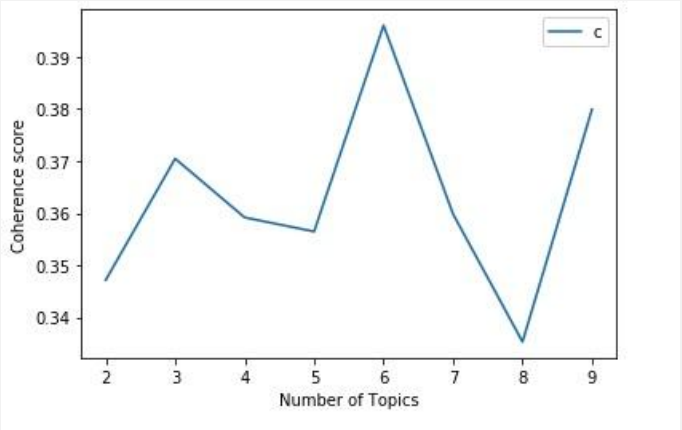
- K is the number of topics
- N is the number of words in the document
- M is the number of documents to analyse
- α is the Dirichlet-prior concentration parameter of the per-document topic distribution
- β is the same parameter of the per-topic word distribution
- φ(k) is the word distribution for topic k
- θ(i) is the topic distribution for document i
- z(i,j) is the topic assignment for w(i,j)
- w(i,j) is the j-th word in the i-th document
- φ and θ are Dirichlet distributions, z and w are multinomials.

http://chdoig.github.io/pytexas2015-topic-modeling/#/3/4

Modeling

LDA -Finding the Optimal Number of Topics

Topic Coherence: Determining Optimal number of Topics



3 Topics:

```
(0, '0.281*"love" + 0.055*"babi" + 0.037*"girl" + 0.024*"like" + 0.022*"heart" + 0.020*"lone"')
(1, '0.042*"fire" + 0.036*"god" + 0.025*"burn" + 0.025*"war" + 0.023*"like" + 0.022*"fight"')
(2, '0.058*"home" + 0.038*"christma" + 0.036*"time" + 0.024*"like" + 0.021*"sun" + 0.020*"blue"')
```

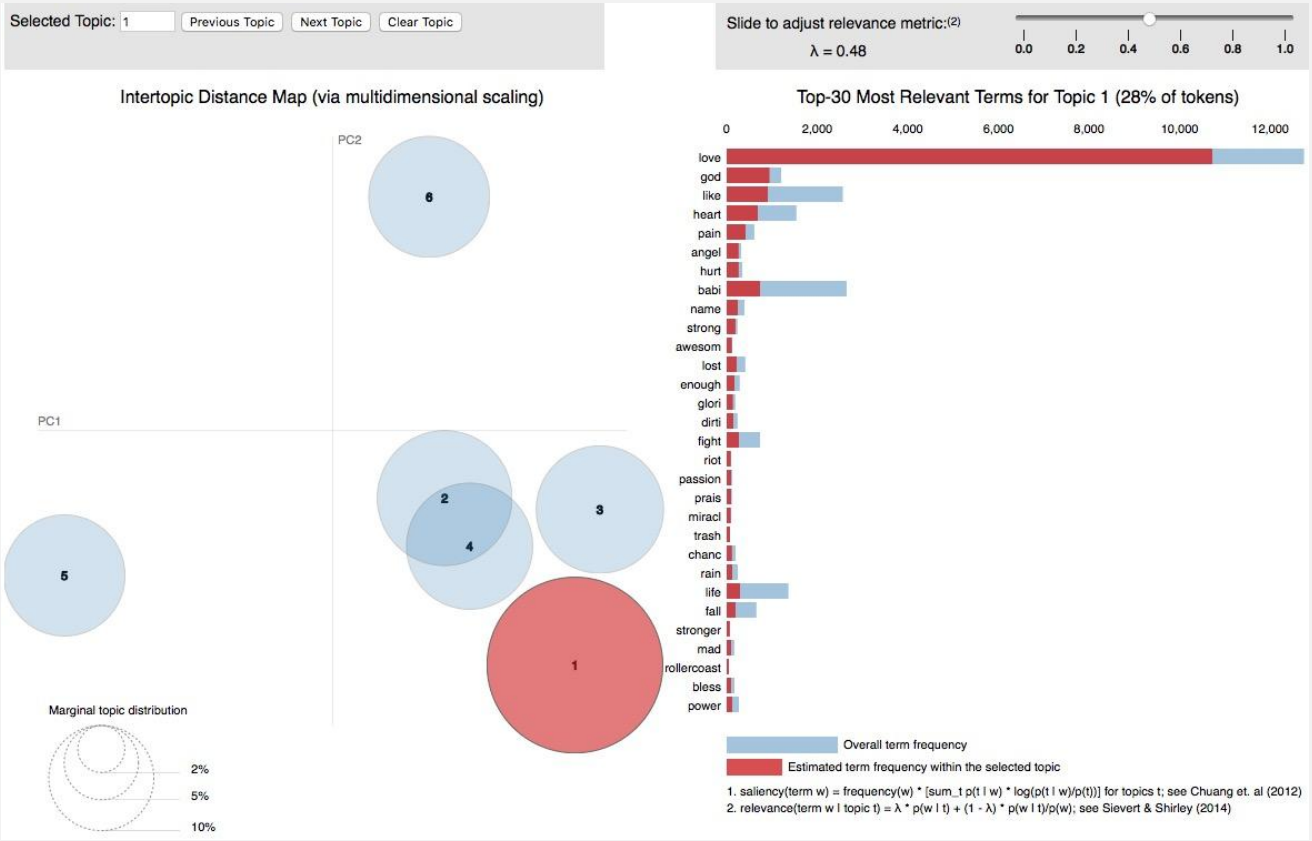
6 Topics:

```
(0, '0.057*"war" + 0.051*"burn" + 0.035*"sun" + 0.027*"sin" + 0.027*"fight" + 0.025*"easi"')
(1, '0.360*"love" + 0.032*"god" + 0.031*"like" + 0.025*"babi" + 0.023*"heart" + 0.014*"pain"')
(2, '0.060*"babi" + 0.057*"fire" + 0.049*"time" + 0.030*"happi" + 0.027*"heart" + 0.026*"love"')
(3, '0.075*"girl" + 0.061*"christma" + 0.056*"love" + 0.053*"lone" + 0.032*"merri" + 0.027*"babi"')
(4, '0.097*"home" + 0.047*"joy" + 0.043*"good" + 0.023*"babi" + 0.018*"time" + 0.017*"love"')
(5, '0.044*"like" + 0.042*"life" + 0.025*"lie" + 0.019*"love" + 0.017*"sick" + 0.017*"slow"')
```

8 Topics:

```
(0, '0.557*"love" + 0.042*"heart" + 0.019*"kiss" + 0.017*"like" + 0.016*"life" + 0.013*"time"')
(1, '0.240*"babi" + 0.084*"like" + 0.031*"parti" + 0.031*"rock" + 0.031*"honey" + 0.019*"gun"')
(2, '0.122*"fire" + 0.064*"time" + 0.052*"fight" + 0.043*"fall" + 0.036*"easi" + 0.025*"sick"')
(3, '0.102*"god" + 0.069*"burn" + 0.057*"sun" + 0.053*"good" + 0.050*"life" + 0.037*"like"')
(4, '0.107*"home" + 0.097*"girl" + 0.071*"christma" + 0.047*"joy" + 0.039*"blue" + 0.036*"happi"')
(5, '0.092*"war" + 0.047*"sin" + 0.033*"name" + 0.028*"dead" + 0.028*"death" + 0.027*"con"')
(6, '0.051*"lone" + 0.033*"pain" + 0.029*"lie" + 0.027*"hate" + 0.023*"like" + 0.021*"lost"')
(7, '0.052*"world" + 0.051*"peopl" + 0.038*"slow" + 0.033*"shake" + 0.030*"bird" + 0.029*"sky"')
```

Modeling
LDA



Results & Analysis

TOPICS

0, Angry

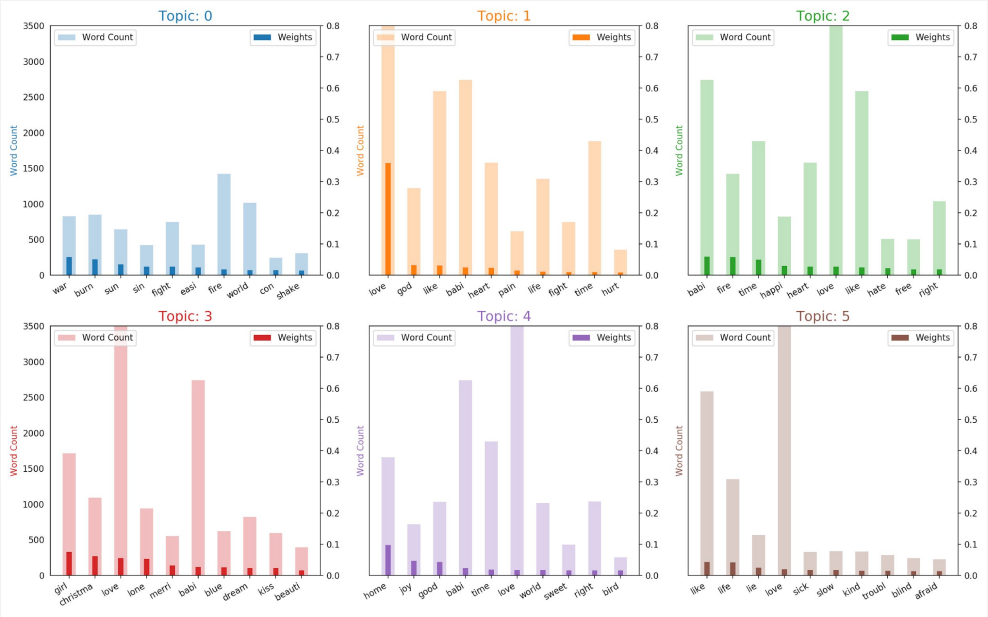
1, Love, Sad

2, Excited

3, Love, Happy

4,Peace

5, Love,Sad



Demo

Eminem —— 《One Shot, two shot》

I'm tryna slip through the exit and get to where my car is at
Bitches screamin' everywhere and niggas is wildin'
Two minutes ago we was all jokin' and smilin'
This chick is clingin' onto me sobbin' and sighin'
Sayin' she didn't mean to diss me earlier and she cryin'
But its real and cats is gettin' killed
So I hugged her and used her body as a human shield
And she got hit now she yellin'
(Don't leave me!)
I told her I'd be right back and the dumb bitch believed me
I squeezed through the back door and made my escape
I ran and got my 38 I hope its not to late
One shot two shot three shot four shots
……
God damn there just went another damn bullet I'm hit
My vest is barely able to handle it, its to thin
If I get hit again I can't do it, I scoop deep
……
I ain't come in this bitch to party I came in this bitch to fight
Although I can't stay here to fight 'cause I'm poppin' niggas tonight
That's right bitches I'm drunk with revenge
Shot a bouncer in the neck for tryna check when I get in
Swift told me to meet him here so its clear that this fucker
Shoot out the back of his truck goes up in this motherfucker
……

```
Eminem | One shot, two shot - Violence, Anger
Sentiment Words
{'like', 'bar', 'right', 'life', 'odd', 'music', 'hand', 'money', 'mess', 'damn', 'kill', 'believ', 'truck', 'human', 'gun', 'bodi', 'thank', 'peopl', 'fight',
'face', 'drunk', 'collaps', 'vest', 'god', 'wife', 'black', 'floor', 'car', 'tire', 'plenti', 'dumb', 'use', 'miss', 'hit', 'bitch', 'hug', 'argument', 'aliv',
'clear', 'man', 'swift', 'cat', 'fun', 'hope', 'bizarr', 'sound', 'cliqu', 'fast', 'strut', 'hous', 'fuck', 'desert', 'door', 'fire', 'spray', 'shit', 'outsid'
, 'diss', 'save', 'parti', 'bullet', 'reveng'}
[(0, 0.61559546), (1, 0.024770392), (2, 0.00114846), (3, 0.001139392), (4, 0.0011462743), (5, 0.3562001)]
```

0: 62%

Demo

Bob Marley —— 《Three Little Birds》

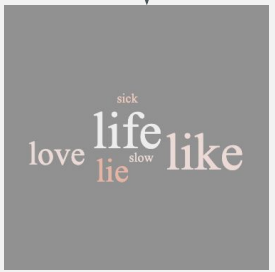
Imagine me and you, I do
I think about you day and night, it's only right
To think about the girl you love and hold her tight
So happy together
If I should call you up, invest a dime
And you say you belong to me and ease my mind
Imagine how the world could be, so very fine
So happy together
I can't see me lovin' nobody but you
For all my life
When you're with me, baby the skies'll be blue
……
So happy together
So happy together
How is the weather
So happy together
We're happy together
So happy together
Happy together
So happy together
So happy together (ba-ba-ba-ba ba-ba-ba-ba)

'''

```
Bob Marley | Three Little Birds - Love Happy
Sentiment Words
{'blue', 'happi', 'babi', 'right', 'mind', 'life', 'girl', 'world', 'eas', 'sky', 'invest', 'fine', 'imagin', 'love'}
[(0, 0.0050705294), (1, 0.0051040943), (2, 0.6410427), (3, 0.13757965), (4, 0.0050775856), (5, 0.2061254)]
```

2: 64%

3: 14%



5: 21%

Conclusion and Limitation

Conclusion

LDA can relatively accurately classify the lyrics with obvious emotional tendency.

Limitation

The emotions in songs are always complex and intertwined, and hardly defined.

Future Research

LSA (Latent semantic analysis) / LSI (Latent Semantic Indexing)

HDP (Hierarchical Dirichlet Process)

REFERENCES

- Kaggle:
- https://www.kaggle.com/mousehead/songlyrics

- Oramas, S.(2018), *Natural Language Processing for Music Knowledge Discovery,*
- https://arxiv.org/abs/1807.02200

- Oramas, S.(2018), *Natural Language Processing for Music Information Retrieval,* https://www.upf.edu/web/mdm-dtic/tutorial-natural-language-processing-for-music-information-retrieval