EZRA ZINBERG

ADVISED BY DR. CHRISTIANE D. FELLBAUM

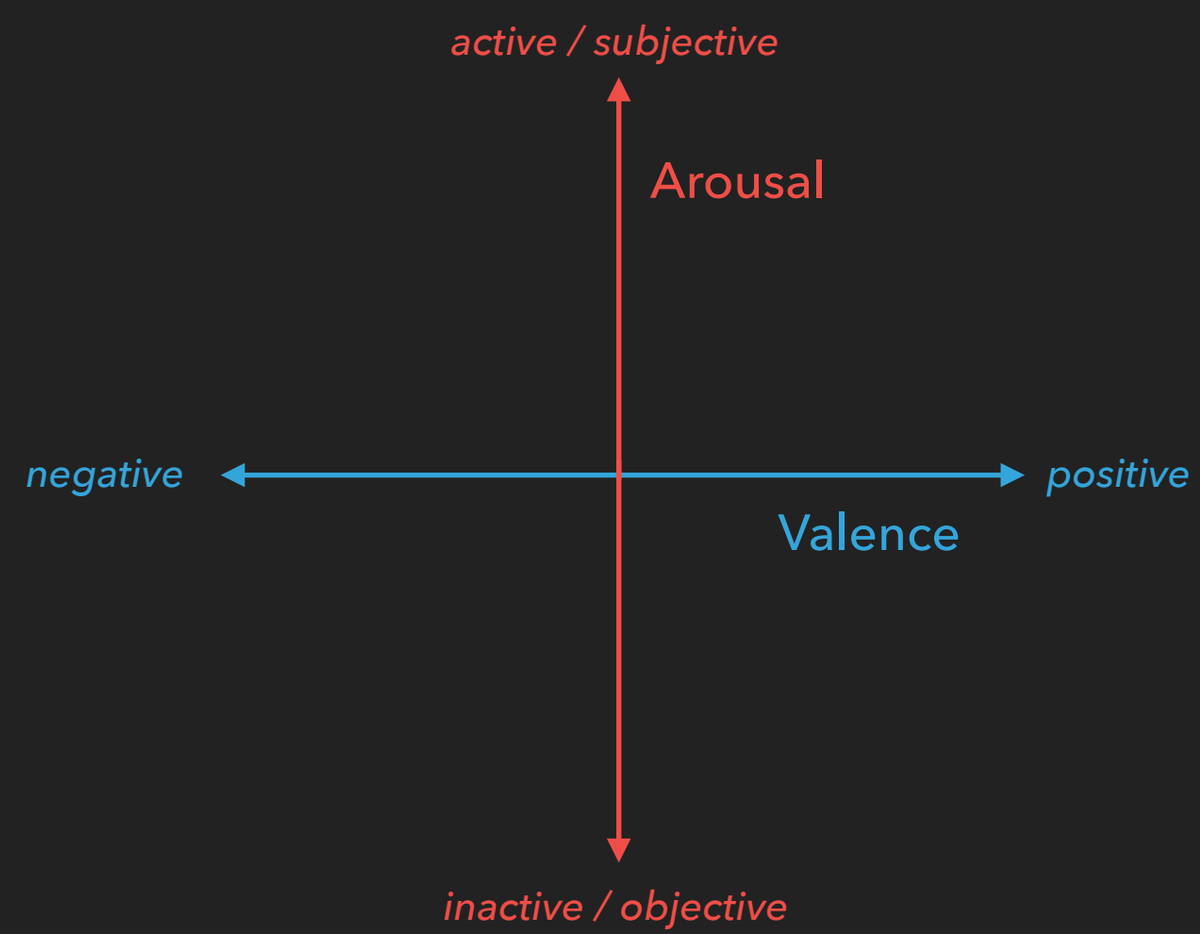
TRACING SENTIMENT IN MUSICAL CAREERS

SENTIMENT AND OPINION MINING

Sentiment can be described as

- Positive / negative
- Subjective / objective

Sentiment values are used to analyze author's opinion about a topic, or general sentiment of a corpus



Russell (1980)

SENTIWORDNET

Dictionary of english words and their sentiment scores

(pos, neg, obj)

=

(pos, neg, [1 - pos+neg])

```
>>> syns = list(swn.senti_synsets("happy"))
>>> word = syns[0]
>>> word.pos_score()
0.875
>>> word.neg_score()
0.0
>>> word.obj_score()
0.125
>>> syns = list(swn.senti_synsets("sad"))
>>> word = syns[0]
>>> word.pos_score()
0.125
>>> word.neg_score()
0.75
>>> word.obj_score()
0.125
>>> syns = list(swn.senti_synsets("table"))
>>> word = syns[0]
>>> word.pos_score()
0.0
>>> word.neg_score()
0.0
>>> word.obj_score()
1.0
```

SENTIWORDNET

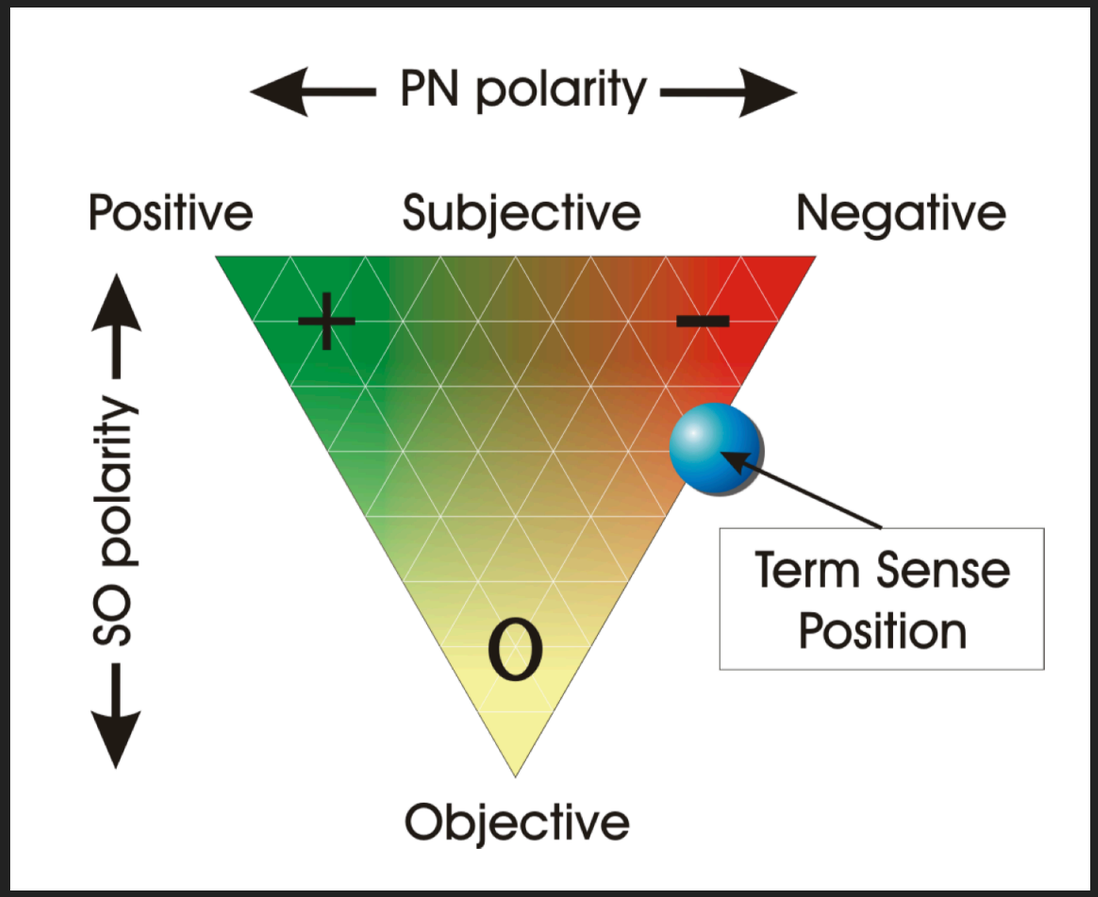
Dictionary of english words and their sentiment scores

(pos, neg, obj)

=

(pos, neg, [1 - pos+neg])

3-coordinate system can be shown on a Sentiment Triangle



Esuli & Sebastiani (2006)

"THE DISCOURSE OF POP SONGS"

- Studied top 50 songs of 1987 for factors such as:
  - Words per minute
  - Gender
  - Time
  - Place
- Early paper in lyrics analysis

Murphey (1992)

"MOODYLYRICS"

- Studied lyrics for mood-tagging
- Categorized songs into moods:
  - Happy, angry, sad, relaxed
- Conclusion: Automated analysis consistent with 'gold standard' human tagging ~75% of the time

Çano & Morisio (2017)

"THE DISCOURSE OF POP SONGS"

- Corpus of only 50 songs
- No sentiment analysis

"MOODYLYRICS"

- No analysis for individual artists

How can we use sentiment analysis tools to answer new questions about the world of music?

MOTIVATION

Use sentiment analysis tools to discover new insights in music history and psychology

Quantify new insights and existing intuitions

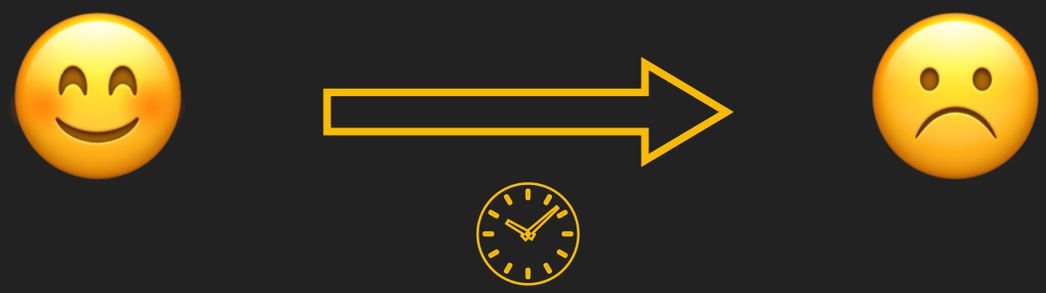
GOAL

Analyze sentiment in the lyrics of individual artists

Trace the change in sentiment for artists over time, from early work to later work

HYPOTHESIS

Sentiment will become increasingly negative over time as an artist's musical career progresses.



APPROACH

1. Acquire large sample size of lyrics over large number of artists
2. Calculate average sentiment values for each artist
    A. Overall sentiment average
    B. Sentiment in early work
    C. Sentiment in later work
3. Evaluate sentiment values to trace change in sentiment over time

EVALUATION

$$\Delta S_{total} = S_{late} - S_{early}$$

$$\Delta S_{early} = S_{avg} - S_{early}$$

$$\Delta S_{late} = S_{late} - S_{avg}$$

$\Delta S \Uparrow \longrightarrow H_1$ ✓

$\Delta S \Downarrow \longrightarrow H_1$ ✗

CORPUS DETAILS

- Lyrics from Kaggle SongLyrics corpus
- Merged with song release dates from MusicBrainz API
- Sample size:
  - ~20K lyrics
  - 485 artists

Metallica	Stone Dead Forever	8/18/08	And didn't you hear me you never listen past And didn't I see you gone to seed And the only reason is that you're too young greed
Metallica	Wherever I May Roam	6/28/06	I have stripped of all but pride So in her I do confide And she keeps me satisfied
Metallica	Whiskey In The Jar	2014-07	The Cork and Kerry Mountains I saw Captain Farrell And his money, he was countin'
Michael Bolton	A Heart Can Only Be So Strong	1997	Just to walk back a hundred times more A thousand nights, I have sworn not to stay, ooh oh I remained tangles in the chains of desire, ooh
Michael Bolton	A Love So Beautiful	1995	Our love long ago But in my heart I feel the same Old afterglow
Michael Bolton	A Time For Letting Go	2005	Isn't really what you need And the dream and all its promise Was never meant to be

PROCESS

1. Clean data
   - Tokenize
   - Remove stop words
2. Calculate individual artist data
   - Calculate average sentiment for all tracks
   - Designate “early work” and “late work” as songs released in the span of first 1/5 of career, last 1/5 of career, respectively
   - Calculate 3 representative metrics
3. Find average of metrics over all artists

```
# iterate through rows and find range of dates for artist's songs
with open('songdata-all-fields.csv') as csv_file:
    csv_reader = csv.reader(csv_file, delimiter=',')
    for row in csv_reader:
        if row[0] == artist:
            if (row[2] == ""):
                continue
            releaseDate = parser.parse(row[2])
            if releaseDate > parser.parse("2019-11-23"):
                continue
            if releaseDate < earliestDate:
                # print("new earliest date: " + str(earliestDate))
                earliestDate = releaseDate
            if releaseDate > latestDate:
                latestDate = releaseDate
                # print("new latest date: " + str(latestDate))

earlyBound = earliestDate + (latestDate-earliestDate) / 5
lateBound = latestDate - (latestDate-earliestDate) / 5

print("early period: " + str(earliestDate) + " thru " + str(earlyBound))
print("late period: " + str(lateBound) + " thru " + str(latestDate))

earlyHits = 0
lateHits = 0

with open('songdata-all-fields.csv') as csv_file:
    csv_reader = csv.reader(csv_file, delimiter=',')
    for row in csv_reader:
        if row[0] == artist:
            if (row[2] == ""):
                continue
            releaseDate = parser.parse(row[2])

            if releaseDate < earlyBound:
                earlyHits += 1
                trackSent = getSentiment.average(row[3])
                for i in range(len(trackSent)):
                    early[i] += trackSent[i]

            elif releaseDate > lateBound:
                lateHits += 1
                trackSent = getSentiment.average(row[3])
                for i in range(len(trackSent)):
```

Python code snippet

RESULTS

Metric	Value	% Change
ΔS_{total}	0.0015	3.28%
ΔS_{early}	0.0011	2.40%
ΔS_{late}	0.0004	0.87%

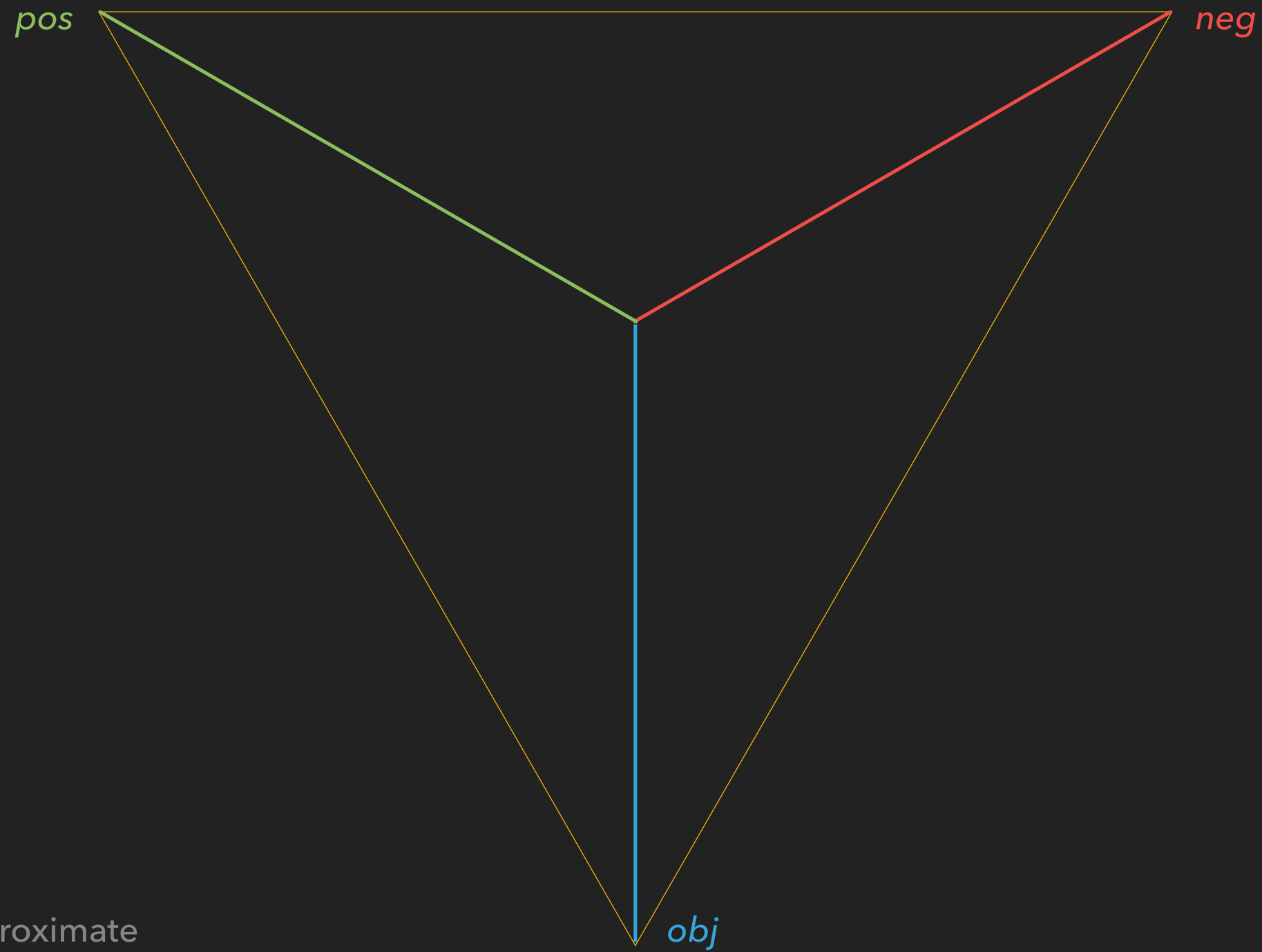
DISCUSSION

Results show weak evidence in favor of hypothesis

All sentiment metrics are positive => negative sentiment increased with time, but only slightly

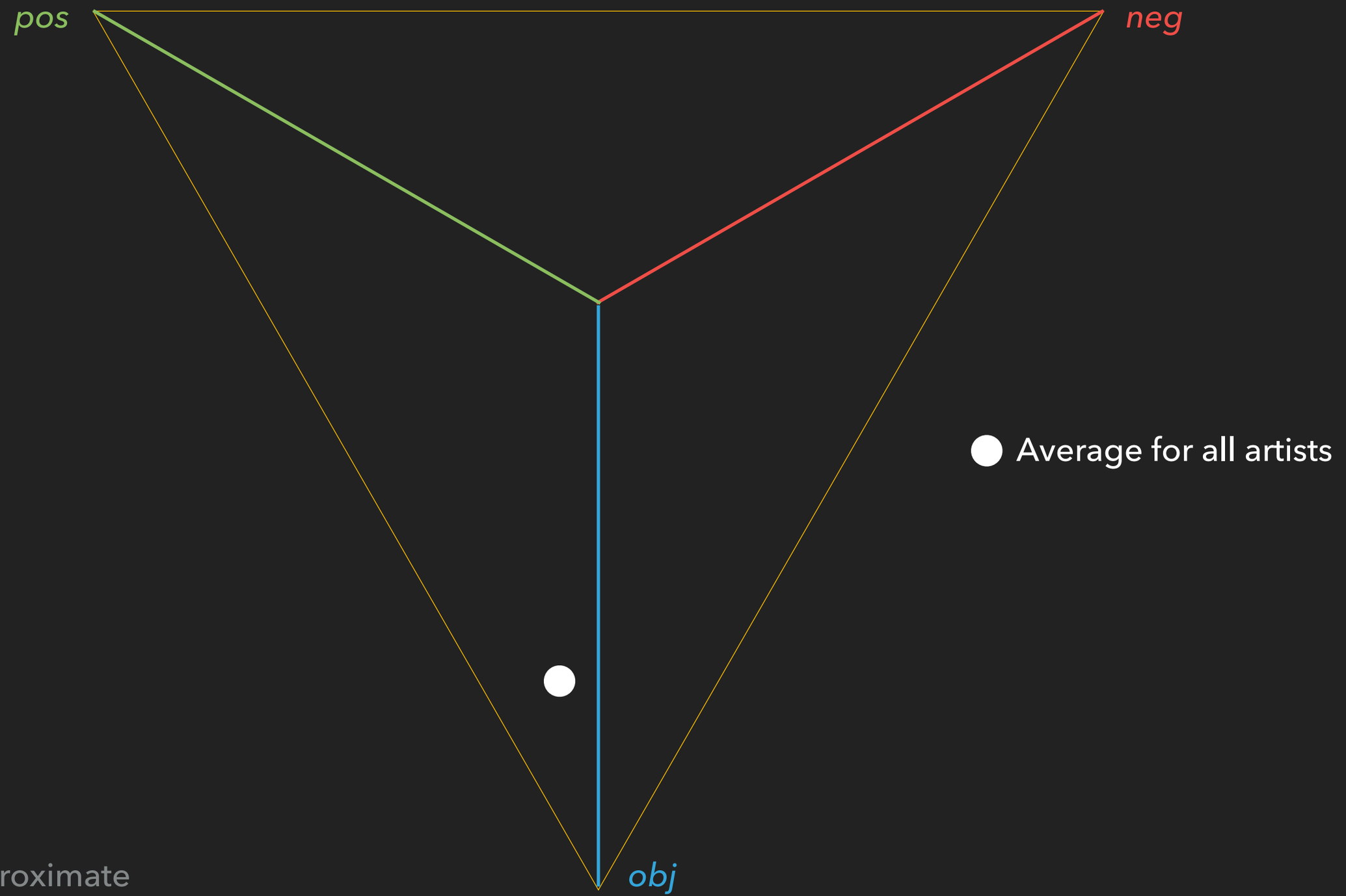
Metric	Value	% Change	Supports H_1
ΔS_{total}	0.0015	3.28%	~ ✓
ΔS_{early}	0.0011	2.40%	~ ✓
ΔS_{late}	0.0004	0.87%	~ ✓

DISCUSSION



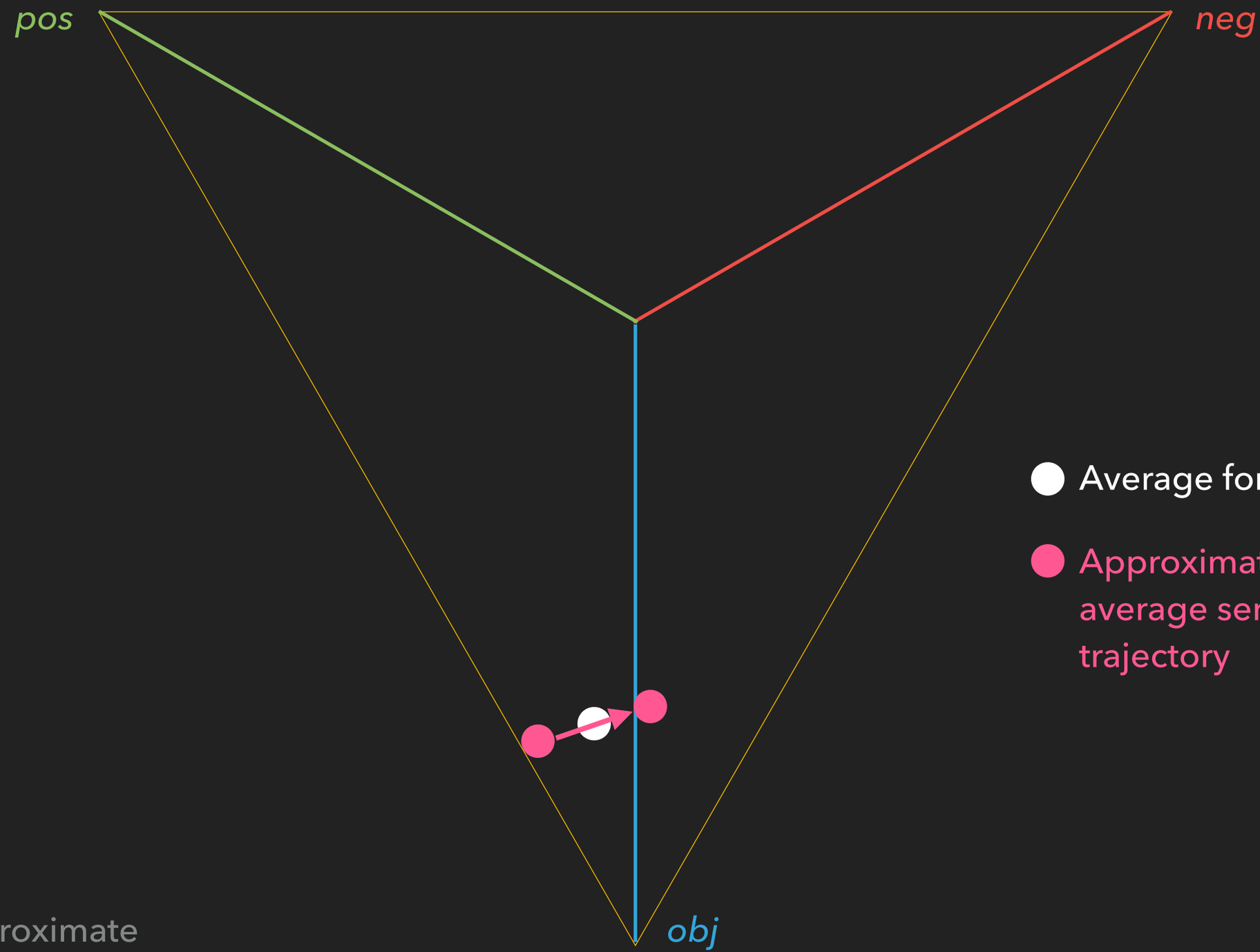
*Plot is approximate

DISCUSSION



*Plot is approximate

DISCUSSION



*Plot is approximate

CONCLUSIONS

Results are inconclusive

Negative sentiment increases on average, as hypothesized, but not significantly

Sentiment and sentiment change in music lyrics is highly varied based on artist and period of artist's career

THANK YOU!

QUESTIONS?

REFERENCES

Russell, James A. "A circumplex model of affect." Journal of personality and social psychology 39.6 (1980): 1161.

Esuli, Andrea, and Fabrizio Sebastiani. "Sentiwordnet: A publicly available lexical resource for opinion mining." LREC. Vol. 6. 2006.

Murphey, Tim. "The discourse of pop songs." Tesol Quarterly 26.4 (1992): 770-774.

Çano, Erion, and Maurizio Morisio. "Moodylyrics: A sentiment annotated lyrics dataset." Proceedings of the 2017 International Conference on Intelligent Systems, Metaheuristics & Swarm Intelligence. ACM, 2017.

Kaggle SongLyrics Database. https://www.kaggle.com/mousehead/songlyrics

MusicBrainz API Python Documentation. https://python-musicbrainzngs.readthedocs.io/en/v0.6/api/