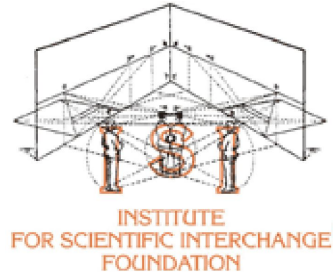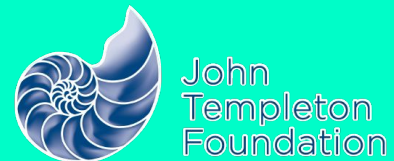# Significance and Popularity in Music Production

**Bernardo Monechi, ISI Foundation**
Kreyon 2017 – Rome

The Struggle For Success

- In many fields concerning artistic creation, there is a constant struggle for success

- Popular music artists want their songs on the top of the billboards..

- Painters want their artworks to be known among the population (or just the population interested in art..)

- Scientists want their work recognized by the scientific community.. And possibly by the rest of the population!

Artworks Value

- ..but is "success" the only value attributed to artworks?

- Maybe not:

  - Experts Appreciation..

  - Impact over other artists..

  - Impact outside the field…

  - Niche..
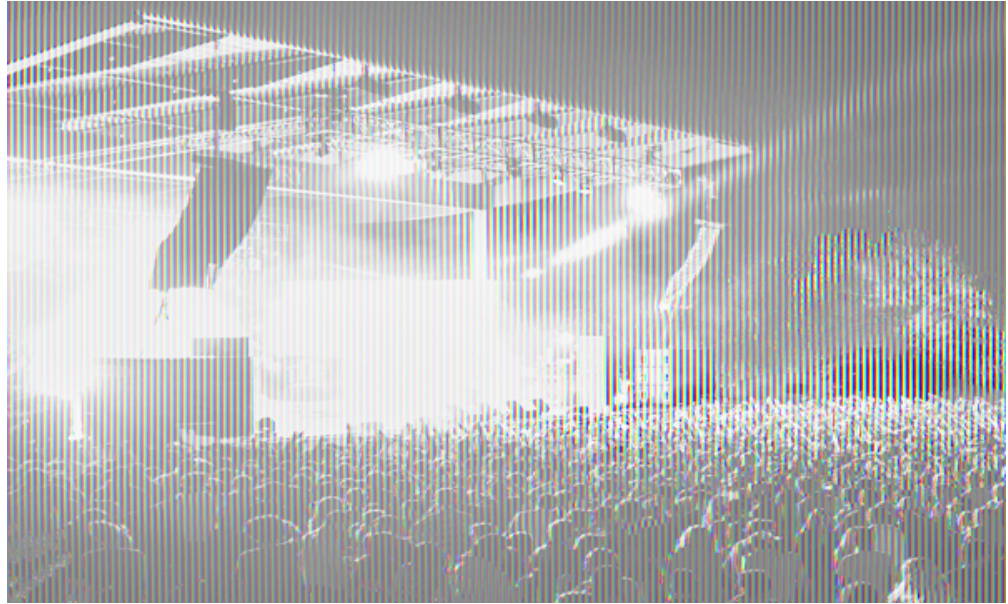
Impact over Other Artists..

- **JB Lenoir** (1929-1965)
- Wrote several Blues Standards..
- "You Shook Me":
  - Muddy Waters
  - Led Zeppelin
  - B.B. King
  - …
- Wim Wenders "The Soul of a Man" (2003)
- Blues All of Fame (2011)

"J.B had a struggle playin' unappreciated blues in vain,
Now the Blues has lost a king and I've lost a friend who died in vain"

John Mayall – The Death of JB Lenoir (1968)

Impact Outside the Field..

- **John Carpenter's The Thing** (1982)
- "Lower-than-expected" box office revenues
- Hard to compete with **Steven Spielberg's ET the Extra-Terrestrial** and **Ridley Scott's Blade Runner** released the same day
- Became a cult film
- Now Considered one of the "greatest horror movies ever made"
- According to IMDb (http://www.imdb.com/) is now referenced in other movies, cartoons, animes, tv shows, videogames..
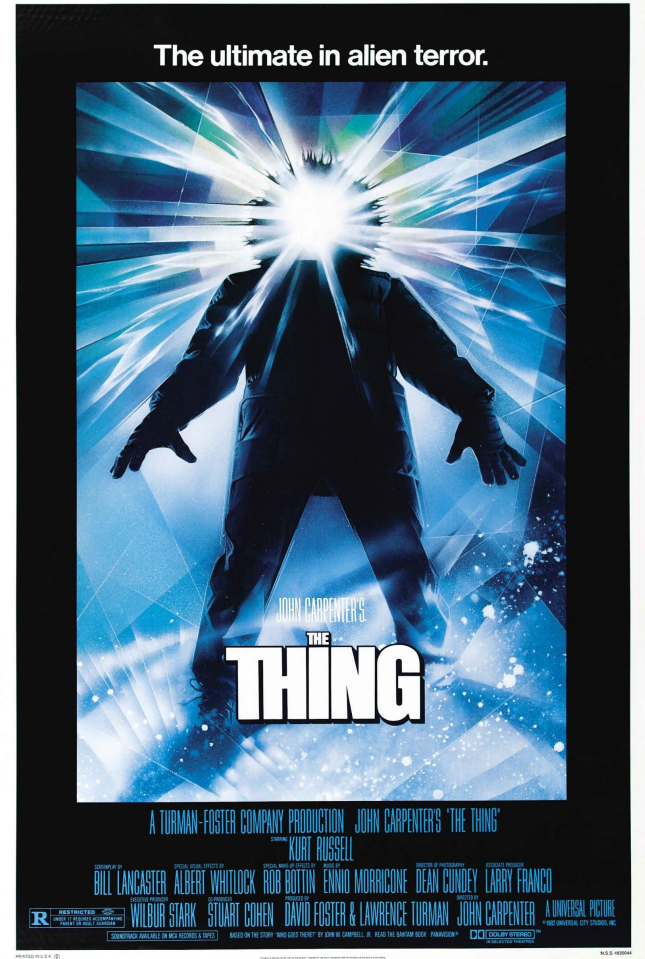
Artworks Value

- Is it possible to identify different levels of "significance" of artworks?

- If it is, which are the criteria used to choose if an album is significant?

- Is it linked with intrinsic features? With the relation between an artwork and the others? With the way it influenced the production of other artworks in the future?

Popular Music

- Popular Music is a worldwide culture: shared perspective over famous artists, music genres, etc..

- Music Recording Industry had an estimated total value of 15 billion US$ in 2014

- Lots of data: Last.fm, MusicBrainz, Discogs..

Popular Music DB: Last.fm

- Last.fm is a website where people can listen to music and share it with each other

- Stores info about **bands/artists**, **albums** and **music tracks**

- Users can **comment** each element and annotate it with **tags** in order to classify them

- Other info: **Playcount**, release date, similar elements...

Significance in Popular Music

- **Aim**: use user defined tags in order to build a growing conceptual space representing music production at a given time:

  - Relation between albums
  - Relation between albums and global music production
  - Albums contribution to the growth of the space in terms of actualized space, adjacent possible (short term impact), long-term impact
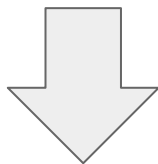
- Use it in order to define metrics capable of predicting success or significance in a broader sense

Data Selection

- Wrong Release Dates: many times not the first one.

- Integration with MusicBrainz DB.

- Over-Tagging: poorly validated tags in non-popular albums

Time Resolution: 1 year
163829 Albums
108984 Distinct Tags
From 1950 to 2015
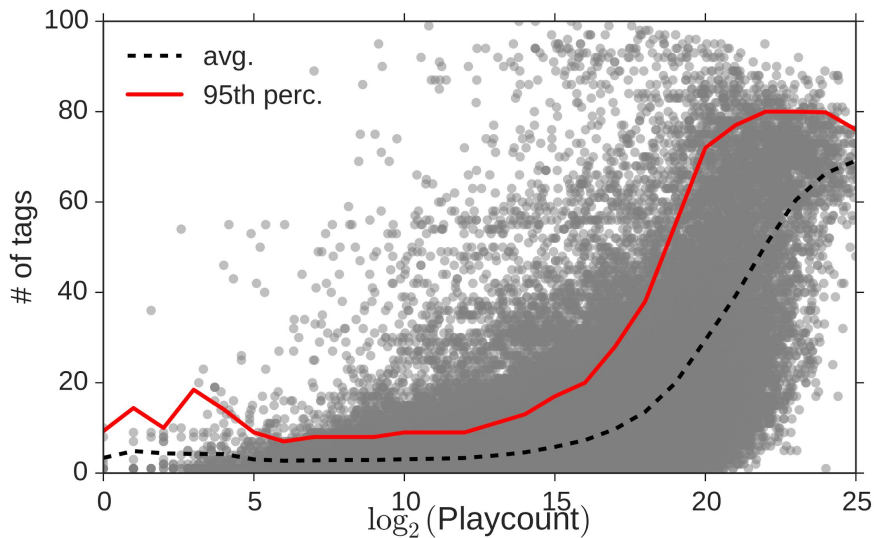
Number of album tags as a function of the album Playcount

Growing Conceptual Space

New Album

t = yyyy

t = yyyy + 1
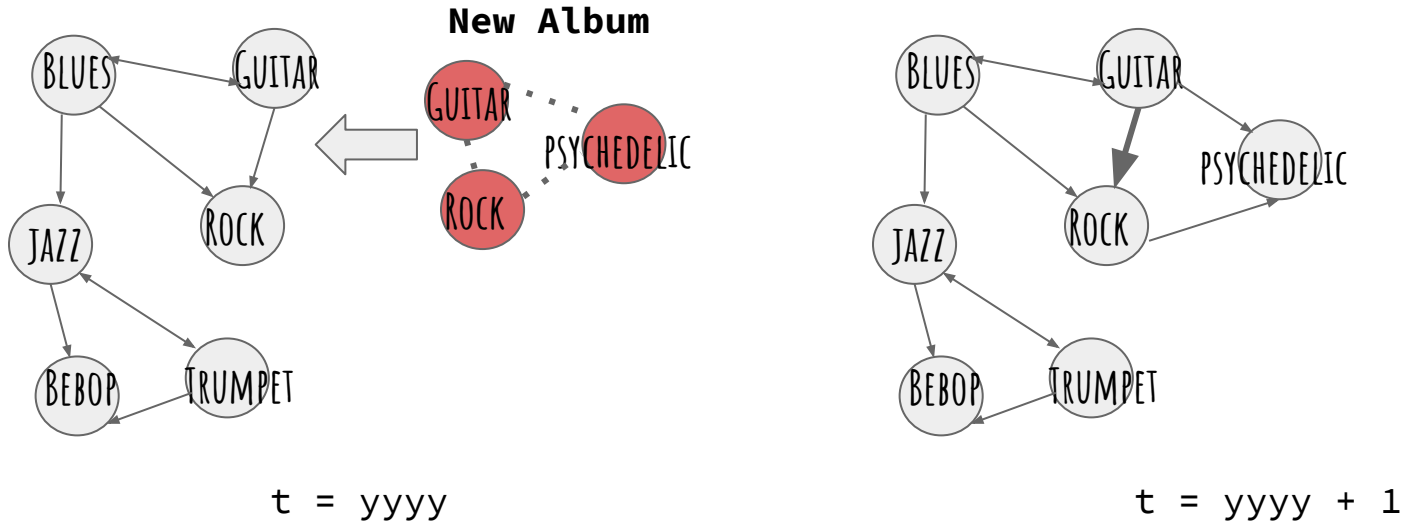
- All the albums released in a certain year are attached to the network
- Add new nodes, Add new links, Reinforce existing links
- Allow new tags to be part of the adjacent possible

Topical Representation

t = 2015
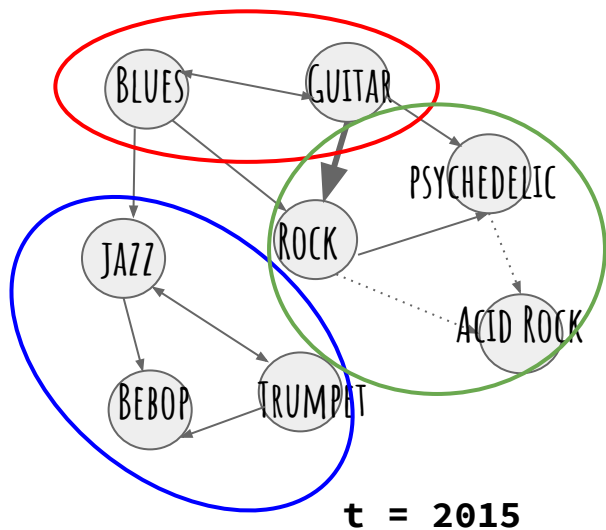
Topical Representation: apply community detection to the final network, each tag will be labelled according to the community it belongs to.

Classification of the tags according to the nowadays point of view.

"R'n'R" and "Rock and Roll" tags are equivalent.

$l \in L_a$: album set of topics

Topical Representation
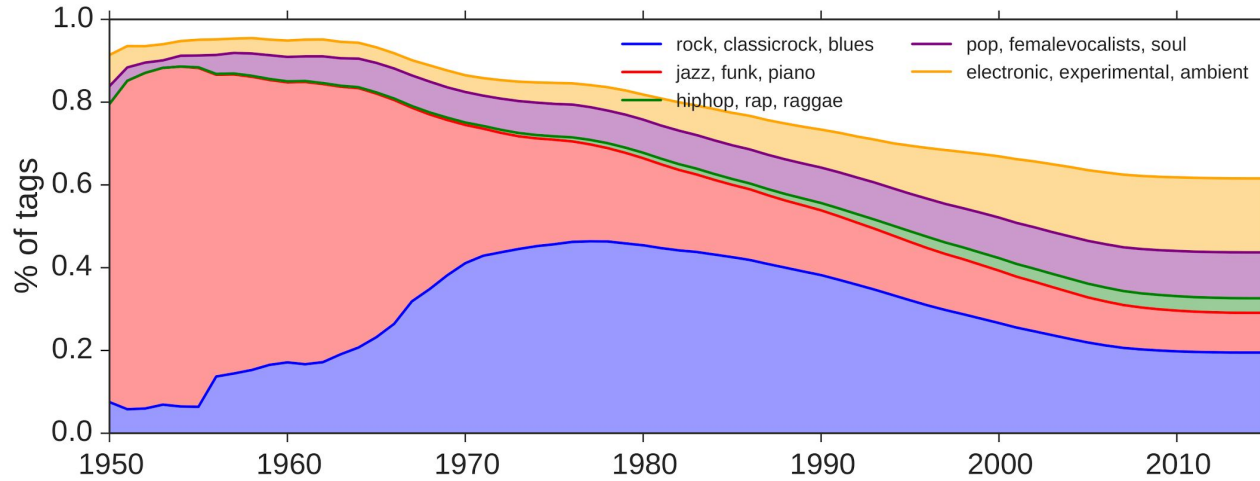
Cumulative percentage of tags in some topics
year by year

- Topics loosely represents music styles
- Few topics are usually dominating sharing more than 50% of the tags

Metrics Definition: Heterogeneity and Time

$$e_a = \frac{-\sum_{i \in L_a} f(l) \log f(l)}{\log |L_a|}$$

Topical Entropy: heterogeneity in the topics of the tags of an album

$$TS = \frac{\sum_{e \in E_a} |y_{tag_2} - y_{tag_1}|}{|E_a|}$$

Average Time Span: heterogeneity in the creation times of the tags of an album

$$A = \frac{\sum_{tag \in T_a} (y_a - y_{tag})/(y_a - 1950)}{|T_a|}$$

Average Tag Age: aging of the tags

Metrics Definition: Global Production Relation

$$M = \frac{1}{\sqrt{2}} \| \sqrt{f} - \sqrt{F} \|_2$$

Mainstreamness: Originality with respect to all the other albums

$$B_a = \sum_{l \in L_a} f(l) \frac{N_l(y_a + 1) - N_l(y_a)}{N_l(y_a)}$$

Burstiness: Contribution to the growth of a particular topic

Metrics Definition: Adjacent Possible and Innovation

$$f_{new} = \frac{|T_a \setminus T(y_a - 1)|}{|T_a|}$$

Novelty: Fraction of brand new tags introduced by the albums

$$Adj(a) = \frac{\sum_{tag \in T_a} f_{Adj}(tag, y_a)}{|T_a|}$$

Adjacent Possible: Average Fraction of Adjacent Possible Unlocked by the tags of the album.

[1] Kauffman, S. A. Investigations: The Nature of Autonomous Agents and the Worlds They Mutually Create. SFI working papers (Santa Fe Institute, 1996).

[2] Gravino, P., et al. "Crossing the horizon: exploring the adjacent possible in a cultural system.", in ICCC 2016, proceedings of the Seventh International Conference on Computational Creativity, Paris, 2016

Metrics Definition: Long-Term Impact

$$uch(a) = \frac{1}{|T_a|} \sum_{tag \in T_a} uch(tag) \left(1 - \frac{N(tag, y)}{N(tag)} \right)$$

Uchronia: Percentage of the network destroyed by the removal of the tags of the album

$$h_{uch}(a) = \frac{1}{|T_a|} \sum_{tag \in T_a} h_{uch}(tag) \left(1 - \frac{N(tag, y)}{N(tag)} \right)$$

Uchronia Entropy: Variation in the entropy of the distribution of the topics after the removal of the tags of the album

Levels of Significance

- Popularity is the most common value attributed to a music album

- Last.fm **Playcount** is a good proxy for this kind of success (small-time bias = everlasting popularity)

- Other possibilities are offered by **expert-made lists:**

  - The 500 Greatest Albums of All Time according to the **Rolling Stone Magazine (RS)**

  - The 500 Greatest Albums of All Time according to the **NME Magazine**

  - **GRAMMY Hall Of Fame** List

  - The **National Recording Registry (NRR)** list

Levels of Significance
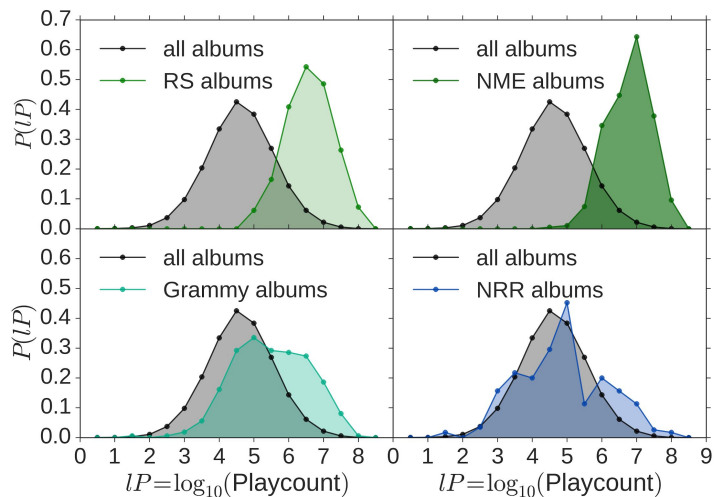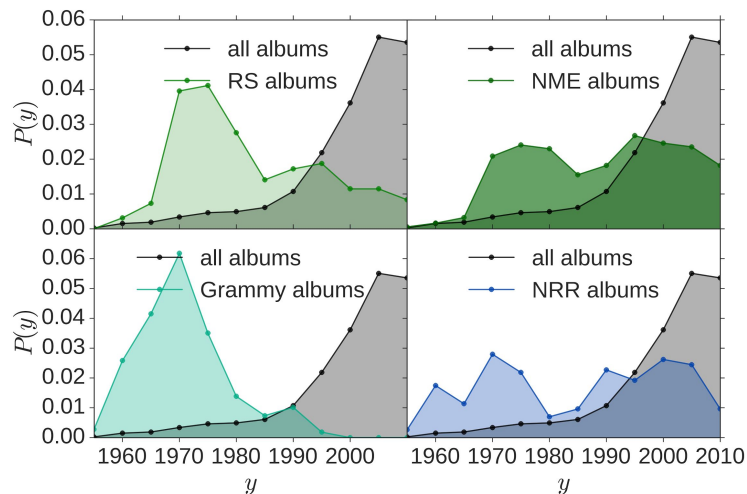
- Bias Towards before 80s albums for Rolling Stone and Grammy
- Bias Towards popularity for Rolling Stone and NME
- Small or no bias for Grammy and NRR

Levels of Significance

- Four Main Categories:

  - **Popular Albums**: Albums with Playcount > 75th percentile of the Playcount distribution

  - **Highly Popular Albums**: Albums with Playcount > 90th percentile of the Playcount distribution

  - **RS/NME Albums**: Expert-made lists of albums with popularity bias

  - **Grammy/NRR Albums**: Expert-made lists of albums without popularity bias

- Are we able to predict if an album belongs to one of these categories?

Classification

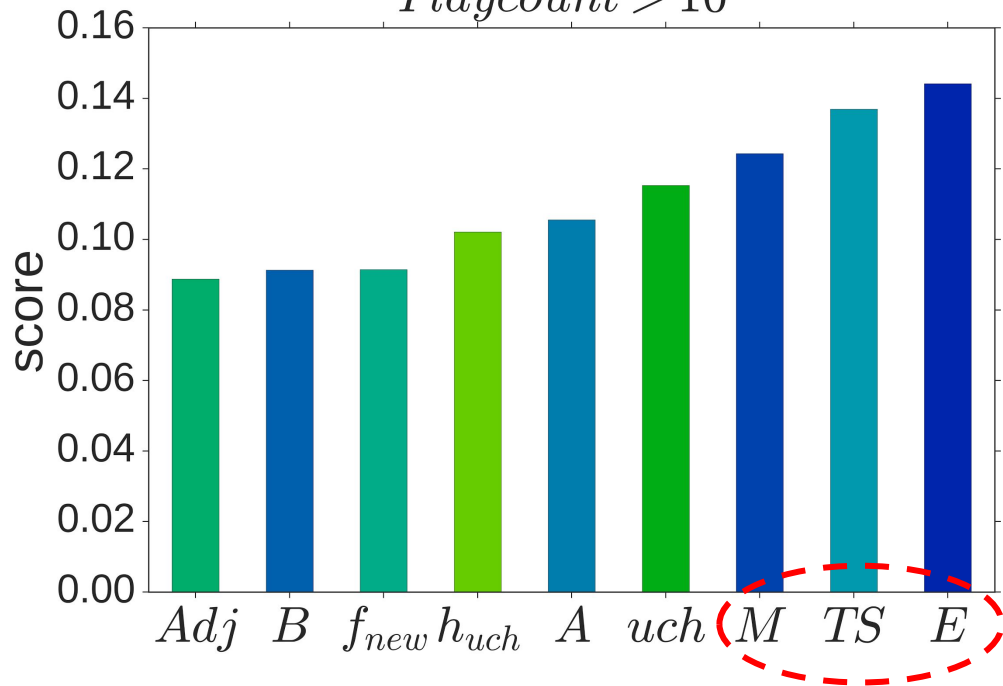**Random Forest Classification:** 80% of the sample as training, correction for small categories.

Popular Category: AuROC 0.86
Highly Popular Category: AuROC 0.92
RS/NME Category: AuROC 0.89
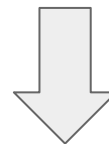**Grammy/NRR Category:** AuROC 0.81
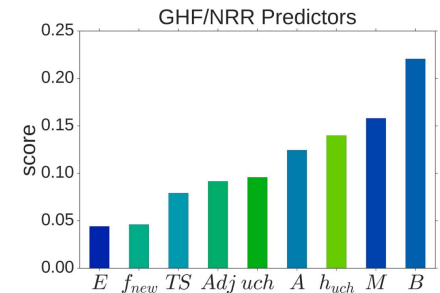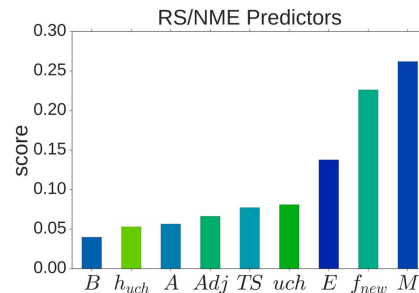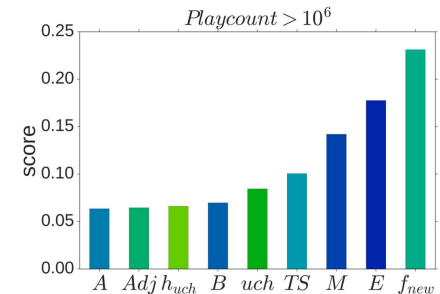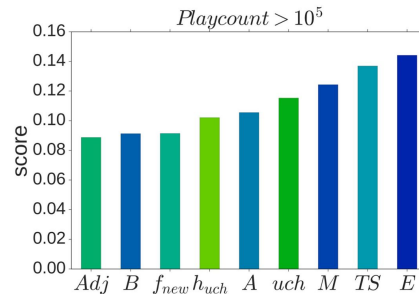
Rankings of the most important predictors in the classification with the Random Forest Algorithm
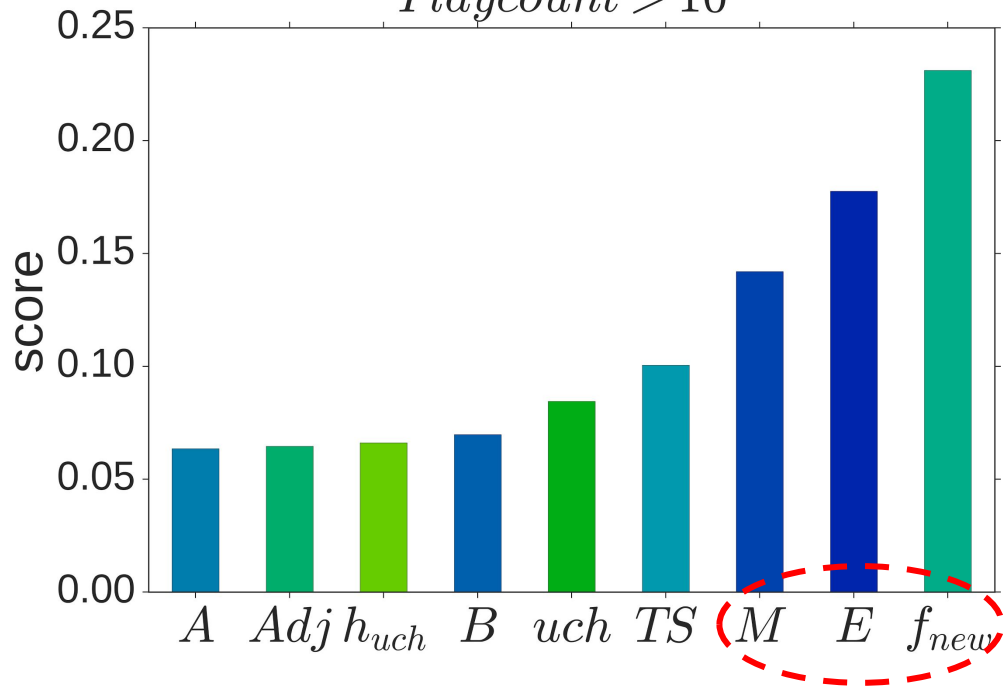
Popular Category

$Playcount > 10^5$

- Topical Entropy and Average Time Span are the best predictor
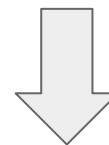
- Avg. Time Span

- Heterogeneity!
- Plus Mainstreamness
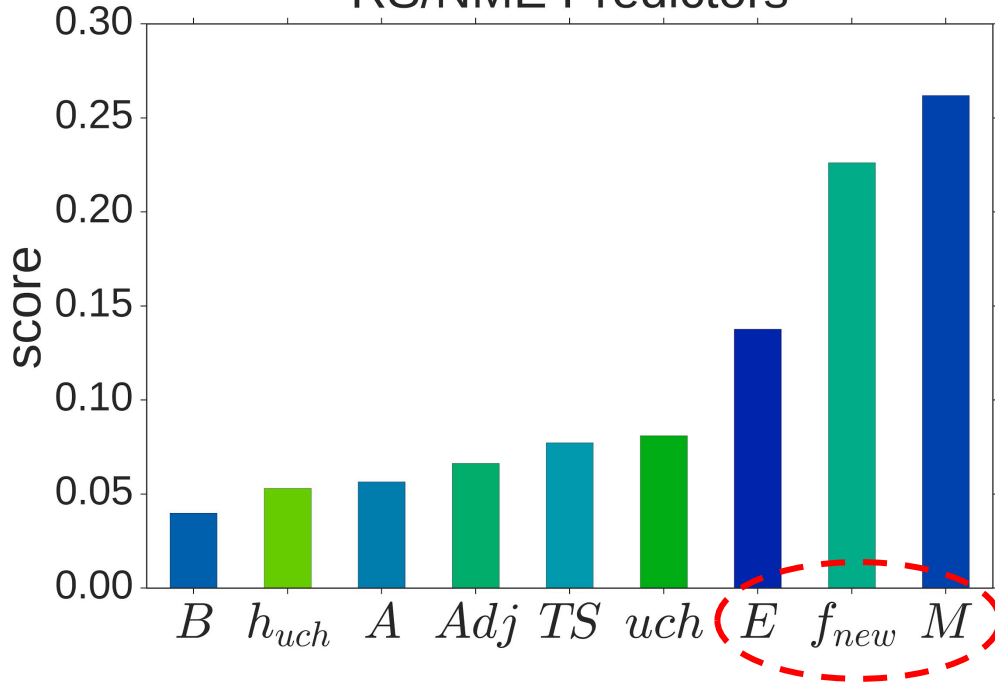
Highly Popular Category

$Playcount > 10^6$

- Innovation comes into play!

- Heterogeneity + Introduction of Novelties!
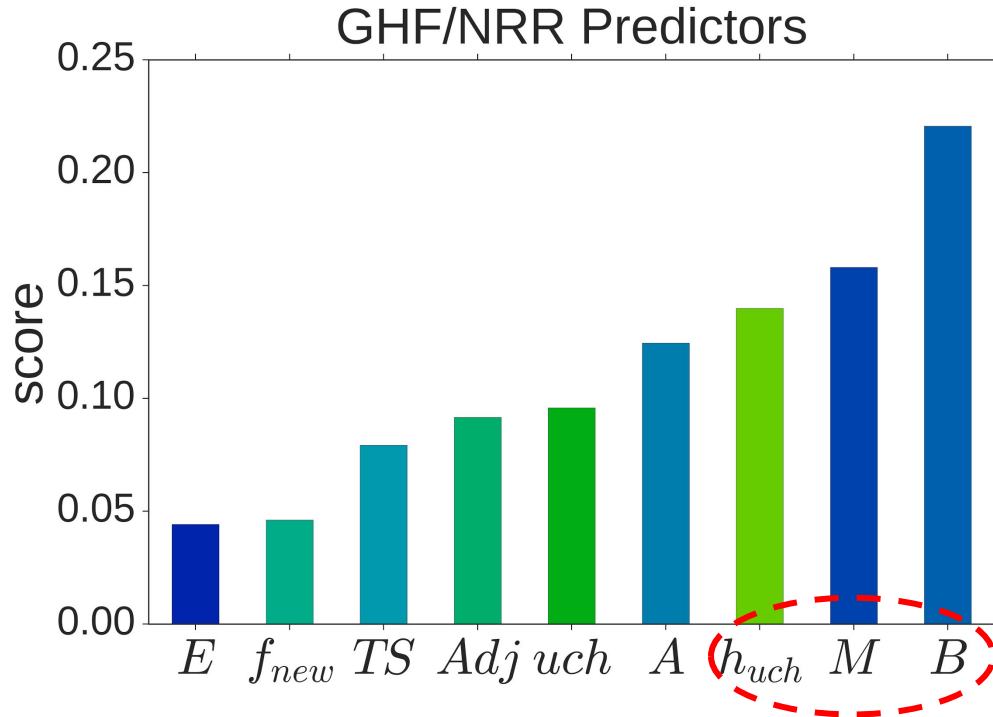
- Plus Mainstreamness

RS/NME Category

RS/NME Predictors

- Prediction quite similar to the highly popular category

- Mainstreamness is the most important predictor

- Maybe not so surprising..

GRAMMY/NRR Category

GHF/NRR Predictors

- **Burstiness** is the most important predictor: at the beginning of a trend!

- **Long-Term impact:** influence over the largest topics

- **Average Tag Age** not shown but indicates tags very related to the release date of the album

⬇

- **Cultural and Historical Value!**

Conclusions and Perspectives

- We introduced a framework based on an expanding conceptual space representing music production

- We are able to identify albums appreciated by the public as well as those who had some impact from a cultural point of view

- Popularity = Heterogeneity + Innovation

- Cultural Value = Long-term Impact + New Trends

- Mainstreamness: don't be too different!

- Can we find and characterize other levels of significance? (is the adjacent possible relevant?)

- Other frameworks? Paintings, Movies, Scientific Papers? "Sleeping Beauties"?
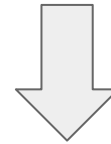
Thanks for the Attention

Monechi, Bernardo, et al. "Significance and popularity in music production." *Open Science* 4.7 (2017): 170433.
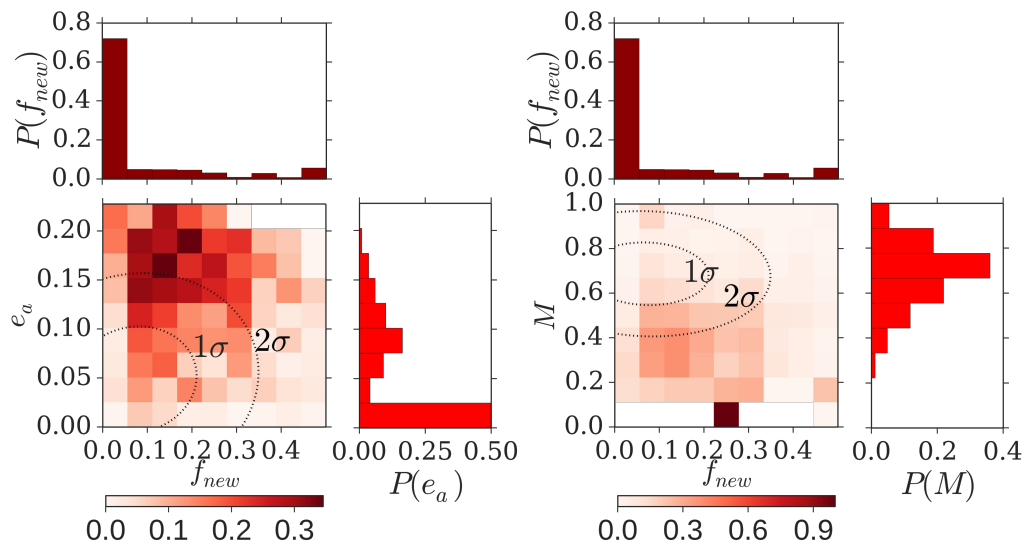
Popular Category
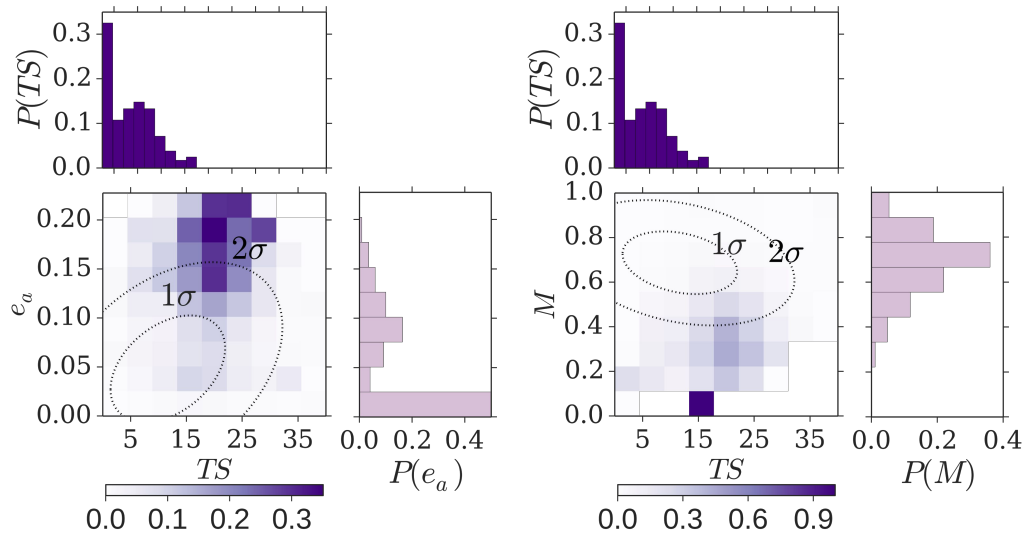
Conditional probabilities of belonging to the Popular Category

- Topical Entropy and Average Time Span are the best predictor
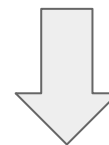
- Avg. Time Span

- Heterogeneity + Similarity to other artworks!
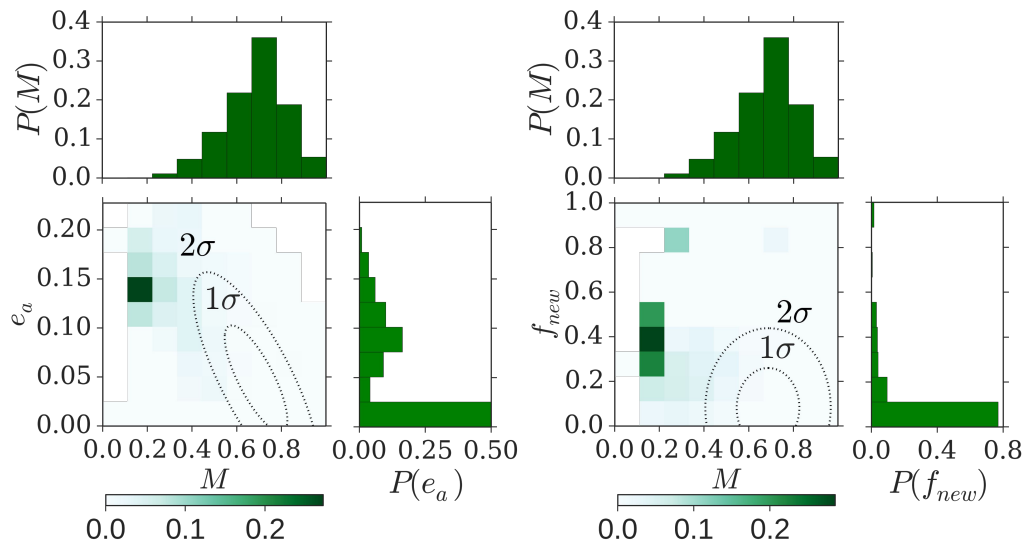
Highly Popular Category

Conditional probabilities of belonging to the
Highly Popular Category

- Innovation comes into play!

- Heterogeneity + Introduction of Novelties!

- Mainstreamness is not shown but has the same effect as before

RS/NME Category

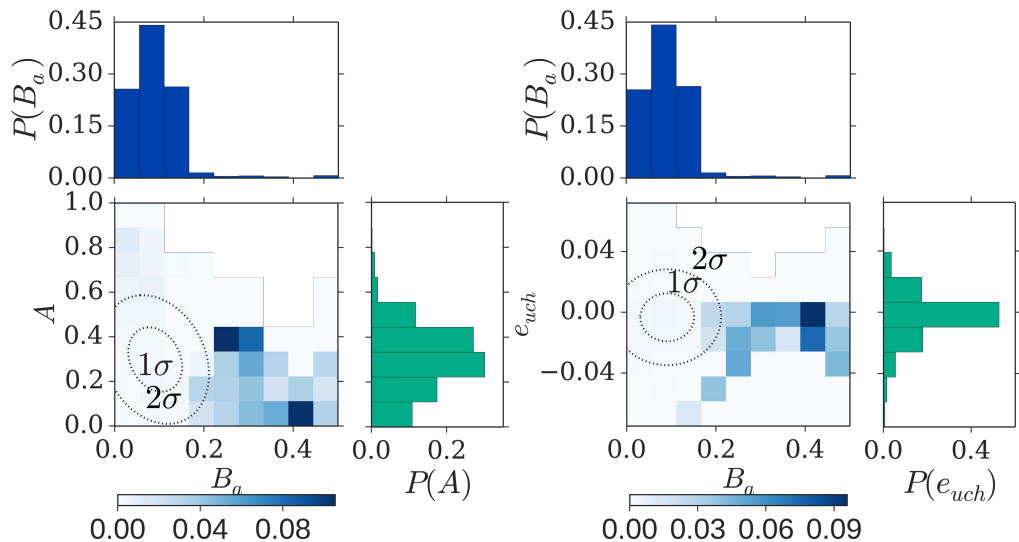Conditional probabilities of belonging to the RS/NME Category

- Prediction quite similar to the highly popular category

- Mainstreamness is the most important predictor

- Maybe not so surprising..

GRAMMY/NRR Category

Conditional probabilities of belonging to the GRAMMY/NRR Category

- **Burstiness** is the most important predictor: at the beginning of a trend!

- **Long-Term impact:** influence over the largest topics

- **Average Tag Age** not shown but indicates tags very related to the release date of the album
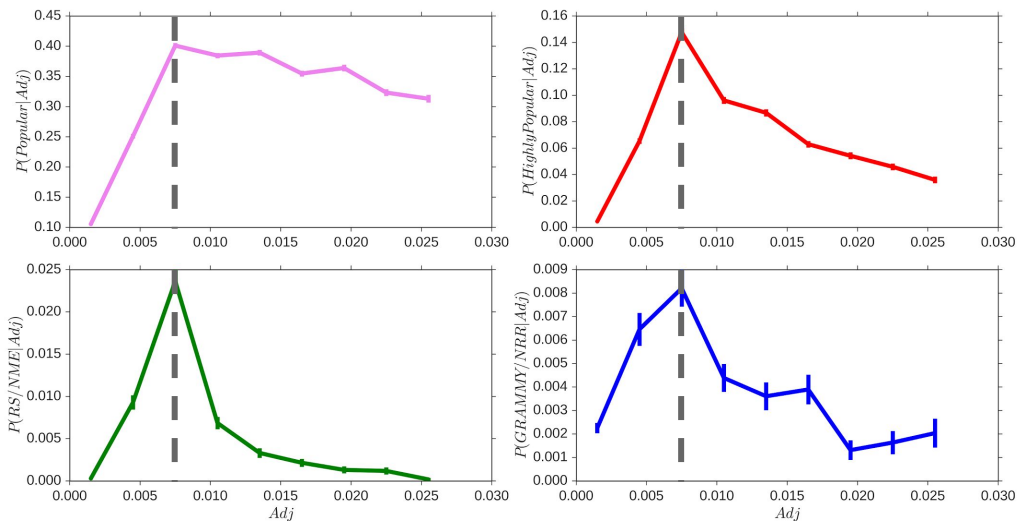
- **Cultural and Historical Value!**

What about the Adjacent Possible Metrics?
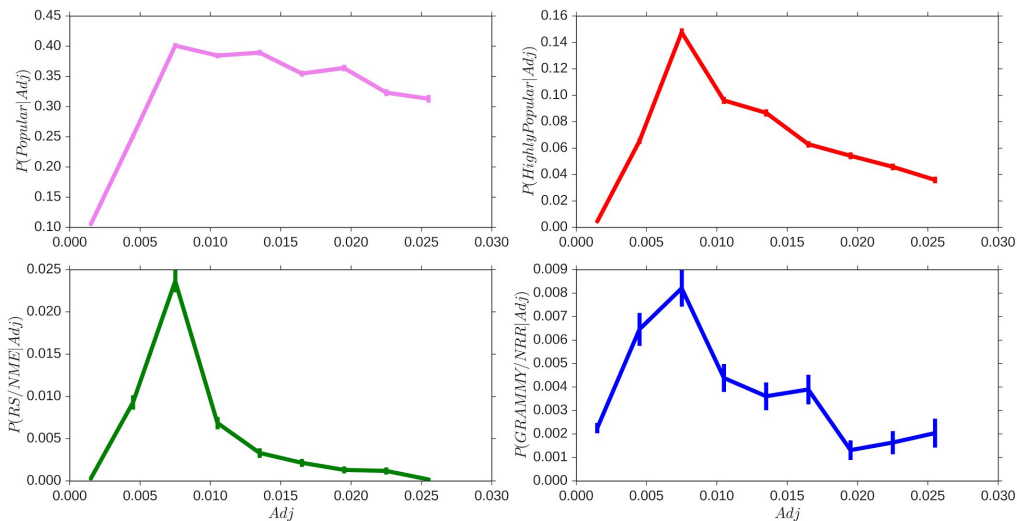
Conditional probabilities of belonging to the one of the four categories as a function of the Adjacent Possible Metrics

- **Adjacent Possible Metric** is never an important predictor
- It is a one-step-away in the future metric
- Related with many possibilities which will not be relevant..

What about the Adjacent Possible Metrics?

Conditional probabilities of belonging to the one of the four categories as a function of the Adjacent Possible Metrics

- **Adjacent Possible Metric** is never an important predictor
- It is a one-step-away in the future metric
- Related with many possibilities which will not be relevant..

- **No-one is rewarding the avant-garde!**