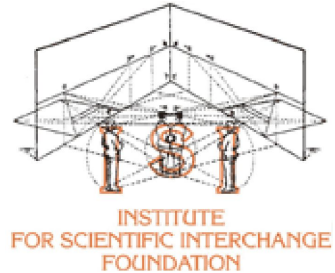
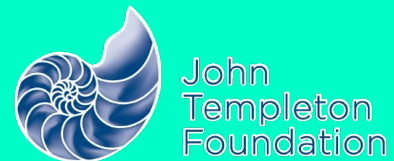


# SIGNIFICANCE AND POPULARITY IN MUSIC PRODUCTION



**Bernardo Monechi, ISI Foundation**  
**Cognition & Socio-Ecology Session**  
**CCS2016 Amsterdam**

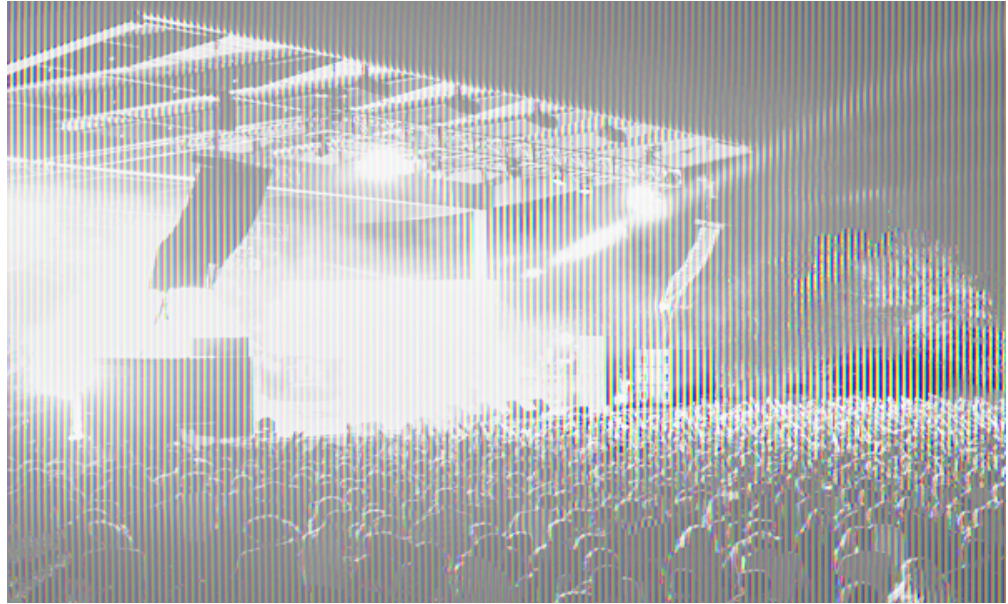


# THE STRUGGLE FOR SUCCESS

- 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..



- 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...

The screenshot shows the Last.fm page for Frank Zappa's album 'Hot Rats'. The album cover is on the left, and the right side contains the album title, artist name, and navigation tabs for Overview, Wiki, Artwork, and Shouts. Below this, there are statistics for 'SCROBBLES' (1.7M) and 'LISTENERS' (129.6K). A 'BUY THIS ALBUM' section includes a 'Buy' button with a 'iTunes' logo and a dropdown arrow. The album is tagged with 'progressive rock', 'jazz fusion', 'jazz-rock', '1969', 'rock', and 'instrumental'. A description follows, stating it's the second solo album by Frank Zappa, released in October 1969, consisting of six songs, five of which are instrumental. Below the description is a 'Tracklist' section with a table of tracks, their durations, and play counts. At the bottom, there is a 'Shoutbox' section with a 'Log in to Last.fm or sign up to leave a shout.' link and a user named 'GoatUser' who has posted a shout.

RELEASE DATE	RUNNING LENGTH	RUNNING TIME	BUY THIS ALBUM
15 October 1969	6 tracks	43:40	<input type="button" value="Buy"/> <input type="button" value="iTunes"/>

[progressive rock](#) · [jazz fusion](#) · [jazz-rock](#) · [1969](#) · [rock](#) · [instrumental](#)

Hot Rats is the second solo album by Frank Zappa. It was released in October 1969. The album consists of six songs, five of which are instrumental (the song "Willie the Pimp" features a short vocal by Captain Beefheart). It was Zappa's first recording project after the dissolution of the original Mothers of Invention. Because it focuses on long

Tracklist			
1	▶️ Peaches en Regalia	3:58	187,256
2	▶️ Willie the Pimp	9:25	93,213
3	▶️ Son of Mr. Green Genes	8:58	66,392
4	▶️ Little Umbrellas	3:09	64,244
5	▶️ The Gumbo Variations	12:55	58,511
6	▶️ It Must Be a Camel	5:15	52,750

Shoutbox Recently Popular ▾

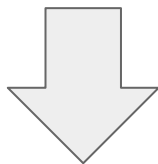
[Log in to Last.fm or sign up to leave a shout.](#)

**GoatUser**  
I don't listen to the rest that much but those first two tracks have to be one of the best two-punch combo openers I've ever heard.



# 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

# SIGNIFICANCE IN POPULAR MUSIC

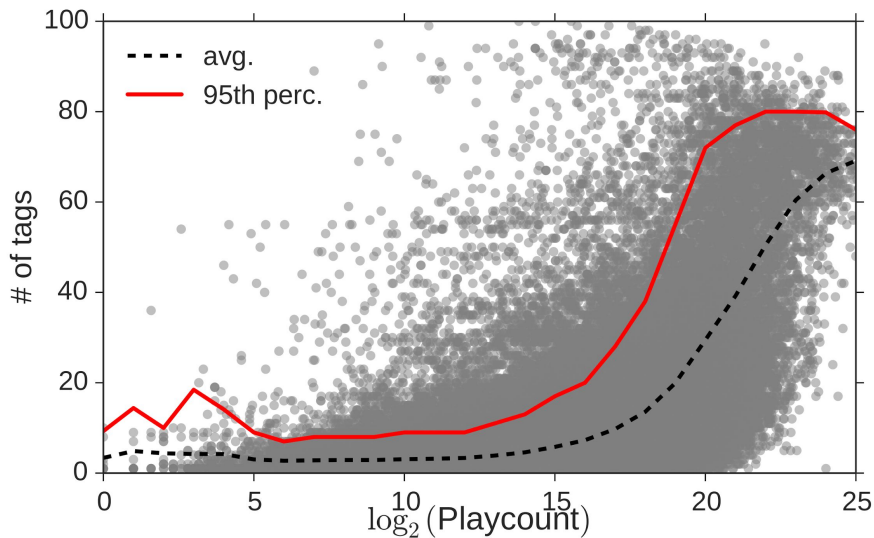
Not a completely brand new approach:

- [1] Cattuto, Ciro, et al. "Collective dynamics of social annotation." *Proceedings of the National Academy of Sciences* 106.26 (2009): 10511-10515.
- [2] Bischoff, Kerstin, et al. "Can all tags be used for search?." *Proceedings of the 17th ACM conference on Information and knowledge management*. ACM, 2008. APA
- [3] Sreenivasan, Sameet. "Quantitative analysis of the evolution of novelty in cinema through crowdsourced keywords." arXiv preprint arXiv:1304.0786 (2013)

# DATA SELECTION

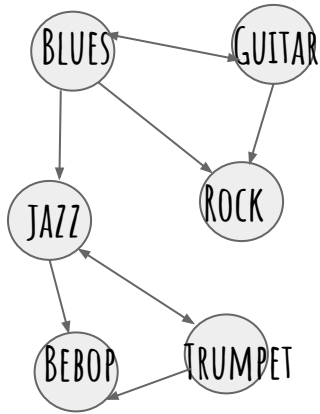
- Wrong Release Dates: many times not the first one.
- Integration with MusicBrainz DB.
- Over-Tagging: poorly validated tags in non-popular albums

Time Resolution: 1 year  
163829 Albums  
108984 Distinct Tags  
From 1950 to 2015



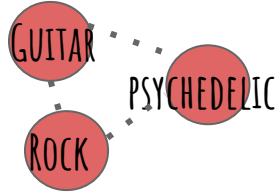
Number of album tags as a function of the album Playcount

# GROWING CONCEPTUAL SPACE



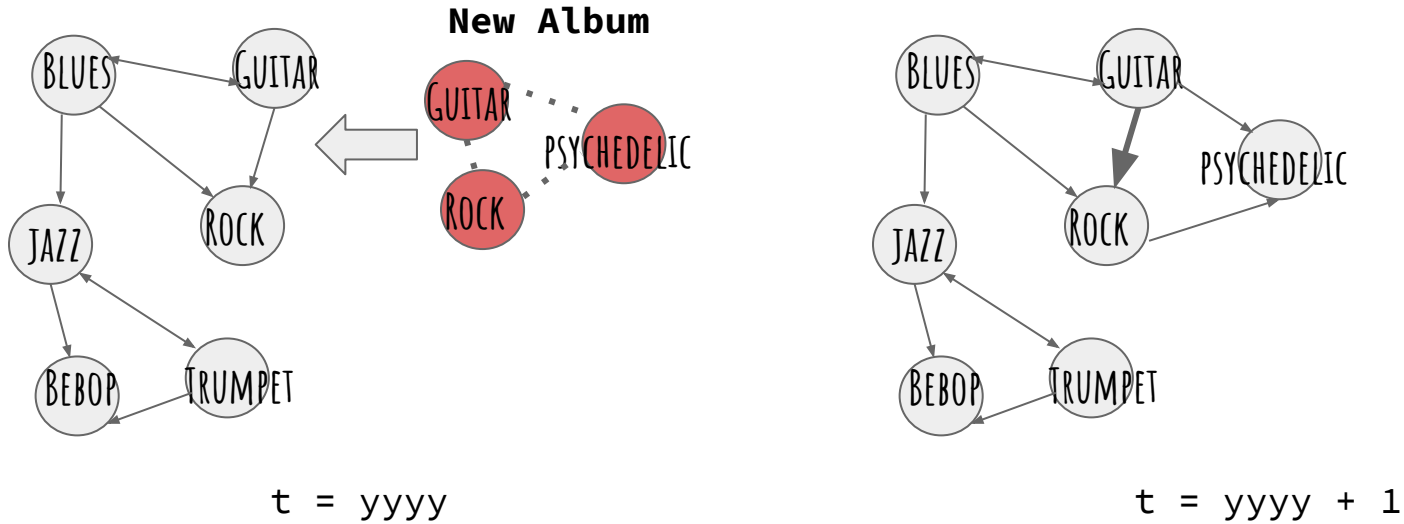
t = yyyy

## New Album



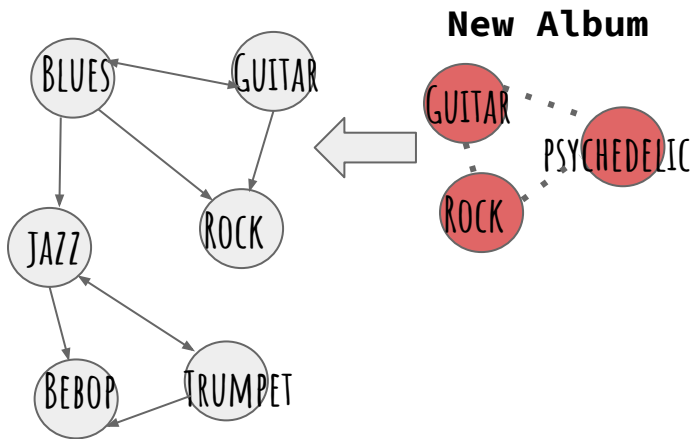
- Tags are the nodes of the network
- A link represents co-occurrence in an album
- Link weight = number of co-occurrences
- Link direction = time ordered

# GROWING CONCEPTUAL SPACE



- 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

# GROWING CONCEPTUAL SPACE

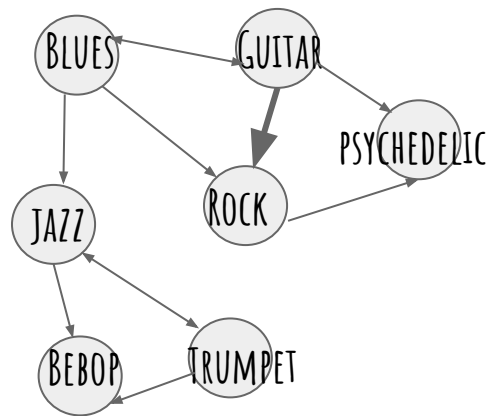


$t = yyyy$

$a$ : an album

$tag \in T_a$ : set of album tags

$\{tag_i, tag_j\} \in E_a$ : set of album tag couples

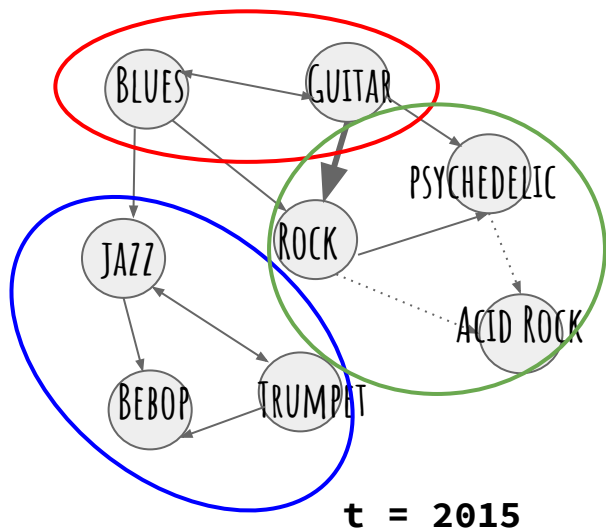


$t = yyyy + 1$

$y_a$ : album release date

$y_{tag}$ : tag creation year

# TOPICAL REPRESENTATION



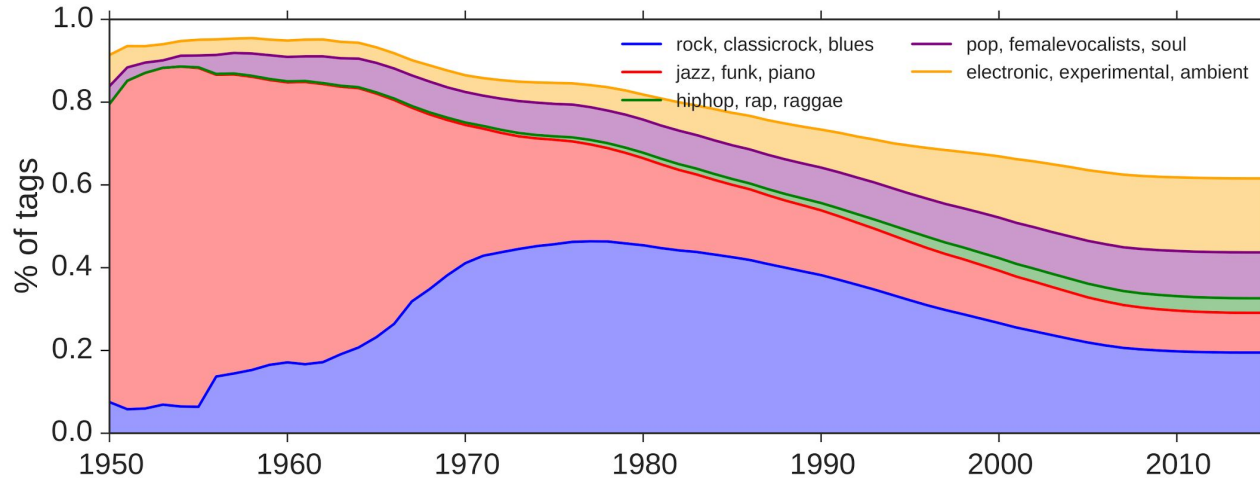
**Topical Representation:** apply community detection to the final network, each tag will be labelled according to the community it belongs to.

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

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

**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

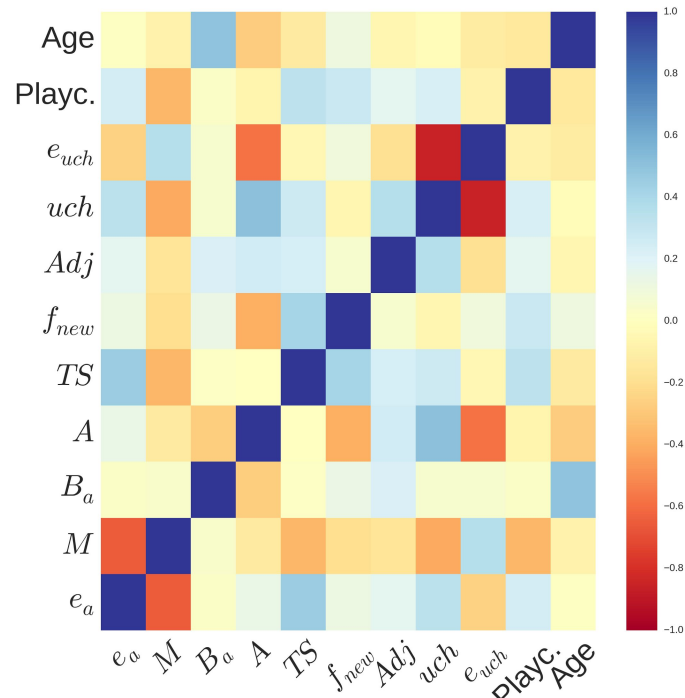
$$uch(a) = \frac{1}{|T_a|} \sum_{tag \in T_a} uch(tag) \left( 1 - \frac{N(tag, y)}{N(tag)} \right)$$

**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

# CORRELATIONS BETWEEN METRICS



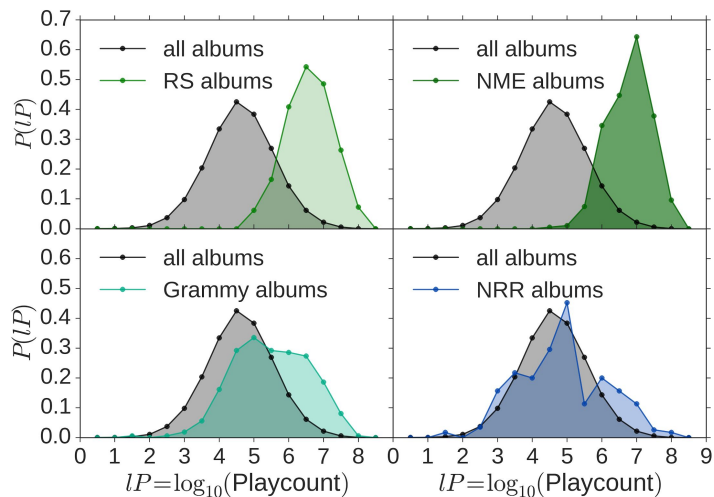
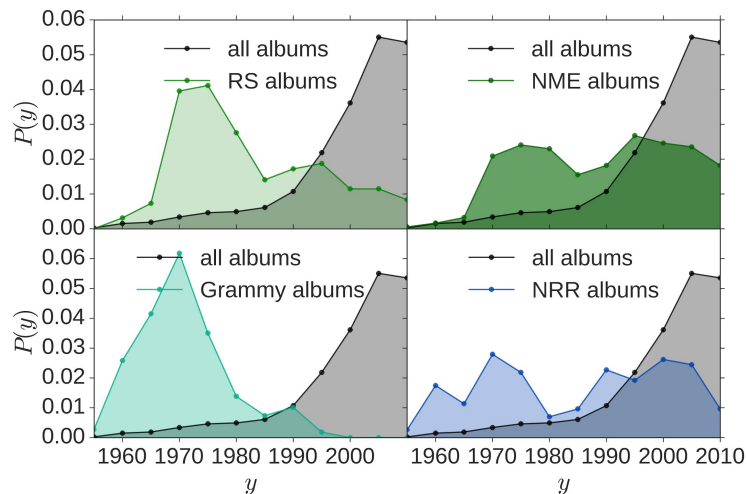
Matrix of Spearman's Correlation Coefficients between metrics

- Strong Correlations between heterogeneity metrics, Mainstreamness and Playcount
- **Playcount slightly anti-correlated with Age**
- Almost every metrics is uncorrelated with the albums age
- Long-Term Impact correlated with Adjacent-Possible

# 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

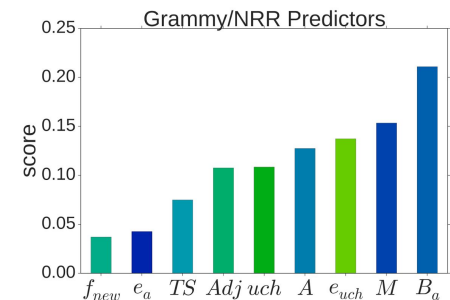
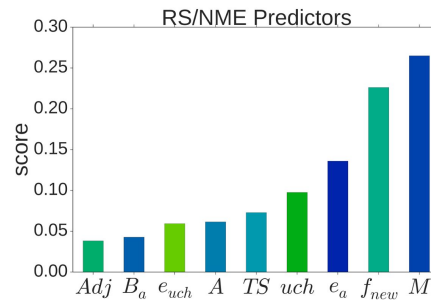
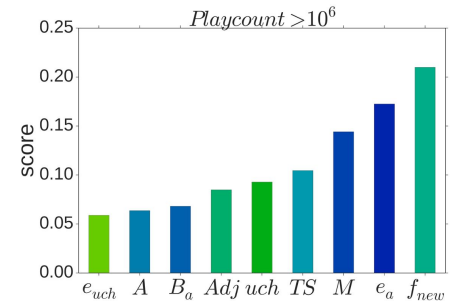
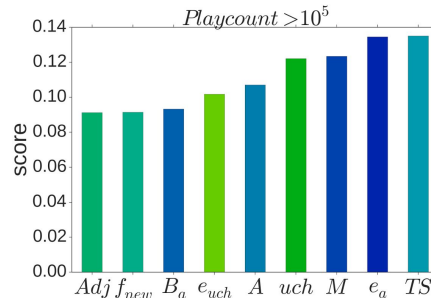
- Four Main Categories:
  - **Popular Albums:** Albums with Playcount  $>$  75th percentile of the Playcount distribution
  - **Highly Popular Albums:** Albums with Playcount  $>$  90th percentile of the Playcount distribution
  - **RS/NME Albums:** Expert-made lists of albums with popularity bias
  - **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

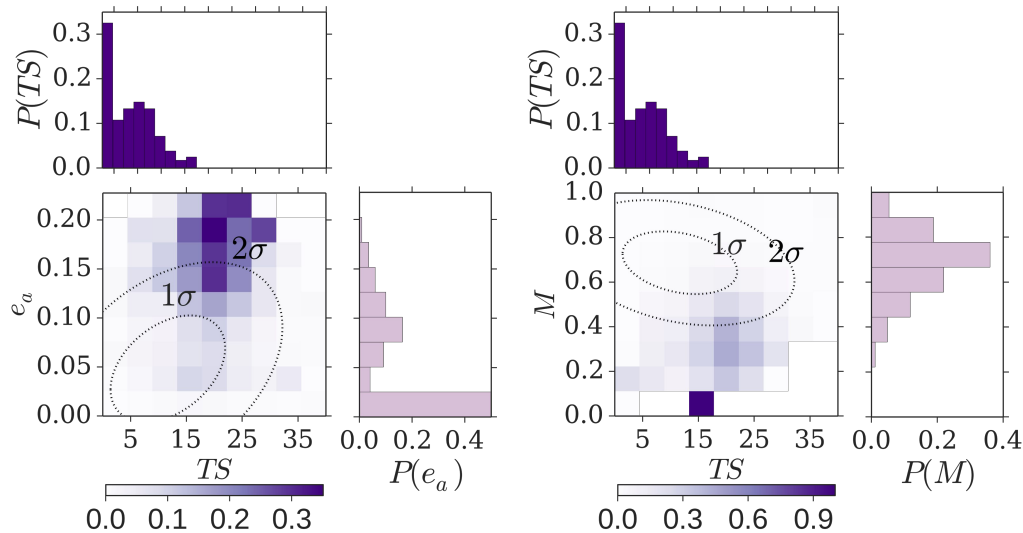
**Random Forest Classification:** 80% of the sample as training, correction for small categories.

**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

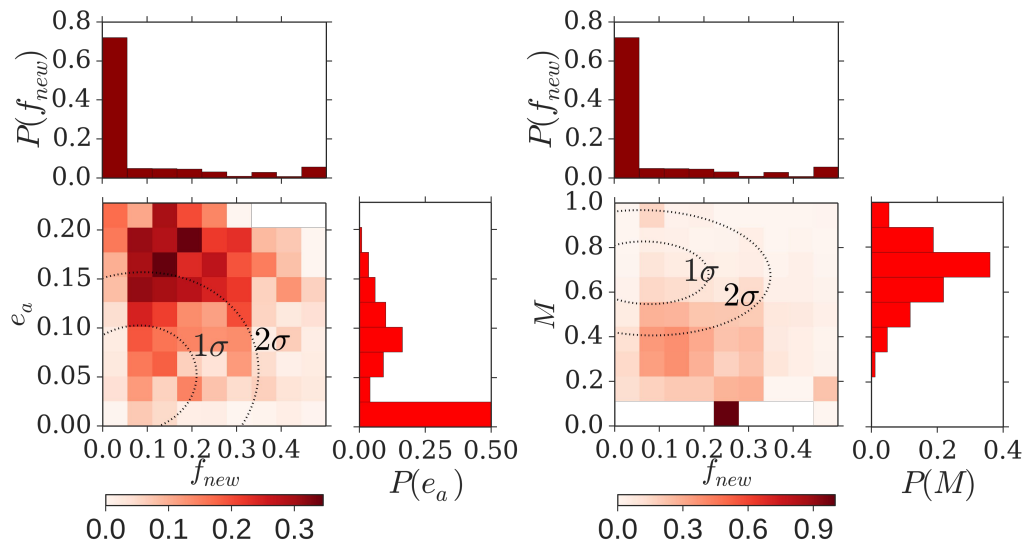
# POPULAR CATEGORY



Conditional probabilities of belonging to the Popular Category

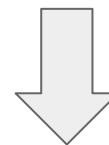
- Topical Entropy and Average Time Span are the best predictor
  - Small Mainstreamness enhance the probability of belonging to the category
- ↓
- Heterogeneity + Similarity to other artworks!

# HIGHLY POPULAR CATEGORY



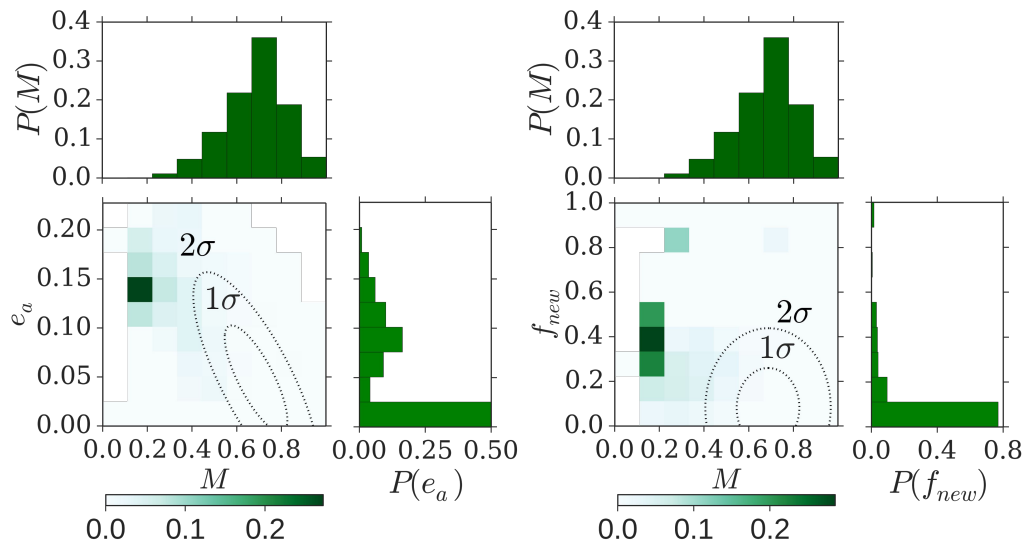
Conditional probabilities of belonging to the Highly Popular Category

- Innovation comes into play!



- Heterogeneity + Introduction of Novelties!
- Mainstreamness is not shown but has the same effect as before

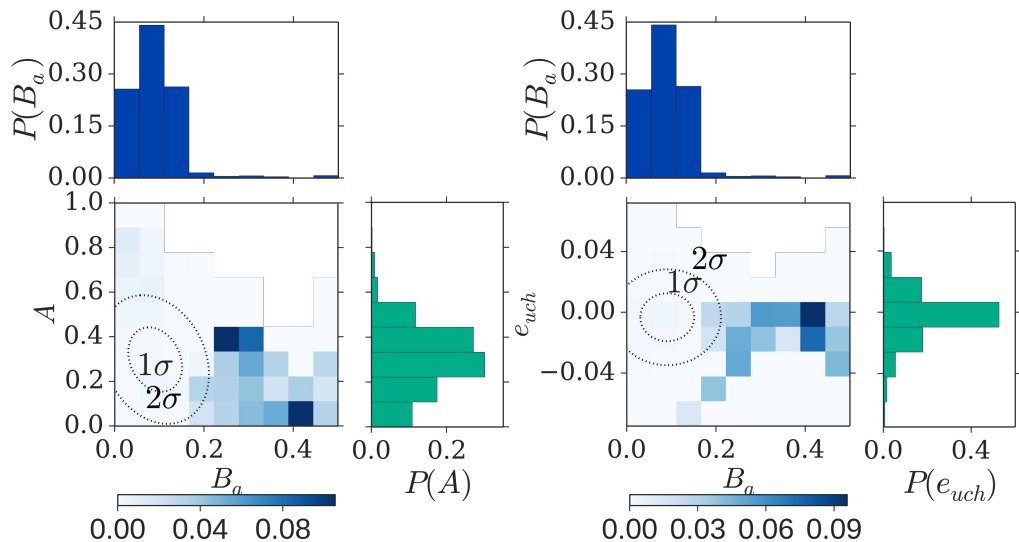
# RS/NME CATEGORY



- Prediction quite similar to the highly popular category
- Mainstreamness is the most important predictor
- Maybe not so surprising..

Conditional probabilities of belonging to the  
RS/NME Category

# GRAMMY/NRR CATEGORY



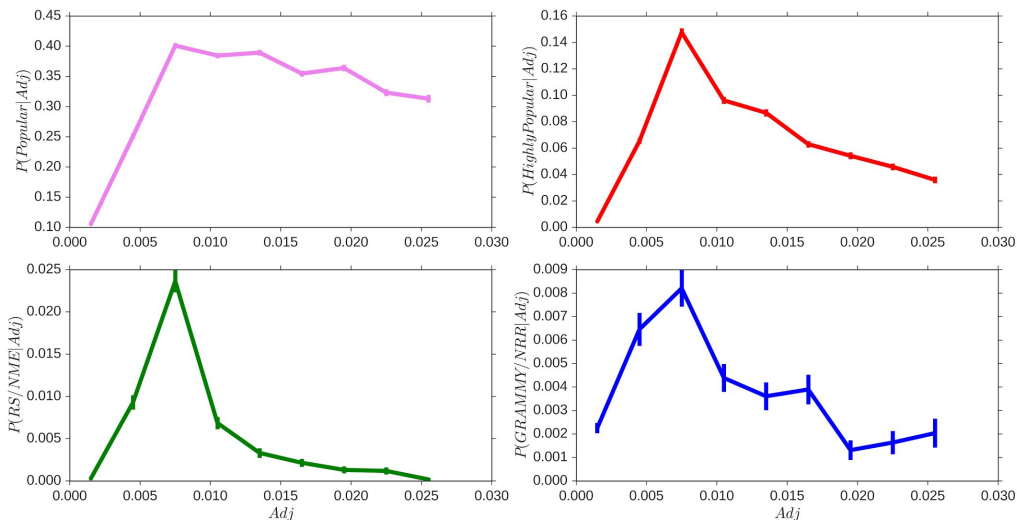
Conditional probabilities of belonging to the  
GRAMMY/NRR Category

- **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!**

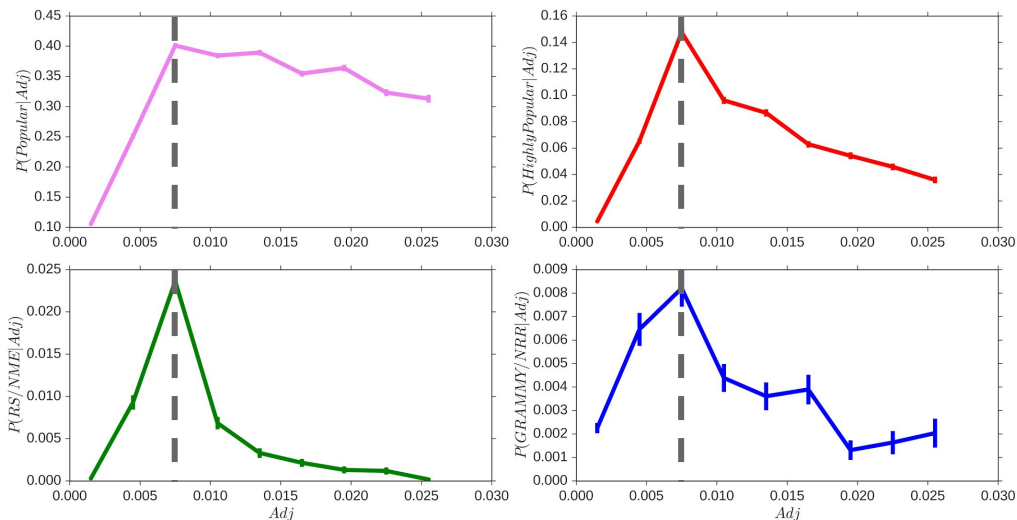
# WHAT ABOUT THE ADJACENT POSSIBLE METRICS?



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!**

# 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

**Collaborators:**

Pietro Gravino, ISI Foundation (Turin)

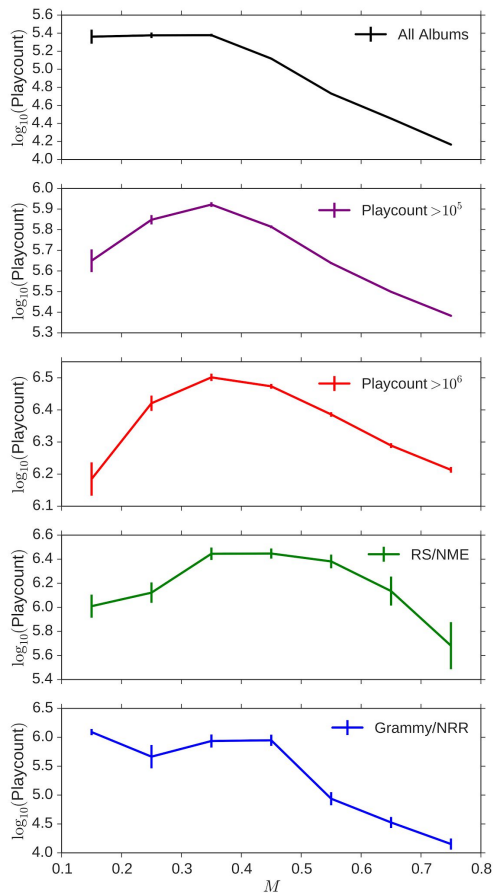
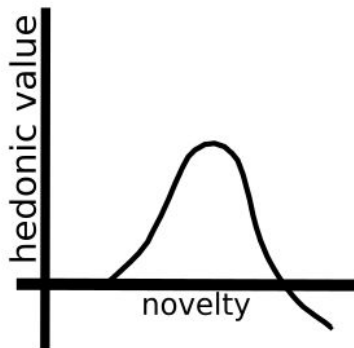
Vito D.P. Servedio, Institute for Complex Systems (Rome)

Francesca Tria, Sapienza University of Rome

Vittorio Loreto, Sapienza University of Rome

# MAINSTREAMNESS VS PLAYCOUNT

- A maximum seems to emerge
- Spearman's correlation coefficient is always negative
- Some similarity with the “Wundt Curve”



# POPULAR MUSIC DB: LAST.FM

- We will focus on the **Albums** in Last.fm
- We will reconstruct the historical music production
- They are (used to be) the final product of the artistic creation activity
- “Death of Album” phenomenon is quite recent

The screenshot shows the Last.fm page for the album 'Hot Rats' by Frank Zappa. The album cover is visible on the left. The main header includes the artist name 'Frank Zappa', the album title 'Hot Rats', and navigation tabs for 'Overview', 'Wiki', 'Artwork', and 'Shouts'. Below this, there are statistics: 'SCROBBLES 1.7M' and 'LISTENERS 129.6K'. A 'BUY THIS ALBUM' section includes a 'Buy' button with a 'iTunes' logo and a dropdown arrow. The album is categorized as 'progressive rock · jazz fusion · jazz-rock · 1969 · rock · instrumental'. A short description follows: 'Hot Rats is the second solo album by Frank Zappa. It was released in October 1969. The album consists of six songs, five of which are instrumental (the song "Willie the Pimp" features a short vocal by Captain Beefheart). It was Zappa's first recording project after the dissolution of the original Mothers of Invention. Because it focuses on long loing, the music sounds very different than earlier Zappa albums'. The tracklist is displayed below, with columns for track number, play button, track name, duration, and listener count. The tracks are: 1. Peaches en Regalia (3:58, 187,256 listeners), 2. Willie the Pimp (9:25, 93,213 listeners), 3. Son of Mr. Green Genes (8:58, 66,392 listeners), 4. Little Umbrellas (3:09, 64,244 listeners), 5. The Gumbo Variations (12:55, 58,511 listeners), and 6. It Must Be a Camel (5:15, 52,750 listeners). A 'Shoutbox' section is at the bottom, with a 'Log in to Last.fm or sign up to leave a shout.' prompt. A user named 'GoatUser' has posted a shout: 'I don't listen to the rest that much but those first two tracks have to be one of the best two-punch combo openers I've ever heard.' The shout is dated '18 Jul 10:55pm'.

FRANK ZAPPA

Frank Zappa  
**Hot Rats**  
Overview Wiki Artwork Shouts

SCROBBLES 1.7M LISTENERS 129.6K

BUY THIS ALBUM  
iTunes Buy

progressive rock · jazz fusion · jazz-rock · 1969 · rock · instrumental

Hot Rats is the second solo album by Frank Zappa. It was released in October 1969. The album consists of six songs, five of which are instrumental (the song "Willie the Pimp" features a short vocal by Captain Beefheart). It was Zappa's first recording project after the dissolution of the original Mothers of Invention. Because it focuses on long loing, the music sounds very different than earlier Zappa albums

Tracklist

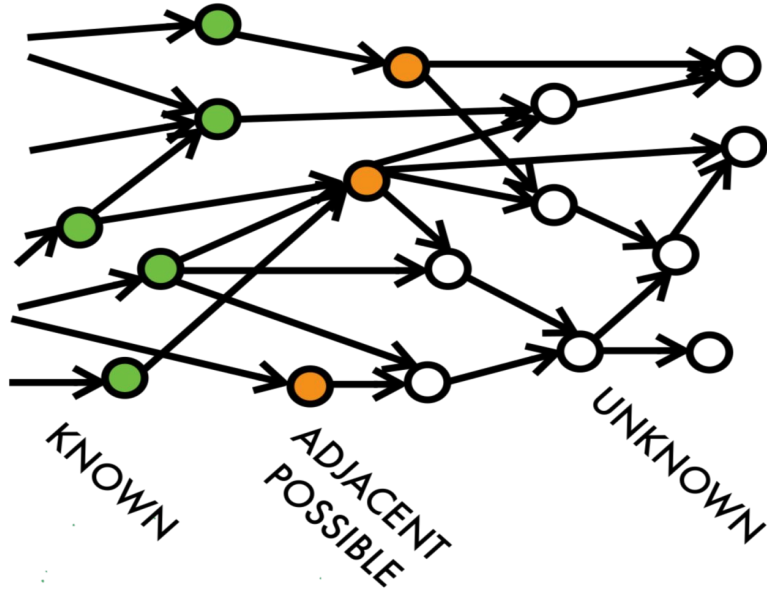
Track	Duration	Listeners
1 Peaches en Regalia	3:58	187,256
2 Willie the Pimp	9:25	93,213
3 Son of Mr. Green Genes	8:58	66,392
4 Little Umbrellas	3:09	64,244
5 The Gumbo Variations	12:55	58,511
6 It Must Be a Camel	5:15	52,750

Shoutbox

Log in to Last.fm or sign up to leave a shout.

GoatUser  
18 Jul 10:55pm  
I don't listen to the rest that much but those first two tracks have to be one of the best two-punch combo openers I've ever heard.

# THE ADJACENT POSSIBLE FRAMEWORK



- The realization of something new (a new movie) is enabled by the presence of other things
- The Adjacent Possible is not real (yet)
- It consists of all those things which are just “one step” from being real
- Actualized by modification or recombination of the existent

The adjacent possible is a kind of shadow future, hovering on the edges of the present state of things, a map of all the ways in which the present can reinvent itself.

Robert Johnson, from the Wall Street Journal article 'The Genius of the Tinkerer'