

Walk the Talk

Analyzing the relation between implicit and explicit feedback for preference elicitation

Denis Parra, University of Pittsburgh

Dr. Xavier Amatriain, Telefonica R&D



Which is the plan of this presentation?



Ah?

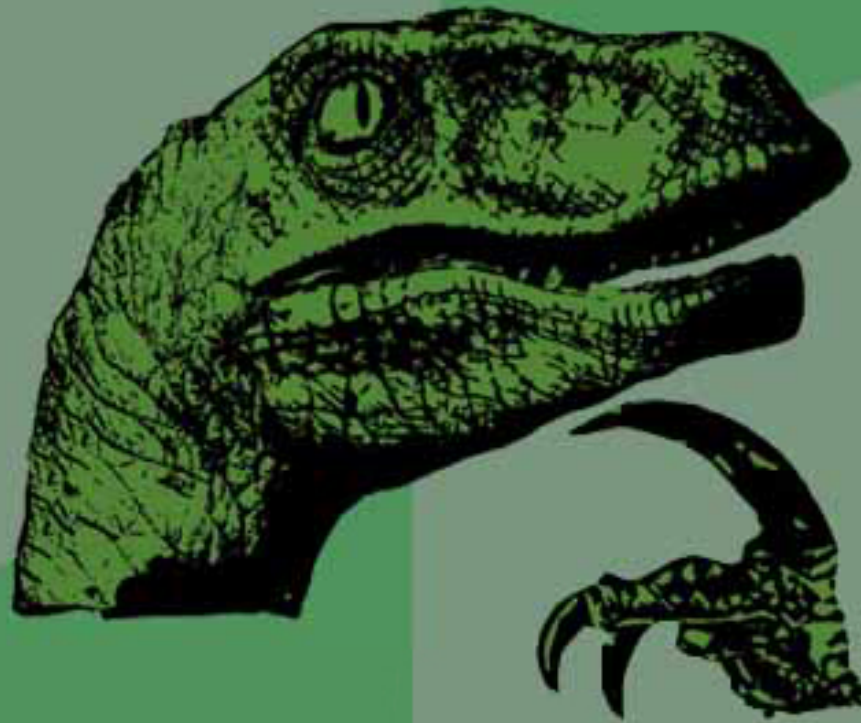


Outline

- Introduction (motivation)
- The survey on last.fm users
 - What did we crawl, what did we ask, sampling strategy, results
- General analysis of the results
- Regression Analysis
- Discussion, ongoing work and (some) conclusions



**Which motivation did you
have**



To do this research?

Introduction (1/2)

- Most of recommender system approaches rely on explicit information of the users, but...
- Explicit feedback: scarce (people are not especially eager to rate or to provide personal info)
- Implicit feedback: Is less scarce, but ([Hu et al., 2008](#))

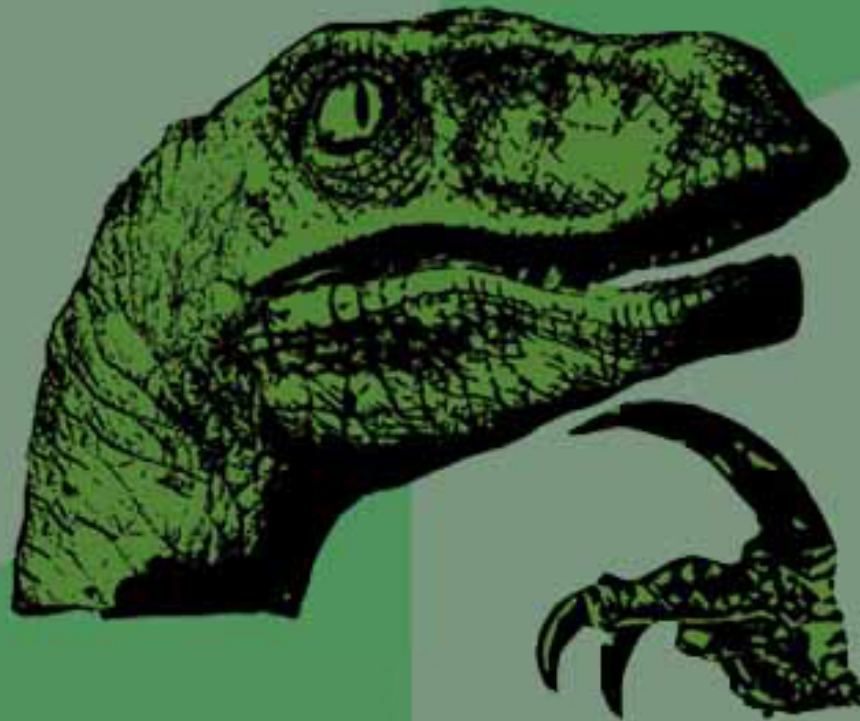
There's no negative feedback	... and if you watch a TV program just once or twice?
Noisy	... but explicit feedback is also noisy (Amatriain et al., 2009)
Preference & Confidence	... we aim to map the I.F. to preference (our main goal)
Lack of evaluation metrics	... if we can map I.F. and E.F., we can have a comparable evaluation



Introduction (2/2)

- Is it possible to map implicit behavior to explicit preference (ratings)?
- Which variables better account for the amount of times a user listens to online albums? [Baltrunas & Amatriain CARS '09 workshop – RecSys 2009.]
- **OUR APPROACH: Study with Last.fm users**
 - Part I: demographics and online music consumption
 - Part II: Rating 100 albums collected from their last.fm user profile

**Which user data did you
obtain from last.fm users?**





2.1 Demographic/Consumption Data

- **Gender**
- **Age**
- **Country**
- Hours per week spent on internet [**int_hrs_per_week**]
- Hours per week listening to music online [**msc_hrs_per_week**]
- Number of concerts per year [**conc_per_year**]
- Do you read specialized music blogs or magazines? [**blogs_mag**]
- Do you have experience evaluating music online? [**rate_music**]
- How frequently do you buy physical music records? [**buy_records**]
- How frequently do you buy music online? [**buy_online**]
- Do you prefer listening to single tracks, whole albums or either way? [**track_or_CD**]

2.2 Albums listened to in last.fm

The screenshot shows the last.fm website interface. The browser address bar displays 'www.last.fm/user/aleixooo/charts?subtype=albums'. The navigation menu includes 'last.fm', 'Music', 'Radio', 'Events', 'Charts', and 'Community'. A search bar is visible with the text 'Music search'. The main content area is titled 'Charts' and features a dropdown menu set to 'albums' and a row of filters: 'Last 7 days', 'Last 3 months', 'Last 6 months', 'Last 12 months', and 'Overall'. A table lists the top 10 albums with their corresponding number of global plays, represented by blue horizontal bars. The 5th album, 'Nick Cave and the Bad Seeds - Murder Ballads', is highlighted with a red box. Two yellow callout boxes provide additional context: one on the left explains the time periods available, and one on the right explains the data points for each album.

Albums they listened to during last:
7days, 3months, 6months, year,
overall

For each album in the list we
obtained: # user plays (in each
period), # of global listeners and # of
global plays

Rank	Album	Global Plays
1	Radiohead – The King of Limbs	72
2	Nick Cave & The Bad Seeds – The Boatman's Call	59
3	Nick Cave & The Bad Seeds – The Best Of (Disc 1)	36
4	Radiohead – Kid A	34
5	Nick Cave and the Bad Seeds – Murder Ballads	30
6	Nick Cave & The Bad Seeds – The Lyre Of Orpheus	24
7	Radiohead – In Rainbows	18
8	Nick Cave and the Bad Seeds – The Boatman's Call	14
9	Life's Decay – Eklaasera	12
10	Nick Cave and the Bad Seeds – The Lyre of Orpheus	9

2.4 Survey Screenshots

- Requirements: 18 y.o., scrobblings > 5000

www.hsli.pitt.edu/mobil... Gmail - Last week to finis... Survey for User Study

estions/index.php?lastfm_subject=alidacon&survey_id=1&ver_hash=e966e24eacf245196eedf74353386cdb

ps TID Pitt real st Entert CHile Craiglist Red Soc AH trips Seminars Conferences research Customize Li

Survey about music taste - Telefonica I+D

Part I: 11 questions about demographics, music experience and consumption.

A) User Consent

Before starting the survey, please tell us if you accept the [terms and conditions of this study](#).

I have read the terms and conditions of this study and I accept voluntarily to participate on it. I also acknowledge that I am 18 years old or older.

B) Demographics

1. Gender

2. Age
 Your age must be a number between 18 and 99.

3. Current Country

C) Media Consumption behavior

1. How many hours per week do you use the internet?

2. How many hours per week do you listen to music?


3. How many concerts do you usually attend per year?

4. How frequently do you read specialized blogs or

cs.cmu.edu/~aarli/... www.hsli.pitt.edu/mobil... Gmail - Last week to finis... Survey for User Study Party II: Items to rate

et/MusicSurvey/questions/ratingItem.php

Blogs misc web apps TID Pitt real st Entert CHile Craiglist Red Soc AH trips Seminars Conferences research Custor



43 of 100

Gold

Artist/Band | The Cranberries

Tracks (up to 12)

1. Dreams
2. Salvation
3. Sunday
4. Free To Decide
5. Pretty
6. When You're Gone
7. How
8. Hollywood
9. Cordell
10. Not Sorry
11. Animal Instinct
12. Linger

Need more info? | [Click here for additional information about this album](#)

How would you rate this album?

2.5 Survey Part I – Results (1/2)

151 users started	127 users completed	114 after filtering outliers
Users from 23 countries	25 users from Spain, 15 from USA, 16 from UK	
80% spent 20 or more hours per week on the internet	50% listened to over 20 hours per week of music	
9% did not attend music concerts	30% went to 11 or more concerts per year	

2.5 Survey Part I – Results (2/2)

35% read sometimes music magazines or blogs

20% read them every week

50% never or seldom rated music online

45% bought between 1 and 10 records per year

18% did not buy any

35% never bought music online

8% bought music online at least once per month

14% preferred listening to single tracks

45% preferred listening to albums, ~40% either way

Ok, ok, now...



**What did you get after
analyzing the results?**

3.6 General Analysis - Assumptions

- We “expect” strong positive correlation between ratings and implicit feedback
- We “expect” some level of positive correlation between ratings and recentness
- We (don’t?) expect a significant relation between ratings and global popularity
- **On demographic data:** Just listening to track or album shows a significant effect (using ANOVA) $F(2, 62) = 3.949, p = 0.024$

So far so good



Are you able to predict ratings with those variables?

M2:
implicit
feedback &
recentness

4 Regression Analysis

M1: implicit feedback

M4:
Interaction of
implicit
feedback &
recentness

- Model 1: $r_{iu} = \beta_0 + \beta_1 \cdot ifiu$
- Model 2: $r_{iu} = \beta_0 + \beta_1 \cdot ifiu + \beta_2 \cdot reiu$
- Model 3: $r_{iu} = \beta_0 + \beta_1 \cdot ifiu + \beta_2 \cdot reiu + \beta_3 \cdot gpi$
- Model 4: $r_{iu} = \beta_0 + \beta_1 \cdot ifiu + \beta_2 \cdot reiu + \beta_3 \cdot ifiu \cdot reiu$

M3: implicit
feedback,
recentness,
global
popularity

Model	R^2	F -value	p -value	β_0	β_1	β_2	β_3
1	0.125	$F(1, 10120) = 1146$	$< 2.2 \cdot 10^{-16}$	2.726	0.499	-	-
2	0.1358	$F(2, 10019) = 794.8$	$< 2.2 \cdot 10^{-16}$	2.491	0.484	0.133	-
3	0.1362	$F(3, 10018) = 531.8$	$< 2.2 \cdot 10^{-16}$	2.435	0.486	0.134	0.0285
4	0.1368	$F(3, 10018) = 534.7$	$< 2.2 \cdot 10^{-16}$	2.677	0.379	0.038	0.053

Table 1. Regression Results. R^2 , F -value, and p -value for the 5 models.

- Including Recentness increases R^2 in more than 10% [1 -> 2]
- Including GP increases R^2 , not much compared to RE + IF [1 -> 3]
- Not Including GP, but including interaction between IF and RE improves the variance of the DV explained by the regression model. [2 -> 4]

4.1 Regression Analysis

Model	RMSE1	RMSE2
User average	1.5308	1.1051
M1: Implicit feedback	1.4206	1.0402
M2: Implicit feedback + recentness	1.4136	1.034
M3: Implicit feedback + recentness + global popularity	1.4130	1.0338
M4: Interaction of Implicit feedback * recentness	1.4127	1.0332

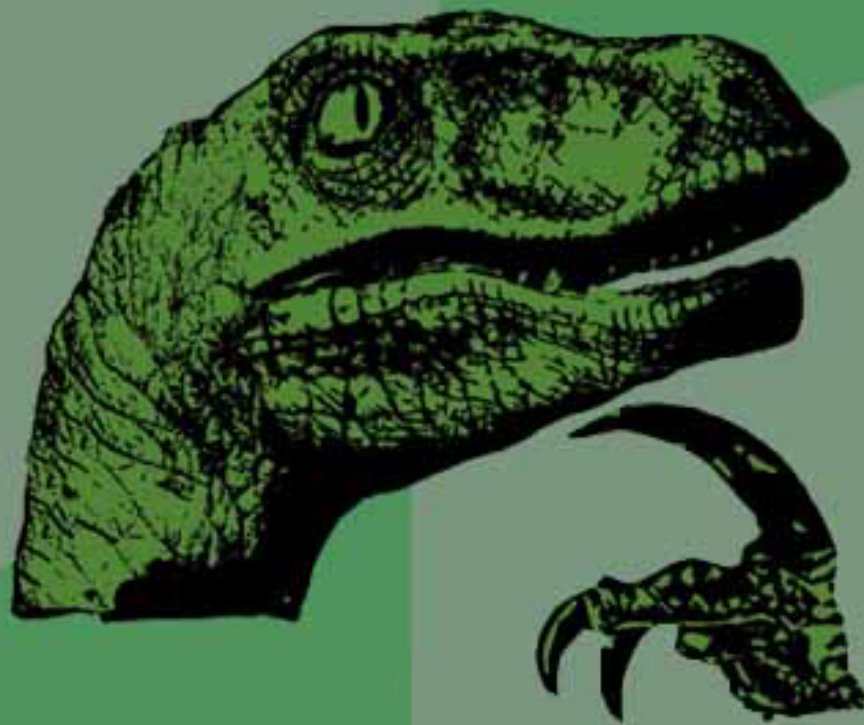
- We tested conclusions of regression analysis by predicting the score, checking RMSE in 10-fold cross validation.
- **Results of regression analysis are supported.**

4.2 Regression Analysis – Track or Album

Model	Tracks	Tracks/ Albums	Albums
User average	1.1833	1.1501	1.1306
M1: Implicit feedback	1.0417	1.0579	1.0257
M2: Implicit feedback + recentness	1.0383	1.0512	1.0169
M3: Implicit feedback + recentness + global popularity	1.0386	1.0507	1.0159
M4: Interaction of Implicit feedback * recentness	1.0384	1.049	1.0159

- Including this variable that seemed to have an effect in the general analysis, helped to improve accuracy of the model

Is this the end?



5. Ongoing Work

- Incorporate the nested nature of the ratings: they are not independent, so our regression will consider the user as a random factor to group ratings by user (mixed model)
- Ratings are not continuous by nature: we will use logistic regression (back up slide for further discussion)
- Using raw data (with some transformations) as the value of some predictors (implicit feedback and global popularity)

6. Conclusions

- Using a linear model, Implicit feedback and recency can help to predict explicit feedback (in the form of ratings)
- Global popularity doesn't show a significant improvement in the prediction task
- Our model can help to relate implicit and explicit feedback, helping to evaluate and compare explicit and implicit recommender systems.

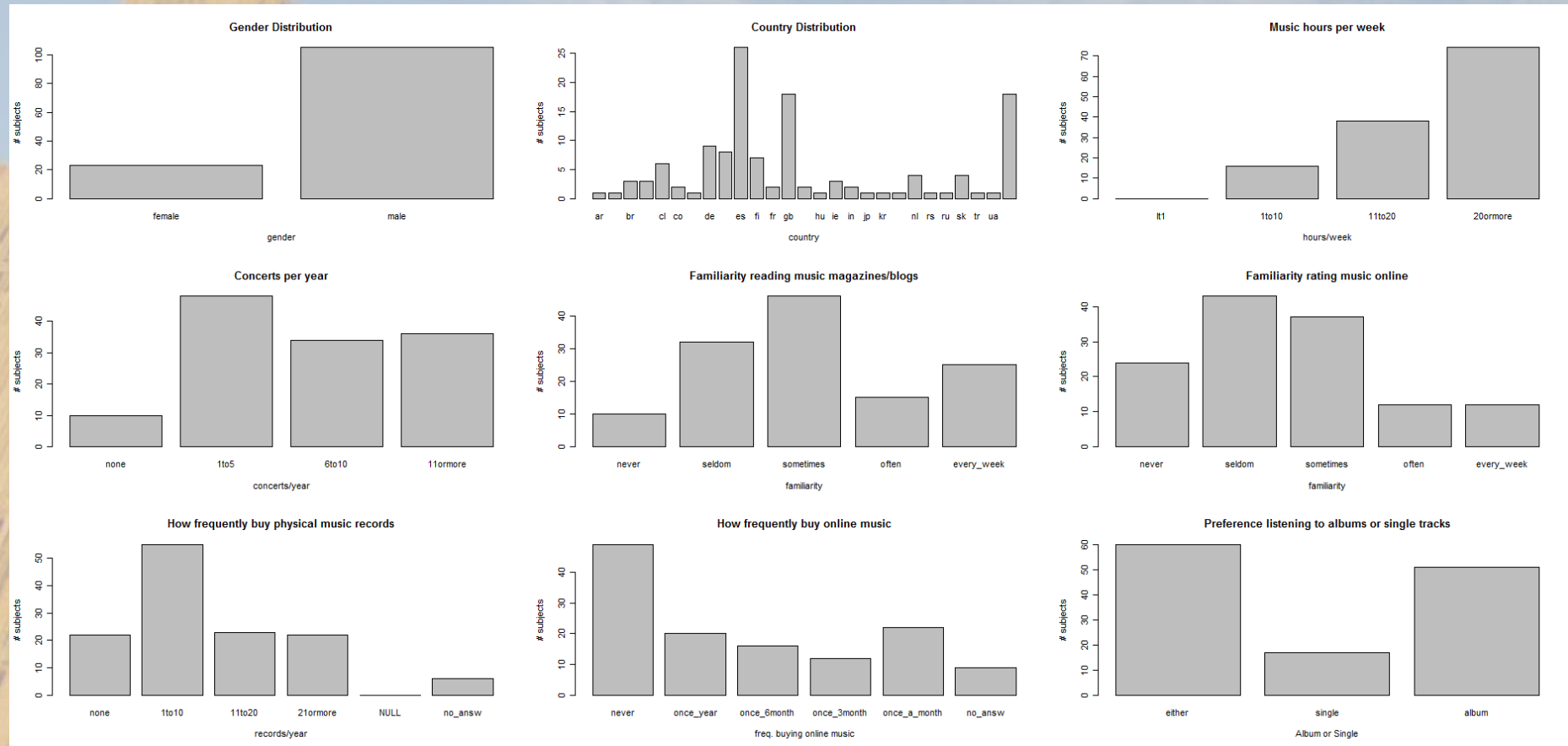
THANKS

- For spending your time listening to this talk 😊
- Questions?

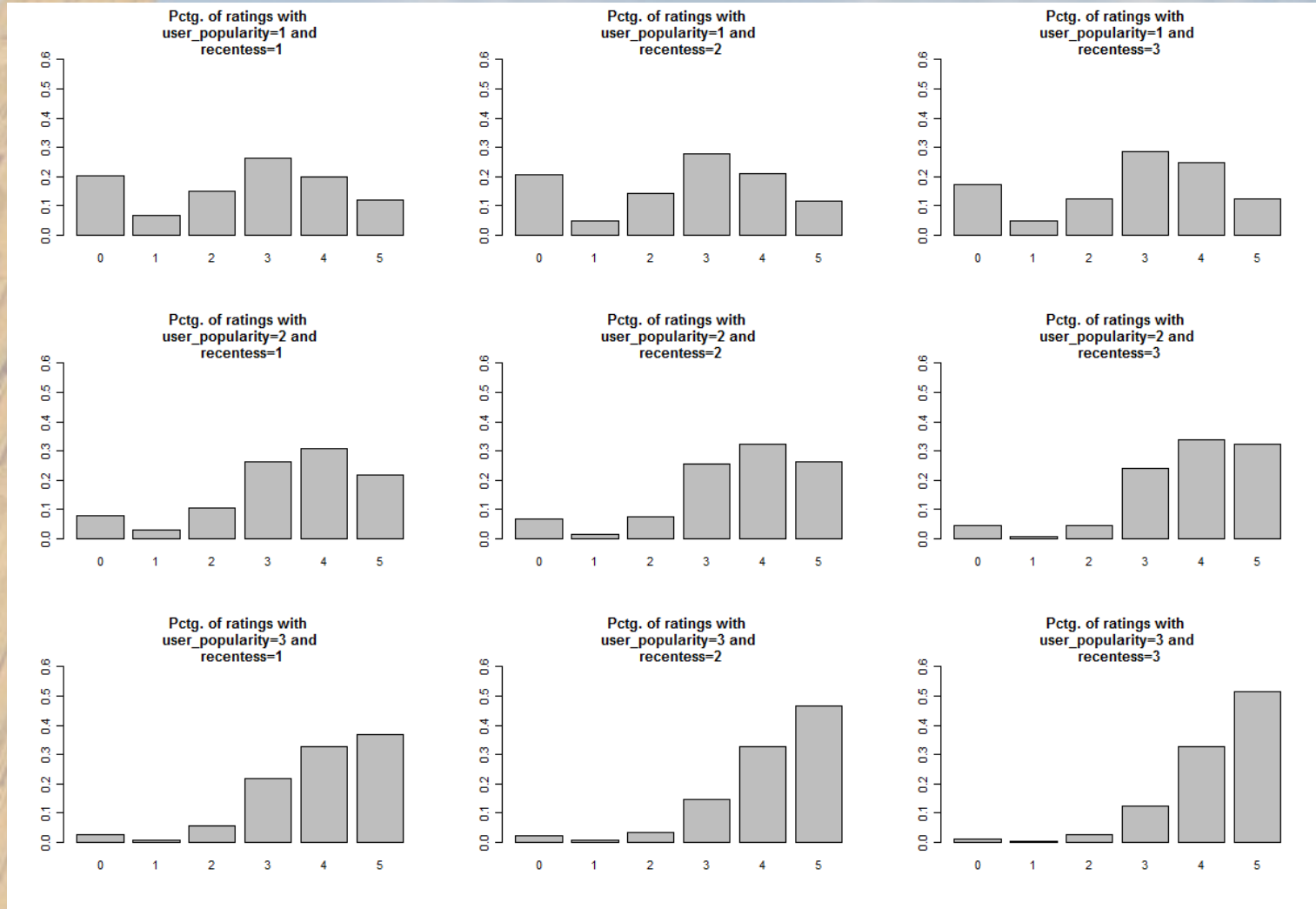
Denis Parra

www.sis.pitt.edu/~dparra/
dap89@pitt.edu

Survey part I results

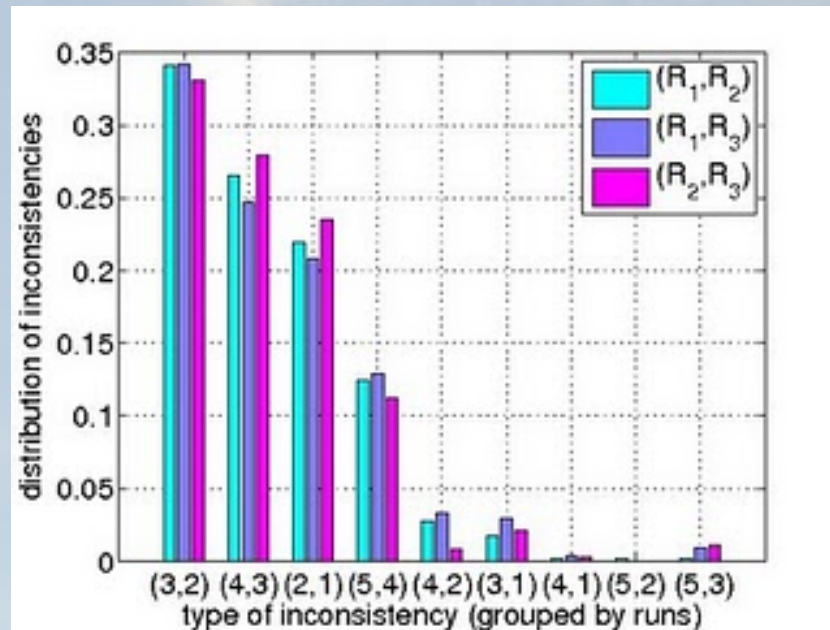


Graphics comparing % of ratings given 2 variables



Discussion: Evaluating with RMSE

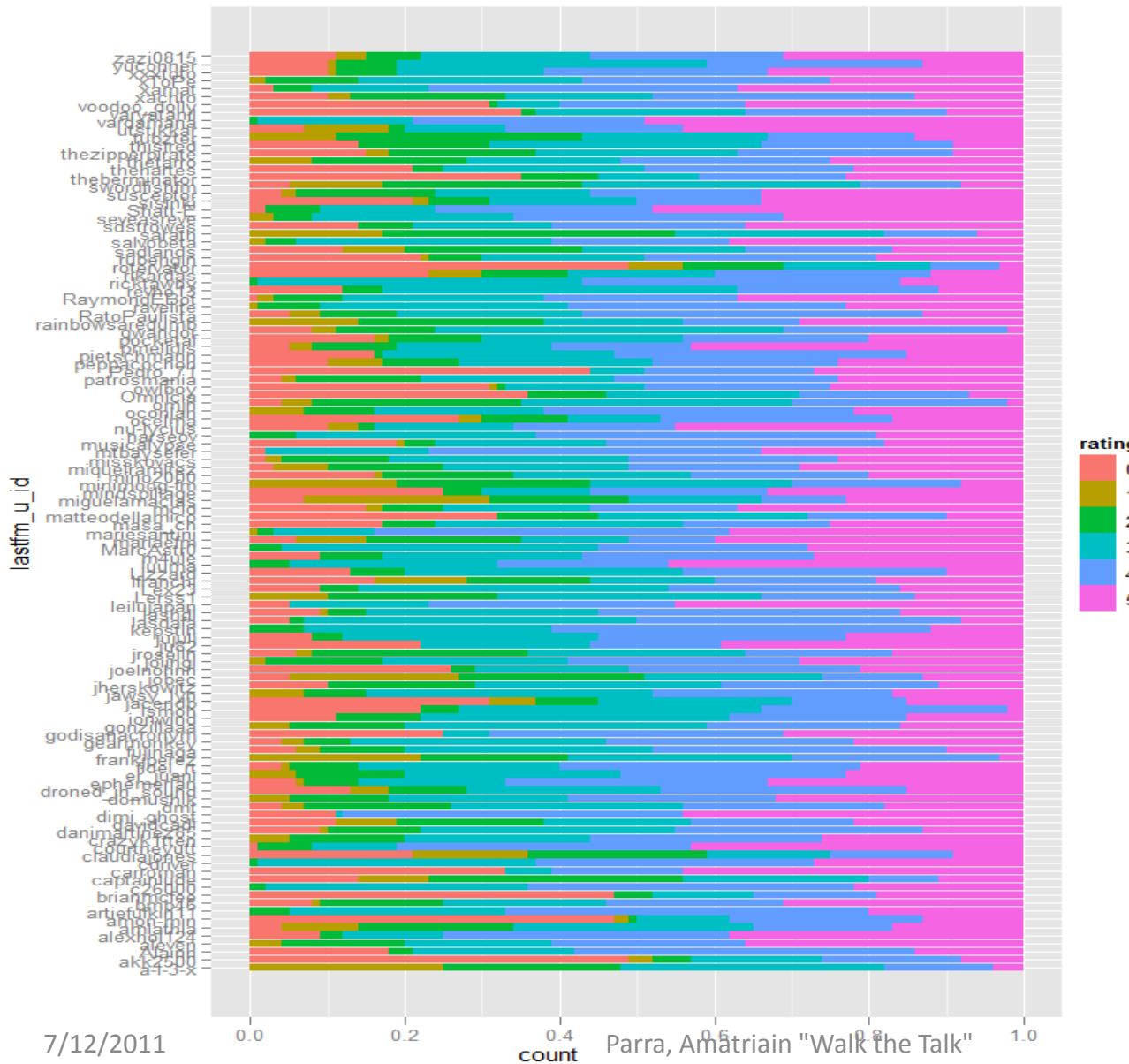
- Discussion after working on this paper: Should we continue using RMSE? IS NOT THE SAME MISPREDICTING a rating 2 by 3 than a rating 4 by 5



Comparison with Study of Duncan Watts

- ... on evolution of artificial markets, where Watts concludes that people is actually influenced by popularity of items
- ... but that are important differences with that study:
 - Watts study was trying to assess music quality, we asked users to avoid judging quality: just if they liked it or not.
 - In Watts study subjects were presented NEW songs that they HAD TO LISTEN TO in order to judge their quality.

3.1 Distribution of ratings



Average rating:

- Considering 0s:

3.206316

- Not considering 0s:

3.611144

3.2 Implicit Feedback

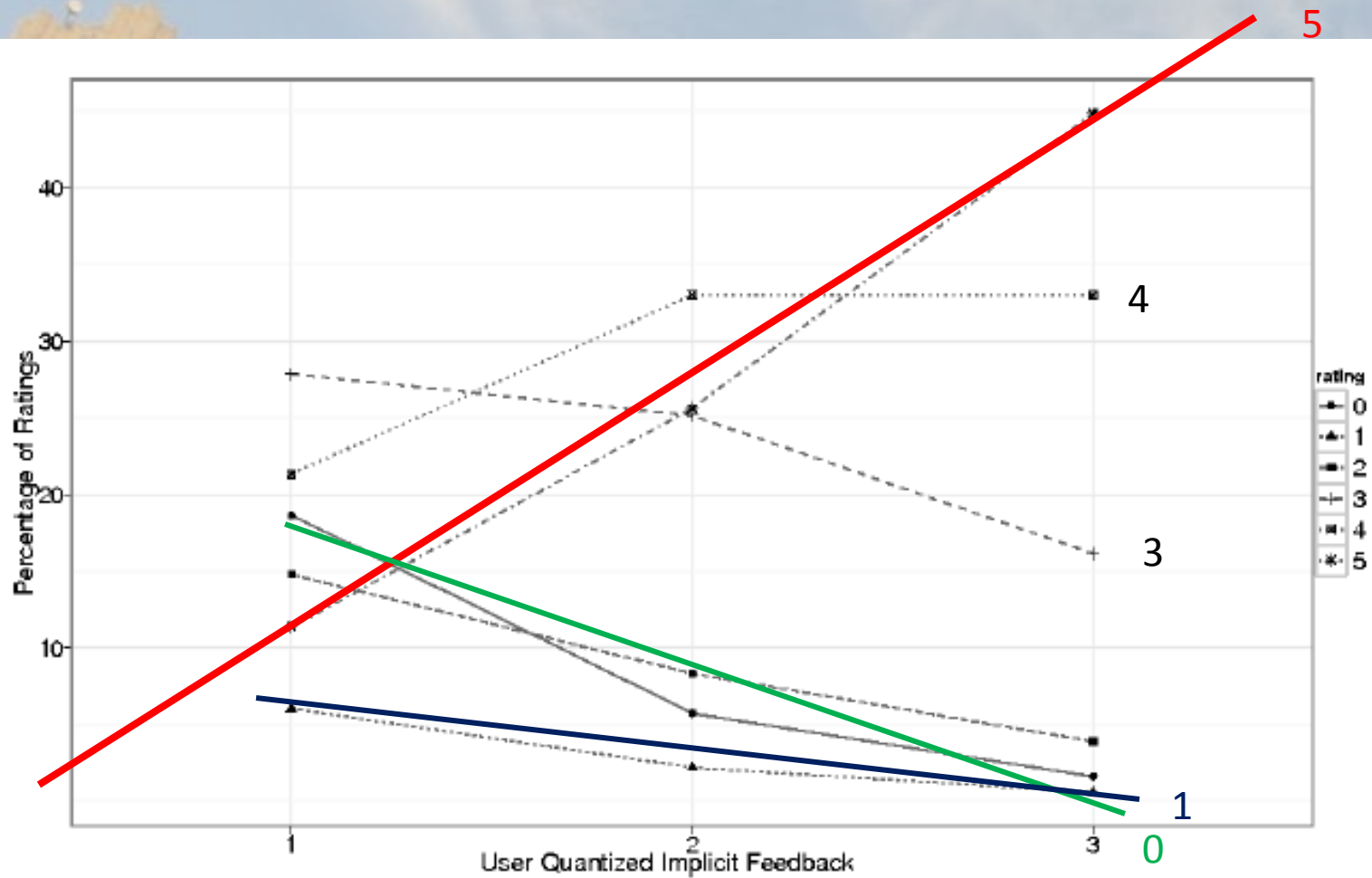


Fig. 3. Distribution of ratings given different values of implicit feedback

3.3 Recentness

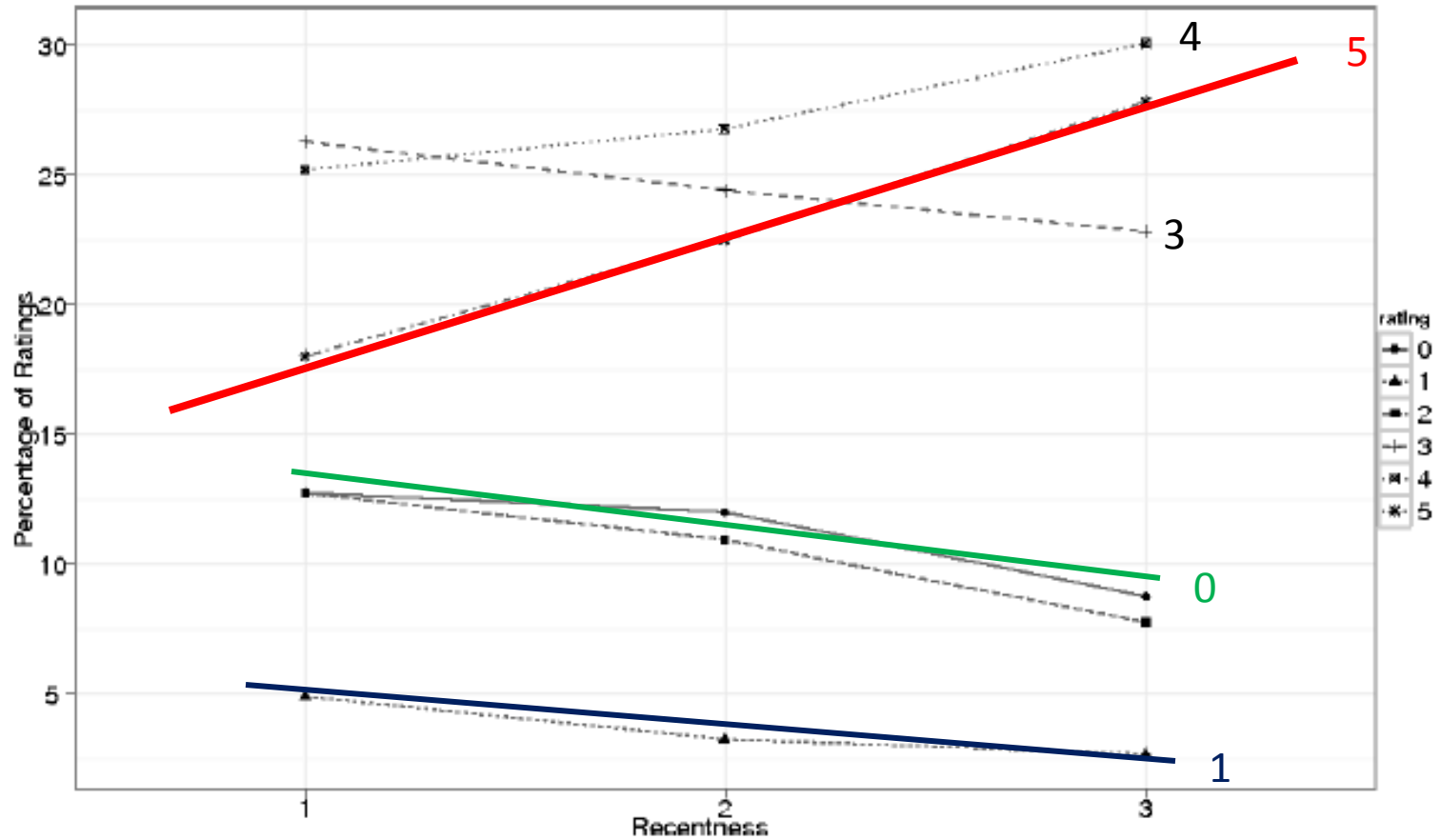


Fig. 4. Distribution of ratings given different values of recentness

3.4 Global Popularity

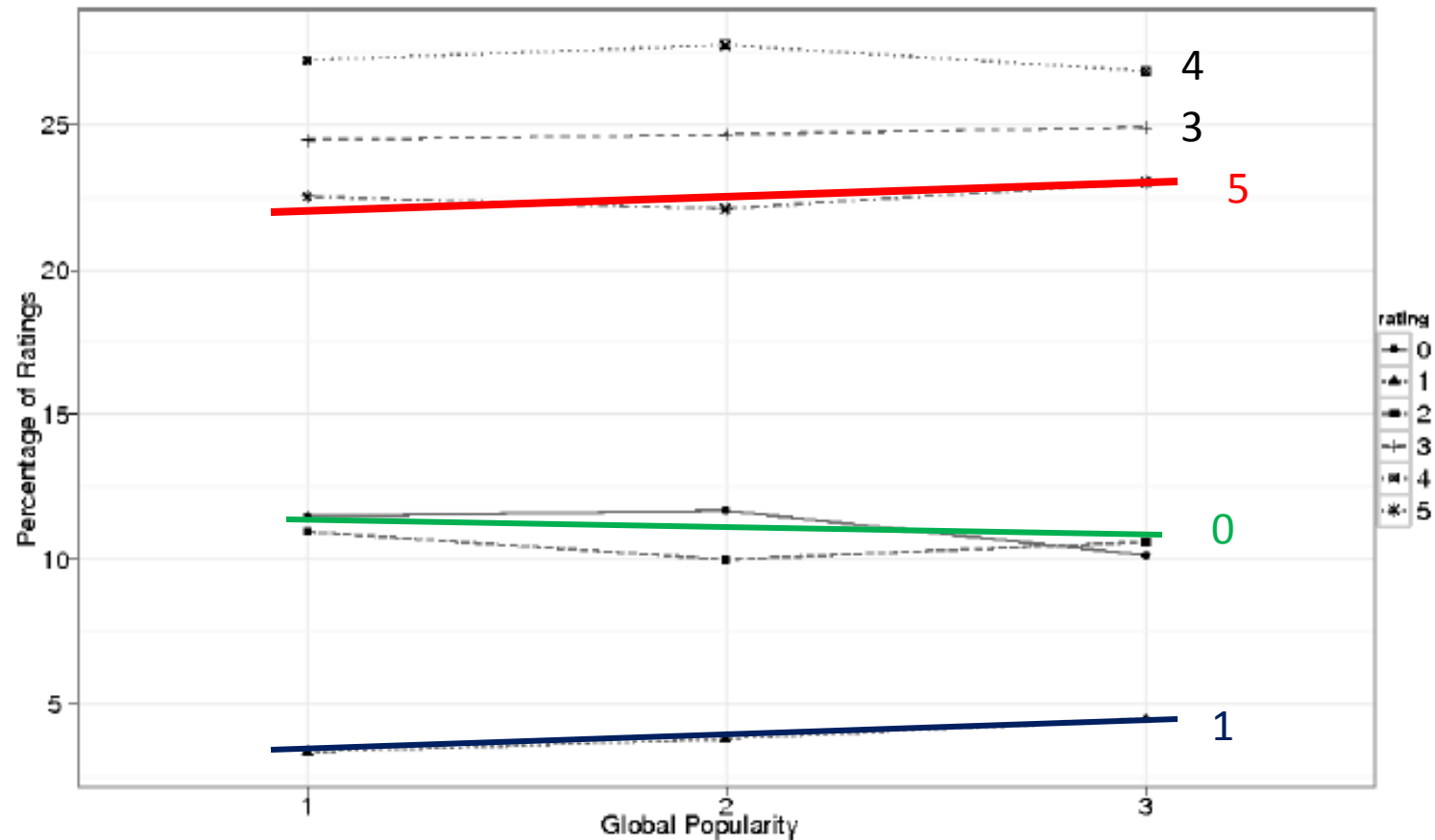


Fig. 5. Distribution of ratings given different values of overall popularity

3.5 Effect of Track or CD

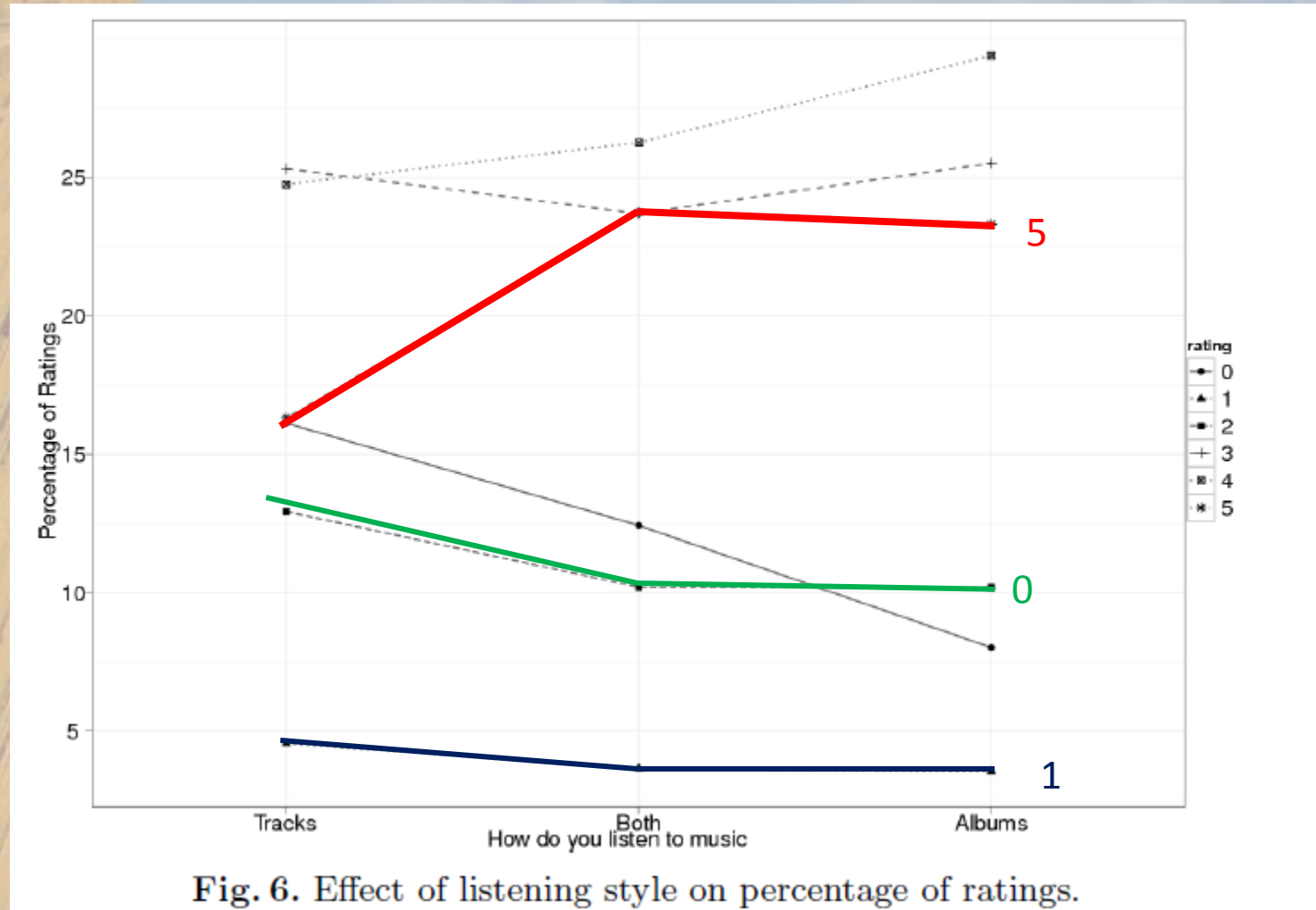


Fig. 6. Effect of listening style on percentage of ratings.

3 General Analysis

- Initial assumption: Rating and IF (# playcount) must be strongly correlated.

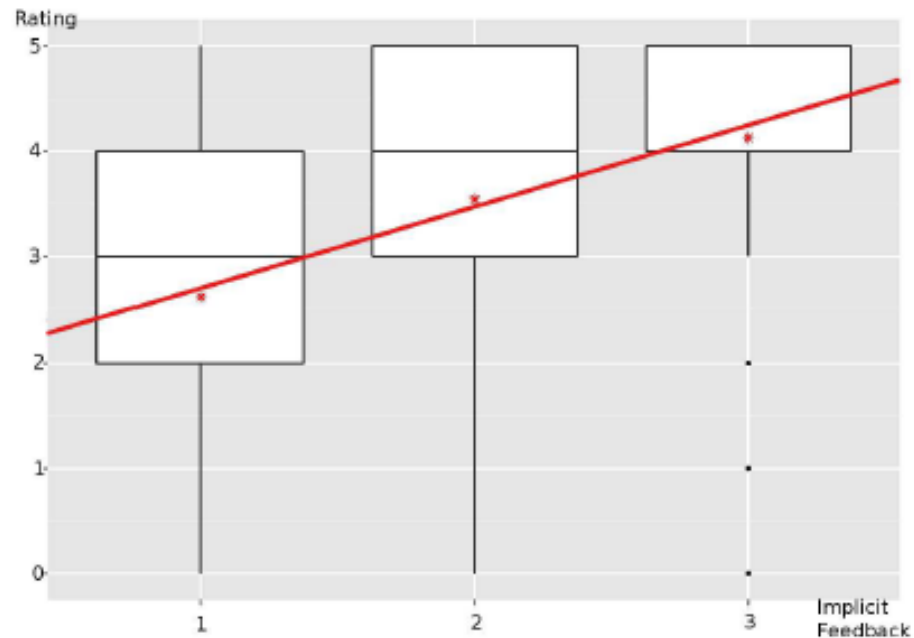


Fig. 2. Relation between implicit feedback and explicit ratings