(How) Does Data-based Music Discovery Work?

Akarapat Charoenpanich, London School of Economics
Aleksi Aaltonen, Warwick Business School

Context and motivation

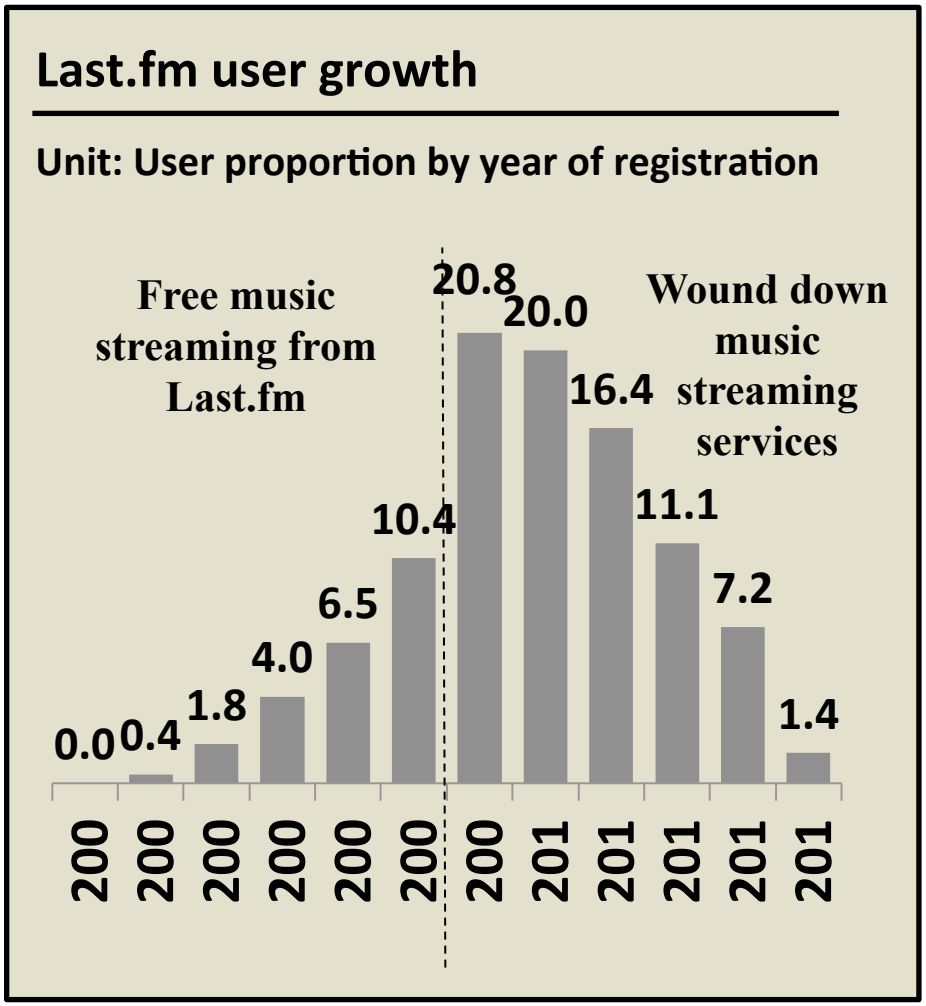
- Music industry has been significantly transformed by digital technology.
  - Digital ecosystem effectively abolish constraints on music distribution
  - Music / product space available to consumers become extremely large
- Key problem: Music discovery.
  - Given an extremely large music / product space, how do people learn what they want to listen to?
  - Exploratory search becomes very important, contrast to known-item search
- Business opportunity
  - New form of music discovery to help people to listen to music that better match their taste

Research question

- **Empirical object**: Last.fm
- **Research question**: Does the new form of music discovery in Last.fm provide value to consumers?
- Amount of music consumption as indicator for value
  - Main dependent variable in our statistical analysis
  - Not direct measure of commercial success
  - But possibly associated with public valuation (Brynjolfsson, Kim and Oh, 2013)
- Able to obtain a representative sample of user
  - Rejection sampling. Select user by their identifiers randomly
  - End up with random sample of around 13K

Briefly about Last.fm

- Platform for music discovery founded in 2002
  - Deploy collaborative filtering recommender system and social media feature
  - Users submit data to Last.fm
  - Last.fm provide personalized music recommendation
- Policy change in 2009
  - Restrict free music streaming services
  - Impact on user growth



Music discovery (and consumption) through Last.fm: 3 crucial factors can be identified

1 Collaborative filtering recommender system

- Popular approach for product recommendation
- Construct similarity network linking music items together utilizing social data submitted by its users
- Recommend products similar to those that users like in the past according to constructed similarity network

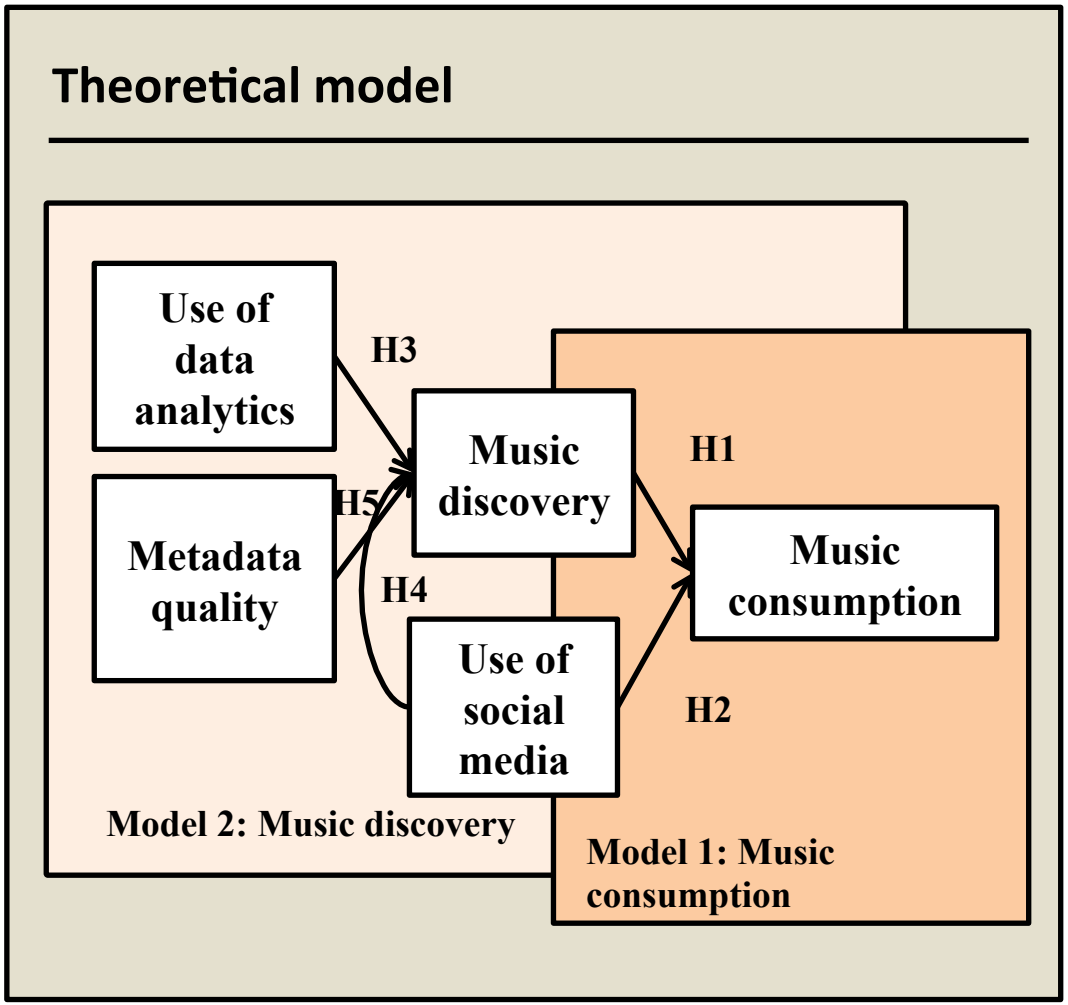
2 Social media features

- Enhance both music discovery and music consumption
- Idiosyncratic links by posting, commenting etc. and these links can be seen as their personal recommendation
- Similarity network + social network = dual network

3 Autocorrection system (metadata infrastructure)

- Help to improve quality of social data as there exist no authority to control for music metadata standard
- Metadata need to be associated with the right content . Identifiers become very important. But there are more than 100 ways to spell Guns N' Roses
- Last.fm implement autocorrection system in January 2009 to alleviate the problem

Theoretical model and estimation



- With Last.fm data, estimate the model twice
  - Before 2009
  - 2009 and onward
- Why?
  - To see impact of policy changes on consumer offering in 2009 through intercept terms (**negative**)
  - To assess impact of autocorrection system which has been implemented in 2009 (**positive**)

Data collection: Assembly of dataset

1 **PLAYCOUNT**	- Measure **music consumption** - Count number of listening event during different years
2 **LISTENING CONCENTRATION**	- Measure **success of music discovery** - Successful music discovery more concentrated listening profile at level of artist. Apply normalized **Herfindahl Index (HI)**
3 **NUMBER OF FRIENDS**	- Measure **social media engagement** - Compute a proxy variable to track number of friends through time
4 **NUMBER OF CORRECTION**	- Measure **metadata quality** - With auto-correction mapping applied to listening event data of user, compute a proxy variable to trace number of correction through time
5 **PAST LISTENING SIMILARITY**	- Measure **use of data analytics** - Why? Crux of recommender system => recommend products similar to those users like in the past - With similarity network, compute a proxy variable to trace past listening similarity of user through time

Estimation result

	Music discovery model		
	Dependent variable: Success of music discovery		
		Beta	Sig.
Before 2009	(Constant)	3.66	0.00
	Use of data analytics	0.40	0.00
	Metadata quality	0.10	0.00
	Social media engagement	0.08	0.06
2009 and onward	(Constant)	2.25	0.00
	Use of data analytics	0.81	0.00
	Metadata quality	0.33	0.00
	Social media engagement	0.13	0.01

	Music consumption model		
	Dependent variable: Music consumption		
		Beta	Sig.
Before 2009	(constant)	3.64	0.00
	Success of music discovery	0.38	0.00
	Social media engagement	1.05	0.00
2009 and onward	(constant)	2.38	0.00
	Success of music discovery	0.47	0.00
	Social media engagement	1.36	0.00

Estimation result: direct, indirect and total effect on music consumption

The direct and indirect impact of data analytics, data quality, social media engagement and policy changes upon music consumption

	Before 2009			2009 and after		
	Direct	**Indirect**	**Total**	**Direct**	**Indirect**	**Total**
Increase in use of data analytics by 1%	n/a	+0.15%	+0.15%	n/a	+0.30%	+0.30%
Increase in metadata quality by 1%	n/a	+0.04%	+0.04%	n/a	+0.12%	+0.12%
Increase in use of social media by 1%	+1.05%	+0.03%	+1.08%	+1.36%	+0.05%	+1.41%
Changes to consumer offerings	n/a	n/a	n/a	-71.7%	-35.2%	-81.3%

Although new form of music discovery is valuable its value is relatively modest as compared to that derived from music acquisition

Discussion and conclusion

- Call attention upon assumption underpinning product/service innovations based on big data
  - Can big data resources underlying Last.fm lead to enough competitive advantage?
  - Last.fm made a loss of 2 million GBP last year
  - Relatively small proportion of music enthusiast, who value successful music discovery highly
- Theorize on dynamics of this kind of business venture
  - Changing user dynamics and unexpected secondary effects
  - Initial reduction of user base lower value of social media, and cannot counter that by improving music discovery features
  - Self-propagating decline of user base
  - Network externalities in the realm of big data