Unpredictability of Popularity

Machine Learning with Graphs

Department of Computer Science
University of Massachusetts, Lowell
Spring 2021

Hadi Amiri
hadi@cs.uml.edu

UMASS
LOWELL

Lecture Topics

- Unpredictability of popularity

Unpredictability

- If we replay the history:
  - Do you think the set of most popular items will be the same as the current set of popular items?

Unpredictability

- If we replay the history:
  - Do you think the set of most popular items will be the same as the current set of popular items?
  - Do we observe power law?

Unpredictability

- If we replay the history:
  - Do you think the set of most popular items will be the same as the current set of popular items?
    - Random effects early in the process play a role in the future popularity.
  - Do we observe power law?
    - Power-law distribution of popularity would probably exist in each replay!

How to properly investigate unpredictability in the contents of RGR?

Unpredictability

- Music download site
  - 48 obscure songs/bands.
  - >14K visitors
    - can participate only once and can't share opinions.
  - Visitors can listen and download songs
  - "download count" for each song is shown to visitors.
    - the number of times it had been downloaded thus far.

Unpredictability

- Music download site
  - Parallel World - two settings:
    1. Visitors upon arrival were being assigned at random to one of eight copies of the site.

    2. Visitors upon arrival were being assigned to a copy of the site in which "download counts" info was removed.

Unpredictability

- Music download site

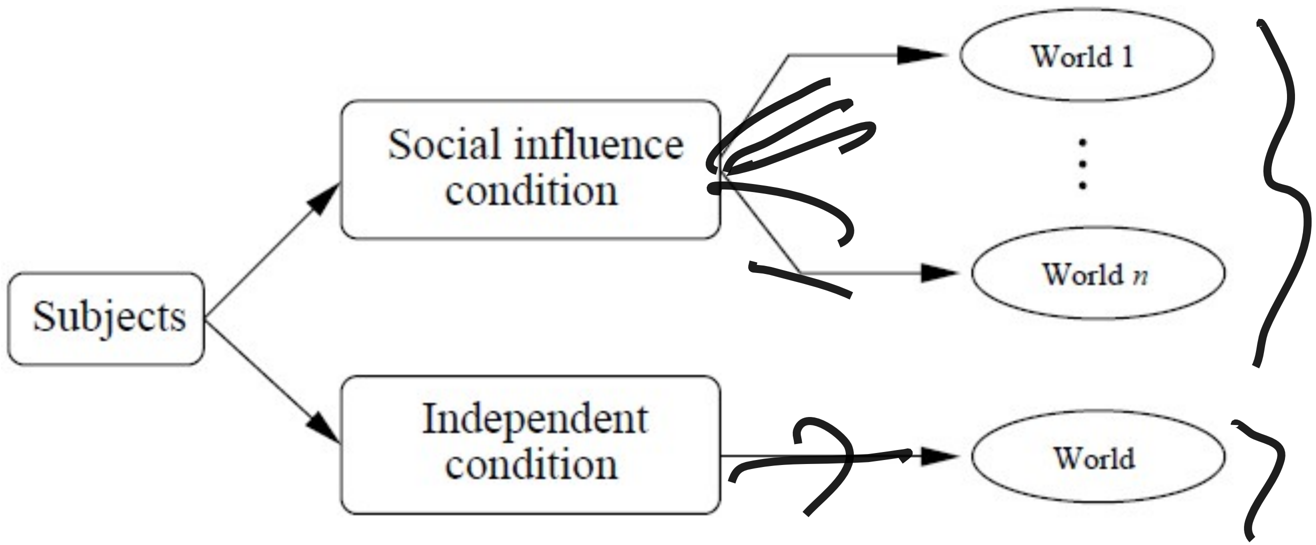


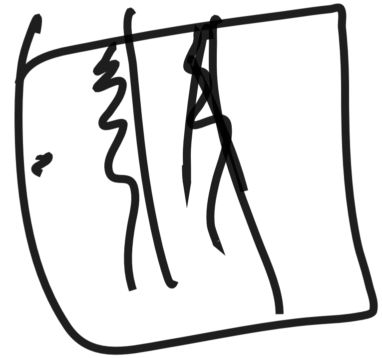
Figure S1: Schematic of the experimental design.

Unpredictability

	# of downloads		# of downloads		# of downloads
HARTSFIELD: "enough is enough"	20	GO MOREDCAI: "it does what its told"	12	UNDO: "while the world passes"	24
DEEP ENOUGH TO DIE: "for the sky"	17	PARKER THEORY: "she said"	47	UP FOR NOTHING: "in sight of"	13
THE THRIFT SYNDICATE: "2003 a tragedy"	20	MISS OCTOBER: "pink agression"	27	SILVERFOX: "gnaw"	17
THE BROKEN PROMISE: "the end in friend"	19	POST BREAK TRAGEDY: "florence"	14	STRANGER: "one drop"	10
THIS NEW DAWN: "the belief above the answer"	12	FORTHFADING: "fear"	24	FAR FROM KNOWN: "route 9"	18
NOONER AT NINE: "walk away"	6	THE CALEFACTION: "trapped in an orange peel"	20	STUNT MONKEY: "inside out"	46
MORAL HAZARD: "waste of my life"	8	52METRO: "lockdown"	17	DANTE: "lifes mystery"	14
NOT FOR SCHOLARS: "as seasons change"	27	SIMPLY WAITING: "went with the count"	16	FADING THROUGH: "wish me luck"	10
SECRETARY: "keep your eyes on the ballistics"	5	STAR CLIMBER: "tell me"	38	UNKNOWN CITIZENS: "falling over"	34
ART OF KANLY: "seductive intro, melodic breakdown"	10	THE FASTLANE: "til death do us part (i dont)"	31	BY NOVEMBER: "if i could take you"	20
HYDRAULIC SANDWICH: "separation anxiety"	20	A BLINDING SILENCE: "miseries and miracles"	17	DRAWN IN THE SKY: "tap the ride"	12
EMBER SKY: "this upcoming winter"	25	SUM RANA: "the bolshevik boogie"	15	SELSIUS: "stars of the city"	22
SALUTE THE DAWN: "i am error"	13	CAPE RENEWAL: "baseball warlock v1"	12	SIBRIAN: "eye patch"	14
RYAN ESSMAKER: "detour_(be still)"	14	UP FALLS DOWN: "a brighter burning star"	11	EVAN GOLD: "robert downey jr"	10
BEERBONG: "father to son"	12	SUMMERSWASTED: "a plan behind destruction"	17	BENEFIT OF A DOUBT: "run away"	38
HALL OF FAME: "best mistakes"	19	SILENT FILM: "all i have to say"	61	SHIPWRECK UNION: "out of the woods"	16

Subjects could participate only once and could not share opinions.

Unpredictability



Experiment 1

- Social Influence:
  - Each visitor was given information only about the behavior of others in its copy of the site!
    - Opportunity to contribute to RGR dynamics!
    - **Songs presented in grid & were not ordered by download counts!**
  - The parallel copies started out identically
    - same songs, download counts for all songs set to zero.
- Independent:
  - No direct contribution to RGR dynamics!
  - **Songs presented in grid & in random order.**

Source: Experimental study of inequality and unpredictability in an artificial cultural market. Salganik et. al. science 2006.

Unpredictability

Experiment 2

- Social Influence:
  - Each visitor was given information only about the behavior of others in its copy of the site!
    - Opportunity to contribute to RGR dynamics!
    - **Songs presented in one column & in descending order of download counts!**
  - The parallel copies started out identically
    - same songs, download counts for all songs set to zero.
- Independent:
  - No direct contribution to RGR dynamics!
  - **Songs presented in one column & random order.**

Unpredictability

Music Lab – Song Selection - Mozilla Firefox

File Edit View Go Bookmarks Tools Help

http://www.musiclab.columbia.edu/me/songs

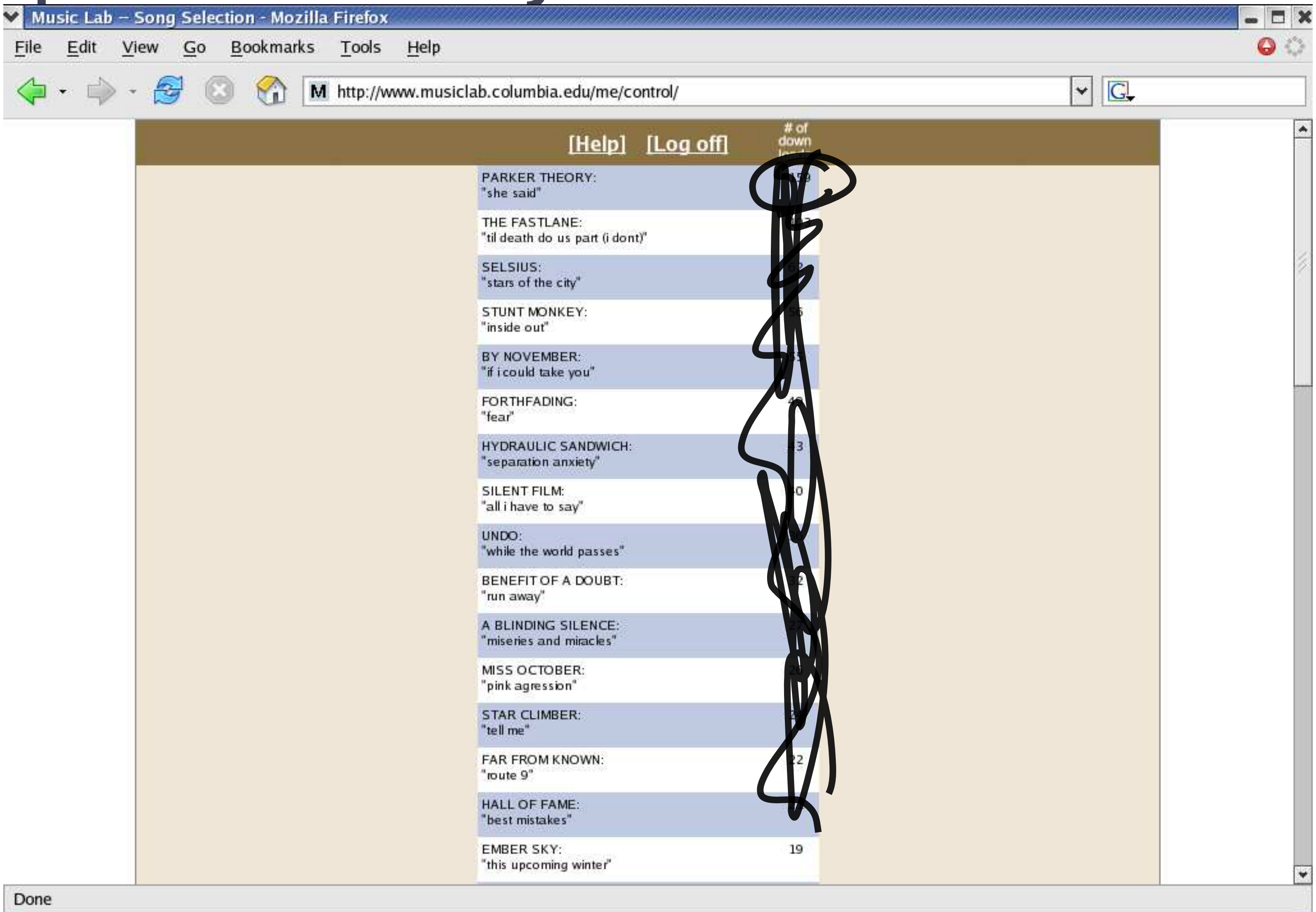
[Help] [Log off]

	# of down loads		# of down loads		# of down loads
HARTSFIELD: "enough is enough"	20	GO MOREDCAI: "it does what its told"	12	UNDO: "while the world passes"	24
DEEP ENOUGH TO DIE: "for the sky"	17	PARKER THEORY: "she said"	47	UP FOR NOTHING: "in sight of"	13
THE THRIFT SYNDICATE: "2003 a tragedy"	20	MISS OCTOBER: "pink agression"	27	SILVERFOX: "gnaw"	17
THE BROKEN PROMISE: "the end in friend"	19	POST BREAK TRAGEDY: "florence"	14	STRANGER: "one drop"	10
THIS NEW DAWN: "the belief above the answer"	12	FORTHFADING: "fear"	24	FAR FROM KNOWN: "route 9"	18
NOONER AT NINE: "walk away"	6	THE CALEFACTION: "trapped in an orange peel"	20	STUNT MONKEY: "inside out"	46
MORAL HAZARD: "waste of my life"	8	52METRO: "lockdown"	17	DANTE: "lifes mystery"	14
NOT FOR SCHOLARS: "as seasons change"	27	SIMPLY WAITING: "went with the count"	16	FADING THROUGH: "wish me luck"	10
SECRETARY: "keep your eyes on the ballistics"	5	STAR CLIMBER: "tell me"	38	UNKNOWN CITIZENS: "falling over"	34
ART OF KANLY: "seductive intro, melodic breakdown"	10	THE FASTLANE: "til death do us part (i dont)"	31	BY NOVEMBER: "if i could take you"	20
HYDRAULIC SANDWICH: "separation anxiety"	20	A BLINDING SILENCE: "miseries and miracles"	17	DRAWN IN THE SKY: "tap the ride"	12
EMBER SKY: "this upcoming winter"	25	SUM RANA: "the bolshevik boogie"	15	SELSIUS: "stars of the city"	22
SALUTE THE DAWN: "i am error"	13	CAPE RENEWAL: "baseball warlock v1"	12	SIBRIAN: "eye patch"	14
RYAN ESSMAKER: "detour_(be still)"	14	UP FALLS DOWN: "a brighter burning star"	11	EVAN GOLD: "robert downey jr"	10
BEERBONG: "father to son"	12	SUMMERSWASTED: "a plan behind destruction"	17	BENEFIT OF A DOUBT: "run away"	38
HALL OF FAME: "best mistakes"	19	SILENT FILM: "all i have to say"	61	SHIPWRECK UNION: "out of the woods"	16

Subjects could participate only once and could not share opinions.

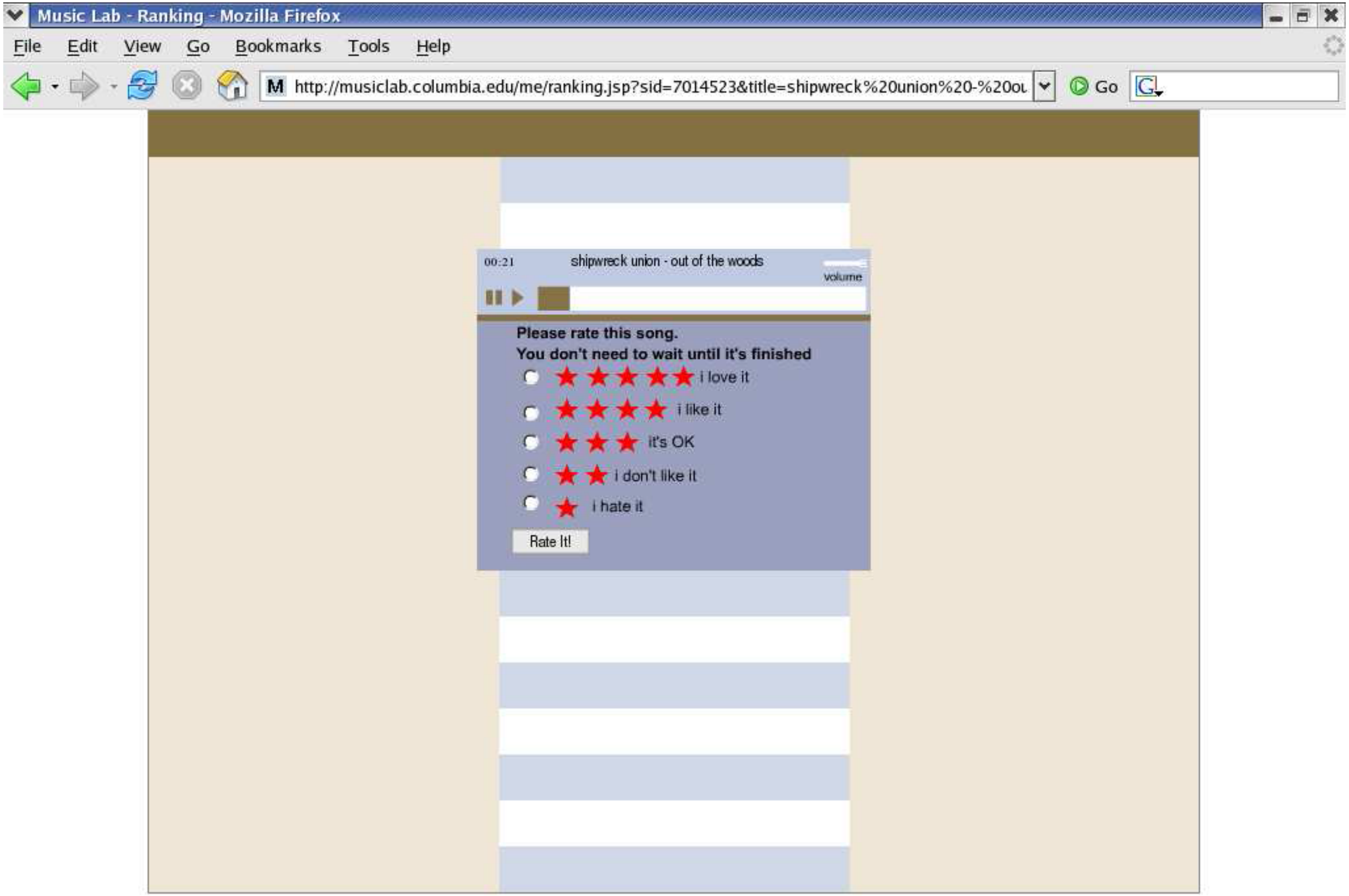
Source: Experimental study of inequality and unpredictability in an artificial cultural market. Salganik et. al. science 2006.

Unpredictability

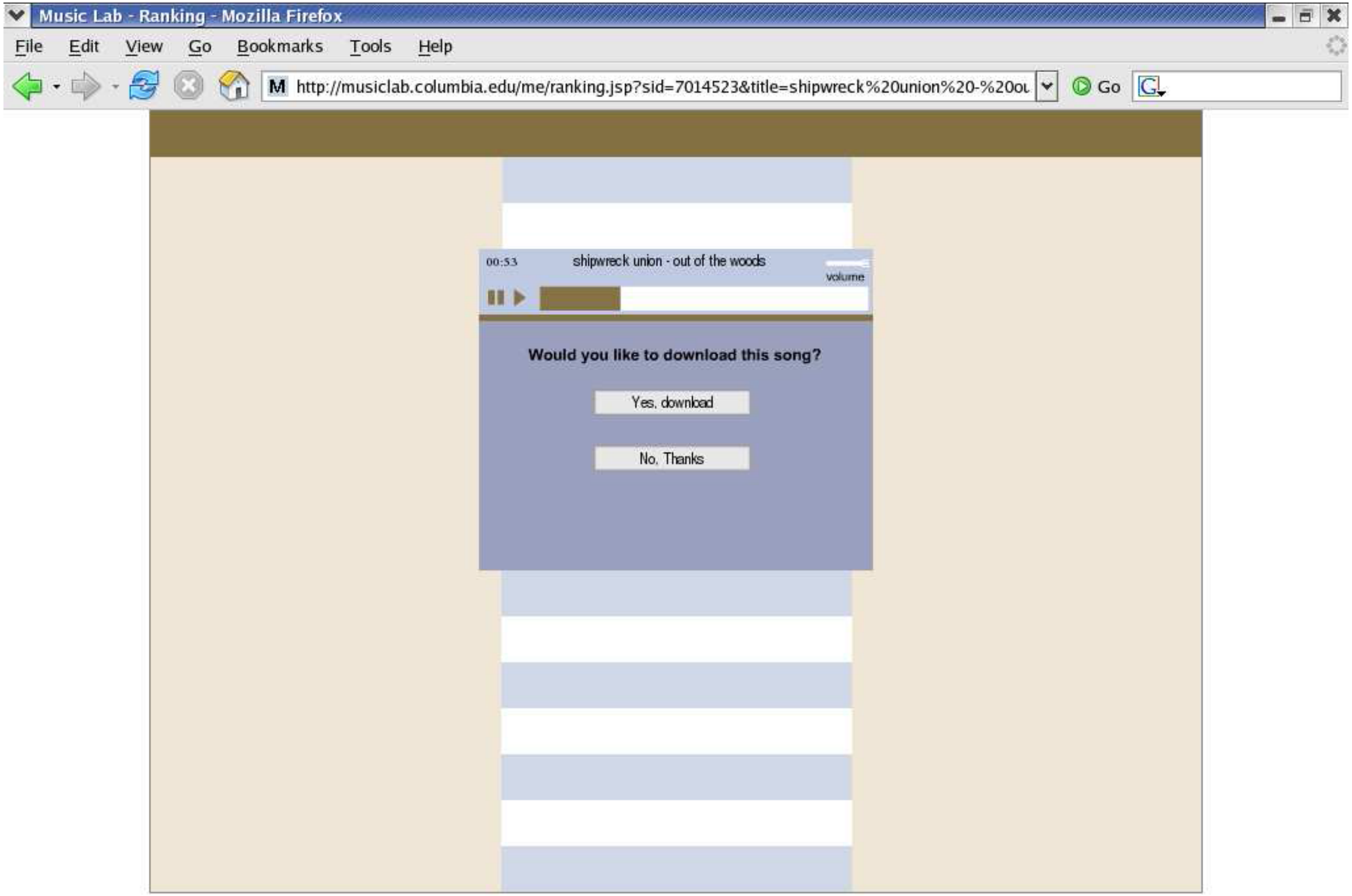


Subjects could participate only once and could not share opinions.

Unpredictability



Unpredictability



Source: Experimental study of inequality and unpredictability in an artificial cultural market. Salganik et. al. science 2006.

Unpredictability

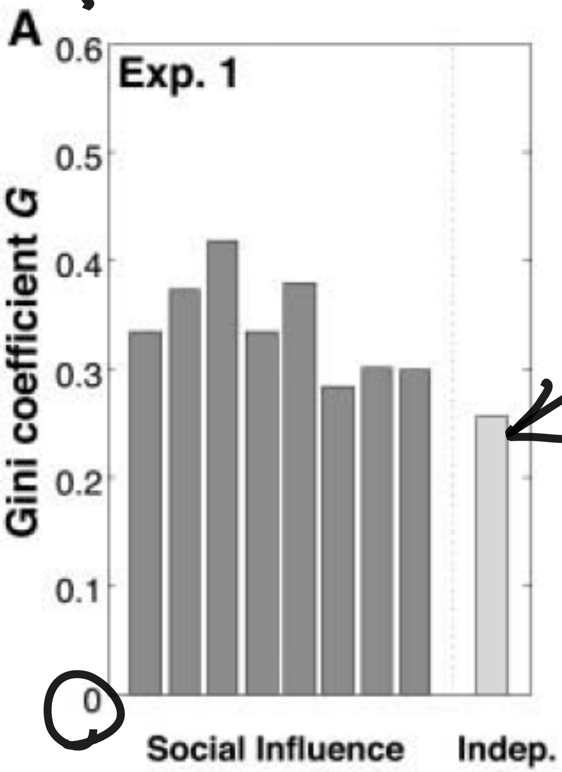


Fig. 1. Inequality of success for social influence (dark bars) and independent (light bars) worlds for (**A**) experiment 1 and (**B**) experiment 2. The success of a song is defined by m_i, its market share of downloads ($m_i = d_i / \sum_{k=1}^{S} d_k$, where d_i is song i's download count and S is the number of songs). Success inequality is defined by the Gini coefficient $G = \sum_{i=1}^{S} \sum_{j=1}^{S} |m_i - m_j| / 2S \sum_{k=1}^{S} m_k$, which represents the average difference in market share for two songs normalized to fall between 0 (complete equality) and 1 (maximum inequality). Differences between independent and social influence conditions are significant ($P < 0.01$) (*18*).

1. The social influence worlds exhibit greater inequality—popular songs are more popular and unpopular songs are less popular—than the independent world.

Unpredictability

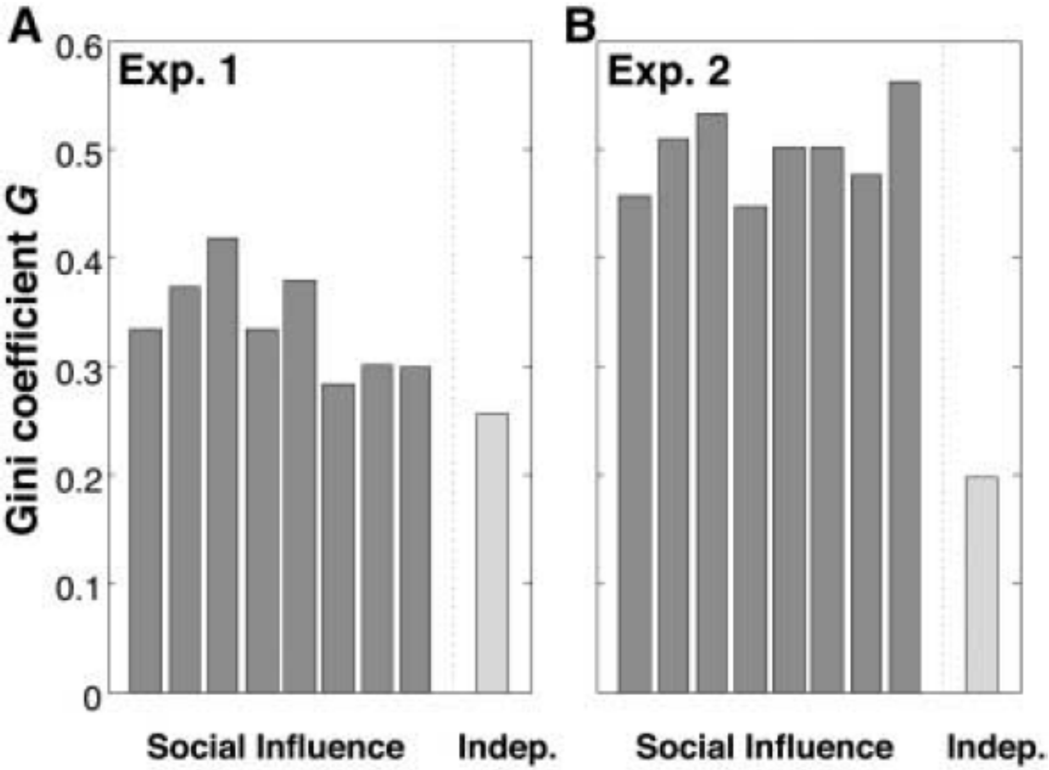
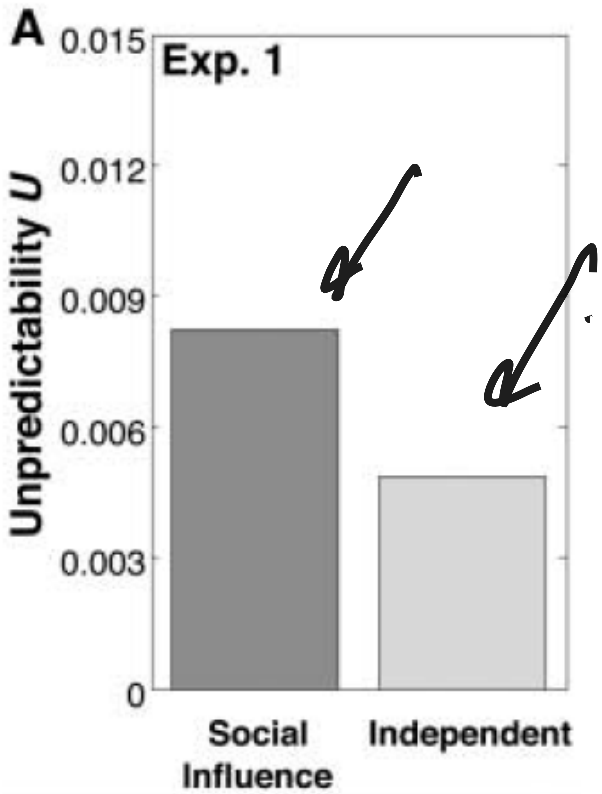


Fig. 1. Inequality of success for social influence (dark bars) and independent (light bars) worlds for (**A**) experiment 1 and (**B**) experiment 2. The success of a song is defined by m_i, its market share of downloads ($m_i = d_i / \sum_{k=1}^{S} d_k$, where d_i is song i's download count and S is the number of songs). Success inequality is defined by the Gini coefficient $G = \sum_{i=1}^{S} \sum_{j=1}^{S} |m_i - m_j| / 2S \sum_{k=1}^{S} m_k$, which represents the average difference in market share for two songs normalized to fall between 0 (complete equality) and 1 (maximum inequality). Differences between independent and social influence conditions are significant ($P < 0.01$) (*18*).

1. The social influence worlds exhibit greater inequality—popular songs are more popular and unpopular songs are less popular—than the independent world.
2. Inequality increased from experiment 1 to experiment 2: not only that social influence contributes to inequality, but as individuals are subject to stronger forms of social influence, the collective outcomes will become increasingly unequal.

Unpredictability

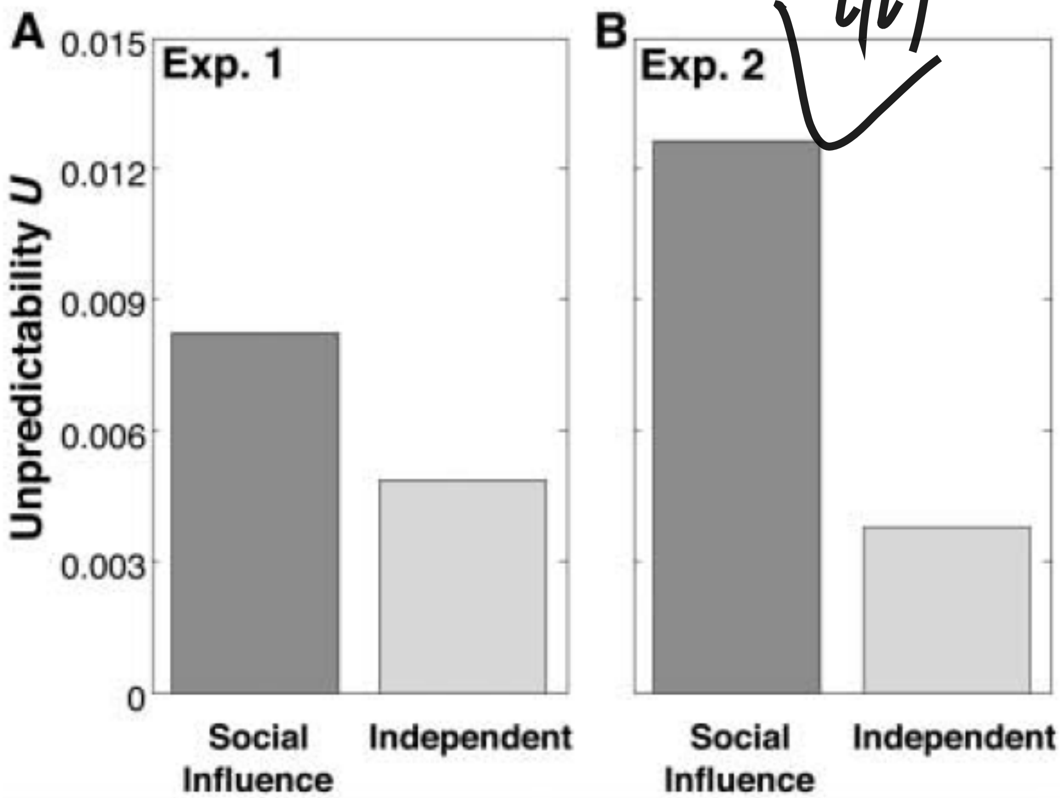
Fig. 2. Unpredictability of success for (**A**) experiment 1 and (**B**) experiment 2. In both experiments, success in the social influence condition was more unpredictable than in the independent condition. Moreover, the stronger social signal in experiment 2 leads to increased unpredictability. The measure of unpredictability u_i for a single song i is defined as the average difference in market share for that song between all pairs of realizations; i.e., $u_i = \sum_{j=1}^{W} \sum_{k=j+1}^{W} |m_{i,j} - m_{i,k}| / \binom{W}{2}$, where $m_{i,j}$ is song i's market share in world j and W is the number of worlds. The overall unpredictability measure $U = \sum_{i=1}^{S} u_i/S$ is then the average of this measure over all S songs. For the independent condition, we randomly split the single world into two subpopulations to obtain differences in market shares, and we then averaged the results over 1000 of these splits. All differences are significant ($P < 0.01$) (*18*).



- Success case of social influence is more unpredictable than success in independent case.

Unpredictability

Fig. 2. Unpredictability of success for (**A**) experiment 1 and (**B**) experiment 2. In both experiments, success in the social influence condition was more unpredictable than in the independent condition. Moreover, the stronger social signal in experiment 2 leads to increased unpredictability. The measure of unpredictability u_i for a single song i is defined as the average difference in market share for that song between all pairs of realizations; i.e., $u_i = \sum_{j=1}^{W} \sum_{k=j+1}^{W} |m_{i,j} - m_{i,k}| / \binom{W}{2}$, where $m_{i,j}$ is song i's market share in world j and W is the number of worlds. The overall unpredictability measure $U = \sum_{i=1}^{S} u_i / S$ is then the average of this measure over all S songs. For the independent condition, we randomly split the single world into two subpopulations to obtain differences in market shares, and we then averaged the results over 1000 of these splits. All differences are significant ($P < 0.01$) (*18*).



- Success case of social influence is more unpredictable than success in independent case.
- Stronger social influence leads to increased unpredictability.

Unpredictability

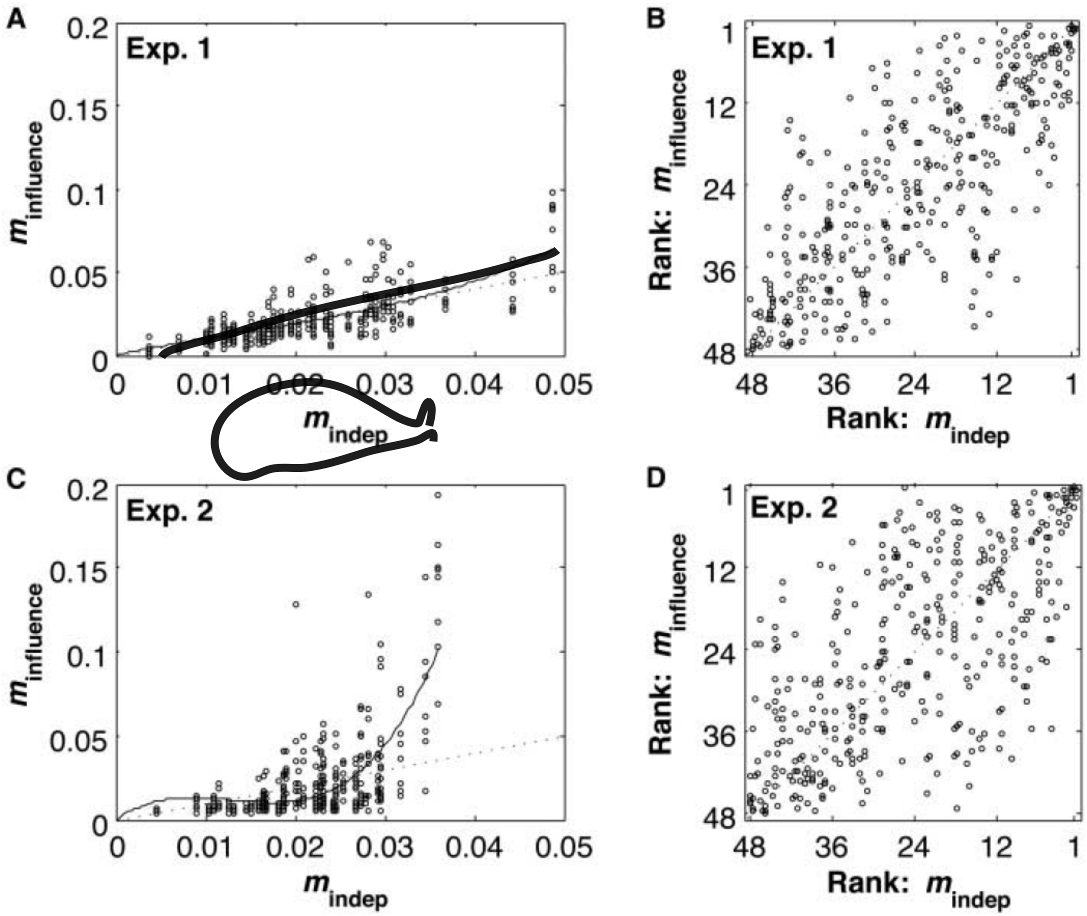


Fig. 3. Relationship between quality and success. (**A**) and (**C**) show the relationship between m_{indep}, the market share in the one independent world (i.e., quality), and $m_{influence}$, the market share in the eight social influence worlds (i.e., success). The dotted lines correspond to quality equaling success. The solid lines are third-degree polynomial fits to the data, which suggest that the relationship between quality and success has greater convexity in experiment 2 than in experiment 1. (**B**) and (**D**) present the corresponding market rank data.

- On average, quality is positively related to success.
- Songs of any given quality can experience a wide range of success.
- The best songs never do very badly, and the worst songs never do extremely well, but almost any other result is possible.
- Unpredictability also varies with quality, the best songs are the most unpredictable, whereas when measured in terms of rank, intermediate songs are the most unpredictable.

Reading

- Experimental study of inequality and unpredictability in an artificial cultural market. Salganik et. al. Science'06.