Lecture 15:
CS 5306 / INFO 5306:
Crowdsourcing and
Human Computation

Anno Regni
A N N Æ
REGINÆ
Magnæ Britanniæ, Franciæ, & Hiberniæ,
DUODECIMO.

At the Parliament Summoned to be Held at *Weſtminſter*, the Twelfth Day of *November*, *Anno Dom.* 1713. In the Twelfth Year of the Reign of our Sovereign Lady *ANNE*, by the Grace of God, of *Great Britain*, *France*, and *Ireland*, Queen, Defender of the Faith, &c.

And by ſeveral Writs of Prorogation Begun and Holden on the Sixteenth Day of *February*, 1713. Being the Firſt Seſſion of this preſent Parliament.

LONDON,
Printed by *John Baskett*, Printer to the Queens moſt Excellent Majeſty, And by the Aſſigns of *Thomas Newcomb*, and *Henry Hills*, deceas'd. 1714.

Anno Regni

A N N Æ

REGINÆ

Magnæ Britanniæ, Franciæ, & Hiberniæ,

DUODECIMO.

At the Parliament Summoned to be Held at *Weſtminſter*, the Twelfth Day of *November*, *Anno Dom.* 1713. In the Twelfth Year of the Reign of our Sovereign Lady *ANNE*, by the Grace of God,

Sixteenth Day of *February*, 1713. Being the Firſt Seſſion of this preſent Parliament.

LONDON,

Printed by *John Baskett*, Printer to the Queens moſt Excellent Majeſty, And by the Aſſigns of *Thomas Newcomb*, and *Henry Hills*, deceas'd. 1714.

1714 British Longitude Act

Anno Regni

A N N Æ

REGINÆ

Magnæ Britanniæ, Franciæ, & Hiberniæ,

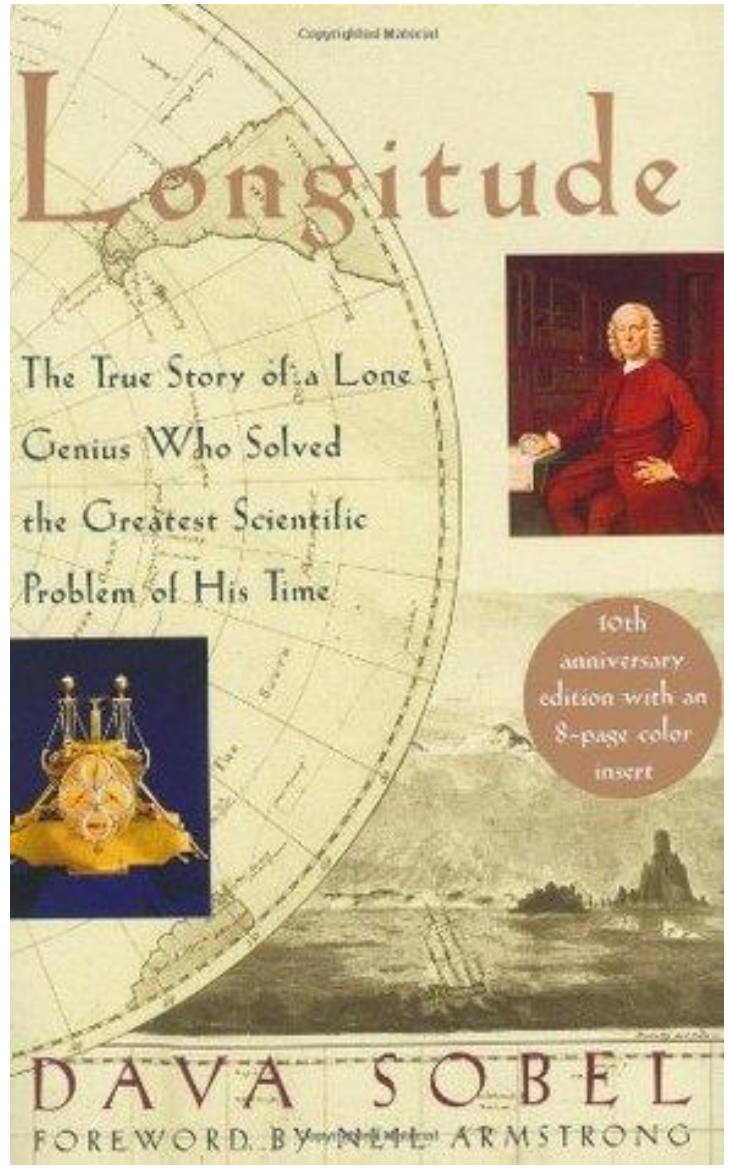
DUODECIMO.

At the Parliament Summoned to be Held at *Weſtminſter*, the Twelfth Day of *November*, *Anno Dom.* 1713. In the Twelfth Year of the Reign of our Sovereign Lady *ANNE*, by the Grace of God,

1714 British Longitude Act
(1598 Prince Philip II Spain)

LONDON,

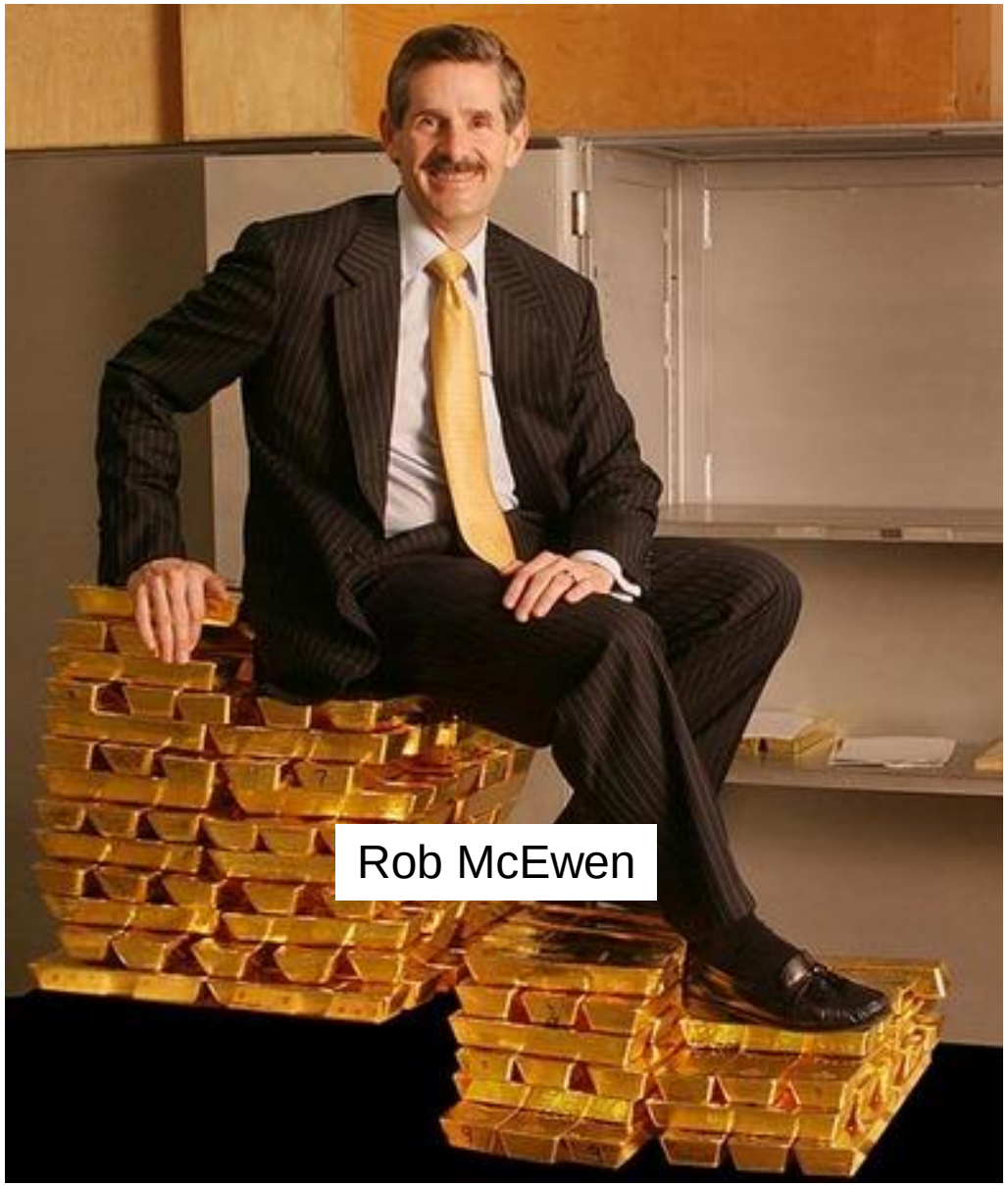
Printed by *John Baskett*, Printer to the Queens moſt Excellent Majeſty, And by the Aſſigns of *Thomas Newcomb*, and *Henry Hills*, deceas'd. 1714.


Copyrighted Material
Longitude
The True Story of a Lone Genius Who Solved the Greatest Scientific Problem of His Time
10th anniversary edition with an 8-page color insert
DAVA SOBEL
FOREWORD BY NEIL ARMSTRONG

Incentive Contests: History

- 1714 British Longitude: measure a ship's longitude while at sea
- 1734 Sweden: A method for stopping the progress of fires
- 1775 France Alkali Prize: Produce alkali from sea salt
- 1795 Napolean: Prize for preserving food
- 1810 Napolean: A flax spinning machine
- 1833 Societe de Encouragement pour le Industrie National: Large-scale commercial hydraulic turbines
- 1852 Royal Agricultural Society of Britain Guano Prize: a fertilizer as effective as Peruvian guano
- 1863 Phelan and Collender billiard ball company: Non-ivory billiard ball

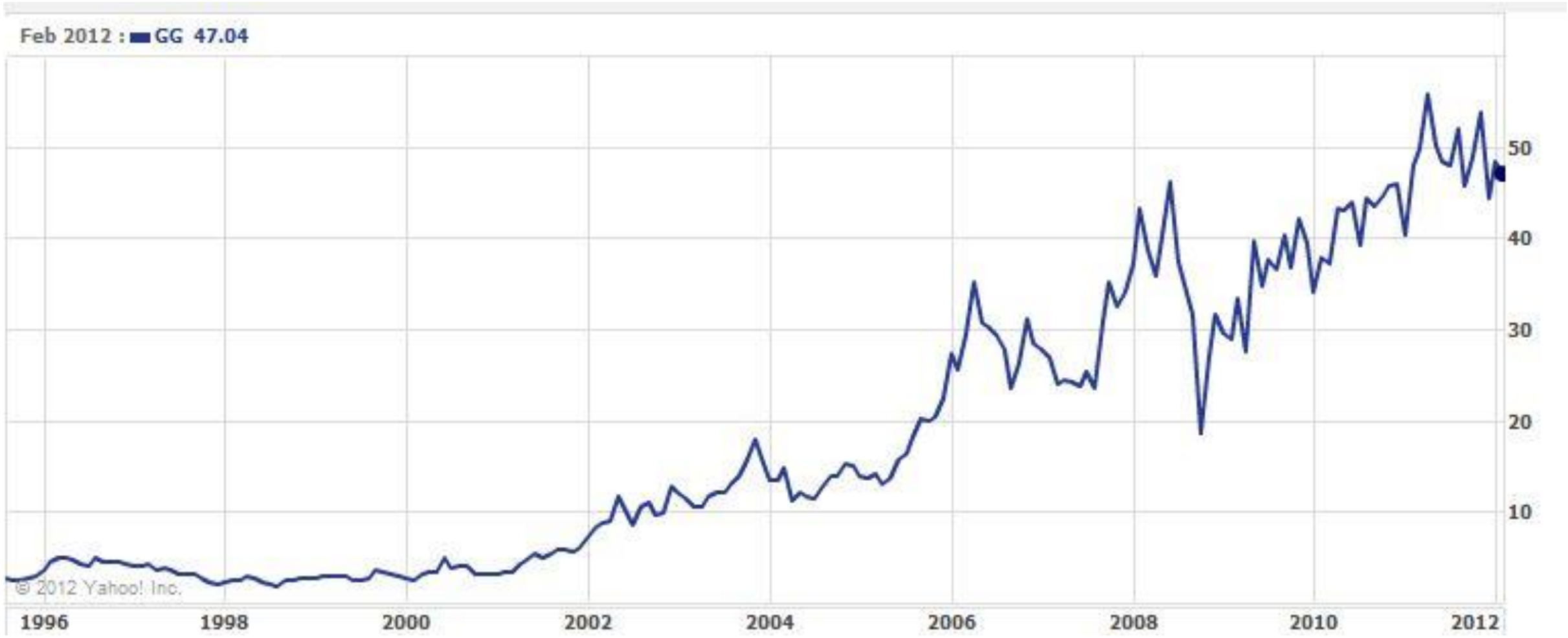
GOLDCORP | CHALLENGE.COM
CHALLENGE 1
Take the Challenge
Win the Gold
Forbes
THE GLOBAL SEARCH CHALLENGE
GLOBAL SEARCH
GOLDCORP


Rob McEwen

Goldcorp Challenge

- 400 megabytes of data
- >1000 participants
- $575,000 in prizes

- Identified over 100 sites
- 50% previously unknown
- >80% yielded significant gold reserves
- Value > $6B
- Estimated time savings of 2-3 years


Feb 2012 : GG 47.04
50
40
30
20
10
© 2012 Yahoo! Inc.
1996
1998
2000
2002
2004
2006
2008
2010
2012

DARPA
GRAND CHALLENGE

Leaderboard

10.05%

Display top 20 leaders.

Rank	Team Name	Best Score	% Improvement	Last Submit Time
1	BellKor's Pragmatic Chaos	0.8558	10.05	2009-06-26 18:42:37
	Grand Prize - RMSE <= 0.8563			
2	PragmaticTheory	0.8582	9.80	2009-06-25 22:15:51
3	BellKor in BigChaos	0.8590	9.71	2009-05-13 08:14:09
4	Grand Prize Team	0.8593	9.68	2009-06-12 08:20:24
5	Dace	0.8604	9.56	2009-04-22 05:57:03
6	BigChaos	0.8613	9.47	2009-06-23 23:06:52

ANSARI
3
People
100 Km
(62 Miles)
14 Days
PRIZE®

XPRIZE
Making the Impossible POSSIBLE

What We Do

REINVENT THE TOILET CHALLENGE

STRATEGY OVERVIEW

Print

GENERAL INFORMATION

Strategy Overview →

Strategy Leadership

RESOURCES

Press Releases (22)

QUICK LINKS

Reinvent the Toilet Fair: India 2014 Program

Reinvent the Toilet Fair: India 2014 Technical Guide

Reinvent the Toilet Fair 2012 Program

Reinvent the Toilet Challenge Fact Sheet

Reinvent the Toilet Videos

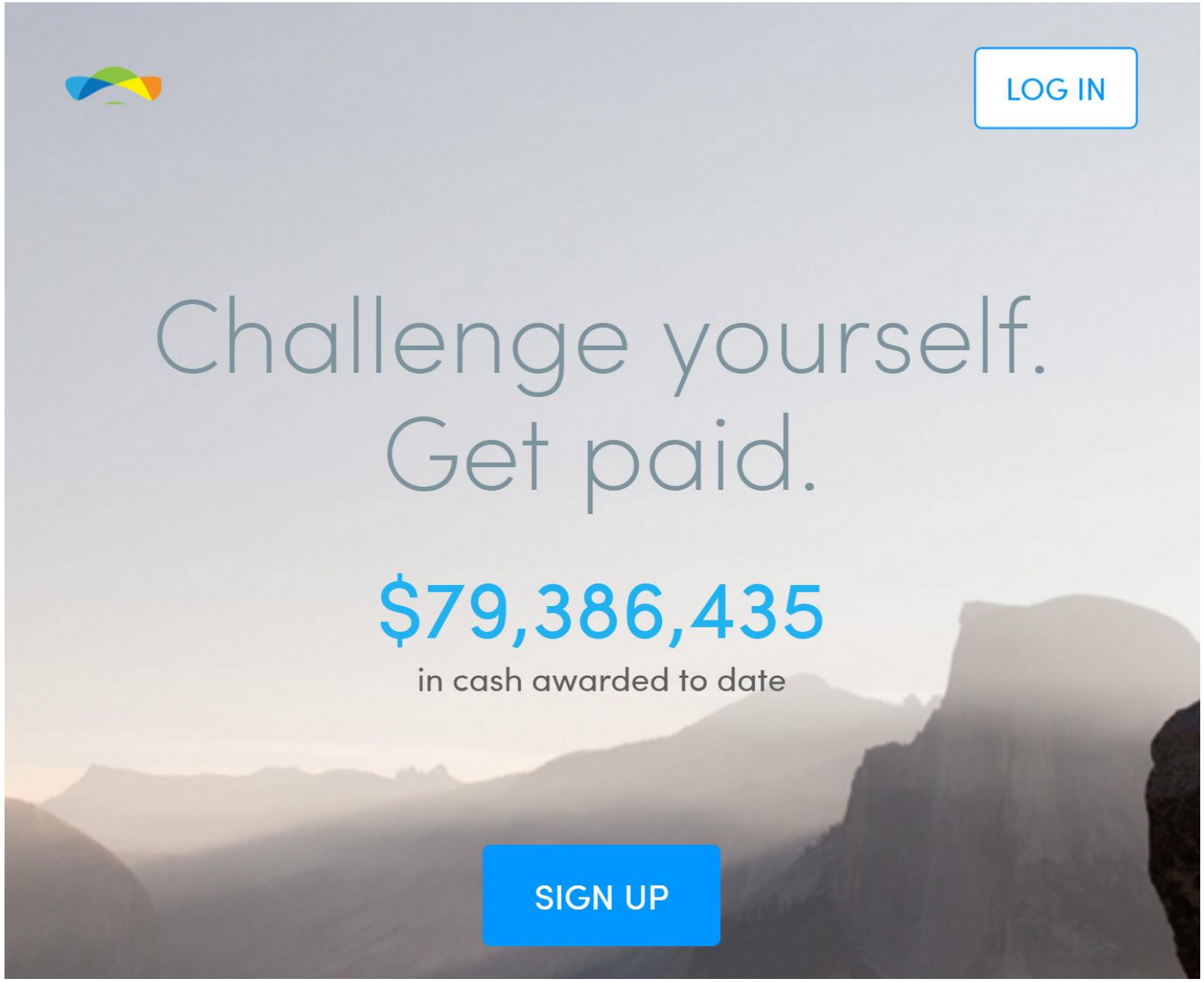
Researchers at the California Institute of Technology (Caltech) have built a toilet that uses the sun to power an electrochemical reactor.

About the Reinvent the Toilet Challenge

In 2011, the Water, Sanitation & Hygiene program initiated the Reinvent the Toilet Challenge to bring sustainable sanitation solutions to the 2.5 billion people worldwide who don't have access to safe, affordable sanitation.

In This Page

ABOUT THE REINVENT THE TOILET CHALLENGE


LOG IN
Challenge yourself.
Get paid.
$79,386,435
in cash awarded to date
SIGN UP

INNOCENTIVE®

Innocentive Statistics

- Total Registered Solvers: >375,000 from nearly 200 countries
- Total Challenges Posted: >2,000 External Challenges
- Total Award Dollars Posted: >$48M
- Range of Awards: $5,000 to $1M
- Total Awards Given: >2,400
- Premium Challenge Success Rate: 85%

Other Examples

- Threadless
- 99designs
- OpenIdeo

Netflix Prize

- 2006:
  Netflix Cinematch algorithm: 0.9525 RMSE
  (Just give mean rating for a movie: 1.054)
- $1M if you could improve this by 10%, to 0.8572
- $50,000 per year to best attempt if 10% not reached, as long as it's 1% better than the previous year
- Ineligible countries: Cuba, Iran, Syria, North Korea, Myanmar, Sudan, and

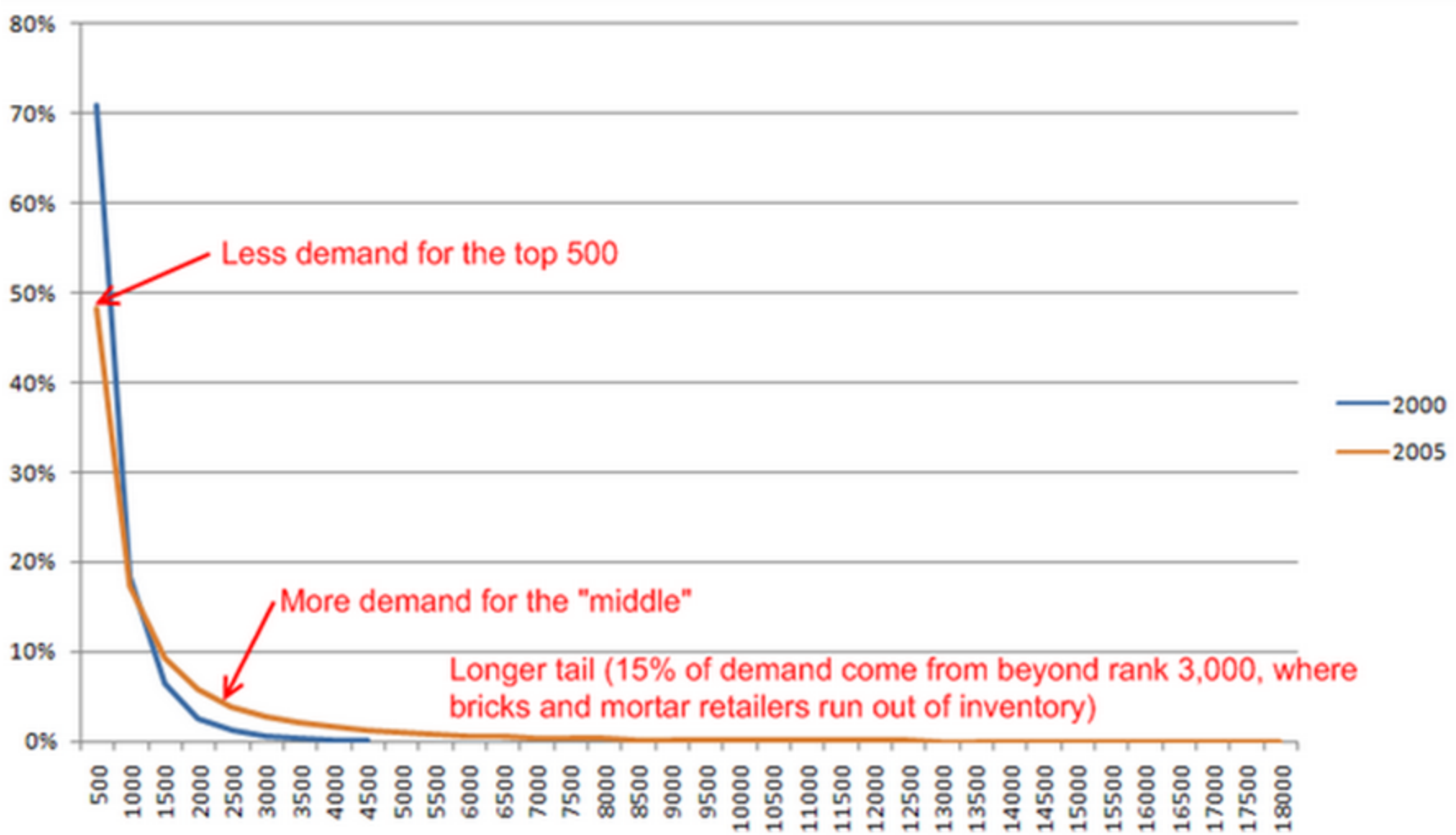
Netflix Prize

- 2006:
  Netflix Cinematch algorithm: 0.9525 RMSE
  (Just give mean rating for a movie: 1.054)
- $1M if you could improve this by 10%, to 0.8572
- $50,000 per year to best attempt if 10% not reached, as long as it's 1% better than the previous year
- Ineligible countries: Cuba, Iran, Syria, North Korea, Myanmar, Sudan, and Quebec

Netflix Prize Risks

- Privacy violation
- Prize won immediately
- No winners
- Labor overhead
- Lowering barrier to entry from competitors
- $1M

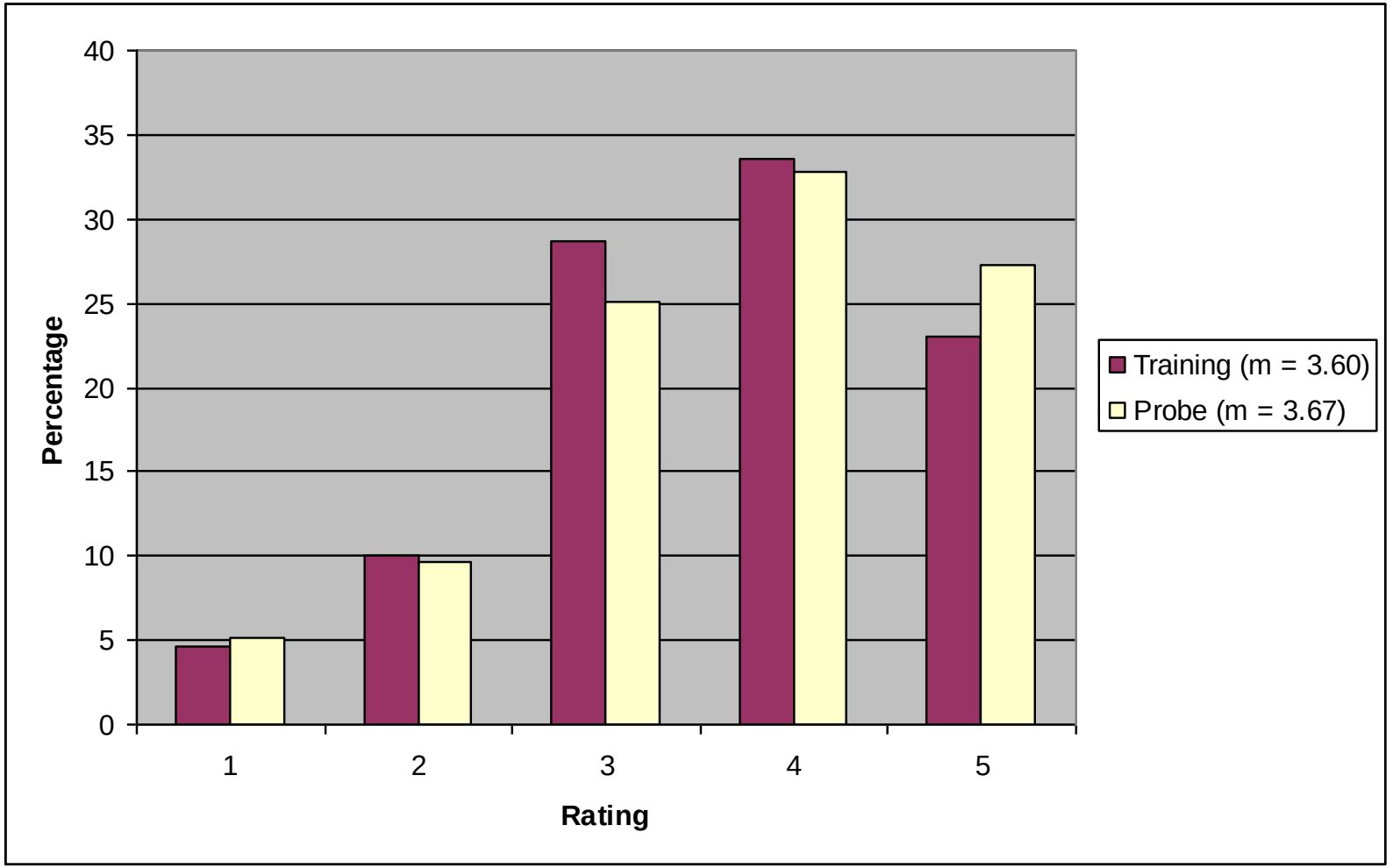
Netflix Prize



Netflix Prize Data

- 100,480,507 ratings for training (2000-2005)
  - 480,189 users
  - 17,770 movies
  - Each item: <user, movie, date, {1:5}>
  - 1,408,395 “probe” items in training set with distribution similar to test data
  - Average user rated > 200 movies
    One user rated over 17,000
  - Average movie rated by > 5000 users
    Some movies had only 3 ratings
- Hidden data:
  - 1,408,342 items you could get error rate on
  - 1,408,789 items on which Netflix rated submissions

Higher Mean Rating in Probe Data



Data about the Movies

Most Loved Movies	Avg rating
The Shawshank Redemption	4.593
Lord of the Rings :The Return of the King	4.545
The Green Mile	4.306
Lord of the Rings :The Two Towers	4.460
Finding Nemo	4.415
Raiders of the Lost Ark	4.504

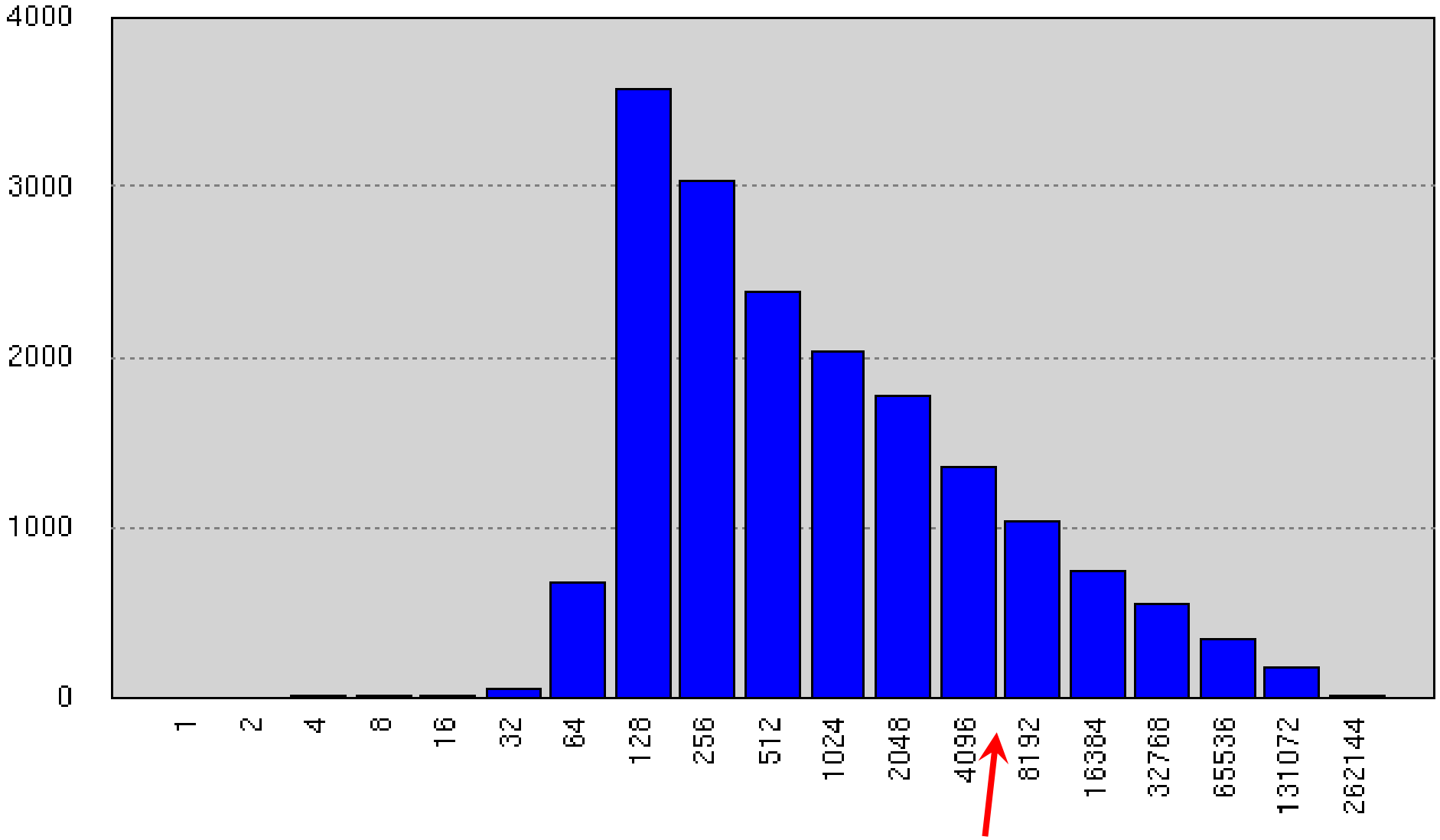
Most Rated Movies
Miss Congeniality
Independence Day
The Patriot
The Day After Tomorrow
Pretty Woman
Pirates of the Caribbean

Highest Variance
The Royal Tenenbaums
Lost In Translation
Pearl Harbor
Miss Congeniality
Napolean Dynamite
Fahrenheit 9/11

Most Active Users

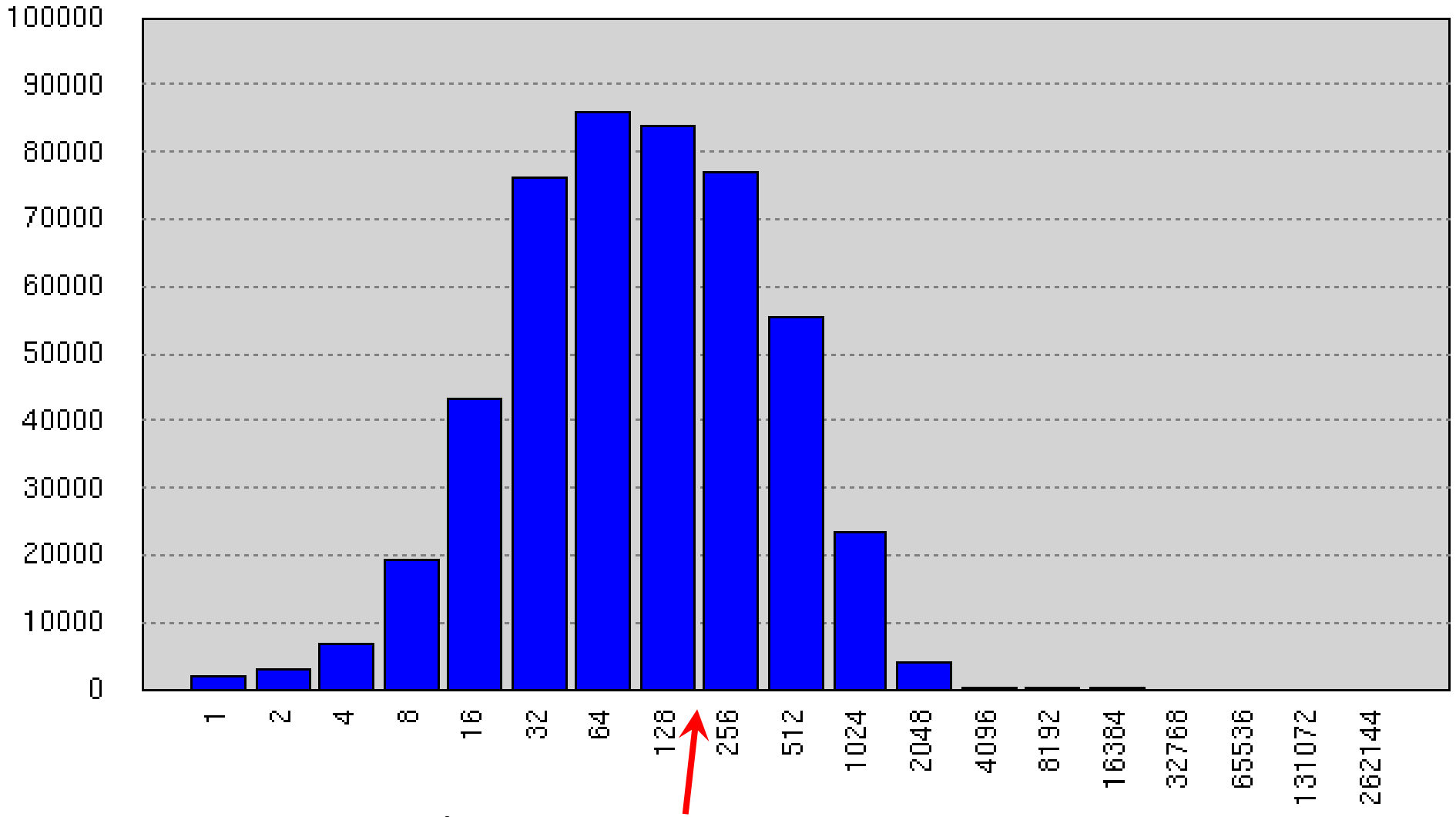
User ID	# Ratings	Mean Rating
305344	17,651	1.90
387418	17,432	1.81
2439493	16,560	1.22
1664010	15,811	4.26
2118461	14,829	4.08
1461435	9,820	1.37
1639792	9,764	1.33
1314869	9,739	2.95

Ratings per Movie in Training Data



Avg #ratings/movie: 5627

Ratings per User in Training Data

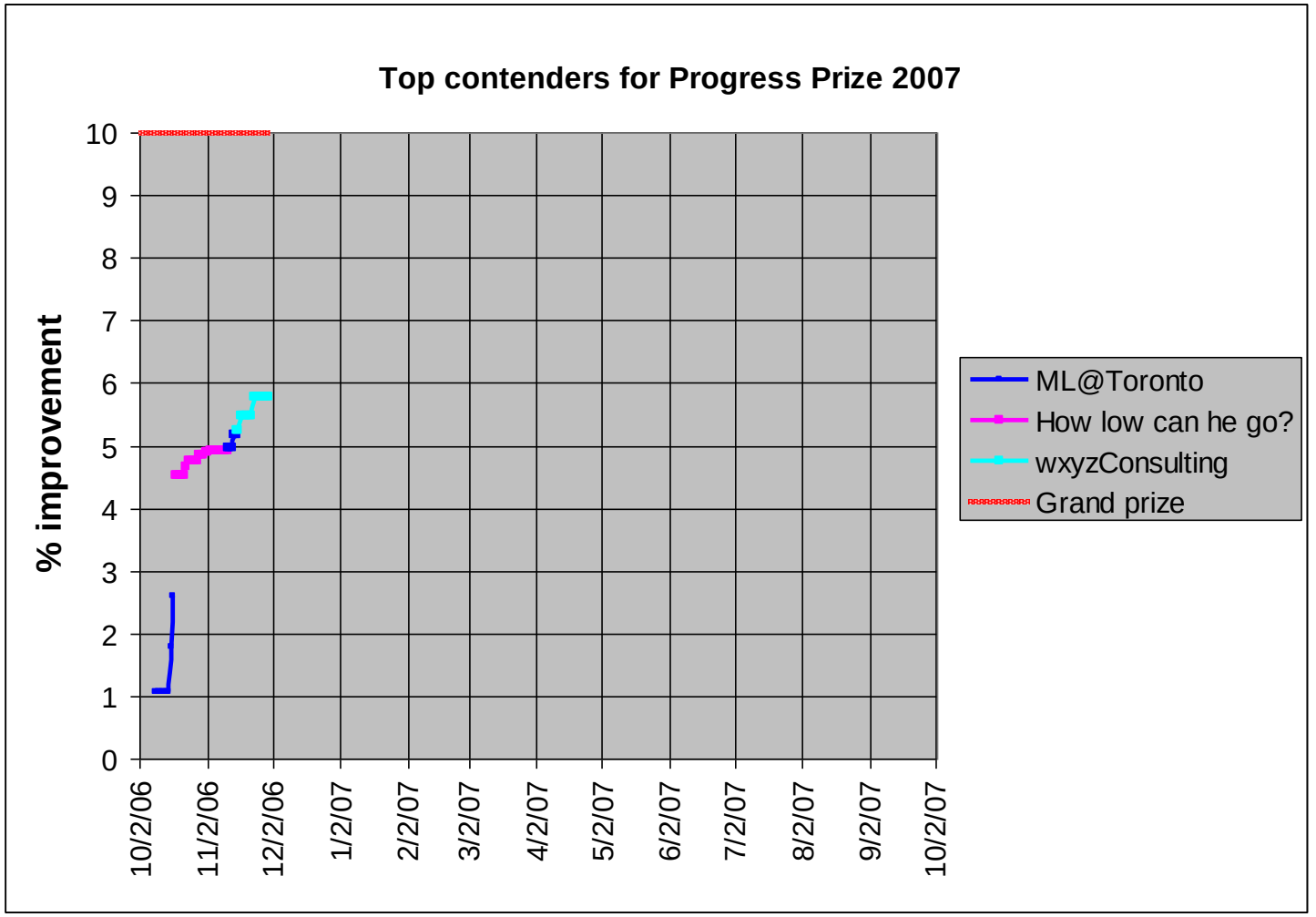


Avg #ratings/user: 208

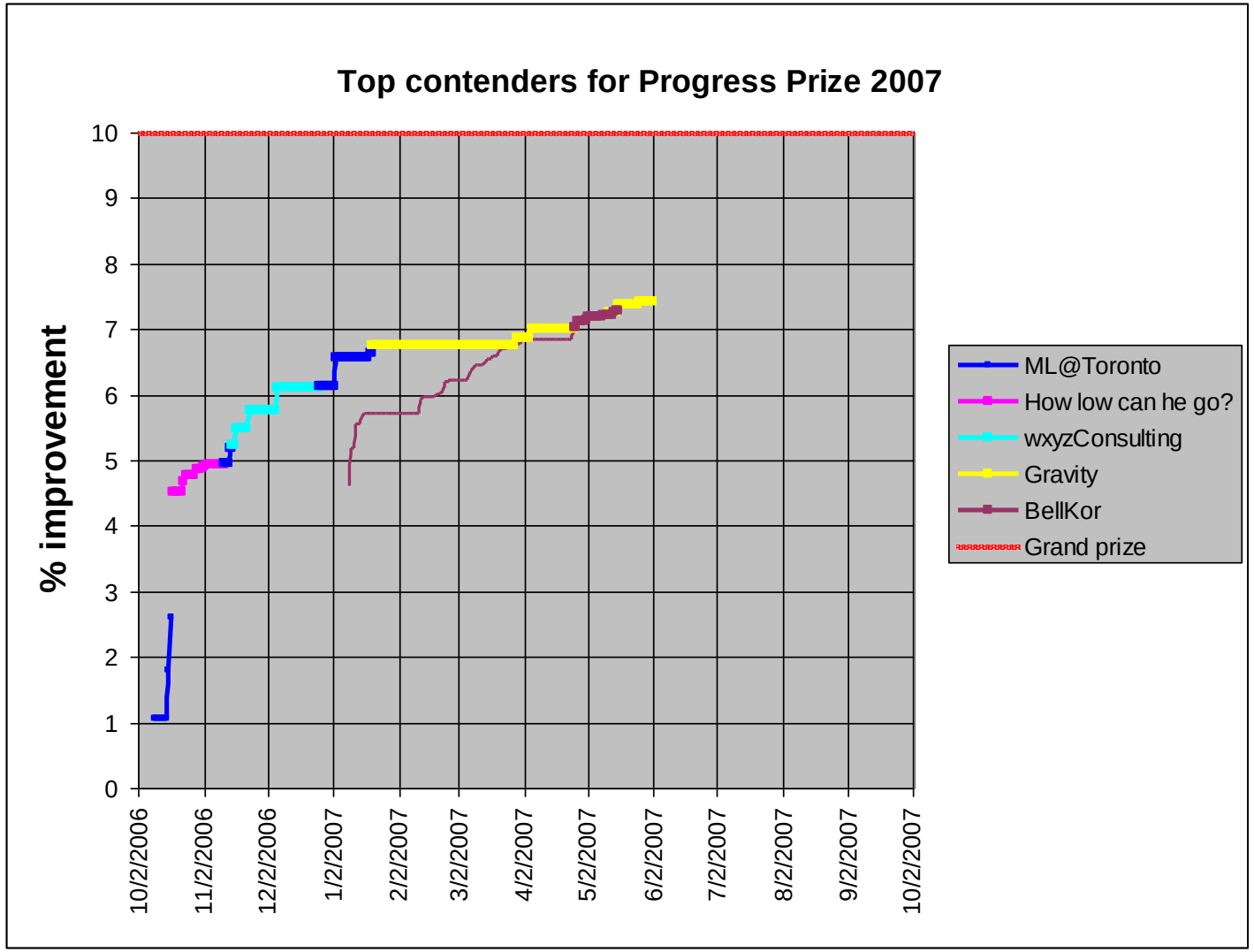
Netflix Prize Timeline

- October 2, 2006: Contest opened
- October 8, 2006: WXYZConsulting beats Cinematch
- October 15, 2006: 3 teams had beaten Cinematch, one by >1%
- June 2007: over 20,000 teams (150 countries) registered

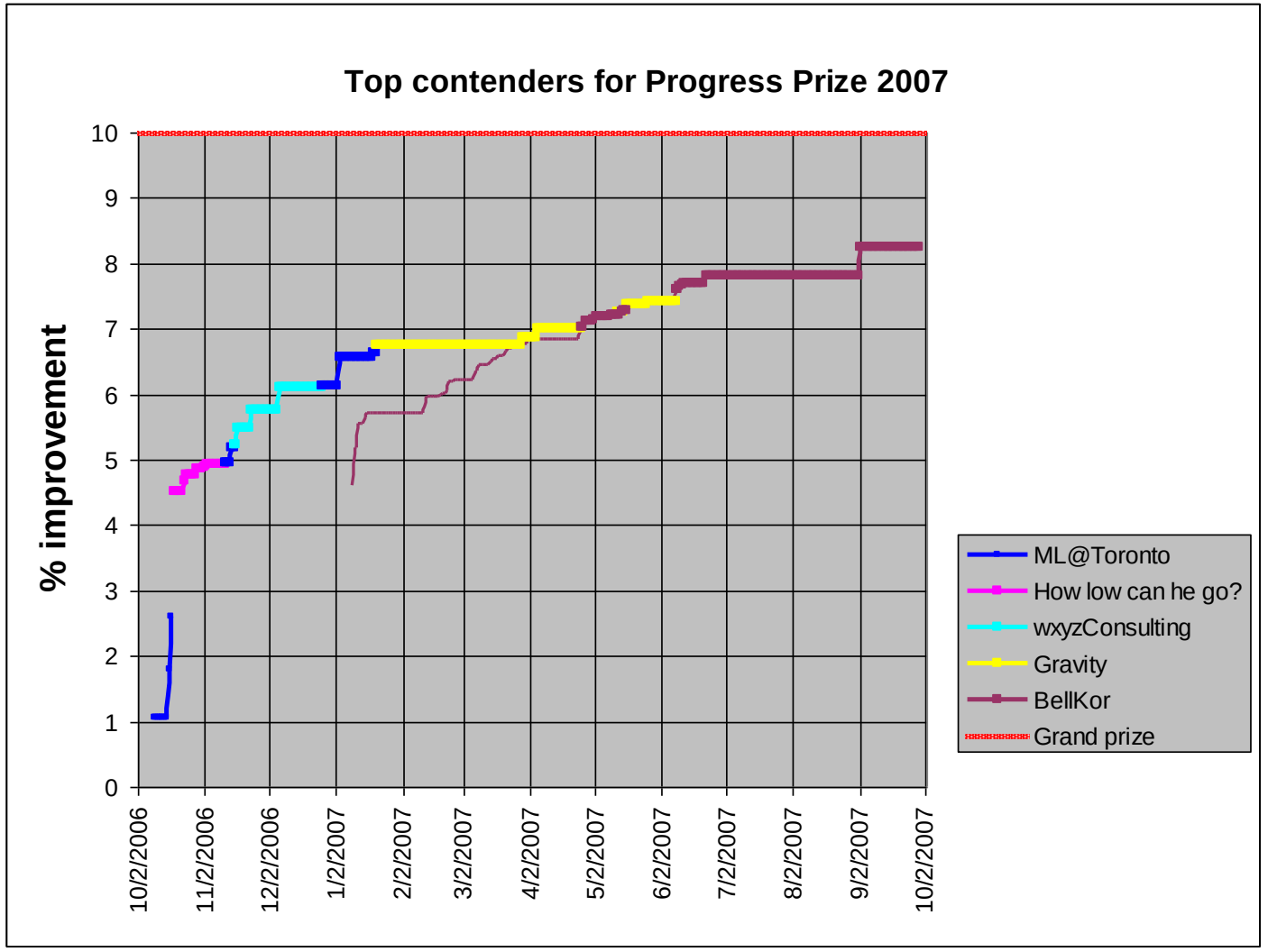
Progress after 2 Months



Progress after 8 Months



Progress after 1 Year



Netflix Prize Timeline

- Year 1:
  - Progress prize:
    - KorBell (aka BellKor): 8.43% improvement
    - Publish description of their algorithm
    - Linear combination of 107 different factors

Netflix Prize Timeline

- Year 1:
  - Progress prize:
    - KorBell (aka BellKor): 8.43% improvement
    - Publish description of their algorithm
    - Linear combination of 107 different factors
- Year 2:
  - Progress prize:
    - Only 3 teams qualify (>1%)
    - BellKor in BigChaos: 9.44%
    - Publish description of their algorithm

Netflix Prize Timeline

- Year 1:
  - Progress prize:
    - KorBell (aka BellKor): 8.43% improvement
    - Publish description of their algorithm
    - Linear combination of 107 different factors
- Year 2:
  - Progress prize:
    - Only 3 teams qualify (>1%)
    - BellKor in BigChaos: 9.44%
    - Publish description of their algorithm
- Year 3:
  - Top 2 candidates:
    - BellKor's Pragmatic Chaos: 10.05%
    - The Ensemble: 10.09%

Netflix Prize Timeline

Rank	Team Name	Best Test Score	% Improvement	Best Submit Time
Grand Prize - RMSE = 0.8567 - Winning Team: BellKor's Pragmatic Chaos				
1	BellKor's Pragmatic Chaos	0.8567	10.06	2009-07-26 18:18:28
2	The Ensemble	0.8567	10.06	2009-07-26 18:38:22
3	Grand Prize Team	0.8582	9.90	2009-07-10 21:24:40
4	Opera Solutions and Vandelay United	0.8588	9.84	2009-07-10 01:12:31
5	Vandelay Industries !	0.8591	9.81	2009-07-10 00:32:20
6	PragmaticTheory	0.8594	9.77	2009-06-24 12:06:56
7	BellKor in BigChaos	0.8601	9.70	2009-05-13 08:14:09
8	Dace_	0.8612	9.59	2009-07-24 17:18:43
9	Feeds2	0.8622	9.48	2009-07-12 13:11:51
10	BigChaos	0.8623	9.47	2009-04-07 12:33:59
11	Opera Solutions	0.8623	9.47	2009-07-24 00:34:07
12	BellKor	0.8624	9.46	2009-07-26 17:19:11

Netflix Prize Timeline

- Year 1:
  - Progress prize:
    - KorBell (aka BellKor): 8.43% improvement
    - Publish description of their algorithm
    - Linear combination of 107 different factors
- Year 2:
  - Progress prize:
    - BellKor in BigChaos: 9.44%
    - Publish description of their algorithm
- Year 3:
  - Top 2 candidates:
    - BellKor's Pragmatic Chaos: 10.05%
    - The Ensemble: 10.09%

Netflix Prize Timeline

- Year 1:
  - Progress prize:
    - KorBell (aka BellKor): 8.43% improvement
    - Publish description of their algorithm
    - Linear combination of 107 different factors
- Year 2:
  - Progress prize:
    - BellKor in BigChaos: 9.44%
    - Publish description of their algorithm
- Year 3:
  - Top 2 candidates:
    - BellKor's Pragmatic Chaos: 10.05%
    - The Ensemble: 10.09%
    - Over 800 factors

References

- Y. Koren, Collaborative filtering with temporal dynamics, ACM SIGKDD Conference 2009
- Koren, Bell, Volinsky, Matrix factorization techniques for recommender systems, IEEE Computer, 2009
- Y. Koren, Factor in the neighbors: scalable and accurate collaborative filtering, ACM Transactions on Knowledge Discovery in Data, 2010

Netflix Prize Postscript: 2007

Robust De-anonymization of Large Sparse Datasets

Arvind Narayanan and Vitaly Shmatikov
The University of Texas at Austin

Abstract

We present a new class of statistical de-anonymization attacks against high-dimensional micro-data, such as individual preferences, recommendations, transaction records and so on. Our techniques are robust to perturbation in the data and tolerate some mistakes in the adversary's background knowledge.

We apply our de-anonymization methodology to the Netflix Prize dataset, which contains anonymous movie ratings of 500,000 subscribers of Netflix, the world's largest online movie rental service. We demonstrate that an adversary who knows only a little bit about an individual subscriber can easily identify this subscriber's record in the dataset. Using the Internet Movie Database as the source of background knowledge, we successfully identified the Netflix records of known users, uncovering their apparent political preferences and other potentially sensitive information.

1 Introduction

and sparsity. Each record contains many attributes (*i.e.*, columns in a database schema), which can be viewed as dimensions. Sparsity means that for the average record, there are no "similar" records in the multi-dimensional space defined by the attributes. This sparsity is empirically well-established [7, 4, 19] and related to the "fat tail" phenomenon: individual transaction and preference records tend to include statistically rare attributes.

Our contributions. Our first contribution is a formal model for privacy breaches in anonymized micro-data (section 3). We present two definitions, one based on the probability of successful de-anonymization, the other on the amount of information recovered about the target. Unlike previous work [25], we do not assume *a priori* that the adversary's knowledge is limited to a fixed set of "quasi-identifier" attributes. Our model thus encompasses a much broader class of de-anonymization attacks than simple cross-database correlation.

Our second contribution is a very general class of de-anonymization algorithms, demonstrating the fundamental limits of privacy in public micro-data (section 4). Under very mild assumptions about the distribution from

Netflix Prize Postscript: 2009

NETFLIX

Press Releases

Current Press Releases

<< [Back to Press Releases]

Netflix Awards $1 Million Netflix Prize and Announces Second $1 Million Challenge

NEW YORK, Sept. 21 /PRNewswire/ -- After almost three years and submissions by more than 40,000 teams from 186 countries, Netflix, Inc., the world's largest online movie rental service (NASDAQ: NFLX), today awarded the $1 million Netflix Prize to a team of engineers, statisticians and researchers who achieved the competition's goal of a 10 percent improvement over the accuracy of the Netflix movie recommendation system when the competition was launched in Oct. 2006. Netflix members already are benefiting from improvements Netflix Prize contestants have contributed to the recommendations system.

Moments after bestowing the $1 million prize, Netflix announced a second $1 million challenge, asking the world's computer science and machine learning communities to keep the improvements coming.

The team "BellKor's Pragmatic Chaos," the merging of three teams that had previously competed against one another in the contest, received the $1 million Netflix Prize in an award ceremony hosted here today by Netflix Co-Founder and CEO Reed Hastings and Chief Product Officer Neil Hunt.

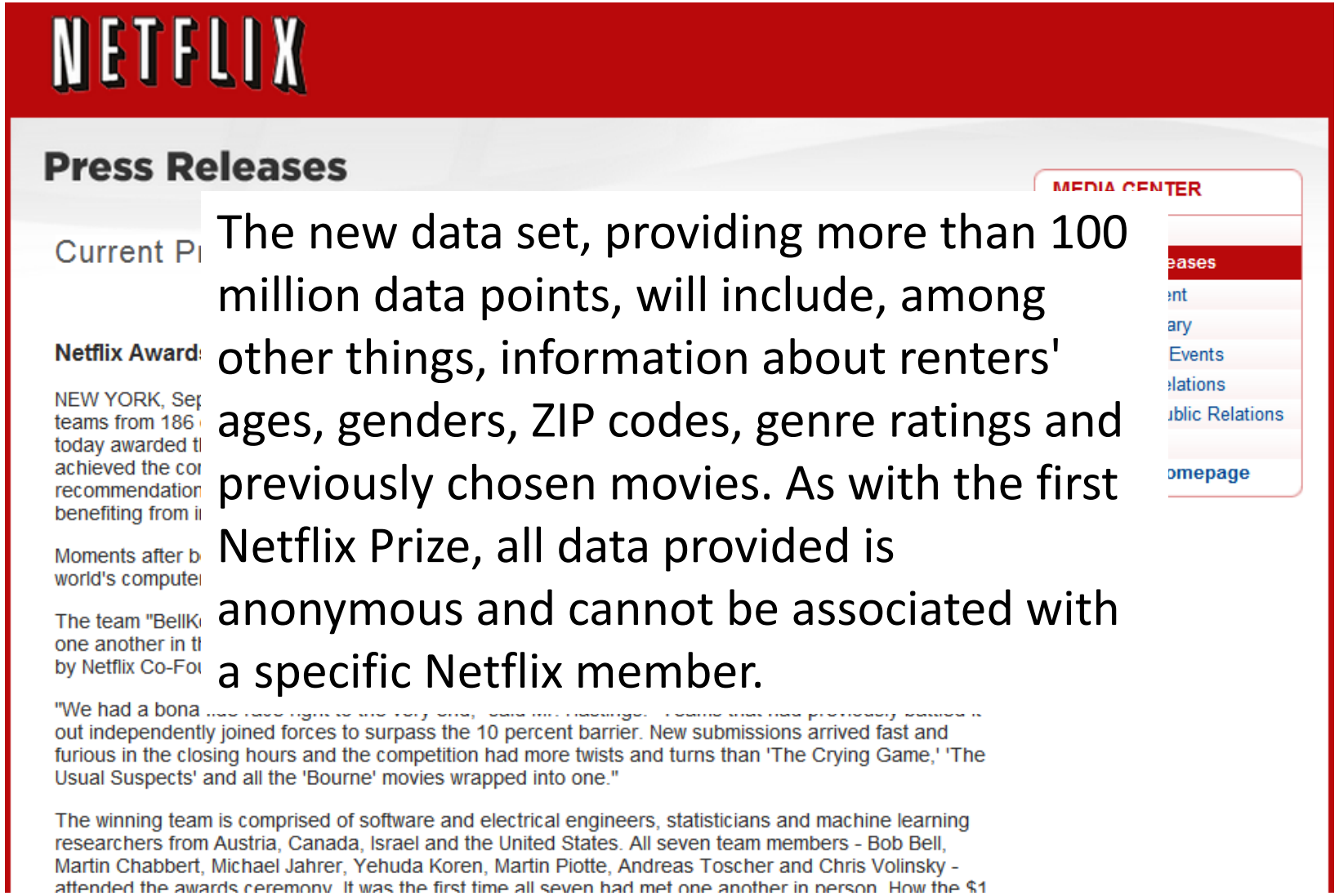
"We had a bona fide race right to the very end," said Mr. Hastings. "Teams that had previously battled it out independently joined forces to surpass the 10 percent barrier. New submissions arrived fast and furious in the closing hours and the competition had more twists and turns than 'The Crying Game,' 'The Usual Suspects' and all the 'Bourne' movies wrapped into one."

The winning team is comprised of software and electrical engineers, statisticians and machine learning researchers from Austria, Canada, Israel and the United States. All seven team members - Bob Bell, Martin Chabbert, Michael Jahrer, Yehuda Koren, Martin Piotte, Andreas Toscher and Chris Volinsky - attended the awards ceremony. It was the first time all seven had met one another in person. How the $1

MEDIA CENTER

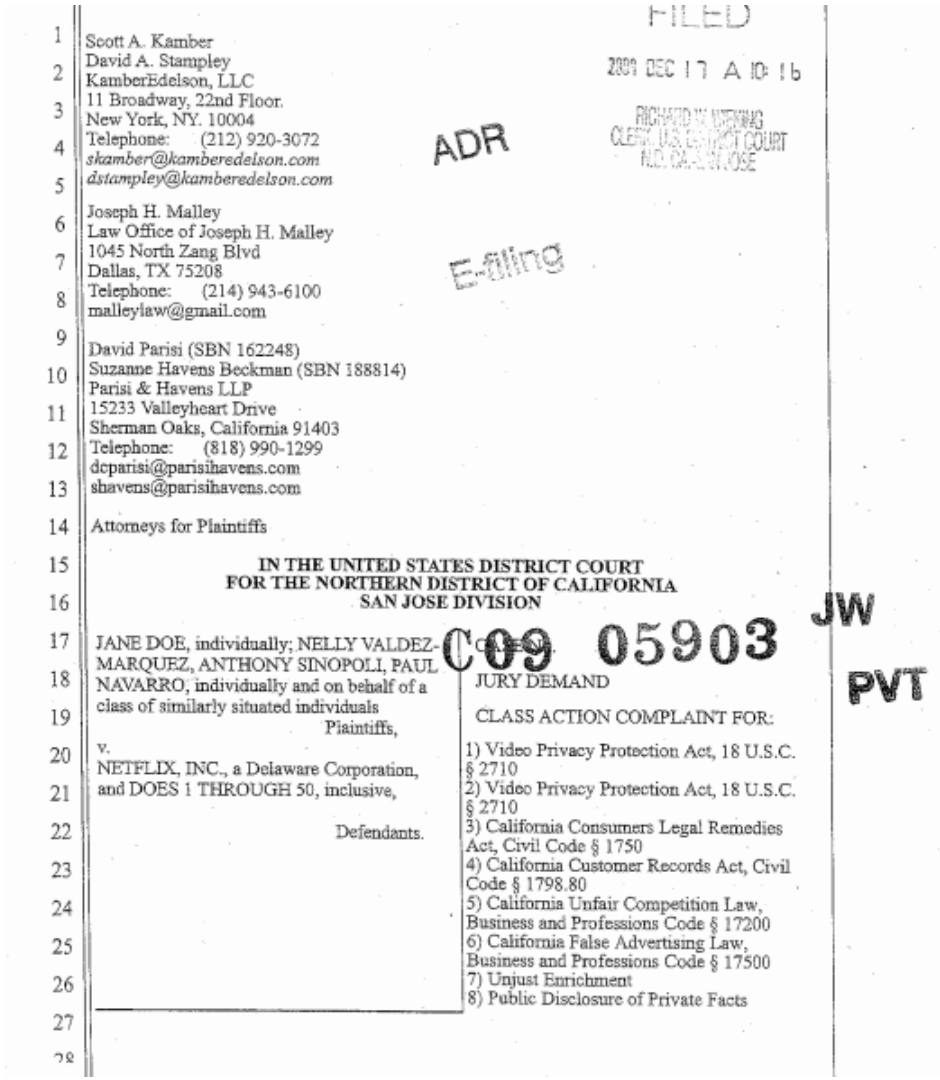
- Press Kit
- Press Releases
- Management
- Image Library
- Upcoming Events
- Investor Relations
- Contact Public Relations
- RSS
- Back to Homepage

Netflix Prize Postscript: 2009



The new data set, providing more than 100 million data points, will include, among other things, information about renters' ages, genders, ZIP codes, genre ratings and previously chosen movies. As with the first Netflix Prize, all data provided is anonymous and cannot be associated with a specific Netflix member.

Netflix Prize Postscript: 2009



FILED

2009 DEC 17 A 10:16

RICHARD W. WIEKING
CLERK, U.S. DISTRICT COURT
N.D. CA, S.J.

ADR

E-filing

Scott A. Kamber
David A. Stampley
KamberEdelson, LLC
11 Broadway, 22nd Floor
New York, NY. 10004
Telephone: (212) 920-3072
skamber@kamberedelson.com
dstampley@kamberedelson.com

Joseph H. Malley
Law Office of Joseph H. Malley
1045 North Zang Blvd
Dallas, TX 75208
Telephone: (214) 943-6100
malleylaw@gmail.com

David Parisi (SBN 162248)
Suzanne Havens Beckman (SBN 188814)
Parisi & Havens LLP
15233 Valleyheart Drive
Sherman Oaks, California 91403
Telephone: (818) 990-1299
dcparisi@parisihavens.com
shavens@parisihavens.com

Attorneys for Plaintiffs

IN THE UNITED STATES DISTRICT COURT
FOR THE NORTHERN DISTRICT OF CALIFORNIA
SAN JOSE DIVISION

JANE DOE, individually; NELLY VALDEZ-MARQUEZ, ANTHONY SINOPOLI, PAUL NAVARRO, individually and on behalf of a class of similarly situated individuals

Plaintiffs,

v.

NETFLIX, INC., a Delaware Corporation, and DOES 1 THROUGH 50, inclusive,

Defendants.

Case No. C09 05903 JW PVT

JURY DEMAND

CLASS ACTION COMPLAINT FOR:

1) Video Privacy Protection Act, 18 U.S.C. § 2710
2) Video Privacy Protection Act, 18 U.S.C. § 2710
3) California Consumers Legal Remedies Act, Civil Code § 1750
4) California Customer Records Act, Civil Code § 1798.80
5) California Unfair Competition Law, Business and Professions Code § 17200
6) California False Advertising Law, Business and Professions Code § 17500
7) Unjust Enrichment
8) Public Disclosure of Private Facts

Netflix Prize Postscript: 2009

I. NATURE OF THE ACTION

1. On October 2, 2006, Netflix perpetrated the largest voluntary privacy breach to date, disclosing sensitive and personal indentifying consumer information. The information was not compromised by malicious intruders. Rather, it was given away to the world freely, and with fanfare, as part of a contest intended to benefit its trusted custodian, Netflix.

2. This right to privacy does not appear to be significant to Netflix. This lawsuit is brought as a class action by and on behalf of similarly situated Netflix subscribers, qualified by the class definition and class period, whose privacy was violated by the actions of Netflix, Inc., ("Netflix") pursuant to their contest, "Netflix Prize." Jane Doe, a lesbian, who does not want her sexuality nor interests in gay and lesbian themed films broadcast to the world, seeks anonymity in this action. Paul Navarro files this action to prevent Netflix from going through with its announced intentions to make additional disclosures of personal identifying information including, but not limited to, users' video renting history and rating habits.

3. Netflix knowingly authorized, directed, ratified, approved, acquiesced, or participated in the disclosure to third parties of the sensitive information and/or personal identifying information derived from the activity of the Netflix subscribers' online electronic communications, when they accessed the Netflix website to rent and rate videos.

4. Netflix is an "Electronic Communication Service Provider" to its subscribers and knowingly disclosed to third parties the contents of Netflix's subscribers' communications, including but not limited to, subscribers' rental and rating videos information, while in elec-

Netflix Prize Postscript: 2010

UNITED STATES OF AMERICA
FEDERAL TRADE COMMISSION
WASHINGTON, D.C. 20580

Maneesha Mithal
Associate Director
Division of Privacy & Identity Protection

Direct Dial: 202.326.2771
Fax : 202.326.3062
E-mail: mmithal@ftc.gov

March 12, 2010

BY E-MAIL & FEDERAL EXPRESS

Reed Freeman
Morrison & Foerster LLP
2000 Pennsylvania Ave., NW
Washington, DC 20006

Dear Mr. Freeman:

On October 13, 2009, staff from the FTC's Division of Privacy and Identity Protection contacted your client, Netflix, Inc. ("Netflix"), regarding the privacy implications of Netflix's planned release of customer movie viewing data in connection with the company's efforts to improve its movie recommendation algorithm. Specifically, staff expressed concern that, despite Netflix's efforts to "anonymize" the customer data prior to its release, it would be possible to re-identify specific customers and thereby associate them with their movie viewing histories and preferences.

Staff's concerns about Netflix's planned release stemmed from research published after the

Netflix Prize Postscript: 2010

NETFLIX The Netflix Blog

FRIDAY, MARCH 12, 2010

Netflix Prize Update

This is Neil Hunt, Chief Product Officer for Netflix.

6 Tweet 36 Like

About five months ago we announced that Netflix would sponsor a sequel to the Netflix Prize. We've given a lot thought to how to sponsor a contest that discovers more about the predictability of Netflix members' movie watching behavior while always ensuring we protect Netflix members' privacy.

In the past few months, the Federal Trade Commission (FTC) asked us how a Netflix Prize sequel might affect Netflix members' privacy, and a lawsuit was filed by KamberLaw LLC pertaining to the sequel. With both the FTC and the plaintiffs' lawyers, we've had very productive discussions centered on our commitment to protecting our members' privacy.

We have reached an understanding with the FTC and have settled the lawsuit with plaintiffs. The resolution to both matters involves certain parameters for how

LINKS

Netflix Tech Blog
Netflix America Latina
Netflix Brasil
Facebook Netflix Page
Netflix Website
---> RSS Feed Page
---> DVDs Releasing This Week
---> Logo and Media Materials
---> Jobs at Netflix

ABOUT THE NETFLIX BLOG

Hello and welcome to the official Netflix Blog! We the blog authors are various members of the Netflix team. We're also rabid movie fans. We hope this will be a great forum for us to talk about what we are doing, and for you to tell us what you think.

BLOG ARCHIVE

► 2011 (40)
▼ 2010 (27)
► December (3)
► November (4)
► October (3)