- [Start-Ups](#)
- [Business Computing](#)
- [Companies](#)

[Bits](#)
[Blog »](#)

[Personal Tech »](#)

- [Digital Cameras](#)
- [Cellphones](#)
- [All Products](#)

Bits

Business ▪ Innovation ▪ Technology ▪ Society
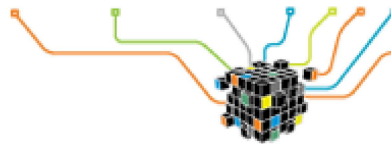
March 12, 2010, *2:46 pm*

Netflix Cancels Contest Plans and Settles Suit

By [STEVE LOHR](#)

[Netflix](#)'s $1 million prize contest was such a [research and business hit](#)that when the winners were declared last September the company immediately announced plans for another one.

But it turned out that letting very smart computer scientists and statisticians dig through the video rental site's data had one major, unforeseen, drawback. A pair of researchers at the [University of Texas](#) showed that the supposedly anonymized data released for the contest, which included movie recommendations and choices made by hundreds of thousands of customers, [could in fact be used to identify them](#).

That brought the attention of the Federal Trade Commission, and also drew a lawsuit from KamberLaw L.L.C. On Friday, bowing to privacy concerns, Netflix said it was shelving its plans for a sequel to the prize contest.

In a [post](#)on a company blog, Neil Hunt, Netflix's chief product officer, explained that the company had "reached an understanding" with the F.T.C. and settled the lawsuit. "The resolution of both matters involves certain parameters on how we use Netflix data in any future research programs," he said.

Netflix, Mr. Hunt wrote, will continue to "explore ways to collaborate with the research community" to improve its recommendation system.

The three-year competition, begun in 2006, awarded the $1 million prize to the first group that could improve the predictive accuracy of the company's internal recommendation software. A seven-person team, led by two current and one former employee of AT&T Labs, made it first.

During the contest, teams joined up and combined their software and statistical models to improve their results. Experts hailed the outcome as proof of the power of new crowdsourcing techniques for finding nuggets of value in vast sets of data.

But the big data sets also carry the potential to inadvertently reveal a person's identity, even without a name attached. Data about online personal behavior can increasingly be sifted, correlated and analyzed for patterns that often point to an individual.

So-called de-anonymization work is an emerging field in university research, and has the potential for use and abuse in industry.

"Slices of our online behavior can be identifying, even though you don't think so," said Vitaly Shmatikov, a computer scientist at the University of Texas.

Mr. Shmatikov and Arvind Narayanan, who is now a post-graduate researcher at Stanford University, were the authors of the de-anonymization research using the Netflix data.

- E-mail This
- Print
- Share
    Twitter
- Sign in to Recommend

- Previous post Plastic Logic Delays Its Que Tablet Next post When Will Location-Based Coupons Take Off?

35 Readers' Comments

Post a Comment »

- All Comments
- Highlights
- Readers' Recommendations
- Replies

- Oldest
- Newest

1   of 2 Next

1.
dh27
nyc
March 12th, 2010
3:45 pm
I'm sympathetic to data privacy concerns, but this FTC action seems misplaced. For one, the harm that might flow from disclosure of this data seems minimal, mostly because Netflix really doesn't have controversial or embarrassing

titles; no one is going to learn that you rented porn. Social network users voluntarily disclose their movie preferences all the time; this isn't a SSN, a credit card number, or something potentially more sensitive like your google searches. Two, nobody is forcing you to use Netflix and there are plenty of (less convenient, but more private) alternatives. If it's important, you can consume movies you don't want the world to know you've watched through one of these safer vehicles (e.g., library, video store -- to which the Bork Bill still applies).
Three, the research and intellectual mobilization generated by the study was fascinating and the data set was extremely robust. Much more could have come of it.
[Recommend](#) Recommended by 13 Readers
[2](#).
Eva
NYC
March 12th, 2010
3:45 pm
I'm not sure I understand. If a list looked like this:

Person 1: Titanic, 101 Dalmations, Crash
Person 2: My Fair Lady, Bye Bye Birdie, Dancing in the Rain, West Side Story
Person 3....

how could that be a problem? Were locations also included? Even so, I find this confusing, unless there was additional web-browsing/cookie-collected data provided as well
[Recommend](#) Recommended by 3 Readers
[3](#).
umm
chicago
March 12th, 2010
3:58 pm

You can't rent anything weird on Netflix, so if this is at the zip code level, who cares?
[Recommend](#) Recommended by 5 Readers
[4](#).
Dan Clestine
Brooklyn
March 12th, 2010
4:24 pm
People who enjoyed this article might also enjoy.....
[Recommend](#) Recommended by 19 Readers
[5](#).
Thomas Zaslavsky
Binghamton, N.Y.
March 12th, 2010
4:24 pm
I think the point is that "anonymization" is at best uncertain and probably a myth. It depends on our happening not to know how to identify people, and case after case has shown that this lack of knowledge is temporary and has no possible guarantee.

As for the privacy issue: I just don't want my movie selections released to the general public without my consent. It's no one's business, regardless of whether I have anything you or I believe needs hiding. You may feel differently, and

you're free to issue a general permission to release all information.
Recommend Recommended by 7 Readers
6.
Dan
New York
March 12th, 2010
4:24 pm
When they say, "reached a settlement", obviously that's code for "money changed hands." I'm just curious: how much? Don't we live in a great system where you can pay off your regulators and your problems disappear?
Recommend Recommended by 2 Readers
7.
Lisa
Berkeley, CA
March 12th, 2010
4:24 pm
Like Eva, I also don't understand why this would raise privacy concerns. So you figure out that person 1 rented movies X, Y and Z, and loved Z most of all. You figure out the same sorts of things for person 2, and so on. How does this help you figure out who persons 1 and 2 and so on actually are?
Recommend Recommended by 0 Readers
8.
Bill
Ithaca
March 12th, 2010
4:24 pm
I dare anyone to identify me simply from the titles I have rented from Netflix.
Recommend Recommended by 4 Readers
9.
J.S.
New Jersey
March 12th, 2010
4:24 pm
Lawyers, nuf said.
Recommend Recommended by 5 Readers
10.
Susan Brandon
Manhattan
March 12th, 2010
4:24 pm
The point is that one's shopping expeditions and purchases are private transactions. In a world where internet companies are making tons of money, it is part of their responsibility to their customers to protect their privacy.
Recommend Recommended by 1 Readers
11.
Suhas
Houston Tx
March 12th, 2010
4:24 pm
Eva, the researchers at UT were able to de-anonymize some of the entries by comparing users' movie rankings to other non-anonymous ranking sites like IMDB.

So they notice that this entry from the Netflix data:
Person 19: Titanic rating 10, bye bye birdie rating 5, west side story rating 2

Matches this entry from IMDB:
joe johnson: Titanic rating 10, west side story rating 2, bye bye birdie rating 5

and they deduce that Person 19 is Joe Johnson. This would not have been possible w/o similar datasets with non-anonymous data.
Recommend Recommended by 3 Readers
12.
krp
ny, ny
March 12th, 2010
4:39 pm
@ dh27 and umm - - It all depends on what you call "controversial" or "weird." Someone might not want anyone to know that he or she rented torture porn or erotic thrillers. Someone else might not care. Personally, I don't mind the research, but the privacy issues are real.
Recommend Recommended by 0 Readers
13.
Smash
US
March 12th, 2010
4:39 pm
If you follow the link to the research paper and scroll past all the jargon, the authors compared Public ratings entered in the Internet Movie Database (IMDb) with the "anonymous" ratings contained in the Netflix data and could find a relationship, especially for those people that reviewed movies that were much less popular. They also gave an example of "water cooler" discussions where someone reveals their movie-watching likes/dislikes that could then be used to datamine the Netflix data. Me? I could care less as I don't participate in either online ratings or idle office chitcat.
Recommend Recommended by 3 Readers
14.
1st Timer
Washington DC
March 12th, 2010
4:39 pm
The study assumed that you had some knowledge of an individuals rental and ratings history. From that limited knowledge, say six approximate ratings and six approximate rental dates, you could potentially identify an individuals entire rental history in the data set.
Recommend Recommended by 0 Readers
15.
Jeff
Washington, DC
March 12th, 2010
4:39 pm
Netflix doesn't rent porn, but you might be able to tell that a person was renting movies with, say, gay characters or that a member of the current administration is a big fan of "24". If the data set contained the dates the movies were rented, you might be able to infer information about travel, etc. The issue is that Netflix promised its customers

privacy and for a sub-set of them broke the promise.

Recommend Recommended by 1 Readers
16.
janetvan
Detroit, MI
March 12th, 2010
4:39 pm
I'm happy to disclose that Superbad, Hurt Locker and OngBak 2 are my most recent rentals. No harm, no foul.

@Suhas: I don't see how your explanation works, as IMDB uses a 1-10 scale and Netflix uses a 1-5 scale with no halfsies.
Recommend Recommended by 0 Readers
17.
Jonathan Rochkind
Baltimore, MD
March 12th, 2010
4:39 pm
Commenters should view the "FAQ" by the U of Texas paper about what they showed or didn't show, answers a lot of questions and addresses some misconceptions in the comments.

http://userweb.cs.utexas.edu...

The new netflix prize data was going to include zip code and birth date as well, some of the implications of which are discussed here:

http://www.freedom-to-tinker.com...
Recommend Recommended by 4 Readers
18.
Publius
NY
March 12th, 2010
4:39 pm
Scary stuff man.

Not Netflix - but the immense power of computer algorithms to nail us.

I am a religious person, and I think there is much that is mysterious in the human person - but I also know that we are much, much more transparent than we think we are. It used to take a wise person some time to figure a person out.

Now, they can sift your basic demographic and personal info along with your browsing habits and predict what color underwear you'll wear on Saturday night to go see a movie you don't even know is out yet.

We all better get used to a little more humility. We're not as original and special as we think we are.
Recommend Recommended by 2 Readers
19.
Kate

New York
March 12th, 2010
4:47 pm
The Times had an interesting graphic a while back about what movies Netflix customers were renting broken down by Zip Codes. It was a revealing look at the populations of an area compared to other areas.

That said, there are dozens, if not hundreds, of titles I've rented over the years that I've never bothered to watch. Some were rented for others in my household, some were rented on a lark. I'm a big fan of privacy, but I really don't think my movie rental history, even paired with my Zip Code, could reveal my identity.
Recommend Recommended by 0 Readers

20.
Suhas
Houston Tx
March 12th, 2010
4:47 pm
janetvan, they didn't look for exact matches, just matches that were close enough.
Recommend Recommended by 0 Readers

21.
Josh Goldman
Boston MA
March 12th, 2010
4:57 pm
We live in a fairly free society (now) where there aren't usually problems if people know what movies you rent or books you read. But in other places, this clearly isn't true. If someone is living in China, public knowledge that he or she watches movies about the Dalai Lama, Falun Gong, or other unapproved topics could have very serious consequences.

Part of maintaining a free society is keeping some information private. If you were known to read Communist-leaning books in the 50's, you could lose your job. I wouldn't want George W. Bush's executive branch to be able to guess my political views by knowing the books I read (if I actually had much time to read books) and I'm sure there are other people who feel the same way about Obama.
Recommend Recommended by 2 Readers

22.
Wyman Rembert
Myrtle Beach, SC
March 12th, 2010
5:16 pm
If, as Suhas (comment at 4:24pm) says, they used movie rankings from IMDB, for example, which the persons ranking the films had used their real names, and matched them to the 'anonymous' rankings from Netflix then I don't see the problem. If the person, Joe Johnson in Suhas' example, used his name to rank the movies on IMDB then who cares if some computer/forensic/statistician genius who've been given the info by Netflix match up Joe Johnson who rated some movies at IMDB with the ratings he gave the movies he rented at Netflix? It isn't like these really, really smart scientists/geeks/nerds/businessmen/whatever you want to call them (some working for AT&T Labs, I think) are going to use this for nefarious purposes. Persons that smart could find out private things about you from computer records a lot more easily than matching up movie recommendations/ratings. And why can't Netflix users opt in or out of the contest? Why can't the teams who have the anonymous data can sign legally binding contracts stating that they won't misuse/sell the data they uncover? After all, aren't these legitimate scientists, mostly in the USA, and not Russian or Chinese hackers out to steal your identity? More info/details are needed to understand

this. Details please, NYTimes. Thanks.
[Recommend](#) Recommended by 0 Readers
[23](#).
[kenbron](#)
Corvallis, Oregon
March 12th, 2010
5:16 pm
How confounding is it, when my son comes home and does a marathon veiwing of streaming dowload content on my Netflix account? Sure, maybe I watch gay content movies, but so what? I am straight and married. I am amazed that Netflix would cancel a second contest, save for the fact that improving their
predictive algoritm at this point may not help their business at all.
[Recommend](#) Recommended by 0 Readers
[24](#).
ex-baller
Albany
March 12th, 2010
5:16 pm
"Just because you're paranoid doesn't mean that someone isn't actually after you." That may or may not be attributable to Woody Allen. But, really, this is all about lawyers looking to make a big score. Pathetic.
[Recommend](#) Recommended by 0 Readers
[25](#).
The Geezer Brigade
Ocean Springs, MS
March 12th, 2010
5:16 pm
Boy, maybe I'm mellowing in my dotage, but I kind of enjoy getting suggestions as to what I might enjoy from outfits like Netflix, Amazon and The Teaching Company. If y'all can figure out who I am from all of that data, so what? The only time I resent being "found out" is when I've been hiding from myself.

Been there. Done that. Real tired of it.
[Recommend](#) Recommended by 1 Readers

| 1 | of 2 [Next](#)

Post a Comment

You are currently logged in as .

Display Name [(What's this)](#)	
Location	
Characters Remaining: 5000	
Comment (Required)	
Submit	

Comments are moderated and generally will be posted if they are on-topic and not abusive. For more information, please see our [Comments FAQ](#).