Dan Fanelli
DATA 643 Special Topics: Recommender Systems
Discussion 4

"Read the article below and consider how to handle attacks on recommender systems. "

This will be a battle that goes on forever, just like the battle of spammers. Once a defense is found, a more sophisticated recommender will try to beat that defense. If I recall, it was Macy's that actually beat google during the holiday season by buying up a bunch of domain names and linking to the Macy's holiday pages to trick google into thinking those were the most popular deals. Some basic thoughts on some strategies:

- Use OpenID as your log on provider - companies like google and facebook are spending millions to make sure that they are not being tricked into thinking bots are people, so use the OAuth that they offer to also be more sure your users are real
- Monitor your data: Log statistics on a regular basis of the landscape of your data relating to the recommendations. In the example of the movie we just read about, the early review dates were the big read flag that this was not legitimate. Other flags might be out of the ordinary high volumes of recommendations in a short span of time or too many users from a given ip address or location casting the votes.

"Can you think of a similar example where a collective effort to alter the workings of content recommendations have been successful?"

- A certain mobile app developer that I "heard about" (I would never do such a thing) may or may not have found sites in the past where you could trade positive reviews of your apps with other developers looking to get an edge for their apps. In this case, very vague reviews that don't mention anything specific about the product could have been the red flag, since people using this scheme would not have the time to actually try the apps before posting the reviews.

"How would you design a system to prevent this kind of abuse?"

- To prevent the abuse mentioned above, I think the proprotion of reviews to total downloads could have told Google Play that these reviews were not legitimate. If the average mobile app gets 5 reviews for every 1000 downloads, but then a bunch of apps start getting a review per download all of a sudden, that would be a pretty clear indicator that the data is not legitimate.