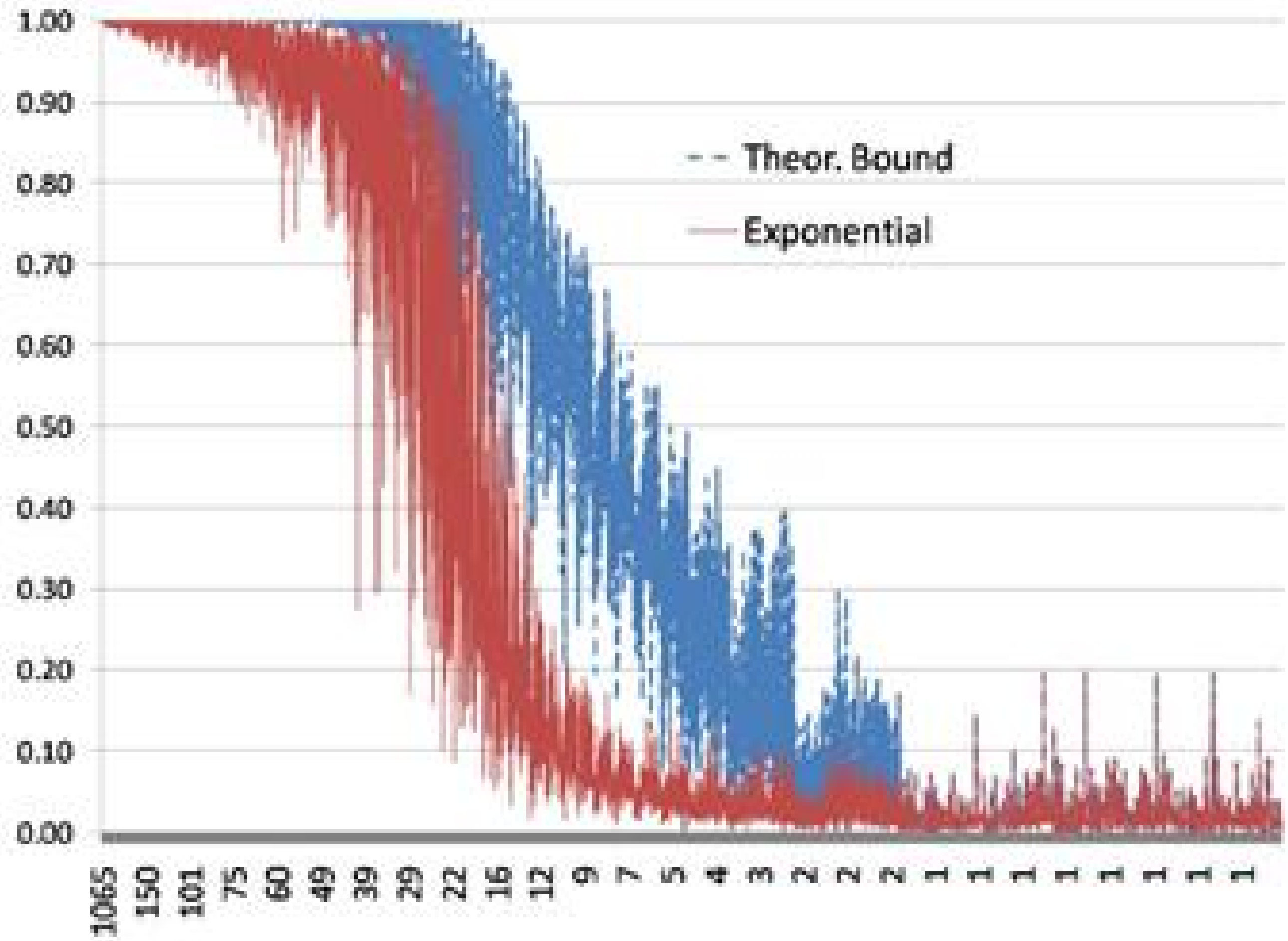
A View from **Emerging Technology from the arXiv**

The Fundamental Limits of Privacy For Social Networks

Using social networks to make recommendations will always compromise privacy, according to a mathematical proof of the limits of privacy.

May 5, 2010

Recommendation engines are among the hottest properties on the web. These sites make make recommendations by mining the pattern of links that crop up in social networks.



Facebook recommends new contacts based on the pattern of connections between existing users, Amazon recommends books and other products based on purchase histories and Netflix recommends movies based historical ratings.

To be sure, these sites produce useful results for users which can dramatically increase sales for a merchant. But they can also compromise privacy. For example, a social network recommendation might reveal that one person has been in email contact with another or that an individual has bought a certain product or watched a specific film. It may even be a breach of privacy to discover that your friend doesn't trust your judgement in books.

In fact there's a long history of privacy controversies associated with social networks. In 2007, Facebook caused a storm by revealing people's purchase history to their friends. At about the same time, a team of researchers de-anonymised a dataset of movie recommendations released by Netflix by comparing it to a publicly available dataset of movie ratings on the Internet Movie Database. And more recently, Google drew a storm of criticism when launching its social network Buzz because it revealed details about people's email network to others.

Today, Aleksandra Korolova at Stanford University with Ashwin Machanavajjhala and Atish Das Sarmait, say that privacy breaches are inevitable when networks are exploited in this way. In fact, they've worked out a fundamental limit to the level of privacy that is possible when social networks are mined for recommendations.

That's quite a task given that there are various different approaches to making recommendations. However, Korolova, Machanavajjhala and Sarmait have come up with a general model that captures the essence of the problem.

Their approach is to consider a general graph consisting of various nodes and the links between them. This may be network in which the nodes are books, say, and a link between two nodes represents the purchase of one book by the owner of another. The team consider all these links to be private information.

Korolova, Machanavajjhala and Sarmait then consider an attacker who wants to work out the existence of a link in the graph from a particular recommendation. So given the knowledge that people who bought book x also bought book y, is it possible to determine a purchase decision made by a specific individual?

To do this, Korolova, Machanavajjhala and Sarmait define a privacy differential as the ratio of the likelihoods that the website makes such a recommendation with the using the private purchase decision in question and without it.

The question they then ask is to what extent can recommendaitons be made while preserving this privacy differential.

It turns out that there is a trade off between the accuracy of the recommendation and the privacy of the network. So a loss of privacy is inevitable for a good recommendation engine.

The group also look at ways of preserving privacy by anonymizing data, for example by adding noise to it. They even compare different privacy preserving algorithms using a dataset of voting patterns on Wikipedia.

The results are not entirely encouraging. the trade off between the accuracy of the recommendations and privacy is always apparent. “This finding throws into serious question the feasibility of developing social recommendation algorithms that are both accurate and privacy-preserving for many real-world settings,” say the team.

That’s a potentially explosive result. But it would be unfair to jump to conclusions too quickly. It’s fair to say that this group’s definition of privacy is enormously strict (as it should be). BUt that makes it all the more important to quantify exactly what privacy issues are at stake in each kind of social network before making a judgement.

That will be a tricky task but one that recommendation engines may be forced to pursue.

Ref: arxiv.org/abs/1004.5600: On the (Im)possibility of Preserving Utility and Privacy in Personalized Social Recommendations