MNM
TEAM
Munich Network Management Team

Ludwig-Maximilians-Universität München

Prof. Dr. D. Kranzlmüller, Dr. N. gentschen Felde

Data Science & Ethics

– The Netflix example I (advanced!) –

Exercise 1: *How To Break Anonymity of the Netflix Prize Dataset*

Try to verify the results of the paper *Robust De-anonymization of Large Sparse Datasets* by Arvind Narayanan and Vitaly Shmatikov[1] in form of an implemenatation!

due date:	21.07.2017 (EOB)
no. of students:	2–3
deliverables:	1. Implemenatation (including source code(s))
	2. Documentation (max. 10 pages)
	3. Presentation (10 – max. 15 minutes)

[1] In proceedings of 29^{th} IEEE Symposium on Security and Privacy, Oakland, CA, May 2008, pp. 111-125. IEEE Computer Society, 2008, `https://www.cs.utexas.edu/~shmat/shmat_oak08netflix.pdf`

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.