

Similarity Ranking in Large-Scale Bipartite Graphs

Alessandro Epasto, Università di Roma "La Sapienza"

This talk addresses the problem of computing similarity rankings in large-scale bipartite graphs. The problem has several real-world applications, including identifying competing advertisers in online advertising system and suggesting related queries or finding users with similar interests.

In these settings bipartite graphs represent the relationships between actors and items and often the items are partitioned into an arbitrary set of categories. We are interested in computing on-the-fly rankings of similar actors, given an actor and an arbitrary subset of categories of interest.

We present a novel algorithmic framework for the efficient computation of several graph-theoretical similarity measures including measures based on random walks.

Among these measures, the most challenging to compute turned out to be the Personalized PageRank (PPR) on which the talk will focus. Our solution makes use of some interesting results from the Markov chain literature to speed up the computation of the PPR stationary distribution in our setting.

Finally, we show some experiments with real-world data including a very large dataset from Google AdWords.

This is a joint work with Jon Feldman, Silvio Lattanzi, Vahab Mirrokni (Google Research) and Stefano Leonardi (Sapienza University of Rome).