Personalized PageRank Dimensionality and Algorithmic Implications (AGSR_21)

Daniel Vial, Vijay Subramanian

Department of Electrical and Computer Engineering, University of Michigan, Ann Arbor, MI

Many systems can be represented as graphs, sets of objects (called *nodes*) and pairwise relations between these objects (called *edges*). These include Internet, which contains websites (nodes) that are connected via hyperlinks (edges); social networks like Twitter, which contains users (nodes) that follow one another (edges); and the human brain, which contains neurons (nodes) that exchange signals through chemical pathways (edges). To study graphs, researchers in diverse domains have used *Personalized PageRank* (PPR). Informally, PPR assigns to each node v a vector π_v, where $\pi_v(w)$ describes the relevance of node w to node v. PPR has proven in many applications. For example, Twitter has used PPR to recommend who users should follow (user v may wish to follow user w if $\pi_v(w)$ is large). Unfortunately, computing all n PPR vectors for a graph of n nodes has complexity $O(n^3)$, which is infeasible for massive graphs arising in modern domains (like Twitter).

In this work, we argue that the situation is not so dire. In particular, we provide an algorithm to estimate all n PPR vectors with bounded error (in the l_1 norm) and with sub-quadratic complexity (i.e. complexity $O(n^c)$ for some $c < 2$). To the best of our knowledge, our scheme improves upon all algorithms found in the literature, the most competitive of which have complexity $O(n^2 log(n))$. We note that our accuracy guarantee holds for any graph, while our complexity guarantee holds for a certain class of graphs. We believe that this class contains realistic models of real-world networks; as an example, we devise a Twitter-like model, which contains a few highly-connected nodes -- modeling celebrities with millions of Twitter followers -- and far more moderately-connected nodes -- modeling “normal” users.

The fundamental reason why our algorithm outperforms existing methods is that, while existing methods estimate each of the n PPR vectors separately, our algorithm exploits the structure of these vectors to estimate them jointly (and thus more efficiently). Specifically, we prove that the effective dimension of the set of PPR vectors vanishes relative to n in our setting; put differently, we show that only a vanishing fraction of these vectors are truly independent. This structural insight allows us to directly estimate only the vanishing fraction of independent vectors, and then to use these vectors to indirectly estimate the others. We believe that this structural insight can used to design more efficient algorithms in other settings, suggesting that our analysis may lead to further advancements.