

MACHINE LEARNING FOR PREDICTING PERFORMANCE OF GRAPH
ALGORITHMS WITH APPLICATIONS IN GRAPH DATABASES
AND NETWORK SIMILARITY

Abstract

by Keyvan Sasani, M.S.
Washington State University
December 2018

Chair: Assefaw H. Gebremedhin

Graphs have long been used in contexts such as social, biological, infrastructural and knowledge networks to represent complex entities and their relationships. The growing use of network modeling urges new methods of computation to overcome the challenges. On the other hand, machine learning has become ubiquitous in solving complex problems, and even helped humans understand semantics of real world by rethinking how information becomes knowledge in an intelligent agent. This thesis studies the effectiveness of machine learning algorithms in solving computationally expensive graph problems.

At first, we propose a general framework for predicting graph query performance with respect to three performance metrics: execution time, query answer quality, and memory consumption. The learning framework generates and makes use of informative statistics from data and query structure and employs a multi-label regression model to predict the multi-metric query performance. We apply the framework to study two common graph query classes—*reachability* and graph *pattern matching*; the two classes differ significantly in their query complexity. For both query classes, we develop suitable models to predict the performance. We demonstrate the efficacy of our framework via experiments on real-world information and social networks. Furthermore, by leveraging the framework, we propose a novel workload optimization algorithm and show that it improves the efficiency of workload management.

The second part of the work is devoted to the problem of tracking graph similarities when the graphs are continuously changing over time. We aim to leverage the temporal information in the graphs to avoid recalculating the similarities from scratch. We propose a new approach, namely approximation of similarity measures in time-evolving graphs where correspondence between the nodes of the two graphs is known. In particular, we use machine learning to build a model to predict the graph similarity by considering only the changes in the graphs. Our experimental results over both real-world and synthetic graphs show the effectiveness and agility of the proposed approaches.