Combinatorial Optimization on Massive Datasets: Streaming, Distributed, and Massively Parallel Computation

Sepehr Assadi
University of Pennsylvania

2/15/2018 at 10:30 am
CoRE 301

Abstract

With the emergence of massive datasets across different application domains, there is a rapidly growing need to solve various optimization tasks over such datasets. To cope with the sheer size of these big data problems, one needs to design algorithms that are highly efficient in their resource usage: for example, their internal memory in case of streaming algorithms, their communication overhead in case of distributed algorithms, and both memory and communication for massively parallel algorithms (such as MapReduce).

In this talk, I will give an overview of my work in this area. As an illustrative example, I will focus on the problem of finding large matchings, namely the maximum matching problem, over massive graphs. Maximum matching is one of the most prominent problems in combinatorial optimization and has important applications in modern big data analysis, e.g., in online advertising. I will describe the previous techniques for designing efficient algorithms for graph optimization problems over massive graphs and prove an inherent limitation of these techniques for designing resource efficient algorithms for maximum matching. I will then describe a general approach for solving the maximum matching problem in a unified way across different models of computation over massive graphs such as streaming, distributed computing, and MapReduce, which simultaneously improves the state-of-the-art in all these models.

Bio

Sepehr Assadi is a PhD student in Computer Science at University of Pennsylvania, advised by Sanjeev Khanna. He is broadly interested in the theoretical foundations of big data analysis and particularly in designing efficient algorithms for combinatorial optimization over massive datasets. His dissertation work has primarily focused on designing algorithms in streaming, distributed, and massively parallel computation (such as MapReduce) models

for fundamental graph optimization problems as well as for submodular function optimization. Sepehr obtained his B.Sc degree from Sharif University of Technology in 2013. He has received the best paper awards at WINE 2015 and SPAA 2017, and the best student paper award at PODS 2017.

Faculty Host: Pranjal Awasthi