Title: Algebraic approaches for coded caching and distributed computing

Abstract: Caching is a popular technique for facilitating large scale content delivery over the Internet. Traditionally, caching operates by storing popular content closer to the end users. Recent work demonstrates that allowing coding in the cache and coded transmission from the server (referred to as coded caching) to the end users can allow for significant reductions in the number of bits transmitted from the server to the end users.

However, the original formulation of the coded caching problem assumes that the server and the end users are connected via a single shared link. In Chapter 2, we consider a more general topology where there is a layer of relay nodes between the server and the users. We propose novel schemes for a class of such networks that satisfy a so-called resolvability property and demonstrate that the performance of our scheme is strictly better than previously proposed schemes. Moreover, the original coded caching scheme requires that each file hosted in the server be partitioned into a large number (i.e., the subpacketization level) of non-overlapping subfiles. From a practical perspective, this is problematic as it means that prior schemes are only applicable when the size of the files is extremely large. In Chapter 3, we propose a novel coded caching scheme that enjoys a significantly lower subpacketization level than prior schemes, while only suffering a marginal increase in the transmission rate. We demonstrate that several schemes with subpacketization levels that are exponentially smaller than the basic scheme can be obtained.

Distributed matrix multiplication is an important problem, especially in domains such as deep learning of neural networks. It is well recognized that the computation times on distributed clusters are often dominated by the slowest workers (called stragglers). Recently, techniques from coding theory have found applications in straggler mitigation in the specific context of matrix-matrix and matrix-vector multiplication. The computation can be completed as long as a certain number of workers (called the recovery threshold) complete their assigned tasks.

In Chapter 4, we consider matrix multiplication under the assumption that the absolute values of the matrix entries are sufficiently small. Under this condition, we present a method with a significantly smaller recovery threshold than prior work. Besides, the prior work suffers from serious numerical issues owing to the condition number of the corresponding real Vandermonde-structured recovery matrices; this condition number grows exponentially in the number of workers. In Chapter 5, we present a novel approach that leverages the properties of circulant permutation matrices and rotation matrices for coded matrix computation. In addition to having an optimal recovery threshold, we demonstrate an upper bound on the worst case condition number of our recovery matrices grows polynomially in the number of workers.