Column subset selection from partially observed data

The problem of matrix column subset selection deals with selecting a subset of columns from an input matrix such that the input can be well approximated by the span of the selected columns. This is an instantiation of prototype or feature selection in unsupervised setting, and has received much attention from CS theory community in the fully observed case. However, In many applications the complete data matrix is unavailable or we have choice of which selective entries to observe. In this talk, I will present the first provably correct column subset selection algorithms for partially observed data matrices with both additive and relative error guarantees. I will also present some open questions, comparison with matrix completion, as well as detailed empirical comparisons with some heuristic methods. The methods exhibit different merits and drawbacks in terms of statistical accuracy, computational efficiency, sample complexity and model assumptions, which provides a nice exploration of the tradeoff between these desired properties for column subset selection.