


Entry-Specific Matrix Estimation Under Arbitrary Sampling Patterns Through the Lens of Network Flows

Yudong Chen 

Department of Computer Sciences, University of Wisconsin-Madison, WI, USA

Xumei Xi 

School of Operations Research and Information Engineering, Cornell University, Ithaca, NY, USA

Christina Lee Yu 

School of Operations Research and Information Engineering, Cornell University, Ithaca, NY, USA

Abstract

Matrix completion tackles the task of predicting missing values in a low-rank matrix based on a sparse set of observed entries. It is often assumed that the observation pattern is generated uniformly at random or has a very specific structure tuned to a given algorithm. There is still a gap in our understanding when it comes to arbitrary sampling patterns. Given an arbitrary sampling pattern, we introduce a matrix completion algorithm based on network flows in the bipartite graph induced by the observation pattern. For additive matrices, we show that the electrical flow is optimal, and we establish error upper bounds customized to each entry as a function of the observation set, along with matching minimax lower bounds. Our results show that the minimax squared error for recovery of a particular entry in the matrix is proportional to the effective resistance of the corresponding edge in the graph. Furthermore, we show that the electrical flow estimator is equivalent to the least squares estimator. We apply our estimator to the two-way fixed effects model and show that it enables us to accurately infer individual causal effects and the unit-specific and time-specific confounders. For rank-1 matrices, we use edge-disjoint paths to form an estimator that achieves minimax optimal estimation when the sampling is sufficiently dense. Our discovery introduces a new family of estimators parametrized by network flows, which provide a fine-grained and intuitive understanding of the impact of the given sampling pattern on the difficulty of estimation at an entry-specific level. This graph-based approach allows us to quantify the inherent complexity of matrix completion for individual entries, rather than relying solely on global measures of performance.

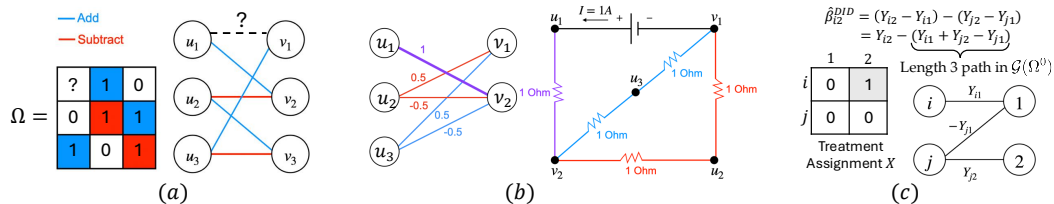



Figure 1 (a) Example of transforming a matrix observation pattern to a bipartite graph. Each a path from u_1 to v_1 defines an unbiased estimator for the ? entry. (b) Example of a network flow, in particular the electrical flow that corresponds to the optimal unbiased estimator. This flow sends a unit of current from u_1 to v_1 on the electrical network constructed by treating each edge in the graph as a unit resistor, thus putting high weight on edges that appear on multiple paths. (c) Equivalence of Difference-in-differences estimators to flow estimators constrained to length-3 paths.

2012 ACM Subject Classification Networks → Network algorithms

Keywords and phrases Matrix estimation, network flows, panel data analysis, non-uniform sampling

Digital Object Identifier 10.4230/LIPIcs.ITCS.2025.36

 © Yudong Chen, Xumei Xi, and Christina Lee Yu; licensed under Creative Commons License CC-BY 4.0
16th Innovations in Theoretical Computer Science Conference (ITCS 2025).
Editor: Raghu Meka; Article No. 36; pp. 36:1–36:2



Leibniz International Proceedings in Informatics
Schloss Dagstuhl – Leibniz-Zentrum für Informatik, Dagstuhl Publishing, Germany

36:2 Entry-Specific Matrix Estimation

Category Extended Abstract

Related Version *Full Version*: <https://www.arxiv.org/pdf/2409.03980>

Funding National Science Foundation grants CCF-1704828, CCF-2233152, CCF-2337796 and CNS-1955997, and AFOSR grant FA9550-23-1-0301