

COLLOQUIUM

MATHEMATICS AND STATISTICS
QUEEN'S UNIVERSITY



COUNTING MOTIFS AND CONNECTED COMPONENTS OF LARGE GRAPHS VIA SUBGRAPH SAMPLING

Abstract. Learning properties of large graphs from samples is an important problem in statistical network analysis. We revisit the problem of estimating the numbers of connected components in a graph of N vertices based on the subgraph sampling model, where we observe the subgraph induced by n vertices drawn uniformly at random. The key question is whether it is possible to achieve accurate estimation, i.e., vanishing normalized mean-square error, by sampling a vanishing fraction of the vertices. We show that it is possible by accessing only sublinear number of samples if the graph does not contain high-degree vertices or long induced cycles; otherwise it is impossible. Optimal sample complexity bounds are obtained for several classes of graphs including forests, cliques, and chordal graphs. The methodology relies on topological identities of graph homomorphism numbers, which, in turn, also play a key role proving minimax lower bounds based on constructing random instances of graphs with matching structures of small subgraphs. We will also discuss results for the neighborhood sampling model, where we observe the edges between the sampled vertices and their neighbors.

Jason Klusowski (Yale University)

Jason M. Klusowski obtained his B.Sc. in Mathematics and Statistics in 2013 from the University of Manitoba, receiving the *Robert Ross McLaughlin Scholarship in Mathematics*, the *Faculty of Science Medal in B.Sc.*, and the *Governor General's Silver Medal*. In 2013 he was received the *NSERC Alexander Graham Bell Canada Graduate Scholarship* and joined Yale University where he is currently a Ph.D. candidate in Statistics and Data Science, under the supervision of Prof. Andrew Barron. Jason Klusowski's research interests include the theoretical and computational aspects of neural networks, approximation algorithms for networks, high-dimensional function estimation, mixture models, and shape constrained estimation.

234 JEFFERY HALL
3:30pm · NOVEMBER 29 · 2017