

Goodness of fit for log-linear network models: dynamic Markov bases using hypergraphs

Elizabeth Gross¹ · Sonja Petrović² ·
Despina Stasi²

Received: 27 January 2015 / Revised: 4 February 2016 / Published online: 5 April 2016
© The Institute of Statistical Mathematics, Tokyo 2016

Abstract Social networks and other sparse data sets pose significant challenges for statistical inference, since many standard statistical methods for testing model/data fit are not applicable in such settings. Algebraic statistics offers a theoretically justified approach to goodness-of-fit testing that relies on the theory of Markov bases. Most current practices require the computation of the entire basis, which is infeasible in many practical settings. We present a dynamic approach to explore the fiber of a model, which bypasses this issue, and is based on the combinatorics of hypergraphs arising from the toric algebra structure of log-linear models. We demonstrate the

E. Gross is supported by the NSF Postdoctoral Research Fellowship, NSF award #DMS-1304167. S. Petrović and D. Stasi acknowledge partial support from Grants #FA9550-12-1-0392 and #FA9550-14-1-0141 from the U.S. Air Force Office of Scientific Research (AFOSR) and the Defense Advanced Research Projects Agency (DARPA). Some of the computations were performed on a cluster provided by an NSF-SCREMS Grant to IIT. Part of this work was completed while D. Stasi was a postdoc at Pennsylvania State University Statistics Department.

Electronic supplementary material The online version of this article (doi:[10.1007/s10463-016-0560-2](https://doi.org/10.1007/s10463-016-0560-2)) contains supplementary material, which is available to authorized users.

✉ Sonja Petrović
Sonja.Petrovic@iit.edu

Elizabeth Gross
elizabeth.gross@sjsu.edu

Despina Stasi
despina.stasi@gmail.com

¹ Department of Mathematics and Statistics, San José State University, One Washington Square, San Jose, CA 95192, USA

² Department of Applied Mathematics, Illinois Institute of Technology, Rettaliata Engineering Center room 208, 10 West 32nd Street, Chicago, IL 60616, USA

approach on the Holland–Leinhardt p_1 model for random directed graphs that allows for reciprocation effects.

Keywords Algebraic statistics · Markov basis · Hypergraph · Toric ideal · Contingency table · Network model · Random graph · Sampling algorithm