

The public reporting burden for this collection of information is estimated to average 1 hour per response, including the time for reviewing instructions, searching existing data sources, gathering and maintaining the data needed, and completing and reviewing the collection of information. Send comments regarding this burden estimate or any other aspect of this collection of information, including suggestions for reducing this burden, to Washington Headquarters Services, Directorate for Information Operations and Reports, 1215 Jefferson Davis Highway, Suite 1204, Arlington VA, 22202-4302. Respondents should be aware that notwithstanding any other provision of law, no person shall be subject to any penalty for failing to comply with a collection of information if it does not display a currently valid OMB control number.  
PLEASE DO NOT RETURN YOUR FORM TO THE ABOVE ADDRESS.

1. REPORT DATE (DD-MM-YYYY) 27-10-2021	2. REPORT TYPE Final Report	3. DATES COVERED (From - To) 3-Jul-2018 - 1-Sep-2021
---	--------------------------------	---

4. TITLE AND SUBTITLE Final Report: Statistical Methods for Percolation in Practice: Random Graph Hidden Markov Models	5a. CONTRACT NUMBER W911NF-18-1-0237
	5b. GRANT NUMBER
	5c. PROGRAM ELEMENT NUMBER 611102

6. AUTHORS	5d. PROJECT NUMBER
	5e. TASK NUMBER
	5f. WORK UNIT NUMBER

7. PERFORMING ORGANIZATION NAMES AND ADDRESSES Boston University Office of Sponsored Program 881 Commonwealth Avenue Boston, MA 02215 -1300	8. PERFORMING ORGANIZATION REPORT NUMBER
---	--

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS (ES) U.S. Army Research Office P.O. Box 12211 Research Triangle Park, NC 27709-2211	10. SPONSOR/MONITOR'S ACRONYM(S) ARO
	11. SPONSOR/MONITOR'S REPORT NUMBER(S) 71321-CS.1

12. DISTRIBUTION AVAILABILITY STATEMENT Approved for public release; distribution is unlimited.
--

13. SUPPLEMENTARY NOTES The views, opinions and/or findings contained in this report are those of the author(s) and should not be construed as an official Department of the Army position, policy or decision, unless so designated by other documentation.
---

14. ABSTRACT
--------------

15. SUBJECT TERMS
-------------------

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	15. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
a. REPORT UU	b. ABSTRACT UU	c. THIS PAGE UU	UU		Eric Kolaczyk
					19b. TELEPHONE NUMBER 617-353-5208

# RPPR Final Report

## as of 20-Dec-2021

Agency Code: 21XD

Proposal Number: 71321CS

Agreement Number: W911NF-18-1-0237

### INVESTIGATOR(S):

**Name:** Eric Kolaczyk  
**Email:** kolaczyk@bu.edu  
**Phone Number:** 6173535208  
**Principal:** Y

Organization: **Boston University**

Address: Office of Sponsored Program, Boston, MA 022151300

Country: USA

DUNS Number: 049435266

EIN: 042103547

**Report Date:** 01-Dec-2021

Date Received: 27-Oct-2021

**Final Report** for Period Beginning 03-Jul-2018 and Ending 01-Sep-2021

**Title:** Statistical Methods for Percolation in Practice: Random Graph Hidden Markov Models

**Begin Performance Period:** 03-Jul-2018

**End Performance Period:** 01-Sep-2021

**Report Term:** 0-Other

Submitted By: Eric Kolaczyk

Email: kolaczyk@bu.edu

Phone: (617) 353-5208

**Distribution Statement:** 1-Approved for public release; distribution is unlimited.

### STEM Degrees:

### STEM Participants:

**Major Goals:** The goal of this program of research is to develop a class of random graph hidden Markov models (RG-HMMs) for characterizing percolation in noisy, dynamically evolving networks, as well as a set of procedures for statistical estimation and hypothesis testing with these models, in a computationally scalable manner.

Specifically, our goals are the following:

1. Develop a complete framework for likelihood-based inference under HMM variants of network growth models.
2. Extend that framework to the practically relevant context where (a) the duration spent in states varies, and (b) we observe only a longitudinal subsampling of time points.
3. Achieve computationally efficient scaling of these frameworks to large networks through the adaptation of recent developments in spectral and method-of-moments techniques for high-dimensional HMMs.

In addition, we explored several major directions related to individual aspects of our RG-HMM setting, including dynamic random graph models with attractors and the impact of noise on various network tasks (e.g., subgraph counting, estimating  $R_0$  in networks, etc.).

**Accomplishments:** With respect to our first three goals:

- We have developed a comprehensive framework for likelihood-based parameter estimation and hypothesis testing under both the initial version of our proposed HMM (with equi-spaced time points) and the extended version (where, more realistically, the duration spent in states varies and we observe only a longitudinal subsampling of time points). These RG-HMM models are developed using variants of two network growth models – classical Erdos-Renyi and the so-called product-rule due to Achlioptas and colleagues (as an example of so-called ‘explosive’ percolation). We develop an extension of the classical EM algorithm for HMMs, with use of MCMC (based on the missing-information principle) and particle filtering to approximate requisite distributions and moments in a stochastic manner. Implementation is in the R statistical software environment.
- Computational complexity of the necessary likelihood-based calculations for statistical inference has been characterized. Code has been reasonably optimized, in the sense of making use of best-practice in data structures, parallelization, and the like. For additional speedups, one could implement approximations to certain key random sampling steps in the algorithm (e.g., randomly selecting an edge to delete in a dense network). Recent work by Alon and others in this space suggest that sublinear algorithms can be used for speed up at the cost of some sacrifice in distributional accuracy.
- Simulations, on networks of 30-100 nodes observed over a similar number of time points, indicate it is possible to estimate the underlying model parameters quite accurately. Several nuances regarding scaling of percolation laws, and our ability to accurately discriminate between laws, are explored in detail.
- We have conducted an in-depth case study of the methodology with high-frequency EEG data from subjects with

## RPPR Final Report as of 20-Dec-2021

epilepsy. This means an automated construction of multivariate time series data from EEG measurements, an automated extraction of percolation subseries, automated parameter estimation, and, finally, hypothesis testing comparing standard percolation versus a form of “explosive” percolation. The results of this analysis suggest that focal versus generalized epileptic seizures may potentially display explosive versus classical percolation, respectively. Additional analysis with additional seizures would be needed to confirm this relationship.

In addition, with respect to major directions related to individual aspects of our RG-HMM setting, we accomplished the following:

- We have pursued develop of a new class of dynamic networks where, instead of the relatively simple Markov-like versions of simple percolation models in our hidden Markov layer, we instead use latent position models. Motivated in part by the type of “clumping” that can occur in epilepsy during seizure onset, we have extended the usual approach to latent layer modeling to incorporate the notion of “attractors” (and repulsion, as the antithesis of attraction) from dynamic systems. These latent, dynamic network models with attractors have been explored by our group in simulation, and estimation procedures designed and implemented for parameter estimation. With an emphasis on better understanding their value-added when multiple forces of attraction / repulsion are present, we have explored their application to Twitter and reddit data in the context of affective partisanship in US politics. (This research will be continued under recently awarded NSF funding.)
- Our HMMs purposefully have a noise layer (i.e., dynamic random graphs but observed in noisy form). We have pursued several related projects seeking to better understand the implications on downstream network science tasks of having noisy network data, rather than non-noisy data. These tasks include counting of subgraphs, network recovery, estimation of network-based epidemic reproduction numbers (i.e.,  $R_0$ ), and causal inference under network interference. For most of these projects, a generic signal+noise model was employed. However, we have also explored briefly the specific context in which noise is incurred due to record linkage in large databases.

**Training Opportunities:** In the first year of this award, one graduate student was been supported on this work during the academic year and another during the summer. Specifically, Ms. Xiaojing ZHU contributed the work on elements 1 through 5 above. Ms. Wenrui LI contributed the work on element 6 above.

In the second year of this award, three graduate students were supported at various levels (all female). Ms. Xiaojing ZHU contributed to elements 1-4 and 6 above, Ms. Wenrui Li contributed to element 6 above. And Ms. Shurong LIN contributed to element 7 above.

In the third year of this award, two graduate students were supported at various levels (both female). Ms. Xiaojing ZHU contributed to elements 1, 2, 5, and 6 above, while Ms. Shurong LIN contributed to element 7.

## RPPR Final Report as of 20-Dec-2021

**Results Dissemination:** The following manuscripts are available, as well as the corresponding software.

- “Network recovery from unlabeled noisy samples” (2021). Josephs, N., Li, W., and Kolaczyk, E.D. Proceedings of the Asilomar Conference on Signals, Systems, and Computers. [Manuscript: arXiv:2104.14952; GitHub Site: KolaczykResearch/NetworkRecovery]
- “Inferring the type of phase transitions undergone in epileptic seizures using random graph hidden Markov models for percolation in noisy dynamic networks” (2021). Zhu, X., Shappell, H. Kramer, M.A., Chu, C.J., and Kolaczyk, E.D. [Manuscript: arXiv:2102.11948; GitHub Site: KolaczykResearch/RG-HMM]
- “Disentangling positive and negative partisanship in affective polarization using a coevolving latent space network with attractors model” (2021). Zhu, X., Caliskan, C., Christenson, D.P., Spiliopoulos, K., Walker, D., and Kolaczyk, E.D. [Manuscript: arXiv:2109.13129v2; GitHub Site: KolaczykResearch/CLSNA-2Party-Polarization]
- “Causal inference under network interference with noise” (2021) Li, W., Sussman, D.L., and Kolaczyk, E.D. [Manuscript: arXiv:2105.04518; GitHub Site: KolaczykResearch/CausInfNoisyNet]
- “Estimation of local time-varying reproduction numbers in noisy surveillance data” (2021). Li, W., Bulekova, K., Gregor, B., White, L.F., and Kolaczyk, E.D. [medRxiv, doi: <https://doi.org/10.1101/2021.04.23.21255958>; GitHub Site: KolaczykResearch/EstimLocalRt]

In addition, the following manuscript has been accepted for publication in the American Journal of Public Health, but is currently under embargo until its scheduled publication date of January 2022: “Projecting Quarantine Utilization During a Pandemic”, Li, W., Kolaczyk, E.D., and White, L.F. Both this article and the last one above were motivated by research needs emerging during the early stages of the pandemic.

The following invited talks were given to date on topics relating, in whole or in part, to the work on this award:

“Random Graph Hidden Markov Models for Percolation in Noisy Dynamic Networks”?Xiaoqing Zhu. Joint Statistical Meetings, Denver, Colorado. July 2019.

“Causal Inference Under Network Interference with Noise”. Wenrui Li. Joint Statistical Meetings, Denver, Colorado. July 2019.

“Dynamic Latent Space Network Models with Attractors for Flocking and Polarization”. Xiaoqing Zhu. Joint Statistical Meetings, Philadelphia, PA. August, 2020. (virtual)

“Quantitative Methods for Understanding Coalescence and Fragmentation in Dynamics Networks of Epileptic Seizures.” Eric Kolaczyk. Data Science in Social and Behavioral Sciences Workshop. Statistical and Applied Mathematical Sciences Institute (SAMSI). January, 2021.

“Why Aren’t Network Statistics Accompanied by Uncertainty Statements?” Eric Kolaczyk. TRIPODS Winter School & Workshop on Graph Learning and Deep Learning. Mathematical Institute for Data Science, Johns Hopkins University. January, 2021.

“Causal Inference Under Network Interference with Noise.” Eric Kolaczyk. 2021 Conference on Econometrics. Cowles Foundation, Yale University. June, 2021. (virtual)

“Quantitative methods for understanding coalescence and fragmentation in dynamic networks of epileptic seizures.” Eric Kolaczyk. Satellite Workshop on Nonstandard Brain Image Analysis, Organization for Human Brain Mapping. June, 2021. (virtual)

“Accounting for Network Noise: Counting, Experimenting, and Epidemic Control.” Eric Kolaczyk. Statistical Inference for Network Models, Satellite Workshop of Networks 2021. June, 2021. (virtual)

“Quantitative methods for understanding coalescence and fragmentation in dynamic networks of epileptic seizures.” Eric Kolaczyk. International Statistical Institute, World Statistics Conference. July, 2021. (virtual)

## RPPR Final Report as of 20-Dec-2021

“Accounting for Network Noise: Counting, Experimenting, and Epidemic Control.” Eric Kolaczyk. Joint Statistical Meetings. August, 2021. (virtual)

“On the Asymptotics of Temporal Motif Estimation via Sampling”. Xiaojing Zhu. Joint Statistical Meetings. August, 2021. (virtual)

**Honors and Awards:** A grant to expand upon the work on dynamic latent space network models with attractors has been awarded (for 3 years) by the National Science Foundation. (“Understanding Social Dynamics Through Coevolving Latent Space Networks With Attractors”, NSF SES-2120115, Kolaczyk PI).

The PI received the 2020 Mosteller Statistician of the Year award from the Boston Chapter of the American Statistical Association.

The PI was named director of the Rafik B. Hariri Institute for Computing and Computational Science and Engineering, Boston University’s computing and data science institute.

### Protocol Activity Status:

**Technology Transfer:** Nothing to Report

### PARTICIPANTS:

**Participant Type:** PD/PI

**Participant:** Eric Kolaczyk

**Person Months Worked:** 1.00

Project Contribution:

National Academy Member: N

**Funding Support:**

**Participant Type:** Graduate Student (research assistant)

**Participant:** Xiaojing Zhu

**Person Months Worked:** 8.00

Project Contribution:

National Academy Member: N

**Funding Support:**

**Participant Type:** Graduate Student (research assistant)

**Participant:** Shurong Lin

**Person Months Worked:** 2.00

Project Contribution:

National Academy Member: N

**Funding Support:**

**Participant Type:** Graduate Student (research assistant)

**Participant:** Wenrui Li

**Person Months Worked:** 2.00

Project Contribution:

National Academy Member: N

**Funding Support:**

**RPPR Final Report**  
as of 20-Dec-2021

**Partners**

,

I certify that the information in the report is complete and accurate:

Signature: Eric D. Kolaczyk

Signature Date: 10/27/21 5:22PM

"Nothing to report in the uploaded pdf (see accomplishments)"