AFRL-AFOSR-VA-TR-2019-0226

Statistical Foundations for Measurement-based System Verification in Complex Networks

Eric Kolaczyk
TRUSTEES OF BOSTON UNIVERSITY

08/02/2019
Final Report

Air Force Research Laboratory
AF Office Of Scientific Research (AFOSR)/ RTA2
Arlington, Virginia 22203
Air Force Materiel Command

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

1. REPORT DATE *(DD-MM-YYYY)*
07/24/2019

2. REPORT TYPE
FINAL

3. DATES COVERED *(From - To)*
06/15/2012 - 01/31/2018

4. TITLE AND SUBTITLE
"Statistical Foundations for Measurement-Based System Verification in Complex Networks"

5a. CONTRACT NUMBER

5b. GRANT NUMBER
FA9550-12-1-0102

5c. PROGRAM ELEMENT NUMBER

6. AUTHOR(S)
DR. ERIC D KOLACZYK

5d. PROJECT NUMBER

5e. TASK NUMBER

5f. WORK UNIT NUMBER

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)
TRUSTEES OF BOSTON UNIVERSITY
25 BUICK STREET, SUITE 200
BOSTON MA 02215-1300

8. PERFORMING ORGANIZATION REPORT NUMBER

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)
AF OFFICE OF SCIENTIFIC RESEARCH
875 RANDOLPH STREET
ARLINGTON, VA22203

10. SPONSOR/MONITOR'S ACRONYM(S)
DOD/AFOSR

11. SPONSOR/MONITOR'S REPORT NUMBER(S)

12. DISTRIBUTION/AVAILABILITY STATEMENT

13. SUPPLEMENTARY NOTES

14. ABSTRACT
Complex networks of various types are central to the current and future mission readiness of the Air Force. Particularly important are communication networks, which can span multiple temporal and spatial scales and connect a variety of human, semi-autonomous, and autonomous agents, each playing an important role towards the overall mission success. For successful deployment of such networks, a reliable paradigm for measurement and system verification is critical. At the core of such a paradigm must be an appropriate statistical methodology, for everything from network sampling, to network characterization, to modeling, inference, and prediction of network-oriented parameters associated with performance and, ul

15. SUBJECT TERMS

16. SECURITY CLASSIFICATION OF:			**17. LIMITATION OF ABSTRACT**	**18. NUMBER OF PAGES**	**19a. NAME OF RESPONSIBLE PERSON**
a. REPORT	**b. ABSTRACT**	**c. THIS PAGE**			Eric Kolaczyk
					19b. TELEPHONE NUMBER *(Include area code)* 617-353-5208

Final Report

Grant Title: Statistical Foundations for Measurement-based System Verification in Complex Networks
Grant Number: FA9550-12-1-0102
Principle Investigator: Eric D. Kolaczyk, Department of Mathematics & Statistics, Boston University
Date: July 25, 2019

Primary Goals:

Complex networks of various types are central to the current and future mission readiness of the Air Force. Particularly important are communication networks, which can span multiple temporal and spatial scales and connect a variety of human, semi-autonomous, and autonomous agents, each playing an important role towards the overall mission success. In order to preserve such network-based information infrastructures (and, more generally, the structure and integrity of the information itself traversing these networks), questions of their design and robustness are critical. But upon deployment, a corresponding paradigm of measurement and system verification becomes equally – if not more – critical. And with this need for measurement in complex mobile communication networks comes the need for an appropriate statistical methodology, for everything from sampling, to characterization, to modeling, inference, and prediction of network-oriented parameters associated with performance and, ultimately, risk. However, statistics for many of the most basic and fundamental tasks, while developed to an arguably mature state in now-canonical contexts like signal and image processing, data mining, etc., are still comparatively immature in the context of complex networks. In fact, tools for even some of the seemingly most innocuous of statistical tasks, such as equipping summary statistics (e.g., density, clustering coefficients, etc.) of measured networks with confidence intervals, are almost non-existent for complex networks.

Accordingly, under this award we pursued a broad-based, multi-faceted program of research to systematically lay key pieces of the statistical foundation necessary to pursue measurement-based systems validation in complex networks. This included work on estimation under network sampling, propagation of uncertainty in network summary statistics, robust dynamic community detection, network topology inference, and network averaging under repeated sampling.

Major Accomplishments Under Goals:

While a broad range of results were obtained under this research program, much of these can be grouped under the following five main technical categories.

1. *Estimation under sampling of networks*. Under a variety of canonical network sampling plans, recovery of the degree distribution of the original, underlying network reduces to a linear inverse problem. For typical designs of interest, this inverse problem is ill-posed. Hence estimation is challenging. Nevertheless, setting up the problem as a constrained, complexity-penalized weighted least-squares problem, we have demonstrated [Pub#15] that it is possible to produce surprisingly accurate estimates of the degree distribution -- even under rather extreme levels of sampling (e.g., induced subgraph sampling from only 10% of the vertices in a network). A similar problem is that of estimating individual network degrees, i.e., estimating the degree sequence. For the specific case of induced subgraph sampling, we have shown [Pub#17] that knowledge of the network degree distribution (e.g., through an estimation procedure like the

above) can significantly enhance estimation accuracy (over the naïve approach of inverse probability weighting) through a Bayesian procedure. Finally, we have conducted initial exploration of the theoretical characterization of the degree distribution estimation problem [Dissertation #3].

2. *Propagation of uncertainty in noisy networks*. Uncertainty in the presence / absence status of edges in a network necessarily must propagate to uncertainty in higher-level statements about the network, such as subgraph counts and other related network summary statistics. Nevertheless, despite the ubiquity of network summary statistics in empirical research, our understanding of the nature of this propagation remains quite rudimentary. Our work has advanced this understanding in several fundamental ways. First, under variants of a certain graph-based `signal-plus-noise' model, we established a set of central limit theorems for the distribution of subgraph counts [Pubs#10,12]. Specifically, we showed that for low levels of noise in the limit of large networks, these counts tend in distribution to a so-called Skellum distribution – that is, a two-tailed analogue of the Poisson distribution. Second, under the same model, we showed [Pub#4] that while (a) it is impossible to simultaneously estimate both the network edge density (for example) and unknown edge noise levels, nevertheless (b) it is possible to do such estimation with as little as three noisy replicates of a network. More generally, we offer estimators for certain arbitrary network subgraph densities. Confidence intervals for these estimates may be produced, using asymptotic normality and a novel network-bootstrap algorithm. Empirical results show that the effects of `Type II' error bias (i.e., not observing edges actually present in the true network) can be quite pronounced in the context of gene coexpression networks. While these are parametric approaches, we have also explored nonparametric approaches in the spirit of wavelet-based denoising [Pub#16]. We have also done empirical work that explores the impact of noise on network percolation [Pub#13]. We have demonstrated that while network noise can easily inhibit our ability to distinguish between two starkly different percolation regimes (e.g., classical percolation versus `explosive' percolation) using the standard measure of the growth of the largest connected component as a metric, we can nevertheless achieve decidedly better results by tracking the *second* largest connected component.
3. *Dynamic community detection with robustness to noise*. One of the most common of higher-level network analysis tasks is graph partitioning, aka `community detection'. The detection of communities in networks that are dynamically changing over time is particularly challenging. And, again, the role of `noise' in the presence/absence status of edges in the observed network is poorly understood. We have developed [Pub#5 and Dissertation #6] a method of dynamic community detection that uses the notion of a plex (i.e., a small network in which each of, say, m nodes is connected to at least, say, k of the others) as the local element of larger communities, and defines such communities as the extent to which such a plex can be `walked' in a certain way throughout the network. This approach is, by design, naturally robust to `Type II' error in the observed edges. The method substantially outperforms two popular and representative competitors (i.e., the clique percolation method of Palla et al. and the multilayer modularity method of Mucha et al.). This improvement, in turn, appears to provide decided advantages in the study of dynamic network representations of epileptic seizures.
4. *Network topology inference*. One of the primary sources of uncertainty in network topology is the fact that many of the most commonly used networks are necessarily inferred from more fundamental measurements. A canonical example of such is the class of association networks, i.e., where an edge between two nodes in the network indicates some sufficiently strong level of association between node characteristics (e.g., based on correlation, mutual information, etc.). As such, it is important to both develop methods of network topology inference for which we

understand their properties, and – more ambitiously – to study the impact of inferred topology on how other observations of network behavior or function behave. Under this award, we made fundamental contributions of both types. In the context of network topology inference, we have developed two new approaches to inferring the topology of dynamic networks that connect to the notion of dynamic Granger causality. The first blends network models with traditional multi-scale modeling to produce a method of dynamic topology inference that is robust to changepoints [Pub#5]. The impact of this new approach has been demonstrated in the context of cognitive task analysis through network-based representations of MRI data. We also provide theoretical guarantees characterizing the accuracy of network inference. The second blends network models with the notion of latent factor modeling, in a Bayesian context, and is illustrated in the context of multivariate financial time series analysis [Pub#6]. In contrast, we have also studied in a number of contexts the problem of how best to infer the source of diffusions propagating across a network [Pub#3, 9, and 14]. The approaches are quite distinct and allow different strengths to be leveraged. For example, in [Pub#14] we provide a multi-level model of network diffusion observations in a Bayesian context that integrates multiple sources of biological information with latent subspace structure, to produce estimates of diffusion source with probabilistic weight of evidence. On the other hand, in [Pub#9], we use methods from high-dimensional penalized regression to infer a `perturbed mean' vector whose entries can be ranked to prioritize putative source locations. In this setting, we are able to precisely characterize the impact of network topology uncertainty on the results. Finally, in [Pub#3], we combine multivariate Gaussian mixture models and stochastic differential equations, in the spirit of data assimilation, to predict diffusion sources using first-arrival time data. This is illustrated in the context of predicting cholera sources in developing areas.

5. *Average behavior of networks under many noisy samples*. Mirroring the traditional perspective in complex network analysis, the majority of the work described above pertains to the context in which one or a few (typically large) networks are observed. But in recent years, a nontrivial body of work is beginning to emerge around a complementary context, i.e., that in which we observe a large number of (possibly small) networks. This setting is still quite undeveloped, and our group has laid seminal groundwork on a number of problems. In particular, our work is directed at developing the analogue of `Statistics 101' types of tools for an era where we have data sets of network-data objects – that is, where each data point is in fact an entire network. We have focused in particular on the canonical problem of defining a sample mean, characterizing its behavior under sampling, and using that characterization to inform decision making (e.g., through hypothesis testing). In our initial work [Pub#11], we characterized the geometry of the space of all labeled networks, associated with it a nonparametric notion of average based on the corresponding natural distance, established a central limit theorem, and provided a hypothesis test with asymptotic chi-square behavior for determining, say, differences in network averages between groups. The utility of this approach was demonstrated in the context of computational neuroscience, where our approach can provide substantially more power than the naïve approach that compares individual vertex pairs across networks one by one. Recently, we have extended this work to the case of unlabeled networks [Pub#8]. There the mathematics is considerably more involved, since the lack of vertex labels translates to necessarily analyzing the previous geometric space with the permutation group modded out. Nevertheless, a strong law limit can be established for `mean sets', and a central limit theorem, for distributions satisfying certain more stringent conditions. This work has direct relevance to the problem of working with data sets of networks under privacy restrictions.

Training Opportunities:

Seven PhD graduate students at Boston University received training during the reporting period (4 women, 3 men). Lisa Pham worked on network-based methods to identify sources of propagation along a network in the context of noisy biological network data. She is currently a managing partner in a hedge fund. Weston Viles worked on (i) methods of dynamic community detection displaying robustness to noise, and (ii) characterizing propagation of noise in network topology to network subgraph counts. He is currently an assistant professor at the University of Southern Maine. Yaonan Zhang worked on estimation of degree distributions from sampled networks. She is currently a Quantitative Finance Researcher at State Street Global. Paula Griffin worked on identification of sources of diffusion propagation along a network in the context of noisy biological network data. She is currently a Group Product Manager at Quora. Aleksandrina Goeva also worked on the estimation of degree distributions from sampled networks. She is currently a Postdoctoral Fellow at the Broad Institute. Jun Li worked on statistical inference problems where we have many (noisy) networks. He is currently a Data Scientist at Google. Finally, Xinyu Kang worked on bootstrapping in networks and also multi-scale analysis of dynamic network time series. He is currently a Data Scientist at New York Life Insurance Company.

In addition, three postdoctoral fellows were supported on this award (one under-represented minority). Cedric Ginestet worked on statistical inference problems where we have many (noisy) networks. He is currently a Statistician with Kings College London. Apratim Ganguly worked on estimation of individual node degrees under network sampling, as well as multi-scale analysis of dynamic network time series data. He is currently a Quantitative Analyst at Google. Finally, Daniel Ahelegbey worked on dynamic latent space models for multivariate time series data. He is currently an assistant professor at the University of Pavia, Italy.

Results Dissemination

Books Published:

1. Kolaczyk, E.D. (2017). Topics at the Frontier of Statistics and Network Analysis: (Re)Visiting the Foundations. Cambridge University Press.
2. Kolaczyk, E.D. and Csardi, G. (2014). *Statistical Analysis of Network Data with R.* Springer, New York.

Papers Under Revision/Review:

3. Li, J., Manitz, J., Bertuzzo, E., and Kolaczyk, E.D. (2019). Sensor-based localization of epidemic sources on human mobility networks. Under review at *PLoS Computational Biology*.
4. Chang, J., Kolaczyk, E.D., and Yao, Q. (2019) Estimation of subgraph density in noisy networks. (arxiv:1803.02488; submitted to the *Journal of the American Statistical Association*.
5. Martinet, L.-E., Kramer, M.A., Viles, W., Perkins, N., Spencer, E., Chu, C.J., Cash, S.S., and Kolaczyk, E.D. (2018). Robust dynamic community tracking with applications to human brain functional networks. Under invited revision for *Nature Communications*.

6. Kang, X., Ganguly, A., and Kolaczyk, E.D. (2018) Dynamic networks with multi-scale temporal structure. (arxiv:1712.08586; under invited revision for *Journal of the Royal Statistical Society, Series B*)
7. Ahelegbey, D.F., Carvalho, L.E., Kolaczyk, E.D. (2018) A Bayesian covariance graphical and latent position model for multivariate financial time series. (arxiv:1712.06797; under invited revision for *Annals of Applied Statistics*)

Papers Published:

8. Kolaczyk, E.D., Lin, L., Rosenberg, S., Walters, J., and Xu, J. (2019). Averages of unlabeled networks: geometric characterization and asymptotic behavior. *Annals of Statistics*, (to appear).
9. Griffin, P.J., Zhang, Y., Johnson, W.E., and Kolaczyk, E.D. (2018). Detection of multiple perturbations in multi-omics biological networks. *Biometrics*, 74(4), 1351 -- 1361.
10. Gan, H.L. and Kolaczyk, E.D. (2018). Approximation of the difference of two Poisson-like counts by Skellam. *Journal of Applied Probability*, 55(2), 416 – 430.
11. Ginestet, C., Liu, J., Balachandran, P., Rosenberg, S., and Kolaczyk, E.D. (2017). Hypothesis testing for network data in functional neuroimaging. *Annals of Applied Statistics*, 11(2), 725-750.
12. Balachandran, P., Kolaczyk, E.D., and Viles, W. (2017). On the propagation of low-rate measurement error to subgraph counts in large networks. *Journal of Machine Learning Research*, 18(61):1 - 33.
13. Viles, W., Ginestet, C.E., Tang, A., Kramer, M.A., and Kolaczyk, E.D. (2016). Percolation under noise: Detecting explosive percolation using the second-largest component. *Physical Review E*, DOI: 10.1103/PhysRevE.93.052301.
14. Pham, L.M., Carvalho, L., Schaus, S., and Kolaczyk, E.D. (2016). Perturbation detection through modeling of gene expression on a latent biological pathway network: a Bayesian hierarchical approach. *Journal of the American Statistical Association,* 111, 73-92.
15. Zhang, Y. Kolaczyk, E.D., and Spencer, B.D. (2015). Estimating network degree distributions under sampling: an inverse problem, with applications to monitoring social media networks. *Annals of Applied Statistics*, 9(1), 166-199.

Refereed Conference Proceedings

16. Kang, X., Ganguly, A., and Kolaczyk, E.D. (2017). Dynamic networks with multi-scale temporal structure. *Proceedings of the 51st Asilomar Conference on Signals, Systems, and Computers*.
17. Ganguly, A., and Kolaczyk, E.D. (2017). Estimation of vertex degrees in a sampled network. *Proceedings of the 51st Asilomar Conference on Signals, Systems, and Computers*.
18. Zhang, Y., Lappas, T., Crovella, M., and Kolaczyk, E.D. (2014). Online ratings: convergence towards a positive perspective? *Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*.

Invited Book Chapters:

19. Ginestet, C., Kramer, M., and Kolaczyk, E.D. (2016). Network analysis. In *Handbook of Neuroimaging Data Analysis*, Ombao, H., Lindquist, M., Thompson, W., and Aston, J. (ed.). CRC Press.

Theses/Dissertations:

1. Kang, Xinyu. *Statistical methods for topology inference, denoising, and bootstrapping in networks*. Department of Mathematics & Statistics, Boston University. May, 2018.
2. Li, Jun. *Statistical methods for certain large, complex data challenges*. Department of Mathematics & Statistics, Boston University. May, 2018.
3. Goeva, Aleksandrina. *Complexity-penalized methods for structured and unstructured data*. Department of Mathematics & Statistics, Boston University. May, 2017.
4. Griffin, Paula. *Biological network models for inferring mechanism of action, characterizing cellular phenotypes, and predicting drug response*. Biostatistics Department, Boston University. May, 2016.
5. Zhang, Yaonan. Statistical analysis of network data motivated by problems in online social media. Department of Mathematics & Statistics, Boston University. May, 2015.
6. Viles, Weston D. *Network data analysis*. Department of Mathematics & Statistics, Boston University. May, 2014.
7. Pham, Lisa M. *Network-based methods to identify mechanisms of action in disease and drug perturbation profiles using high-throughput genomic data*. Bioinformatics Program, Boston University. May, 2013.

Invited Presentations (on topics relating in whole or in part to this award)[1]

Plenaries / Keynotes:

1. "Why Aren't Network Statistics Accompanied by Uncertainty Statements?" 3rd Graph Signal Processing Workshop (Keynote Speaker). EPFL. Lausanne, Switzerland. June, 2018.
2. "On the Impact of Network Inference on Network Science: Propagation of Uncertainty." SIAM Workshop on Inferring Networks from Non-network Data (Keynote Speaker). Austin, Texas. April, 2017.
3. "Estimating network degree distributions from sampled networks: an inverse problem." Conference on Applied Statistics in Defense (Plenary Speaker). Washington, DC. October, 2016.
4. "Statistical analysis of network data in the context of `Big Data': Large networks and many networks." IMA-HK-IAS Joint Program on Statistics and Computational Interface to Big Data (Keynote speaker). Hong Kong University of Science & Technology. January, 2015.
5. "Inference of network summary statistics through network denoising". Symposium on Graph Signal Processing (Keynote speaker). 1St IEEE Global Conference on Signal and Information Processing. Austin, Texas. December, 2013.

[1] For conciseness, only presentations given by the PI are listed here. Another 1-3 presentations per year were given by the 10 graduate students and postdoctoral fellows supported at various stages of this award.

Conference/Workshop Presentations:

6. "Statistical Analysis of Network Data: Foundations (Still!) Under Construction." Workshop on Statistical Inference, Learning, and Models in Data Science. The Fields Institute, Toronto, CA. September, 2018.
7. "Why Aren't Network Statistics Accompanied by Uncertainty Statements?" ICSA Applied Statistics Symposium. New Brunswick, New Jersey. June, 2018.
8. "On the Propogation of Uncertainty in Network Summaries." Workshop on Statistics of Network Analysis. Alan Turing Institute, London, England. May, 2018.
9. "Dynamic Networks with Multi-scale Temporal Structure." GraphEx Symposium. MIT Lincoln Labs. Lincoln, MA. April, 2018.
10. "Dynamic Networks with Multi-scale Temporal Structure." Workshop on Complex Time Series Modeling and Forecasting: Dynamic Networks, Spatio-temporal Data, and Functional Processes." Tsinghua-Sanya International Mathematics Forum, China. January, 2018.
11. "Estimation of Vertex Degrees in a Sampled Network." 51st Asilomar Conference on Signals, Systems, and Computers. Asilomar, CA. October, 2017.
12. "Dynamic Causal Networks with Multi-scale Temporal Structure." 51st Asilomar Conference on Signals, Systems, and Computers. Asilomar, CA. October, 2017.
13. "Statistics and Network Science: Overview and Open Problems" Annual Joint Statistical Meetings. Baltimore, Maryland. August, 2017.
14. "Challenges in Network Sampling: Open Problems and Some Progress." Annual Joint Statistical Meetings. Baltimore, Maryland. August, 2017.
15. "Dynamic causal networks with multi-scale temporal structure." Cowles Foundation / Yale Econometric Conference on Networks. Yale University. June, 2017.
16. "Dynamic causal networks with multi-scale temporal structure." Workshop on Dynamic Networks. Isaac Newton Institute. Cambridge, England. December, 2016.
17. Comment on "A regularization scheme on word occurrence rates that improves estimation and interpretation of topical content." Best paper session, Journal of the American Statistical Association – Applications & Case Studies. Annual Joint Statistical Meetings. Chicago, Illinois. August 2016.
18. "Estimating Network Degree Distributions from Sampled Networks: An Inverse Problem." SIAM Annual Meeting. Boston, Massachusetts. July, 2016.
19. "Dynamic causal networks with multi-scale temporal structure." 4th IMS-APR Meeting. Hong Kong, China. June, 2016.
20. "Dynamic causal networks with multi-scale temporal structure." Workshop on a Celebration of Statistics at Chicago (60th anniversary celebration). University of Chicago. Chicago, Illinois. May, 2016.
21. "Estimating network degree distributions from sampled networks: an inverse problem." Workshop on Networks, Random Graphs, and Statistics. Columbia University. New York, New York. May, 2016.
22. "Dynamic causal networks with multi-scale temporal structure." Workshop on Complex Systems in Time Series. London School of Economics. London, England. December, 2015.
23. "Statistical analysis of network data objects, with applications in functional neuroimaging." Annual Joint Statistical Meetings. Seattle, WA. August, 2015.
24. "Inference of network summary statistics through network denoising." Annual Meeting of the Institute of Mathematical Statistics. Sydney, Australia. July, 2014.
25. "Online ratings: Convergence towards a positive perspective?" IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP). Florence, Italy. May, 2014.

26. “Modeling and prediction of financial time-series: should we network?” McGill/Bellairs Research Workshop on Financial Data Modeling. Barbados. January, 2014.
27. “Estimating network degree distributions from sampled networks: An inverse problem.” Workshop on Social Network Data: Collection and Analysis. Statistical and Applied Mathematical Sciences Institute (SAMSI). RTP, North Carolina. October, 2013.
28. “Estimating network degree distributions from sampled networks: An inverse problem.” World Statistics Congress. Hong Kong. August, 2013.
29. “Detecting perturbed biological pathways through latent network modeling of gene expression.” Annual Joint Statistical Meetings. Montreal, Canada. August, 2013.
30. “Characterizing evolving patterns of cohesiveness in high-frequency dynamic networks.” 2Nd Workshop on Industry and Practices for Forecasting. Paris, France. June, 2013.
31. “The effect of noise and uncertainty on the analysis of large networks.” SIAM Conference on Computational Science and Engineering. Boston, MA. March, 2013.

Research Seminars:

32. “Why Aren’t Network Statistics Accompanied by Uncertainty Statements?” Stochastics and Statistics Seminar, MIT. March, 2019.
33. “Statistical Analysis of Network Data – Three Vignettes.” Department of Mathematics, Dartmouth College. February, 2019.
34. “Dynamic Networks with Multi-scale Temporal Structure.” Statistics Seminar, CREST. Paris, France. April, 2018.
35. “Statistical Analysis of Network Data in the Context of `Big Data’: Large Networks and Many Networks.” Department of Mathematics, Northwestern University. May, 2017.
36. “Estimating Network Degree Distributions Under Sampling: An Inverse Problem, with Application to Monitoring Social Media Networks.” Department of Economics, University of Maryland. March, 2017.
37. “Network-based Statistical Models and Methods for Identification of Cellular Mechanisms of Action.” Department of Statistics, Oxford University. December, 2016.
38. “Estimating Network Degree Distributions from Sampled Networks: An Inverse Problem”. Probability & Statistics Seminar, School of Mathematics, Bristol University. December, 2016
39. “Statistical Analysis of Network Data in the Context of 'Big Data': Large Networks and Many Networks”. Center for Statistics and Machine Learning, Princeton University. April 2016.
40. “Estimating Network Degree Distributions from Sampled Networks: An Inverse Problem.” Department of Statistics, North Carolina State University. April, 2016.
41. “Network-based Statistical Models and Methods for Identification of Cellular Mechanisms of Action.” Theodore L. Badger Lectures in Network Medicine. Channing Division of Network Medicine, Brigham and Women's Hospital. Boston, MA. September, 2015.
42. “Statistical Analysis of Network Data in the Context of `Big Data': Large Networks and Many Networks.” Big Data Initiative Seminar Series. London School of Economics (LSE), London. March, 2015.
43. “Statistical Analysis of Network Data: (Re)visiting the Foundations.” Department of Statistics, University of Chicago. October, 2014.

44. “Statistical Analysis of Network Data: (Re)visiting the Foundations. Laboratory for Information and Decision Sciences (LIDS), MIT. September, 2014.
45. “Statistical Analysis of Network Data.” Centre for Statistics, Gottingen University. March, 2014.
46. “Estimating Network Degree Distributions from Sampled Networks: An Inverse Problem. Université des Artes et Metier, Paris. November, 2013.
47. “Estimating Network Degree Distributions from Sampled Networks: An Inverse Problem. Department of Statistics, University of Georgia. September, 2013.
48. “Network-based Statistical Models and Methods for Identification of Cellular Mechanisms of Action.” Department of Statistics, University of California-Davis. March, 2013.

Honors and Awards

1. Eric Kolaczyk (PI) was elected a fellow of the Institute of Mathematical Statistics (IMS). (2017)
2. Eric Kolaczyk (PI) was elected a fellow of the American Association for the Advancement of Science (AAAS). (2017)