Risk Identification & Quantification in Complex Human-Natural Systems via Convergent Data Intensive Research

Toryn L. J. Schafer
tls255@cornell.edu
Cornell University
Ithaca, New York, USA

Ryan M. McGranaghan
rmcgranaghan@astraspace.net
Atmosphere Space Technology Research Associates
Colorado, USA

Mila Getmansky Sherman
msherman@isenberg.umass.edu
UMASS Amherst
Massachusetts, USA

Mei-Ling E. Feng
emfeng@lpzoo.org
Lincoln Park Zoo
Chicago, Illinois, USA

Olukunle O. Owolabi
olukunle.owolabi@tufts.edu
Tufts University
Medford, Massachusetts, USA

Sean E. Ryan
sean.ryan@cornell.edu
Cornell University
Ithaca, New York, USA

Marie-Christine Düker
md2224@cornell.edu
Cornell University
Ithaca, New York, USA

Michael Jauch
maj225@cornell.edu
Cornell University
Ithaca, New York, USA

David S. Matteson
matteson@cornell.edu
Cornell University
Ithaca, New York, USA

ABSTRACT

Human-natural systems involve complex interdependent processes, but domain specific processes are traditionally studied in non-overlapping research silos. The Predictive Risk Investigation SysteM (PRISM) for multi-layer dynamic interconnection analysis is a group of collaborators across multiple domains who work to discover data driven connections specifically among domain risks. We bring our inter-disciplinary approach to risk assessment to our KDD'21 workshop. Our workshop is a step toward a holistic approach to systemic risk analysis by welcoming speakers in applied and technical research at the forefront of risk and complex systems.

CCS CONCEPTS

• **Applied computing** → *Mathematics and statistics*;

KEYWORDS

associated anomalies, complex systems; data-intensive risk assessment; human-natural systems; systemic risk; volatility

ACM Reference Format:
Toryn L. J. Schafer, Ryan M. McGranaghan, Mila Getmansky Sherman, Mei-Ling E. Feng, Olukunle O. Owolabi, Sean E. Ryan, Marie-Christine Düker, Michael Jauch, and David S. Matteson. 2021. Risk Identification & Quantification in Complex Human-Natural Systems via Convergent Data Intensive Research. In *Proceedings of the 27th ACM SIGKDD Conference on Knowledge Discovery and Data Mining (KDD '21), August 14–18, 2021, Virtual Event, Singapore.* ACM, New York, NY, USA, 2 pages. https://doi.org/10.1145/3447548.3469480

KDD '21, August 14–18, 2021, Virtual Event, Singapore

ACM ISBN 978-1-4503-8332-5/21/08.
https://doi.org/10.1145/3447548.3469480

1 INTRODUCTION AND MOTIVATION

Many key areas of social and scientific importance, such as climate, finance, energy, transportation, and ecology, can be viewed as a complex network of interdependent processes. These connections mean that small events in one area may accrue through the network and wreak havoc on the entire system. Currently, there is no single discipline that is equipped to identify broader signs of systemic risk and mitigation targets. In this workshop, we will look to study how risks in different domains are connected. For instance, what risks in agriculture, ecology, energy, finance, and hydrology are heightened by climate variability and change? How might risks in, for example, space weather, be connected with energy, water, and finance? Recent advances in computing and data science and the data revolution in each of these domains have provided a means to address these questions.

This workshop focuses on techniques and tools that enable us to explore large-scale complex systems via multi-resolution dynamic datasets. We seek ideas and advances that will allow scientists to collaborate across disparate domains and integrate analyses of datasets on different time scales and resolutions in order to improve the prediction of risks (potentials for extreme outcomes and system failures). Topics may include but are not limited to new approaches to data representation and integration that facilitate cross disciplinary collaboration, case studies demonstrating collaboration of multiple domains via data science as well as cutting-edge applications of complex systems analyses. Through this workshop we hope to identify interrelationships between domains and propose novel approaches to understanding and quantifying systemic risk in our human-natural systems.

From climate change to political turbulence, our human-natural systems face a mounting set of interconnected risks. Data mining and science techniques offer a unique way to capture and understand these risks by combining information from different domains. To meet this challenging task, we gather knowledge from a diverse set of experts.

2 OBJECTIVE AND TOPICS OF INTEREST

The objective of this workshop is to showcase techniques and tools that unravel complex inter-connections and inspire ways to advance our understanding and preservation of human-natural systems. The workshop will address this objective by inviting contemporary research and data science that facilitate collaboration and cross-disciplinary efforts through the following topics:

- Defining & quantifying risk (assessment) from data: Identification of new and existing measures of quantifying critical risk in human natural systems, approaches to improve existing measures risk quantification
- Combining multiresolution data: Methods and approaches for resolving temporal and geographical disparities in critical risk measures, Data science/Machine Learning techniques for correcting bias in dirty real-world datasets for multidisciplinary risk analysis.
- Knowledge discovery for human-natural systems: Knowledge framework and definition for reliability and resiliency in risk quantification, Knowledge framework for causation and correlation in the context of interdisciplinary risk assessment.
- Complex systems/network analysis: Single pathways between interdisciplinary measures of critical risk, two-way(dual) pathways between interdisciplinary critical risk
- Systemic risk measures: Holistic systemic analysis and understanding of critical risk, framework and design for translating systemic measures of critical risk to improve resiliency and reliability of real-world infrastructure.
- Data science methods for facilitating collaboration across domains: Existing techniques for interdisciplinary analysis and framework for improving existing methods, novel methods or interdisciplinary analysis
- Cross-disciplinary research: Framework for improving data driven cross-disciplinary research methods.
- Linking data and domains via data science: Framework for improving data science and domain collaboration collaboration for interdisciplinary risk assessment

3 PROGRAM

The half day virtual workshop will consist of a series of 6 to 8 research talks, followed by a one hour poster session with discussion encouraged throughout. At the time of writing, research talks will include presentations on original work, in addition to the following invited speakers:

- **Animashree Anandkumar, Caltech and NVIDIA** As a director of machine learning at NVIDIA, she leads the development of next-generation AI algorithms. As a Bren professor at Caltech CMS department, she co-leads the AI4science initiative with Yisong Yue, bringing together AI researchers with experts from other disciplines to push modern AI tools into every area of science and engineering. Her research has spearheaded the development of tensor algorithms, central to effectively processing multidimensional and multimodal data, and for large-scale AI applications.
- **Catherine Calder, University of Texas at Austin** Catherine "Kate" Calder is a Statistician, Chair of Statistics and Data Sciences, and faculty research associate in the Population Research Center at the University of Texas at Austin. Her research interests are motivated by applications in the environmental, social, and health sciences and lie in areas of spatial statistics, Bayesian modeling and computation, and statistical network analysis. She has contributed to advancing methods that address multi-variate, multi-dimensional spatio-temporal data and models, and model-based comparisons of networks.
- **Megan Konar** Megan Konar is an associate professor in the Department of Civil and Environmental Engineering at the University of Illinois at Urbana-Champaign. Prof Konar's research focuses on the intersection of water, food, and trade. Her research is interdisciplinary and draws from hydrology, environmental science, and economics. She was recently awarded the NSF CAREER award and Early Career Award from AGU Hydrologic Sciences.

Additionally, **David S. Matteson** will give an overview of the PRISM team's ongoing work, including a case study on linking trans-domain critical risks with the electricity domain [1]. For more details on the PRISM team's activities visit https://sites.google.com/view/prism-prj/home.

4 PROGRAM COMMITTEE

4.1 Organizing Committee

David S. Matteson, Ryan M. McGranaghan, Sean E. Ryan, Marie-Christine Düker, Toryn L. J. Schafer, Michael Jauch, Mei-Ling E. Feng, Olukunle O. Owolabi

4.2 Additional PRISM Members

Judy Che-Castaldo, Rémi Cousin, Rajesh K. Gupta, Wei Ren, Chaopeng Shen, Deborah Sunter, Lan Wang, Joe Hughes, Drew Resnick

ACKNOWLEDGMENTS

Funding provided by the NSF Harnessing the Data Revolution (HDR) program, "Collaborative Research: Predictive Risk Investigation SysteM (PRISM) for Multi-layer Dynamic Interconnection Analysis" (awards #1940160, 2023755, 1940176, 1940190, 1940208, 1940223, 1940276, 1940291, and 1940696). R. McGranaghan was partially supported under the NSF Convergence Accelerator Award to the Convergence Hub for the Exploration of Space Science (CHESS) team (NSF Award Number: 1937152).

REFERENCES

[1] Judy P. Che-Castaldo, Rémi Cousin, Stefani Daryanto, Grace Deng, Mei-Ling E. Feng, Rajesh K. Gupta, Dezhi Hong, Ryan M. McGranaghan, Olukunle O. Owolabi, Tianyi Qu, Wei Ren, Toryn L. J. Schafer, Ashutosh Sharma, Chaopeng Shen, Mila Getmansky Sherman, Deborah A. Sunter, Lan Wang, and David S. Matteson. 2021. Critical Risk Indicators (CRIs) for the electric power grid: A survey and discussion of interconnected effects. *To Appear Environment Systems and Decisions* (2021). arXiv:2101.07771 [stat.AP]