

Xiaohan Kang

MACHINE LEARNING ENGINEER

☎ (+1) 515-509-6693 | ✉ xkang515@gmail.com | 🏠 veggente.github.io | 📷 Veggente | 📄 xiaohankang | 🎓 Xiaohan Kang

Work Experience

Rockfish Data Inc.

Pittsburgh, PA

MACHINE LEARNING ENGINEER

Jul. 2023–present

- Engage with customers to understand their use cases, pain points, data characteristics, and synthetic data needs, translating insights into well-defined problem formulations.
- Design and implement **synthetic data solutions**, delivering end-to-end proofs of concept using advanced deep learning architectures, including *GANs, transformers, and diffusion models*, with a focus on high-dimensional time-series data under stringent *privacy constraints*.
- Identify *key technical features* that address critical synthetic data challenges based on customers' needs and requirements, collaborate with stakeholders to conceptualize these solutions, and work closely with software engineers to implement and deliver them effectively.
- Coauthored *patent applications* for synthetic data generation methods and systems.
- Provide technical support to customers' engineering teams, facilitating deployment and scaling of synthetic data workflows for seamless onboarding and operational efficiency.

University of Illinois Urbana–Champaign

Urbana, IL

POSTDOCTORAL RESEARCH ASSOCIATE, DEPT. OF ELECTRICAL AND COMPUTER ENGINEERING, HOST: PROF. BRUCE HAJEK

Mar. 2016–Jun. 2023

- Developed a novel machine learning framework for causal network inference, the first to generate reliability scores for edge predictions by using Gaussian approximations of bootstrap aggregating.
- Applied the framework to reconstruct gene regulatory networks by analyzing large-scale time-series RNA-seq data, uncovering regulatory interactions and identifying key genes influencing soybean flowering and circadian rhythms.
- Characterized fundamental limits of classification errors and causal network inference, employing mathematical modeling and rigorous statistical methods to provide insights into the behavior of complex systems under uncertainty.
- Taught over 60 students in ECE 313 (Probability with Engineering Applications), using engaging demonstrations and visualizations to explain probabilistic concepts, earning positive teaching evaluation.

Cisco Systems, Inc.

San Jose, CA

SOFTWARE ENGINEERING INTERN

May–Aug. 2015

- Developed a Django-based web application for debugging distributed networking systems.
- Led team to win the first place at Cisco Intern Hackathon.

Education

Arizona State University

Tempe, AZ

PH.D. IN ELECTRICAL ENGINEERING, ADVISOR: PROF. LEI YING

2015

- Dissertation: "Performance analysis of low-complexity resource-allocation algorithms in stochastic networks using fluid models."

Tsinghua University

Beijing, China

B.E. IN ELECTRONIC ENGINEERING

2009

Skills

PYTHON, PYTORCH, UNIX, GIT, C/C++, SQL, PANDAS, KERAS, TENSORFLOW

Selected Projects

Enhancing ML models with synthetic data in data silos (Rockfish Data)

- Designed a synthetic data generation framework with tailored and optimized generative model architectures and configurations, improving ML models with distributed, non-shareable data silos.
- Achieved detection accuracy comparable to centralized models trained on real data, while maintaining data privacy compliance.
- Demonstrated superior performance over existing approaches, offering more flexible workflows.

Causal network inference and rare event detection (Postdoctoral Research)

- Designed CausNet, a comprehensive framework to identify causal interactions within networks with reliability scores for enhanced accuracy.
- Developed MLEROC, an estimator of the optimal ROC curve characterizing the tradeoff between the true positive rate and the false positive rate.
- Discovered a novel photoperiodic flowering pathway of soybean using CausNet on large-scale biological datasets.

Scheduling algorithms in stochastic networks (Ph.D. Research)

- Characterized sharp bounds on the performance of low-complexity algorithms for scheduling real-time traffic in wireless networks.
- Proposed batch-filling, a dispatching algorithm for large computing systems with strong performance guarantees and low messaging overhead.
- Provided optimality stability guarantees for a low-complexity algorithm for scheduling multi-hop traffic in linear wireless networks.

Selected Research Talks

- [T10] “Finite-sample lower bounds on information requirements for causal network inference,” *BIRS CMO Workshop on Learning in Networks: Performance Limits and Algorithms*, invited talk, Oaxaca, Mexico, 2022.
- [T9] “Lower bounds on information requirements for causal network inference,” *INFORMS Annual Meeting*, invited talk, Anaheim, CA, 2021.
- [T8] “On modeling the circadian clock gene regulatory network in soybean,” *Finding Your Inner Modeler Workshop IV (FYIM)*, University of Illinois at Chicago, Chicago, IL (virtual), 2021.
- [T7] “Time series experimental design under one-shot sampling: The importance of condition diversity,” *Energy & Information Systems Seminar*, invited talk, Carnegie Mellon University, Pittsburgh, PA, 2019.
- [T6] “On the challenge of gene regulatory network reconstruction from high-throughput sequencing data,” *Network Science Seminar Series*, invited talk, Arizona State University, Tempe, AZ, 2018.
- [T5] “CausNet: a causal inference algorithm for gene regulatory network reconstruction,” *The Plant and Animal Genome XXVI Conference (PAG 2018)*, San Diego, CA, 2018.
- [T4] “The power of slightly more than one sample in randomized load balancing,” *SINE Seminar*, invited talk, University of Illinois at Urbana-Champaign, Urbana, IL, 2016.
- [T3] “The power of slightly more than one sample in randomized load balancing,” guest lecture (hosted by Prof. Rhonda Righter), University of California, Berkeley, CA, 2016.
- [T2] “The power of slightly more than one sample in randomized load balancing,” *INFORMS Annual Meeting*, invited talk, Philadelphia, PA, 2015.
- [T1] “On the performance of largest-deficit-first for scheduling real-time traffic in wireless networks,” invited talk (hosted by Prof. Eytan Modiano), Massachusetts Institute of Technology, MA, 2015.

Publications

Conference publications

- [C9] Bruce Hajek and Xiaohan Kang, “Maximum likelihood estimation of optimal receiver operating characteristic curves from likelihood ratio observations,” *IEEE International Symposium on Information Theory (ISIT)*, 2022. [\[DOI\]](#) [\[arXiv\]](#)
- [C8] Xiaohan Kang and Bruce Hajek, “Lower bounds on information requirements for causal network inference,” *IEEE International Symposium on Information Theory (ISIT)*, 2021. [\[DOI\]](#) [\[arXiv\]](#)
- [C7] Honghao Wei, Xiaohan Kang, Weina Wang, and Lei Ying, “QuickStop: A Markov optimal stopping approach for quickest misinformation detection,” *ACM International Conference on Measurement and Analysis of Computer Systems (SIGMETRICS)*, 2019. [\[DOI\]](#) [\[arXiv\]](#)
- [C6] Xiaohan Kang, I-Hong Hou, and Lei Ying, “On the capacity requirement of largest-deficit-first for scheduling real-time traffic in wireless networks,” *ACM International Symposium on Mobile Ad Hoc Networking and Computing (MobiHoc)*, 2015. [\[DOI\]](#)
- [C5] Lei Ying, R. Srikant, and Xiaohan Kang, “The power of slightly more than one sample in randomized load balancing,” *IEEE Conference on Computer Communications (INFOCOM)*, 2015. **(Best Paper Award)** [\[DOI\]](#)
- [C4] Xiaohan Kang, Juan José Jaramillo, and Lei Ying, “Stability of longest-queue-first scheduling in linear wireless networks with multihop traffic and one-hop interference,” *IEEE Conference on Decision and Control (CDC)*, 2013. [\[DOI\]](#)
- [C3] Xiaohan Kang, Weina Wang, Juan José Jaramillo, and Lei Ying, “On the performance of largest-deficit-first for scheduling real-time traffic in wireless networks,” *ACM International Symposium on Mobile Ad Hoc Networking and Computing (MobiHoc)*, 2013. [\[DOI\]](#)
- [C2] Xiaohan Kang, Juan José Jaramillo, and Lei Ying, “Impacts of peer churn on P2P streaming networks,” *Annual Allerton Conference on Communication, Control and Computing (Allerton)*, 2012. [\[DOI\]](#)
- [C1] Xiaohan Kang, Juan José Jaramillo, “A strategy-proof and non-monetary admission control mechanism for wireless access networks,” *International Conference on Heterogeneous Networking for Quality, Reliability, Security and Robustness (QShine)*, 2010. [\[DOI\]](#)

Journal publications

- [J6] Xiaohan Kang, Bruce Hajek, and Yoshie Hanzawa, “From graph topology to ODE models for gene regulatory networks,” *PLOS ONE*, vol. 15, no. 6, pp. e0235070, 2020. [\[DOI\]](#)
- [J5] Faqiang Wu, Xiaohan Kang, Minglei Wang, Waseem Haider, William B. Price, Bruce Hajek, and Yoshie Hanzawa, “Transcriptome-enabled network inference revealed the *GmCOL1* feed-forward loop and its roles in photoperiodic flowering of soybean,” *Frontiers in Plant Science*, vol. 10, pp. 1221, 2019. [\[DOI\]](#)
- [J4] Xiaohan Kang, Bruce Hajek, Faqiang Wu, and Yoshie Hanzawa, “Time series experiment design under one-shot sampling: The importance of condition diversity,” *PLOS ONE*, vol. 14, no. 10, pp. e0224577, 2019. [\[DOI\]](#)
- [J3] Lei Ying, R. Srikant, and Xiaohan Kang, “The power of slightly more than one sample in randomized load balancing,” *Mathematics of Operations Research*, vol. 42, no. 3, pp. 692–722, 2017. [\[DOI\]](#)
- [J2] Xiaohan Kang, Weina Wang, Juan José Jaramillo, and Lei Ying, “On the performance of largest-deficit-first for scheduling real-time traffic in wireless networks,” *IEEE/ACM Transactions on Networking*, vol. 24, pp. 72–84, Feb. 2016. [\[DOI\]](#)
- [J1] Xiaohan Kang, Juan José Jaramillo, Lei Ying, “Stability of longest-queue-first scheduling in linear wireless networks with multihop traffic and one-hop interference,” *Queueing Systems*, vol. 80, no. 3, pp. 273–291, Jul. 2015. [\[DOI\]](#)

Academic Services

- Technical Program Committee member for ACM MobiHoc 2019–2022, and WiOpt 2021.
- Reviewer for *IEEE/ACM Transactions on Networking*, *Queueing Systems*, *IEEE Transactions on Mobile Computing*, *IEEE Communications Letters*, *IEEE Transactions on Vehicular Technology*, *IEEE Signal Processing Letters*, *IEEE Transactions on Network Science and Engineering*, and *IEEE International Symposium on Information Theory*.