

Cong Mu

[cmu2@jhu.edu](mailto:cmu2@jhu.edu)

<https://congm.github.io>

## Education

Johns Hopkins University (Baltimore, MD)	08/2017 – 05/2023
Doctor of Philosophy   Major: Applied Mathematics and Statistics	(Expected)
Master of Science in Engineering   Major: Applied Mathematics and Statistics	
Master of Science in Engineering   Major: Computer Science	
Sun Yat-Sen University (Guangzhou, China)	08/2013 – 06/2017
Bachelor of Science   Major: Statistics   Minor: Finance	

## Research Experience

### Dynamic Network Sampling for Subsequent Inference | Johns Hopkins University

- Designed dynamic network sampling scheme to optimize block recovery for SBM in the case where it is expensive to observe the entire graph; provided justification for using Chernoff information in subsequent inference for graphs.
- Key words: **Dynamic Network Sampling, Stochastic Blockmodel (SBM), Community Detection, Chernoff Information.**
- Mu, C., Park, Y., & Priebe, C. E. (2022). Dynamic network sampling for subsequent inference. [[Preprint](#)]

### Community Detection for SBM with Vertex Covariates | Johns Hopkins University

- Developed model-based spectral algorithms for clustering vertices in SBM with vertex covariates; assessed effects of observed and unobserved vertex heterogeneity on block recovery; employed Chernoff information to analytically compare the performance and derive the Chernoff ratio expression for certain models of interest.
- Key words: **Spectral Graph Inference, Community Detection, Stochastic Blockmodel (SBM), Vertex Covariates, Chernoff Ratio.**
- Mu, C., Mele, A., Hao, L., Cape, J., Athreya, A., & Priebe, C. E. (2022). On spectral algorithms for community detection in stochastic blockmodel graphs with vertex covariates. IEEE Transactions on Network Science and Engineering. [[Paper](#)] [[Preprint](#)]

### Statistical Models for Large Networks | Johns Hopkins University

- Built statistical models and algorithms that could be scaled to analyze large networks; estimated and simulated network formation models using high performance computing; developed R package with research objectives such as identifying the community structure.
- Key words: **(Generalized) Random Dot Product Graph, SBM with Covariates, Parallel Computation.**
- Dynamic Network Structure [[Shiny App](#)]
- (G)RDPG with Covariates [[R Package](#)]

## Automatic Tools for Dash Cam Video | Johns Hopkins University

- Developed automatic tools for analyzing and annotating video stream with relevant information such as timing, speed, traffic, accidents, objects and etc.
- Key words: **Structural Similarity Index, Oriented FAST and Rotated BRIEF, Image Hashing, Robust Image Similarity Measure, Deep Neural Networks.**
- **Mu, C., & Budavári, T. (2018).** Dash Cam Video Analysis: Laptimes and Beyond. Poster presented at 2018 IDIES Annual Symposium, Baltimore, MD. [[Poster](#)]

## Professional Experience

---

<b>Data &amp; Applied Scientist Intern   Microsoft</b> (Redmond, WA)	05/2022 – 08/2022
--	-------------------

- Provided a general guideline of pre-training and fine-tuning language models using Hugging Face and different hardware. Used pre-training language models for question generation (answer-agnostic) in Korean as running example. (**Transformer, ProphetNet**)
- Created a demo using pre-trained and fine-tuned models. Presented in Team Data Science Discussion and prepared a [blog](#). (**Python Dash**)

<b>Data &amp; Applied Scientist Intern   Microsoft</b> (Virtual)	06/2021 – 08/2021
--	-------------------

- Pre-trained language models for a few selected languages using Microsoft Turing technology. Fine-tuned on different downstream tasks and built subsequent applications. (**Transformer, DeBERTa, DeepSpeed**)
- Created a demo using pre-trained and fine-tuned models. Presented in Azure Data ML Talk Series. (**Python Dash**)

<b>Analyst Intern   GF Fund Management</b> (Guangzhou, China)	11/2016 – 04/2017
---	-------------------

- Selected features to build market emotional indicators and developed model to predict market, achieved 92% accuracy. (**XGBoost, Random Forest, Lasso**)
- Mined key business data and constructed data reporting system; analyzed and visualized product and user data to provide decision support. (**R Markdown, R Shiny**)

## Honors & Awards

---

MINDS Data Science Fellowship, Johns Hopkins University	2021 & 2022 Spring
Creel Family Engineering Fellowship, Johns Hopkins University	2019 – 2020
University Scholarship for Outstanding Student, Sun Yat-Sen University	2014 – 2016