JAEWON CHUNG

Ph.D. student at the intersection of causal inference, machine learning, and human brains

✉ j1c@jhu.edu   🖥 j1c.org   👤 he/him   in linkedin.com/in/j1chung   ⭗ github.com/j1c   🎓 Google Scholar

EDUCATION

Ph.D., Biomedical Engineering – Johns Hopkins University   Jun. 2019 - Jun. 2024
Highlighted Courses: Matrix Analysis, Real Analysis, Statistical Model Selection.

M.S.E., Biomedical Engineering – Johns Hopkins University   Sept. 2017 - May. 2019
Highlighted Courses: Machine Learning, Deep Learning, Data Mining, Probability, Statistics.

B.A., Neuroscience & Economics – Wesleyan University   Sep. 2009 - May. 2013

EXPERIENCE

Graduate Researcher - Ph.D.   Jun. 2019 - Prresent
Johns Hopkins University   *Baltimore, MD*

- Research and develop algorithms for statistical modeling of networks with applications to connectomes; implement in graspologic in collaboration with Microsoft researchers (**Git, CI, software development**) [P5, P6, P7, P10, L1, L2]
- Research and develop algorithms for multivariate hypothesis testing specifically tailored for biomedical datasets with high dimensions and low sample sizes; implement in hyppo (**Numpy, Numba**) [P1, P3]
- Develop causal machine learning algorithms to estimate causal effects of genomes on populations of connectomes (brain networks) [P2, P3, P4]
- Aggregate, process, and analyze terabytes' of MRI images to generate connectomes in the cloud (**AWS Batch, AWS S3**) [P8]

Machine Learning Intern   Jun. 2023 - Aug. 2023
Generate Biomedicines   *Boston, MA*

- Researched protein surface-based representation learning algorithm for scoring protein binding affinity and identifying protein binding sites (**PyTorch, PyTorch Geometric, AWS EC2**)
- Trained surfaced-based models on publicly available and proprietary antibodies distributed on 16+ GPUs
- Evaluation showed a 10% improvement over in-house binding affinity algorithms
- Implemented a generative protein algorithm conditioned on the signature of protein surfaces

Graduate Researcher - M.S.   Sep. 2017 - May. 2019
Johns Hopkins University   *Baltimore, MD*

- Developed an unsupervised convolutional covariance algorithm to detect synapses in big electron microscopy datasets (**Python, Scipy, PyTorch**), resulting in **NeurIPS BigNeuro Workshop presentation** [L11]
- Engineered a data analysis pipeline using parallel programming to detect synapses, classify synapses into different types using Random Forest algorithm, and visualize exploratory data analysis of detected synapses (**scikit-learn, AWS EC2, AWS S3**) [L11]
- Created online synapse annotation tool for neuroscientists used to compare algorithmically detected synapses to the ground truth of human annotations (**Javascript**)
- Aggregated and maintained terabytes of electron microscopy data for volumetric storage using BossDB

Data Analyst   Jun.2016-Aug.2017
NYU Langone Orthopedic Hospital   *New York, NY*

- Supervised nurse-led data-driven projects across five hospital floors, assisting them with data analysis, management, and visualization, which led to significant decreased sepsis rate and shortened hospital stays (**pandas, matplotlib, Excel**)
- Played an integral role in the hospital's achievement of the prestigious Magnet Recognition designation in 2017 through my significant contribution to quality of care projects
- Maintained multiple databases, and created business and quality of care reports for the hospital leadership (**Tableau**)

PUBLICATIONS & PREPRINTS

[P1] C. Shen, **J. Chung**, R. Mehta, T. Xu, and J. T. Vogelstein. "Independence Testing for Temporal Data". *arXiv* (2024). URL: https://arxiv.org/abs/1908.06486.

[P2] E. W. Bridgeford, M. Powell, G. Kiar, S. Noble, **J. Chung**, S. Panda, R. Lawrence, T. Xu, M. Milham, B. Caffo, and J. T. Vogelstein. "When no answer is better than a wrong answer: a causal perspective on batch effects". *bioRxiv* (2024). DOI: 10.1101/2021.09.03.458920.

[P3] **J. Chung**, E. W. Bridgeford, M. Powell, D. Pisner, T. Xu, and J. T. Vogelstein. "The Heritability of Human Connectomes: a Causal Modeling Analysis". *bioRxiv* (Dec. 2023). DOI: 10.1101/2023.04.02.532875.

[P4] E. W. Bridgeford, **J. Chung**, B. Gilbert, S. Panda, A. Li, C. Shen, A. Badea, B. Caffo, and J. T. Vogelstein. "Learning sources of variability from high-dimensional observational studies". *arXiv preprint arXiv:2307.13868* (July 2023).

[P5] **J. Chung**, B. Varjavand, J. Arroyo-Relión, A. Alyakin, J. Agterberg, M. Tang, C. E. Priebe, and J. T. Vogelstein. "Valid two-sample graph testing via optimal transport Procrustes and multiscale graph correlation with applications in connectomics". *Stat* (2022). DOI: 10.1002/sta4.429.

[P6] V. Gopalakrishnan, **J. Chung**, E. Bridgeford, B. D. Pedigo, J. Arroyo, L. Upchurch, G. A. Johnson, N. Wang, Y. Park, C. E. Priebe, et al. "Discovery of Multi-Level Network Differences Across Populations of Heterogeneous Connectomes". *arXiv* (2022). URL: https://arxiv.org/abs/2011.14990.

[P7] **J. Chung**, E. Bridgeford, J. Arroyo, B. D. Pedigo, A. Saad-Eldin, V. Gopalakrishnan, L. Xiang, C. E. Priebe, and J. T. Vogelstein. "Statistical connectomics". *Annual Review of Statistics and Its Application* (2021). DOI: 10.1146/annurev-statistics-042720-023234.

[P8] A. S. Charles, B. Falk, N. Turner, T. D. Pereira, D. Tward, B. D. Pedigo, **J. Chung**, R. Burns, S. S. Ghosh, J. M. Kebschull, et al. "Toward Community-Driven Big Open Brain Science: Open Big Data and Tools for Structure, Function, and Genetics". *Annual Review of Neuroscience* (2020). DOI: 10.1146/annurev-neuro-100119-110036.

[P9] T. M. Tomita, J. Browne, C. Shen, **J. Chung**, J. L. Patsolic, B. Falk, C. E. Priebe, J. Yim, R. Burns, M. Maggioni, et al. "Sparse projection oblique randomer forests". *Journal of Machine Learning Research* (2020). URL: http://jmlr.org/papers/v21/18-664.html.

[P10] **J. Chung**, B. D. Pedigo, E. W. Bridgeford, B. K. Varjavand, H. S. Helm, and J. T. Vogelstein. "GraSPy: Graph Statistics in Python." *Journal of Machine Learning Research* (2019). URL: http://jmlr.org/papers/v20/19-490.html.

[P11] J. T. Vogelstein, E. W. Bridgeford, B. D. Pedigo, **J. Chung**, K. Levin, B. Mensh, and C. E. Priebe. "Connectal coding: discovering the structures linking cognitive phenotypes to individual histories". *Current opinion in neurobiology* (2019). DOI: 10.1016/j.conb.2019.04.005.

LECTURES & TALKS

[L1] **J. Chung**, E. Bridgeford, and J. T. Vogelstein. "Hands on Graph Machine Learning". Johns Hopkins University - Advanced Graph Analytics Workshop, Sept. 2023.

[L2] **J. Chung**, E. Bridgeford, and J. T. Vogelstein. "Hands on Graph Machine Learning". Joint Statistical Meeting - Continuing Education Course, Toronto, CA, Aug. 2023.

[L3] **J. Chung**, M. Powell, E. Bridgeford, and J. T. Vogelstein. "Statistical Modeling of Structural Connectomes Reveal High Genetic Influence on Connectivity". Organization for Human Brain Mapping, Glasgow, UK, June 2022.

[L4] V. Gopalakrishnan, **J. Chung**, E. Bridgeford, B. D. Pedigo, C. E. Priebe, and J. T. Vogelstein. "Statistical Methods for Multiscale Comparative Connectomics". Neuromatch Conference 3.0, Virtual, Oct. 2020.

[L5] **J. Chung**, M. Powell, E. Bridgeford, and J. T. Vogelstein. "Human Structural Connectomes are Heritable". Neuromatch Conference 3.0, Virtual, Oct. 2020.

[L6] J. Arroyo, A. Athreya, J. Cape, G. Chen, **J. Chung**, C. E. Priebe, and J. T. Vogelstein. " Statistical inference for joint embeddings of multiple connectome data." Organization for Human Brain Mapping, Virtual, June 2020.

[L7] **J. Chung**, B. D. Pedigo, C. E. Priebe, and J. T. Vogelstein. "Structural Connectomes are Heritable". Society for Neuroscience, Chicago, IL, USA, Oct. 2019.

[L8] **J. Chung**, B. D. Pedigo, C. E. Priebe, and J. T. Vogelstein. "Clustering Multi-Modal Connectomes". Organization for Human Brain Mapping, Rome, Italy, June 2019.

[L9] **J. Chung**. "Statistical Methods for Population of Connectomes". Organization for Human Brain Mapping - Open Science SIG Meeting, Rome, Italy, June 2019.

[L10] B. D. Pedigo, **J. Chung**, E. W. Bridgeford, B. Varjavand, C. E. Priebe, and J. T. Vogelstein. "GraSPy: an Open Source Python Package for Statistical Connectomics". Max Planck /HHMI Connectomics Meeting, Berlin, Germany, Apr. 2019.

[L11] B. Duderstadt, **J. Chung**, F. Collman, and J. T. Vogelstein. "NOMADS: Neurodata's Opensource Method for Automatic Detection of Synapses". NeurIPS - BigNeuro Workshop, Long Beach, CA, USA, Dec. 2017.

[L12] J. Barrientos, M. Kaur, A. Mark, **J. Chung**, N. Driscoll, A. Bender, and K. Rai. "Outcomes of Patients with Chronic Lymphocytic Leukemia (CLL) after Idelalisib Therapy Discontinuation". Vol. 126. Dec. 2015, pp. 4155–4155. DOI: 10.1182/blood.V126.23.4155.4155.

OPEN-SOURCE PROJECTS

I have extensive experience working in asynchronous teams on code reviews, unit testing with **pytest**, continuous integration, API designs and discussion and implementing robust code in **Python**. All our organizations are dedicated to diversity, equity and inclusion and commonly host office hours, community development and public forums.

graspologic | GitHub | *Creator/Core Developer*   2019 - now
- Collaborated with Microsoft Research to develop and implement statistical network analysis algorithms in Python, currently in production at Microsoft. Project has 369 GitHub stars and 250K downloads (formerly GraSPy) [P10].

m2g | GitHub | *Developer*   2020 - now
- Turn-key pipeline that provides robust and reliable estimates of MRI connectivity, session and group-level analysis, and summary statistic computation. Used to generate connectomes for [P3].

hyppo | GitHub | *Developer*   2021 - now
- Develop and implement algorithms for multivariate hypothesis testing in high-dimensional and nonlinear datasets with a unified API. Project has 221 GitHub stars and 160k downloads [P1].

dipy | GitHub | *Contributor*   2019 - 2020
- Optimized Python-based 3D/4D+ medical image analysis package by implementing parallelization in preprocessing algorithms and resolving numerous bugs.

lollipop | GitHub | *Creator/Core Developer*   2019 - 2020
- Python package for supervised dimensionality reduction algorithm.

NOMADS | GitHub | *Core Developer*   2017 - 2018
- Neurodata's Opensource Method for Autonomous Detection of Synapses is an end-to-end Python pipeline for detecting synapses using an unsupervised detection algorithm, and classifying each synapse as excitatory or inhibitory from immunofluorescence brain images [L11].

PyMEDA | GitHub | *Creator/Core Developer*   2017 - 2018
- Python package for exploratory data analysis using multiple data visualizations.

HONORS & AWARDS

Trainee Highlight Award	9th Annual BRAIN Initiative Meeting	2023
AWS Research Grant	Developing and Optimizing Diffusion Magnetic Resonance Imaging Processing Pipelines for Multishell Images	2021
AWS Research Grant	Connectomics in the Cloud	2020
AWS Research Grant	Neurodata's Opensource Method for Autonomous Detection of Synapses (NO-MADS)	2018
Wolfram Award	MedHacks - Top 10 out of 375 projects	2017

SKILLS

Languages Python, Cython, R, C++, HTML, CSS, JavaScript, SQL, MatLab

AI Tools PyTorch, PyTorch Geometric, Scikit-learn, Numpy, Pandas, Weights & Biases

Other AWS (EC2, S3, Batch), Boto3, Docker, Jupyter, Git

Academic & Research	**ML/DL**	**Leadership**	**Communication**
Open ended research	Graph modeling	Project management	Public speaking
Human subjects research	Geometric deep learning	Agile software development	Technical writing
In-depth literature review	Transfer learning	Conflict resolution	Effective Visualization

Last generated on **February 19, 2024**