

Alex Hayes

alexpghayes@gmail.com • 775 338 8842 • Madison, WI, USA • [LinkedIn](#) • [Github](#) • www.alexpghayes.com

WORK EXPERIENCE

University of Wisconsin-Madison

August 2018 - Present

PhD Student, Department of Statistics

- Developed statistical methods based on principal components analysis to cluster networks with missing data, to perform regression on networks, and to construct, interpret, and regularize network embeddings.
- Developed causal inference methods to estimate mediation and spillover effects in social networks, and to determine when product changes have harmful side-effects on behaviors that are difficult to measure. Used causal machine learning to improve precision of estimates while reducing computational requirements by a factor of 5000.
- Implemented research methods in user-friendly software. Released nine open source R packages to CRAN (notable: [fastRG](#), [vsp](#), [distributions3](#), [gdim](#), [aPPR](#), [fastadi](#)).
- Resolved computational bottlenecks in matrix completion algorithms by designing and implementing sparse matrix methods in R and C++. Scaled methods by three orders of magnitude to handle networks with millions of nodes.
- Designed an approach to find localized clusters of Twitter users via Personalized PageRank. Managed unreliable Twitter API behavior by caching data in a Neo4J database running in Docker.
- Collaborated with ROpenSci to design software development standards for statistical software. Reviewed scientific software for ROpenSci, the R Journal, and the Journal of Open Source Software.

Facebook

Summer 2020 & Summer 2021

Research Intern, Core Data Science

- Prototyped a pipeline to automatically suggest relationships between hashtags, for a team using manual labeling. Prototype embedded a hashtag co-occurrence network and was implemented with Python, PyTorch and SQL.
- Conducted experiments on hyperbolic embeddings for knowledge graphs and determined non-viability of hyperbolic methods. Advised against additional R&D investment, potentially saving \$200k+ in compute costs.
- Designed a metric, based on calibration of machine learning models, to help product teams understand reliability of prevalence estimates. Metric reported daily on multiple dashboards. Implemented with sklearn, Numpy, pandas.

RStudio

Summer 2018

Intern, tidymodels team

- Re-factored thousands of lines of R code and developed a new test suite for the [broom](#) package (600k+ downloads/month, part of the tidyverse), improving behavioral consistency and reducing maintenance burden.
- Shipped a major new release of the package ([broom 0.5.0](#)). Resolved 80+ open issues and coordinated 40+ pull requests from open source contributors.

EDUCATION

University of Wisconsin-Madison

2018–2024 (expected)

Ph.D. Statistics

Rice University

2014–2018

B.A. Statistics, with *Distinction in Research and Creative Work*

SKILLS

- Specialized: Network analysis, embeddings, clustering, causal machine learning, interference, mediation
- Statistics: Data analysis, visualization, modeling, regression, generalized linear models, hypothesis testing
- Software: Proficient in R, Python, tidyverse, bash/unix, git; familiar with SQL, C++, Docker, AWS, Julia, Stan