Alex Hayes

alexpghayes@gmail.com • 775 338 8842 • Madison, WI, USA • Github • www.alexpghayes.com

I am a PhD Candidate in Statistics at the University of Wisconsin-Madison, where I study networks and causal inference. I am co-advised by Keith Levin and Karl Rohe.

PROFESSIONAL EXPERIENCE

University of Wisconsin-Madison **August 2018 - Present**
PhD Candidate, Department of Statistics

- Developed statistical methods 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.
- 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.

Facebook **Summer 2020 & Summer 2021**
Research Intern, Core Data Science

- Prototyped a pipeline to automatically suggest relationships between hashtags, replacing a manual labeling workflow. Prototype embedded hashtag co-occurrence network and was implemented in 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.

Rice University **Fall 2017**
Undergraduate researcher with Genevera Allen

Fred Hutchinson Cancer Research Center **Summer 2017**
Undergraduate researcher with Elizabeth Brown

Houston Parks and Recreation Department **Spring 2016**
Undergraduate researcher

EDUCATION

University of Wisconsin-Madison **2018–2024**
Ph.D. Statistics **(expected)**

Rice University **2014–2018**
B.A. Statistics, with *Distinction in Research and Creative Work*

MANUSCRIPTS

1. (in preparation) **Alex Hayes** and Kevin Levin. *Latent contagion in low-rank networks*. 2024+
2. (in preparation) **Alex Hayes** and Kevin Levin. *Peer effects in the linear-in-means model may be inestimable even when identified.* 2024+
3. (under review) **Alex Hayes**, Mark M. Fredrickson, and Keith Levin. *Estimating network-mediated causal effects via principal components network regression*. http://arxiv.org/abs/2212.12041. 2024+

PUBLICATIONS

1. **Alex Hayes** and Karl Rohe. *Co-factor analysis of citation networks*. Accepted at the Journal of Computational and Graphical Statistics. 2024
2. Hadley Wickham, Mara Averick, Jennifer Bryan, Winston Chang, Lucy D'Agostino McGowan, Romain François, Garrett Grolemund, **Alex Hayes**, Lionel Henry, Jim Hester, Max Kuhn, Thomas Lin Pedersen, Evan Miller, Kirill Müller, David Robinson, Dana Paige Seidel, Vitalie Spinu, Kohske Takahashi, Davis Vaughan, Claus Wilke, Kara Woo, Hiroaki Yutani. *Welcome to the Tidyverse*. Journal of Open Source Software. 2019

RESEARCH & STATISTICAL SOFTWARE

1. fastRG (CRAN, Github): Sample Generalized Random Dot Product Graphs in Linear Time. Useful to investigate properties of network models and spectral estimators. Easily control expected degree of sampled networks, and compute population SVDs. Leverages RSpectra to efficiently eigendecompose large, structured matrices. Implements https://www.jmlr.org/papers/v19/17-128.html
2. vsp (CRAN, Github): Vintage Sparse PCA for Semi-Parametric Network Analysis. Estimate latent positions in random dot product graphs via spectral embeddings and interpret them via varimax rotation. Easily regularize networks to handle noise. Implements https://doi.org/10.1093/jrsssb/qkad029
3. aPPR (Github): Approximate Personalized PageRank. Local clustering of networks based on degree-regularized PageRank estimates. Designed specifically for large networks only available via an API, such as the Twitter following graph. Leverages S3 and R6 object systems. Implements https://doi.org/10.1111/rssb.12349
4. gdim (CRAN, Github): Estimate Graph Dimension using Cross-Validated Eigenvalues. Determine the number of communities in stochastic blockmodels and variants. Implements https://doi.org/10.48550/arXiv.2108.03336
5. fastadi (CRAN, Github): Self-Tuning Data Adaptive Matrix Imputation. Estimating singular subspaces of sparsely observed matrices. Includes specialized computation and C++ implementation for upper triangular data. Implements https://doi.org/10.1080/10618600.2018.1518238 and https://doi.org/10.5705/ss.202016.0205
6. broom (CRAN, Github): Convert Statistical Objects into Tidy Tibbles. Part of the tidyverse. Puts hundreds of types of statistical estimates into a consistent format to make programming with them easier.
7. distributions3 (CRAN, Github): Probability Distributions as S3 Objects. An object-oriented interface to probability computations, with emphasis on careful documentation, beginner friendliness and classroom applicability.

In addition to writing code, I collaborated with ROpenSci to design software development standards for statistical software, I reviewed scientific software for ROpenSci, the R Journal, and the Journal of Open Source Software, and I helped organize the Chicago R Unconference in 2019.

TALKS

1. *Estimating network-mediated causal effects via spectral embeddings*. NetSci 2024. 2024-06-17
2. *Asymptotic identification of peer effects in linear models.* Thesis defense. 2024-04-04
3. *Peer effects are parametrically indistinguishable from baseline behaviors in the asymptotic limit*. Statistics Graduate Student Seminar, UW-Madison. 2023-11-27
4. *Peer effects are parametrically indistinguishable from baseline behaviors in the asymptotic limit*. Statistics Graduate Student Seminar, UW-Madison. 2023-11-27
5. *Latent contagion in low-rank networks*. Levin Lab Meeting, UW-Madison. 2023-10-11
6. *Peer diffusion over uncertain networks*. IFDS Ideas Seminar, UW-Madison. 2023-09-18
7. *Estimating network-mediated causal effects via spectral embeddings*. JSM 2023. 2023-08-09
8. *Estimating network-mediated causal effects via spectral embeddings*. IFDS Ideas Seminar, UW-Madison. 2023-04-24
9. *Estimating network-mediated causal effects via spectral embeddings*. Statistics Graduate Student Seminar, UW-Madison. 2022-10-14
10. *Estimating indirect effects induced by homophily via spectral network regression*. Tianxi Li and Can Le Joint Lab Meeting. 2022-07-07
11. *distributions3: From basic probability to probabilistic regression*. UseR 2022. 2022-06-23
12. *The Low Hanging Fruit of the Twitter Following Graph*. JSM 2021. 2021-08-11
13. *Solving the model representation problem with broom*. rstudio::conf(2019). 2019-01-25
14. *Solving the model representation problem with broom*. Statistics Graduate Student Seminar, UW-Madison. 2018-11-30
15. *Convenient data analysis with broom*. RStudio Webinar Series. 2018-11-14
16. *Solving the model representation problem with broom*. Madison R User Group. 2018-09-19

POSTER PRESENTATIONS

1. *Estimating network-mediated causal effects via spectral embeddings*. IFDS Annual Meeting. 2023-08-07
2. *Estimating network-mediated causal effects via spectral embeddings*. ACIC 2023. 2023-05-24
3. *Using data to support real-time decision making by the Hurricane Harvey crisis management team*. Rice Data Science Conference. 2017-10-10
4. *An exploratory analysis of the effect of waiting room interactions on adherence in clinical trials*. Fred Hutch Intern Poster Competition. 2017-08-10

TEACHING

Graduate Teaching Assistant

- STAT 340 Intro to Data Modeling II, UW-Madison — Fall 2022
- Statistics Department Outstanding TA Award — 2018-2019
- STAT 324 Intro to Statistics for Engineers, UW-Madison — Spring 2019
- STAT 324 Intro to Statistics for Engineers, UW-Madison — Fall 2018

Guest Lecturer

- Confidence intervals. STAT 340, UW-Madison — 2022-10-25 & 2022-10-27
- Sampling with Twitter following graph with aPPR. STAT 992, UW-Madison — 2020-10-08
- Hypothesis testing. STAT 324, UW-Madison — 2018-10-18

Co-instructor

- Applied Machine Learning Workshop. rstudio::conf(2019). — 2019-01-15 & 2019-01-16

Undergraduate Teaching Assistant

- COMP 540 Statistical Machine Learning, Rice University — Spring 2018
- COMP 330 Data Science: Tools & Models, Rice University — Fall 2017

MENTORING

- Nathan Kolbow (undergraduate research assistant), currently a PhD student in Biostatistics at UW-Madison