Justin Sybrandt, Ph.D.

Email: justin@sybrandt.com Web: sybrandt.com
GitHub: github.com/JSybrandt Cell: (484) 354-8692

TL;DR: Expert in Natural Language Processing. Focus in Automatic Biomedical Hypothesis Generation.

Education

Ph.D. in CS, Clemson University (Aug. 2016 - May 2020)

- Adviser: Dr. Ilya Safro.
- Thesis: *Exploiting Latent Features of Text and Graphs*.
- Recipient of the GAANN DAISE & NRT RIES fellowships.
- Received School of Computing's "Outstanding Ph.D. Student in Computer Science" award.

BS in CS, Minor in Math, Grove City College (Aug. 2012 - May 2016)

- Graduated Summa Cum Laude (GPA 3.85/4).
- Top of class in computer science (in-major GPA 3.95/4).

Publications

Peer-Reviewed Publications

- J. Sybrandt, I. Tyagin, M. Shtutman, and I. Safro. Agatha: Automatic graph mining and transformer based hypothesis generation approach. In *Proceedings of the 29th ACM International Conference on Information & Knowledge Management*, CIKM '20, page 2757–2764, New York, NY, USA, 2020. Association for Computing Machinery
- F. Ding, X. Zhang, J. Sybrandt, and I. Safro. Unsupervised hierarchical graph representation learning by mutual information maximization. 2020
- J. Sybrandt and I. Safro. First-and high-order bipartite embeddings. In *Proceedings of MLG 2020: 16th International Workshop on Mining and Learning with Graphs*, MLG'20, New York, NY, USA, 2020. Association for Computing Machinery
- J. Sybrandt, R. Shaydulin, and I. Safro. Hypergraph partitioning with embeddings. *IEEE Transactions on Knowledge and Data Engineering*, pages 1–1, 2020
- J. Sybrandt, M. Shtutman, and I. Safro. Moliere: Automatic biomedical hypothesis generation system. In *Proceedings of the 23rd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pages 1633–1642, 2017
- J. Sybrandt, M. Shtutman, and I. Safro. Large-scale validation of hypothesis generation systems via candidate ranking. In *2018 IEEE International Conference on Big Data (Big Data)*, pages 1494–1503. IEEE, 2018
- J. Sybrandt, A. Carrabba, A. Herzog, and I. Safro. Are abstracts enough for hypothesis generation? In *2018 IEEE International Conference on Big Data (Big Data)*, pages 1504–1513. IEEE, 2018
- M. Aksenova, J. Sybrandt, B. Cui, V. Sikirzhytski, H. Ji, D. Odhiambo, M. D. Lucius, J. R. Turner, E. Broude, E. Peña, et al. Inhibition of the dead box rna helicase 3 prevents hiv-1 tat and cocaine-induced neurotoxicity by targeting microglia activation. *Journal of Neuroimmune Pharmacology*, pages 1–15, 2019
- W. Locke, J. Sybrandt, L. Redmond, I. Safro, and S. Atamturktur. Using drive-by health monitoring to detect bridge damage considering environmental and operational effects. *Journal of Sound and Vibration*, 468:115088, 2020

Preprints, Extended Abstracts, and Technical Reports

- J. Sybrandt and I. Safro. Cbag: Conditional biomedical abstract generation. 2020
- J. Sybrandt and J. Hick. Rapid replication of multi-petabyte file systems. *Parallel Data Storage Workshop (PDSW) 2015*, 2015

- R. Shaydulin and J. Sybrandt. To agile, or not to agile: A comparison of software development methodologies. *arXiv preprint arXiv:1704.07469*, 2017

Development Skills and Technologies

Programming Languages	C++, Python, Bash, SQL, Matlab, Java, Scala
Tools	Git, Linux, VIM, LaTeX, Mercurial
ML-Libraries	PyTorch, Tensorflow, Keras, Scikit-Learn, Horovod
Parallel/Distributed Programming	Dask, OpenMP, Spark, Flume, Dataswarm, GNU-Parallel, MPI

Work Experience

Google Brain

- "*Make machines intelligent. Improve people's lives.*"
- Google Brain focuses on introducing machine learning across Google products. Prior projects include Tensorflow, MapReduce, and BERT.

Summer 2019, Ph.D. SWE Intern, Facebook

- Attented the Intern Executive Dinner hosted by Mark Zuckerberg, awarded to only 13 of the over 3,000 interns in 2019.
- Improved the precision and recall of models that detect violating content on Instagram by exploring and producing embedding-based features.
- Demonstrated high productivity and fast learning speed, as evidenced by formal peer feedback, while adapting to the workflow at Facebook.

Summer 2018, Ph.D. SWE Intern, Google

- Proposed and produced a graph-mining solution for identifying product attributes that could decreased the need for human oversight by over 50%.
- Worked efficiently, developing my proposed system from a whiteboard idea to an in-production pipeline ahead of schedule.
- Performed comprehensive validation, ensuring classifier performance across product categories.
- Presented work to senior research scientists within Google's graph-mining team.

Summer 2017, Graduate Research Assistant, Los Alamos National Lab

- Developed high performance software in Julia for non-negative matrix factorization to be released in the open-source scientific computing library *madsjulia*.
- Evaluated the ability for my research project MOLIERE to extend to water resources research with the computational environmental science group.

2015-2016, Programming Intern, Vigilant Cyber Systems, Inc.

- Developed a visualization library in Scala using ScalaFX for a DoD contract.
- Worked independently as a self-led remote employee, while simultaneously finishing my last year of undergrad.

Summer 2015, Student Researcher, UC Berkeley & NERSC

- Designed a tool in Java to quickly synchronize multi-petabyte Parallel File Systems.
- Created a poster that was accepted at the ACM Student Poster Session at the Supercomputing conference in 2015.
- Presented a work in progress paper at the Parallel Data Storage Workshop.

Summer 2014, Student Researcher, Grove City College

- Prototyped multiple temporal distribution preferences for Data Stream Management Systems.

- Implemented research ideas using Python, and deployed large-scale simulations on Linux.
- Created a poster that was presented at the Grove City student poster session.

2012-2014, Programming Intern, Gravic Inc.

- Worked closely with a small team to create tools for teachers to more easily administer exams.
- Implemented new features using VB.NET to allow for OCR detection of student IDs, fractional responses, and free-form textual answers.
- Collaborated directly with corporate partners to implement features that allowed our products to interoperate.

Key Talks

- Presented "AGATHA" at CIKM'20.
- Presented "First- and Higher-Order Bipartite Embeddings" at MLG'20.
- Invited online talk with the AI Socratic Circles group to present Agatha. Online at: `https://aisc.ai.science/events/2020-04-01/`.
- Presented thesis research at the Google PhD. Intern Research Conference (only 30 accepted).
- Presented research to senior data scientists from Pfizer as an invited webinar.
- Presented two research papers at the session on "Big Data Applications: Health & Science Discovery" at BigData'18 in Seattle WA.
- Presented thesis research as a part of Clemson's exhibitor booth presentations at SC'18 in Dallas TX.
- Presented the research paper on "Moliere" at the KDD'17 in Halifax NS Canada.

Teaching Experience

- F. 2019: Supervisor of project team in CPSC 8480, Network Science.
- S. 2018, F. 2018, F. 2019, S. 2020: Guest Lecture: Applied Data Science
- Fall 2017 - Spring 2018, Project Supervisor for senior projects in Seminar in Professional Issues II

Honors and Awards

- Member of Upsilon Pi Epsilon CS Honor Society.
- Member of the Kappa Mu Epsilon National Mathematics Honor Society.
- Member of the Alpha Tau Mu chapter of Mortarboard, a service-oriented honor society.
- Recipient of the KDD'17 ACM student travel award.
- Recipient of the BigData'18 IEEE student travel award.
- Recipient of the Clemson CCIT Super Computing'18 travel award.
- Recipient of the GAANN DAISE & NRT RIES fellowships.
- Received School of Computing's "Outstanding Ph.D. Student in Computer Science" award.

Assisted Preparation of Funded Proposals

- NSF RAPID: Automated discovery of COVID-19 related hypotheses using publicly available scientific literature, $104K, PI: Dr. Ilya Safro