Tom Magelinski

Washington, DC · tmagelinski.github.io · tmagelin@cs.cmu.edu

Education

Carnegie Mellon University — Pittsburgh, PA
PhD Computer Science (Societal Computing) *GPA: 3.9* — August 2017 - May 2023
Thesis: "Contextualized Conversational Network Dynamics on Social Media"
Committee: Kathleen M. Carley (Chair), Renaud Lambiotte, Patrick Park, and Osman Yagan

Virginia Tech — Blacksburg, VA
Honors Baccalaureate Engineering Science and Mechanics *GPA: 3.9* — August 2013 - May 2017
Minors in Math and Physics

University of Oxford — Oxford, UK
Visiting Student — January 2015 - April 2015

Experience

Johns Hopkins Applied Physics Lab — Washington, DC
Senior Data Scientist - Generative AI and Information Extraction — July 2023 - Present

- Implemented a robust validation framework for an advanced retrieval augmented generation (RAG) pipeline on a large-scale document stream through prompting of large language models
- Partnered in the development of a multi-lingual document clustering service, capable of finding relevant documents to a user's query, clustering documents into stories, and summarizing them
- Researching methods of incorporating multiple media representation techniques such as Whisper and ImageBind to improve upon various multi-modal tasks like any-to-any media search
- Improved a large-scale information extraction pipeline for threat detection using a mixture of LLMs and classic NLP techniques

Spotify Research — New York, NY
Research Scientist Intern — June 2021 - September 2021

- Demonstrated the potential of social signals to improve podcast search and recommendation
- Improved podcast understanding by integrating heterogeneous social network embedding into a transformer-based pipeline using StellarGraph

CASOS Lab — Pittsburgh, PA
Graduate Research Assistant — August 2017 - May 2023

- Improved unsupervised multi-modal tweet representational learning through *Deep Tweet Infomax*, which leverages state of the art multi-lingual natural language processing techniques, the conversational graph, hashtags, and URLs. Implemented in PyG and trained on GPU
- Developed highly scalable graph algorithms for bot detection and coordinated actor detection on Twitter datasets with tens of millions of Tweets using igraph and PyTorch
- Built a distributed data analysis pipeline to clean and learn representations of $\sim$ 100 million Tweets in PySpark and BigGraph
- Improved SotA graph classification accuracy by 1-2% on social media datasets by creating and implementing a novel deep graph-convolutional architecture in PyTorch
- Built an interactive dashboard in Plotly to analyze Twitter hashtag network dynamics using diachronic node embeddings
- Designed and implemented a scalable centrality measure which improves the ranking of 1 million PA road intersections by a factor of 8, as measured by ability to fragment the road network
- Published in venues like *AAAI*, *TheWebConf*, *ICWSM*, *The Journal of Online Trust and Safety*, *Applied Network Science*, and *IEEE Transactions on Network Science and Engineering*

Ross Dynamics Lab — Blacksburg, VA
Undergraduate Researcher — November 2015 - May 2017

Mathematical Institute — Oxford, UK
Undergraduate Research Assistant — January 2015 - April 2015

- Built and coded a mechanical material stretcher, complete with GUI and image-based software to measure material stress and strain

Bio-Inspired Fluids Lab — Blacksburg, VA
Undergraduate Research Assistant — September 2014 - December 2014

- Collected and organized data from experiments to understand drinking mechanisms of dogs

Technical Skills

Languages and Tools:	Python, R, Java, MATLAB, SQL (MySQL, BigQuery), LaTeX, Git
ML Frameworks and Libraries:	PyTorch, TensorFlow, PyG, StellarGraph, BigGraph, PySpark MLlib
Network Science Libraries:	igraph, NetworkX, Graph-Tool
Data and Statistics Libraries:	PySpark, NumPy, Pandas, NLTK, spaCy, SciPy, statsmodels, scikit-learn
Visualization:	Matplotlib, Seaborn, Plotly
Research Areas:	Graph Representation Learning, Node Embedding, Node Classification, Graph Classification, Network Science, Community Detection, Multi-Modal Learning, Natural Language Processing (with graph data)

Awards

Knight Foundation Fellow Fall 2020, Spring 2021
Tuition and stipend funding to support research on coordinated information campaigns on Social Media, particularly surrounding COVID-19 and the 2020 US Election

ARCS Foundation Scholar August 2017 - August 2020
Supplemental funding to develop dynamic network analysis techniques to discover changes in community structure

Outstanding Senior: Engineering Science and Mechanics May 2017

Virginia Tech Rhode Scholar Nominee November 2016

Virginia Tech Marshall Scholar Nominee November 2016

Tau Beta Pi: Engineering Honors Society September 2015 - May 2017

Academic Service

Societal Computing Seminar Chair September 2020 - May 2023
The SC seminar is a platform for PhD students to develop their public speaking skills, primarily through research talks. However, I often organize talks on special topics such as panels on academic vs. industry jobs, and town halls meetings. This position also requires I represent the PhD students in faculty meetings, where I advocate on their behalf.

Co-Organizer: Ethics for Technologists Lecture Series November 2018 - November 2021
Our lecture series aims to equip technologists and engineers with tools and frameworks for handling the ethical aspects of their work. We have secured funding from the Dean's office to hold monthly talks from a wide range of speakers. The average attendance is 30 people.

PhD Applicant Mentor for Historically Underrepresented Groups Fall 2021

Faculty Hiring Committee Member 2020-2021 Hiring Year

PhD Admissions Committee Member 2018, 2019

Biomedical Engineering and Mechanics Ambassador August 2016 - May 2017

Review Activities

I believe serving as a peer reviewer is an important act of service as a researcher. I have reviewed multiple papers in ICWSM, WebSci, Applied Network Science, Social Network Analysis and Mining, IEEE Transactions on Computational Social Systems, Quality & Quantity, PLOS ONE, Computer Networks, Computational and Mathematical Organization Theory, IEEE Access, and SBP-BRiMS.

Teaching and Demos

Dynamic Network Analysis Pittsburgh, PA
Teaching Assistant Spring 2021
I led recitations, held office hours for homework help, and advised students for their project in the graduate-level course. Multiple student's projects were submitted to major scientific conferences. I also updated syllabus's reading list to include relevant publications in social networks from the past year, and updated assignments.

CASOS Summer Institute Pittsburgh, PA
Teaching Assistant 2018-2022
The Summer Institute is a week-long course in Network Science tools and applications, with students from academia, industry, and government. I have lectured on Contextualized Network Analysis, Clustering, Community Detection, Network Robustness, Immediate Impact Analysis, Network Simulations, and Network Regression. I have given case-studies on finding factions with Ukrainian Voting Networks. In this role I also served as a mentor for student's research projects, spanning a range of topics including social science studies, transportation optimization, and organizational management.

Detecting Coordinated Actors on Twitter Remote

Knight Research Network Tool Demonstration Day October 2021

Bot Detection with BotHunter Remote
Demo at CMU IDeaS Institute Conference July 2021

Networked Time Series Analysis and Clustering Remote
Tutorial at SBP-BRiMS 2020
This tutorial was co-instructed with Mihovil Bartulovic.

Undergraduate Engineering Courses Remote
Tutor 2019-2021

Preprints and Working Papers

Magelinski, T., & Carley, K. M. (2022). Identity-Based Attribute Prototypes Distinguish Communities on Twitter. *Under Review.*

Magelinski, T., & Carley, K. M. (2022). Network Models of Vector-Contextualized Interaction Data. *Working Paper.*

Publications - Journals

Magelinski, T., Ng, L., & Carley, K. M. (2022). A Synchronized Action Framework for Detection of Coordination on Social Media. Journal of Online Trust and Safety.

- We give a highly scalable algorithm for the detection of communities exhibiting synchronized behavior, potentially in order to manipulate online discussions. The method is multi-modal, allowing for the detection of groups coordinating along multiple modalities, e.g., those simultaneously targeting a collection of URLs, hashtags, and users in their information maneuver.

Magelinski, T., Bartulovic, M., & Carley, K. M. (2021). Measuring Node Contribution to Community Structure with Modularity Vitality. IEEE Transactions on Network Science and Engineering.

- We give a highly scalable algorithm for calculating Modularity-Vitality, which measures centrality of nodes with respect to communities. MV improves the ranking of 1 million PA road intersections by a factor of 8, as measured by ability to fragment the road network, when comparing to techniques like PageRank.

Uyheng, J., **Magelinski, T.**, Villa-Cox, R., Sowa, C., & Carley, K. M. (2019). Inter-operable pipelines for social cyber-security: assessing Twitter information operations during NATO Trident Juncture 2018. Computational and Mathematical Organization Theory.

- We advocate for interoperability when developing tools for analyzing information operations and apply one such pipeline to the discussion of the Trident Juncture Exercise. We identify several anti-NATO narratives painting it as brutal, incompetent, or unwanted by the public, and looked at how those narratives complemented that of Russian state-sponsored media.

Magelinski, T., & Carley, K. M. (2019). Community-based time segmentation from network snapshots. Applied Network Science.

- We give an algorithm for determining time segments of community stability. These segments can then be analyzed with static network analysis. Applied to the Ukrainian Parliament, we can see a massive change in political alliance following the revolution of 2014.

Magelinski, T., & Carley, K. M. (2019). Analytic Models of Roll Call Voting Dynamics. IEEE Transactions on Computational Social Systems.

- Ukrainian parliament has an interesting structure, where bills require many votes to pass. We show that this system is well-modeled with an ordinary differential equation, showing that the first 2 votes are indicative of the bill's fate.

Publications - Conferences

Magelinski, T., & Carley, K. M. (2023). Contextualized Conversational Network Dynamics on Social Media. In ICWSM 2023.

- We demonstrate that interactional context plays an important role in modeling social media conversations as networks. We separate a large twitter discussion into discrete conversations by learning tweet-level vector representations with a custom heterogeneous graph neural network and then clustering them. Contextualized networks are shown to have differing nodesets, topology, and central actors.

Tsuchiya, T., Cuevas Villalba, A., **Magelinski, T.**, & Christin, N. (2023). Misbehavior and Account Suspension in an Online Financial Communication Platform. In The Web Conference 2023.

- We characterize rule violations and account suspensions of users on TradingView, the largest online communication platform for online trading. We find that suspended accounts tend to form more closely-knit communities than non-suspended accounts, and that the formation of these communities may be driven by price fluctuations in certain aspects with meme-stocks receiving disproportionate levels of rule violations. Lastly, we find that paying accounts are less likely to be suspended than free accounts even when posting similar levels of policy-violating content.

Magelinski, T., Beskow, D. M., & Carley, K. M. (2020). Graph-Hist: Graph Classification from Latent Feature Histograms with Application to Bot Detection. In AAAI (pp. 5134-5141).

- We introduce a differentiable histogram layer to a deep graph-classification architecture written in PyTorch. Graph-Hist improves SotA graph-classification by 1-2% on Reddit Benchmarks. We show this leads to more generalizable, but less scalable bot-detection on Twitter.

Magelinski, T., Bartulovic, M., & Carley, K. M. (2020). Canadian Federal Election and Hashtags That Do Not Belong. In International Conference on Social Computing, Behavioral-Cultural Modeling and Prediction and Behavior Representation in Modeling and Simulation. (pp. 161-170).

- We show that hashtags with the highest Modularity Vitality scores are more interpretable labels for hashtag clusters than existing measures like degree. We apply this method to the Canadian Election discussion with 3 networks of 10 million hashtags each and find that both of the major players had strong social media presence during the election, but discussion quickly changed to specific policy problems afterward.

Magelinski, T., Stine, Z., Marcoux, T., Agarwal, N., & Carley, K.M (2020). Artifacts of Crisis: Textual Analysis of Euromaidan. In International Conference on Social Computing, Behavioral-Cultural Modeling and Prediction and Behavior Representation in Modeling and Simulation. (pp. 329-339).

Magelinski, T., Hou, J., Mylovanov, T., & Carley, K. M. (2019). Detecting Disruption: Identifying Structural Changes in the Verkhovna Rada. In International Conference on Social Computing, Behavioral-Cultural Modeling and Prediction and Behavior Representation in Modeling and Simulation (pp. 194-203).

Magelinski, T., & Carley, K. M. (2018). Legislative voting dynamics in Ukraine. In International Conference on Social Computing, Behavioral-Cultural Modeling and Prediction and Behavior Representation in Modeling and Simulation (pp. 82-88).

Talks, Posters, and Workshop Papers

Magelinski, T. & Carley, K. M. (2022). Contextualized Networks. Oral. Sunbelt 2022.

Magelinski, T. & Carley, K. M. (2022). Community Prototypes in Large Twitter Conversations. Oral. Sunbelt 2022.

Magelinski, T., Ng, L., & Carley, K. M. (2021). A Synchronized Action Framework for Responsible Detection of Coordination on Social Media. Workshop paper. Responsible Social Media Mining - MAISoN, Co-Located with IJCAI.

Magelinski, T. & Carley, K. M. (2021). Modularity Vitality for Bipartite Networks and Projections. Oral. Networks 21.

Angelopoulos, S., Brik, T., **Magelinski, T.**, & Carley, K. M. (2020). What you gonna do when they come for you? Network effect of information exposure on coalition formation. Sunbelt.

Harder, N., Brashears, M., Brik, T., Carley, K.M., **& Magelinski, T.** (2020). Understanding and Predicting Legislative Behavior in the Verkhovna Rada through New Methods of Ecological Modeling. Sunbelt.

Bhutani, M., **Magelinski, T.**, & Kolter, Z. (2019). Sinkhorn-Flow: Predicting Probability Mass Flow in Dynamical Systems Using Optimal Transport. Optimal Transport & Machine learning Workshop at NeurIPS.

Uyheng, J., **Magelinski, T.**, Cox, R. V., Sowa, C., & Carley, K. M. (2019). Information Operations Analysis of NATO Trident Juncture Exercise 2018. International Conference on Social Computing, Behavioral-Cultural Modeling, and Prediction and Behavior Representation in Modeling and Simulation.

Hou, J., **Magelinski, T.**, & Mylovanov, T. (2019). Minsk II Agreement between Russia and Ukraine and Polarization of the Ukrainian Parliament. Advancing Research through Computing Student Poster

Contest. **Winning Entry**.

Magelinski, T., Cruickshank, I., & Carley, K. M. (2018). Comparison of faction detection methods in application to Ukrainian parliamentary data. International Conference on Social Computing, Behavioral-Cultural Modeling, and Prediction and Behavior Representation in Modeling and Simulation.

Magelinski, T., & Carley, K. M. (2019). Effects of Network Aggregation in Simple Diffusion Simulations. International Conference on Social Computing, Behavioral-Cultural Modeling, and Prediction and Behavior Representation in Modeling and Simulation.

Magelinski, T., & Ross, S. (2016). Sources of uncertainty and inaccuracy in airdrop operations. Fall Fluid Mechanics Symposium.

Invited Talks

South Big Data Hub - Social Cybersecurity Working Group April 2023
Identity-Based Attribute Prototypes Distinguish Communities on Twitter. *Talk.*

Johns Hopkins Applied Physics Lab April 2023
Identity-Based Attribute Prototypes Distinguish Communities on Twitter. *Talk.*

NYU Center for Social Media and Politics January 2023
Adding Context to Models of Online Communities. *Talk.*

IDeaS Workshop on Linking Online Activity to Offline Behavior January 2023
Politics Panel. *Moderator.*

Journal of Trust and Safety Webinar February 2022
Synchronized Action Framework for Detection of Coordination on Social Media. *Interview.*

IDeaS Conference October 2021
Discovering Prototypical Attributes of Online Communities Through MultiView Modularity Vitality. *Talk.*

Hack4Impact's Bytes of Good Series February 2021
Discussed research and PhD life to get undergraduates interested in careers in social cybersecurity. *Interview.*

IDeaS Conference on Social-Cybersecurity in Times of Crisis and Change November 2020
Detecting Coordinated Behavior in the Twitter Campaign to Reopen America. *Talk.*