

Semih Cantürk

semihcanturk00@gmail.com | +90 (544) 925 4085
github.com/semihcanturk | linkedin.com/in/semihcanturk

EDUCATION

PhD & MSc in Computer Science | Mila Institute & Université de Montréal 2020–Present

Supervisor: Guy Wolf

MSc completed & PhD started: September 2022

MSc Thesis: *Taxonomy of Datasets in Graph Learning: A Data-Driven Approach to Improve GNN Benchmarking*.

GPA: 4.30/4.30

BSE in Systems Science & Engineering | University of Pennsylvania 2014–2018

BSE Thesis: *Motor Task Prediction through fMRI Data*

Supervisor: Victor Preciado

Minors in Computer Science & Mathematics

GPA: 3.68/4.00, Dean's List, *magna cum laude*

RESEARCH

My interests cover theory and applications of machine learning with a focus on graph representation learning (GRL). In addition, I am interested in domain adaptation, out-of-distribution generalization & reinforcement learning. More recently, I have been working on applying GRL on molecular data, improving the scalability of graph neural networks, and self-supervised learning on graphs.

Efficiently learning positional/structural embeddings for graph transformers: Graph transformers owe their recent advent as state-of-the-art in GRL to leveraging precomputed positional & structural embeddings (PE/SE). This reliance on PE/SEs are a significant bottleneck in the scalability of transformers, as they are expensive to compute and are not generalizable. My recent research focuses on how to *learn* robust and representations of PE/SEs that can generalize across datasets (as opposed to explicitly computing them), thus improving the scalability of graph transformers and potentially enabling generalizable pretrained models in molecular graph learning.

Taxonomy of graph learning benchmarks: Despite extensive research on developing superior GNN architectures, the set of graph datasets commonly used for model evaluation are insufficient to serve as distinguishing benchmarks. In particular, it is currently not well-understood what aspects of a given model these benchmarks probe. A main research interest of mine is to improve the state of benchmarking in graph learning. Thus, in addition to my recent Oral acceptance paper in LoG 2022 (see Publications) which taxonomizes existing benchmark datasets by a novel framework that uses graph perturbations to infer the essential data characteristics that GNNs models leverage in their predictions, I am currently working on several follow-up studies.

Supply chain analysis via graph representation learning: Supply chain data provides useful and unique insights about the interactions between companies and has been studied in academic finance literature in development of investment strategies. Our work constructs temporal graphs from supply chain data and aims to extract insights by applying graph learning and graph signal processing techniques to guide investment strategies.

Multi-agent reinforcement learning (MARL): Exploring goal-conditioned hierarchical planning (GCP) approaches and improving on existing work in the area. In particular, we aim to extend the current body of work in GCP to multi-agent RL settings.

WORK EXPERIENCE

INDUSTRY

Zetane Systems

Montreal, QC | 2019–Present

Researcher and Software Developer, Machine Learning

- Responsible for development of the machine learning explainability module (XAI) in the *Zetane Engine* for computer vision problems through approaches such as class-activation mapping (CAM) methods (e.g. GradCAM++, ScoreCAM, EigenCAM), as well as game-theoretic or surrogate approaches such as SHAP and LIME.
- Develop the dataset augmentation and model explainability (XAI) modules of *Zetane Protector* for object classification, object detection and semantic segmentation.

- Lead and support machine learning projects with industry partners in robotics, energy, construction and automotive industries.
- Integrated agent-based reinforcement learning algorithms to the Zetane Engine, thus enabling training of 3D agents in a simulated environment in real time.

InfoTRON

Istanbul, Turkey | 2017–2018

Software Development Intern

- Built a machine learning framework using ARToolkit and OpenCV to recognize and classify CAD models in AR/VR environments.
- Enabled InfoTRON automate object recognition in their AR/VR solutions.

SAS Analytics

Istanbul, Turkey | Summer 2016

Data Scientist Intern

- Worked on fraud detection projects with global industry partners in insurance sector.
- Completed training on SAS language as well as SAS Enterprise Guide & Enterprise Miner software

ACADEMIC

University of Pennsylvania

Philadelphia, PA | 2018–2020

Undergraduate Researcher, supervised by Victor Preciado & Cassiano Becker

- Built a machine learning pipeline that predict motor tasks based on brain activity (from fMRI data) using various signal processing techniques and neural networks.
- Won the Penn Engineering Societal Impact Award.
- After graduation, extended the project to use recurrent and graph-based learning techniques.

Imperial College

London, UK | Summer 2016

Undergraduate Researcher

- Built a distributed system that runs acute3D, a 3D-modeling software, in the Imperial College Data Observatory, the largest data visualization studio of its kind in Europe.
- Enabled the exploration of generated 3D models in the observatory through an WebGL-based viewer.

TEACHING ASSISTANTSHIPS

Mila Institute & Université de Montréal

Montréal, QC | Fall 2021

MAT 6495: Spectral Graph Theory/Théorie spectrale des graphes, Guy Wolf

University of Pennsylvania

Philadelphia, PA | Fall 2016

ESE 210: Introduction to Dynamical Systems, Robert Ghrist

AWARDS & LEADERSHIP

Bourse d'exemption, Université de Montréal | 2020-present

Penn Department of Electrical & Systems Engineering | 2018 Societal Impact Award

Study In America | Top 25 College Graduates of 2018, Turkey

Penn Preceptorials Committee | Chair (2015-2017), Treasurer (2014-2015)

SKILLS

Programming: Python, Java, C++, C#, JavaScript, Matlab, HTML, CSS, OCaml, SQL, SAS, Swift

ML Libraries & Software: PyTorch, TensorFlow, Spark, Hadoop, SAS Enterprise Guide & Miner, JMP

Computer Vision & Graphics: OpenCV, WebGL, ARToolkit

3D Graphics & CAD Software: AutoCAD, Rhinoceros 3D, AVEVA, acute3D

Languages: Turkish (Native), English (Fluent), Spanish (B2), Greek (B2), French (A2)

PUBLICATIONS

- **Cantürk** et al., Graph Positional and Structural Encoder. *Molecular ML (MoML)* 2023. May 29, 2023.
- **Cantürk** et al., Taxonomy of Benchmarks in Graph Representation Learning. *Proceedings of the First Learning on Graphs Conference (LoG 2022, Oral)*, PMLR 198, Virtual, December 9–12, 2022.
- **Cantürk** et al., Towards a Taxonomy of Graph Learning Datasets. *Data Centric AI (DCAI) workshop at NeurIPS 2021*. December 14, 2021.
- **Cantürk**, Singh, St-Amant & Behrmann. *Machine-Learning Driven Drug Repurposing for COVID-19*. ArXiv preprint, 2020.

PEER REVIEWING

- ICLR Workshop on Geometrical and Topological Representation Learning: 2022, 2023
- Learning on Graphs (LoG) conference: 2022