

Education

- 2015 **PhD**, *Carnegie Mellon University*, Pittsburgh.
Computational Biology/ School of Computer Science
Dissertation: *Inferring And Analyzing The Present And The Past Of Networks From Limited Information*
Advisor: Prof. Carl Kingsford **Committee:** Russell Schwartz, Seyoung Kim, Guy E. Blelloch
- 2011 **M.S.**, *University of Maryland*, College Park.
Computer Science
- 2008 **BEng**, *Bogazici University*, Istanbul, GPA – 3.8/4.0 (Ranked #1 in class).
Computer Engineering

Current Interests

Data Mining Applications in Finance, Bioinformatics, Social Networks, Machine Learning

Work Experience

Ozyegin University, Computer Science Department

Istanbul

2020 Sep - **Assistant Professor.**
Present

J.P. Morgan Chase, CIB

New York City

- 2020 Mar - **AI & Applied ML Researcher, Lead.**
- 2020 Sep
- Develop algorithms to detect fraudulent activities in market-making & trading through time-series analysis
 - Develop techniques to identify entitlement anomalies through interaction graph analysis

Goldman Sachs Asset Management, IMD

New York City

- 2018 Jan - **Vice President, Quantitative Strategist. Fundamental Equity (FE), \$60 billion mutual fund.**
- 2019 Nov
- Generated over 1.5 Sharpe signal on supply-chain dataset for biweekly-rebalanced quantamental fund
 - Developed optimal hierarchical portfolio construction for \$5 billion Exchange Fund, by backtesting historical inflows
 - Increased annual Exchange Fund inflow to \$1 billion, by developing rule-based framework to decide stock acceptance under 3 seconds
 - Designed & led the whole IMD Exchange Fund portfolio construction project
- 2015 Dec - **Associate, Quantitative Strategist. Goldman Sachs Investment Partners (GSIP), flagship**
- 2018 Jan **multistrategy hedge fund.**
- Developed systematic allocation strategies for Event driven fund by backtesting merger data
 - Developed volatility trading strategies for Chinese market through index options
 - Developed min-cost replication & hedging algorithms to rebalance \$1 billion Liquid Alt. funds under 2 minutes
 - Responsible for risk management of \$4 billion hedge fund

CMU, School of Computer Science

Pittsburgh

- 2015 Mar - **Machine Learning Postdoctoral Researcher,** CARNEGIE MELLON UNIVERSITY.
- Dec
- Focused on time-series analysis of lung development with Ziv Bar-Joseph at CMU Machine Learning Department
 - Examined the tradeoffs between dense and replicate sampling strategies for high-throughput time series experiments
- 2011-2015 **Research / Teaching Assistant,** CARNEGIE MELLON UNIVERSITY.
- Developed a method to deconvolve ensemble chromatin interaction data in cell subpopulations
 - Developed convex optimization-based method to predict information diffusion history over social network
 - Developed linear optimization-based approach to predict information diffusion network
 - Led recitation hours, and graded exams in 02-713: Algorithms & Data Structures for Scientists course

University of Maryland, School of Computer Science

College Park

- Developed MRF-based method to predict protein annotation with provably optimal guarantees
- Developed unsupervised tree-based method to reconstruct species interaction network history
- Led recitation hours, and graded exams in CMSC 131, CMSC 433, CMSC 423 courses

Publications

Journal Publications under Review

- [1] Batuhan Eralp and **Emre Sefer**. High order chromatin structure connect sQTLs with the splicing of distant genes. *Cells and Bioscience*, Under Review 2022
- [2] Zehra Erva Ergun and **Emre Sefer**. RiskBERT: Risk Prediction from Financial Reports through Transfer Learning. *Journal of Finance and Data Science*, Under Review 2022
- [3] Necla Nisa Soyulu and **Emre Sefer**. BERT2OME: Prediction of 2'-O-methylation Modifications from RNA Sequence by Transformer Architecture Based on BERT. *IEEE/ACM Transactions on Computational Biology and Bioinformatics*, Under Review 2022
- [4] **Emre Sefer**. SCMIN: Convex Inference of Scale-free Contact Networks over COVID Diffusion Data. *IEEE/ACM Transactions on Computational Biology and Bioinformatics*, Under Review 2022
- [5] Zehra Erva Ergun and **Emre Sefer**. FinSentiment: Predicting Financial Sentiment Through Transfer Learning. *Intelligent Systems in Accounting, Finance and Management* Under Review 2022
- [6] Batuhan Eralp and **Emre Sefer**. Joint Analysis of sQTL and Hi-C Reveals Spatial Proximity Between sQTLs and Target Genes in Cancer Tissues. *BMC Genomics*, Under Review 2022
- [7] **Emre Sefer**. A Comparison of Alignment Algorithms for Whole Genome Bisulfite Sequencing in Mammals. *Algorithms for Molecular Biology*, Under Review 2022

Journal Publications

- [13] **Emre Sefer**. MOCMIN: Convex Inferring of Modular Contact Networks over COVID Diffusion Data. *Turkish Journal of Electrical Engineering and Computer Sciences*, 2022
- [12] **Emre Sefer**. A Comparison of Topologically Associating Domain Callers over Mammals at High Resolution. *BMC Bioinformatics*, April 2022
- [11] **Emre Sefer**. Biocode: A Data-Driven Procedure to Learn the Growth of Biological Networks. *IEEE/ACM Transactions on Computational Biology and Bioinformatics*, April 2022
- [10] **Emre Sefer**. ProbC: Joint Modeling of Epigenome and Transcriptome Effects in 3D Genome. *BMC Genomics*, April 2022
- [9] **Emre Sefer**. Hi-C Interaction Graph Analysis Reveals the Impact of Histone Modifications in Chromatin Shape. *Applied Network Science*, July 2021
- [8] **Emre Sefer** and Carl Kingsford. Metric Labeling and Semi-metric Embedding for Protein Annotation Prediction. *Journal of Computational Biology*, May 2021
- [7] **Emre Sefer** and Carl Kingsford. Semi-nonparametric Modeling of Topological Domain Formation From Epigenetic Data. *Algorithms for Molecular Biology* 14 (1), 4. 2019
- [6] M Kleyman, **Emre Sefer**, Nicola, T., Espinoza, C., Chhabra, D., Hagood, J. S., Kaminski, N., Ambalavanan, N., and Ziv Bar-Joseph. Selecting the most appropriate time points to profile in high-throughput studies. *eLife Sciences* 2017
- [5] **Emre Sefer**, M Kleyman, and Ziv Bar-Joseph. Tradeoffs between Dense and Replicate Sampling Strategies for High-Throughput Time Series Experiments. *Cell systems* 3 (1), 35-42. 2016
- [4] **Emre Sefer**, Geet Duggal, and Carl Kingsford. Deconvolution Of Ensemble Chromatin Interaction Data Reveals The Latent Mixing Structures In Cell Subpopulations. *Journal of Computational Biology* 23 (6), 425-438. 2016
- [3] **Emre Sefer** and Carl Kingsford. Diffusion Archaeology for Diffusion Progression History Reconstruction. *Knowledge and Information Systems* 2016(2):530-539

- [2] Geet Duggal, Rob Patro, **Sefer, Emre**, Hao Wang, Darya Filippova, Samir Khuller, and Carl Kingsford. Resolving spatial inconsistencies in chromosome conformation measurements. *Algorithms for Molecular Biology*, 8(1):8, 2013
- [1] Rob Patro, **Emre Sefer**, Justin Malin, Guillaume Marcais, Saket Navlakha, Carl Kingsford. Parsimonious reconstruction of network evolution. *Algorithms for Molecular Biology* 2012 7:25

Conference Publications with Proceedings (Refereed)

- [15] Tuna Tuncer, Uygur Kaya, **Emre Sefer**, Onur Uralcam, Tugcan Hoser. Asset Price and Direction Prediction via Deep 2D Transformer and Convolutional Neural Networks. ICAIF 2022, 3rd ACM International Conference on AI in Finance
- [14] Ugur Dolu and **Emre Sefer**. A Novel GBT-based Approach for Cross-Channel Fraud Detection on Real-World Banking Transactions. AIAI 2022, 18th International Conference on Artificial Intelligence Applications and Innovations
- [13] **Emre Sefer**. Joint Modeling of Histone Modifications in 3D Genome Shape Through Hi-C Interaction Graph. *Complex Networks* 2020
- [12] **Emre Sefer**, and Ziv Bar-Joseph. Shall we dense? Comparing design strategies for time series expression experiments*. RECOMB 2016, ***Winner of the Best Paper Award**
- [11] **Emre Sefer**, Geet Duggal, and Carl Kingsford. Deconvolution Of Ensemble Chromatin Interaction Data Reveals The Latent Mixing Structures In Cell Subpopulations. RECOMB 2015
- [10] **Emre Sefer** and Carl Kingsford. Convex Risk Minimization To Infer Networks From Probabilistic Diffusion Data At Multiple Scales. ICDE 2015
- [9] **Emre Sefer** and Carl Kingsford. Semi-nonparametric Modeling of Topological Domain Formation From Epigenetic Data. WABI 2015
- [8] **Emre Sefer** and Carl Kingsford. Diffusion Archaeology for Diffusion Progression History Reconstruction. ICDM 2014
- [7] Geet Duggal, Rob Patro, **Sefer, Emre**, Hao Wang, Darya Filippova, Samir Khuller, and Carl Kingsford. Resolving spatial inconsistencies in chromosome conformation measurements. WABI 2012
- [6] Robert Patro, Geet Duggal, **Emre Sefer**, Hao Wang, Darya Filippova, and Carl Kingsford. The missing models: a data-driven approach for learning how networks grow*. KDD 2012, ***Winner of Best Video Award**
- [5] Rob Patro, **Emre Sefer**, Justin Malin, Guillaume Marcais, Saket Navlakha, Carl Kingsford. Parsimonious reconstruction of network evolution. WABI 2011
- [4] Robert Gove, Nick Gramsky, **Emre Sefer**, Ben Shneiderman. NetVisia: Heat map & matrix visualization of dynamic social network statistics & content. SocialCom 2011
- [3] **Emre Sefer** and Carl Kingsford. Metric labeling and semi-metric embedding for protein annotation prediction. RECOMB 2011
- [2] Dana Nau, **Emre Sefer**, Ugur Kuter. Thinking ahead in real-time search. ICAPS 2009
- [1] **Emre Sefer**, Ugur Kuter, Dana Nau. Real-time A* search with depth-k lookahead. International Symposium on Combinatorial Search, SoCS 2009

Journal & Conference Abstracts with Proceedings

- [1] Teodora Nicola, **Emre Sefer**, et al. Identification Of Optimal Time Points And Proteomic Profiling During Murine Lung Alveolar Septation [abstract]. *American Journal of Respiratory and Critical Care Medicine* 2016;193:A6561
- [2] C. R. Espinoza, D. Chhabra, T. Nicola, N. Ambalavanan, N. Kaminski, **Emre Sefer**, Z. Bar-Joseph, J. S. Hagood. Dynamic Changes of DNA Methylation During Different Stages of Normal Mouse Lung Development [abstract]. *American Journal of Respiratory and Critical Care Medicine* 2016;193:A2344

Supervised Thesis

- Ugur Dolu, 2022, Novel Sampling Technique and Gradient Boosting Tree-based Approach for Cross-channel fraud detection, Ozyegin University Data Science Department

Invited Talks & Tutorials

Peer-reviewed Conference Presentations

- Biocode: A Data-Driven Approach for Learning How Biological Networks Grow. BIOKDD in KDD 2021, Singapore. 15/08/2021
- MOCMIN: Convex Inferring of Modular Contact Networks over COVID Diffusion Data. Communities in Networks Satellite in Networks 2021, Indiana, USA. 01/07/2021
- Joint Modeling of Histone Modifications in 3D Genome Shape Through Hi-C Interaction Graph. Complex Networks 2020, Madrid, Spain. 02/12/2020
- Shall we dense? Comparing design strategies for time series expression experiments. RECOMB 2016, Los Angeles, USA. 10/04/2016
- Semi-nonparametric Modeling of Topological Domain Formation From Epigenetic Data. WABI 2015, Atlanta, USA. 15/09/2015
- Deconvolution Of Ensemble Chromatin Interaction Data Reveals The Latent Mixing Structures In Cell Subpopulations. RECOMB 2015, Warsaw, Poland 18/04/2015
- Convex Risk Minimization To Infer Networks From Probabilistic Diffusion Data At Multiple Scales. ICDE 2015, Seoul, South Korea. 12/04/2015
- Diffusion Archaeology for Diffusion Progression History Reconstruction. ICDM 2014, Shenzhen, China. 18/12/2014
- Metric labeling and semi-metric embedding for protein annotation prediction. RECOMB 2011, Vancouver, Canada. 10/04/2011
- Real-time A* search with depth-k lookahead. International Symposium on Combinatorial Search. SoCS 2009, Los Angeles, USA. 14/08/2009

Tutorials

- Finding Topological Domains in Genome. ACM-BCB 2015, Atlanta, USA. 12/09/2015

Special Invited Talks

- Analyzing The Present And The Past Of The Networks From Limited Information. Ozyegin University Computer Science Department, Istanbul, Turkey. 14/04/2020
- Analyzing The Present And The Past Of The Networks From Limited Information. Istanbul Technical University, Faculty of Computer Engineering, Istanbul, Turkey. 04/04/2020
- Analyzing The Present And The Past Of The Networks From Limited Information. Yeditepe University, Computer Engineering Department, Istanbul, Turkey. 27/03/2020
- Analyzing The Present And The Past Of The Networks From Limited Information. Kadir Has University Computer Science Department, Istanbul, Turkey. 20/03/2020
- Deconvolution Of Ensemble Chromatin Interaction Data Reveals The Latent Mixing Structures In Cell Subpopulations. Ziv-Bar Joseph's Systems Biology Research Group at CMU Machine Learning Department, Pittsburgh, PA. 14/02/2015

Posters

- Convex Risk Minimization To Infer Networks From Probabilistic Diffusion Data At Multiple Scales. GLBIO 2015, Pittsburgh, USA. 12/04/2015
- Real-time A* search with depth-k lookahead. International Symposium on Combinatorial Search. SoCS 2009, Los Angeles, USA. 14/08/2009

Professional Contributions

- Serving as an editor for Frontiers in Artificial Intelligence journal
- Serving as an editor for Journal of Advances in Management Sciences & Information Systems journal
- Gave **tutorial** about topological domains in 3D genome at ACM-BCB 2015
- Served as a **Program Committee Member** for ACM-BCB 2015
- Serve as a **Program Committee Member** for IJCAI 2021, IJCAI 2022
- Serve as a **Program Committee Member** for Complex Networks 2021, 2022
- Served as a **Program Committee Member** for ASYU 2022
- Worked as a part of NIH funded **LungMAP** project to map lung dynamics (<https://www.lungmap.net/>)
- Reviewed papers for RECOMB 2012, RECOMB 2013, RECOMB 2014, RECOMB 2015, ISMB 2012, ISMB 2014, WABI 2015, ACM-BCB 2015 conferences
- Reviewed papers for Plos Computational Biology, International Review of Financial Analysis, Bioinformatics Advances, Frontiers in Artificial Intelligence, Frontiers in Psychology, Frontiers in Cell and Developmental Biology, Genome Research, Bioinformatics, BMC Bioinformatics, Journal of Computational Biology, Nucleic Acids Research, IEEE Transactions on Computational Biology and Bioinformatics, Optimization Letters, Briefings in Bioinformatics journals

Teaching Experience

Assistant Professor,

OZYEGIN UNIVERSITY.

- Fall, 2020. CS 201: Data Structures and Algorithms
- Spring, 2021. CS 333: Algorithms Analysis
- Spring, 2021. CS 440/540: Machine Learning in Finance
- Spring, 2021. CS 533: Advanced Algorithms
- Fall, 2021. CS 104: Introduction to Programming
- Fall, 2021. CS 440/540: Machine Learning in Finance
- Spring, 2022. CS 333: Algorithms Analysis
- Spring, 2022. CS 540: Machine Learning in Finance
- Spring, 2022. CS 412/512: Bioinformatics Algorithms

Other Achievements

- Received **Best Paper Award** at RECOMB 2016
- Received University of Maryland Computer Science Fellowship during graduate studies
- Graduated from Bogazici University in the **1st** rank with High Honors
- Received Bogazici University Full Scholarship during undergraduate studies
- Received **Best Video Award** at KDD 2012 conference techtalks.tv/talks/57076/
- Received fellowships for RECOMB 2011, RECOMB 2015, ICDM 2014, ICDE 2015, ICAPS 2009 conferences
- Ranked 394th among 1.500.000 people in Turkey University Entrance Exam (OSS)
- Ranked 17th among 100.000 people in Turkey Graduate Education Test (ALES) 2008

Computer skills

Advanced	PYTHON, SLANG/SECDB (Goldman Sachs' proprietary language), Optimization Software (AXIOMA, CPLEX, GUROBI, AMPL, IMSL)
Intermediate	C++, MATLAB, R, SQL, SPARK