

Xiao (Thomas) Zhang

<http://www.umich.edu/~tomzhang>

203-856-1378

tomzhang@umich.edu

CAREER I am looking for challenging opportunities in data science/research/software engineering.
STATEMENT I am authorized to work in the US under F1-OPT and I will need H1B/O1 visa sponsorship.

SKILLS **Programming Language:**

- *Advanced:* Python, Matlab, Gephi, Latex.
- *Intermediate:* C++/C, Mathematica, R, Git, Go, Spark, Javascript, SQL, MongoDB, Neo4j.

Analytical — Statistical Inference, Machine Learning, Network Science, Graph algorithms, Bayesian Inference, Natural Language Processing, Monte Carlo methods, Advanced Linear Algebra.

EXPERIENCE **Data Scientist**, Owler, San Mateo, CA, Aug, 2017 - Current
Leading data scientist working with a dataset of 12M+ companies and 1.5M+ users; Implemented product based on a competition network of 1.5M+ companies and 40M+ edges using network analysis and machine learning techniques; Collaborated with engineers building microservice using grpc and collaborated with product team for product specs.

- Implemented community detection and graph analysis algorithms for clustering companies to provide insights for market analysis and sales optimization.
- Used NLP techniques for business news categorization and information retrieval.
- Built news recommendation system applying data mining techniques and Bayesian models.
- Implemented systems for data cleaning using anomaly detection machine learning algorithms.

Graduate Student Research Assistant, University of Michigan, Sep, 2012 - June, 2017
My research focused on network science and resulted in 5 top-tier journal publications and received 200 citations as of January 2018.

- Analyzed and proved limitations spectral clustering algorithm using random matrix theory.
- Implemented message passing inference algorithm for importance-based division on random graph models. (Python, C++)
- Proposed a new regularized ranking/centrality measure in networks to avoid anomalous measure. (Python)
- Proposed a new spectral clustering algorithm that outperforms k-means for community detection in networks. (Python)
- Derived models for time-varying network data and proposed learning algorithm for fitting real-world data into the models. (Python, C++)
- Goodness of fit and model selection of network models. (Python)

Selected Class Projects, Ann Arbor, MI, Sep, 2012 - current

- Message passing algorithm for computing graph spectra. (Python, EECS 598, Random Matrix Theory)
- Numerical percolation thresholds for lattices using Monte Carlo algorithm. (Python, PHYS 514, Computational Physics)

Santa Fe Complex Systems Summer School, Santa Fe, NW, June, 2016

- Used regression methods to study the network properties of a large movie data set(over 5000 movies and 200Mb in size). (Python)
- Studied the data of epidemics spreading in South Africa's ostrich farm using dynamic models. (Python)

Undergraduate Student Researcher, Troy, NY, Jan, 2011 - May, 2012

- Worked on the Daya Bay Neutrino Experiment, a large international particle physics collaboration with over 100 scientists.

EDUCATION **University of Michigan**, Ann Arbor, MI
Ph.D in Physics, Sep, 2012 - June, 2017
Thesis: Statistical Inference and Spectral Methods for Network Analysis. Advisor: Prof. Mark Newman
Rensselaer Polytechnic Institute, Troy, NY
B.S. in Physics, Sep, 2009 - May, 2012 – GPA: 3.87, Major: 3.93