Noemi Derzsy

E-mail: noemiderzsy@gmail.com
Mobile: (518) 248-0470
Website: www.noemiderzsy.com
Twitter: @NoemiDerzsy
LinkedIn: www.linkedin.com/in/NoemiDerzsy
GitHub: github.com/nderzsy

EDUCATION

2008 - 2012 *Ph.D in Physics*
Faculty of Physics, Babeş - Bolyai University, Cluj-Napoca, Romania

2007 - 2009 *M.S. in Computational Physics*
Faculty of Physics, Babeş - Bolyai University, Cluj-Napoca, Romania

2003 - 2007 *B.S. in Physics and Computer Science*
Faculty of Physics, Babeş - Bolyai University, Cluj-Napoca, Romania

PROFESSIONAL EXPERIENCE

2018 – Present **Data Science Fellow**
Insight Data Science, New York, New York

- Developed "*Tariff Code Finder: NLP to Detect Loopholes in the US Tariff Code,*" a project to identify and reduce ambiguities in the current tariff schedule for items that can be mislabeled or misused.
- Built a tariff code classifier in Python using tf-idf and word2vec embedding of item descriptions to accurately classify products into the correct tariff category and avoid mislabeling.
- Built a web app using Flask with PostgreSQL, and deployed it on AWS to showcase the project: http://item2tariffcode.site/

2017 – Present **Course Instructor**
DataCamp

- Prepared coursework/curriculum for course titled "*Hypothesis Testing in Python*" (forthcoming).
- Developed material and code exercises in Python for one sample, two sample, multiple sample hypothesis testing and A/B testing.

2013 – Present **Postdoctoral Research Associate**
Social Cognitive Networks Academic Research Center (SCNARC), Rensselaer Polytechnic Institute, Troy, New York
Department of Physics, Applied Physics, and Astronomy, Rensselaer Polytechnic Institute, Troy, New York

- Analyzed criminal activities in city neighborhoods using public crime records from the City of Chicago, and revealed distinct crime patterns in high- vs. low crime regions using Python and Bokeh visualizations.
- Constructed city neighborhood network from geodata and modeled criminal activities on this network as an infectious epidemic spread using C++ and Python, and revealed from census data the demographic properties that correlate with criminal activities.
- Studied the impact of criminal activities in city neighborhoods on real-estate market using Zillow real-estate data in Python.

- Developed a predictive algorithm in Python to predict criminal activities from real-estate market fluctuations, and vice-versa, to predict from criminal activities real-estate market fluctuations in city neighborhoods.
- Developed code/algorithms in C++ to efficiently influence, monitor and control nodes in complex social networks, methods that are more cost-efficient computationally and financially, and also provide 60% higher resilience against targeted attacks and random failures in networks compared to previously existing methods.
- Analyzed the dynamics of cascading failures in complex networks, applicable to opinion formation and influencing in social networks, and developed mitigation strategies in C++ to reduce system damage by 30%.

2016 – Present **NASA Datanaut**
NASA - National Aeronautics and Space Administration
- Used NLP and topic modeling techniques in Python to understand and model open government data associations.
- Used network analysis to build term-keyword network structures to gain understanding of open NASA data ontologies.
- Analyzed the accuracy of human-supplied keywords and descriptions of open NASA datasets using NLP tools in Python.

2011 - 2011 **Visiting Research Student**
Universidade do Porto, Porto, Portugal
- Analyzed and modeled income distributions in Cluj district, Romania in SQL and C++ using a large-scale social security database.
- Modeled the student mobility social network between European universities using C++ and a dataset of the Erasmus scholarship exchange program.
- Investigated and modeled the user communication behavior in the Enron e-mail communication datasets using C++.

2010 - 2012 **Physicist**
Department of Legal Metrology, Cluj-Napoca, Romania
- Verified, calibrated and authorized measuring instruments used in public interest.

2007 - 2010 **IT Specialist**
Social Security Administration, Cluj-Napoca, Romania
- Implemented new legislation and organization procedures in existing database applications and developed new user-friendly applications using SQL. Recognized the need and provided technical support and training to colleagues.
- Generated periodical and on-request reports, queries from the social security databases using SQL.

BOOKS
N. Derzsy, *Understanding Crime Through Science - Interdisciplinary Analysis and Modeling of Criminal Activities*, Springer, 2018 (forthcoming)

BOOK CHAPTERS
N. Derzsy, B. K. Szymanski, G. Korniss, *Neighborhood crime and real estate market: from analysis to prediction*, in "Understanding Crime Through Science - Interdisciplinary Analysis and Modeling of Criminal Activities", Springer, 2018 (forthcoming)

SCIENTIFIC JOURNAL PUBLICATIONS

Derzsi, N. Derzsy, E. Káptalan, Z. Néda, *Topology of the Erasmus student mobility network*, Physica A 390, 2601 (2011)

N. Derzsy, Z. Néda, M. A. Santos, *Income distribution patterns from a complete social security database*, Physica A 391, 5611 (2012)

N. Derzsy, *Employer mobility network revealed from social security data*, STUDIA UBB PHYSICA, Vol. 57(LVII), 1, 25 (2012)

F. Molnár Jr., N. Derzsy, É. Czabarka, L. Székely, B. K. Szymanski, G. Korniss, *Dominating Scale-Free Networks Using Generalized Probabilistic Methods*, Scientific Reports 4, 6308 (2014)

F. Molnár Jr., N. Derzsy, B. K. Szymanski, G. Korniss, *Building Damage-Resilient Dominating Sets in Complex Networks against Random and Targeted Attacks*, Scientific Reports 5, 8321 (2015)

A. Moussawi, N. Derzsy, X. Lin, B. K. Szymanski, G. Korniss, *Limits of Predictability of Cascading Overload Failures in Spatially-Embedded Networks with Distributed Flows,* Scientific Reports 7, 11729 (2017)

X. Niu, A. Moussawi, N. Derzsy, B. K. Szymanski, G. Korniss, *Evolution of the Global Risk Network Mean-Field Stability Point*, Proceedings of Complex Networks 2017, The Sixth International Conference on Complex Networks and Their Applications (2018)

N. Derzsy, F. Molnár Jr., B. K. Szymanski, G. Korniss, *Utilizing Maximal Independent Sets as Dominating Sets in Scale-Free Networks* (preprint)

SCIENTIFIC JOURNAL PEER-REVIEWING

- Journal of Statistical Mechanics: Theory and Experiment
- Chaos: An Interdisciplinary Journal of Nonlinear Science
- Theoretical Computer Science
- Computational Social Networks

PROGRAM COMMITTEE

- NetCrime 2018 - 3rd Symposium on the Structure and Mobility of Crime, Paris, France, June 12, 2018
- CompleNet'17 - 8th Conference on Complex Networks, Dubrovnik, Croatia, March 21-24, 2017
- ChASM 2016 - Computational Approaches to Social Modeling, Bellevue, WA, November 14, 2016
- ChASM 2014 - Computational Approaches to Social Modeling, Bloomington, IN, June 24-26, 2014

JUDGING

- Review Panel Member, U.S. Department of Homeland Security Science & Technology Directorate's *Hidden Signals Data Science Challenge*

MENTORING

- Mentor, U.S. Department of Homeland Security Science & Technology Directorate's *Hidden Signals Data Science Challenge*

ORGANIZED TUTORIALS

- *Network Analysis in Python*, workshop for Data Science & Data Engineer Fellows at Insight NYC
- *Analyzing Geolocated Data with Twitter*, WebSci'17 - 9th International ACM Web Science Conference 2017
- *Machine Learning for Open NASA Data* (scheduled), NASA Datanaut program

ORGANIZED EVENTS

- NASA Datanauts Community Event: Working with NASA' s Open Data, Rensselaer Polytechnic Institute, April 6, 2017

INVITED TALKS

2018	February	Panel Member at ***Women in Data Science Careers***, at NYU Center for Data Science
2017	December	***Topic Modeling Open NASA Data***, at NASA Datanaut program
2017	November	***Data Science in a Networked Era***, at NASA Datanaut program
2017	November	***Chicago Crime: From Analysis to Prediction,*** at Rensselaer Polytechnic Institute
2017	October	***Accessing and Analyzing Open NASA Data***, at NASA Datanaut program
2017	October	***Dominating Sets***, in Frontiers of Network Science class, at Rensselaer Polytechnic Institute
2017	May	***Social Contagion of Criminal Activities in City Neighborhood Networks***, at Army Research Lab NS CTA Annual Technical Meeting
2017	April	***Open NASA Data: From API to Data Analysis,*** at SpaceApps Bootcamp
2017	April	***Open NASA Data: Visualization and Analysis,*** at NASA Johnson Space Center
2017	April	***Open NASA Data: From API to Data Analysis***, at NASA Datanaut program
2017	April	***Open NASA Data Analysis***, at Rensselaer Polytechnic Institute
2017	March	***Open NASA Data: From API to Data Analysis***, at NYU Center For Data Science
2017	February	***Network Science & Data Science Project***, at NASA Datanaut Kickoff 2017, at NASA HQ
2016	December	***Network Science Meets Data Science***, at NASA Datanaut program
2016	October	***Dominating Sets Part 1: Algorithms and Stability***, at Rensselaer Polytechnic Institute
2016	October	***Dominating Sets Part 2: Overdomination and Stability***, at Rensselaer Polytechnic Institute
2015	October	***Dominating Sets,*** at Rensselaer Polytechnic Institute
2014	November	***Dominating Sets****: Algorithms and Stability*, at Rensselaer Polytechnic Institute
2011	June	***Employee mobility generated network and Pareto's law - revealed from social security***

data, Physics Seminar talk at University of Porto, Portugal

CONTRIBUTED TALKS / CONFERENCE PARTICIPATIONS

2017 November ***Data Science Keys to Open Up OpenNASA Datasets***, PyData, New York, NY, USA

2017 October ***Natural Language Processing from Scratch***, PyGotham, New York, NY, USA

2017 September ***Topic Modeling Open NASA Data***, STRATA Data Conference, New York, NY, USA

2017 August JupyterCon Conference, New York, NY, USA

2017 June Artificial Intelligence Conference, New York, NY

2017 June ***Analyzing Geolocated Data with Twitter***, WebSci, Troy, NY, USA

2017 June ***Chicago Crime and House Prices: From Analysis to Prediction***, NetCrime – 2nd Symposium on the Structure and Mobility of Crime, Indianapolis, IN, USA

2017 May Open Data Science Conference (ODSC) East, Boston, MA, USA

2017 May ***Social Contagion of Criminal Activities in City Neighborhood Networks***, ARL NS CTA Annual Technical Meeting, Newark, DE, USA

2017 April ***Open NASA Data: Visualization & Analysis***, JSC Data Science Day 2.0, Houston, TX, USA

2017 April ***Open NASA Data: From API to Data Analysis***, SpaceApps Bootcamp, New York, NY, USA

2017 April MariaDB User Conference, New York City, USA

2017 March 11th Annual Machine Learning Symposium, New York Academy of Sciences, New York, USA

2017 March ***Cascading Failures in Flow-Driven Networks Induced by Multiple Initiators***, APS March Meeting, New Orleans, Louisiana, USA

2016 July ***Cascading Failures in N Stable and N-1 Stable System Configurations of the European Power Grid,*** DTRA Meeting, Springfield, Virginia, USA

2016 March ***Utilizing Maximal Independent Sets as Dominating Sets in Scale-Free Networks***, APS March Meeting, Baltimore, Maryland, USA

2015 July ***Cascading Failures and Stochastic Analysis for Mitigation in Spatially-Embedded Random Networks,*** DTRA Meeting, Springfield, Virginia, USA

2015 June ***Highly Damage-Resilient Dominating Sets in Complex Networks against Random and Targeted Attacks***, NetSci - The International School and Conference on Network Science, Zaragoza, Spain

2015 June ***Cascading Failures and Stochastic Mitigation in Spatially-Embedded Random Networks***, NetSci Satellite Symposium - Information, Self-Organizing Dynamics, and Synchronization on Complex Networks II, Zaragoza, Spain

2015 March ***Cascading Failures and Stochastic Analysis for Mitigation in Spatially-Embedded Random***

Networks, APS March Meeting, San Antonio, Texas, USA

2014 July DARPA Workshop, Arlington, Virginia, USA

2014 June ***Stability of Dominating Sets in Complex Networks against Random and Targeted Attacks***, NetSci - The International School and Conference on Network Science, San Francisco, California, USA

2014 March ***Scaling of Various Dominating Sets in Scale-Free and Empirical Networks,*** APS March Meeting, Denver, Colorado, USA

2013 May ***Topology of the Erasmus Student Mobility Network***, IMI Summer School on Network Science, Columbia, South Carolina, USA

2012 June Aalto Complex Networks Factory, Helsinki, Finland

2011 June ***Topology of the Erasmus Student Mobility Network***, NetSci - The International School and Conference on Network Science, Budapest, Hungary

AWARDS, HONORS

- 2017 O'Reilly Conference Diversity and Inclusion Scholarship to Artificial Intelligence Conference
- 2017 Advanced Python Programming scholarship by RMORT & WWC
- 2017 Open Data Science Conference (ODSC) East scholarship
- 2016 AI with the Best & WWC scholarship
- 2016 Datanaut (Data Scientist) at NASA's Datanauts program
- 2003 Physics Olympiad, 3rd prize, regional
- 2003 English Language and Literature Olympiad, 2nd prize, regional
- 2002 Physics Olympiad, 2nd prize, regional
- 2002 English Language and Literature Olympiad, 2nd prize, regional
- 2001 Physics Olympiad, 2nd prize, regional
- 2000 Physics Olympiad, 1st prize, regional

CERTIFICATIONS

- Cambridge Certificate of Advanced English (2007)
- Professional Leadership Series - Rensselaer Polytechnic Institute (2015)
- Stanford Machine Learning MOOC Coursera (2016)
- Python Data Structures MOOC Coursera (2016)
- Advanced Python Programming RMOTR (2017)
- Interactive Data Visualization with Python and Bokeh (2017)

TEACHING EXPERIENCE

- 2011-2012 Teaching Assistant Spectroscopy and Lasers

TECHNICAL SKILLS

- Programming languages: C/C++, Python, SQL
- Machine Learning tools/packages: Scikit-learn, Keras, Tensorflow
- Visualization tools: Matplotlib, Bokeh, Seaborn, OriginLab, Graphviz, Gephi, QGIS, CARTO
- Other tools: Pandas, SciPy, NumPy, NLTK, SpaCy, BeautifulSoup, Flask, PostgreSQL, GitHub

ACTIVITIES, MEMBERSHIPS

- New York Academy of Sciences (NYAS)
- American Physical Society (APS)
- Data Management Association (DAMA) NY Capital Region
- Women Techmakers
- Active member of Data Science, Deep Learning and AI meetups in New York

HACKATHONS/WORKSHOPS

- Major League Hacking RPI (2017)
- STAN Workshop NYC (2017)
- NYC Open Data Workshop NYC (2017)
- NASA SpaceApps Hackathon NYC (2017)
- Smart Cities Hackathon NYC (2017)
- Markdown, Jupyter, GitHub as a new publishing paradigm, Rensselaer Polytechnic Institute (2017)
- Scikit-learn Sprint: Contributing to Open Source Workshop, WiMLDS, NYC (2017)
- Neural Networks Workshop, WiMLDS, NYC (2017)
- Major League Hacking RPI (2016)
- Git/GitHub Workshop, Digital Ocean, NYC (2016)

BLOG POSTS

- Women in Machine Learning and Data Science event: http://wimlds.org/noemi-derzsy-scikit-learn-sprint/

LANGUAGES

- Hungarian: Native proficiency
- Romanian: Native proficiency
- English: Bilingual proficiency
- German: Elementary proficiency