

# Cagri Ozcaglar

Research Scientist  
Amazon  
345 Boren Avenue North  
Seattle, WA 98109

Mobile: +1-(518)-694-6920  
Email: [cagriozcaglar@gmail.com](mailto:cagriozcaglar@gmail.com)  
<http://www.cs.rpi.edu/~ozcagc2>

- Education**
- **Ph.D., Computer Science** August 2012  
Rensselaer Polytechnic Institute Troy, NY  
**Thesis:** Algorithmic data fusion methods for tuberculosis  
**Advisor:** Prof. Bulent Yener
  - **M.S., Computer Science** May 2008  
Rensselaer Polytechnic Institute Troy, NY  
**Thesis:** Classification of email messages into topics using Latent Dirichlet Allocation  
**Advisors:** Prof. Sibel Adali, Prof. Boleslaw Szymanski
  - **B.S., Computer Science** June 2006  
Bilkent University Ankara, Turkey

**Research Interests** Data Mining, Machine Learning, Data Fusion, Multiway Analysis, Bioinformatics, Population Biology, Disease Spread Modeling, Social Network Analysis

- Skills**
- **Languages:** Java, C++, Python, R, MATLAB, Perl, PHP, HTML, Shell Scripting.
  - **Database Management:** SQL, PostgreSQL, MySQL, Oracle 10g/11g.
  - **Big Data Processing Technologies:** Hadoop, Pig, AWS.
  - **Tools and Applications:** Latex, Eclipse, Visual Studio .NET, CVS, SVN.
  - **Operating Systems:** Windows, Unix/Linux.

- Experience**
- **Amazon** Seattle, WA  
*Research Scientist* September 2013 - Present
    - ◊ Large-scale machine learning with Consumer Marketing Analytics team.
  - **Bank of America Merrill Lynch** New York, NY  
*Software Developer* August 2012 - August 2013
    - ◊ Contributed to the design and implementation of TESS real time feeds (RTF).
    - ◊ Implemented a connector between Access Request Management (ARM) and RAM DB.
  - **Rensselaer Polytechnic Institute** Troy, NY  
*Research Assistant* 2006 - August 2012
    - ◊ **TB-Insight:** <http://tbinsight.cs.rpi.edu/>  
*Supervisor:* Prof. Bulent Yener (PI: Prof. Kristin Bennett)
      - UBF: Developed the Unified Biclustering Framework (UBF) to find host-pathogen associations among *M. tuberculosis* complex strains and TB patients.
      - Developed an algorithm to find the mutation history in the DR region of *M. tuberculosis* complex and found topological attributes of the resulting phylogenetic tree. Built two new models, Starting Point Model (SPM) and Longest Block Model (LBM) for the mutation length frequency.

- TCF: Developed the Tensor Clustering Framework (TCF) to cluster *M. tuberculosis* complex strains into coherent groups using multiple-biomarker tensors.
- TB-Vis: Designed and implemented a visualization program. The visualiations include: 1) Spoligoforests which display spoligotype evolution using various distance measures of genomic data of *M. tuberculosis* complex, 2) Host-pathogen maps which display patient characteristics classified by the genotype of *M. tuberculosis* complex which infects the patients. Spoligoforests are used in TB-Lineage tool at [http://tbinsight.cs.rpi.edu/about\\_tb\\_lineage.html](http://tbinsight.cs.rpi.edu/about_tb_lineage.html). Both spoligoforests and host-pathogen maps are used in the design of interactive version of TB-Vis at [http://tbinsight.cs.rpi.edu/about\\_tb\\_vis.html](http://tbinsight.cs.rpi.edu/about_tb_vis.html).
- ◊ **Enron email classification into topics**  
*Supervisor:* Prof. Sibel Adali, Prof. Boleslaw Szymanski
  - Classified email messages of Enron into topics using the Latent Dirichlet Allocation. Described new metrics for classification assessment of email distribution into topics.
- ◊ **MetPetDB: <http://metpetdb.rpi.edu/>**  
*Supervisor:* Prof. Sibel Adali, Prof. Boleslaw Szymanski
  - Contributed to the design and implementation of a client-server architecture of a database for Metamorphic Petrology.
- **TUBITAK (Scientific and Technological Research Council)** Ankara, Turkey  
*Software Engineer Intern* Summer 2005
  - ◊ Designed and implemented a program which calculates the salaries of the employees using their entrance and exit time to/from the building.
- **TRT (Turkish Radio and Television)** Ankara, Turkey  
*Software Engineer Intern* Summer 2004
  - ◊ Contributed to implementation of the user interface of Eurovision Song Contest website of Turkey.
- Teaching**
  - **Rensselaer Polytechnic Institute** Troy, NY  
*Teaching Assistant*
    - ◊ Introduction to Algorithms Fall 2010
    - ◊ Computer Science II: Data Structures Spring 2008
    - ◊ Computer Organization Spring 2007
    - ◊ Artificial Intelligence Fall 2006
  - **Bilkent University** Ankara, Turkey  
*Undergraduate Teaching Assistant*
    - ◊ Discrete Mathematics Spring 2006
- Journal Papers**
  - K. P. Bennett, C. Ozcaglar, J. Ranganathan, S. Raghavan, J. Katz, D. Croft, B. Yener, A. Shabbeer. **TB-vis: Visualizing TB patient-pathogen relationships.** *Tuberculosis*, 2013.
  - A. Shabbeer, C. Ozcaglar, K. P. Bennett. **Crossing minimization within graph embeddings.** *arXiv*, 2012.
  - C. Ozcaglar, A. Shabbeer, N. Kurepina, N. Rastogi, B. Yener, K. P. Bennett. **Inferred spoligoforest topology unravels spatially bimodal distribution of mutations in the DR region.** *IEEE Transactions on NanoBioscience*, 2012.
  - C. Ozcaglar, A. Shabbeer, S. Vandenberg, B. Yener, K. P. Bennett. **Epidemiological models of *Mycobacterium tuberculosis* complex infections.** *Mathematical Biosciences*, 2012. (Most accessed paper of *Mathematical Biosciences* journal between March 2012 and December 2012).

- A. Shabbeer, L. Cowan, C. Ozcaglar, N. Rastogi, S. L. Vandenberg, B. Yener, K. P. Bennett. **TB-Lineage: an online tool for classification and analysis of strains of *Mycobacterium tuberculosis* complex.** *Infection, Genetics and Evolution*, 2012.
- A. Shabbeer, C. Ozcaglar, B. Yener, K. P. Bennett. **Web tools for molecular epidemiology of tuberculosis.** *Infection, Genetics and Evolution*, 2012. (Most accessed paper of *Infection, Genetics and Evolution* journal between December 2011 and June 2012).
- C. Ozcaglar, A. Shabbeer, S. Vandenberg, B. Yener, K. P. Bennett. **Sublineage structure analysis of *Mycobacterium tuberculosis* complex strains with multiple-biomarker tensors.** *BMC Genomics*, 2011.

Conference/  
Workshop  
Papers

- S. Salem, C. Ozcaglar. **MFMS: Maximal frequent module set mining from multiple human gene expression datasets.** *ACM SIGKDD International Workshop on Data Mining in Bioinformatics (BIOKDD)*, Chicago, August 2013.
- C. Ozcaglar, A. Shabbeer, N. Kurepina, B. Yener, K. P. Bennett. **Data-driven insights into deletions of *Mycobacterium tuberculosis* complex chromosomal DR region using spoligoforests.** *IEEE International Conference on Bioinformatics and Biomedicine (BIBM)*, Atlanta, November 2011.
- K. P. Bennett, C. Ozcaglar, J. Ranganathan, S. Raghavan, J. Katz, D. Croft, B. Yener, A. Shabbeer. **Visualization of tuberculosis patient and *Mycobacterium tuberculosis* complex genotype data via host-pathogen maps.** *IEEE BIBM Workshop on Computational Advances in Molecular Epidemiology*, Atlanta, November 2011.
- M. Aminian, A. Shabbeer, K. Hadley, C. Ozcaglar, S. Vandenberg, K. P. Bennett. **Knowledge-based Bayesian network for the classification of *Mycobacterium tuberculosis* complex sublineages** *ACM Conference on Bioinformatics, Computational Biology and Biomedicine (BCB)*, Chicago, August 2011.
- C. Ozcaglar, A. Shabbeer, S. Vandenberg, B. Yener, K. P. Bennett. **Examining the sublineage structure of *Mycobacterium tuberculosis* complex strains with multiple-biomarker tensors.** *IEEE International Conference on Bioinformatics and Biomedicine (BIBM)*, Hong Kong, December 2010.
- A. Shabbeer, C. Ozcaglar, M. Gonzalez, K. P. Bennett. **Optimal embedding of heterogeneous graph data with edge crossing constraints.** *NIPS Workshop on Challenges of Data Visualization*, Whistler, BC, Canada, December 2010.
- C. Ozcaglar, A. Shabbeer, S. Vandenberg, B. Yener, K. P. Bennett. **Multiple-biomarker tensor analysis for tuberculosis lineage identification.** *NIPS Workshop on Tensors, Kernels and Machine Learning*, Whistler, BC, Canada, December 2010.
- J. M. Pyle, F. S. Spear, S. Adali, B. K. Szymanski, S. Pearce, A. Waters, Z. Linder, C. Ozcaglar. **MetPetDB: The Unique Aspects of Metamorphic Geochemical Data and Their Influence on Data Model, User Interface and Collaborations.** *Geological Society of America Abstracts with Programs*, Vol. 39, No. 6, 2007.

Technical  
Reports

- C. Ozcaglar, B. Yener, K. P. Bennett. **Host-pathogen association analysis of tuberculosis patients via Unified Biclustering Framework.** Rensselaer Polytechnic Institute. TR-12-05, 2012.
- C. Ozcaglar, A. Shabbeer, S. Vandenberg, B. Yener, K. P. Bennett. **A clustering framework for *Mycobacterium tuberculosis* complex strains using multiple-biomarker tensors.** Rensselaer Polytechnic Institute. TR-10-08, 2010.
- F. S. Spear, J. M. Pyle, S. Adali, B. K. Szymanski, A. Waters, Z. Linder, C. Ozcaglar, S. O. Pearce. **MetPetDB: A database for metamorphic geochemistry.** Rensselaer Polytechnic Institute. TR-08-14, 2008.

Working  
Papers

- C. Ozcaglar, S. Salem. **Ensampling Methods.**

- C. Ozcaglar, S. Salem, B. Yener. **DIQB: Density-invariant quasi-biclustering**.
- Theses**
- **Algorithmic data fusion methods for tuberculosis**, *Ph.D. thesis*, Rensselaer Polytechnic Institute, 2012.
  - **Classification of email messages into topics using Latent Dirichlet Allocation**, *M.S. thesis*, Rensselaer Polytechnic Institute, 2008.
- Poster Presentations**
- C. Ozcaglar, B. Yener, K. P. Bennett. **UBF: Unified Biclustering Framework**. *New York Academy of Sciences (NYAS) 7th Annual Machine Learning Symposium*, NYC, October 2012.
  - C. Ozcaglar, A. Shabbeer, S. Vandenberg, B. Yener, K. P. Bennett. **A clustering framework for *Mycobacterium tuberculosis* complex strains using multiple-biomarker tensors**. *RPI-NSF Workshop on Multiscale Modeling of Complex Data*, Troy, NY, September 2011.
  - C. Ozcaglar, A. Shabbeer, S. Vandenberg, B. Yener, K. P. Bennett. **Insights into Camin-Sokal Parsimony and evolution of spoligotypes via spoligoforests**. *New York Academy of Sciences (NYAS) Symposium on Imaging, Visualization and Simulation: New Tools for Technology and Healthcare*, NYC, June 2011.
  - C. Ozcaglar, A. Shabbeer, S. Vandenberg, B. Yener, K. P. Bennett. **Multiple-biomarker tensor analysis for tuberculosis lineage identification**. *NIPS Workshop on Tensors, Kernels and Machine Learning*, Whistler, BC, Canada, December 2010.
  - C. Ozcaglar, B. Yener, A. Shabbeer, M. Aminian, K. P. Bennett. **A clustering framework for *Mycobacterium tuberculosis* complex strains using multiple-biomarker tensors**. *New York Academy of Sciences (NYAS) 5th Annual Machine Learning Symposium*, NYC, October 2010.
  - C. Ozcaglar, B. Yener, A. Shabbeer, M. Aminian, K. P. Bennett. **Examining sublineage structure of *Mycobacterium tuberculosis* complex strains with multiway modeling**. *Eigenvector University*, Seattle Washington, May 2010. (Best poster award)
- Talks**
- **MFMS: Maximal frequent module set mining from multiple human gene expression datasets**. *ACM SIGKDD International Workshop on Data Mining in Bioinformatics (BIOKDD)*, Chicago, August 2013.
  - **TCF: Tensor clustering framework on multiple-biomarker tensors**. *Bogazici University*, January 2012.
  - **TCF: Tensor clustering framework on multiple-biomarker tensors**. *Middle East Technical University*, January 2012.
  - **TCF: Tensor clustering framework on multiple-biomarker tensors**. *Bilkent University*, January 2012.
  - **Data-driven insights into deletions of *Mycobacterium tuberculosis* complex chromosomal DR region using spoligoforests**. *IEEE International Conference on Bioinformatics and Biomedicine (BIBM)*, Atlanta, November 12-15, 2011.
  - **Examining the sublineage structure of *Mycobacterium tuberculosis* complex strains with multiple-biomarker tensors**. *IEEE International Conference on Bioinformatics and Biomedicine (BIBM)*, Hong Kong, December 18-21, 2010.
  - **Extracting Associations from Activities**. *NSF Cyber-Enabled Discovery and Innovation Symposium*, Troy, NY, September 2007.
- Awards**
- Student Travel Award to attend IEEE BIBM 2010.
  - Best poster award, Eigenvector University 2010.
  - Full scholarship awarded by Rensselaer Polytechnic Institute for graduate study, 2006 - 2012.

*Cagri Ozcaglar*

- Full scholarship awarded by Bilkent University for undergraduate education, 2002 - 2006.
- Top 0.01% in Nationwide University Entrance Exam among 1.5 million candidates, 2002.
- Ranked 1st in the Mediterranean Region in 8th and 9th Turkish Mathematics Olympiad, 2000, 2001.
- Bronze medal in 3rd and 4th Turkish Secondary School Mathematics Olympiad, 1998, 1999.

**Activities** • **Reviewer for journals:** ACM Transactions on Algorithms, Computational Intelligence, Network Modeling Analysis in Health Informatics and Bioinformatics, Computational and Mathematical Methods in Medicine.

- **Member:** IEEE, ACM.

**Affiliations** • Treasurer of Turkish Student Association at RPI, 2011-2012.  
• Member of Building Planning Committee, Computer Science, RPI, 2011-2012.  
• RPI School of Science Graduate Council, Computer Science Representative, 2010-2011.  
• Member of Graduate Recruiting Committee, Computer Science, RPI, 2007-2008.

**Personal** • **Date of Birth:** 08/14/1984

**Information** • **Languages:** Turkish (native), Spanish

- **Visa status:** H1-B

**References** Available upon request.