Maintaining An Online Publication List

Tamara G. Kolda
Sandia National Labs

* Webpage Expert*

* Self-proclaimed
Importance of Online Publication List

- Take advantage of your rights to make your work accessible
  - Many journals allows the Author’s Final Version to be posted on preprint servers like arXiv *
  - Some journals allow the Published Version to be posted on the author’s home page *
- Make preprints available
  - Especially useful during job search
- Accuracy
  - Unusual names (O’Leary, van de Geijn)
  - Common names (Meza, Conroy)
- Two options
  - Use a tool like Google Scholar
  - Maintain your own list
  - Not mutually exclusive options

* SIAM allows authors these rights.
Interlude: Home Page Basics

- Critical Information
  - Actual clickable email
  - Work or mobile phone
  - Snail mail address
  - Brief bio
  - Publication list

- Optional Information
  - Picture
  - Skype, Twitter, Google+ ids
  - Software downloads
  - Awards
  - Service activities
  - Copies of talks

- Keep it up to date
  - Don’t make it too complicated
  - Update with each new paper!

- Avoid personal information

Mathematicians generally fail to provide short bios, requiring the visitor to piece together a profile by clicking many links.

Evrim Acar (Ataman)
Assistant Professor
Address:
Faculty of Life Sciences
University of Copenhagen
Rigshospitalet 30,
DK-1568 Frederiksberg C
E-mail: evrim.acaratan@gmail.com
evrim@life.ku.dk

Evrim Acar is an Assistant Professor at LIFE at University of Copenhagen, Denmark. Her research interests are data mining, machine learning, and mathematical modeling; in particular, tensor decompositions and their applications in social network analysis, computational neuroscience and chemometrics. Evrim received the Danish Council for Independent Research Sapere Aude Young Researcher Award in 2012.

Evrim received her MS and PhD in Computer Science from Rensselaer Polytechnic Institute (Troy, NY) in December, 2006 and May, 2008. She got her BS in Computer Engineering from Bogazici University (Istanbul, Turkey) in July, 2003. [See CV]

David F. Gleich
Assistant Professor
Computer Science
Purdue University

INTERESTS
Matrix-based network computations
PageRank
Spectral graph theory
Permutation models of physical systems
Large scale data computations

CLASSES
Spring 2013, CS520 - Computational Methods in Optimization
Fall 2012, CS561 - Matrix Computations
Spring 2012, Computational Methods in Optimization
Fall 2011, Network and Matrix Computations

Copyright © David F. Gleich 2011
This info was created with Asymptote static site generator
“HELPER” SITES
Google Scholar

- **Benefits**
  - Easy to set up
  - Automatically finds articles
  - Easy to correct
  - Counts citations (generously)
  - People can “follow” you
  - You can track citations to your work

- **Cons**
  - Doesn’t always link to the correct PDF
  - Generates poor BibTeX (no DOI)

- **How to get started**
  - Requires Google (aka Gmail) account

- **Highly Recommended**
  - Extremely low maintenance

---

July 11, 2013
Kolda - SIAM Annual
Microsoft Academic Search

- **Benefits**
  - Most of what Google Scholar offers
  - Includes DOI (unlike Google)
  - Co-author graph
  - Top keywords, conferences, journals

- **Cons**
  - Slow to find papers and preprints
  - Lots of mistakes
  - Fewer statistics (e.g., H-index)

- **How to Get Started**
  - academic.research.microsoft.com

- **No Recommendation**
  - Lags behind Google in finding papers
  - But really easy to set up...
ResearcherID (Thomson Reuters)

- **Benefits**
  - Import from Web of Science, etc.
  - Some useful metrics
  - BibTeX export

- **Cons**
  - No preprints!
  - Fewer cites than Google Scholar (partly due to mistakes)
  - Manual updating
  - No merge capability
  - Stupidly confused by upper/lowercase
  - Only includes researchers that have manually set up an account

- **How to Set Up**
  - [www.researcherid.com](http://www.researcherid.com)

- **Not Recommended**
  - Too hard to make corrections
ResearchGate

- **Pros**
  - Social network for publications
  - Allows user to upload PDF versions
  - Recommends papers and keeps you updated on publications of colleagues, etc.

- **Cons**
  - Insists that I work in Albuquerque!
  - No export to BibTeX!
  - Does not link to DOI!
  - Makes up its own “impact” score
  - “Spam” emails

- **Getting Started**
  - [www.researchgate.net](http://www.researchgate.net)

- **No Recommendation**
  - Might be useful for technical social networking
  - Can be overwhelming
Other Sites

- **ArXiv.org**
  - Highly recommended as preprint server
  - Many researchers follow certain topics
  - Quickly indexed by Google
  - Easy to post revisions
  - Can append final citation and comments

- **LinkedIn**
  - Manual input
  - Hard to maintain

- **Institutional Repositories**
  - Common in the UK
  - Work much like arXiv, but local
  - Benefit: Can post published version in some cases
MAINTAINING YOUR OWN LIST
Maintaining Your Own List

- **Pros**
  - Accuracy
  - Includes published PDF (if allowed)
  - Additional comments
    - Pointers to software
    - Notes about paper prizes, special issues, etc.
    - Corrections

- **Cons**
  - Requires effort

- **Systems**
  - Maintain publication database
  - Manually update web page and CV
JabRef BibTeX Reference Manager

- **Features**
  - Highly-customizable, advanced BibTeX editor
  - Java based, so works on Windows, Mac, Linux, etc.
  - Open source, active project, continually being improved
  - Import from arXiv, Google Scholar, IEEEXplore, etc.
  - Understands DOIs and hyperlinks
  - Downloads and manages PDFs and related files
  - Custom export to HTML (*)

- **Getting Started**
  - http://jabref.sourceforge.net/

- **Tammy’s Export Filter**
  - http://tinyurl.com/ljwnqz5

**Abstract:** Tensors have found application in a variety of fields, ranging from chemometrics to signal processing and beyond. In this paper, we consider the problem of multilinear modeling of sparse count data. Our goal is to develop a descriptive tensor factorization model of such data, along with appropriate algorithms and theory. To do so, we propose that the random variation is best described via a Poisson distribution, which better describes the zeros observed in the data as compared to the typical assumption of a Gaussian distribution. Under a Poisson assumption, we fit a model to observed data using the negative log-likelihood score. We present a new algorithm for Poisson tensor factorization called CANDECOMP-PARAFAC alternating Poisson regression (CP-APR) that is based on a majorization-minimization approach. It can be shown that CP-APR is a generalization of the Lee-Seung multiplicative updates. We show how to prevent the algorithm from converging to non-KKT points and prove convergence of CP-APR under mild conditions. We also explain how to implement CP-APR for large-scale sparse tensors and present results on several data sets, both real and simulated.

**Keywords:** nonnegative tensor factorization, nonnegative CANDECOMP-PARAFAC, Poisson tensor factorization, Lee-Seung multiplicative updates, majorization-minimization algorithms

Abstract: Tensors have found application in a variety of fields, ranging from chemometrics to signal processing and beyond. In this paper, we consider the problem of multilinear modeling of sparse count data. Our goal is to develop a descriptive tensor factorization model of such data, along with appropriate algorithms and theory. To do so, we propose that the random variation is best described via a Poisson distribution, which better describes the zeros observed in the assumption of a Gaussian distribution. Under a Poisson observed data using the negative log-likelihood score. We present a tensor factorization called CANDECOMP-PARAFAC alternating based on a majorization-minimization approach. It can be shown that CP-APR is a generalization of the Lee-Seung multiplicative updates. We show how to prevent the algorithm from converging to non-KKT points and prove convergence of CP-APR under additional conditions. We also provide a way to use CP-APR for large-scale problems and applications.

Keyword: factorization

BibTeX:
```
@article{ChKo12,
  author = {Eric C. Chi and Tamara G. Kolda},
  title = {On Tensors, Sparsity, and Nonnegative Factorizations},
  journal = {SIAM Journal on Matrix Analysis and Applications},
  month = {December},
  year = {2012},
  volume = {33},
  number = {4},
  pages = {1272-1299},
  doi = {10.1137/110859063}
}
```
@Inproceedings{SePiKo13a,
  Title = {Triadic Measures on Graphs: The Power of Wedge Sampling},
  Author = {C. Seshadhri and Ali Pinar and Tamara G. Kolda},
  Booktitle = {SDM13: Proceedings of the 2013 SIAM International Conference on Data Mining},
  Pages = {10--18},
  Year = {2013},
  Month = may,
  Abstract = {Graphs are used to model interactions...},
  Annote = {Winner of SDM13 Best Paper Prize!},
  Keywords = {triangle counting, directed triangle counting, clustering coefficient, Hoeffding's inequality},
  Oldversion = {Triangles-arXiv-1202.5230},
  Pdf = {SePiKo13a-SDM13.pdf},
  Url = {http://knowledgecenter.siam.org/338SDM/338SDM/1}
}

**Abstract:** Graphs are used to model interactions in a variety of contexts, and there is a growing need to quickly assess the structure of a graph. Some of the most useful graph metrics, especially those measuring social cohesion, are based on triangles. Despite the importance of these triadic measures, associated algorithms can be extremely expensive. We discuss the method of wedge sampling. This versatile technique allows for the fast and accurate approximation of all current variants of clustering coefficients and enables rapid uniform sampling of the triangles of a graph. Our methods come with provable and practical time-approximation tradeoffs for all computations. We provide extensive results that show our methods are orders of magnitude faster than the state-of-the-art, while providing nearly the accuracy of full enumeration. Our results will enable more wide-scale adoption of triadic measures for analysis of extremely large graphs, as demonstrated on several real-world examples.

**Keywords:** triangle counting, directed triangle counting, clustering coefficient, Hoeffding's inequality

*Winner of SDM13 Best Paper Prize!*
Where to Post My Publication List?

- **10 free minutes?**
  - Yes: Create a MS Academic Search profile
  - No: 3 free minutes?
    - Yes: Create a Google Scholar profile
    - No: Last homepage update?
      - Yes: Plead the 5th
      - No: Just last week
        - Yes: Create Your Own Publication Page
        - No: Want to post final version?
          - Yes: Love lots of emails?
            - Yes: Of course!
            - No: Looking for 1st job?
              - Yes: Post on arXiv
              - No: Join ResearchGate
          - No: Want to share your preprints with the world?
            - Yes: Post on arXiv
            - No: Join ResearchGate

Tammy Kolda
tgkolda@sandia.gov