

## Contact

Google - Santa Monica  
1333 Second Street  
Santa Monica, CA 90401  
<http://www.GuanYuqiang.net/>

Work Phone: 1 (310) 310-6105  
Work Fax: 1 (310) 310-6404  
Business Email: [yguan@google.com](mailto:yguan@google.com)  
Personal Email: [yuqiang\\_guan@yahoo.com](mailto:yuqiang_guan@yahoo.com)

## Work Experience

Google Santa Monica, CA  
Software Engineer Mar. 2008 - now  
Content/Search Ads relevance.

Microsoft Redmond, WA  
Software Engineer Jun. 2006 - Feb. 2008  
Search quality metrics, text/log mining, business/customer intelligence, question answering

Syncata El Segundo, CA  
Intern May 2005 - Aug. 2005  
Information retrieval and text mining

## Professional Activities

### Referee for

Computational Intelligence in Bioinformatics (book)  
IEEE Transactions on Pattern Analysis and Machine Intelligence  
IEEE Transactions on Neural Networks  
SIAM Journal on Matrix Analysis and Applications  
International Journal on Document Analysis and Recognition

### Program committee member for

IEEE Data Mining Conference 2006

### Reviewer for

IEEE Data Mining Conferences (ICDM 2003-05)  
SIAM Data Mining Conferences (SDM 2002-06)  
ACM KDD Conferences (KDD 2004-05)  
Workshop on Clustering High-Dimensional Data and its Applications (at SDM 2002-04 & ICDM 2003)  
Data Mining and Knowledge Discovery Journal  
ACM Computing Surveys  
IEEE Transactions on Knowledge and Data Engineering  
IEEE Transactions on Pattern Analysis and Machine Intelligence

## Education

Ph.D. Computer Science, The University of Texas at Austin, 2006  
B.S. Computer Science, University of Science and Technology of China, 1997

## Scientific Software

**Graclus**, A new efficient graph clustering software that does normalized cut, ratio association without eigenvector computation..

**Co-clustering**, It contains information-theoretic and minimum-residue co-clustering.

**Gmeans**, It contains spherical k-means, information-theoretic clustering, diametric clustering and Euclidean k-means algorithms with 6 different initializations and local search.

## Research Interest and Experience

Large-scale data mining, machine learning, pattern recognition, information retrieval, bioinformatics, scientific computing and graph theory. I have

Developed a fast, high-quality multilevel kernel-based graph clustering algorithm;

Obtained new theoretical connections between spectral clustering and weighted kernel k-means;

Investigated the minimum residue co-clustering algorithms on gene-expression data in bioinformatics;

Proposed a time and memory efficient technique for entire preprocessing and clustering large document collections;

Proposed a local search strategy to improve clustering results;

Studied profile minimization on triangulated triangles and integral computation using Quasi-Monte Carlo Methods.

## Papers

Ph.D. dissertation

- Large-Scale Clustering: Algorithms and Applications, *Yuqiang Guan*, May 3rd 2006.

Book chapters

- Clustering with Entropy-like k-means Algorithms, *M. Teboulle, P. Berkhin, I. Dhillon, Y. Guan and J. Kogan*, Book chapter in Grouping Multidimensional Data : Recent Advances in Clustering, J. Kogan, C. Nicholas, M. Teboulle(Eds.), pages 127-160, Springer, 2006.
- Efficient Clustering of Very Large Document Collections, *I. Dhillon, J. Fan and Y. Guan*, Invited book chapter in Data Mining for Scientific and Engineering Applications, R. L. Grossman, C. Kamath, P. Kegelmeyer, V. Kumar, R. R. Namburu (Eds.), pages 357-381, Kluwer, 2001.

Conference paper

- A Fast Kernel-based Multilevel Algorithm for Graph Clustering, *Dhillon, Y. Guan and B. Kulis*, Proceedings of The 11th ACM SIGKDD, Chicago, IL, Aug. 21 - 24, 2005.
- Kernel k-means, Spectral Clustering and Normalized Cuts, *I. Dhillon, Y. Guan and B. Kulis*, Proceedings of The 10th ACM SIGKDD, Seattle, WA, August 22-25, 2004.
- Minimum Sum-Squared Residue Co-clustering of Gene Expression Data, *H. Cho, I. Dhillon, Y. Guan and S. Sra*, Proceedings of The 4th SIAM Data Mining Conference, Lake Buena Vista, Florida, April 22-24, 2004.
- Information Theoretic Clustering of Sparse Co-Occurrence Data, *I. Dhillon and Y. Guan*, Proceedings of The 3rd IEEE International Conference on Data Mining, Melbourne, Florida, November 19 - 22, 2003.
- Iterative Clustering of High Dimensional Text Data Augmented by Local Search, *I. Dhillon, Y. Guan and J. Kogan*, Proceedings of The 2nd IEEE Data Mining Conference, Maebashi TERRSA, Maebashi City, Japan December 9 - 12, 2002.
- Resource Allocation for Clusters, *E. de Doncker, L. Cucos and Y. Guan*, Proceedings of the High Performance Computing Symposium, pp. 122-125, 2001.

- Distributed Quasi-Monte Carlo Methods in a Heterogeneous Environment, *E. de Doncker, R. Zanny, M. Ciobanu and Y. Guan*, Proceedings of the IPDPS Heterogeneous Computing Workshop 2000, pp. 200-206, 2000.
- Asynchronous Quasi-Monte Carlo Methods, *E. de Doncker, R. Zanny, M. Ciobanu and Y. Guan*, Proceedings of the High Performance Computing Symposium, pp. 130-135, 2000.
- On diameter D edge Deletion problems, *Y. Guan and K. Williams*, Presentation on Combinatorics, Graph Theory, Computing 30th Conference, Boca Raton FL, March 1998.

#### Journal paper

- Weighted Graph Cuts without Eigenvectors: A Multilevel Approach, *I. Dhillon, Y. Guan and B. Kulis*, IEEE Transactions on Pattern Analysis and Machine Intelligence (PAMI), Vol. 29, No. 11, pp. 1944-1957, 2007.
- Error Bounds for the Integration of Singular Functions using Equidistributed Sequences, *E. deDoncker and Y. Guan*, Journal of Complexity. 19/3 pp. 259-271, Elsevier Science, April 2003.
- Profile Minimization on Triangulated Triangles, *Y. Guan and K. Williams*, Discrete Mathematics, vol 260C, pp 69-76, Elsevier Science, Jan. 2003.

#### Technical reports

- A Unified View of Kernel k-means, Spectral Clustering and Graph Cuts, *I. Dhillon, Y. Guan and B. Kulis*, UTCS Technical Report #TR-04-25, 2004.
- Information Theoretic Clustering of Sparse Co-Occurrence Data, *I. Dhillon and Y. Guan*, UTCS Technical Report #TR-03-39, Sept., 2003. Also appeared in the 3rd SIAM International Conference on Data Mining (Workshop on Clustering High-Dimensional Data and its Applications), 2003.
- Refining Clusters in High-dimensional Text Data, *I. Dhillon, Y. Guan, and J. Kogan*, UTCS Technical Report #TR-02-03, Jan, 2002. Also appeared in the 2nd SIAM International Conference on Data Mining (Workshop on Clustering High-Dimensional Data and its Applications), April, 2002.

#### Teaching Experience

I have been a TA for graduate and undergraduate courses. My TA duties involve grading, holding office hours, leading discussion sessions, working out solutions for homework and exams, helping professors edit lecture notes, maintaining class web pages, etc.

Scientific Computing (Spring and Fall 2004)

System Modeling and Scientific Computing (Spring 2003)

Advanced Programming (Fall 2002)

Elements of Software Design (Spring 2002)

Large Scale Data Mining (Spring and Fall 2001)

Abstract Data Type (Spring and Fall 2000)

#### Honors and Awards

Excellent in Research Award recipient (2000)

College-wide "Outstanding Graduate Research Award" recipient (1999)

#### References

Upon request.

**Citizenship and Visa status**  
China and H1B