

---

**George Karypis**  
Associate Professor

Department of Computer Science & Engineering  
4-192 EE/CS  
200 Union Street SE  
Minneapolis, MN 55455

phone: (612) 626-7524  
fax: (612) 626-1597  
email: karypis@cs.umn.edu  
URL: <http://www.cs.umn.edu/~karypis>

---

**George Karypis's** research interests spans the areas of data mining, bioinformatics, cheminformatics, high performance computing, information retrieval, collaborative filtering, and scientific computing. His research has resulted in the development of software libraries for serial and parallel graph partitioning (METIS and ParMETIS), hypergraph partitioning (hMETIS), for parallel Cholesky factorization (PSPASES), for collaborative filtering-based recommendation algorithms (SUGGEST), clustering high dimensional datasets (CLUTO), finding frequent patterns in diverse datasets (PAFI), and for protein secondary structure prediction (YASSPP). He has coauthored over 150 journal and conference papers on these topics and a book title "Introduction to Parallel Computing" (Publ. Addison Wesley, 2003, 2<sup>nd</sup> edition). In addition, he is serving on the program committees of many conferences and workshops on these topics, and on the editorial boards of the International Journal of Data Mining and Bioinformatics, the journal on Current Proteomics, Advances in Bioinformatics, and Biomedicine and Biotechnology.

## PUBLICATIONS

### Books

1. "Introduction to Parallel Computing" (2<sup>nd</sup> edition). Ananth Grama, Anshul Gupta, George Karypis, and Vipin Kumar. Addison-Wesley, ISBN: 0-2016-4865-2, 2003.
2. "Introduction to Parallel Computing: Design and Analysis of Algorithms". Vipin Kumar, Ananth Grama, Anshul Gupta, and George Karypis. Benjamin/Cumming, ISBN: 0-8053-3170-0, 1994.

### Book Chapters (Invited)

1. "Scientific Data Analysis", Chandrika Kamath, Nikil Wale, George Karypis, Gaurav Pandey, Vipin Kumar, Krishna Rajan, Nagiza F. Samatova, Paul Bremyer, Guruprasad Kora, Chongle Pan, and Srikanth Yoginath. In "Scientific Data Management", Arie Shoshani and Doron Rotem (ed), CRC Press/Taylor and Francis Books, 2008 (in press).
2. "Towards a Scalable kNN CF Algorithm: Exploring Effective Applications of Clustering". Al. Mamunur Rashid, Shyong K. Lam, Adam LaPitz, George Karypis, and John Riedl. In "Web Mining and Web Usage Analysis", O. Nasraoui, M. Spiliopoulou, J. Srivastava, B. Mobasher, and B. Masand, Springer, 2007 (in press).
3. "Document Clustering". Ying Zhao and George Karypis. In "Encyclopedia of Machine Learning", Claude Sammut (ed), Springer, 2007 (in press).
4. "Protein Structure Prediction Using String Kernels". Huzefa Rangwala, Kevin DeRonne, and George Karypis. In "Knowledge Discovery in Bioinformatics: Techniques, Methods, and Applications", Y. Pan and T. Hu (eds). John Wiley and Sons, 2007.
5. "Data Mining Algorithms for Virtual Screening of Bioactive Compounds". Mukund Deshpande, Michihiro Kuramochi, and George Karypis. In "Data Mining in Biomedicine", P. Pardalos (eds). Springer-Verlag London Ltd, 2007.
6. "Finding Topological Frequent Patterns from Graphs Datasets". Michihiro Kuramochi and George Karypis. In "Mining Graph Data", L.B. Holder and D. Cook (eds). John Wiley & Sons, 2006.
7. "Criterion Functions for Clustering on High Dimensional Data". Ying Zhao and George Karypis. In "Grouping Multidimensional Data: Recent Advances in Clustering", Jacob Kogan Charles Nicholas, Marc Teboulle (eds). Springer-Verlag London Ltd, 2006.
8. "Partitioning and Load Balancing For Emerging Parallel Applications and Architectures", Karen Devine, Erik G. Boman, and George Karypis. In "Parallel Processing for Scientific Computing", M. Heroux. P. Raghavan, and H. D. Simon (eds) SIAM, 2006.
9. "Mining Scientific Datasets Using Graphs". Michihiro Kuramochi, Mukund Deshpande, and George Karypis. In "Data Mining: Next Generation Challenges and Future Directions", H. Kargupta, A. Joshi, K. Sivakumar, and Y. Yesha (eds.). AAAI Press, 2004.
10. "Mining Chemical Compounds". Mukund Deshpande, Michihiro Kuramochi, and George Karypis. In "Data Mining in Bioinformatics", J. Wang, M. Zaki, H. Toivonen, and D. Shasha (eds.). Springer-Verlag London Ltd, 2004.

11. “*Clustering in Life Sciences*”. Ying Zhao and George Karypis. In “*Functional Genomics: Methods and Protocols*”, M. Brownstein, A. Khodursky and D. Conniffe (editors). Humana Press, 2003.
12. “*Multilevel Hypergraph Partitioning*”. George Karypis. In “*Multilevel Optimization in VLSI CAD*”, J. Cong and J. R. Shinnerl (editors). Kluwer Academic Publishers, Boston, 2003.
13. “*Graph Partitioning For High Performance Scientific Simulations*”. Kirk Schloegel, George Karypis, and Vipin Kumar. In “*Sourcebook of Parallel Computing*”, J. Dongarra, I. Foster, G. Fox, K. Kennedy, A. White, L. Torczon, and W. Gropp (eds.). Morgan Kaufmann, 2002.
14. “*Parallel Data Mining Algorithms*”. Mahesh Joshi, Eui-Hong Han, George Karypis, and Vipin Kumar. In “*Sourcebook of Parallel Computing*”, J. Dongarra, I. Foster, G. Fox, K. Kennedy, A. White, L. Torczon, and W. Gropp (eds.). Morgan Kaufmann, 2002.
15. “*Data Mining for Turbulent Flows*”. Eui-Hong Han, George Karypis, and Vipin Kumar. In “*Data Mining for Scientific and Engineering Applications*”, C. Kamath, P. Kegelmeyer, V. Kumar, and R. Namburu (eds.). Kluwer Academic Publishers, 2001.
16. “*Parallel Association Rules*”. Mahesh Joshi, Eui-Hong Han, George Karypis, and Vipin Kumar. In “*Large-scale Parallel and Distributed Data Mining*”, M. Zaki, C. Ho (eds.). Lecture Notes in Computer Science/Lecture Notes in Artificial Intelligence (LNCS/LNAI), vol. 1759, 2000, Springer-Verlag
17. “*Scalable Parallel Algorithms for Sparse Linear Systems*”, Anshul Gupta, George Karypis, and Vipin Kumar. In “*Parallel Computing in Optimization*”, A. Migdalas, P. Pardalos, S. Story (eds.). Kluwer Academic Publishers; pp73—98, 1997.
18. “*Scalable Parallel Algorithms for Unstructured Problems*”. Vipin Kumar, Ananth Grama, Anshul Gupta, and George Karypis. A. Ferreira and J.D.P. Rolim (eds.), “*Parallel Algorithms for Irregular Problems: State of the Art*”; Kluwer Academic Publishers; pp. 99—113, 1995.

## Journal Papers

1. “*Learning Preferences of New Users in Recommender Systems: An Information Theoretic Approach*”. Al Mamunur Rashid, George Karypis, and John Riedl. SIGKDD Explorations, 2009 (in press).
2. “*Mining Transcriptome Data for Function-Trait Relationship of Hyper Productivity of Recombinant Antibody*”, Salim Charaniya, George Karypis, and Wei-Shou Hu. Biotechnology and Bioengineering, 2008 (in press).
3. “*Common Pharmacophore Identification Using Frequent Clique Detection Algorithm*”, Yevgeniy Podolyan and George Karypis. Journal of Chemical Information and Modeling, 2008 (in press).
4. “*Mining Bioprocess Data: Challenges and Opportunities*”, Salim Charaniya, Wei-Shou Hu, and George Karypis. Trends in Biotechnology, 26(12), 690—699, 2008.
5. “*CONTOUR: An Efficient Algorithm for Discovering Discriminating Subsequences*”, Jianyong Wang, Yuzhou Zhang, Lizhu Zhou, George Karypis, and Charu C. Aggarwal. Data Mining and Knowledge Discovery, 2008 (in press).
6. “*Comparison of Descriptor Spaces for Chemical Compound Retrieval and Classification*”, Nikil Wale, Ian A. Watson, and George Karypis. Knowledge and Information Systems (KAIS) Journal, Vol 14, No. 3, pp. 347—375, 2008.
7. “*Indirect Similarity based Methods for Effective Scaffold-Hopping in Chemical Compounds*”, Nikil Wale, Ian Watson, and George Karypis. Journal of Chemical Information and Modeling, 48 (4), 730—741, 2008.
8. “*fRMSDPred: Predicting local RMSD between structural fragments using sequence information*”, Huzefa Rangwala and George Karypis. PROTEINS: Structure, Function, and Bioinformatics, 72(3), 1005—1018, 2008.
9. “*Conserved GU-rich elements mediate mRNA decay by binding to CUG-binding protein 1*”, I.A. Vlasova, N.M. Tahoe, D. Fan, O. Larsson, B. Rattenbacher, J.R. SternJohn, J. Vasdewani, G. Karypis, C.S. Reilly, P. Bitterman, and P.R. Bohjanen. Molecular Cell, 29, 263—270, 2008.
10. “*Transcriptome Dynamics Based Operon Prediction and Verification in Streptomyces Coelicolor*”, Salim Charaniya, Sarika Mehra, Wei Lian, Karthik Jayapal, George Karypis, and Wei-Shou Hu. Nucleic Acids Research, Vol. 35, No. 21, pp. 7222—7236, June 2007.
11. “*Clustering methodologies for identifying country core competencies*”. R.N. Kostoff, J.A. del Rio, H. D. Cortes, C. Smith, A. Smith, C. Wagner, L. Leydesdorff, G. Karypis, G. Malpohl, and R. Tshiteya. Journal of Information Science, Vol. 33, No. 1, pp. 21—40, 2007.
12. “*Out-of-Core Coherent Closed Quasi-Clique Mining from Large Dense Graph Databases*”, Zhiping Zeng, Jianyong Wang, Lizhu Zhou, and George Karypis. ACM Transactions on Database Systems, Vol 32, Issue 2, 2007.

13. “*Effective Optimization Algorithms for Fragment-Assembly Based Protein Structure Prediction*”. Kevin DeRonne and George Karypis. *Journal of Bioinformatics and Computational Biology*, Vol. 5, No. 2, pp. 335—352, 2007.
14. “*Discovering Frequent Geometric Subgraphs*”. Michihiro Kuramochi and George Karypis. *Information Systems Journal*, 32, pp. 1101-1120, 2007.
15. “*QCRNA 1.0: A database of quantum calculations for RNA catalysis*”, T.J. Giese, B.A. Gregersen, Y. Liu, K. Nam, E. Mayaan, A. Moser, K. Range, O.N. Faza, C.S. Lopez, A.R. de Lera, G. Schaftenaar, X. Lopze, T.S. Lee, G. Karypis, D.M. York. *Journal of Molecular Graphics & Modeling*, Vol. 25, No. 4, pp. 423—433, 2006.
16. “*Building Multiclass Classifiers for Remote Homology Detection and Fold Recognition*”. Huzefa Rangwala and George Karypis. *BMC Bioinformatics*, 7:455, 2006.
17. “*On Mining Instance-Centric Classification Rules*”. Jianyong Wang and George Karypis. *IEEE Transactions on Knowledge and Data Engineering*, 18(11), pp. 1497—1511, 2006.
18. “*Genomic view of systemic autoimmunity in MRL/lpr mice*”. J. Liu, G. Karypis, KL. Hippen, AL. Vegoe, P. Ruiz, GS. Gilkeson, and TW. Behrens. *Genes and Immunity*, 7, pp. 156—168, 2006.
19. “*YASSPP: Better Kernels and Coding Schemes Lead to Improvements in Protein Secondary Structure Prediction*”. George Karypis. *PROTEINS: Structure, Function, and Bioinformatics*, 64(3), pp. 575—586, 2006.
20. “*On Efficiently Summarizing Transactions for Clustering*”. Jianyong Wang and George Karypis. *Knowledge and Information Systems*, Vol. 9, No. 1, pp. 19—37, 2006.
21. “*Multi-Objective Hypergraph Partitioning Algorithms for Cut and Maximum Subdomain Degree Minimization*”. Navaratnasothie Selvakkumaran, and George Karypis. *IEEE Transactions on CAD*, 25(3), pp. 504—517, 2006.
22. “*Prediction of Contact Maps Using Support Vector Machines*”. Ying Zhao and George Karypis. *International Journal on Artificial Intelligence Tools*. Vol. 14, Issue 5, pp.849—865, 2005.
23. “*Profile Based Direct Kernels for Remote Homology Detection and Fold Recognition*”. Huzefa Rangwala and George Karypis. *Bioinformatics*, Vol. 31, No. 23, pp. 4239—4247, 2005.
24. “*Finding Frequent Patterns in a Large Sparse Graph*”. Michihiro Kuramochi and George Karypis. *Data Mining and Knowledge Discovery*, Vol. 11, No. 3, 243—271, 2005.
25. “*Frequent Substructure Based Approaches for Classifying Chemical Compounds*”. Mukund Deshpande, Michihiro Kuramochi, Nikhil Wale, and George Karypis. *IEEE Transactions on Knowledge and Data Engineering*, Vol. 17, No. 8, 1036—1050, 2005.
26. “*Gene Classification Using Expression Profiles: A Feasibility Study*”. Michihiro Kuramochi and George Karypis. *International Journal on Artificial Intelligence Tools*. Vol. 14, No. 4, pp. 641—660, 2005.
27. “*Data Clustering in Life Sciences*”. Ying Zhao and George Karypis. *Molecular Biotechnology*, 31(1), pp. 55—80, 2005.
28. “*Hierarchical Clustering Algorithms for Document Datasets*”. Ying Zhao and George Karypis. *Data Mining and Knowledge Discovery*, Vol. 10, No. 3, pp. 141—168, 2005.
29. “*Finding Frequent Patterns Using Length-Decreasing Support Constraints*”. Masakazu Seno and George Karypis. *Data Mining and Knowledge Discovery*, Vol. 10, No. 3, pp. 197—228, 2005.
30. “*Expression Levels for Many Genes in Human Peripheral Blood Cells are Highly Sensitive to ex Vivo Incubation*”. E. Baechler, F. Batliwalla, G. Karypis, P. Gaffney, K. Moser, W. Ortmann, K. Espe, S. Balasubramanian, K. Hughes, J. Chan, A. Begovich, S. Chang, P. Gregersen, and T. Behrens. *Genes and Immunity*, Vol. 5, No. 5, pp. 347—353, 2004.
31. “*A Boolean Algorithm for Reconstructing the Structure of Regulatory Networks*”. Sarika Mehra, Wei-Shou Hu, and George Karypis. *Metabolic Engineering*, Vol. 6, No. 4, pp. 326—339, 2004.
32. “*Macromolecule Mass Spectrometry: Citation Mining of User Documents*”. Ronald Kostoff, Clifford Bedford, Antonio del Rio, Hector Cortes, and George Karypis. *Journal of the American Society for Mass Spectrometry*, Vol. 15, No. 3, pp. 281—287, March 2004.
33. “*Parallel Formulations of Tree-Projection-Based Sequence Mining Algorithm*”. Valerie Guralnik and George Karypis. *Parallel Computing*, Vol. 30, pp. 443—472, 2004.
34. “*wCLUTO: A Web-Enabled Clustering Toolkit*”. Matthew Rasmussen, Mukund Deshpande, George Karypis, Jim Johnson, John Crow, and Ernest Retzel. *Plant Physiology*, October 2003, Vol. 133, pp. 510—516.
35. “*Item-based Top-N Recommendation Algorithms*”. Mukund Deshpande and George Karypis. *ACM Transactions on Information Systems*. Volume 22, Issue 1, pp. 143—177, January 2004.
36. “*An Efficient Algorithm for Discovering Frequent Subgraphs*”. Michihiro Kuramochi and George Karypis. *IEEE Transactions on Knowledge and Data Engineering*, Vol. 16, No. 9, 1038—1051, 2004.
37. “*Empirical and Theoretical Comparisons of Selected Criterion Functions for Document Clustering*”. Ying Zhao and George Karypis. *Machine Learning*, 55, pp. 311-331, 2004.

38. “*Interferon-Inducible Gene Expression Signature in Peripheral Blood Cells of Patients with Severe SLE*”. E. Baechler, F. Batliwalla, G. Karypis, P. Gaffney, W. Ortmann, K. Espe, K. Shark, W. Grande, K. Hughes, V. Kapur, P. Gregersen, and T. Behrens. *Proceedings of the National Academies of Sciences (PNAS)*, Vol. 100, No. 5, pp. 2610—2615, March 2003.
39. “*Selective Markov Models for Predicting Web-Page Accesses*”. Mukund Deshpande and George Karypis. *ACM Transactions on Internet Technology*, Vol. 4, Issue 2, pp. 163—184, May 2004.
40. “*Predicting the Performance of Randomized Parallel Search: An Application to Robot Motion Planning*”. Daniel Challou, Maria Gini, Vipin Kumar, and George Karypis. *Journal of Intelligent and Robotic Systems*. Volume 38, Number 1, pp. 31—53, September 2003.
41. “*Parallel Static and Dynamic Multi-Constraint Graph Partitioning*”. Kirk Schloegel, George Karypis, and Vipin Kumar. *Concurrency and Computation: Practice and Experience*. Volume 14, Issue 3, pages 219—240, 2002.
42. “*Privacy Risks in Recommender Systems*”. Naren Ramakrishnan, Benjamin J. Keller, Batul J. Mirza, Ananth Y. Grama, and George Karypis. *IEEE Internet Computing*, 54—62, Vol 5, No. 6, 2001.
43. “*Wavefront Diffusion and LMSR: Algorithms for Dynamic Repartitioning of Adaptive Meshes*”. Kirk Schloegel, George Karypis, and Vipin Kumar. *IEEE Transactions on Parallel and Distributed Systems*. Vol. 12, No. 5, 451—466, May 2001.
44. “*Multilevel k-way Hypergraph Partitioning*”. George Karypis and Vipin Kumar. *VLSI Design*, Vol. 11, No. 3, pp. 285—300, 2000.
45. “*Scalable Parallel Data Mining for Association Rules*”. Sam Han, George Karypis, and Vipin Kumar. *IEEE Transactions on Knowledge and Data Engineering*, Vol. 12, No. 3, pp337—352, 2000.
46. “*Chameleon: A hierarchical Clustering Algorithms Using Dynamic Modeling*”. George Karypis, Eui-Hong Han, and Vipin Kumar. *IEEE Computer*, Special Issue on Data Analysis and Mining. Vol. 32, No. 8, pp68—75, August 1999.
47. “*Parallel Multilevel k-way Partition Scheme for Irregular Graphs*”. George Karypis and Vipin Kumar. *SIAM Review*, Vol. 41, No. 2, pp. 278—300, 1999.
48. “*Document Categorization and Query Generation on the World Wide Web Using WebACE*”. D. Boley, M. Gini, R. Gross, E. Han, K. Hastings, G. Karypis, V. Kumar, B. Mobasher, and J. Moore. *AI Review*, Vol. 11, pp 365—391, December 1999.
49. “*Partitioning-Based Clustering for Web document Categorization*”. Daniel Boley, Maria Gini, Robert Gross, Eui-Hong Han, Kyle Hastings, George Karypis, Vipin Kumar, Bamshad Mobasher, and Jerome Moore. *Decision support Systems*, Vol. 27, No. 3, pp. 329—341, 1999.
50. “*Multilevel Hypergraph Partitioning: Application in VLSI Domain*”. George Karypis, Rajat Aggarwal, Vipin Kumar, and Shashi Shekhar. *IEEE Transactions on VLSI Systems*, Vol. 7, No. 1, pp. 69—79, 1999.
51. “*A Fast and Highly Quality Multilevel Scheme for Partitioning Irregular Graphs*”. George Karypis and Vipin Kumar. *SIAM Journal on Scientific Computing*, Vol. 20, No. 1, pp. 359—392, 1999.
52. “*A Parallel Algorithm for Multilevel Graph Partitioning and Sparse Matrix Ordering*”. George Karypis and Vipin Kumar. *Journal of Parallel and Distributed Computing*, Vol. 48, pp. 71—85, 1998.
53. “*Multilevel k-way Partitioning Scheme for Irregular Graphs*”. George Karypis and Vipin Kumar. *Journal of Parallel and Distributed Computing*, Vol. 48, pp. 96—129, 1998.
54. “*Multilevel Diffusion Schemes for Repartitioning of Adaptive Meshes*”. Kirk Schloegel, George Karypis, and Vipin Kumar. *Journal of Parallel and Distributed Computing*, Vol. 47, pp. 109—124, 1997.
55. “*Highly Scalable Parallel Algorithms for Sparse Matrix Factorization*”. Anshul Gupta, George Karypis, and Vipin Kumar. *IEEE Transactions on Parallel and Distributed Systems*, Vol. 8, No. 5, pp. 502—520, 1997.
56. “*Unstructured Tree Search on SIMD Parallel Computers*”. George Karypis and Vipin Kumar. *IEEE Transactions on Parallel and Distributed Systems*, Vol. 5, No. 10, pp. 1057—1072, October 1994.

### Conference Papers

1. “*The Set Classification Problem and Solution Methods*”. Xia Ning and George Karypis. *SIAM International Conference on Data Mining*, 2009.
2. “*A Kernel Framework for Protein Residue Annotation*”. Huzefa Rangwala, Chris Kauffman, and George Karypis. *Proceedings of the 13<sup>th</sup> Pacific-Asia Conference on Knowledge Discovery (PAKDD)*, 2009 (in press).
3. “*Genome Alignments using MPI-LAGAN*”. Ruinan Zhang, Huzefa Rangwala, and George Karypis. *IEEE International Conference on Bioinformatics and Biomedicine*, 2008 (in press).
4. “*TOPTMH: Topology Predictor for Transmembrane Alpha Helices*”. Rezwana Ahmed, Huzefa Rangwala, and George Karypis. *ECML PKDD*, pp. 23—38, 2008.

5. *“Improving Homology Models for Protein-Ligand Binding Sites”*. Chris Kauffman, Huzefa Rangwala, and George Karypis. LSS Computational systems Bioinformatics Conference (CSB), 2008.
6. *“Architecture Aware Partitioning Algorithms”*. Irene Moulitsas and George Karypis. 8<sup>th</sup> Intl. Conference on Algorithms and Architectures for Parallel Processing (ICA3PP), pp. 42—52, 2008.
7. *“An Analysis of Information Content Present in Protein-DNA Interactions”*. Chris Kauffman and George Karypis. Pacific Symposium on Biocomputing, pp. 477—488, 2008.
8. *“fRMSDAlign: Protein Sequence Alignment Using Predicted Local Structure Information”*. Huzefa Rangwala and George Karypis. 6<sup>th</sup> Asia Pacific Bioinformatics Conference, pp. 111—122, 2008.
9. *“Interleaving of Gate Sizing and Constructive Placement for Predictable Performance”*. Sungjae Kim, Eugene Shragowitz, George Karypis, and Rung-Bin Lin. International Symposium on VLSI Design, Automation, and Test, pp. 1—4, 2007.
10. *“Methods for Effective Virtual Screening and Scaffold-Hopping in Chemical Compounds”*. Nikil Wale, Ian A. Watson, and George Karypis. LSS Computational Systems Bioinformatics Conference (CSB), pp. 403—416, 2007.
11. *“fRMSDPred: Predicting Local RMSD Between Structural Fragments using Sequence Information”*. Huzefa Rangwala and George Karypis. LSS Computational Systems Bioinformatics Conference (CSB), pp. 311—322, 2007.
12. *“Discriminating Subsequence Discovery for Sequence Clustering”*. Jianyong Wang, Yuzhou Zhang, Lizhu Zhou, George Karypis, and Charu C. Aggarwal. SIAM International Conference on Knowledge and Data Discovery, 2007.
13. *“Comparison of Descriptor Spaces for Chemical Compound Retrieval and Classification”*. Nikil Wale and George Karypis. IEEE International Conference on Data Mining (ICDM), pp. 678—689, 2006.
14. *“Incremental Window-based Protein Sequence Alignment Algorithms”*. Huzefa Rangwala and George Karypis. Bioinformatics Special Issue on the 5<sup>th</sup> European Conference on Computational Biology (ECCB), Vol 15, No. 4, e17—23, 2007.
15. *“Coherent Closed Quasi-Clique Discovery from Large Dense Graph Databases”*. Zhiping Zeng, Jianyong Wang, Lizhu Zhou, and George Karypis. 12<sup>th</sup> ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, pp. 797—802, 2006.
16. *“Effective Optimization Algorithms for Fragment-Assembly Based Protein Structure Prediction”*. Kevin W. DeRonne and George Karypis. LSS Computational Systems Bioinformatics Conference (CSB2006), 2006.
17. *“Multilevel Algorithms for Partitioning Power-Law Graphs”*. Amine Abou-Rjeili and George Karypis. IEEE International Parallel & Distributed Processing Symposium (IPDPS), 2006.
18. *“Feature-based Recommendation System”*. Eui-Hong Han and George Karypis. Proceedings of the 14<sup>th</sup> Conference of Information and Knowledge Management (CIKM), pp. 446—452, 2005.
19. *“Effective Document Clustering for Large Heterogeneous Law Firm Collections”*. Jack G Conrad, Khalid Al-Kofahi, Ying Zhao, and George Karypis. 10<sup>th</sup> International Conference on Artificial Intelligence and Law (ICAIL), pp. 177—187, 2005.
20. *“HARMONY: Efficiently Mining the Best Rules for Classification”*. Jianyong Wang and George Karypis. Proceedings of the 2005 SIAM International conference on Data Mining, pp. 205—216, 2005.
21. *“Topic-Driven Clustering for Document Datasets”*. Ying Zhao and George Karypis. Proceedings of the 2005 SIAM International conference on Data Mining, pp. 358—369, 2005.
22. *“Influence in Ratings-Based Recommender Systems: An Algorithm-Independent Approach”*. Al Mamunur Rashid, George Karypis, and John Riedl. Proceedings of the 2005 SIAM International conference on Data Mining. 2005.
23. *“Soft Clustering Criterion Functions for Partitional Document Clustering”*. Ying Zhao and George Karypis. Proceedings of the 13<sup>th</sup> Conference of Information and Knowledge Management (CIKM), pp. 246—247, 2004.
24. *“SUMMARY: Efficiently Summarizing Transactions for Clustering”*. Jianyong Wang and George Karypis. Proceedings of the 4<sup>th</sup> IEEE Conference on Data Mining (ICDM), 2004.
25. *“GREW—A Scalable Frequent Subgraph Discovery Algorithm”*. Michihiro Kuramochi and George. Proceedings of the 4<sup>th</sup> IEEE Conference on Data Mining (ICDM), pp.439—442, 2004.
26. *“Efficient Closed Pattern Mining in the Presence of Tough Block Constraints”*. Krishna Gade, Jianyong Wang, and George Karypis. Proceedings of the 10<sup>th</sup> ACM SIGKDD Conference on Knowledge Discovery and Data Mining (KDD), pp. 138—147, 2004.
27. *“Multi-Resource Aware Partitioning Algorithms for FPGAs with Heterogeneous Resources”*. Navaratnasothie Selvakkumaran, Abishek Ranjan, Salil Rajee, George Karypis. Proceedings of the 41<sup>st</sup> Design and Automation Conference (DAC), pp. 741—746, 2004.

28. “*Finding Frequent Patterns in a Large Sparse Graph*”. Michihiro Kuramochi and George Karypis. Proceedings of the 2004 SIAM International Conference on Data Mining, 2004.
29. “*BAMBOO: Accelerating Closed Itemset mining by Deeply Pushing the Length Decreasing Support Constraint*”. Jianyong Wang and George Karypis. Proceedings of the 2004 SIAM International Conference on Data Mining, 2004.
30. “*Frequent Sub-Structure-Based Approaches for Classifying Chemical Compounds*”. Mukund Deshpande, Michihiro Kuramochi, and George Karypis. Proceedings of the 3<sup>rd</sup> IEEE Conference on Data Mining (ICDM), 2003.
31. “*Intelligent Meta-Search Engine for Knowledge Management*”. Eui-Hong Han, George Karypis, and Doug Mewhort. Proceedings of the 12<sup>th</sup> Conference of Information and Knowledge Management (CIKM), pp. 492—495, 2003.
32. “*Multi-Objective Hypergraph Partitioning Algorithms for Cut and Maximum Subdomain Degree Minimization*”. Navaratnasothie Selvakumaran and George Karypis. IEEE/ACM International Conference on Computer Aided Design (ICCAD), 2003.
33. “*Multi-Constraint Mesh Partitioning for Contact/Impact Computations*”. George Karypis. Proceedings of the 2003 ACM/IEEE Conference on Supercomputing 2003.
34. “*Prediction of Contact Maps Using Support Vector Machines*”. Ying Zhao and George Karypis. Proceedings of the 3<sup>rd</sup> IEEE International Conference on Bioinformatics and Bioengineering (BIBE), pp. 26—33, 2003.
35. “*Discovering Frequent Geometric Subgraphs*”. Michihiro Kuramochi and George Karypis. Proceedings of the 2<sup>nd</sup> IEEE Conference on Data Mining (ICDM), pp. 258—265, 2002.
36. “*SLPminer: An Algorithm for Finding Frequent Sequential Patterns Using a Length-Decreasing Support Constraint*”. Masakazu Seno and George Karypis. Proceedings of the 2<sup>nd</sup> IEEE Conference on Data Mining (ICDM), pp. 418-425, 2002.
37. “*A Polynomial Time Approximation Scheme for Rectilinear Steiner Minimum Tree Construction in the Presence of Obstacles*”, Jian. Liu, Ying. Zhao, Eugene Shragowitz, and George Karypis. In 9<sup>th</sup> International Conference on Electronics, Circuits and Systems, pp. 781—784, 2002.
38. “*Evaluation of Hierarchical Clustering Algorithms for Document Datasets*”. Ying Zhao and George Karypis. Proceedings of the 11<sup>th</sup> Conference of Information and Knowledge Management (CIKM), pp. 515-524, 2002.
39. “*Using Conjunction of Attribute Values for Classification*”. Mukund Deshpande and George Karypis. Proceedings of the 11<sup>th</sup> Conference of Information and Knowledge Management (CIKM), pp. 356—364, 2002.
40. “*Multi-objective Circuit Partitioning for Cutsizes and Path-Based Delay Minimization*”. Cristinel Ababei, Navaratnasothie Selvakumaran, Kia Bazargan, and George Karypis. IEEE/ACM International Conference on Computer Aided Design (ICCAD), pp. 181—185, 2002.
41. “*Evaluation of Techniques for Classifying Biological Sequences*”. Mukund Deshpande and George Karypis, Proceedings of the 6<sup>th</sup> Pacific-Asia Conference on Knowledge Discovery (PAKDD), 2002.
42. “*Expert Agreement and Content Based Reranking in a Meta Search Environment using Mearf*”. Uygur Oztekin, George Karypis, and Vipin Kumar. Proceedings of the 11<sup>th</sup> WWW Conference, pp. 333—344, 2002.
43. “*Incremental SVD-Based Algorithms for Highly Scalable Recommender Systems*”. Badrul Sarwar, George Karypis, Joe Konstan, and John Riedl. Proceedings of the 5<sup>th</sup> International Conference on Computer and Information Technology (ICCIT), 2002.
44. “*Recommender Systems for Large-Scale E-Commerce: Scalable Neighborhood Formation Using Clustering*”. Badrul Sarwar, George Karypis, Joe Konstan, and John Riedl. Proceedings of the 5<sup>th</sup> International Conference on Computer and Information Technology (ICCIT), 2002.
45. “*Improve Precategorized Collection Retrieval by Using Supervised Term Weighting Schemes*”. Ying Zhao and George Karypis, International Conference on Information Technology Coding and Computing, pp. 16—21, April 2002.
46. “*Gene Classification Using Expression Profiles: A Feasibility Study*”. Michihiro Kuramochi and George Karypis. Proceedings of the 2<sup>nd</sup> IEEE International Conference on Bioinformatics and Bioengineering (BIBE), pp. 191-200, 2001.
47. “*Evaluation of Item-based Top-N Recommendation Algorithms*”. George Karypis, Proceedings of the 10<sup>th</sup> Conference of Information and Knowledge Management (CIKM), pp. 247—254, 2001.
48. “*Graph Partitioning for Dynamic, Adaptive and Multi-phase Scientific Simulations*”, Kirk Schloegel, George Karypis, and Vipin Kumar. IEEE International Conference on Cluster Computing, pp. 271—273, 2001.
49. “*A Scalable Algorithm for Clustering Sequential Data*”. Valerie Guralnik and George Karypis. Proceedings of the 1<sup>st</sup> IEEE Conference on Data Mining, pp. 179—186, 2001.

50. “*LPMiner: An Algorithm for Finding Frequent Itemsets Using Length Decreasing Support Constraints*”. Masakazu Seno and George Karypis. Proceedings of the 1<sup>st</sup> IEEE Conference on Data Mining, pp. 505-512, 2001.
51. “*Frequent Subgraph Discovery*”. Michihiro Kuramochi and George Karypis. Proceedings of the 1<sup>st</sup> IEEE Conference on Data Mining, pp. 313-320, 2001.
52. “*Multilevel Algorithms for Generating Coarse Grids in Multigrid Methods*”. Irene Moulitsas and George Karypis. Proceedings on Supercomputing 2001.
53. “*Parallel Algorithms for Sequence Mining*”. Valerie Guralnik, Nivea Garg, and George Karypis. Proceedings of Europar, pp. 310—320, 2001.
54. “*Selective Markov Models*”. Mukund Deshpande and George Karypis. SIAM Conference on Data Mining, 2001.
55. “*Item-Based Collaborative Filtering Recommendation Algorithms*”. Badrul Sarwar, George Karypis, Joseph Konstan, and John Riedl. WWW10, pp. 285—295, 2001.
56. “*Text Categorization Using Weight adjusted k-Nearest Neighbor Classification*”. Eui-Hong Han, George Karypis, and Vipin Kumar. Proceedings of the 5<sup>th</sup> Pacific-Asia Conference on Knowledge Discovery and Data Mining (PAKDD), pp. 53—65, 2001.
57. “*Analysis of Recommendation Algorithms for E-Commerce*”. Badrul Sarwar, George Karypis, Joseph Konstan, and John Riedl. Proceedings of the 2<sup>nd</sup> ACM Conference on Electronic Commerce, pp. 158—167, 2000.
58. “*Fast Dimensionality Reduction Algorithm with Applications to Document Retrieval & Categorization*”. George Karypis and Eui-Hong Han. Proceedings of the 9<sup>th</sup> International Conference on Information and Knowledge Management, pp. 12—19, 2000.
59. “*A Unified Algorithm for Load-balancing Adaptive Scientific Simulations*”. Kirk Schloegel, George Karypis, and Vipin Kumar. Proceedings of the 2000 ACM/IEEE Conference on Supercomputing, 2000.
60. “*Centroid-Based Document Classification: Analysis & Experimental Results*”. Eui-Hong Han and George Karypis. Proceedings of the 4<sup>th</sup> European Conference on Principles and Practice of Knowledge Discovery in Databases (PKDD), pp. 424—431, 2000.
61. “*Memory Management Techniques for Gang Scheduling*”. William Leinberger, George Karypis, and Vipin Kumar. Europar 2000.
62. “*Parallel Multilevel Algorithms for Multi-Constraint Graph Partitioning*”. Kirk Schloegel, George Karypis, and Vipin Kumar. Europar , pp. 296—310, 2000, “*Distinguished Paper*” award.
63. “*Job Scheduling in the Presence of Multiple Resource Requirements*”. William Leinberger, George Karypis, and Vipin Kumar. Proceedings of the 1999 ACM/IEE Conference on Supercomputing, 1999.
64. “*Multi-Capacity Bin Packing Algorithms with Applications to Job Scheduling under Multiple Constraints*”. William Leinberger, George Karypis, and Vipin Kumar. Proceedings of the International Conference on Parallel Processing, pp. 404—412, 1999.
65. “*A New Algorithm for Multi-objective Graph Partitioning*”. Kirk Schloegel, George Karypis, and Vipin Kumar. Proceedings of Europar, pp. 322-331, 1999.
66. “*Multilevel k-way Hypergraph Partitioning*”. George Karypis and Vipin Kumar. Proceedings of the 36<sup>th</sup> Design Automation Conference, pp. 343—348, 1999.
67. “*PSPASES: An Efficient and Scalable Parallel Direct Solver*”. Mahesh V. Joshi, George Karypis, Vipin Kumar, Anshul Gupta, and Fred Gustavson. Proceedings of 9<sup>th</sup> SIAM Conference on Parallel Processing and Scientific Computing, 1999.
68. “*Dynamic Repartitioning of Adaptively Refined Meshes*”. Kirk Schloegel, George Karypis, and Vipin Kumar. Proceedings of 9<sup>th</sup> SIAM Conference on Parallel Processing and Scientific Computing, 1999.
69. “*Multilevel Algorithms for Multi-Constraint Graph Partitioning*”. George Karypis and Vipin Kumar. Proceedings of 10<sup>th</sup> Supercomputing Conference, pp. 1—13, 1998.
70. “*Dynamic Repartitioning of Adaptively Refined Meshes*”. Kirk Schloegel, George Karypis, and Vipin Kumar. Proceedings of 10<sup>th</sup> Supercomputing Conference, pp. 1—8, 1998.
71. “*A Performance Study of Diffusive vs. Remapped Load-Balancing Schemes*”. Kirk Schloegel, George Karypis, Vipin Kumar, Rupak Biswas, and Leonid Oliker. Proceedings of the 11<sup>th</sup> Intl. Conference on Parallel and Distributed Computing Systems, 1998.
72. “*ScalParC: A new Efficient and Scalable Parallel Classification Algorithm for Mining Large Datasets*”. Mahesh Joshi, George Karypis, and Vipin Kumar. Proceedings of the 12<sup>th</sup> Intl. Parallel Processing Symposium, pp. 573—579, 1998.

73. "A High Performance Two Dimensional Scalable Parallel Algorithm for Solving Sparse Triangular System". Mahesh Joshi, Anshul Gupta, George Karypis, and Vipin Kumar. Proceedings of the 4<sup>th</sup> Intl. Conference on High Performance Computing, pp. 137—143, 1997.
74. "Scalable Parallel Data Mining for Association Rules". Eui-Hong Han, George Karypis, and Vipin Kumar. Proceedings of the 1997 ACM-SIGMOD Intl. Conference on Management of Data, pp. 277—288, 1997.
75. "Parallel Threshold-based ILU Factorization". George Karypis and Vipin Kumar. Proceedings of 9<sup>th</sup> Supercomputing Conference, pp. 1—24, 1997.
76. "Repertitioning of Adaptive Meshes: Experiments with Multilevel Diffusion". Kirk Schloegel, George Karypis, and Vipin Kumar. Proceedings of the Third Intl. Euro-Par Conference, 1997.
77. "Design and Implementation of a Scalable Parallel Direct Solver for Sparse Symmetric Positive Definite Systems: Preliminary Results". Anshul Gupta, Fred Gustavson, Mahesh Joshi, George Karypis, and Vipin Kumar. Proceedings of the 8<sup>th</sup> SIAM Conference on Parallel Processing for Scientific Computing, 1997.
78. "A Coarse-Grain Parallel Formulation of Multilevel  $k$ -way Graph Partitioning Algorithm". George Karypis and Vipin Kumar. Proceedings of the 8<sup>th</sup> SIAM Conference on Parallel Processing for Scientific Computing, 1997.
79. "WebACE: A Web Agent for Document Categorization and Exploration". J. Moore, E. Han, D. Boley, M. Gini, R. Gross, K. Hastings, G. Karypis, V. Kumar, B. Mobasher. Proceedings of the 2<sup>nd</sup> Intl. Conference on Autonomous Agents, pp. 408—415, 1997.
80. "Multilevel Hypergraph Partitioning: Application in VLSI Domain". George Karypis, Rajat Aggarwal, Vipin Kumar, and Shashi Shekhar. Proceedings of the 34<sup>th</sup> Design and Automation Conference, pp. 526—529, 1997.
81. "Parallel Multilevel  $k$ -way Graph Partitioning". George Karypis and Vipin Kumar. Proceedings of 8<sup>th</sup> Supercomputing Conference, 1996.
82. "Architecture, Algorithms and Applications for Future Generation Supercomputers". Vipin Kumar, Ahmed Sameh, Ananth Grama, and George Karypis. Proceedings of the 6<sup>th</sup> Symposium on the Frontiers of Massively Parallel Computing, pp. 346—354, 1996.
83. "Parallel Multilevel Graph Partitioning". George Karypis and Vipin Kumar. Proceedings of the 10<sup>th</sup> Intl. Parallel Processing Symposium, pp. 314—319, 1996.
84. "Analysis of Multilevel Graph Partitioning". George Karypis and Vipin Kumar. Proceedings of 7<sup>th</sup> Supercomputing Conference, 1995.
85. "Multilevel Graph Partitioning and Sparse Matrix Ordering". George Karypis and Vipin Kumar. Proceedings of the 1995 Intl. Conference on Parallel Processing, 1995.
86. "A High Performance Sparse Cholesky Factorization Algorithm for Scalable Parallel Computers". George Karypis and Vipin Kumar. Proceedings of the 5<sup>th</sup> Symposium on the Frontiers of Massively Parallel Computation, pp. 204—213, 1995.
87. "A Highly Parallel Interior Point Algorithm: Extended Abstract". George Karypis, Anshul Gupta, and Vipin Kumar. Proceedings of the 7<sup>th</sup> SIAM Conference on Parallel Processing, 1995.
88. "A Parallel Formulation of Interior Point Algorithms". George Karypis, Anshul Gupta, and Vipin Kumar. Proceedings of 6<sup>th</sup> Supercomputing Conference, pp. 1057—1072, 1994.
89. "Efficient Parallel Mappings of a Dynamic Programming Algorithm: A Summary of Results". George Karypis and Vipin Kumar. Proceedings of the 7<sup>th</sup> Intl. Parallel Processing Symposium, pp. 563—568, 1993.
90. "Unstructured Tree Search on SIMD Parallel Computers: A Summary of Results". George Karypis and Vipin Kumar. Proceedings of the 4<sup>th</sup> Supercomputing Conference, pp. 453—462, 1992.

### Workshop Papers

1. "The Set Classification Problem and Solution Methods". Xia Ning and George Karypis. ICDM Workshop on Foundations of Data Mining, 2008.
2. "Learning Preferences of New Users in Recommender Systems: An Information Theoretic Approach". Al M Rashid, George Karypis, and John Riedl. SIGKDD Workshop on Web Mining and Web Usage Analysis (WEBKDD), 2008.
3. "A Segment-based Approach to Clustering Multi-Topic Documents". Andrea Tagarelli and George Karypis. Text Mining Workshop, SIAM Data mining Conference, 2008.
4. "A Multi-Level Parallel Implementation of a Program for Finding Frequent Patterns in a Large Sparse Graph". Steve Reinhardt and George Karypis. 12<sup>th</sup> International Workshop on High-Level Parallel Programming Models and Supportive Environments (HIPS), 2007.
5. "ClustKNN: A Highly Scalable Hybrid Model- and Memory-Based CF Algorithm". Al Mamunur Rashid, Shyong K. Lam, George Karypis, and John Riedl. WebKDD 2006 Workshop.

6. “*Finding Functionally Related Genes by Local and Global Analysis of MEDLINE Abstracts*”. Sigve Nakken and Christopher Kauffman, and George Karypis. SIGIR04 Bio Workshop: Search and Discovery in Bioinformatics. 2004.
7. “*Perimeter-Degree: A priori metric for directly measuring and homogenizing interconnection complexity in multilevel placement*”. Navaratnasothie Selvakumaran, Phiroze Parakh, and George Karypis. IEEE Conference on System Level Interconnect Prediction (SLIP), pp. 53—59, 2003.
8. “*Mining Scientific Datasets Using Graphs*”. Michihiro Kuramochi, Mukund Deshpande, and George Karypis. NSF Workshop on Next Generation Data-mining, 2002.
9. “*Automated Approaches for Classifying Structures*”. Mukund Deshpande, Michihiro Kuramochi, and George Karypis. SIGKDD Workshop on Bioinformatics, BIOKDD 2002.
10. “*A Scalable Algorithms for Clustering Protein Sequences*”. Valerie Guralnik and George Karypis. Workshop on Bioinformatics, KDD 2001.
11. “*Efficient Algorithms for Creating Product Catalogs*”. Michael Steinbach, George Karypis, and Vipin Kumar. KDD-2000 Workshop on Web Mining, SIAM Data Mining Conference, 2001.
12. “*A Feature Weight Adjustment Algorithm for Document Classification*”. Shrikanth Shankar and George Karypis. KDD-2000 Workshop on Text Mining.
13. “*Application of Dimensionality Reduction in Recommender System – A Case Study*”. Badrul Sarwar, George Karypis, Joseph Konstan, and John Riedl. WebKDD-2000 Workshop.
14. “*A Comparison of Document Clustering Techniques*”. Michael Steinbach, George Karypis, and Vipin Kumar. KDD-2000 Workshop on Text Mining.
15. “*Load Balancing Across Near-Homogeneous Multi-Resource Servers*”. William Leinberger, George Karypis, Vipin Kumar, Rupak Biswas. In 9<sup>th</sup> Heterogeneous Computing Workshop, pp. 60—71, 2000.
16. “*Clustering Based on Association Rule Hypergraphs*”. Eui-Hong Han, George Karypis, Vipin Kumar, and Bamshad Mobasher. Proceedings of the Workshop on Research Issues on Data Mining and Knowledge Discovery, 1997.
17. “*Web Page Categorization and Feature Selection Using Association Rule and Principal Component Clustering*”. J. Moore, E. Han, D. Boley, M. Gini, R. Gross, K. Hastings, G. Karypis, V. Kumar, B. Mobasher. Proceedings of the 7<sup>th</sup> Workshop on Information Technologies and Systems, 1997.
18. “*Experiences with A Parallel Formulation of An Interior Point Algorithm*”. George Karypis, Anshul Gupta, and Vipin Kumar. DIMACS Series in Discrete Mathematics and Theoretical Computer Science. Vol. 22, pp 163—180, 1995.

## INVITED TALKS

1. “*Algorithms for Graph and Hypergraph Partitioning and They Applications*”, Conference on Graph Theory and Its Applications, Coimbatore, India, December 2008.
2. “*Biclustering Methods meets Formal Concept Analysis*”. Concept Lattices and Their Applications, Olumouc, Czech Republic, October 2008.
3. “*Drug and Probe Discovery and its Mathematical Challenges*”. DOE/NSF Workshop on the Mathematics for Analysis of Petascale Data, June 2008.
4. “*Trends in Bioinformatics*”. Tech Tune-up, University of Minnesota, June 2008.
5. “*Accelerating Drug Discovery: Methods for Effective Virtual Screening and Scaffold Hopping*”. Colloquium, University of Huston, April 2008.
6. “*Indirect Similarity Measures in Cheminformatics*”. Eli-Lilly, December 2007.
7. “*Mining Large Graphs*”, DyDAn Workshop on Associating Semantics with Graphs, Rutgers, April 2007.
8. “*Data Mining for Bioprocess Optimization*”. Genentech Corporation, March 2007.
9. “*Sub-structure-Based Virtual Screening and Retrieval Algorithms in Drug Discovery*”. Agency for Science, Technology, and Research, Bioinformatics Institute, Singapore, April 2006.
10. “*Discovering Knowledge from Life Sciences Literature: Opportunities, Challenges, and Success Stories*”. Keynote speech at the “Workshop in Knowledge Discovery from Life Sciences Literature” at PAKDD, Singapore, April 2006.
11. “*Data-Mining Opportunities in Bioinformatics*”. SAS Data-Mining Conference, October 2003.
12. “*Genomic Grid: Distributed Resources, Data, and Services*”. Data Mining and Exploration Middleware for Distributed and Grid Computing, September 2004, Minnesota Supercomputing Institute, University of Minnesota.
13. “*Classifying Chemical Compounds*”. Eli-Lilly, August 2003.

14. “*Data-Mining and Bioinformatics*”. St. Cloud State University, January 2003.
15. “*Data-Mining and Bioinformatics*”. Minnesota IT Leadership Forum, October 2002.
16. “*Clustering Documents and its Applications*”. 7<sup>th</sup> Annual Text Summit, Thompson Publishing, September 2002 (keynote speech).
17. “*Frequent Subgraph Discovery: Mining Scientific and Relational Data Sets*”. IPAM workshop on Scientific Data Mining, UCLA, January 2002.
18. “*Multilevel Algorithms for Circuit Partitioning*”, IPAM workshop on Multilevel Methods for VLSI Design, UCLA, December 2001.
19. “*Selective Markov Models*”. Honeywell Laboratories, March 2001.
20. “*Concept Indexing: A Fast Dimensionality Reduction Algorithm with Applications to Document Retrieval & Categorization*”. IMA Workshop on Text Mining, Minneapolis, April 2000.
21. “*Text Mining*”. Purdue, Computer Science Department, April 2000.
22. “*Data Mining in Genomics*”. Incyte Pharmaceuticals, Palo Alto, April 2000.
23. “*Genome Computing Issues and Mining Gene Expression Data*”. IEEE CS/IEEE EMBS, Minneapolis, November 1999.
24. “*Multi-Constraint and Multi-Objective Graph Partitioning*”. AHPCRC workshop on Graph Partitioning, Minneapolis, October 1999.
25. “*Chameleon: Clustering Using Dynamic Modeling*”. AHPCRC workshop on Scientific Data Mining, Minneapolis, September 1999.
26. “*Data Mining Research at AHPCRC*”. Center for Army Analysis, Washington, D.C., September 1999.
27. “*Clustering and Classification of High Dimensional Data-Sets*”. Lawrence Livermore National Lab, November 1998.
28. “*Multi-Constraint Graph Partitioning*”. Lawrence Livermore National Lab, October 1998.
29. “*Multi-label Classification of Statutes Documents*”. WEST Publishing Group, September 1998.
30. “*Multilevel Nested Dissection: Experiences with Parallel Formulations*”. SIAM Conference on Linear Algebra, October 1997.
31. “*Multilevel Repartitioning of Adaptive Meshes*”. Army HPC Research Center Workshop on Unstructured Mesh Generation and Partitioning, October 1997.
32. “*Parallel and Adaptive Graph Partitioning*”. Lawrence Livermore National Lab, April 1997.
33. “*Graph Algorithms and Data Mining*”. Pataflops Algorithm Workshop, April 1997.
34. “*Parallel k-way Mesh Partitioning. Workshop on Parallel Unstructured Grid Computations*”. Argonne National Lab, September 1996.
35. “*Experiences with a Parallel Formulation of an Interior Point Algorithm*”. DIMACS Workshop on Parallel Processing of Discrete Optimization Problems, February 1995.
36. “*Multilevel Graph Partitioning Algorithms*”. Cray Research, September 1994.

## TUTORIALS

1. “*Computational Methods for DNA and Protein Sequence Analysis*”. Genomics Signal Processing and Statistics, College Station, TX, 2006.
2. “*Parallel Partitioning Software for Static, Dynamic, and Multi-phase Computations*”. Supercomputing 2001, November 2001, Denver, CO.
3. “*Data mining for Genomics*”. 1<sup>st</sup> SIAM Conference on Data Mining, April 2001, Chicago, IL.
4. “*Using METIS and ParMETIS*”. Army HPC Research Center’s Workshop on “*Graph Partitioning and Applications: Current and Future Directions*”, October 1999

## RESEARCH GRANTS

1. “*Functional Genomics of Nectar Production in Brassicaceae*”, NSF, \$1,336,289, 9/1/2008—8/31/2012 (with Clay Carter).
2. “*Discerning Pivotal High Productivity Characteristics through Recognition of Patterns in Process Data*”, GenenTech, \$108,750, 12/1/2007—12/1/2008 (with Wei-Shou Hu).
3. “*Effective & Efficient Whole Genome Alignment Algorithms*”, IBM Rochester, \$35,000, 6/1/2006—6/1/2007.
4. “*Classification Algorithms for Chemical Compounds*”, NIH, \$1,149,001, 9/30/2005—9/30/2009.
5. “*SEI: Virtual Screening Algorithms for Bioactive Compounds Based on Frequent Substructures*”, NSF, \$405,498, 9/1/2004—8/31/2009.

6. “*ITR: Graph Partitioning Algorithms for Complex Problems & Applications*”. NSF, \$122,000, 8/25/2003—8/24/2005.
7. “*Summer Bioinformatics Institute*”, NSF/NIH, \$498,596, 01/01/03—12/31/05 (with V. Kumar, J. Carlis, L. Ellis, A. Grosberg, V. Kapur, A. Odlyzko, H. Othmer, W. Pan, R. Phillips, E. Retzel, K. Silverstein, D. Truhlar, N. Young).
8. “*CAREER: Scalable Algorithms for Knowledge Discovery in Scientific Data Sets*”. NSF, \$320,900, February 2002—January 2008.
9. “*Scalable Algorithms for Scientific Computations*”, Army Research Office, \$520,000, Fall 2001—Fall 2006 (as part of AHPERC).
10. “*Pathogenesis and Therapy of Chronic Lung Rejection*”, National Institute of Health, \$1,479,387, Fall 2001—Fall 2006 (with M. Hertz, R. King, V. Kapur, E. Retzel, H. Chen, and K. Savik).
11. “*Autoimmune Biomarkers Collaboratory*”, NIH, \$1,525,454 Fall 2001—Fall 2006 (with T. Behrens).
12. “*Discovery of Changes from the Global Carbon Cycle and Climate System Using Data Mining*”. NASA, \$525,091, Spring 2001- Spring 2004 (with V. Kumar, S. Shekhar, S. Klooster, C. Potter, and A. Torregrosa).
13. “*CISE Research Instrumentation: Cluster Computing for Knowledge Discovery in Diverse Data Sets*”. National Science Foundation, \$121,618, February 2000—January 2003 (with M. Gini, J. Riedl, J. Konstan, S. Shekhar, J. Srivastava).
14. “*Parallelization of KIVA*”. Army Research Office, \$240,000, August 2000—July 2003 (with S. Garrick and V. Kumar)
15. “*Scientific Data Mining*”. Department of Energy, \$120,000, March 2000—February 2001 (with V. Kumar).
16. “*Dynamic Feature Extraction and Data Mining for Analysis of Turbulent Flows*”. National Science Foundation, \$1,462,500, October 1999—September 2002 (with V. Kumar, V. Interrante, G. Candler, I. Marusic, Longmire, S. Garrick).
17. “*Multi-Constraint Multi-Objective Graph Partitioning*”. National Science Foundation, \$386,544, September 1999—August 2002 (with V. Kumar).
18. “*Scalable Parallel Algorithms for Irregular & Adaptive Computations*”. Department of Energy (Level II ASCI Initiative), \$578,000; October 1998 – September 2001; (with V. Kumar).
19. “*Scalable Parallel Algorithms for Solving Sparse Linear Systems*”. Army Research Office, \$230,000; September 1998 – August 2001; (with V. Kumar).
20. “*Graph Partitioning for Dynamic, Adaptive and Multi-Phase Computations*”. SGI/Cray, \$55,000; January 1998 – December 1999; (with V. Kumar).
21. “*Load Balancing on the Information Power Grid*”. NASA, \$40,000; May 1998 – September 1998; (with V. Kumar).
22. “*Scalable Data Mining Algorithms*”. Army Research Office (ASSERT); \$75,000; May 1997 – April 2000; (with V. Kumar).

## SOFTWARE DEVELOPED

- METIS** Serial software package for partitioning unstructured graphs and for computing fill reducing matrix re-orderings. METIS is used extensively in numerous application areas including scientific computing, parallel and distributed processing, operations research, geographical information systems, molecular biology, and data mining.  
URL: <http://www.cs.umn.edu/~metis/metis>.
- hMETIS** Serial software package for partitioning hypergraphs. hMETIS is based on the multilevel paradigm and is able to quickly compute very high quality partitions of very large and irregular hypergraphs. It is used extensively to partition hypergraphs corresponding to VLSI circuits, in data mining for clustering, and to optimize the storage of databases on disks.  
URL: <http://www.cs.umn.edu/~metis/hmetis>.
- PARMETIS** An MPI-based parallel library for partitioning unstructured and adaptively refined meshes and for computing fill-reducing matrix re-orderings. It is a highly parallel implementation of the serial METIS package; with additional functionality to accommodate needs for partitioning and load balancing that exist only on parallel computations.  
URL: <http://www.cs.umn.edu/~metis/parmetis>.
- PSPASES** An MPI-based library that implements a parallel sparse Cholesky-based direct solver. It incorporates a highly parallel multi-frontal Cholesky algorithm, as well as highly parallel

algorithms for computing fill reducing orderings, symbolic factorization, and forward and backward substitution.

URL: <http://www.cs.umn.edu/~mjoshi/pspases>.

**SUGGEST**

A collaborative filtering based top-*N* recommendation engine. It uses an efficient item-based model that adapts to the sparsity of the data set that leads to real-time high quality recommendations.

URL: <http://www.cs.umn.edu/~karypis/suggest>.

**MGRIDGEN**

A highly optimized serial and parallel library for obtaining a sequence of successive coarse grids that is well suited for geometric multigrid methods. The quality of the elements of the coarse grids is optimized using a multilevel framework. The parallel library is based on MPI and is portable to a wide-range of architectures.

URL: <http://www.cs.umn.edu/~moulitsa/software.html>.

**CLUTO**

A software package for clustering low- and high-dimensional data sets. It treats data clustering as an optimization problem that tries to optimize a particular clustering criterion function. It provides a variety of clustering criterion functions and various partitional and agglomerative clustering algorithms.

URL: <http://www.cs.umn.edu/~cluto>.

**gCLUTO**

A cross-platform graphical user interface tool on top of the CLUTO library that allows the users to interactively load, cluster, and visualize their datasets. One of its key features is the extensive cluster visualization capabilities that include, tree, matrix, and an OpenGL-based mountain-view of the clustering solution.

URL: <http://www.cs.umn.edu/~cluto/gcluto>.

**wCLUTO**

wCLUTO is a web-enabled data clustering application that is designed for the clustering and data-analysis requirements of gene-expression analysis. wCLUTO is also built on top of the CLUTO clustering library. Users can upload their datasets, select from a number of clustering methods, perform the analysis on the server, and visualize the final results.

URL: <http://cluto.ccg.umn.edu>.

**PAFI**

A software package for discovering frequent patterns in diverse datasets. It contains three main frequent pattern discovery algorithms that can be used to find frequent itemset, sequences, and graph patterns in large databases.

URL: <http://www.cs.umn.edu/~pafi>.

**YASSPP**

A web-server for predicting the secondary structure of proteins from primary sequence. It is based on a cascaded SVM-based machine learning model that combines custom-designed kernel functions with evolutionary information.

URL: <http://yasspp.cs.umn.edu>

**AFGEN**

AFGen is a program that takes as input a set of chemical compounds and generates their vector-space representation based on the set of fragment-based descriptors they contain. This vector-based representation can be used for different tasks in cheminformatics including similarity search, virtual screening, and library design.

URL: <http://glaros.dtc.umn.edu/gkhome/afgen/overview>

**MONSTER**

A web-based server that provides a set of services for annotating residues with functional and structural properties from sequence information only. The structural and functional annotations that are currently provided are secondary structure, transmembrane helices, disorder regions, solvent accessible surface area, DNA binding residues, contact order, and protein blocks.

URL: <http://bio.dtc.umn.edu/monster>

## PROFESSIONAL ACTIVITIES

### Editorships

1. Editorial Board Member, Journal of Biomedicine and Biotechnology; 2008—present.
2. Editorial Board Member, Advances in Bioinformatics; 2007—present.

3. Editorial Advisory Board Member, Current Proteomics; 2007—present.
4. Editorial Board Member, International Journal of Data Mining and Bioinformatics; 2005—present.
5. Associate Editor, IEEE Transactions on Parallel and Distributed Systems; 2003—2007.
6. Guest editor of the special issue of the ACM Transactions on Knowledge Discovery from Data on “Bioinformatics”; 2007.
7. Guest editor of the special issue of IEEE Computing in Science & Engineering on “*Data Mining in Science*”; 2002.
8. Guest editor of the special issue of *Parallel Computing Journal* on “*Graph Partitioning and Parallel Computing*”; 1999.

#### **Conference Chair**

1. Area chair for SIAM Data Mining Conference, Minneapolis, MN, 2007.
2. Area chair for ECML/PKDD-06 Conference, Berlin, Germany, 2006.
3. General Chair of the 6<sup>th</sup> IEEE Symposium on Bioinformatics and Bioengineering (BIBE), Washington, 2006.
4. Chair of the 5<sup>th</sup> IEEE Symposium on Bioinformatics and Bioengineering Conference (BIBE), Minneapolis, 2005.
5. Co-Chair of the 4<sup>th</sup> IEEE Bioinformatics and Bioengineering Conference (BIBE), Taiwan, 2004.
6. Vice Chair of the Program Committee for the 5<sup>th</sup> IEEE International Conference on Data Mining, New Orleans, Louisiana, November 2005.

#### **Workshop Organizer**

1. Member of the organizing committee of the 6<sup>th</sup> SIGKDD workshop on Data Mining in Bioinformatics, which occurred during the SIGKDD 2006 Conference, August 2006.
2. Member of the organizing committee of the 3<sup>rd</sup> International Workshop on Mining Graphs, Trees, and Sequences (MGTS), which occurred during the ECML/PKDD 2005 Conference, October 2005.
3. Member of the organizing committee of the PAKDD workshop on “*Text Mining*”, which occurred during the 6<sup>th</sup> Pacific Asia Conference on Knowledge Discovery and Data Mining, May 2002.
4. Member of the organizing committee of the SIAM workshop on “*Data mining for Genomics*”, which occurred during the 1<sup>st</sup> SIAM Conference on Data Mining, April 2001.
5. Member of the organizing committee of the Army HPC Research Center’s Workshop on “*Graph Partitioning and Applications: Current and Future Directions*”, October 1999.
6. Organizer of a mini-symposium on “*High Performance Data Mining*” at the “9<sup>th</sup> SIAM Conference on Parallel Processing for Scientific Computing”, 1999.
7. Member of the organizing committee of the Army HPC Research Center Workshop on “*Unstructured Mesh Generation and Partitioning*”, 1998.

#### **Conference Program Committee Memberships**

1. International Conference on Bioinformatics and Computational Biology (BICoB): 2009.
2. International Conference on Machine Learning and Applications (ICMLA): 2008.
3. European Conference on Computational Biology (ECCB): 2008
4. International Conference on Database and Expert Systems (DEXA): 2008.
5. International Symposium on Bioinformatics Research and Applications (ISIBRA): 2008.
6. Pacific-Asia Conference on Knowledge Discovery and Data Mining (PAKDD): 2007—present.
7. International Conference on Genome Informatics (GIW): 2007—present.
8. ECML/PKDD Conference: 2006—present.
9. IEEE International Conference on Bioinformatics and biomedicine (BIBM): 2007—present.
10. ACM SIGKDD Conference on Knowledge Discovery and Data Mining: 2004—present.
11. IEEE International Conference on Data Mining (ICDM): 2004—present.
12. IEEE Symposium on Bioinformatics and Bioengineering (BIBE): 2004—present.
13. SIAM Data Mining Conference: 2003—present.
14. Conference of the American Association of Artificial Intelligence (AAAI): 2006.
15. ACM Conference on Information and Knowledge Management (CIKM): 2006—present..
16. International Conference on Database Systems for Advance Applications (DAFSAA): 2006—2007.
17. International Parallel and Distributed Processing Symposium (IPDPS): 2004, 2006—present.
18. International World-Wide-Web Conference (WWW): 2003.
19. International Conference on High Performance Computing (HiPC): 2004.

20. International Conference on Parallel Processing (ICPP): 2003.
21. Supercomputing Conference: 2002, 2007.

### **Workshop Program Committee Memberships**

1. Workshops held in conjunction with the SIGKDD conference:
  1. Large Scale Recommender Systems and the Netflix Prize Competition: 2008.
  2. Workshop on Link discovery: Issues, Approaches and Applications (LinkKDD): 2005—2006.
  3. Open Source Data Mining Workshop (OSDM): 2005.
  4. Multi-Relational Data Mining (MRLDM): 2005.
  5. Workshop on Knowledge Discovery in the Web (WebKDD): 2005—2006, 2008.
  6. Workshop on Data Mining in Bioinformatics (BIOKDD): 2002—2006.
2. Workshops held in conjunction with the ICDE conference.
  1. Workshop on Data Engineering Methods in Bioinformatics (DEBI): 2009.
3. Workshops held in conjunction with the ICDM conference:
  1. High Performance Data Mining Workshop: 2009.
  2. Workshop on Data Mining in Bioinformatics: 2004.
4. Workshops held in conjunction with the SIAM Data Mining conference:
  1. Bioinformatics Workshop: 2004.
  2. Workshop on Clustering High Dimensional Data Sets and its Applications: 2002—2003.
  3. Spatial Data Mining: 2006
5. Workshops held in conjunction with VLDB:
  1. Workshop on Data Mining and Bioinformatics: 2006.
6. Workshops held in conjunction with ECML/PKDD:
  1. Parallel Data Mining (PDM): 2006.
  2. Mining and Learning on Graphs (MLG): 2007—2008.
7. Workshops held in conjunction with IPDPS:
  1. Workshop on High-Performance Grid Computing: 2003—2006.
8. International Workshop on “*Biological Data Management*”, (BIDM): 2004—2005.
9. International workshop on Geographic and Biological Data Management (GBDM): 2004.
10. International workshop on Distributed Data Mining in Life Sciences (LifeDDM): 2005.

### **Reviewer**

1. Served as the reviewer for over five hundred papers in various journals (including ACM Transactions on Computational Biology and Bioinformatics, ACM Transactions on Information Systems, ACM Transactions on Internet Technology, Statistical Analysis and Data Mining, Bioinformatics, BMC Bioinformatics, Proteins, Data Mining and Knowledge Discovery, Journal of Combinatorics, Machine Learning Journal, Data and Knowledge Engineering, Pattern Analysis and Applications, Pattern Recognition, Knowledge and Information Systems, Parallel Computing, SIAM Journal on Scientific Computing, Acta Informatica, International Journal of Computer Mathematics, IEEE Transactions on Computers, IEEE Transactions on Knowledge and Data Engineering, IEEE Transactions on Computer Aided Design, IEEE Transactions on Pattern Analysis and Machine Intelligence, IEEE Transactions on Parallel and Distributed Systems, Journal of Parallel and Distributed Computing, IEEE Concurrency, Journal of Experimental Algorithms, Image and Vision Computing, IEEE Signal Processing Letters, IEEE Journal of Selected Topics in Signal Processing, IEEE Communications Letters, IEEE Systems, Man and Cybernetics) and conferences for which I have served on their program committee.
2. Served as an external reviewer for proposals submitted to NSF, DOE, ARL, ARO, NASA, State of Louisiana, and Science Foundation of Ireland (SFI), on ten NSF review panels, on three NIH study sections, and participated on a site visit for SFI.

## **DEGREES UNDER MY SUPERVISION**

### **Ph.D. Current**

1. Kevin DeRonne (passed preliminary oral exam on 2/2008)
2. Amine Abou-Rjeli (passed WPE)
3. Chris Kauffman (passed WPE)

4. Rezwan Ahmed (passed WPE)
5. Xia Ning (passed WPE)
6. Yevgeniy Podolyan
7. Santosh Kabbur

### Completed

1. Sam Han (Fall 1999, with V. Kumar, currently employed at iXmatch Inc.)
2. Kirk Schloegel (Fall 1999, with V. Kumar, currently employed at Honeywell)
3. Valery Guralnik (2001, with J. Srivastava, currently employed at Honeywell)
4. William Leinberger (2001, with V. Kumar, currently employed at General Dynamics)
5. Mukund Deshpande (2003, with J. Srivastava, currently employed at Persistent Systems Ltd, India)
6. Navaratnasothie Selvakumaran (2005, currently employed at Xilinx)
7. Irene Moulitsas (2005 with Y. Saad, currently at University of Cyprus, Cyprus)
8. Michihiro Kuramochi (2005, currently employed at Google Inc.)
9. Ying Zhao (2005, with D. Du, currently at Tsinghua University, China)
10. Irina Makarevitch (2005) (Applied Plant Sciences)
11. Huzefa Rangwala (2008) (currently at George Mason University)
12. Nikil Wale (2008) (currently employed at Pfizer Corp.)

### M.S. Completed

1. Sushrut Karanjkar (Spring 1998)
2. Dalvinder Malhotra (Winter 1998)
3. Kapil Surlekar (Spring 1999)
4. William Leinberger (Spring 1999)
5. Shrikanth Shankar (spring 2000)
6. Md. Al Hasan (Fall 2001)
7. Ekta Sirohi (Fall 2002)
8. Masakazu Seno (spring 2002)
9. Qing Zhang (Fall 2002)
10. Chang Liu (Fall 2002)
11. Sai Chen (Summer 2003)
12. Rezwan Ahmed (Spring 2003)
13. Nivea Garg (Fall 2003)
14. Krishna Gades (Spring 2004)
15. Eunah Cho (Spring 2004)
16. Jay Vasdewani (Spring 2004)
17. Mahbubur Rahim Khan (Fall 2004)
18. Aris Goulalas-Divanis (Spring 2005)
19. Brian Wallenfelt (Spring 2006)

### HONORS

- Distinguished Paper Award at EuroPar 2000.
- Honorable Mention (2<sup>nd</sup> Place) at KDDCup 2000 competition.
- First Prize Award at Mannheim SuParCup 95 (European Supercomputing Conference).
- Cray Research Fellow for 1995-96.
- Graduate School Fellow University of Minnesota for 1992-93.

### EDUCATION

- 1992-1996      UNIVERSITY OF MINNESOTA, Minneapolis, MN  
 Ph.D. in Computer Science, Spring 1996. GPA 4.0/4.0  
 Dissertation title: “*Graph Partitioning and Its Applications to Scientific Computing*”  
 Dissertation advisor: Vipin Kumar
- 1988-1992      UNIVERSITY OF MINNESOTA, Minneapolis, MN

BS in Computer Science, Spring 1992, Cum Laude, GPA 4.0/4.0

## PROFESSIONAL EXPERIENCE

Summer 2004 to present Computer Science Department, University of Minnesota  
**ASSOCIATE PROFESSOR**

Fall 1999 to Spring 2004 Computer Science Department, University of Minnesota  
**ASSISTANT PROFESSOR**

Summer 1996 to Fall 1999 Computer Science Department, University of Minnesota  
**RESEARCH ASSOCIATE**

## TEACHING EXPERIENCE

1. “*Algorithms & Data Structures I*”. CS 3321/CS4041.
2. “*Introduction to Parallel Programming*”. CS 5151/CS5451.
3. “*Introduction to Data Mining*”. CS 8475.
4. “*Computational Techniques for Genomics*”. CS8980/CS5481
5. “*Summer Institute—Army HPC Research Center*”. Summers of 1997 & 1998.