LAWRENCE
LIVERMORE
NATIONAL
LABORATORY

FINAL REPORT (MILESTONE DATE 9/30/13) FOR SUBCONTRACT NO. B603393: "CLUSTERING AND RANDOMIZATION FOR LARGE GRAPHS AND HYPERGRAPHS"

H. De Sterck

September 12, 2013

Disclaimer

This document was prepared as an account of work sponsored by an agency of the United States government. Neither the United States government nor Lawrence Livermore National Security, LLC, nor any of their employees makes any warranty, expressed or implied, or assumes any legal liability or responsibility for the accuracy, completeness, or usefulness of any information, apparatus, product, or process disclosed, or represents that its use would not infringe privately owned rights. Reference herein to any specific commercial product, process, or service by trade name, trademark, manufacturer, or otherwise does not necessarily constitute or imply its endorsement, recommendation, or favoring by the United States government or Lawrence Livermore National Security, LLC. The views and opinions of authors expressed herein do not necessarily state or reflect those of the United States government or Lawrence Livermore National Security, LLC, and shall not be used for advertising or product endorsement purposes.

FINAL REPORT (MILESTONE DATE 9/30/13)
FOR SUBCONTRACT NO. B603393: "CLUSTERING AND RANDOMIZATION FOR LARGE GRAPHS AND HYPERGRAPHS"

Description of work performed for the subcontract

The following work has been performed by PI Hans De Sterck, graduate student Manda Winlaw, postdoc Rasha Kashef, and graduate student Haifeng Xu, for the required tasks 1-5 (as listed in the Statement of Work):

1. Graduate student Manda Winlaw visited LLNL March-September, 2013, working with Geoffrey Sanders and Van Emden Henson on randomized graph generators. In particular, algorithms have been devised for improving triangle counts and clustering coefficients in the randomized graphs. The first approach that has been taken is to improve triangle counts for a random graph that is initially generated using a Chung-Lu process using a new switching mechanism. The switching mechanism adds or removes triangles to adjust the triangle count without changing the degree distribution. A working prototype has been developed and results have been obtained that will form the basis of a research paper to be submitted in Fall 2013. Second, a prototype implementation has been developed of a two-phase mechanism to first generate triangles using dense blocks, and then correct the degree distribution by a bipartite random process. These methods have the ability to significantly extend the fidelity with which random graphs can be generated that match important properties of input graphs, with applications in computer network modeling, information search and retrieval, and social networking.
2. Investigations have been carried out by Winlaw and Sanders to develop randomized graph generation methods for directed graphs and hypergraphs. It was considered how the principles from the graph generators developed for Task 1 can be extended to these problems, and approaches to this end were identified.
3. Graduate student Manda Winlaw and PI Hans De Sterck have developed new nonlinearly preconditioned nonlinear conjugate gradient methods for optimization. When applied to rectangular tensorial data sets, the new method performs much better than current state-of-the-art methods like standard, non-preconditioned nonlinear conjugate gradients, and alternating least-squares. A manuscript has been prepared that will be submitted in Fall 2013.
4. Postdoc Rasha Kashef with PI Hans De Sterck has developed new multilevel clustering methods for detecting nested overlapping communities in social networks, with application to LinkedIn data and other social network and textual data. Graduate student Haifeng Xu, with Sanders and De Sterck, has developed novel multilevel co-clustering methods for social network data and microarray date. The algorithm co-clusters, for example, users and keywords of social networks for finding nested overlapping communities in the bipartite users-keywords graph (rectangular feature matrix). Both developments have resulted in new algorithms that provide an attractive balance between clustering accuracy and algorithm performance and scalability. Manuscripts are being prepared for both new algorithms, to be submitted in Fall 2013.
5. PI Hans De Sterck has consulted with Sanders and Henson on advanced algorithms for the general problems of developing scalable multilevel factorizations for the deep

analysis of data sets represented by nonsymmetric square and rectangular matrices. Various algorithms from the recent literature were considered and were reported on for guiding directions of future algorithm development. (Task 5 pertains to the consulting portion of the subcontract.)

Publications enabled by the subcontract:
1. Hans De Sterck, Geoff Sanders and Haifeng Xu, 'Fast Multilevel Co-Clustering', manuscript.
2. Hans De Sterck, Geoff Sanders and Manda Winlaw, 'Randomized Graph Generation with Proper Triangle Statistics', manuscript.
3. Hans De Sterck and Manda Winlaw, 'A Nonlinearly Preconditioned Nonlinear Conjugate Gradient Algorithm for Canonical Tensor Decomposition', manuscript.
4. Hans De Sterck and Rasha Kashef, 'An Algebraic Multigrid Method for Finding Multilevel Overlapping Communities in Social Networks: Application to LinkedIn Skills and Expertise', manuscript.
5. Hans De Sterck and Killian Miller, 'An Adaptive Algebraic Multigrid Algorithm for Low-Rank Canonical Tensor Decomposition', SIAM J. Sci. Comput. 35, B1–B24, 2013.
6. Hans De Sterck, 'Steepest Descent Preconditioning for Nonlinear GMRES Optimization', Numerical Linear Algebra with Applications 20, 453-471, 2013.

Presentations enabled by the subcontract:
1. Hans De Sterck, First Canadian Symposium on numerical Analysis and Scientific Computing, Quebec City, June 17, 2013. 'Extending Preconditioned GMRES to Nonlinear Optimization', invited.
2. Hans De Sterck, Weizmann Workshop on Multilevel Computational Methods and Optimization, Rehovot, Israel, May 1, 2013. 'Adaptive Algebraic Multigrid for Canonical Tensor Decomposition', invited.
3. 16th Copper Mountain Conference on Multigrid Methods, March 21, 2013. 'Adaptive Algebraic Multigrid for Singular Value Decomposition'.