

Partitioning hypergraphs for multiple communication metrics

Mehmet Deveci ¹, Kamer Kaya ¹, Bora Uçar ², Ümit V. Çatalyürek ¹

¹ Dept. Biomedical Informatics, The Ohio State University

² CNRS and LIP, ENS Lyon, France

The datasets in many fields of science and engineering are growing rapidly with the recent advances that enable generating and storing big data. These datasets usually have irregular patterns, and there exist complex interactions between their elements. These interactions make it difficult to develop simple and efficient load-balancing mechanisms for the data analysis kernels, forcing researchers to study more complex partitioning methods. Thus, a good partitioning of the data, which is necessary to obtain a scalable performance, a shorter makespan, a minimized energy usage, and a better resource utilization, is more important and harder to obtain.

In scientific computing, computational tasks are usually modeled using connectivity-based topological models, such as graphs and hypergraphs [1, 2, 3]. These models transform the problem at hand to a balanced vertex partitioning problem. The balance restriction on part weights in conventional partitioning corresponds to the load balance in a parallel environment, and the minimization objective for a given metric relates to the minimization of the communication cost among the processing units. A good partitioning should minimize the inter-processor communication while distributing the computation load evenly to the processors. In this work, we study these connectivity-based models and methods, specifically hypergraph models and methods, designed to partition the communicating tasks for an efficient parallelization.

Most of the existing state-of-the-art hypergraph partitioners aim to minimize the total communication volume while balancing the load for efficient parallelization [1, 4, 5]. However, the other communication metrics, such as the total number of messages, the maximum amount of data transferred by a single processor, or combinations of multiple metrics, are equally, if not more, important. For example, the latency-based metrics, which model the communication by using the number of messages sent/received throughout the execution, become more and more important as the number of processors increases [6]. Ideal partitions yield perfect computational load balance and minimize the communication requirements by minimizing the communication overhead. On the other hand, most of the existing hypergraph partitioning methods aim to minimize only the traditional total communication volume metric, with the hope that it improves the other communication metrics as a side effect.

In this work, we argue that the general hypergraph model used by the state-of-art hypergraph partitioners is not sufficient to model and capture other communication metrics than the total volume of communication. We propose a directed hypergraph model that can simultaneously capture multiple communication metrics. Given an application, our main objective is to partition the tasks evenly among processing units and to minimize the communication overhead by minimizing several communication cost metrics by using the proposed directed hypergraph model. Previous studies addressing multiple metrics [7, 8] with the traditional hypergraph model work in two phases where the phases are concerned with disjoint subsets of communication metrics. Generally, the first phase tries to obtain a proper partitioning of data for which the total communication volume is reduced. Starting from this partitioning, the second phase tries to optimize another communication met-

ric. Even though such two-phase approaches allow the use of state-of-the-art techniques in one or both phases, since the solutions sought in one phase are oblivious of the metric used in the other, the search can get stuck in a local optima that cannot be improved in the other phase. Instead, we present a novel approach to treat the minimization of multiple communication metrics as a multi-objective minimization problem which is solved in a single phase with the help of the proposed directed hypergraph model. Addressing all the metrics in a single-phase allows a trade-off between the cost associated with one metric and the cost associated with another one. Inherently, the standard hypergraph model cannot see the communication metrics that are defined per-processor basis. Therefore, the balance on the communication loads of the processors cannot be modeled and formulated in a natural way. Furthermore, since almost all the state-of-the-art partitioners use iterative-improvement-based heuristics for the refinement, a single-phase approach increases the explored search space by avoiding local optima for a single metric.

We have materialized our approach in UMPa [9], which is a multi-level partitioner employing a directed hypergraph model and novel K -way refinement heuristics, since balancing per-processor communication cost metrics requires a global view of the partition. We present methods for minimizing the maximum communication volume, the total and maximum number of messages per processor, as well as the traditional total communication volume in a generalized framework that simultaneously minimizes multiple prioritized communication overhead metrics at the same time. Compared to the state-of-art hypergraph partitioners, we show on a large number of problem instances that UMPa produces much better partitions in terms of several communication metrics with 128, 256, 512, and 1024 processing units; UMPa reduces the maximum communication volume, the total number of messages, and the maximum number of messages sent by a single part up to %85, %45, and %43, respectively.

References

- [1] Ü. V. Çatalyürek, C. Aykanat, Hypergraph-partitioning based decomposition for parallel sparse-matrix vector multiplication, *IEEE Transactions on Parallel and Distributed Systems* 10 (7) (1999) 673–693.
- [2] B. Hendrickson, T. G. Kolda, Graph partitioning models for parallel computing, *Parallel Computing* 26 (2000) 1519–1534.
- [3] Ü. V. Çatalyürek, C. Aykanat, A fine-grain hypergraph model for 2D decomposition of sparse matrices, in: *Proceedings of 15th International Parallel and Distributed Processing Symposium (IPDPS)*, San Francisco, CA, 2001.
- [4] B. Vastenhouw, R. H. Bisseling, A two-dimensional data distribution method for parallel sparse matrix-vector multiplication, *SIAM Review* 47 (1) (2005) 67–95.
- [5] K. Devine, E. Boman, R. Heaphy, R. Bisseling, Ü. V. Çatalyürek, Parallel hypergraph partitioning for scientific computing, in: *Proceedings of 20th International Parallel and Distributed Processing Symposium (IPDPS)*, IEEE, 2006.
- [6] K. Kaya, B. Uçar, U. V. Çatalyürek, On analysis of partitioning models and metrics in parallel sparse matrix-vector multiplication, in: *Proc. of the 10th Int’l Conf. on Parallel Processing and Applied Mathematics (PPAM)*, 2013.
- [7] B. Uçar, C. Aykanat, Encapsulating multiple communication-cost metrics in partitioning sparse rectangular matrices for parallel matrix-vector multiplies, *SIAM J. Sci. Comput.* 25 (2004) 1837–1859.
- [8] R. H. Bisseling, W. Meesen, Communication balancing in parallel sparse matrix-vector multiplication, *Electronic Transactions on Numerical Analysis* 21 (2005) 47–65.
- [9] Ü. V. Çatalyürek, M. Deveci, K. Kaya, B. Uçar, UMPa: A multi-objective, multi-level partitioner for communication minimization, in: *10th DIMACS Implementation Challenge Workshop: Graph Partitioning and Graph Clustering*, 2012, pp. 53–66, published in *Contemporary Mathematics*, Vol. 588, 2013.