Title: Maximal Strip Recovery based on Interval Graphs

Vicky Choi
Department of Computer Science, Virginia Tech, USA
vchoi@cs.vt.edu

Chunfang Zheng ,
Département de recherche opérationnelle et informatique, Université de Montréal, Canada

David Sankoff
Department of Mathematics & Statistics, University of Ottawa, Canada

Abstract:
The Maximal Strip Recovery (MSR) problem was formulated in 2007 for the extraction of synteny blocks from noisy genomic maps or gene orders for comparative purposes [1]. Our objective was to identify a set of consistent maximal strips (or equivalently to remove a minimum set of noisy data) that are credible from a genome rearrangement viewpoint. Over the last few years, there have been a number of follow-up papers [3-7] on MSR that showed the hardness of the problem and its variants. In particular, they proposed some approximation and FPT algorithms for these problems.

For this poster, we revisit the problem and propose a biologically more relevant approximation algorithm. Originally [1, 2] formulated MSR in terms of the equivalent NP-hard problems of Maximum Weight Clique or Maximum Weight Independent Set (MWIS). A solution was obtained by 1), resolving the problem graph in terms of an interval graph and 2), a linear-time algorithm for the MWIS problem on the interval graph. The optimization problem in step 1), namely removing a minimum number of vertices for the graph to become interval, known as Interval Vertex Deletion (IVD) problem, is NP-hard. In [2], we employed an efficient heuristic to solve IVD (and consequently our solution to MSR is also a heuristic). In Chen et al. [3], a 4-approximation algorithm was given based on the observation that the problem graph is a 2-interval graph, since a 4-approximation algorithm for MWIS on 2-interval graphs is known. While it is correct that the problem graph is a special, constrained, case of a 2-interval graph, the problem graph, discounting noise, is more appropriately considered an interval graph, as the genome rearrangement operations invoked in the objective function will not destroy the interval property. In other words, it is only noise that interferes with the interval graph property. This was our biological motivation in [2]. In this poster, we describe a closer approximation algorithm for IVD directly, and thus to MSR, which avoids a relaxation to the unconstrained 2-interval property.

[1] C. Zheng, Q. Zhu, and D. Sankoff. Removing noise and ambiguities from comparative maps in rearrangement analysis. IEEE/ACM Transactions on Computational Biology and Bioinformatics, 4:515–522, 2007.
[2] V. Choi, C. Zheng, Q. Zhu, and D. Sankoff. Algorithms for the extraction of syntheny blocks from comparative maps. In Proceedings of the 7th International Workshop on Algorithms in Bioinformatics (WABI'07), LNCS 4645, pages 277–288, 2007.
[3] Z. Chen, B. Fu, M. Jiang, and B. Zhu. On recovering synthetic blocks from comparative maps. Journal of Combinatorial Optimization, 18:307–318, 2009.
[4] M. Jiang. Inapproximability of maximal strip recovery. Proceedings of the 20th Annual International Symposium on Algorithms and Computation (ISAAC'09), LNCS 5878, pages 616-625, 2009.
[5] L. Wang and B. Zhu. On the tractability of maximal strip recovery. J. of Computational Biology, 17(7):907-914, 2010.
[6] H. Jiang, Z. Li, G. Lin, L. Wang and B. Zhu. Exact and approximation algorithms for the complementary maximal strip recovery problem. Journal of Combinatorial Optimization, In press.
[7] L. Bulteau, G. Fertin and I. Rusu. Maximal strip recovery problem with gaps: hardness and approximation algorithms. Proceedings of the 20th Annual International Symposium on Algorithms and Computation (ISAAC'09), LNCS 5878, pages 710-719, 2009.