A Note on Minimum Flip Consensus and Maximum Compatible Subset

June 25, 2009

Philippe Gambette

Département informatique, L.I.R.M.M., C.N.R.S. - Université Montpellier II.

Abstract. Minimum Flip Consensus is a supertree method introduced in 2002 [CEFBS02, CDE+03, CEFBS06, CEFBB06, BBT08, KU08], which aims at performing the smallest number of changes (taxa deletion or addition) in a set of clusters so that it becomes compatible with a tree. There exist FPT algorithms [CEFBS02, BBT08, KU08], fixed ratio approximation schemes [CEFBS02], and heuristics [CEFBB06] to solve this problem, which can be coded as an edge-editing problem to transform a bipartite graph into an "M-free graph" [CEFBS02]. Maximum Compatible Subset removes the minimum number t of taxa so that the cluster set on the remaining taxa becomes compatible with a tree [HRB+09]. This problem can be coded as a vertex-deletion problem to transform a bipartite graph into an M-free graph.

1 Coding Maximum Compatible Subset as an M-free graph editing problem

Definition 1.1 (character graph). *Let $\mathcal{C} = \{C_1, \ldots, C_r\}$ be a set of clusters on the set X of n leaves. The* character graph *of $\mathcal{C}$ is the bipartite graph $G(\mathcal{C}) = (\mathcal{C}, X, E)$, and $E = \{\{x, C_i\} | x \in C_i\}$.*

Definition 1.2 (M-free graph property). *An M-graph is a cycle-free path of length 4. A bipartite graph $G = (X, Y, E)$ is an M-free graph if it does not have an induced M-graph whose degree-1 nodes are in Y.*

Note that an incompatibility of two clusters C_1 and C_2 of $\mathcal{C}$ forces the existence of three taxa x, y and z such that $x \in C_1 \setminus C_2$, $y \in C_1 \cap C_2$, $z \in C_2 \setminus C_1$, which corresponds exactly to the presence of an induced M-graph with three taxa vertices in the character graph of $\mathcal{C}$, as illustrated in Figure 1.

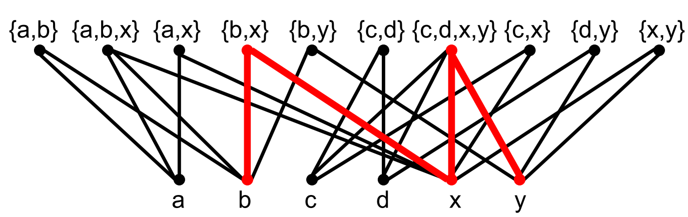

Fig. 1. The character graph of $\mathcal{C} = \{\{a, b\}, \{a, b, x\}, \{a, x\}, \{b, x\}, \{b, y\}, \{c, d\}, \{c, d, x, y\}, \{c, x\}, \{d, y\}, \{x, y\}\}$. Clusters $\{b, x\}$ and $\{c, d, x, y\}$ being incompatible, they induce an M-graph (for example the one on taxa vertices $\{b, x, y\}$ shown in bold red edges) in the character graph.

Theorem 1.1. *Given a cluster set $\mathcal{C}$, solving the MCS problem on $\mathcal{C}$ is equivalent to removing the minimum number t of vertices in the character graph $G(\mathcal{C})$ so that it becomes an M-free graph.*

It was shown in [HRB+09] that the MCS problem could be solved by an $O^*(3^t)$ FPT-algorithm. A simple bounded search tree algorithm on the induced M-graphs with 3 taxa vertices in the character graph also gives the same theoretical complexity (for each M-graph, try to delete one of its three vertices corresponding to a taxa). The problem can also be solved by applying a 3-Hitting Set FPT algorithm on the sets of 3 taxa vertices involved in the M-graphs[1].

Acknowledgments

Anthony Perez told me about the Minimum Flip Consensus problem, and Sylvain Guillemot pointed out the reduction to 3-Hitting Set.

References

BBT08. Sebastian Böcker, Quang Bao Anh Bui, and Anke Truss. An improved fixed-parameter algorithm for minimum-flip consensus trees. In *Proceedings of the Third International Workshop on Parameterized and Exact Computation (IWPEC'08)*, volume 5018 of *Lecture Notes in Computer Science*, pages 43–54, 2008. http://dx.doi.org/10.1007/978-3-540-79723-4_6.

CDE+03. D. Chen, L. Diao, O. Eulenstein, David Fernández-Baca, and M. Sanderson. Flipping: A supertree construction method. In *DIMACS Working Group Meetings on Bioconsensus*, volume 61 of *DIMACS Series in Discrete Mathematics and Theoretical Computer Science*, pages 135–160, 2003. http://books.google.fr/books?id=ARy6MpHxbw0C.

CEFBB06. D. Chen, O. Eulenstein, David Fernández-Baca, and J. Gordon Burleigh. Improved heuristics for minimum-flip supertree construction. *Evolutionary Bioinformatics*, 2:401–410, 2006. http://www.la-press.com/redirect_file.php?fileId=1083&filename=EBO-2-Chen-et-al-(2)&fileType=pdf.

CEFBS02. D. Chen, O. Eulenstein, David Fernández-Baca, and M. Sanderson. Supertrees by flipping. In *Proceedings of the eighth Annual International Computing and Combinatorics Conference (COCOON'02)*, volume 2387 of *Lecture Notes in Computer Science*, pages 128–137, 2002. http://dx.doi.org/10.1007/3-540-45655-4_42.

CEFBS06. D. Chen, O. Eulenstein, David Fernández-Baca, and M. Sanderson. Supertrees by flipping. *IEEE/ACM Transactions in Computational Biology and Bioinformatics*, 3(2):165–173, 2006. http://loco.biosci.arizona.edu/PDF/Chen.2006b.pdf.

HRB+09. Daniel H. Huson, Regula Rupp, Vincent Berry, Philippe Gambette, and Christophe Paul. Computing galled networks from real data. In *Proceedings of the 17th Annual Conference on Intelligent Systems for Molecular Biology & 8th European Conference on Computational Biology (ISMB/ECCB'09)*, volume 25(12) of *Bioinformatics*, pages i85–i93, 2009.

KU08. Christian Komusiewicz and Johannes Uhlmann. A cubic-vertex kernel for flip consensus tree. In Ramesh Hariharan, Madhavan Mukund, and V Vinay, editors, *IARCS Annual Conference on Foundations of Software Technology and Theoretical Computer Science (FSTTCS 2008)*, Dagstuhl, Germany, 2008. Schloss Dagstuhl - Leibniz-Zentrum fuer Informatik, Germany. http://drops.dagstuhl.de/opus/volltexte/2008/1760.

[1] http://fpt.wikidot.com/fpt-races gives a best complexity of $O^*(2.076^k)$ in May 2009