An (almost complete) state of the art around the Graph Motif problem

Florian Sikora
Université Paris-Est, LIGM - UMR CNRS 8049, France
florian.sikora@univ-mlv.fr

March 30, 2012

Abstract

Here is a (tentative) resume of results around the Graph Motif problem. For any mistakes, missing results, or comments on this document, please feel free to contact me !

Contents

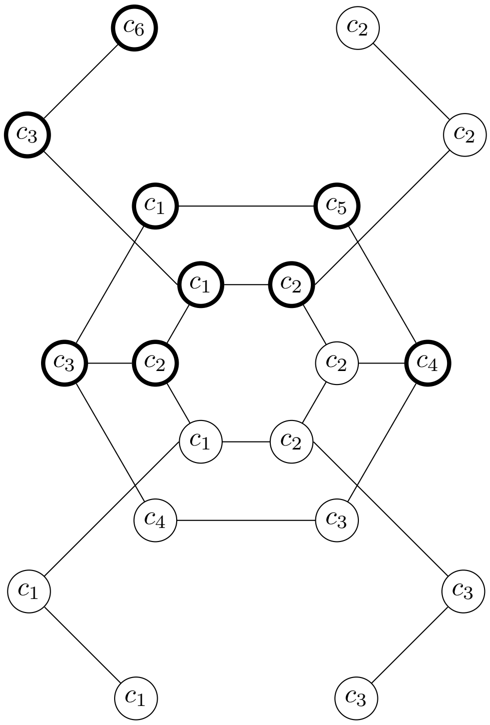

Figure 1: Toy example for the Graph Motif problem, where the graph is vertex-colored by colors $c_i, 1 \leq i \leq 6$. A possible solution for the motif $M = \{c_1, c_1, c_2, c_2, c_3, c_3, c_4, c_5, c_6\}$ is given in bold .

1 Notations

- Let $G = (V, E)$ be the vertex colored target network, $n = |V|$, $m = |E|$
- Let M be the motif. Let k be the size of the solution. Let c be number of colors in the motif.
- If $|M| = c$, then M is *colorful*. (Otherwise, it is a *multiset*)

2 The Graph Motif problem

Input : A vertex-colored graph G, a multiset of colors M.
Question : Does G have a connected subset of vertices whose multiset of colors equals M?

2.1 Network is a tree

- NP-Complete [LFS06]
- NP-Complete, even for colorful motifs and for trees of maximum degree 3. [FFHV07]

- Polynomial when the motif is colorful on caterpillars [ABH+10]
- Polynomial when the motif is colorful and each color occurs at most 2 times in G [DFV11, Sik11]
- NP-Complete for colorful motifs and rooted trees of height two [ABH+10]
- NP-Complete for colorful motifs on trees, even if a specific node (a root) is asked [ABH+10]
- W[1]-hard when the parameter is c. [FFHV07]
- No polynomial kernel on comb-graphs [ABH+10]

2.2 Network is a graph

- NP-Complete for motifs with 2 colors, even if G is bipartite with maximum degree 4[FFHV07]
- NP-Complete for colorful motifs on graphs of diameter two [ABH+10]
- Polynomial-time solvable when c is constant and G has a constant treewidth $\mathcal{O}(n^{2cw+2})$, where w is the treewidth of G [FFHV07]
- For colorful motifs : FPT : $\mathcal{O}(3^k.m)$ [BFKN08]
- For colorful motifs : FPT : $\tilde{\mathcal{O}}(2^k k^2 m)$ time and $\tilde{\mathcal{O}}(kn)$ space [GS10]
- For colorful motifs : FPT : $\mathcal{O}(3^k.m.N_{ins})$ [BHK+09]
- For colorful motifs but allows multiset in the solution : FPT : $\mathcal{O}(3^k.m.N_{ins})$ [BHK+09]
- For multiset motifs : FPT : $\mathcal{O}(87^k.k.n^2)$ [FFHV07]
- For multiset motifs : FPT : $\mathcal{O}(4.32^k.k^2.m)$ time, $\mathcal{O}(2.47^k.n)$ space [BFKN08]
- For multiset motifs: FPT : $\tilde{\mathcal{O}}(4^k k^2 m)$ time and $\tilde{\mathcal{O}}(kn)$ space [GS10]
- For multiset motifs with deletions and r insertions: FPT : $\tilde{\mathcal{O}}(4^k(k+r)^2 m)$ time and $\tilde{\mathcal{O}}((k+r)n)$ space [GS10]

3 The List-colored Graph Motif problem

A set of colors is associated to each network node

Network is a graph

- For colorful motifs but allows multiset in the solution : FPT : $\mathcal{O}(k!.3^k.m.N_{ins})$ [BHK+09]
- For multiset motifs : FPT : $\mathcal{O}(10.88^k.m)$ [BFKN08]
- For multiset motifs: FPT : $\tilde{\mathcal{O}}(4^k k^2 m)$ time and $\tilde{\mathcal{O}}(kn)$ space (implicit algorithm) [GS10]

4 The Biconnected subgraph problem

The solution must be biconnected instead of connected

- $\mathsf{W}[1]$-complete with respect to k [BFKN08]

5 The Bridge-connected subgraph problem

The solution must be bridge connected instead of connected

- $\mathsf{W}[1]$-complete with respect to k [BFKN08]

6 The Max Motif problem

Want an maximum sized connected occurrence of M is G.

Hardness results of the Graph Motif problem hold since it is a special case of the Max Motif problem. For the same reason, FPT algorithm is unlikely if the parameter is the number of deletions.

6.1 Network is a tree

- APX-Hard even when G is a tree of maximum degree 3, colorful motif and each color occurs at most twice in G [DFV09]
- Not approximable within factor $|V|^{\frac{1}{3}-\epsilon}$, for any $\epsilon > 0$, even if the motif is colorful and each color occurs at most twice in G [RS12]
- Not approximable within factor $2^{\log^{\delta} n}$, for any $\delta < 1$ (equivalent to no constant approximation ratio) even if the motif is colorful[DFV09]
- Colorful Motifs : Exponential algorithm : $\mathcal{O}(1.33^n.poly(n))$ [DFV09]
- Multiset Motifs : Exponential algorithm : $\mathcal{O}(1.62^n.poly(n))$ [DFV09]
- Multiset motifs : FPT : $\mathcal{O}(k2^k n^3 \log n)2^{\mathcal{O}(k)}$ [DFV09]

6.2 Network is a graph

- Multiset motifs : FPT : $\mathcal{O}(2^{5k} kn^2 \log^2 n)4^{\mathcal{O}}(k)$ [DFV09]
- For multiset motifs: FPT : $\tilde{\mathcal{O}}(4^k k^2 m)$ time and $\tilde{\mathcal{O}}(kn)$ space [GS10]

7 The Min Add problem

Want an occurrence of M in G with the minimum number of insertions. Equivalent to the Graph Motif problem with a bounded number of insertions.

Hardness results of the Graph Motif problem hold since it is a special case of the Min Add problem. For the same reason, FPT algorithm is unlikely if the parameter is the number of additions.

7.1 Network is a tree

- NP-hard, even with G is a tree of max degree 4, the motif is colorful and each color occurs twice in G [DFV11]

7.2 Network is a graph

- For colorful motifs : FPT : $\mathcal{O}(3^k.m.N_{ins})$ [BHK+09]
- For colorful motifs but allows multiset in the solution : FPT : $\mathcal{O}(3^k.m.N_{ins})$ [BHK+09]
- For multiset motifs with deletions and r insertions: FPT : $\tilde{\mathcal{O}}(4^k(k+r)^2m)$ time and $\tilde{\mathcal{O}}((k+r)n)$ space [GS10]

8 The Min Substitution problem

Want an occurrence of M in G with the minimum number of substitutions

Hardness results of the Graph Motif problem hold since it is a special case of the Min Substitution problem. For the same reason, FPT algorithm is unlikely if the parameter is the number of substitutions.

8.1 Network is a tree

- NP-hard, even with G is a tree of max degree 4, the motif is colorful and each color occurs twice in G [DFV11]
- Not approximable within factor $c \log |V|$, c a constant, even if the motif is colorful and G a tree of depth 2. [RS12]
- W[2]-hard when parametrized by the number of substitutions even is M is colorful [RS12]

8.2 Network is a graph

- FPT $\mathcal{O}^*((3e)^{O(k)})$ [DFV11]

9 The Min-CC problem

Want an occurrence of the motif with the minimum number of connected components in the solution

Hardness results of the Graph Motif problem hold since it is a special case of the Min-CC problem. For the same reason, FPT algorithm is unlikely if the parameter is the number of CC.

9.1 Network is a path

- Polynomial-time solvable if c is a constant in $\mathcal{O}(n^{c+4})$ [DFV07]
- W[2]-hard when parametrized by the number of connected components [BFKN08]

- APX-Hard even for colorful motifs, each color appears exactly twice in G [DFV07]

9.2 Network is a tree

- Solvable in $\mathcal{O}(n^2 2^{\frac{2n}{3}})$ [DFV07]
- FPT in $\mathcal{O}(n^2 k^{(c+1)^2+1})$ [DFV07]
- W[1]-hard when parametrized by c [FFHV07]
- W[2]-hard when parametrized by the number of connected components even is M is colorful [DFV07]
- Not approximable within $c \log n$ for a constant $c > 0$ even if M is colorful [DFV07]

9.3 Network is a graph

- For colorful motifs (search for r connected components): FPT : $\tilde{\mathcal{O}}(2^k k^2 r^2 m)$ time and $\tilde{\mathcal{O}}(krn)$ space [GS10]
- FPT by k [DFV07]
- FPT in $\mathcal{O}(|\ln(\epsilon)|.4.32^k k^2 m)$ [BFKN08]
- For multiset motifs (search for r connected components): FPT : $\tilde{\mathcal{O}}(4^k k^2 r^2 m)$ time and $\tilde{\mathcal{O}}(krn)$ space [GS10]

10 The Edge-Weighted Graph Motif problem

G is weighted on the edges.

10.1 Network is a graph

- These three results want to minimize the weight of the edge-cut between the solution and the rest of the graph.
- Multiset motifs : FPT : $\mathcal{O}(|\log(\epsilon)|2^{k+kd}.n)$, with d maximum degree of G [BRS09]
- Multiset motifs : Branch and bound algorithm : $\mathcal{O}(m\log(m) + n^b)$, with b maximum number of bounds [BRS09]
- Multiset motifs : FPT : $\mathcal{O}(m.c^k.2^\omega.(m.k.2^k.3^\omega + \omega + d))$ time, $\mathcal{O}(m.c^k.2^w)$ space, with d maximum degree of G, w treewidth of G [BRS09]
- These two results want to minimize the weight of the edges in the solution.
- For colorful motifs (sum of weight in the solution $< r$): FPT : $\tilde{\mathcal{O}}(2^k k^2 r^2 m)$ time and $\tilde{\mathcal{O}}(krn)$ space [GS10]
- For multiset motifs (sum of weight in the solution $< r$): FPT : $\tilde{\mathcal{O}}(4^k k^2 r^2 m)$ time and $\tilde{\mathcal{O}}(krn)$ space [GS10]

11 The #Graph Motif problem

Network is a graph

- For colorful motifs : FPT : $\mathcal{O}(2^k k^3 m)$ time and $\mathcal{O}(k^2 n)$ space [GS10]
- For multiset motifs : #W[1]-hard for parameter k, even with two colors in the motif [GS10]

12 The Constrained Graph Motif problem

Given a set of mandatory vertices $V_M \subseteq V$, find an occurrence of M s.t. all the mandatory vertices are in the solution.

12.1 Network is a graph

- FPT with the same complexity of the Graph Motif problem, when the size of the solution is the parameter [DFV11]
- FPT with parameter $t = |M| - |V_M|$ if G is of bounded treewidth [DFV11] $\mathcal{O}(w^{3w} n 2^{4.4427t})$, where w is the treewidth of the graph.[DFV11]
- W[2]-hard if the parameter is $t = |M| - |V_M|$, even when G is a graph of diameter 2 [DFV11]

13 The Module Graph Motif problem

The solution is G must be a graph module (instead of being simply connected)

13.1 Network is a forest

- NP-Complete oven if the motif is colorful and G is a collection of paths of length 3 [RS12]

13.2 Network is a graph

- Polynomial if the module must be strong [RS12]
- FPT in $\mathcal{O}(2^k |V|^2)$ time and $\mathcal{O}(2^k |V|)$ space, where k is the size of the solution [RS12]
- FPT if parameterized by $(k, |C|)$, where C is the set of colors [RS12]
- FPT if a list of colors is given for each node [RS12]

14 Softwares

- Motus [LFS06] `http://genome.crg.es/~vlacroix/motus/`

- Torque [BHK+09]. `http://www.cs.tau.ac.il/~bnet/torque.html`. Internet server, dynamic programming + Integer Linear Programming (CPLEX). Only colorful motifs. Allow insertions and deletions. Allow list colored graph motif. Only one solution.
- GraMoFoNe [BSV10]. `http://igm.univ-mlv.fr/AlgoB/gramofone/`. Cytoscape java plugin. Pseudo Boolean Programming. Multiset motifs, allow insertions and deletions. Allow list colored graph motif.

References

[ABH+10] Abhimanyu M Ambalath, Radheshyam Balasundaram, Chintan Rao H, Venkata Koppula, Neeldhara Misra, Geevarghese Philip, and M S Ramanujan. On the kernelization complexity of colorful motifs. Accepted at the 5th International Symposium on Parameterized and Exact Computation (IPEC 2010), 2010.

[BFKN08] N. Betzler, M.R. Fellows, C. Komusiewicz, and R. Niedermeier. Parameterized algorithms and hardness results for some graph motif problems. In *Proc. 19th Annual Symposium on Combinatorial Pattern Matching (CPM), Pisa, Italy*, volume 5029 of *Lecture Notes in Computer Science*, pages 31–43. Springer, 2008.

[BHK+09] S. Bruckner, F. Hüffner, R.M. Karp, R. Shamir, and R. Sharan. Topology-free querying of protein interaction networks. In *Proc. 13th Annual International Conference on Computational Molecular Biology (RECOMB), Tucson, USA*, page 74. Springer, 2009.

[BRS09] S. Böcker, F. Rasche, and T. Steijger. Annotating Fragmentation Patterns. In *WABI 2009*, volume 5724 of *Lecture Notes in Bioinformatics*, pages 13–24. Springer, 2009.

[BSV10] Guillaume Blin, Florian Sikora, and Stéphane Vialette. GraMoFoNe: a Cytoscape plugin for querying motifs without topology in Protein-Protein Interactions networks. In Hisham Al-Mubaid, editor, *2nd International Conference on Bioinformatics and Computational Biology (BICoB'10)*, pages 38–43, Honolulu, USA, March 2010. International Society for Computers and their Applications (ISCA).

[DFV07] R. Dondi, G. Fertin, and S. Vialette. Weak pattern matching in colored graphs: Minimizing the number of connected components. In *Proc.10th Italian Conference on Theoretical Computer Science (ICTCS), Roma, Italy*, pages 27–38. World-Scientific, 2007.

[DFV09] R. Dondi, G. Fertin, and S. Vialette. Maximum motif problem in vertex-colored graphs. In G. Kucherov and E. Ukkonen, editors, *Proc. 20th Annual Symposium on Combinatorial Pattern Matching (CPM'09), Lille, France*, volume 5577 of *Lecture Notes in Computer Science*, pages 221–235, 2009.

[DFV11] Riccardo Dondi, Guillaume Fertin, and Stéphane Vialette. Finding Approximate and Constrained Motifs in Graphs. In Raffaele Giancarlo and Giovanni Manzini, editors, *Proceedings of the 22nd Annual Symposium on Combinatorial Pattern Matching (CPM)*, volume 6661 of *Lecture Notes in Computer Science*, pages 388–401. Springer, 2011.

[FFHV07] M. Fellows, G. Fertin, D. Hermelin, and S. Vialette. Sharp tractability borderlines for finding connected motifs in vertex-colored graphs. In *Proc. 34th International Colloquium on Automata, Languages and Programming (ICALP), Wroclaw, Poland*, volume 4596 of *Lecture Notes in Computer Science*, pages 340–351. Springer, 2007.

[GS10] Sylvain Guillemot and Florian Sikora. Finding and counting vertex-colored subtrees. In Petr Hlinený and Antonín Kucera, editors, *35th International Symposium on Mathematical Foundations of Computer Science (MFCS'10)*, volume 6281 of *LNCS*, pages 405–416, Brno, Czech Republic, August 2010. Springer.

[LFS06] V. Lacroix, C.G. Fernandes, and M.-F. Sagot. Motif search in graphs: application to metabolic networks. *IEEE/ACM Transactions on Computational Biology and Bioinformatics (TCBB)*, 3(4):360–368, 2006.

[RS12] Romeo Rizzi and Florian Sikora. Some results on more flexible versions of Graph Motif. In *Proc. of the 7th International Computer Science Symposium in Russia (CSR 2012)*, Lecture Notes in Computer Science. Springer, 2012. To appear.

[Sik11] Florian Sikora. *Aspects algorithmiques de la comparaison d'éléments biologiques.* PhD thesis, Université Paris-Est, September 2011. In French.