Maximum Motif Problem in Vertex-Colored Graphs

Riccardo Dondi [1] **Guillaume Fertin** [2] **Stéphane Vialette** [3]

[1]Università di Bergamo, [2]Université de Nantes, [3]Université Paris-Est

CPM 09, June 23, 2009

Outline

Outline

Metabolic Network Analysis

Metabolic network analysis:

- interactions between chemical compounds
- decomposition of metabolic networks into independent functional parts
- understand the modularity and organization in the network

A functional motif is a subnetwork representing a given set of functionalities

GRAPH MOTIF

`[Lacroix, Fernandes, Sagot, TCBB 2006]` introduced a graph-based model:

- vertices $\longrightarrow$ reactions
- edges connect successive reactions
- colors of vertices $\longrightarrow$ reaction types

Motif: a multiset of colors

GRAPH MOTIF - Combinatorial Problem

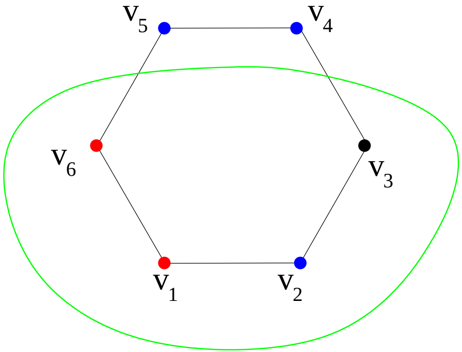
Problem (GRAPH MOTIF)

Input: *a target vertex-colored graph G and a colored motif $\mathcal{M}$ (a multiset of colors).*
Output: *a connected component $G' = (V', E')$ of G so that $\mathcal{C}(V') = \mathcal{M}$.*

G' is an occurrence of $\mathcal{M}$ in G

When $\mathcal{M}$ is a set $\rightarrow$ $\mathcal{M}$ is colorful

GRAPH MOTIF



$$\mathcal{M} = \{blue, red, red, black\}$$

GRAPH MOTIF - Computational Complexity

GRAPH MOTIF is **NP**-complete:

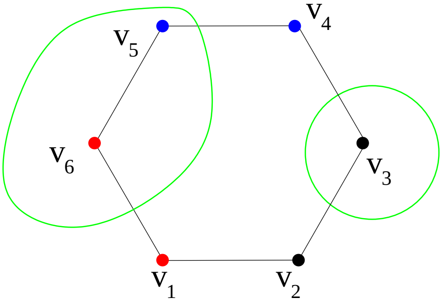
- even if $\mathcal{M}$ is colorful and the target graph G is a tree with maximum degree 3 [Fellows, Fertin, Hermelin, Vialette, ICALP 2007]
- even if $\mathcal{M}$ consists of two colors, and the target graph is bipartite with maximum degree 4 [Fellows, Fertin, Hermelin, Vialette, ICALP 2007]

GRAPH MOTIF - Parameterized Complexity

- FPT algorithms for GRAPH MOTIF, when parameterized by $k = |\mathcal{M}|$:
  - time complexity $\mathcal{O}(87^k k\, poly(n))$ [Fellows, Fertin, Hermelin, Vialette, ICALP 2007]
  - improved to $\mathcal{O}(4.32^k k^2 poly(n))$ [Betzler, Fellows, Komusiewicz, Niedermeier, CPM 2008]
- GRAPH MOTIF is **W[1]**-hard, when parameterized by the number of colors in $\mathcal{M}$ and the target graph is a tree [Fellows, Fertin, Hermelin, Vialette, ICALP 2007]

Outline

Extensions of GRAPH MOTIF



$$\mathcal{M} = \{blue, red, black\}$$

Due to measurement errors: the GRAPH MOTIF problem too stringent $\rightarrow$ no occurrence of $\mathcal{M}$ in G

Extensions of GRAPH MOTIF

Relaxations of the definition of occurrence of $\mathcal{M}$:

- connectivity of an occurrence $\Rightarrow$ MIN-CC [Dondi, Fertin, Vialette, ICTCS 2007]: minimize the number of connected components of the occurrence of $\mathcal{M}$
- colors of an occurrence $\Rightarrow$ MAXIMUM MOTIF

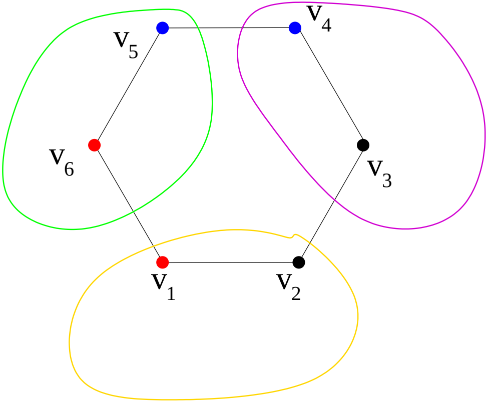
Outline

MAXIMUM MOTIF

Problem (MAXIMUM MOTIF)

Input: *a target vertex-colored graph G and a colored motif* $\mathcal{M}$.
Output: *a connected subgraph* $G' = (V', E')$ *of G of maximum cardinality, so that* $\mathcal{C}(V') \subseteq \mathcal{M}$.

MAXIMUM MOTIF



$$\mathcal{M} = \{blue, red, black\}$$

MAXIMUM MOTIF - Approximation Complexity

Theorem

MAXIMUM MOTIF *is* **APX**-*hard, even when the target graph is a tree T of degree* 3*,* $\mathcal{M}$ *is colorful, each color has at most* 2 *occurrences in* T.

Proof.

L-reduction from Max Independet Set on cubic graphs □

MAXIMUM MOTIF - Approximation Complexity

Theorem

For any constant $\delta < 1$, MAXIMUM MOTIF *cannot be approximated within ratio* $2^{\log^{\delta} n}$ *unless* $\mathbf{NP} \subseteq \mathbf{DTIME}[2^{\mathrm{poly}\log n}]$, *even when the target graph is a tree and* $\mathcal{M}$ *is colorful.*

Proof.

Self-Improvement □

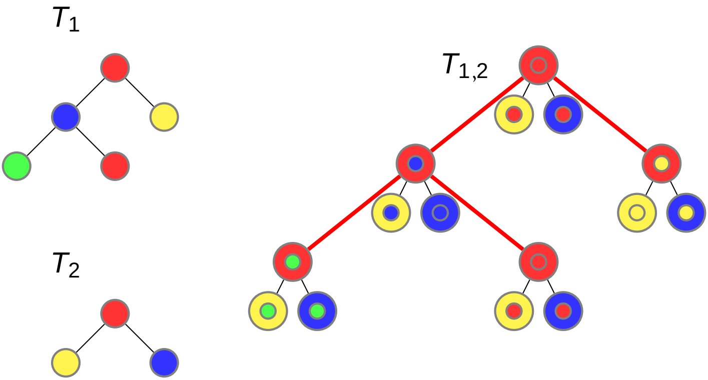
Maximum Motif - Self-Improvement

Definition

Let $I_1 = (\mathcal{M}_1, T_1)$, $I_2 = (\mathcal{M}_2, T_2)$ be two istances of Maximum Motif, the product $I_1 \times I_2 = (T_{1,2}, \mathcal{M}_{1,2})$

- $T_{1,2}$ is a vertex-colored tree
- $\mathcal{M}_{1,2} = \{c_{a,b} : c_a \in \mathcal{M}_1, c_b \in \mathcal{M}_2\}$

MAXIMUM MOTIF - Self-Improvement



MAXIMUM MOTIF - Self-Improvement

Self-improvement:

- Self-product $I^k = I^{k-1} \times I$

Lemma

Let T_k be a solution for MAXIMUM MOTIF *over instance I^k, then we can compute in polynomial time a solution T for* MAXIMUM MOTIF *over instance I, so that $|T|^k \geq |T_k|$*

Lemma

Let T be a solution for MAXIMUM MOTIF *over instance I, then there exists a solution T_k for* MAXIMUM MOTIF *over instance I^k, so that $|T_k| \geq |T|^k$*

MAXIMUM MOTIF - Parameterized Complexity

Lemma

MAXIMUM MOTIF *is in* **FPT** *when parameterized by the size of the solution.*

Two fixed-parameter algorithms based on color-coding technique:

- when the target graph is a tree: time complexity $\mathcal{O}(k2^k n^3 \log n)2^{\mathcal{O}(k)}$
- for general graphs: time complexity $\mathcal{O}(k2^{5k} n^2 \log n)4^{\mathcal{O}(k)}$

MAXIMUM MOTIF- Parameterized Algorithms

Perfect Hash Family - PHF

Definition

A family $\mathcal{F}$ of functions from S to $\{1, \dots, k\}$ is perfect if for any subset $S' \subseteq S$ of k elements, there is an $f \in \mathcal{F}$ which is one-to-one on S'.

$|\mathcal{F}| = \mathcal{O}(2^{\mathcal{O}(k)} \log n)$ [`Alon,Yuster,Zwick, J. ACM 1995`]

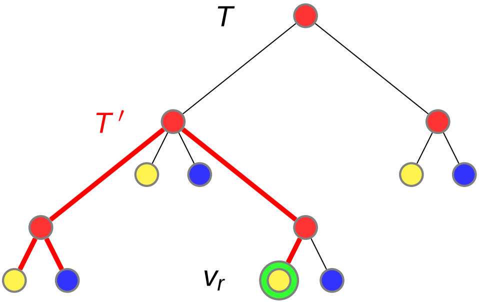
MAXIMUM MOTIF - Parameterized Algorithms

When the target graph tree is a tree T:

Definition

Let T' be a subtree of T, then a vertex $v \in T'$ is defined as the rightmost vertex of T' iff:

- v has no child in T';
- there is no vertex in T' which is a right sibling of a vertex on the path from r to v



v_r is the rightmost vertex of T'

MAXIMUM MOTIF - Parameterized Algorithms

Fixed-parameter algorithm when the target graph tree is a tree T:

1. PHF from $\mathcal{M}$ to $\{1, \ldots, k\}$
2. dynamic programming to compute

$$P_r[v, K]$$

where v is the rightmost vertex of a tree $T' = (V', E')$ and $\mathcal{C}(V')$ is labeled by set $K \subseteq \{1, \ldots, k\}$

MAXIMUM MOTIF - Parameterized Algorithms

For general graphs:

1. combining two PHFs:
   - PHF from $\mathcal{M}$ to $\{1, \ldots, k\}$
   - PHF from V to $\{1, \ldots, k\}$
2. dynamic programming similar to [Fellows, Fertin, Hermelin, Vialette, ICALP 2007]

MAXIMUM MOTIF - Exponential-time Algorithms

Theorem

MAXIMUM MOTIF *for trees of size n can be solved in* $\mathcal{O}(1.62^n\, poly(n))$ *time. In case* $\mathcal{M}$ *is colorful, the time complexity reduces to* $\mathcal{O}(1.33^n\, poly(n))$.

Exact branch-and-bound algorithms

Outline

1. Intorduction
   - Biological Motivations
   - GRAPH MOTIF
2. Extensions of GRAPH MOTIF
3. MAXIMUM MOTIF
   - Approximation Complexity
   - Parameterized Complexity
   - Exponential-time Algorithms
4. Conclusion

Future Directions

Future directions:

- More efficient fixed-parameter algorithms
- Efficient heuristics
- New formulations of the problem

Conclusion

Conclusion:

- GRAPH MOTIF and biological motivations
- Extensions of GRAPH MOTIF $\Rightarrow$ MAXIMUM MOTIF
- Hardness results for MAXIMUM MOTIF
- MAXIMUM MOTIF is in FPT
- Exponential-time algorithms for MAXIMUM MOTIF
- Future directions