The Graph Motif problem parameterized by the structure of its input graph

Édouard Bonnet and Florian Sikora

Institute for Computer Science and Control, Hungarian Academy of Sciences, Budapest, Hungary (MTA SZTAKI)

September 18, 2015
Graph Motif - what is known?
Given a graph $G = (V, E)$ whose vertices are colored by a function $c : V \rightarrow \mathcal{C}$ and a multiset $M$ over $\mathcal{C}$ . . .
Find a connected subgraph colored by

\[ M = \{ \text{yellow, orange, red, pink, blue, gray, black, white, cyan, purple, green, blue, gray, black, white} \} \].
Find a connected subgraph colored by

\[ M = \{ \text{\textcolor{yellow}{yellow}}, \text{\textcolor{orange}{orange}}, \text{\textcolor{red}{red}}, \text{\textcolor{pink}{pink}}, \text{\textcolor{blue}{blue}}, \text{\textcolor{green}{green}}, \text{\textcolor{gray}{gray}}, \text{\textcolor{black}{black}} \}. \]
Find a **connected** subgraph colored by

\[ M = \{\text{\textcolor{yellow}{yellow}}, \text{\textcolor{orange}{orange}}, \text{\textcolor{red}{red}}, \text{\textcolor{purple}{purple}}, \text{\textcolor{black}{black}}, \text{\textcolor{gray}{gray}}, \text{\textcolor{lightgray}{lightgray}}, \text{\textcolor{white}{white}}, \text{\textcolor{pink}{pink}}, \text{\textcolor{green}{green}}\}. \]
Motivations

Lacroix et al. (2005): reaction networks.

Social, technical networks, and mass spectrometry.

Graph pattern matching with only connectivity constraint.
Known results - algorithms

- FPT in the size of the motif: $O^*(2^{|M|})$ [BKK '12, PZ '12].
- FPT in the neighborhood diversity: $O^*(2^{nd})$ [G '12].
- XP in $\text{tw}(G) + |C|$ [FFHV '11].
- Polytime solvable in caterpillar trees [ABHKMPR '10].
Known results - hardness

- $W[1]$-hard on trees w.r.t. the number of colors [FFHV ’11].
- NP-hard on bipartite graphs of degree 4, $|C| = 2$ [FFHV ’11].
- NP-hard on trees of diameter 4 [ABHKMPR ’10].
- NP-hard on comb graphs [CPPW ’12].
Neighborhood diversity

- Least number of subsets in a partition into true or false twins.
- $\kappa$ has linear neighborhood diversity if $\forall G$, $\text{nd}(G) \leq O(\kappa(G))$.
- $\kappa$ has exponential n.d. if $\forall G$, $\text{nd}(G) \leq 2^{O(\kappa(G))}$.
- $\kappa$ has unbounded n.d. if $\forall f$, $\exists G$ such that $\text{nd}(G) > f(\kappa(G))$. 
Neighborhood diversity

- Least number of subsets in a partition into true or false twins.
- $\kappa$ has linear neighborhood diversity if $\forall G$, $\text{nd}(G) \leq O(\kappa(G))$.
- $\kappa$ has exponential n.d. if $\forall G$, $\text{nd}(G) \leq 2^{O(\kappa(G))}$.
- $\kappa$ has unbounded n.d. if $\forall f$, $\exists G$ such that $\text{nd}(G) > f(\kappa(G))$.

Vertex cover number has exponential n.d. since $\text{nd}(G) \leq v\text{c} + 2^{v\text{c}}$. 
Neighborhood diversity

- Least number of subsets in a partition into true or false twins.
- $\kappa$ has linear neighborhood diversity if $\forall G$, $\text{nd}(G) \leq O(\kappa(G))$.
- $\kappa$ has exponential n.d. if $\forall G$, $\text{nd}(G) \leq 2^{O(\kappa(G))}$.
- $\kappa$ has unbounded n.d. if $\forall f$, $\exists G$ such that $\text{nd}(G) > f(\kappa(G))$.

Cluster editing number on connected graphs has
Neighborhood diversity

- Least number of subsets in a partition into true or false twins.
- $\kappa$ has linear neighborhood diversity if $\forall G$, $nd(G) \in O(\kappa(G))$.
- $\kappa$ has exponential n.d. if $\forall G$, $nd(G) \in 2^{O(\kappa(G))}$.
- $\kappa$ has unbounded n.d. if $\forall f$, $\exists G$ such that $nd(G) > f(\kappa(G))$.

Cluster editing number on connected graphs has linear n.d.: $nd(G) \in 3k + 1$. 
Neighborhood diversity

- Least number of subsets in a partition into true or false twins.
- $\kappa$ has linear neighborhood diversity if $\forall G, \text{nd}(G) \leq O(\kappa(G))$.
- $\kappa$ has exponential n.d. if $\forall G, \text{nd}(G) \leq 2^{O(\kappa(G))}$.
- $\kappa$ has unbounded n.d. if $\forall f, \exists G$ such that $\text{nd}(G) > f(\kappa(G))$.

Distance to co-cluster has
Neighborhood diversity

- Least number of subsets in a partition into true or false twins.
- $\kappa$ has linear neighborhood diversity if $\forall G$, $\text{nd}(G) \leq O(\kappa(G))$.
- $\kappa$ has exponential n.d. if $\forall G$, $\text{nd}(G) \leq 2^{O(\kappa(G))}$.
- $\kappa$ has unbounded n.d. if $\forall f$, $\exists G$ such that $\text{nd}(G) > f(\kappa(G))$.

Distance to co-cluster has unbounded n.d.
Aim of the paper

Completing the picture for structural/secondary parameters.

- Does only boundedness of neighborhood diversity count?
- For exponential n.d.: from double to single-exponential time.
Ecological Landscape

Graph Motif - what is known?  
- Bounded neighborhood diversity
- Unbounded neighborhood diversity
- Hardness

- NP-hard with constant parameter values
- FPT but no poly-kernel
  - if NP $\not\subset$ coNP/poly
  - Distance to clique •
  - Distance to co-cluster ◊
  - Distance to clique Cover •
  - Distance to co-cluster Cover ◊
- Max Ind. Set
- Min Dominating Set ◊
  - Distance to perfect

- FPT
  - Vertex Cover •
  - Distance to cluster ◊
  - Distance to disjoint paths ◊
  - Feedback Vertex Set #
  - Distance to chordal
  - Distance to bipartite
  - Distance to cluster

- W[1]-hard, in W[P]
  - Cluster Editing *
  - Max leaf # ◊
  - Feedback edge set # *
  - Bandwidth ◊
  - Pathwidth
  - Max Degree
  - Treewidth
  - h-index
  - Degeneracy

- Chromatic #
- Average Degree
Bounded neighborhood diversity
Parameter: distance to clique

**Theorem**

Graph Motif *can be solved in* $O^*(4^k)$. 
Graph Motif - what is known? Bounded neighborhood diversity Unbounded neighborhood diversity Hardness

$C_1 \leq k$

$M' = M - \{\ldots\}$

Set-colored Set Cover with threshold constraints imposed by $M'$. Solvable in $O^*(m^2 n)$ for $m$ sets and $n$ elements.
Set-colored Set Cover with threshold constraints imposed by $M'$.

Solvable in $O^*(m^2 n)$ for $m$ sets and $n$ elements.

$M' = M - \{\text{yellow, red, red, red, pink, pink, red, green, green, blue, blue, gray, black}\}$
Set-colored Set Cover with threshold constraints imposed by $M'$. Solvable in $O^*(m2^n)$ for $m$ sets and $n$ elements.
Time $O^*(4^k)$ should be improvable to $O^*(2^k)$ but not further:

**Observation**

*Under SCH, Graph Motif cannot be solved in $O^*((2 - \varepsilon)^k)$.***

**SCH:** For any $\varepsilon > 0$, Set Cover cannot be solved in $O((2 - \varepsilon)^n)$.  


Parameter: vertex cover number

Theorem

Graph Motif can be solved in $O^*(2^{2k} \log k)$.

Let’s start as the previous algorithm.

Set-colored Set Cover with threshold constraints and connected intersection graph of the solution.
Parameter: vertex cover number

**Theorem**

*Graph Motif* can be solved in $O^*(2^{2k \log k})$.

- Let’s start as the previous algorithm.
- Guess an ordered partition of the connected components $C_i$s which says how the $C_i$s connect with each other via the IS.
Parameter: vertex cover number

**Theorem**

**Graph Motif** *can be solved in* \( O^*(2^{2k \log k}) \).

- Let’s start as the previous algorithm.
- Guess an ordered partition of the connected components \( C_i \)’s which says how the \( C_i \)’s connect with each other via the IS.
- Maximum matching in an auxiliary bipartite graph.
Again, you may expect to go down to $O^*(2^k)$ but not lower:

**Observation**

*Under SETH, Graph Motif cannot be solved in $O^*((2 - \varepsilon)^k)$.*

SETH: For any $\varepsilon > 0$, SAT cannot be solved in $O((2 - \varepsilon)^n)$.
Unbounded neighborhood diversity
Parameter: distance to co-cluster

Theorem

**Graph Motif** can be solved in $O^*(2^{2k \log k})$.

$H = V - S = I_1 \cup \ldots \cup I_q$ is a co-cluster.

Fix a solution $R$. 

Parameter: distance to co-cluster

**Theorem**

**Graph Motif** can be solved in $O^*(2^{2k \log k})$.

$H = V - S = I_1 \cup \ldots \cup I_q$ is a co-cluster.

Fix a solution $R$.

- case a) $R$ intersects only one $I_j$, solve $S \cup I_j$ vertex cover $k$. 

---

**Graph Motif - what is known?**

Bounded neighborhood diversity

Unbounded neighborhood diversity

Hardness
Parameter: distance to co-cluster

Theorem

Graph Motif can be solved in $O^*(2^{2k \log k})$.

$H = V - S = I_1 \cup \ldots \cup I_q$ is a co-cluster.

Fix a solution $R$.

1. case a) $R$ intersects only one $I_j$, solve $S \cup I_j \Rightarrow$ vertex cover $k$.
2. case b) $R$ intersects two $I_j$s, cliquify $H \Rightarrow$ distance to clique $k$. 
<table>
<thead>
<tr>
<th>Graph Motif - what is known?</th>
<th>Bounded neighborhood diversity</th>
<th>Unbounded neighborhood diversity</th>
<th>Hardness</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td>Hardness</td>
</tr>
</tbody>
</table>

**Hardness**
**Theorem**

**Graph Motif** is \(W[1]\)-hard for parameter max leaf number.
Theorem

**Graph Motif** is $W[1]$-hard for parameter max leaf number.

In fact,

**Theorem**

**Graph Motif** is $W[1]$-hard for parameter number of leaves of the graph + number of colors in subdivisions of stars.
Graph Motif - what is known?  Bounded neighborhood diversity  Unbounded neighborhood diversity  Hardness

\[ M = \{ \bullet, \} \]
$M = \{ \bullet, \, \bullet \times \text{blah}, \, \bullet \times \text{blah}, \}$
Reduction from \( k \)-Multicolored Clique on \( \left( \bigcup_{1 \leq i \leq k} H_i, E \right) \) where \( H_i = \{ u_{i,1}, \ldots, u_{i,t} \} \).
Reduction from \textit{k-Multicolored Clique} on \((\bigcup_{i \in k} H_i, E)\)
where \(H_i = \{u_{i,1}, \ldots, u_{i,t}\}\).

Pairs \((u_{i,a}, u_{j,b})\) and \([t^2 - 1]\) are in one-to-one correspondence.
Reduction from \textit{k-Multicolored Clique} on \((\bigcup_{1 \leq i \leq k} H_i, E)\) where \(H_i = \{u_{i,1}, \ldots, u_{i,t}\}\).

List of edges: \(\{13, 15, 20, 29, \ldots, 80, 81, 92, 97\}\) \((t = 10)\).
Reduction from $k$-Multicolored Clique on $(\bigcup_{1 \leq i \leq k} H_i, E)$ where $H_i = \{u_{i,1}, \ldots, u_{i,t}\}$.

List of edges: \{13, 15, 20, 29, \ldots, 80, 81, 92, 97\} ($t = 10$). Complements to $t^2$: \{3, 8, 19, 20, \ldots, 71, 80, 85, 87\}. 
Reduction from \( k \)-Multicolored Clique on \( (\bigcup_{i=1}^{k} H_i, E) \)
where \( H_i = \{u_{i,1}, \ldots, u_{i,t}\} \).

List of edges: \{13, 15, 20, 29, \ldots, 80, 81, 92, 97\} (\( t = 10 \)).
Complements to \( t^2 \): \{3, 8, 19, 20, \ldots, 71, 80, 85, 87\}.
Consecutive differences: \{3, 5, 11, 1, \ldots, 9, 5, 2\}.
Reduction from $k$-Multicolored Clique on $(\bigcup_{i\in[k]} H_i, E)$ where $H_i = \{u_{i,1}, \ldots, u_{i,t}\}$.

List of edges: $\{13, 15, 20, 29, \ldots, 80, 81, 92, 97\}$ ($t = 10$).
Complements to $t^2$: $\{3, 8, 19, 20, \ldots, 71, 80, 85, 87\}$.
Consecutive differences: $\{3, 5, 11, 1, \ldots, 9, 5, 2\}$.

$M = \{\bullet, \bullet \times \text{blah}, \circ \times \text{blah}, t^2 \times ij\}$
Theorem

Graph Motif is solvable in $O^*(16^{ml} n^{10ml}) = n^{O(ml)}$.

Observation

There is less than $4ml$ vertices of degree at least 3 and removing those vertices leaves a disjoint union of at most $5ml$ paths.

The previous reduction was only ruling out $n^{o(\sqrt{ml})}$ assuming ETH.
Theorem

**Graph Motif** is solvable in $O^*(16^{ml} n^{10ml}) = n^{O(ml)}$.

Observation

There is less than $4ml$ vertices of degree at least 3 and removing those vertices leaves a disjoint union of at most $5ml$ paths.

The previous reduction was only ruling out $n^{o(\sqrt{ml})}$ assuming ETH.

Partitioned Subgraph Isomorphism $\rightarrow$ Graph Motif

$n^{o((ml(G)+|C|)/ \log (ml(G)+|C|))}$ ETH-based lower bound.
Perspectives

- Settle the $2^{O(vc)}$ vs no $2^{o(vc \log vc)}$ under ETH.
- A $2^{O(vc)}$ algorithm would immediately give a single-exponential for parameter distance to co-cluster and edge clique cover number (where the edge clique cover is given with the input).
- Extend the FPT algorithms to the list-colored variant.