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

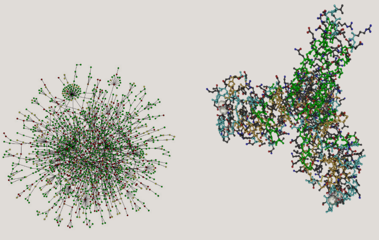
Édouard Bonnet and Florian Sikora

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

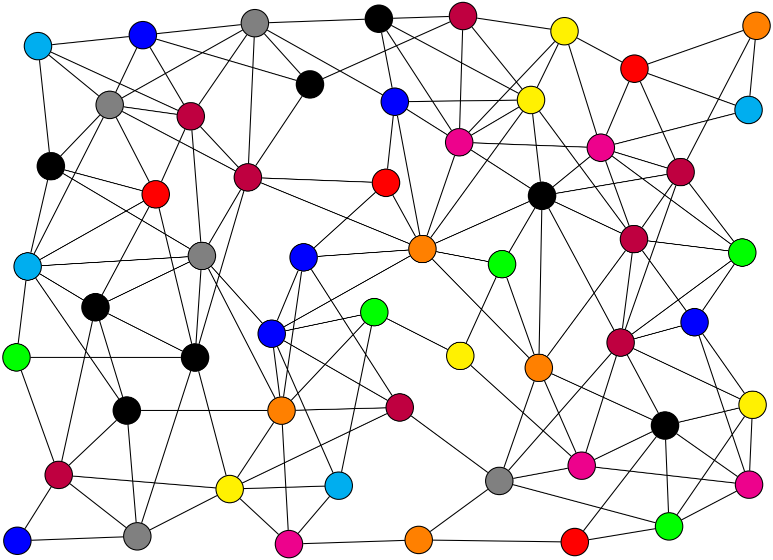
September 18, 2015

Graph Motif - what is known?

Graph Motif

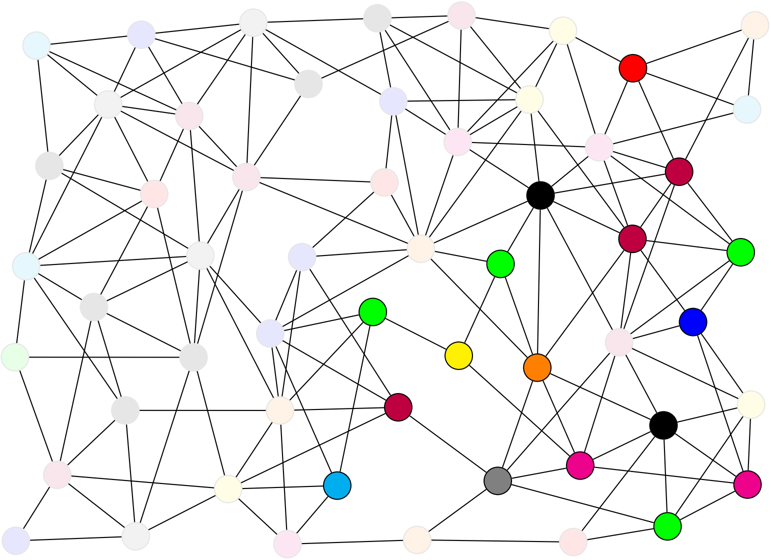


Given a graph $G = (V, E)$ whose vertices are colored by a function $c : V \to \mathcal{C}$ and a multiset M over $\mathcal{C}$ ...



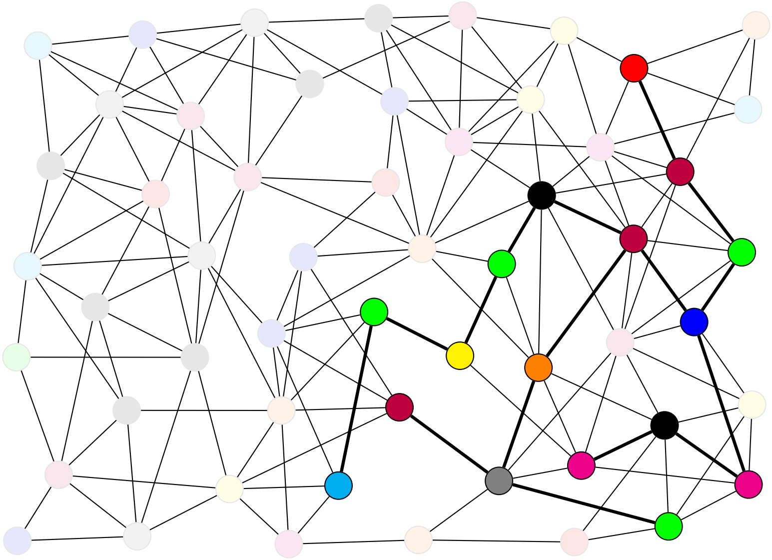
Find a connected subgraph colored by

$M = \{\bullet, \bullet, \bullet, \bullet, \bullet, \bullet, \bullet, \bullet, \bullet, \bullet, \bullet, \bullet, \bullet, \bullet, \bullet, \bullet, \bullet\}$.



Find a connected subgraph **colored by**

$M = \{$●, ●, ●, ●, ●, ●, ●, ●, ●, ●, ●, ●, ●, ●, ●, ●, ●$\}$.



Find a **connected** subgraph colored by

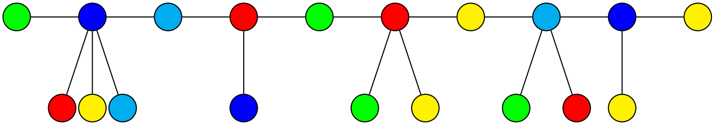
$M = \{$●, ●, ●, ●, ●, ●, ●, ●, ●, ●, ●, ●, ●, ●, ●, ●, ●$\}$.

Motivations

- Lacroix et al. (2005): reaction networks.
- Social, technical networks, and mass spectrometry.
- Graph pattern matching with only connectivity constraint.

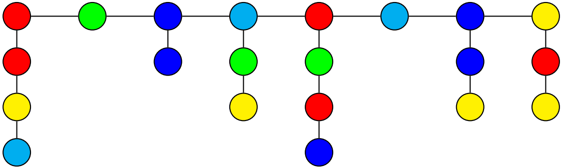
Known results - algorithms

- FPT in the size of the motif: $O^*(2^{|M|})$ [BKK '12,PZ '12].
- FPT in the neighborhood diversity: $O^*(2^{\mathrm{nd}})$ [G '12].
- XP in $\mathrm{tw}(G) + |\mathcal{C}|$ [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, $|\mathcal{C}| = 2$ [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.
- κ has linear neighborhood diversity if $\forall G$, $\mathrm{nd}(G) \leqslant O(\kappa(G))$.
- κ has exponential n.d. if $\forall G$, $\mathrm{nd}(G) \leqslant 2^{O(\kappa(G))}$.
- κ has unbounded n.d. if $\forall f$, $\exists G$ such that $\mathrm{nd}(G) > f(\kappa(G))$.

Neighborhood diversity

- Least number of subsets in a partition into true or false twins.
- κ has linear neighborhood diversity if $\forall G$, $\text{nd}(G) \leqslant O(\kappa(G))$.
- κ has exponential n.d. if $\forall G$, $\text{nd}(G) \leqslant 2^{O(\kappa(G))}$.
- κ 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 $\text{nd}(G) \leqslant \text{vc} + 2^{\text{vc}}$.

Neighborhood diversity

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

Cluster editing number on connected graphs has

Neighborhood diversity

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

Cluster editing number on connected graphs has linear n.d.: $\text{nd}(G) \leqslant 3k + 1$.

Neighborhood diversity

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

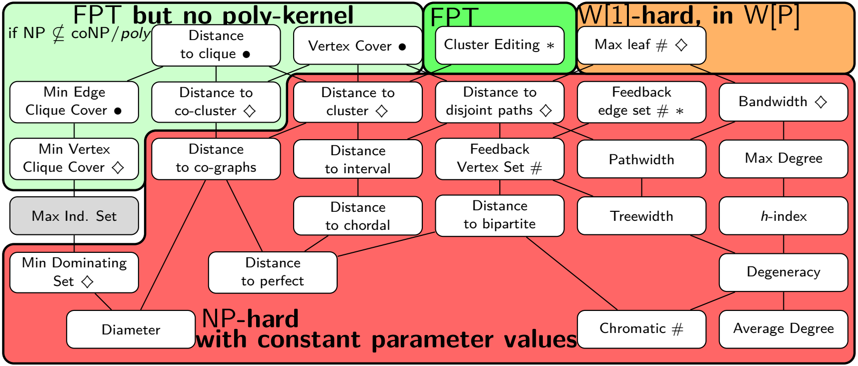
Distance to co-cluster has unbounded n.d.

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

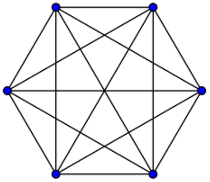


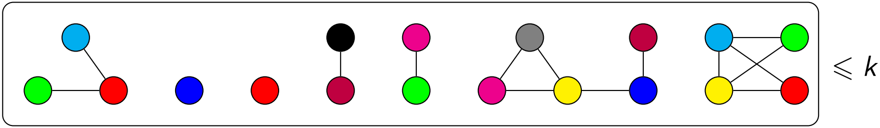
Bounded neighborhood diversity

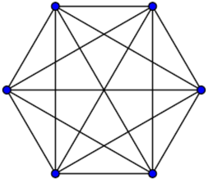
Parameter: distance to clique

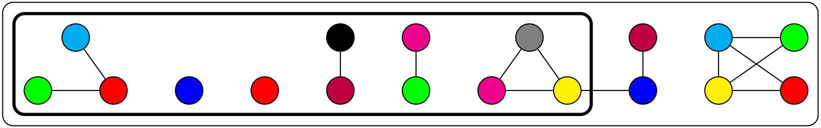
Theorem

GRAPH MOTIF *can be solved in* $O^*(4^k)$.

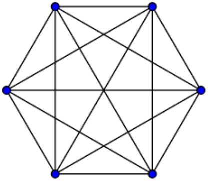


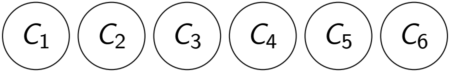

$\leqslant k$





$M' = M -$ 





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* $O^*((2-\varepsilon)^k)$*.*

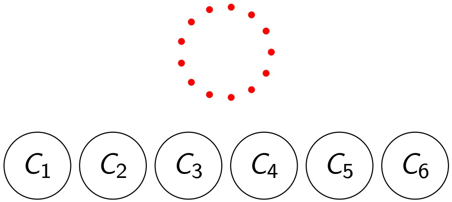
SCH: For any $\varepsilon > 0$, Set Cover cannot be solved in $O((2-\varepsilon)^n)$.

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_is
- which says how the C_is 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_is
- which says how the C_is 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 $S \cup I_j \rightsquigarrow$ vertex cover k.

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 $S \cup I_j \rightsquigarrow$ vertex cover k.
- case b) R intersects two I_js, cliquify $H \rightsquigarrow$ distance to clique k.

Hardness

Theorem

GRAPH MOTIF *is W[1]-hard for parameter max leaf number.*

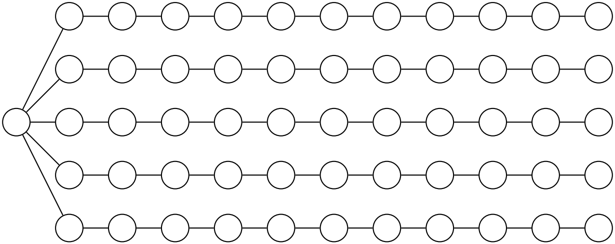
Theorem

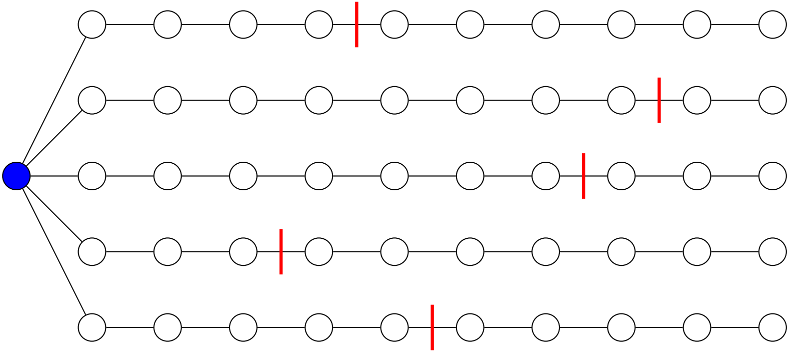
GRAPH MOTIF *is W[1]-hard for parameter max leaf number.*

In fact,

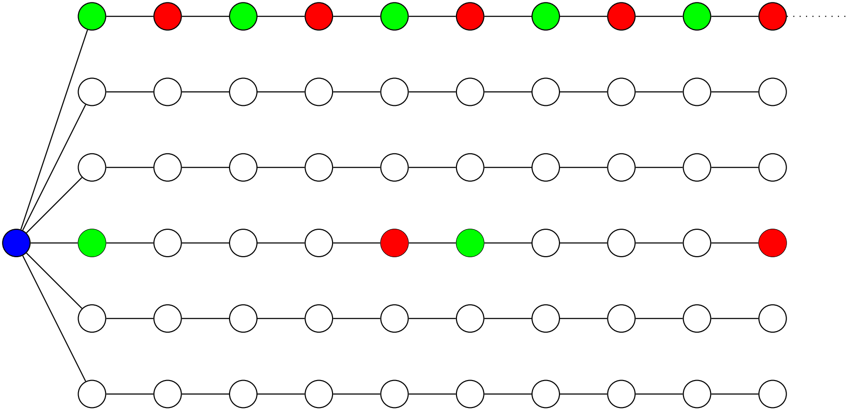
Theorem

GRAPH MOTIF *is W[1]-hard for parameter number of leaves of the graph + number of colors in subdivisions of stars.*



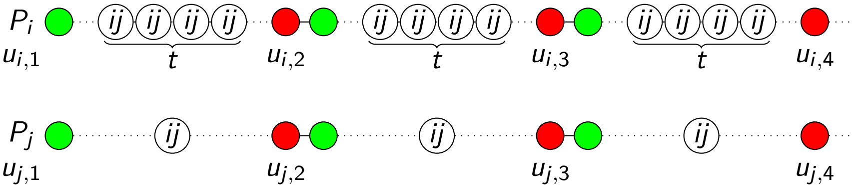


$M = \{$ ,

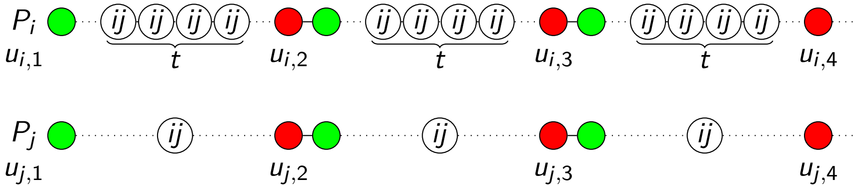


$M = \{$●, ● $\times$ blah, ● $\times$ blah,

Reduction from k-MULTICOLORED CLIQUE on $(\cup_{1 \leqslant i \leqslant k} H_i, E)$ where $H_i = \{u_{i,1}, \ldots, u_{i,t}\}$.

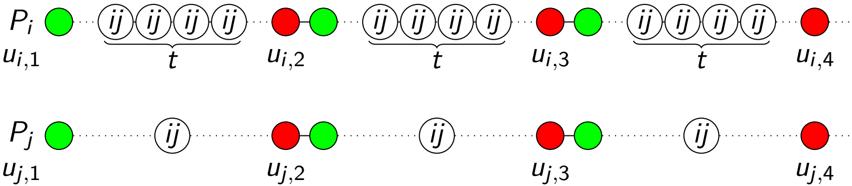


Reduction from k-Multicolored Clique on $(\cup_{1 \leqslant i \leqslant k} H_i, E)$ where $H_i = \{u_{i,1}, \ldots, u_{i,t}\}$.



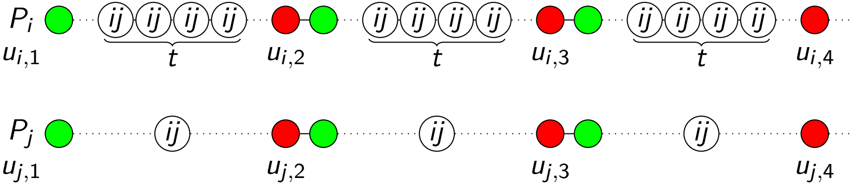
Pairs $(u_{i,a}, u_{j,b})$ and $[t^2 - 1]$ are in one-to-one correspondence.

Reduction from k-Multicolored Clique on $(\cup_{1 \leqslant i \leqslant k} H_i, E)$ where $H_i = \{u_{i,1}, \ldots, u_{i,t}\}$.



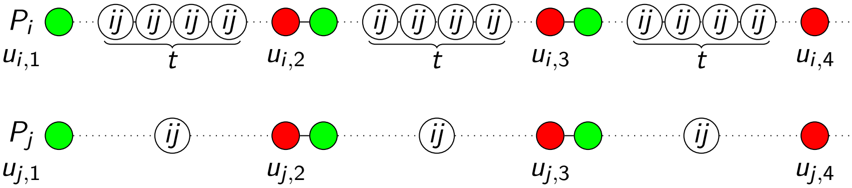
List of edges: $\{13, 15, 20, 29, \ldots, 80, 81, 92, 97\}$ $(t = 10)$.

Reduction from k-MULTICOLORED CLIQUE on $(\cup_{1 \leqslant i \leqslant k} H_i, E)$ where $H_i = \{u_{i,1}, \ldots, u_{i,t}\}$.



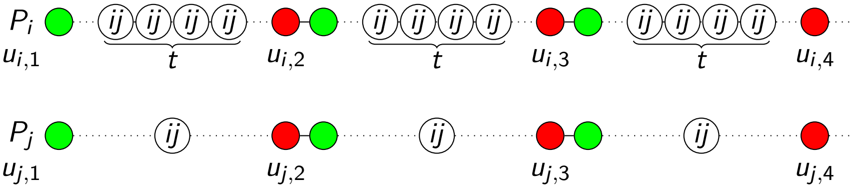
List of edges: $\{13, 15, 20, 29, \ldots, 80, 81, 92, 97\}$ $(t = 10)$.
Complements to t^2: $\{3, 8, 19, 20, \ldots, 71, 80, 85, 87\}$.

Reduction from k-Multicolored Clique on $(\cup_{1 \leqslant i \leqslant k} H_i, E)$ where $H_i = \{u_{i,1}, \ldots, u_{i,t}\}$.

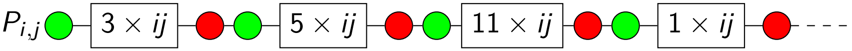


List of edges: $\{13, 15, 20, 29, \ldots, 80, 81, 92, 97\}$ $(t = 10)$.
Complements to t^2: $\{3, 8, 19, 20, \ldots, 71, 80, 85, 87\}$.
Consecutive differences: $\{3, 5, 11, 1, \ldots, 9, 5, 2\}$.

Reduction from k-Multicolored Clique on $(\cup_{1 \leqslant i \leqslant k} H_i, E)$ where $H_i = \{u_{i,1}, \dots, u_{i,t}\}$.



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



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

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 $n^{o(\sqrt{\mathrm{ml}})}$ 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 $n^{o(\sqrt{\mathrm{ml}})}$ assuming ETH.

PARTITIONED SUBGRAPH ISOMORPHISM $\rightarrow$ GRAPH MOTIF
$\rightsquigarrow$ $n^{o((\mathrm{ml}(G)+|\mathcal{C}|)/\log(\mathrm{ml}(G)+|\mathcal{C}|))}$ ETH-based lower bound.

Perspectives

- Settle the $2^{O(\mathrm{vc})}$ vs no $2^{o(\mathrm{vc} \log \mathrm{vc})}$ under ETH.
- A $2^{O(\mathrm{vc})}$ 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.