Parameterized Algorithms and Hardness Results for Some Graph Motif Problems

Nadja Betzler[1] Michael R. Fellows[2]
Christian Komusiewicz[1] Rolf Niedermeier[1]

[1]Friedrich-Schiller-Universität Jena, Germany

[2]University of Newcastle, Australia

Motifs in biological networks

A **motif** in a network is an interesting **subnetwork**. The two main parts of a motif are:

- Components/Vertices
- Topology

The actual definitions differ depending on context:

1. Regulatory networks.

A motif is a small, repeated, and maybe evolutionary conserved subnetwork. [Shen-Orr et al., Nat. Genet. 2002]

This makes assumptions **only about** the **topology** but not about the single **components** of the subnetworks.

Usual algorithmic task: Identifying subnetworks that occur more often than expected in random networks.

Motifs in biological networks

2. Querying of protein interaction networks.

In network querying the goal is to find subnetworks in a given network that are similar to the **query**.

[Sharan & Ideker, Nat. Biotech. 2006]

Here, we make assumptions about the topology **and** the single components.

Typical examples:

- Signalling path queries
- Bounded treewidth queries
- Molecular complex queries

Motifs in biological networks

3. Querying of metabolic networks.

The Graph Motif problem is motivated by motif search in **metabolic reaction graphs**:

- Vertices $\hat{=}$ reactions
- Edges connect successive reactions

For metabolic networks assumptions about the topology of a motif seem less appropriate [Lacroix et al., WABI '05]:

- The static metabolic network that is queried differs from the active metabolism in a living cell.
- Similar topologies give rise to very different functions.

$\rightsquigarrow$ We aim to make **minimal** assumptions about the topology of the motif. We only demand that the components somehow interact.

Graph motif

We model metabolic networks as undirected vertex-colored graphs. The color of a vertex describes its function. A **motif** is a collection of functions:

Definition

A **motif** M is a multi-set of colors. If M is a **set** of colors, then we call M **colorful**.

Imposing **minimal** constraints on the topology of the motif means that we only demand **connectedness**.

Definition

An **occurrence** of a motif M in a vertex-colored graph $G = (V, E)$ is a vertex set $S \subseteq V$, such that $G[S]$ is connected and S has the colors of M.

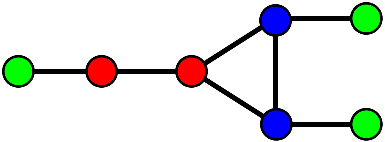
Graph motif problem

Graph Motif

Input: A vertex-colored graph $G = (V, E)$ and a motif M.
Question: Is there a vertex set $S \subseteq V$ such that S is an occurrence of M?

An example...

$M =<$ red, blue, blue, green $>$



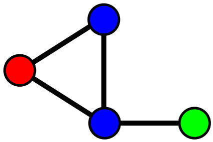
Graph motif problem

GRAPH MOTIF

Input: A vertex-colored graph $G = (V, E)$ and a motif M.
Question: Is there a vertex set $S \subseteq V$ such that S is an occurrence of M?

An example...

$M = <$ red, blue, blue, green $>$



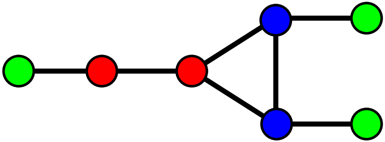
Graph motif problem

GRAPH MOTIF

Input: A vertex-colored graph $G = (V, E)$ and a motif M.
Question: Is there a vertex set $S \subseteq V$ such that S is an occurrence of M?

An example...

$M = <$ red, blue, blue, green $>$



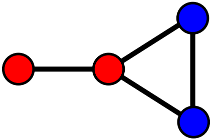
Graph motif problem

Graph Motif

Input: A vertex-colored graph $G = (V, E)$ and a motif M.
Question: Is there a vertex set $S \subseteq V$ such that S is an occurrence of M?

An example...

$M = <$ red, blue, blue, green $>$



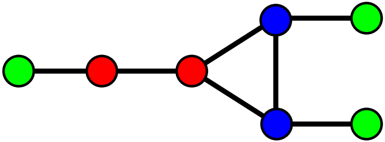
Graph motif problem

Graph Motif

Input: A vertex-colored graph $G = (V, E)$ and a motif M.
Question: Is there a vertex set $S \subseteq V$ such that S is an occurrence of M?

An example...

$M =<$ red, blue, blue, green $>$



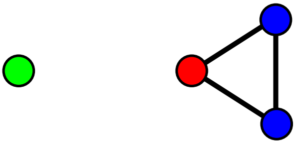
Graph motif problem

Graph Motif

Input: A vertex-colored graph $G = (V, E)$ and a motif M.
Question: Is there a vertex set $S \subseteq V$ such that S is an occurrence of M?

An example...

$M = <$ red, blue, blue, green $>$



NP-hardness

The GRAPH MOTIF problem is NP-hard

- on trees [Lacroix et al., WABI '05]
- on trees of maximum degree 3 even if M is colorful [Fellows et al., ICALP '07]
- on bipartite graphs of maximum degree 3 even if M contains only two colors [Fellows et al., ICALP '07]

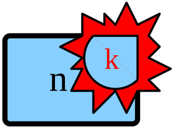
Our approach to deal with the NP-hardness of GRAPH MOTIF: Fixed-parameter algorithmics.

Fixed-parameter tractability

Idea: Problems are specified by size of instance **and** an auxiliary parameter (e.g. size of solution).
Aim: Confine combinatorial explosion to the parameter.

Fixed-parameter tractability:

A problem is called **fixed-parameter tractable** with respect to parameter k if there is an algorithm that solves the problem and has running time $f(k) \cdot poly(n)$.



Parameterized complexity

Completeness program developed by Downey and Fellows [Downey & Fellows 1999]

$$\text{FPT} \subseteq \overbrace{\text{W[1]} \subseteq \text{W[2]} \subseteq \ldots \subseteq \text{W[P]}}^{\text{Presumably fixed-parameter intractable}}$$

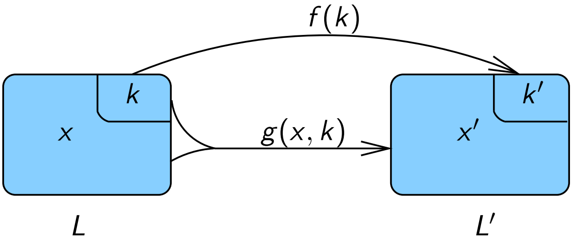
Assumption: FPT $\neq$ W[1]

If W[1]=FPT then 3-SAT for a Boolean formula F with n variables can be solved in $2^{o(n)} \cdot |F|^{O(1)}$ time.

Parameterized reduction

Let L be a parameterized problem.
With a parameterized reduction we reduce L to a problem L' such that

1. $(x, k) \in L$ iff $(x', k') \in L'$.
2. $x' = g(x, k)$ and $k' = f(k)$.



We can show parameterized hardness by reducing from W[1]-hard problems.

Fixed-parameter results

There is an algorithm that solves GRAPH MOTIF in $O(87^k \cdot k \cdot |V|^2)$ time, where $k = |M|$. [Fellows et al., ICALP '07]
$\rightsquigarrow$ GRAPH MOTIF is fixed-parameter tractable with respect to the motif size $k := |M|$.

There is a parameterized reduction from the W[1]-hard CLIQUE problem to GRAPH MOTIF in trees with $f(k)$ colors. [Fellows et al., ICALP '07]
$\rightsquigarrow$ GRAPH MOTIF is W[1]-hard on trees with respect to the parameter "number of colors in M".

Our contributions:

1. Improvement of running time to $O(4.32 \cdot k^2 \cdot |E|)$ for the parameter motif size.
2. Proof of hardness for some variants of GRAPH MOTIF.

Colorful motifs

GRAPH MOTIF

Input: A vertex-colored graph $G = (V, E)$, and a motif M, with $|M| = k$.
Question: Is there a vertex set $S \subseteq V$ such that S is an occurrence of M?

Idea of our algorithm: use the vertex-colors of the graph.

Observation: The problem becomes easier for colorful motifs.

Lemma

Let (G, M, k) be an instance of GRAPH MOTIF such that M is colorful. Then a graph motif of size k can be found in $O(3^k \cdot |E|)$ time.

Colorful motifs

Proof of Lemma...

We can use dynamic programming, storing solutions for the problem of finding occurrences of subsets of M.

The table has entries $D_{v,C}$, where $v \in V$ and $C \subseteq M$ and

$D_{v,C} = 0$ if there is an occurrence S of C with $v \in S$,
$D_{v,C} > 0$ otherwise.

We compute $D_{v,C}$ for increasing $C \subseteq M$. The graph G contains an occurrence of M iff $D_{v,M} = 0$ for some $v \in V$.

Colorful motifs

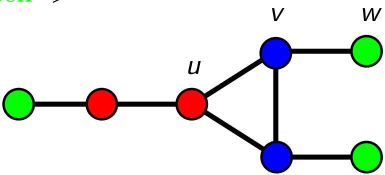
Proof of Lemma...

We initialize the table with

$$D_{v,C} = 0 \text{ if } C = \{\text{col}(v)\}, \text{ and}$$
$$D_{v,C} = 1 \text{ otherwise,}$$

where col(v) is the color of v.

$M =<$ red, blue, green $>$



Colorful motifs

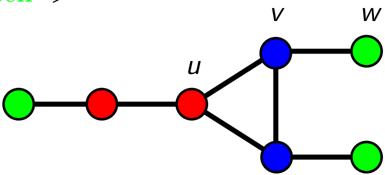
Proof of Lemma...

We initialize the table with

$D_{v,C} = 0$ if $C = \{\text{col}(v)\}$, and
$D_{v,C} = 1$ otherwise,

where $\text{col}(v)$ is the color of v.

$M = <\text{red}, \text{blue}, \text{green} >$



$D_{u,\{red\}} = 0,\ D_{v,\{blue\}} = 0,\ D_{w,\{green\}} = 0$ **but** $D_{u,\{blue\}} = 1$

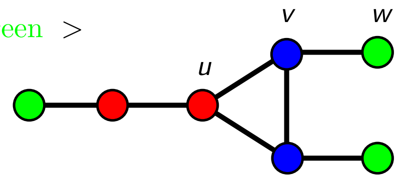
Colorful motifs

Proof of Lemma...

In the recurrence we combine solutions to solved subproblems:

$$D_{v,C} = \min_{u \in N(v),\ C' \subset C} \left\{ \begin{array}{l} D_{u, C \setminus \{\text{col}(v)\}}, \\ D_{v, C' \cup \{\text{col}(v)\}} + D_{v, (C \setminus C') \cup \{\text{col}(v)\}} \end{array} \right\}.$$

$M =<$ red, blue, green $>$



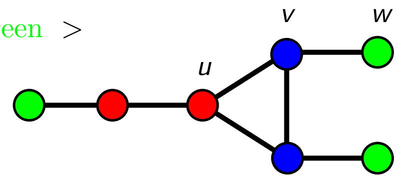
Colorful motifs

Proof of Lemma...

In the recurrence we combine solutions to solved subproblems:

$$D_{v,C} = \min_{u \in N(v),\ C' \subset C} \left\{ \begin{array}{l} D_{u, C \setminus \{\mathrm{col}(v)\}}, \\ D_{v, C' \cup \{\mathrm{col}(v)\}} + D_{v, (C \setminus C') \cup \{\mathrm{col}(v)\}} \end{array} \right\}.$$

$M =< \text{red}, \text{blue}, \text{green} >$



$$D_{v,\{red,blue\}} = D_{u,\{red\}} = 0,\ D_{v,\{green,blue\}} = D_{w,\{green\}} = 0$$

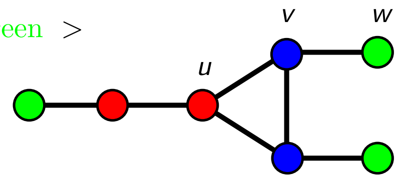
Colorful motifs

Proof of Lemma...

In the recurrence we combine solutions to solved subproblems:

$$D_{v,C} = \min_{u \in N(v),\ C' \subset C} \left\{ \begin{array}{l} D_{u, C \setminus \{\mathrm{col}(v)\}}, \\ D_{v, C' \cup \{\mathrm{col}(v)\}} + D_{v, (C \setminus C') \cup \{\mathrm{col}(v)\}} \end{array} \right\}.$$

$M = <$ red, blue, green $>$



$D_{v,\{red,blue\}} = D_{u,\{red\}} = 0,\ D_{v,\{green,blue\}} = D_{w,\{green\}} = 0$
$D_{v,\{red,blue,green\}} = D_{v,\{red,blue\}} + D_{v,\{green,blue\}} = 0$

Colorful motifs

Proof of Lemma...

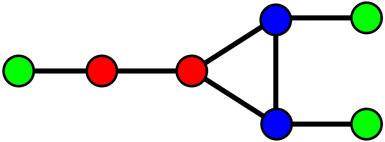
Correctness:
We only combine solutions with different colors
$\rightsquigarrow$ No vertex is used twice.
We only combine solutions between neighbors
$\rightsquigarrow$ The occurrence is connected.

Running time:
We must compute $D_{v,C}$ for all possible subsets of M. For each subset C, we must consider all possibilities to partition C.
$\rightsquigarrow$ 3^k subset combinations.
For each combination and each vertex $v \in V$, we scan once through its adjacency list
$\rightsquigarrow$ $O(|E|)$ time for each combination.

Color coding

Problem for non-colorful motifs: we cannot distinguish vertices by colors. $\rightsquigarrow$ We **randomly** color the **multiple** colors: For a color c that appears $\#(c)$ times in the motif, we introduce $\#(c)$ subcolors.

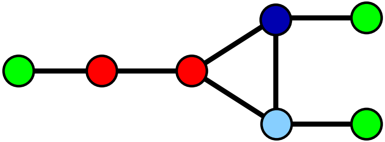
$M =<$ red, blue, blue, green $>$



Color coding

Problem for non-colorful motifs: we cannot distinguish vertices by colors. $\rightsquigarrow$ We **randomly** color the **multiple** colors:
For a color c that appears $\#(c)$ times in the motif, we introduce $\#(c)$ subcolors.

$M =<$ red, light-blue, dark-blue, green $>$



In case the random coloring creates a **colorful** coloring of the occurrence, we can apply the dynamic programming algorithm.

Color coding

Number of trials:
The probability of achieving a colorful coloring of an occurrence of M is $> e^{-k}$.
$\rightsquigarrow$ To achieve an error probability of at most ϵ we have to perform $|\ln(\epsilon)| \cdot O(e^k)$ trials.

Overall running time:
For $|\ln(\epsilon)| \cdot O(e^k)$ trials and $O(3^k \cdot |E|)$ time per trial, we achieve a running time of $O(|\ln(\epsilon)| \cdot (e \cdot 3)^k \cdot |E|) = O(|\ln(\epsilon)| \cdot (8.16)^k \cdot |E|)$.

Running time

Further speedup:
Applying two known speed-up techniques we can further improve the running time.

1. Fast subset convolution [Björklund et al., STOC '07]:
This reduces the running time of the dynamic programming from $O(3^k \cdot |E|)$ to $O(2^k \cdot k^2 \cdot |E|)$.

2. Using more colors [Hüffner et al., APBC '07]:
With $1.3 \cdot \#(c)$ subcolors we achieve a probability of $O(1.76^{-k})$ for a colorful coloring.

Theorem

A graph motif of size k can be found with probability $1 - \epsilon$ in $O(|\ln(\epsilon)| \cdot 4.32^k \cdot k^2 \cdot |E|)$ time.

Fixed-parameter tractable variants

Our algorithm can be adapted to solve two variants of the Graph Motif problem:

In Min-CC-Graph Motif we do not demand connectedness, but try to minimize the number of connected components of the occurrence of M.

Theorem

A min-cc-graph motif of size k can be found with probability $1-\epsilon$ in $O(|\ln(\epsilon)| \cdot 4.32^k \cdot k^2 \cdot |E|)$ time.

In List-colored Graph Motif we allow multiple colors per vertex in G.

Theorem

A graph motif of size k can be found in list-colored graphs with probability $1-\epsilon$ in $O(|\ln(\epsilon)| \cdot 10.88^k \cdot k^2 \cdot |E|)$ time.

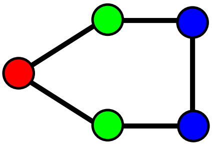
Biconnected motifs

So far: **connectivity**.

Now: **biconnectivity**.

Definition

A graph G is **biconnected** if for each pair of vertices $u, v \in V$, there are 2 vertex disjoint paths from u to v



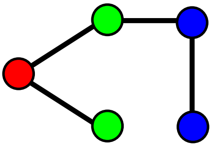
Biconnected motifs

So far: **connectivity**.

Now: **biconnectivity**.

Definition

A graph G is **biconnected** if for each pair of vertices $u, v \in V$, there are 2 vertex disjoint paths from u to v



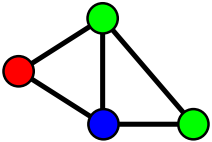
Biconnected graph motif problem

Biconnected Graph Motif

Input: A vertex-colored graph $G = (V, E)$, and a Motif M, with $|M| = k$.
Question: Is there a vertex set $S \subseteq V$ such that $G[S]$ is biconnected and an occurrence of M?

An example...

$M =<$ red, blue, green $>$



Biconnected graph motif problem

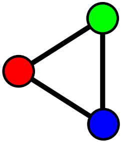
Biconnected Graph Motif

Input: A vertex-colored graph $G = (V, E)$, and a Motif M, with $|M| = k$.

Question: Is there a vertex set $S \subseteq V$ such that $G[S]$ is biconnected and an occurrence of M?

An example...

$M =<$ red, blue, green $>$



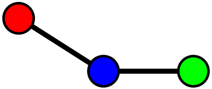
Biconnected graph motif problem

Biconnected Graph Motif

Input: A vertex-colored graph $G = (V, E)$, and a Motif M, with $|M| = k$.
Question: Is there a vertex set $S \subseteq V$ such that $G[S]$ is biconnected and an occurrence of M?

An example...

$M =<$ red, blue, green $>$



W[1]-hardness

Theorem

BICONNECTED GRAPH MOTIF is W[1]-hard with respect to the parameter motif size $k := |M|$.

We show the hardness by a parameterized reduction from CLIQUE.

CLIQUE

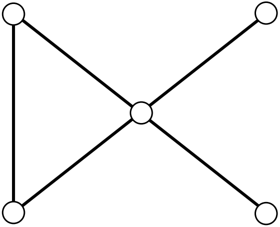
Input: A graph $G = (V, E)$, and an integer $k > 0$.
Question: Is there a vertex set $S \subseteq V$ of size k, such that $G[S]$ is a complete graph?

Theorem [Downey & Fellows, SIAM J. Comput. 1995]

CLIQUE is W[1]-hard.

Parameterized reduction

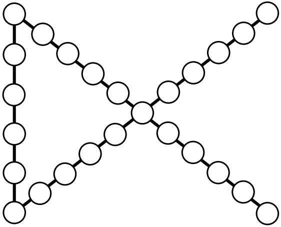
The following is a simple reduction from CLIQUE to the special case of BICONNECTED GRAPH MOTIF, where M has only one color.



Let (G, k) be an instance of CLIQUE. Then...

Parameterized reduction

The following is a simple reduction from CLIQUE to the special case of BICONNECTED GRAPH MOTIF, where M has only one color.



Let (G, k) be an instance of CLIQUE. Then...
...replace every edge of G with a path of length $\binom{k}{2} + 1$.

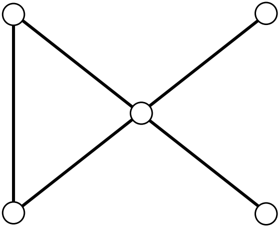
W.l.o.g. assume $k \geq 3$.
Claim: Let G' be the modified graph. Then G has a clique of size k iff G' has a biconnected subgraph of size $k' = k + \binom{k}{2}(\binom{k}{2} + 1)$.

Parameterized reduction

Claim: Let G' be the modified graph. Then G has a clique of size k iff G' has a biconnected subgraph of size $k' = k + \binom{k}{2}(\binom{k}{2} + 1)$.

Idea of proof...

"$\Rightarrow$" Let G contain a clique S of size k. Then $G[S]$ is biconnected and has $\binom{k}{2}$ edges. The corresponding subgraph in G' thus has $k + \binom{k}{2}(\binom{k}{2} + 1)$ vertices.

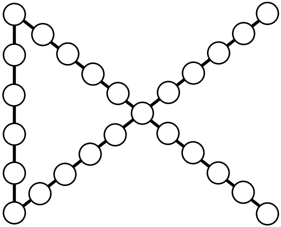


Parameterized reduction

Claim: Let G' be the modified graph. Then G has a clique of size k iff G' has a biconnected subgraph of size $k' = k + \binom{k}{2}(\binom{k}{2} + 1)$.

Idea of proof...

"$\Rightarrow$" Let G contain a clique S of size k. Then $G[S]$ is biconnected and has $\binom{k}{2}$ edges. The corresponding subgraph in G' thus has $k + \binom{k}{2}(\binom{k}{2} + 1)$ vertices.



Parameterized reduction

Claim: Let G' be the modified graph. Then G has a clique of size k iff G' has a biconnected subgraph of size $k' = k + \binom{k}{2}(\binom{k}{2} + 1)$.

Idea of proof...

"$\Leftarrow$" Let $G'[S]$ be a biconnected subgraph of size k'.

1. Each constructed path must be completely included, otherwise we violate the biconnectivity demand. $\rightsquigarrow$
Therefore, $k' := a + b \cdot (\binom{k}{2} + 1)$, where a corresponds to vertices in G and b to edges in G.

2. If $a = k$, then G has a clique of size k.

3. If $a \neq k$, then $G'[S]$ is not biconnected.

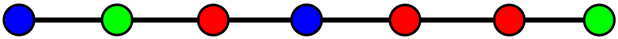
Min-CC-graph motif on paths

Min-CC-Graph Motif

Input: A vertex-colored path $G = (V, E)$, a motif M, and an integer k.
Question: Is there a vertex set $S \subseteq V$ such that $G[S]$ exactly has the colors of M, and induces $\leq k$ connected components?

An example...

$M =<$ red, blue, blue, green, green $>$, $k = 3$



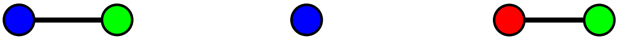
Min-CC-graph motif on paths

Min-CC-Graph Motif

Input: A vertex-colored path $G = (V, E)$, a motif M, and an integer k.
Question: Is there a vertex set $S \subseteq V$ such that $G[S]$ exactly has the colors of M, and induces $\leq k$ connected components?

An example...

$M =<$ red, blue, blue, green, green $>$, $k = 3$



Min-CC-graph motif on paths

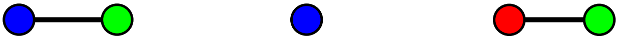
Min-CC-Graph Motif

Input: A vertex-colored path $G = (V, E)$, a motif M, and an integer k.

Question: Is there a vertex set $S \subseteq V$ such that $G[S]$ exactly has the colors of M, and induces $\leq k$ connected components?

An example...

$M =<$ red, blue, blue, green, green $>$, $k = 3$



Theorem [Dondi et al., ICTCS '07]

Min-CC-Graph-Motif on paths is NP-hard.

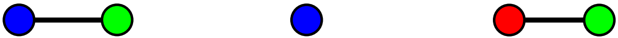
Min-CC-graph motif on paths

MIN-CC-GRAPH MOTIF

Input: A vertex-colored path $G = (V, E)$, a motif M, and an integer k.
Question: Is there a vertex set $S \subseteq V$ such that $G[S]$ exactly has the colors of M, and induces $\leq k$ connected components?

An example...

$M =<$ red, blue, blue, green, green $>, k = 3$



Theorem

MIN-CC-GRAPH-MOTIF on paths is W[1]-hard with respect to the parameter "number of components".

Summary of Results

1. Improvement of running time for Graph Motif and Min-CC-Graph Motif from $O(87^k \cdot k \cdot |V|^2)$ to $O(4.32^k \cdot k^2 \cdot |E|)$.

2. An $O(10.88^k \cdot k^2 \cdot |E|)$ algorithm for List-Colored Graph Motif.

3. Introduction of Biconnected Graph Motif. Proof of W[1]-hardness with respect to the parameter motif size.

4. Proof of W[1]-hardness of Min-CC-Graph Motif on paths with respect to the number of components of the occurrence.

Further result: Proof of W[1]-hardness for higher connectivity demands with respect to the parameter motif size.

Outlook

Interesting open tasks & questions...

1. Application to real world data.
Color coding with increased number of colors has practical potential [Hüffner et al., APBC '07].
Open: efficiency of the fast subset convolution technique.

2. Further exploration of the tractability landscape of motif search in graphs.
For example:

- Which models of "motif occurrence" lead to easy, which to hard problems?
- What are alternative parameterizations for BICONNECTED GRAPH MOTIF?