Busca de padrões (motifs) em redes biológicas

Eloi Araujo

FACOM – UFMS

12 de julho de 2017

Nosso grupo

UFMS

UFABC

UFRN

Universität Bielefeld

Introdução

Problema geral - Estrutura de dados

Motif multiconjunto de cores

Motif topológico-colorido

Problemas estudados nesse trabalho

Esse trabalho

Diego Rubert
Eloi Araujo
Fábio Martinez
Jens Stoye
Marco A Stefanes

Áreas de estudo - aplicação em Biologia

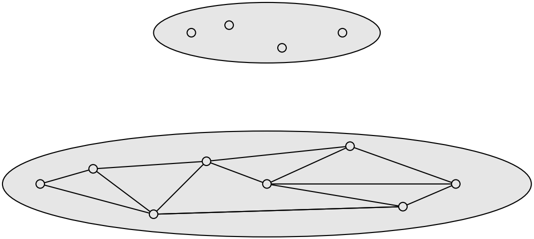
Teoria da computação

1. complexidade
2. algoritmos exatos, aproximados, FPT, paralelos

Aplicações práticas

1. heurísticas
2. implementações

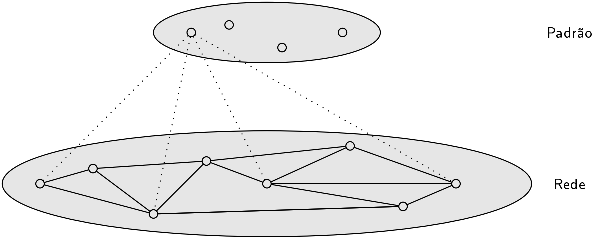
Problema Geral



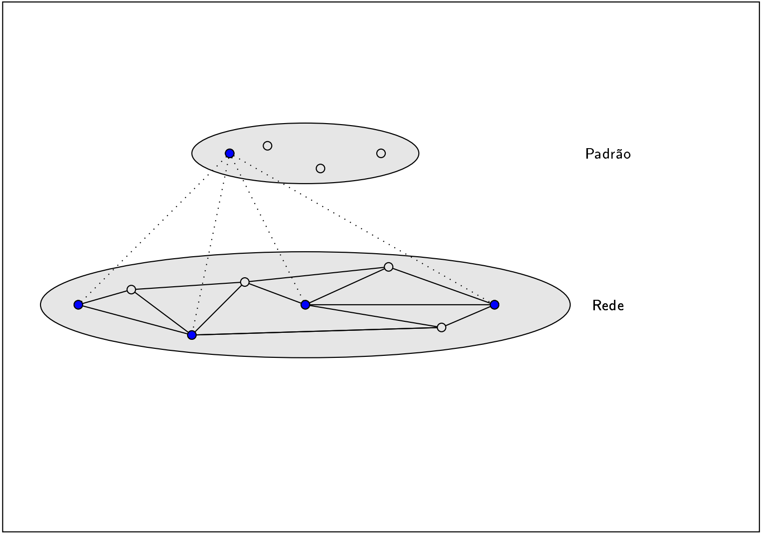
Padrão

Rede

Problema Geral



Problema Geral



Rede biológica - Representação

$\mathcal{C}$: conjunto de cores.

1. Grafo lista-de-vértices colorido $G = (V_G, E_G, \mathcal{L} : V_G \rightarrow 2^{\mathcal{C}})$
2. Grafo vértice-colorido $|\mathcal{L}(u)| = 1$ para todo vértice u da rede.

o grafo $G = (V_G, E_G)$ pode ser orientado ou não

Motif – Representação

é um padrão M que aparece com alta frequência.

1. Motif multiconjunto-colorido: Multiconjunto sobre $\mathcal{C}$;
2. Motif grafo-colorido $T = (V_T, E_G, c : V_T \to \mathcal{C})$.

Operações em Grafos

$G = (V_G, E_G)$

$V' \subseteq V_G$;
$E' \subseteq E_G$

$G - V'$: G removendo os vértices V'
$G - E'$: G removendo as arestas E'

$(G - V') - E'$: *subgrafo de* G;
$G[V'] = G - (V_G - V')$: *grafo* G *induzido* por V'.

Ocorrência de M em G: $S \subseteq V_G$, $G[S]$ é conexo e induz o padrão M.

Problema (Problema geral de busca de motifs)

Entrada: *Rede G e um padrão M*
Objetivo: *M ocorre em G?*

Motif multiconjunto de cores

Eloi Araujo

FACOM – UFMS

12 de julho de 2017

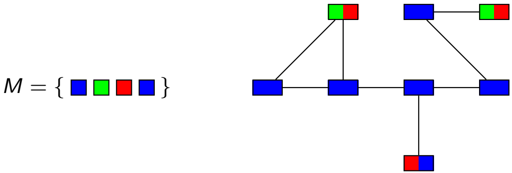
Ocorrência quando motif é um multiconjunto de cores

$\mathcal{C}$: conjunto de cores

$G = (V_G, E_G)$ grafo.
$\mathcal{L} : V_G \to 2^{\mathcal{C}}$ lista de cores
Um *grafo vértice-lista-colorido*: $G + \mathcal{L}$.

$M = \{1, \ldots, m\}$, $c : M \to \mathcal{C}$ padrão

ocorrência de M em G ($M \prec G$): $S \subseteq V_G$:
$G[S]$ é conexo e $\exists$ bijeção $f : S \to M$ tal que $c(f(u)) \in \mathcal{L}(u)$.



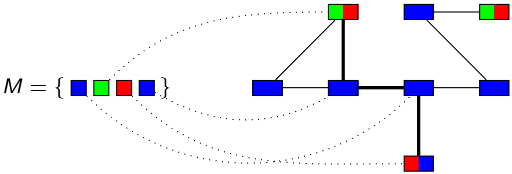
Ocorrência quando motif é um multiconjunto de cores

$\mathcal{C}$: conjunto de cores

$G = (V_G, E_G)$ grafo.
$\mathcal{L} : V_G \to 2^{\mathcal{C}}$ lista de cores
Um *grafo vértice-lista-colorido*: $G + \mathcal{L}$.

$M = \{1, \dots, m\}$, $c : M \to \mathcal{C}$ padrão

ocorrência de M em G ($M \prec G$): $S \subseteq V_G$:
$G[S]$ é conexo e $\exists$ bijeção $f : S \to M$ tal que $c(f(u)) \in \mathcal{L}(u)$.



Extensão do conceito de ocorrência

1. tolerar inserção ou remoção de elementos em M [BHK+10, DFV11].
2. tolerar o emparelhamento de (algumas) cores diferentes limitando o custo total [BHK+10].

Busca de padrão livre de topologia em rede (BPL)

Problema

Dados: G: *grafo vértice-lista-colorido*

M: *multiconjunto de cores*

Encontrar um ocorrência S *de* M *em* G.

Busca de padrão livre de topologia em rede (BPL)

principais resultados:

1. introduzido por [LFS06]
2. ferramenta MOTUS [LFS06]
3. NP-difícil, M conjunto e G árvore, $\Delta \leq 3$ [FFHV11]
4. NP-difícil, M possui 2 cores, G bipartido, $\Delta \leq 4$ [FFHV11]
5. NP-difícil, M conjunto e G possui diâmetro 2 [ABR$^+$10]
6. NP-difícil, G árvore, $\Delta \leq 4$ e M possui 2 cores [DFV13]
7. FPT para o parâmetro $k = |M|$ $O^*(2^k)$ [PZ14],
8. não existe FPT: parâmetro número de cores em M [FFHV11]
9. não existe FPT para o parâmetro $|V_G| - |M|$ [PZ14].

Variantes

indel e substituição [BKK13]: FPT; $O^*(2^k)$; $k =$ tam. da solução.

Em aberto

Complexidade parametrizada (parâmetro $V_G - M$) para árvores [FK16].

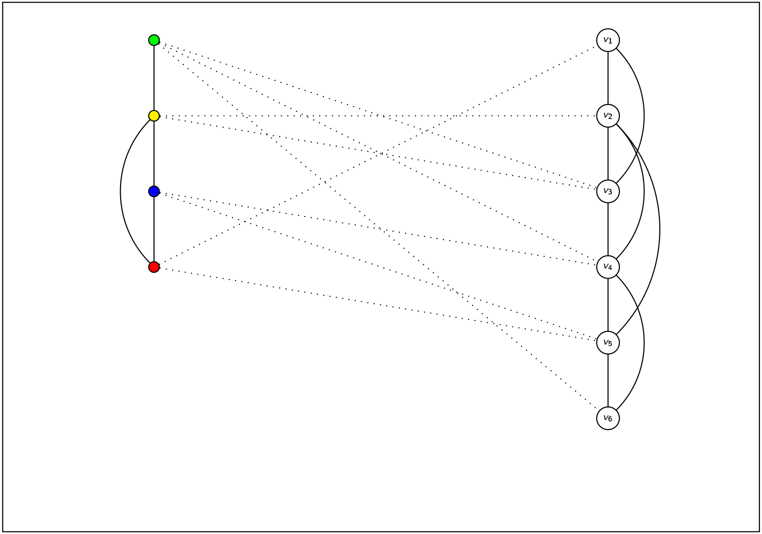
Motif topológico-colorido

Eloi Araujo

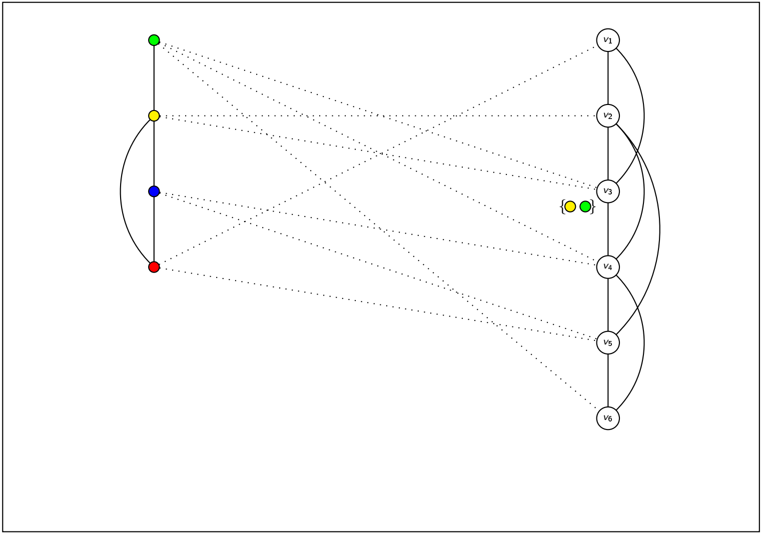
FACOM – UFMS

12 de julho de 2017

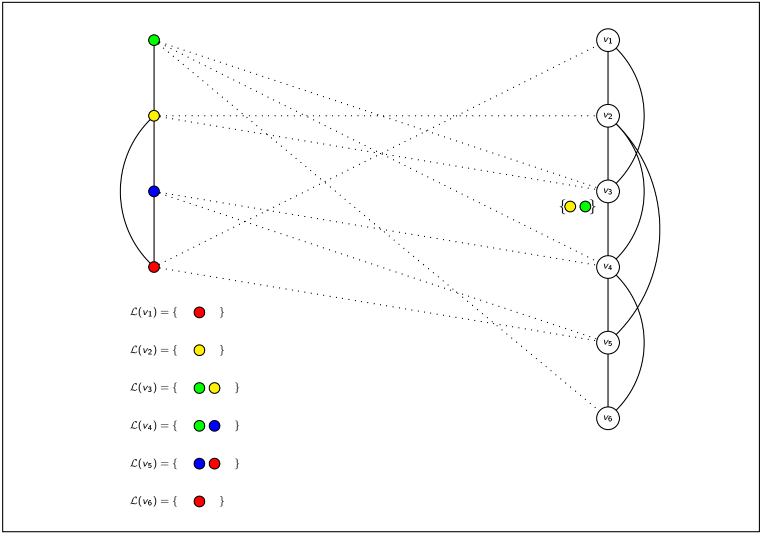
Busca de padrão com topologia de rede



Busca de padrão com topologia de rede



Busca de padrão com topologia de rede



Ocorrência: conjunto de vértices de um subgrafo de G (induzido ou não) isomorfo a M
respeita a atribuição de cores.

Busca de padrão com topologia de rede

Problema

Dados: G*: grafo vértice-lista-colorido*

M*: grafo vértice-colorido colorful (sem repetir cores)*

Objetivo: Encontrar um ocorrência S *de* M *em* G.

Problemas definidos por Fagnot [FLV08].

Busca de padrão com topologia de rede

μ: a cor que mais aparece no grafo G.

Resultados

1. NP-difícil para $\mu > 2$ e polinomial para $\mu \leq 2$ [FLV08].
2. $O^*(\mu^k)$ onde k é o número de cores que aparecem pelo menos duas vezes em G [FLV08].
3. $O^*((\mu - 1)^{k'})$ onde k' é o número de cores que aparecem pelo menos três vezes em G [Kom11].

Problemas estudados nesse trabalho

Eloi Araujo

FACOM – UFMS

12 de julho de 2017

Versão de decisão de busca de padrão com topologia de rede é **NP-completa!**

Versão de decisão de busca de padrão com topologia de rede é **NP-completa!**

redução circuito hamiltoniano: procura de um circuito M em um grafo G onde todos os vértices de G possuem todas as cores de M.

Problema 1 – Busca de padrão com topologia – SM

Problema

Busca de padrão com topologia em grafo vértice-colorido, i. e., $|\mathcal{L}(u)| = 1$ *para todo vértice* v. *(Subgrafo induzido ou não)*

Problema 1 – Busca de padrão com topologia – SM

Problema

Busca de padrão com topologia em grafo vértice-colorido, i. e., $|\mathcal{L}(u)| = 1$ *para todo vértice v. (Subgrafo induzido ou não)*

Simplificação que acreditava-se tornar o problema polinomial.

Teorema. SM é NP-completo. ☹

Teorema. SM é NP-completo. ☹

Prova. $3\text{-}\mathtt{SAT}(\Phi) \leq_P \mathtt{SM}(G, M)$

Teorema. SM é NP-completo. ☹

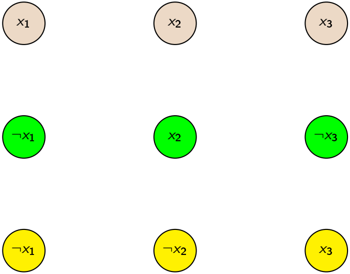
Prova. 3-SAT$(\Phi) \leq_P$ SM(G, M)
$\Phi = C_1 \wedge C_2 \wedge C_3$
$C_1 = x_1 \vee x_2 \vee x_3$, $C_2 = \neg x_1 \vee x_2 \vee \neg x_3$, e $C_3 = \neg x_1 \vee \neg x_2 \vee x_3$.

Teorema. SM é NP-completo. ☹

Prova. $3\text{-}\mathtt{SAT}(\Phi) \leq_P \mathtt{SM}(G, M)$

$\Phi = C_1 \wedge C_2 \wedge C_3$

$C_1 = x_1 \vee x_2 \vee x_3$, $C_2 = \neg x_1 \vee x_2 \vee \neg x_3$, e $C_3 = \neg x_1 \vee \neg x_2 \vee x_3$.

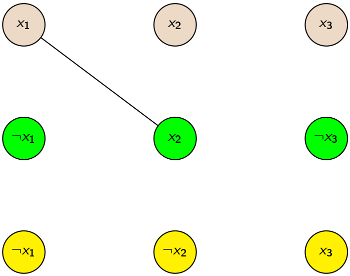


Teorema. $\texttt{SM}$ é NP-completo. ☹

Prova. $3\text{-}\texttt{SAT}(\Phi) \leq_P \texttt{SM}(G, M)$

$\Phi = C_1 \wedge C_2 \wedge C_3$

$C_1 = x_1 \vee x_2 \vee x_3$, $C_2 = \neg x_1 \vee x_2 \vee \neg x_3$, e $C_3 = \neg x_1 \vee \neg x_2 \vee x_3$.



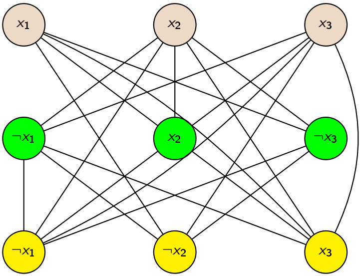
(x, y): $x \neq \neg y$; x e y e estão em clausulas diferentes.

Teorema. SM é NP-completo. ☹

Prova. $3\text{-SAT}(\Phi) \leq_P \text{SM}(G, M)$

$\Phi = C_1 \wedge C_2 \wedge C_3$

$C_1 = x_1 \vee x_2 \vee x_3$, $C_2 = \neg x_1 \vee x_2 \vee \neg x_3$, e $C_3 = \neg x_1 \vee \neg x_2 \vee x_3$.



G

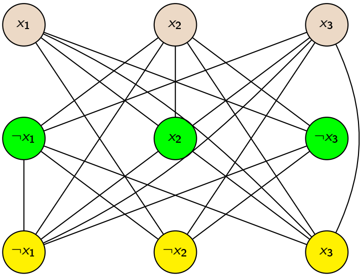
(x, y): $x \neq \neg y$; x e y e estão em clausulas diferentes.

Teorema. $\texttt{SM}$ é NP-completo. ☹

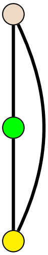
Prova. $3\text{-}\texttt{SAT}(\Phi) \leq_P \texttt{SM}(G, M)$

$\Phi = C_1 \wedge C_2 \wedge C_3$

$C_1 = x_1 \vee x_2 \vee x_3$, $C_2 = \neg x_1 \vee x_2 \vee \neg x_3$, e $C_3 = \neg x_1 \vee \neg x_2 \vee x_3$.



G



M

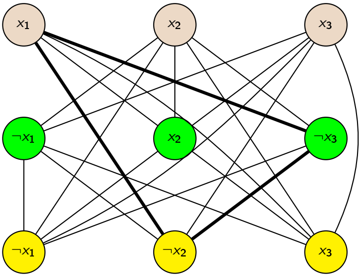
(x, y): $x \neq \neg y$; x e y e estão em clausulas diferentes.

Teorema. SM é NP-completo. ☹

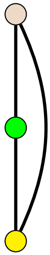
Prova. $3\text{-}\mathtt{SAT}(\Phi) \leq_P \mathtt{SM}(G, M)$

$\Phi = C_1 \wedge C_2 \wedge C_3$

$C_1 = x_1 \vee x_2 \vee x_3$, $C_2 = \neg x_1 \vee x_2 \vee \neg x_3$, e $C_3 = \neg x_1 \vee \neg x_2 \vee x_3$.



G



M

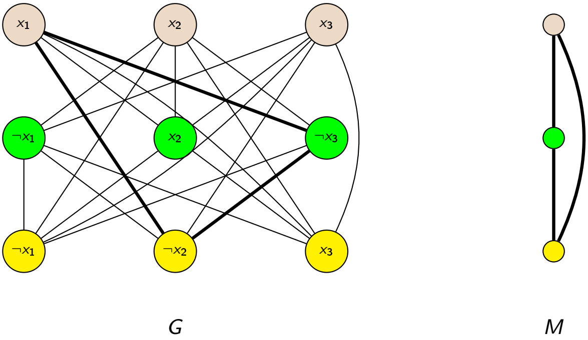
(x, y): $x \neq \neg y$; x e y e estão em clausulas diferentes.

Teorema. SM é NP-completo. ☹

Prova. $3\text{-SAT}(\Phi) \leq_P \text{SM}(G, M)$

$\Phi = C_1 \wedge C_2 \wedge C_3$

$C_1 = x_1 \vee x_2 \vee x_3$, $C_2 = \neg x_1 \vee x_2 \vee \neg x_3$, e $C_3 = \neg x_1 \vee \neg x_2 \vee x_3$.



(x, y): $x \neq \neg y$; x e y e estão em clausulas diferentes.

$x_1, \neg x_2, \neg x_3$ **satisfaz** Φ **e** M **ocorre em** G**!!!**

Problema 2 – Busca de padrão com topologia – ISM

Problema

Dados: G: *grafo vértice-lista-colorido*

M: *árvore.*

Decidir se existe um conjunto $S \subseteq V_G$ *tal que* $G[S] = M$.

Problema 2 – Busca de padrão com topologia – ISM

Problema

Dados: G: *grafo vértice-lista-colorido*

M: *árvore.*

Decidir se existe um conjunto $S \subseteq V_G$ *tal que* $G[S] = M$.

Nova tentativa de simplificar o problema.

Teorema. $\texttt{ISM}$ é NP-completo. ☹

Teorema. $\texttt{ISM}$ é NP-completo. ☹

Prova. $3\text{-}\texttt{SAT}(\Phi) \leq_P \texttt{ISM}(G, M)$

Teorema. $\texttt{ISM}$ é NP-completo. ☹

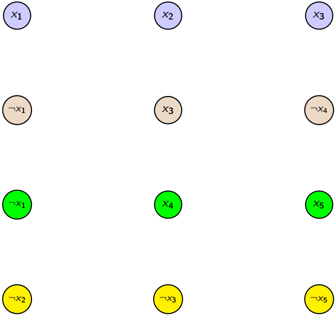
Prova. $3\text{-}\texttt{SAT}(\Phi) \leq_P \texttt{ISM}(G, M)$
$\Phi = C_1 \wedge C_2 \wedge C_3 \wedge C_4$
$C_1 \wedge C_2 \wedge C_3 \wedge C_4,\ C_1 = x_1 \vee x_2 \vee x_3,\ C_2 = \neg x_1 \vee x_3 \vee \neg x_4,$
$C_3 = \neg x_1 \vee x_4 \vee x_5,\ C_4 = \neg x_2 \vee \neg x_3 \vee \neg x_5.$

Teorema. $\mathtt{ISM}$ é NP-completo. ☹

Prova. $\text{3-}\mathtt{SAT}(\Phi) \leq_P \mathtt{ISM}(G, M)$

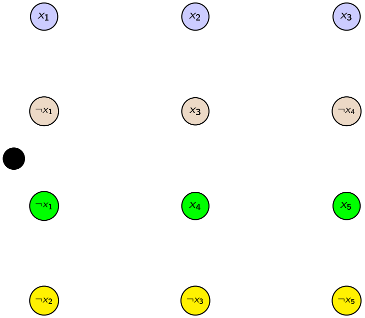
$\Phi = C_1 \wedge C_2 \wedge C_3 \wedge C_4$

$C_1 \wedge C_2 \wedge C_3 \wedge C_4$, $C_1 = x_1 \vee x_2 \vee x_3$, $C_2 = \neg x_1 \vee x_3 \vee \neg x_4$,
$C_3 = \neg x_1 \vee x_4 \vee x_5$, $C_4 = \neg x_2 \vee \neg x_3 \vee \neg x_5$.



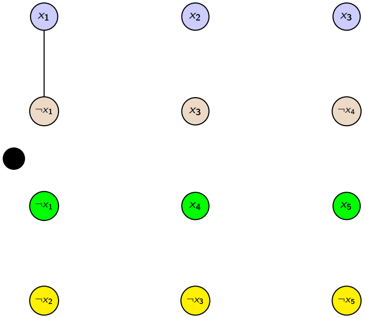
Teorema. ISM é NP-completo. ☹

Prova. $\texttt{3-SAT}(\Phi) \leq_P \texttt{ISM}(G, M)$
$\Phi = C_1 \wedge C_2 \wedge C_3 \wedge C_4$
$C_1 \wedge C_2 \wedge C_3 \wedge C_4$, $C_1 = x_1 \vee x_2 \vee x_3$, $C_2 = \neg x_1 \vee x_3 \vee \neg x_4$,
$C_3 = \neg x_1 \vee x_4 \vee x_5$, $C_4 = \neg x_2 \vee \neg x_3 \vee \neg x_5$.



Teorema. $\texttt{ISM}$ é NP-completo. ☹

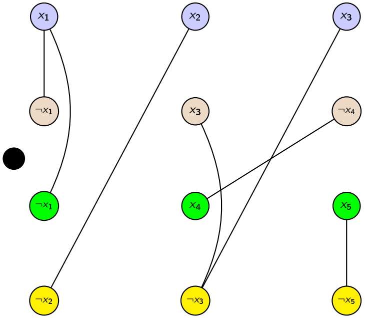
Prova. $3\text{-}\texttt{SAT}(\Phi) \leq_P \texttt{ISM}(G, M)$
$\Phi = C_1 \wedge C_2 \wedge C_3 \wedge C_4$
$C_1 \wedge C_2 \wedge C_3 \wedge C_4$, $C_1 = x_1 \vee x_2 \vee x_3$, $C_2 = \neg x_1 \vee x_3 \vee \neg x_4$,
$C_3 = \neg x_1 \vee x_4 \vee x_5$, $C_4 = \neg x_2 \vee \neg x_3 \vee \neg x_5$.



(x, y): $x = \neg y$; x e y e estão em clausulas diferentes.

Teorema. $\texttt{ISM}$ é NP-completo. ☹

Prova. $3\text{-}\texttt{SAT}(\Phi) \leq_P \texttt{ISM}(G, M)$
$\Phi = C_1 \wedge C_2 \wedge C_3 \wedge C_4$
$C_1 \wedge C_2 \wedge C_3 \wedge C_4$, $C_1 = x_1 \vee x_2 \vee x_3$, $C_2 = \neg x_1 \vee x_3 \vee \neg x_4$,
$C_3 = \neg x_1 \vee x_4 \vee x_5$, $C_4 = \neg x_2 \vee \neg x_3 \vee \neg x_5$.



(x, y): $x = \neg y$; x e y e estão em clausulas diferentes.

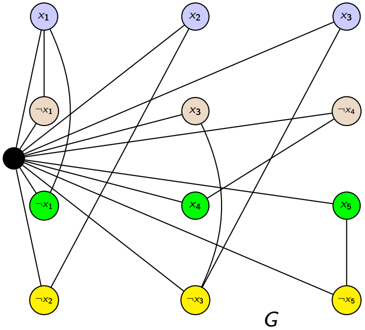
Teorema. $\texttt{ISM}$ é NP-completo. ☹

Prova. $3\text{-}\texttt{SAT}(\Phi) \leq_P \texttt{ISM}(G, M)$

$\Phi = C_1 \wedge C_2 \wedge C_3 \wedge C_4$

$C_1 \wedge C_2 \wedge C_3 \wedge C_4$, $C_1 = x_1 \vee x_2 \vee x_3$, $C_2 = \neg x_1 \vee x_3 \vee \neg x_4$,

$C_3 = \neg x_1 \vee x_4 \vee x_5$, $C_4 = \neg x_2 \vee \neg x_3 \vee \neg x_5$.



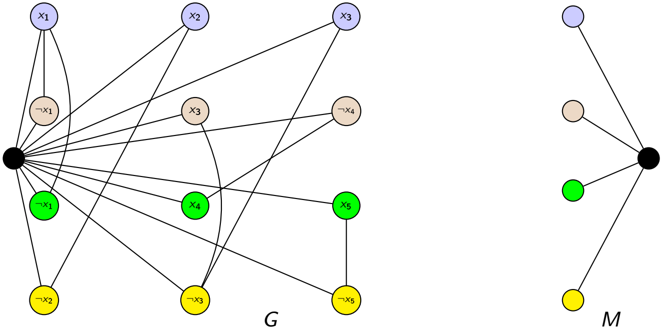
(x, y): $x = \neg y$; x e y e estão em clausulas diferentes.

Teorema. ISM é NP-completo. ☹

Prova. 3-SAT$(\Phi) \leq_P$ ISM(G, M)

$\Phi = C_1 \wedge C_2 \wedge C_3 \wedge C_4$

$C_1 \wedge C_2 \wedge C_3 \wedge C_4$, $C_1 = x_1 \vee x_2 \vee x_3$, $C_2 = \neg x_1 \vee x_3 \vee \neg x_4$,
$C_3 = \neg x_1 \vee x_4 \vee x_5$, $C_4 = \neg x_2 \vee \neg x_3 \vee \neg x_5$.



(x, y): $x = \neg y$; x e y e estão em clausulas diferentes.

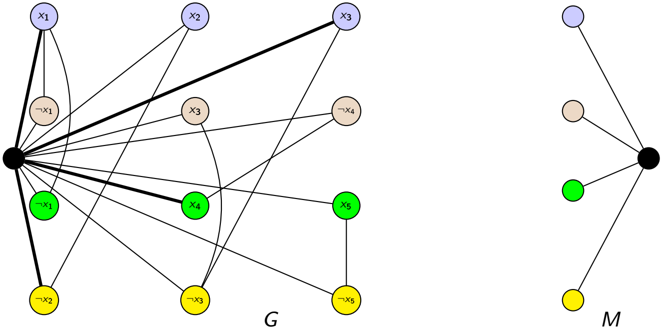
Teorema. ISM é NP-completo. ☹

Prova. $\text{3-SAT}(\Phi) \leq_P \text{ISM}(G, M)$

$\Phi = C_1 \wedge C_2 \wedge C_3 \wedge C_4$

$C_1 \wedge C_2 \wedge C_3 \wedge C_4$, $C_1 = x_1 \vee x_2 \vee x_3$, $C_2 = \neg x_1 \vee x_3 \vee \neg x_4$,

$C_3 = \neg x_1 \vee x_4 \vee x_5$, $C_4 = \neg x_2 \vee \neg x_3 \vee \neg x_5$.



(x, y): $x = \neg y$; x e y e estão em clausulas diferentes.

$x_1, \neg x_2, \neg x_3$ **satisfaz** Φ **e** M **ocorre em** G**!!!**

Problema 3 – Busca de padrão com topologia – TGC

Problema

Problema SM restrito a árvores fazendo o papel de motifs (subgrafo não precisa ser induzido).

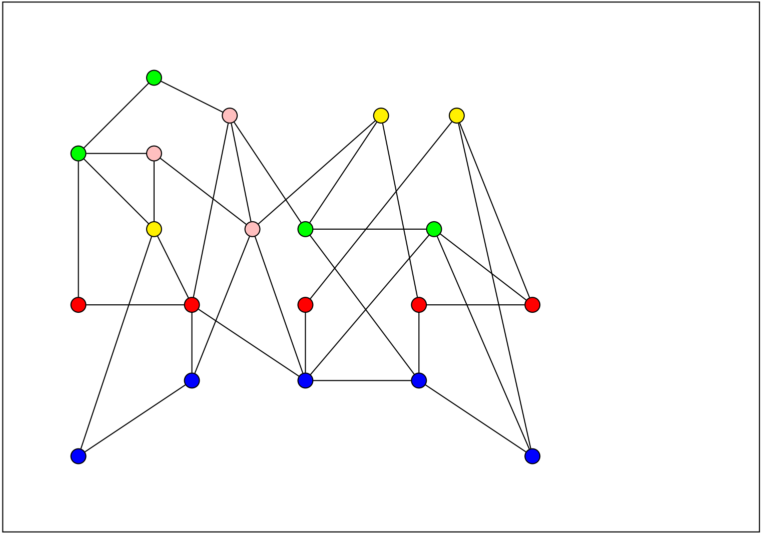
Problema 3 – Busca de padrão com topologia – TGC

Problema

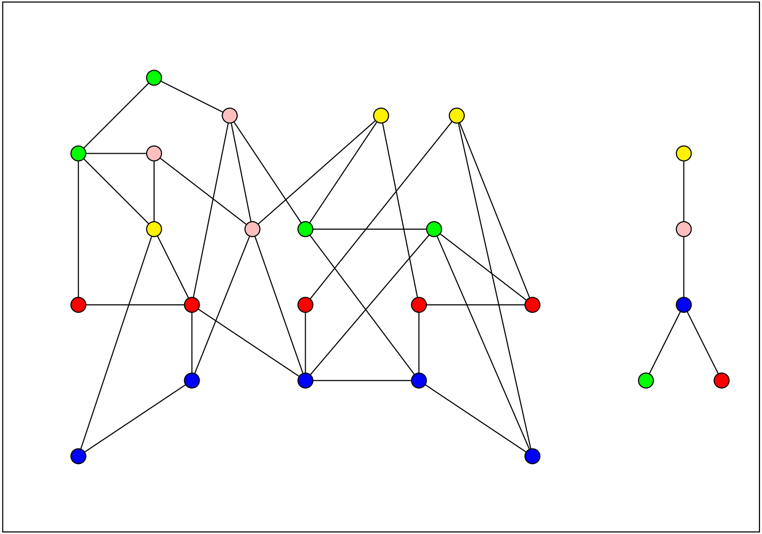
Problema SM restrito a árvores fazendo o papel de motifs (subgrafo não precisa ser induzido).

Polinomial ☺

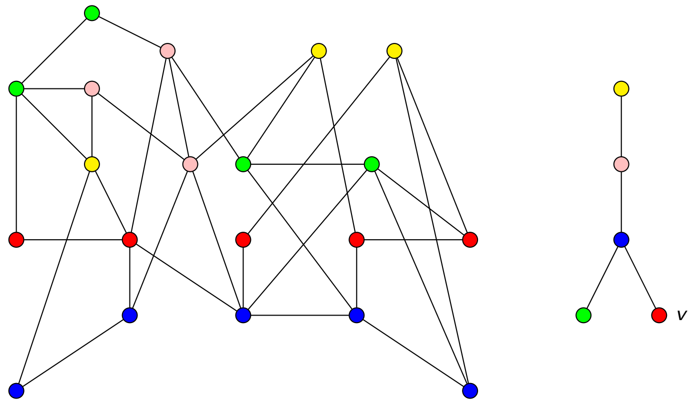
Algoritmo $\texttt{TCG}(G, M)$:=



Algoritmo $\mathtt{TCG}(G, M)$:=

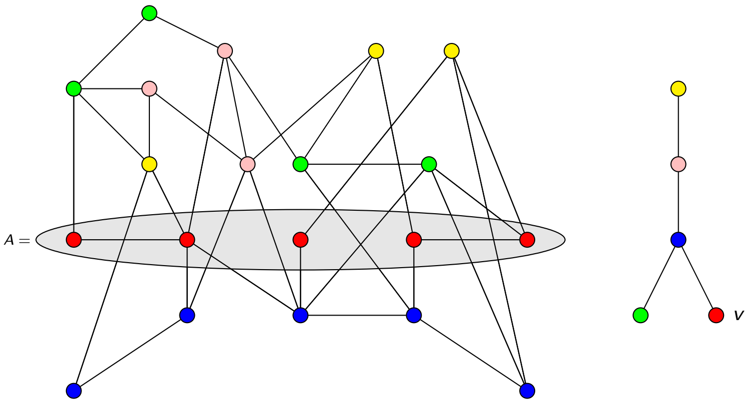


Algoritmo $\mathtt{TCG}(G, M)$:=



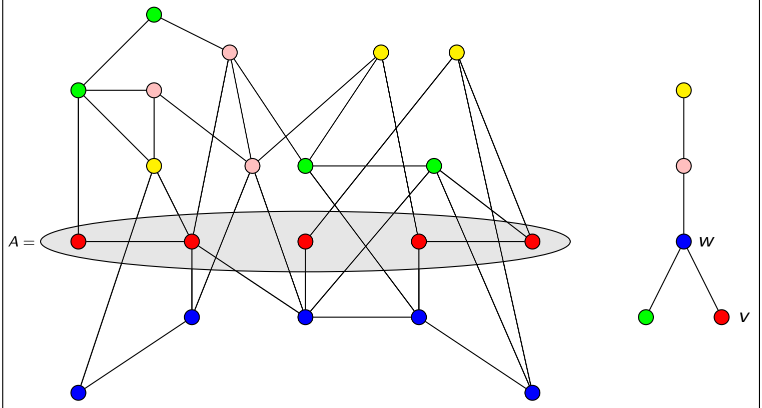
v : **folha de** M

Algoritmo $\texttt{TCG}(G, M)$:=



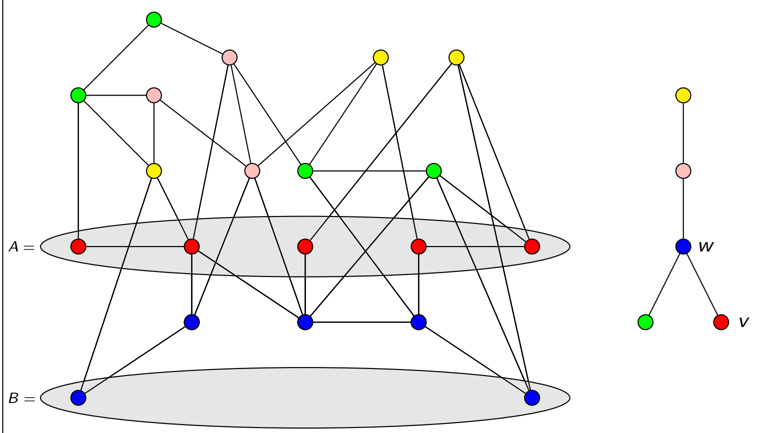
$A := \{a \in V_G : cor(a) = cor(x)\}$

Algoritmo $\mathtt{TCG}(G, M)$:=



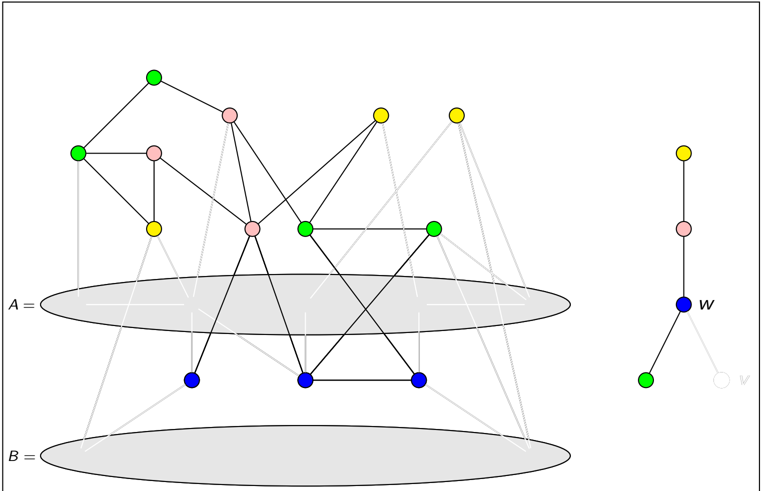
w **adjacente a** v

Algoritmo $\texttt{TCG}(G, M)$:=



$B := \{b \in V_G : cor(b) = cor(v)$ e não possui adjacente de cor de $w\}$

Algoritmo $\texttt{TCG}(G, M)$:=



devolva $\mathrm{TCG}(G - (A \cup B), M - \{x\})$

$\text{TCG}(G, M) :=$
Entrada: grafo G and colorful M
Questão: $M \prec G$?

1. v : folha de M
2. $A := \{a \in V_G : cor(a) = cor(v)\}$
3. **se** $|V_M| = 1$, devolva $A \neq \emptyset$
4. $w \in V_M : vw \in E_M$
5. $B := \{b \in V_G :$

   $cor(b) = cor(v)$ e não possui adjacente de cor de $w\}$
6. devolva $\text{TCG}(G - (A \cup B), M - \{x\})$

TCG(G, M) :=
Entrada: grafo G and colorful M
Questão: $M \prec G$?

1. v : folha de M
2. $A := \{a \in V_G : cor(a) = cor(v)\}$
3. **se** $|V_M| = 1$, devolva $A \neq \emptyset$
4. $w \in V_M : vw \in E_M$
5. $B := \{b \in V_G :$

   $cor(b) = cor(v)$ e não possui adjacente de cor de $w\}$
6. devolva TCG($G - (A \cup B), M - \{x\}$)

Tempo gasto: $m = E_G$

TCG$(G, M) :=$
Entrada: grafo G and colorful M
Questão: $M \prec G$?

1. v : folha de M
2. $A := \{a \in V_G : cor(a) = cor(v)\}$
3. **se** $|V_M| = 1$, devolva $A \neq \emptyset$
4. $w \in V_M : vw \in E_M$
5. $B := \{b \in V_G :$

   $cor(b) = cor(v)$ e não possui adjacente de cor de $w\}$
6. devolva TCG$(G - (A \cup B), M - \{x\})$

Tempo gasto: $m = E_G$

$$T(V_M, V_G, m) = \begin{cases} \Theta(V_G) & \text{se } V_T = 1 \\ T(V_M - 1, V_G, m) + O(V_G + m) & \text{caso contrário.} \end{cases}$$

$\mathrm{TCG}(G, M) :=$

Entrada: grafo G and colorful M

Questão: $M \prec G$?

1. v : folha de M
2. $A := \{a \in V_G : cor(a) = cor(v)\}$
3. **se** $|V_M| = 1$, devolva $A \neq \emptyset$
4. $w \in V_M : vw \in E_M$
5. $B := \{b \in V_G :$

   $cor(b) = cor(v)$ e não possui adjacente de cor de $w\}$
6. devolva $\mathrm{TCG}(G - (A \cup B), M - \{x\})$

Tempo gasto: $m = E_G$

$$T(V_M, V_G, m) = \begin{cases} \Theta(V_G) & \text{se } V_T = 1 \\ T(V_M - 1, V_G, m) + O(V_G + m) & \text{caso contrário.} \end{cases}$$

$$\Rightarrow T(V_M, V_G, m) = O(V_M \cdot (V_G + m))$$

Outros problemas

Considerando grafo vértice-colorido e motif colorful árvore

Outros problemas

Considerando grafo vértice-colorido e motif colorful árvore

1. Contagem de todas as ocorrências disjuntas de M em G – NP-difícil

Outros problemas

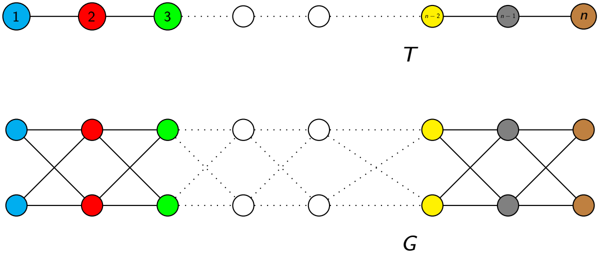
Considerando grafo vértice-colorido e motif colorful árvore

1. Contagem de todas as ocorrências disjuntas de M em G – NP-difícil
2. Enumeração das ocorrências não necessariamente disjuntas de M

Outros problemas

Considerando grafo vértice-colorido e motif colorful árvore

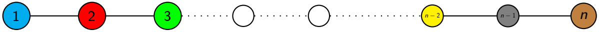
1. Contagem de todas as ocorrências disjuntas de M em G – NP-difícil
2. Enumeração das ocorrências não necessariamente disjuntas de M pode haver um número exponencial



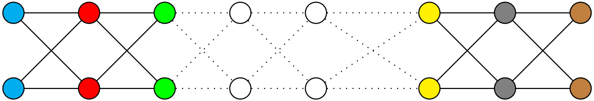
Outros problemas

Considerando grafo vértice-colorido e motif colorful árvore

1. Contagem de todas as ocorrências disjuntas de M em G – NP-difícil
2. Enumeração das ocorrências não necessariamente disjuntas de M pode haver um número exponencial



T



G

3. Contagem de todas as ocorrências de M em G. (inferência)

Trabalhos futuros: ocorrência em motifs topológicos

1. deleção e remoção
2. uso de peso para diferentes pares de cores
3. estratégia com generalização para grafos mais gerais (heurísticas, FPT).

Referências I

A.M. Ambalath, R. Balasundaram, C. Rao, V. Koppula, N. Misra, G. Philip, and M.S. Ramanujan.
On the kernelization complexity of colorful motifs.
In *IPEC*, 2010.

S. Bruckner, F. Hüffner, R.M. Karp, R. Shamir, and R. Sharan.
Topology-free querying of protein interaction networks.
Journal of computational biology, 2010.

A. Björklund, P. Kaski, and L. Kowalik.
Probably optimal graph motifs.
In *LIPIcs-Leibniz Internat. Proceedings in Informatics*, 2013.

R. Dondi, G. Fertin, and S. Vialette.
Complexity issues in vertex-colored graph pattern matching.
Journal of Discrete Algorithms, 2011.

Referências II

R. Dondi, G. Fertin, and S. Vialette.
Finding approximate and constrained motifs in graphs.
Theoretical Computer Science, 2013.

M.R Fellows, G. Fertin, D. Hermelin, and S. Vialette.
Upper and lower bounds for finding connected motifs in vertex-colored graphs.
Journal of Computer and System Sciences, 2011.

G. Fertin and C. Komusiewicz.
Graph motif problems parameterized by dual.
In *LIPIcs-Leibniz Internat. Proceedings in Informatics*, 2016.

I. Fagnot, G. Lelandais, and S. Vialette.
Bounded list injective homomorphism for comparative analysis of protein–protein interaction graphs.
Journal of Discrete Algorithms, 2008.

Referências III

C. Komusiewicz.
Parameterized algorithmics for network analysis: clustering & querying.
Univerlagtuberlin, 2011.

V. Lacroix, C.G. Fernandes, and M-F. Sagot.
Motif search in graphs: application to metabolic networks.
IEEE/ACM Transactions on Comp. Biol. and Bioinformatics, 2006.

R. Y Pinter and M. Zehavi.
Algorithms for topology-free and alignment network queries.
Journal of Discrete Algorithms, 2014.