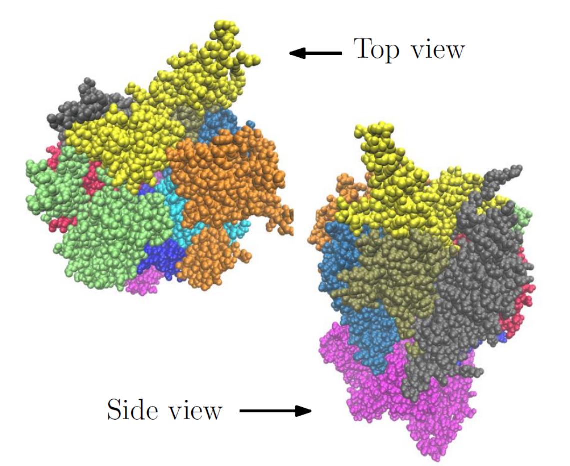
POLYTECH NICE SOPHIA

PROJET SI4 - ALGORITHMES ET COMPLEXITÉ.

RAPPORT DE PROJET

Algorithmes pour l'inférence de connectivité avec application en biologie structurale computationnelle



Author
Louis AMAS
Maxime LERAS

Author
Pierre-Antoine BOULAT
Julien MOLINIER

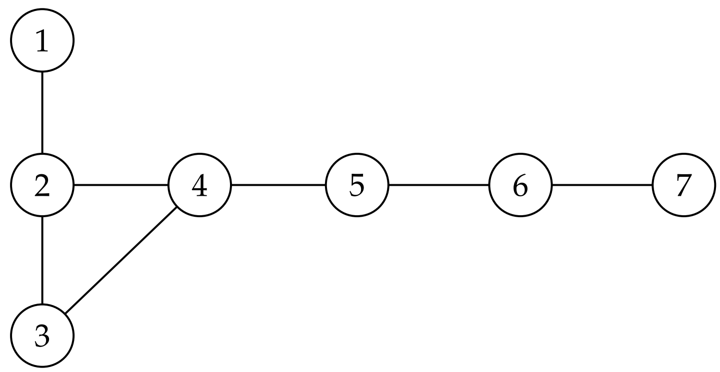
January 5, 2020

1 Question 1

$$Pour\ \Delta = 3 : \tag{1}$$

Une solution existe avec :

$$\Delta(G) = deg(\widehat{S}_2) = deg(\widehat{S}_4) = 3 \tag{2}$$



$$Pour\ \Delta = 2 : \tag{3}$$

Pas de solution car les sommets 4 et 5 sont présent dans 3 sous complexes, ce qui force un degrée minimal sur ces sommets de 3

$$|C_b \cap C_c \cap C_d| = 2 \quad < 3 \tag{4}$$

2 Question 2

Lorsque $\Delta = |V|$, il existe une solution au problème IC lorsque

$$k \geq (\sum_{i=1}^{t} |C_i|) - t \tag{5}$$

Ces valeurs de k ne sont vraies qu'en faisant la supposition que
$|C_i \cap C_j| \leq 1$ pour tout $i \in \{1, ..., t\}$ et pour tout $j \in \{1, ..., t\}$, $i \neq j$.

En effet, si on avait $|C_i \cap C_j| > 1$, la valeur de k minimale pour que le problème IC ait une solution serait inférieure à la valeur donnée en (5)

Si $k = |V|^2$, on observe qu'un graphe complet a au plus $\frac{n*(n+1)}{2}$ arrêtes où $n = |V|$

Or $\forall n \geq 1, \frac{n*(n+1)}{2} \leq n^2$, Cela revient donc à ne pas se préoccuper de k car $|E|$ ne pourra jamais être supérieur à k

Les valeurs de Δ pour lesquelles IC a une solution sont donc toutes les valeurs supérieures ou égales au nombre maximal d'occurences d'un sommet dans tous les sous-complexes

Si on prend l'exemple d'un assemblage dont tous les sous-complexes contiennent le même sommet $\widehat{S}_1$, toujours en respectant la supposition de l'énoncé : $|C_i \cap C_j| \leq 1$ pour tout $i \in \{1, ..., t\}$ et pour tout $j \in \{1, ..., t\}$, $i \neq j$.

On obtient un graphe en forme "d'étoile" avec pour centre le sommet $\widehat{S}_1$. Pour que IC ait une solution dans cette instance, il faut $\Delta = t$ car t est le nombre maximal d'occurences d'un sommet dans les sous-ensembles et ce sommet est ici $\widehat{S}_1$

3 Question 3

Le nombre d'abres couvrants différents pour un graphe complet $G(V, E)$ à p sommets peut directement être calculé avec la formule de Cayley et on obtient :

$$a(G) = p^{p-2} \tag{6}$$

Le nombre de solutions différentes en fonction de la taille des C_i si $\Delta = |V|$ est le produit des p^{p-2} appliqués à chacun des sous-complexes (avec p le nombre de sommets du sous-complexe).

4 Question 4

Les conditions d'existence d'une solution en supposant $\Delta = 2$, $|C_i \cap C_j| \leq 1$ pour tout $i \in \{1, ..., t\}$ et pour tout $j \in \{1, ..., t\}$, $i \neq j$ sont les suivantes :

- Chaque protéine doit appartenir au maximum à 2 sous-complexes ayant 2 protéines au minimum

- Dans un sous-complexe, on ne peut pas avoir plus de 2 protéines si elles appartiennent aussi à un autre sous-complexe.

5 Question 8

Δ k	110	120	130	140	150	160	170	180	190	200
2	0	0	0	0	0	0	0	0	0	0
3	0	0	0	0	0	0	0	0	0	0
4	0	0	0	0	0	0	0	0	0	0
5	0	0	0	0	0	0	0	0	0	0
6	0	0	0	0	0	0	0	30%	10%	10%
7	0	0	0	0	0	0	0	20%	10%	10%
8	0	0	0	0	0	0	0	40%	90%	80%
9	0	0	0	0	0	0	0	70%	80%	90%

6 Question 9

Comme la limite de 200 arrêtes ne suffit pas et que certains tableaux ne contiennent que des 0 la limite a été repoussé à 400 pour cette question.

$t = 20, p = 10$

Δ k	250	260	270	280	290	300	310	320	330	340	350	360	370	380	390	400	410	420	430	440
1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
3	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
4	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
5	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
6	0	0	10%	0	0	0	0	10%	0	0	10%	0	10%	20%	10%	0	0	0	0	0
7	20%	10%	20%	50%	10%	50%	20%	10%	30%	30%	20%	0	10%	10%	20%	0	0	0	0	0
8	70%	50%	50%	70%	80%	70%	80%	100%	90%	70%	70%	70%	80%	50%	80%	0	0	0	0	0
9	80%	70%	60%	80%	90%	90%	80%	80%	70%	90%	70%	90%	100%	80%	80%	0	0	0	0	0

$t = 20, p = 15$

Δ k	250	260	270	280	290	300	310	320	330	340	350	360	370	380	390	400	410	420	430	440
1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
3	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
4	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
5	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
6	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
7	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
8	0	0	0	20%	10%	20%	20%	0	10%	20%	10%	0	10%	20%	0	0	0	0	0	0
9	0	0	0	50%	10%	20%	0	50%	10%	30%	20%	30%	20%	10%	20%	0	0	0	0	0

$t = 20, p = 20$

Δ k	250	260	270	280	290	300	310	320	330	340	350	360	370	380	390	400	410	420	430	440
1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
3	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
4	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
5	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
6	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
7	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
8	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
9	0	0	0	0	0	0	0	0	0	0	0	0	0	10%	0	0	0	0	0	0

$t = 30, p = 10$

Δ k	250	260	270	280	290	300	310	320	330	340	350	360	370	380	390	400	410	420	430	440
1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
3	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
4	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
5	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
6	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
7	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
8	0	0	0	20%	10%	10%	0	0	10%	10%	10%	20%	20%	10%	10%	0	0	0	0	0
9	0	0	20%	0	20%	20%	20%	40%	20%	40%	40%	20%	30%	10%	40%	0	0	0	0	0

$t = 30, p = 15$

Δ k	250	260	270	280	290	300	310	320	330	340	350	360	370	380	390	400	410	420	430	440
1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
3	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
4	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
5	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
6	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
7	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
8	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
9	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0