UN CRIBLE NON-COMMUTATIF EN BIOLOGIE

UNE INTRODUCTION AUX CRIBLES

L M J L
UMR 6629 - Nantes

Pierre-Louis Giscard, Paul Rochet, Richard Wilson

PLAN

1. Contexte

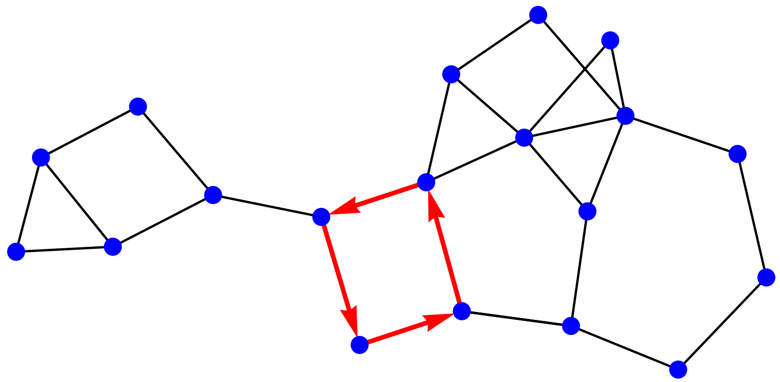
2. La version la plus simple …

… du crible le plus simple

3. En biologie des systèmes

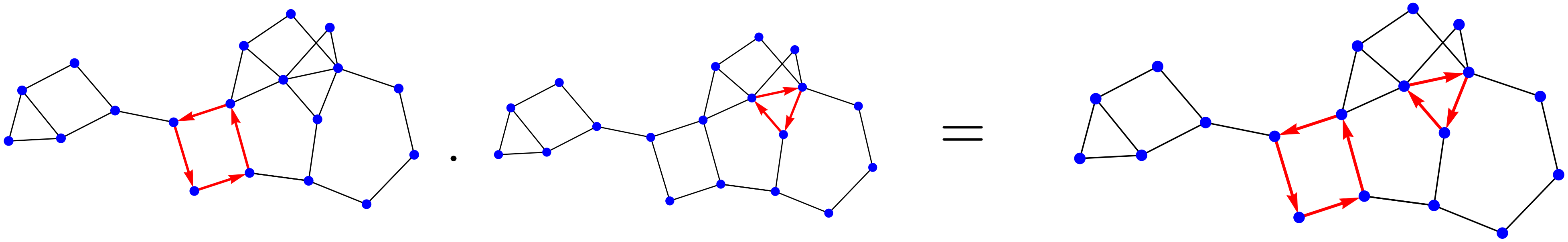
CONTEXTE

Les cycles simples



Point de départ ne compte pas
Orientation compte

Produit `concatenation' $c.c'$



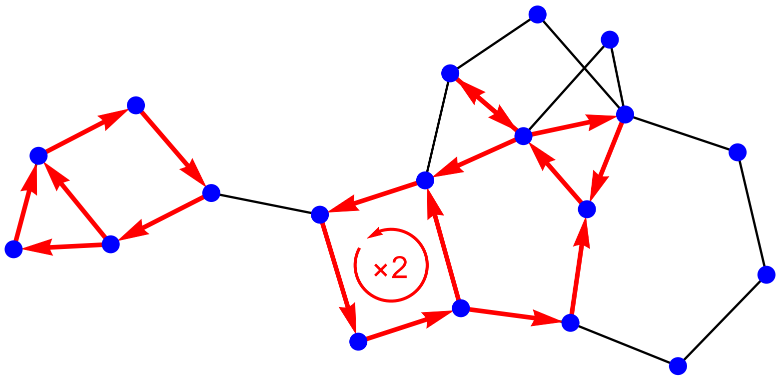
Semi-commutativité $cc' = c'c \iff \mathcal{V}(c) \cap \mathcal{V}(c') = \emptyset$

CONTEXTE

Randonnée (fermée)

$$h = c_1 c_2 c_3 c_4 \cdots c_k$$

à commutation près de cycles successifs disjoints



$$h = c_1 c_2 c_3 c_4 \cdots c_k$$

Si le cycle simple à droite est unique *h* est un chemin

CONTEXTE

Les randonnées obéissent à l'extension semi-commutative de la théorie des nombres

La structure combinatoire de la théorie des nombres est celle des randonnées sur une graphe où tous les cycles simples commutent

CONTEXTE

Les randonnées obéissent à l'extension semi-commutative de la théorie des nombres

La structure combinatoire de la ... nombres est celle des randonnées sur une c... us les cycles simples ...

BOUH ! ! (?)

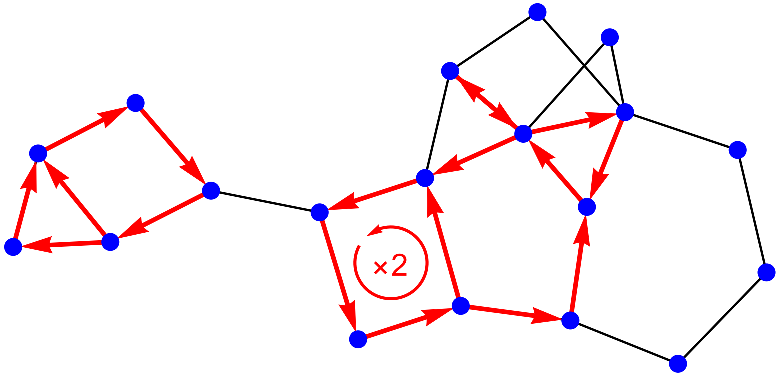
UNE EXTENSION DE LA THÉORIE DES NOMBRES

	Randonnées	Théorie des nombres
Zêta	$\zeta = \mathcal{S}1 = \sum_{h\in\mathcal{H}} h$ $\zeta = \frac{1}{\det(\mathsf{I}-\mathsf{W})}$	$\zeta_R(s) = \sum_{n>0} \frac{1}{n^s}$
Möbius	$\mu(h) = \begin{cases} (-1)^{\Omega(h)}, & h \text{ auto-évitante} \\ 0, & \text{sinon.} \end{cases}$	$\mu(n) = \begin{cases} (-1)^{\Omega(n)}, & n \text{ sans facteur carré} \\ 0, & \text{sinon.} \end{cases}$
Von Mangoldt	$\Lambda(h) = \begin{cases} \ell(p), & h \text{ chemin, } p\|h \text{ à droite} \\ 0, & \text{sinon.} \end{cases}$ $\mathcal{S}\Lambda = \mathrm{Tr}\left[\left(\mathsf{I}-\mathsf{W}\right)^{-1}\right]$	$\Lambda(n) = \begin{cases} \log(p), & n = p^k \\ 0, & \text{sinon.} \end{cases}$
Liouville	$\lambda(h) = (-1)^{\Omega(h)}$ $\mathcal{S}\lambda = \frac{1}{\mathrm{perm}(\mathsf{I}-\mathsf{W})}$ $\vdots$	$\lambda(h) = (-1)^{\Omega(n)}$ $\vdots$

UNE EXTENSION DE LA THÉORIE DES NOMBRES

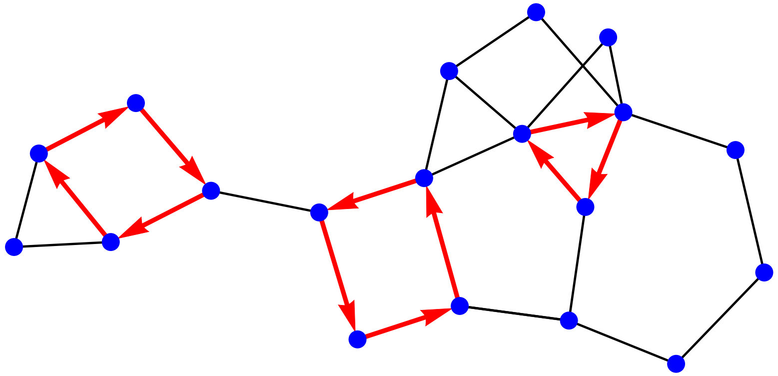
	Randonnées	Théorie des nombres
Nombre de diviseurs	ζ^2	$\zeta_R(s)^2$
Log Zêta	$\log\zeta = \sum_h \frac{\Lambda(h)}{\ell(h)} h$	$\log\zeta_R(s) = \sum_n \frac{\Lambda(n)}{\log(n)} \frac{1}{n^s}$
Log-Mangoldt	$\ell(h) = \sum_{h'\mid h} \Lambda(h')$	$\log(n) = \sum_{d\mid n} \Lambda(n)$
Fonctions totalement multiplicatives	$f^{-1} = \sum_h \mu(h) f(h) h$ $f'/f = -\sum_h \Lambda(h) f(h) h$	$f^{-1} = \sum_n \frac{\mu(n) f(n)}{n^s}$ $f'/f = \sum_n \frac{\Lambda(n) f(n)}{n^s}$
Fonctions totalement additives	$(f*\mu)(h) = \begin{cases} f(p), & h \text{ chemin, } p\mid h \text{ à droite} \\ 0, & \text{sinon.} \end{cases}$	$(f*\mu)(n) = \begin{cases} f(p), & n = p^k \\ 0, & \text{sinon.} \end{cases}$
ζ'/ζ via les premiers	$-\sum_{\gamma:\,\text{cycle simple}} \ell(\gamma) \frac{\det(\mathsf{I}-\mathsf{W}_{\setminus\gamma})}{\det(\mathsf{I}-\mathsf{W})}$	$-\sum_p \log p \, \frac{p^{-s}}{1-p^{-s}}$
Série Ω	$\sum_{w:\,\text{chemin}} w = \det(\mathsf{I}-\mathsf{W}) \sum_{h\in\mathcal{H}} \Omega(h)\, h$ $\vdots$	$\sum_{p,n} \frac{1}{p^{-ns}} = \zeta_R(s)^{-1} \sum_n \frac{\Omega(n)}{n^s}$ $\vdots$

LES RANDONNÉES



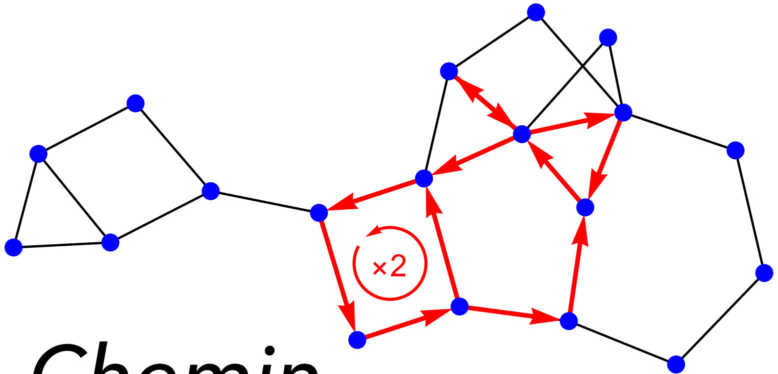
Randonnée h

nombre entier



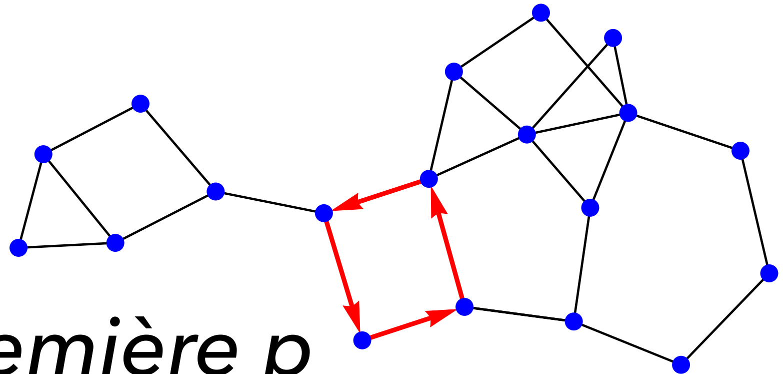
Auto-évitante d

nombre sans facteur carré



Chemin

nombre de la forme p^k



Première p

nombre premier

PLAN

1. Contexte

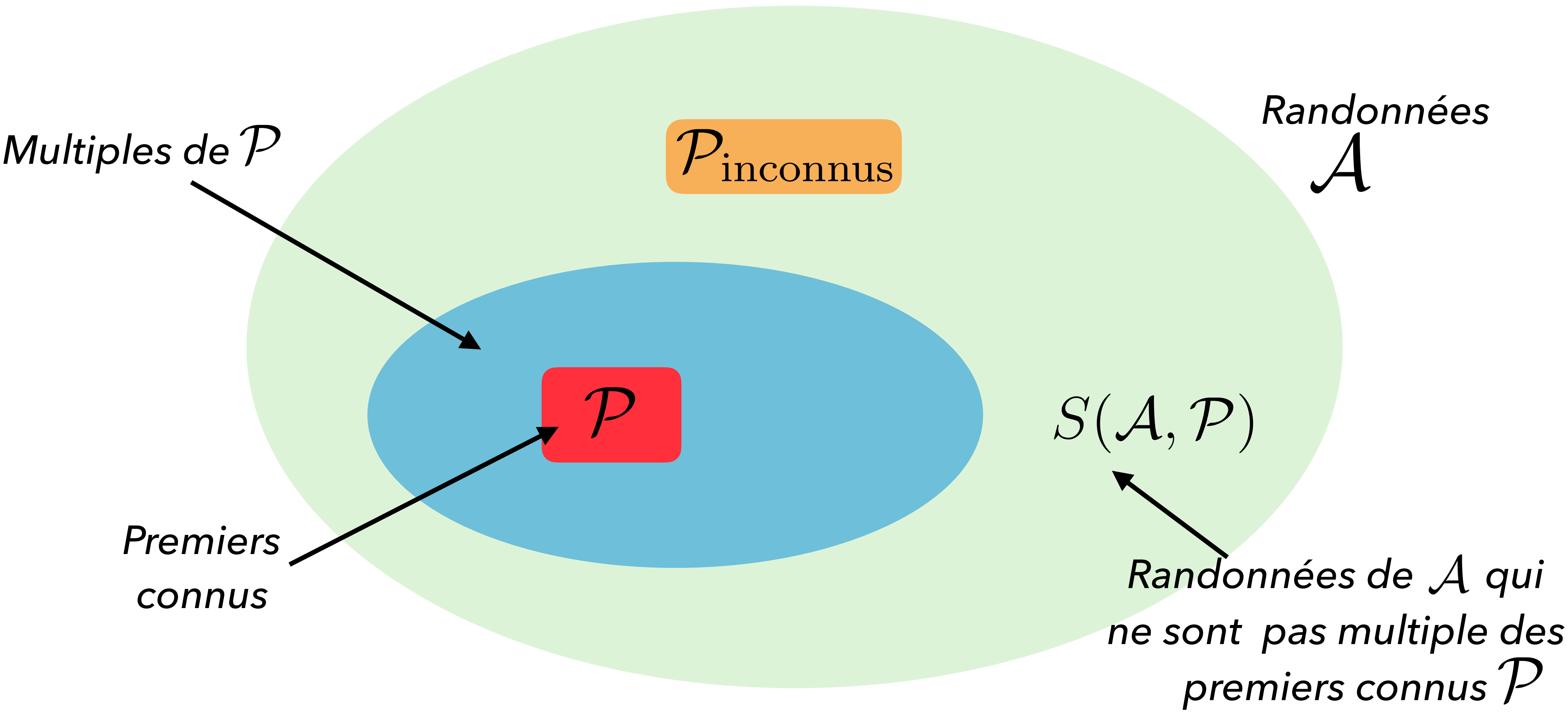
2. La version la plus *simple* ...

... du crible le plus *simple*

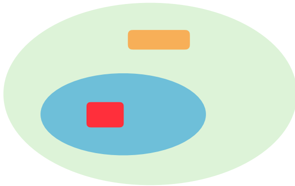
3. En biologie des systèmes

QU'EST CE QU'UN CRIBLE ?

Quelle est la fraction (asymptotique) des randonnées de rang r dans un ensemble $\mathcal{A}$ satisfaisant une propriété donnée ?



OBSERVATIONS GÉNÉRALES

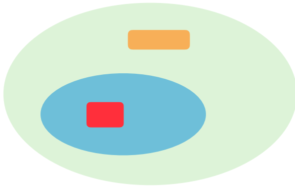


On connait un ensemble $\mathcal{P}$ de premiers

Quel genre d'ensemble?

On verra

OBSERVATIONS GÉNÉRALES



On connait un ensemble $\mathcal{P}$ de premiers

Quel genre d'ensemble?

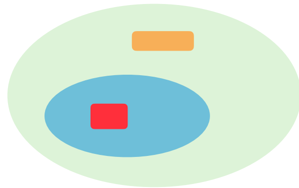
On verra

On dispose d'un ensemble $\mathcal{A}$ de randonnées

A priori, aucune raison que $\mathcal{A}$ *soit spécial*

Randonnées de longueur au plus ℓ $\quad \mathcal{A} \mapsto \mathcal{H}_\ell$

OBSERVATIONS GÉNÉRALES



On connait un ensemble $\mathcal{P}$ de premiers

Quel genre d'ensemble?

On verra

On dispose d'un ensemble $\mathcal{A}$ de randonnées

A priori, aucune raison que $\mathcal{A}$ soit spécial

Randonnées de longueur au plus ℓ $\quad \mathcal{A} \mapsto \mathcal{H}_\ell$

On veux le nombre $S(\mathcal{A}, \mathcal{P})$ d'éléments de $\mathcal{A}$ qui ne sont pas multiples de $\mathcal{P}$

Divisibilité droite ou gauche

Ex. si 2 & 3 ne commutaient pas alors $2 \nmid (3 \times 2 \times 3)$

Si p & p' divisent h, c.a.d. h = Xp = Yp' $\Longrightarrow$ p & p' commutent

p ne touche pas p' $\Longleftrightarrow$ pp' est auto évitant

CRIBLE: PREMIERS PAS

Si p & p' divisent h, c.a.d. h = Xp = Yp' $\Longrightarrow$ p & p' commutent

p ne touche pas p' $\Longleftrightarrow$ pp' est auto évitant

Identité de base : inclusion - exclusion !

$$S(\mathcal{H}_\ell,\mathcal{P}) = \sum_{d\in\mathcal{P}^{s.a}} \mu(d)\,|\mathcal{M}(d)_\ell|$$

Nombre de multiples de d dans $\mathcal{H}_\ell$

Tous les auto-évitants généré par p

Dans l'esprit de la théorie des nombres on cherche…

$$|\mathcal{M}(d)_\ell| = \mathrm{prob}(d)|\mathcal{H}_\ell| + r(d)$$

… telle que $\mathrm{prob}(d)$ soit multiplicative sur $\mathcal{H}$

$$pp' = p'p \Rightarrow \mathrm{prob}(pp') = \mathrm{prob}(p)\mathrm{prob}(p')$$

CRIBLE: PREMIERS PAS

$$S(\mathcal{H}_\ell, \mathcal{P}) = |\mathcal{H}_\ell| \sum_{d \in \mathcal{P}^{s.a}} \underbrace{\mu(d)\,\mathrm{prob}(d)} + \sum_{d \in \mathcal{P}^{s.a}} \mu(d) r(d)$$

Si prob est multiplicative et puisque *d* est auto-évitante

$$d = p_1 p_2 \cdots p_k \implies \mathrm{prob}(d) = \mathrm{prob}(p_1)\mathrm{prob}(p_2)\cdots\mathrm{prob}(p_k)$$

Quelle est prob(p) ?

CRIBLE: PREMIERS PAS

$\forall h \in \mathcal{H}:\ \ hp$ est multiple de *p*

$$\ell(hp) = \ell(h) + \ell(p) \Rightarrow hp \in \mathcal{M}(p)_\ell \iff h \in \mathcal{H}_{\ell-\ell(p)}$$

p a autant de multiples de longueur ℓ qu'il y a de randonnées de longueur $\ell - \ell(p)$, $|\mathcal{M}(p)_\ell| = |\mathcal{H}_{\ell-\ell(p)}|$ donc

$$\mathrm{prob}(p) + \frac{r(p)}{|\mathcal{H}_\ell|} = \frac{|\mathcal{H}_{\ell-\ell(p)}|}{|\mathcal{H}_\ell|}$$

HYPOTHÈSE TEMPORAIRE

On peux trouver $\mathrm{prob}(p)$ multiplicative et asymptotiquement "optimale", c'est à dire

$$\lim_{\ell\to\infty} \frac{r(p)}{|\mathcal{H}_\ell|} = 0 \qquad \boxed{\mathrm{prob}(p) = \lim_{\ell\to\infty} \frac{|\mathcal{H}_{\ell-\ell(p)}|}{|\mathcal{H}_\ell|}}$$

DÉNOMBRER LES RANDONNÉES

SUPPOSITION FONDAMENTALE

Il existe une constante Λ et une fonction $f : \mathbb{N} \to \mathbb{R}$ bornée telle que $\lim_{\ell \to \infty} f(\ell)$ existe et

$$|\mathcal{H}_\ell| = \Lambda^\ell f(\ell)$$

VRAIE POUR LA LONGUEUR

En passant le spectre !

$$|\mathcal{H}_\ell| = \zeta(z)[z^\ell] = \frac{1}{\det(\mathsf{I} - z\mathsf{A}_G)}[z^\ell]$$

donc $\Lambda = \lambda_G$ et

$$\lim_{\ell \to \infty} f(\ell) = \lim_{z \to \frac{1}{\lambda}} (1 - z\lambda)\zeta(z)$$

CRIBLE: PREMIERS PAS

- On identifie

$$\mathrm{prob}(p) = \lim_{\ell\to\infty} \frac{|\mathcal{H}_{\ell-\ell(p)}|}{|\mathcal{H}_\ell|} = \lim_{\ell\to\infty} \frac{\lambda^{\ell-\ell(p)} f\big(\ell-\ell(p)\big)}{\lambda^\ell f(\ell)} = \lambda^{-\ell(p)}$$

- Est bien multiplicative !

$$r(p) = \lambda^{\ell-\ell(p)} \left(f\big(\ell-\ell(p)\big) - f(\ell)\right)$$

$$\lim_{\ell\to\infty} \frac{r(p)}{|\mathcal{H}_\ell|} = \lim_{\ell\to\infty} \lambda^{-\ell(p)} \frac{f\big(\ell-\ell(p)\big) - f(\ell)}{f(\ell)} = 0$$

- Prob(.) est asymptotiquement optimale !

Existence de la limite de *f*

CHOISISSONS LES PREMIERS

$$S(\mathcal{H}_\ell, \mathcal{P}) = |\mathcal{H}_\ell| \sum_{d \in \mathcal{P}^{\text{s.a.}}} \mu(d)\lambda^{-\ell(d)} + \lambda^\ell \sum_{d \in \mathcal{P}^{\text{s.a.}}} \mu(d)\lambda^{-\ell(d)}\big(f(\ell - \ell(d)) - f(\ell)\big)$$

Pas de progrès possible sans plus de précisions sur $\mathcal{P}$

CHOISISSONS LES PREMIERS

$$S(\mathcal{H}_\ell, \mathcal{P}) = |\mathcal{H}_\ell| \sum_{d \in \mathcal{P}^{\mathrm{s.a.}}} \mu(d) \lambda^{-\ell(d)} + \lambda^\ell \sum_{d \in \mathcal{P}^{\mathrm{s.a.}}} \mu(d) \lambda^{-\ell(d)} \big(f(\ell - \ell(d)) - f(\ell) \big)$$

Pas de progrès possible sans plus de précisions sur $\mathcal{P}$

QUE CHOISIR POUR $\mathcal{P}$?

- Tous les premiers de longueur $\ell(p) \leq L$ ← *NP-difficile*
- Tous les premiers d'un sous graphe $H \prec G$

$$\mathcal{P} \equiv \mathcal{P}_H$$

SIMPLIFICATIONS

$$S(\mathcal{H}_\ell, \mathcal{P}_H) = |\mathcal{H}_\ell| \sum_{d \in \mathcal{P}_H^{\text{s.a.}}} \mu(d)\lambda^{-\ell(d)} + \lambda^\ell \sum_{d \in \mathcal{P}_H^{\text{s.a.}}} \mu(d)\lambda^{-\ell(d)} \big(f(\ell - \ell(d)) - f(\ell)\big)$$

$$\det\left(\mathsf{I} - \frac{1}{\lambda}\mathsf{A}_H\right)$$

$$f(\ell - \ell(d)) - f(\ell) = \sum_{k \geq 1} \frac{\nabla^k[f](\ell)}{k!} \big(\ell(d)\big)_{(k)}$$

SIMPLIFICATIONS

$$S(\mathcal{H}_\ell,\mathcal{P}_H)=|\mathcal{H}_\ell|\sum_{d\in\mathcal{P}_H^{\text{s.a.}}}\mu(d)\lambda^{-\ell(d)}+\lambda^\ell\sum_{d\in\mathcal{P}_H^{\text{s.a.}}}\mu(d)\lambda^{-\ell(d)}\big(f(\ell-\ell(d))-f(\ell)\big)$$

$$\det\left(\mathsf{I}-\frac{1}{\lambda}\mathsf{A}_H\right)$$

$$f(\ell-\ell(d))-f(\ell)=\sum_{k\geq 1}\frac{\nabla^k[f](\ell)}{k!}\big(\ell(d)\big)_{(k)}$$

$$S(\mathcal{H}_\ell,\mathcal{P}_H)=|\mathcal{H}_\ell|\,\det\left(\mathsf{I}-\frac{1}{\lambda}\mathsf{A}_H\right)+\lambda^\ell\sum_{k\geq 1}\frac{\nabla^k[f](\ell)}{k!}\det{}^{(k)}\left(\mathsf{I}-\frac{1}{\lambda}\mathsf{A}_H\right)$$

Notation : $\det{}^{(k)}\left(\mathsf{I}-\frac{1}{\lambda}\mathsf{A}_H\right)=\frac{d^k}{dz^k}\det\left(\mathsf{I}-z\mathsf{A}_H\right)\Big|_{z=1/\lambda}$

CRIBLE : RÉSULTAT

$$S(\mathcal{H}_\ell, \mathcal{P}_H) = |\mathcal{H}_\ell| \det\left(\mathsf{I} - \frac{1}{\lambda}\mathsf{A}_H\right) + \lambda^\ell \sum_{k\geq 1} \frac{\nabla^k[f](\ell)}{k!} \det{}^{(k)}\left(\mathsf{I} - \frac{1}{\lambda}\mathsf{A}_H\right)$$

- Premier terme domine asymptotiquement !

$$\frac{S(\mathcal{H}_\ell, \mathcal{P}_H)}{|\mathcal{H}_\ell|} \sim \det\left(\mathsf{I} - \frac{1}{\lambda}\mathsf{A}_H\right) \qquad \ell \to \infty$$

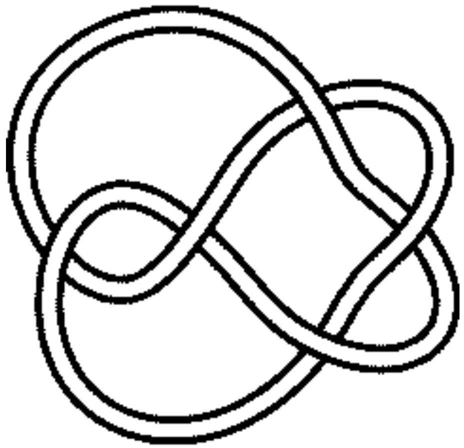
Randonnées qui ne sont pas multiple des premiers connus

- Si $H = G\backslash p$ alors $\det\left(\mathsf{I} - \frac{1}{\lambda}\mathsf{A}_{G\backslash p}\right)$ est la fraction asymptotique de rando.

qui sont des chemins multiples à droite de *p*

CRIBLES : PERSPECTIVES

- Valide pour tout rang r sur les randonnées



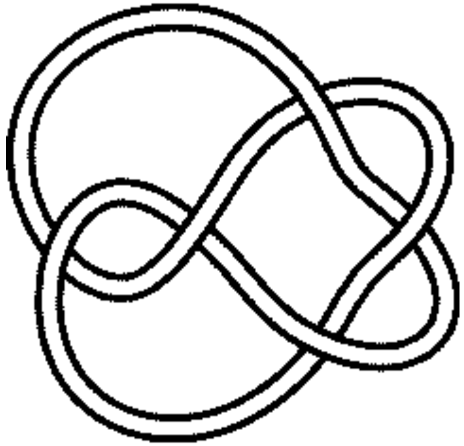
$\rho(h)$ nombre minimal de facteurs auto-évitants d'une randonnée

$$n = p_1^{k_1} p_2^{k_2} \cdots p_n^{k_n} \Longrightarrow \rho(n) = \max\{k_i\}$$

Pas de déterminant pour $r \neq \ell$ sauf cas où $\Lambda_r = 1$

CRIBLES : PERSPECTIVES

- Valide pour tout rang r sur les randonnées



$\rho(h)$ nombre minimal de facteurs auto-évitants d'une randonnée

$$n = p_1^{k_1} p_2^{k_2} \cdots p_n^{k_n} \implies \rho(n) = \max\{k_i\}$$

Pas de déterminant pour $r \neq \ell$ sauf cas où $\Lambda_r = 1$

- La SUPPOSITION FONDAMENTALE s'étend à certains graphes infinis

Travail en densité + opérateurs de troncations

- Crible de Brun permet des encadrements

PLAN

1. Contexte

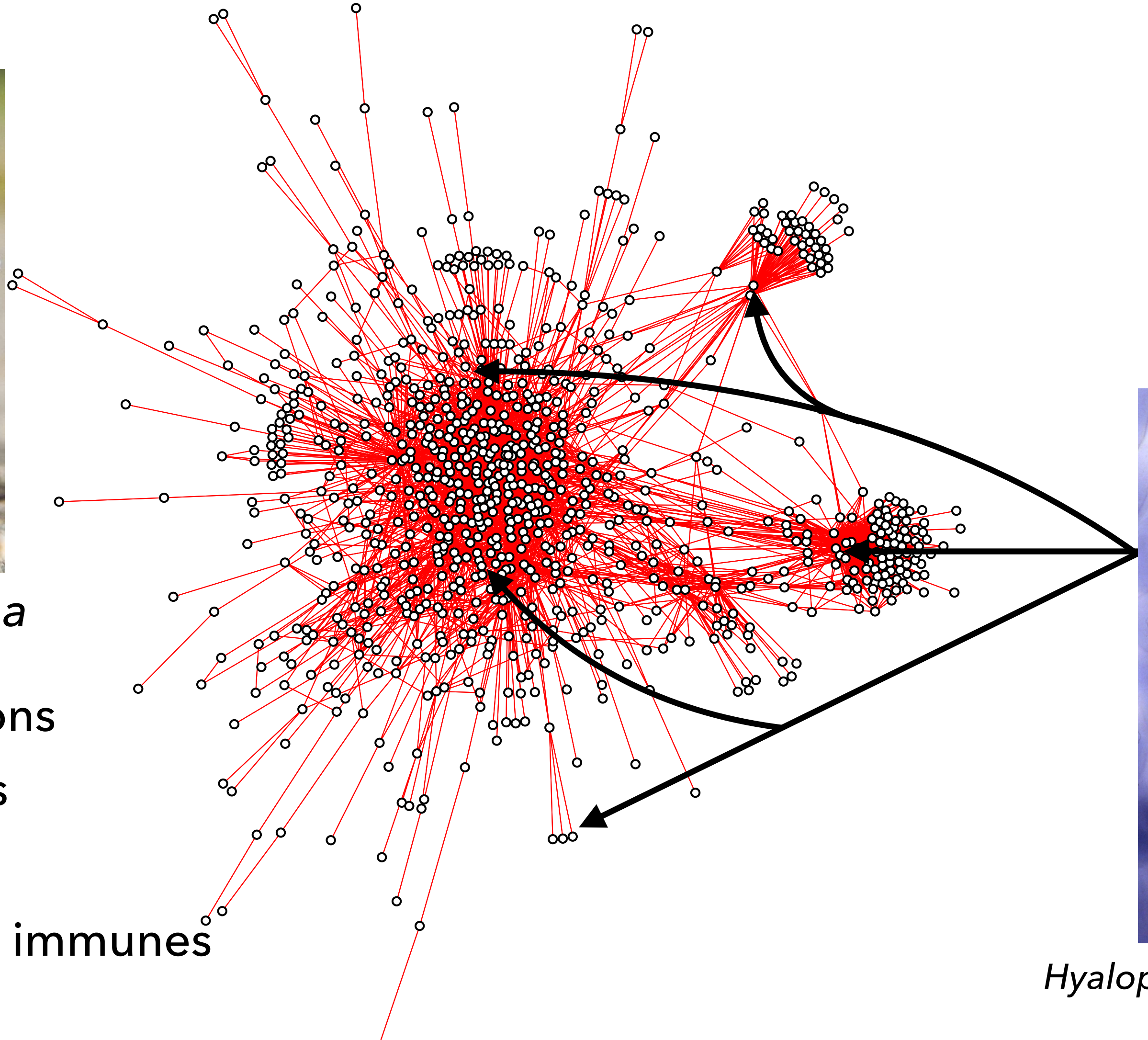
2. La version la plus simple …
… du crible le plus simple

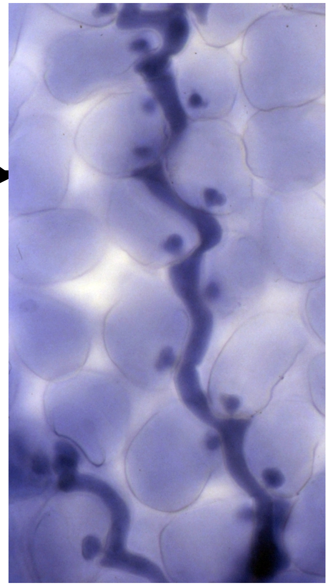
3. En biologie des systèmes

BIOLOGIE DES SYSTÈMES

- Problème : quelles protéines sont ciblées par les pathogènes?

Arabidopsis thaliana





Hyaloperonospora arabidopsidis

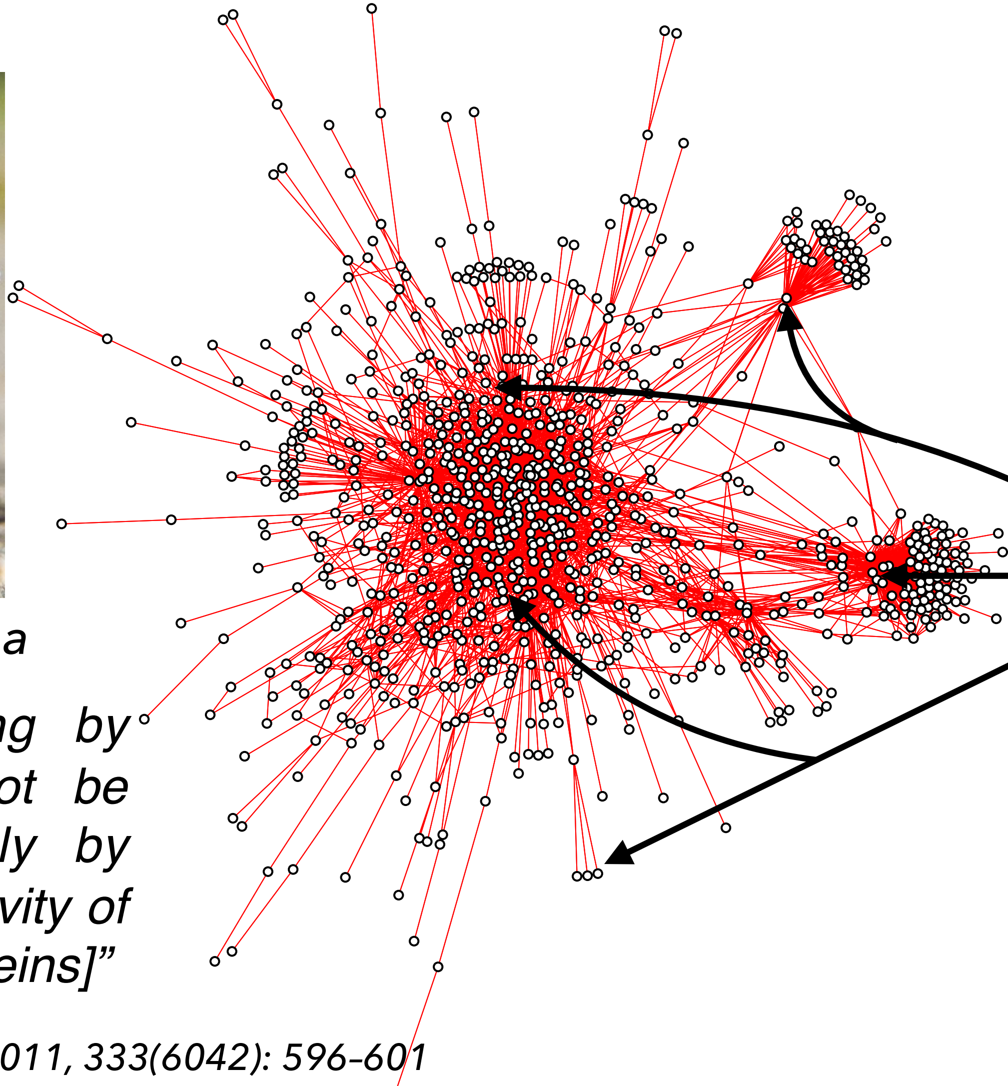
3,148	interactions
926	protéines
137	cibles
170	réactions immunes

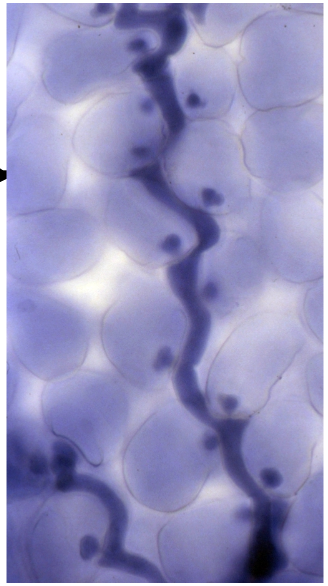
BIOLOGIE DES SYSTÈMES

- Model 1 (Mukhtar 2011) : les cibles sont les hubs du réseau

Arabidopsis thaliana

"protein targeting by pathogens cannot be explained merely by the high connectivity of those target [proteins]"





Hyaloperonospora arabidopsidis

Mukhthar et al., Science. 2011, 333(6042): 596-601

BIOLOGIE DES SYSTÈMES

- Faits : Pathogènes stimulent l'immunité
  Les protéines les plus centrales sont liées à l'immunité

- Hypothèse : Triades "Cible - Centrale - Immune" (CCI) existent et celles qui sont ciblées interceptent le plus de flots sur le réseau

Flots de réactions entre protéines passant par $\triangledown$ = nb. de chemins multiples de $\triangledown$

Crible !

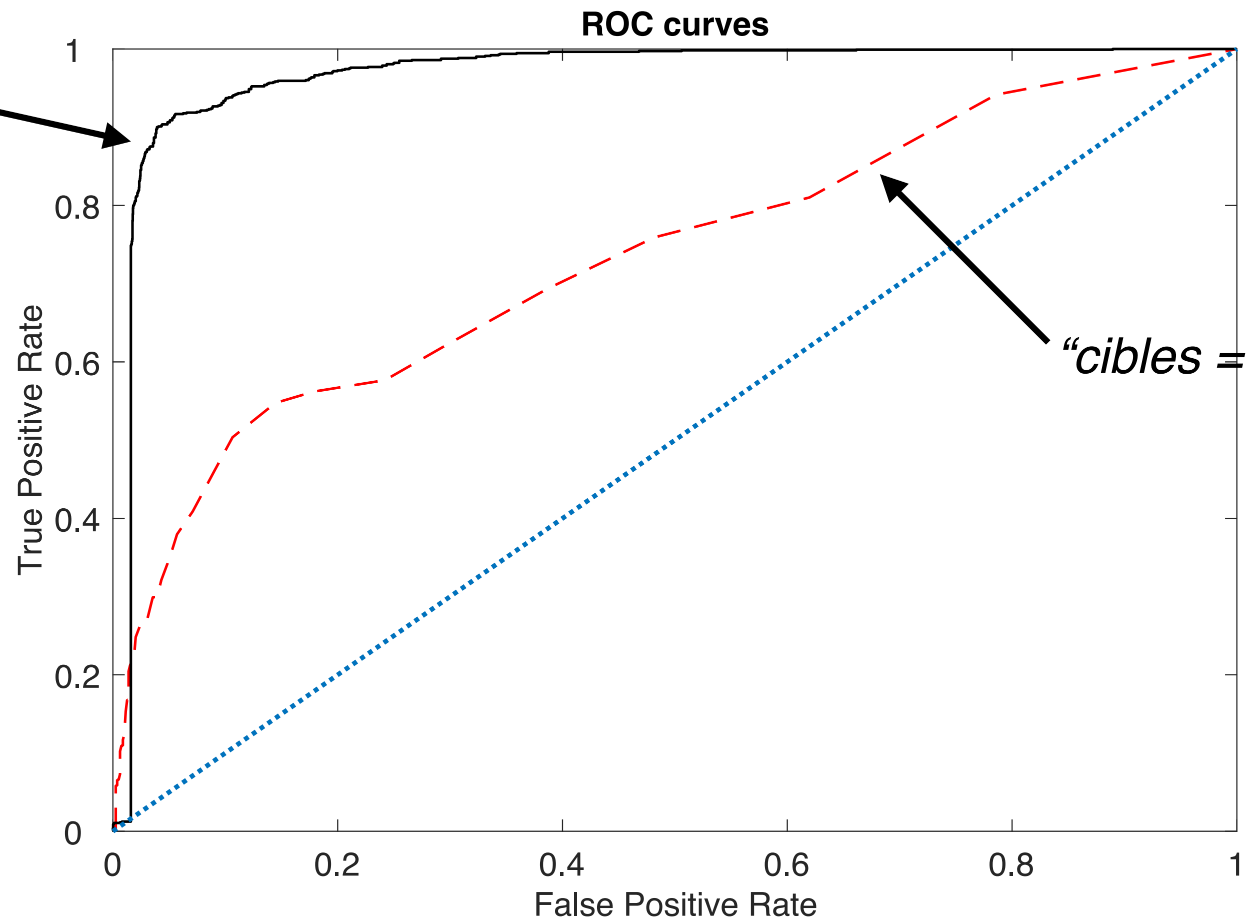
$$f = \det\left(\mathsf{I} - \frac{1}{\lambda}\mathsf{A}_{G\backslash\triangledown}\right)$$

Fraction asymptotique de randonnées qui sont des chemins multiples d'un premier donné

- Model: cibles sont dans des triades CCI et sont f - dominantes

BIOLOGIE DES SYSTÈMES

“pathogènes maximisent la fraction de flots de réactions dans l’hôte qu’ils interceptent”



<u>Bonus:</u> nous avons identifié 2 protéines voisines de 70% de toutes les cibles

Giscard, Wilson, “*Loop-centrality in Complex Networks*” dans Complex Networks & Their Applications VI, Studies in Computational Intelligence, vol 689. Springer (2017).

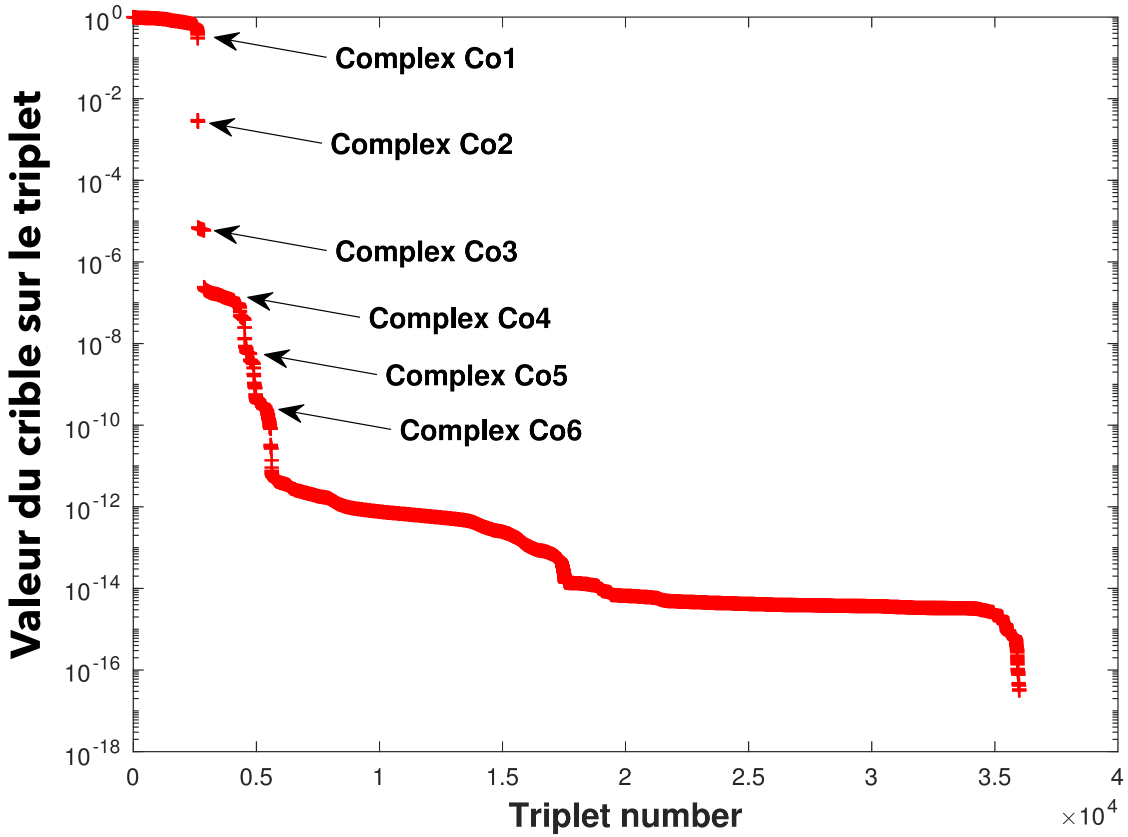
BIOLOGIE DES SYSTÈMES

- Réseau d'interactions protéines -protéines d'une levure

5,303 interactions
1,689 protéines

$$f = \det\left(\mathsf{I} - \frac{1}{\lambda}\mathsf{A}_{G\backslash\triangledown}\right)$$

Fraction asymptotique de toutes les chaines d'interactions entre protéines qui passent par le triplet donné



Giscard, Wilson, "*A Centrality Measure for Cycles and Subgraphs II*" à paraître dans Applied Network Science (2018)

SUR LES CRIBLES

- Stratégie globale pour :

Quelle est la fraction (asymptotique) des randonnées de rang r dans un ensemble $\mathcal{A}$ *satisfaisant une propriété donnée ?*

- Inclut des questions sur les chemins

- Brun, Selberg faisables mais plus difficiles
- Graphes infinis faisables *en densité avec des troncations*

MERCI !