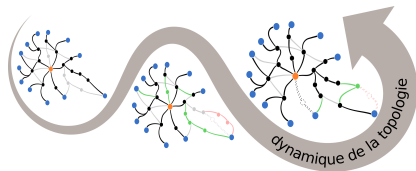


# Analyse empirique et modélisation de la dynamique de la topologie de l'Internet

Sergey Kirgizov

*Directrice de thèse: Clémence Magnien*

Complex Networks, LIP6, (UPMC, CNRS)



Paris, 12 décembre 2014

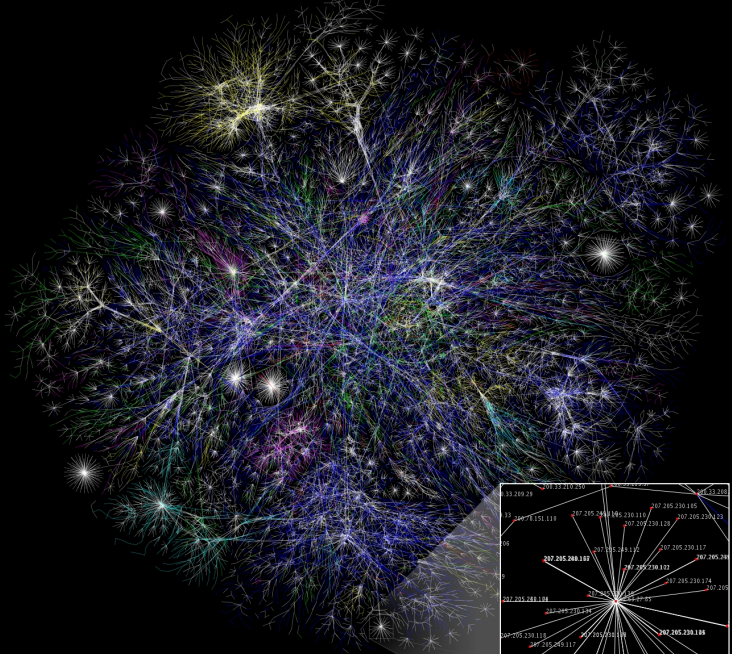
# Plan

---

- 1 Introduction
- 2 Mesures égo-centrées
- 3 Modèle
- 4 Caractérisation de la dynamique
- 5 Taille du sous-graphe observable
- 6 Dynamique réelle et dynamique observée
- 7 Conclusion et perspectives

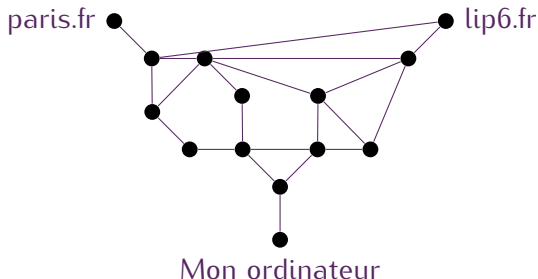
# Introduction :

l'Internet, étude de la topologie de l'Internet



Carte de l'Internet // Barrett Lyon, 2003

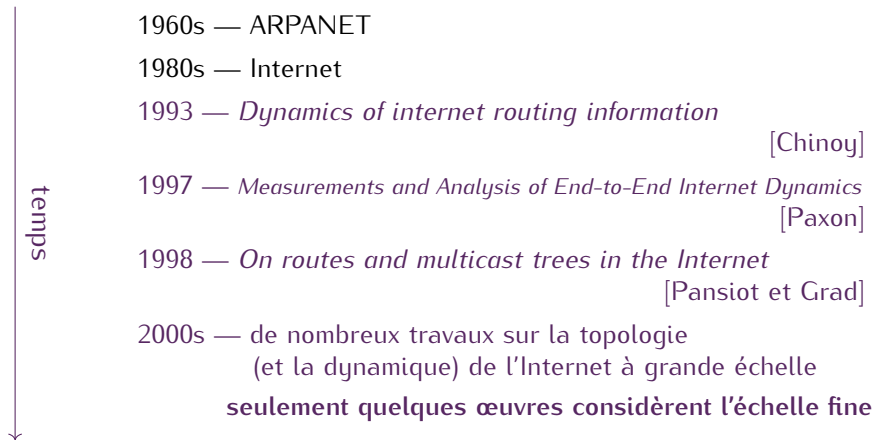
# Topologie au niveau IP



Nœuds : ordinateurs

Liens : connections IP

# Petite note historique



# Étude de la topologie de l'Internet

---

## Pas de carte officielle de l'Internet

Une carte statique ?

Mesures longues et coûteuses;  
Biais sur la structure observée. }  $\implies$  Pas de carte fiable

Étude de la dynamique ?

*Tous les problèmes statiques;*  
Répéter les mesures }  $\implies$  Pas si facile

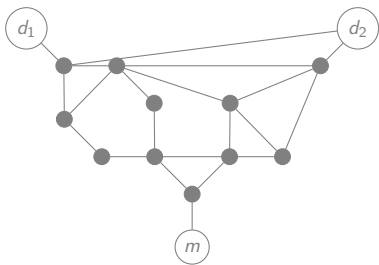
# Mesures égo-centrées :

Routes entre un moniteur et plusieurs destinations fixes

[Latapy, Magnien, Ouédraogo, 2011]

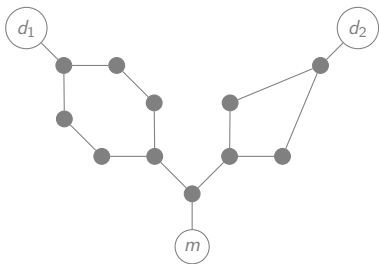


# Vues ego-centrées



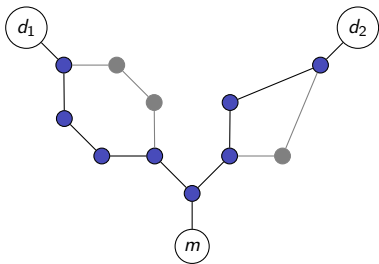
Carte complète

# Vues ego-centrées



Routes entre un moniteur et des destinations

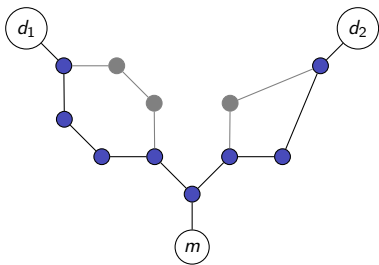
# Vues ego-centrées



Une mesure par l'outil tracetree

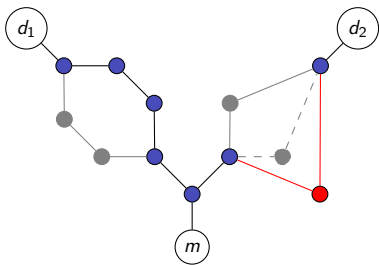
[Latapy, Magnien, Ouédraogo, 2011]

# Vues ego-centrées



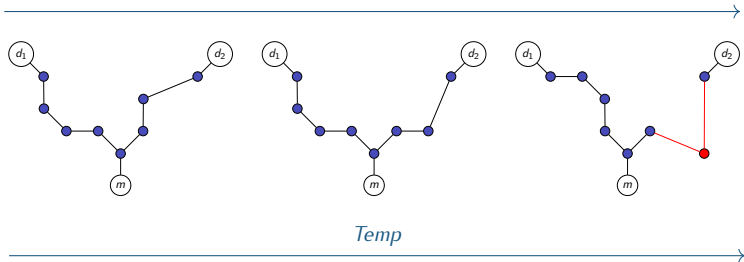
Une autre mesure par l'outil tracetree

# Vues ego-centrées



Une autre mesure par l'outil tracetree

# Dynamique des vues ego-centrées

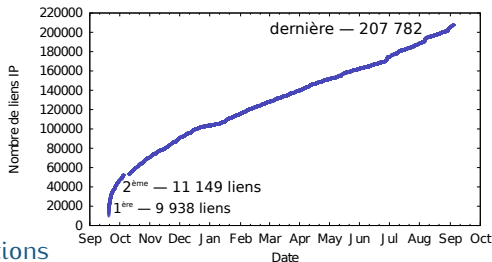


Mesures périodiques rapides  $\implies$  étude de la dynamique

# Caractérisation de la dynamique

Mesures :  $M_1, M_2, \dots, M_i, \dots$

Nombre de liens vus depuis le début :  $t \mapsto \left| \bigcup_{i \in [0, t]} M_i \right|$  # liens



3000 destinations

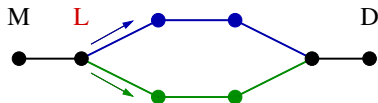
une mesure toutes les 15 minutes pendant 1 an

découverte en permanence de nouveaux liens IP  
à une vitesse élevée

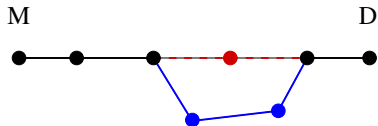
# Causes de la dynamique observée

## Causes

- *Load-balancing*

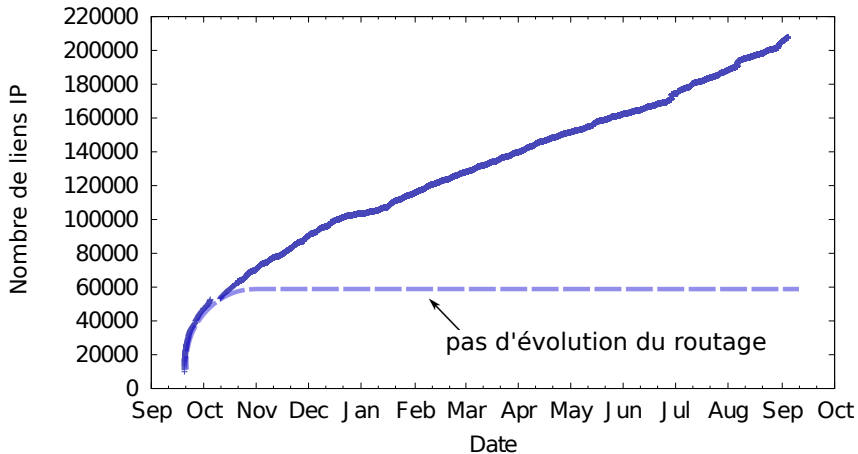


- *Évolution du routage*

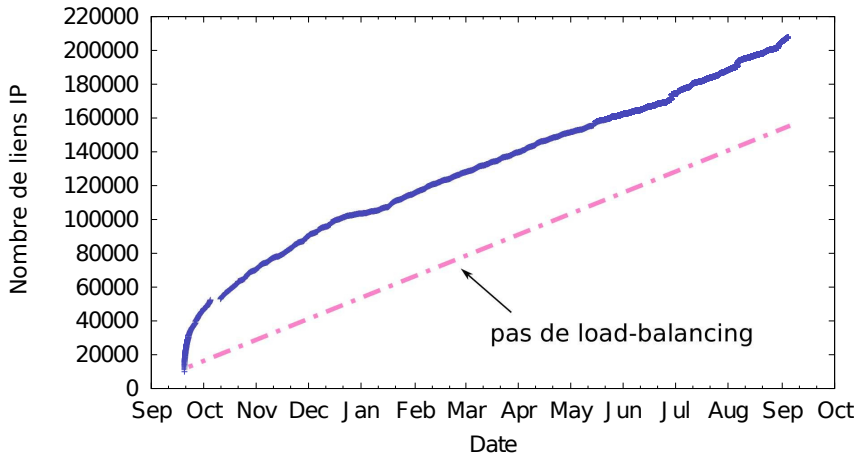




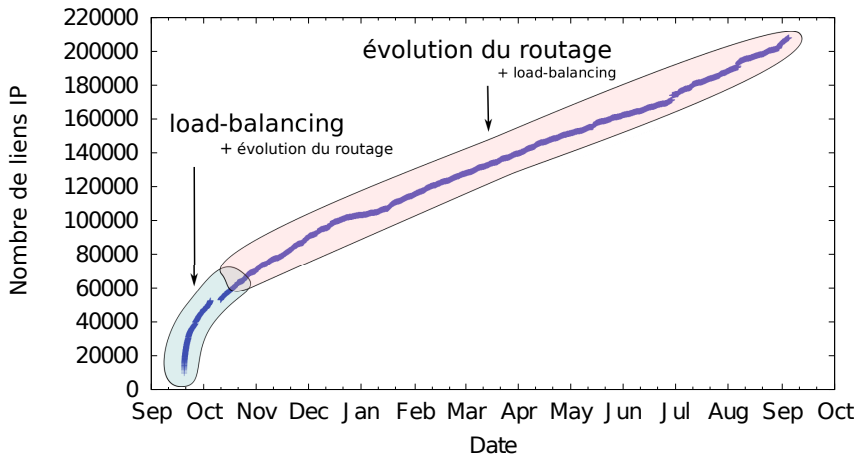
# Load-balancing vs évolution du routage



# Load-balancing vs évolution du routage



# Load-balancing vs évolution du routage



# Modèle :

*Load-balancing* + évolution du routage

# Modélisation

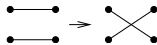
Topologie : Graphe aléatoire (Erdős-Rényi ou configuration model)

Un nœud aléatoire  $\rightarrow$  moniteur  
 $d$  nœuds aléatoires  $\rightarrow$  destinations

Mesure : plus courts chemins (aléatoires)

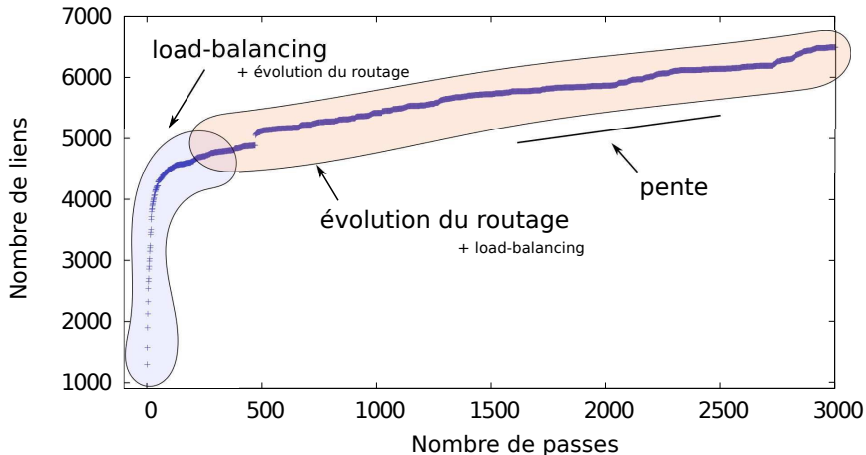


Evolution du routage : “échange de liens”



# Load-balancing vs évolution du routage

Nombre de liens distincts vus depuis le début de la mesure



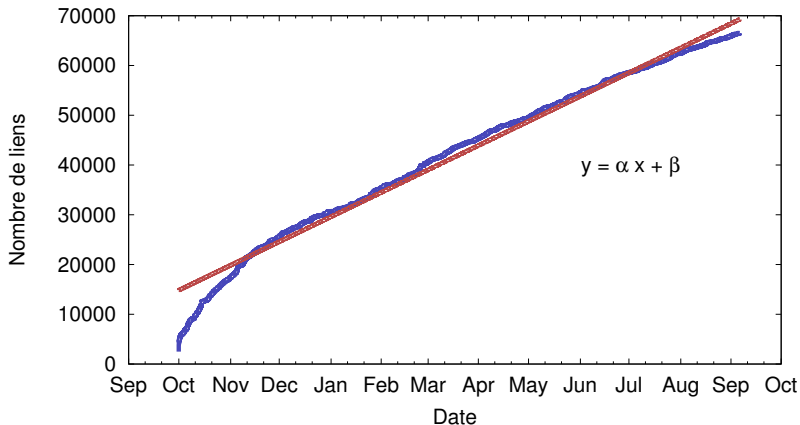
# Caractérisation de la dynamique :

---

la pente d'évolution

[Magnien, Medem, Kirgizov et Tarissan, 2013]

# La pente

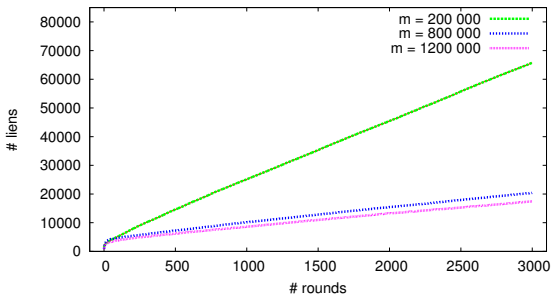


$\alpha$  est la pente de la partie linéaire



# $\alpha = f(\text{paramètres du modèle})$

nombre total de liens

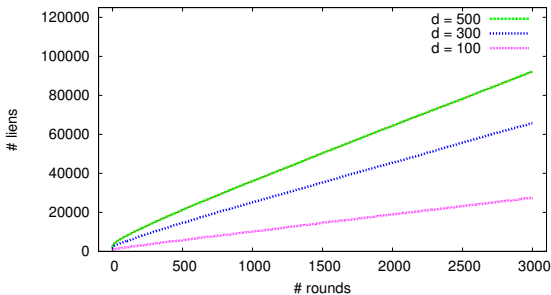


Impact du nombre total de liens

Graphes Erdős–Rényi ( $n = 100\,000$ ,  $d = 100$ ,  $sw = 100$ )

# $\alpha = f(\text{paramètres du modèle})$

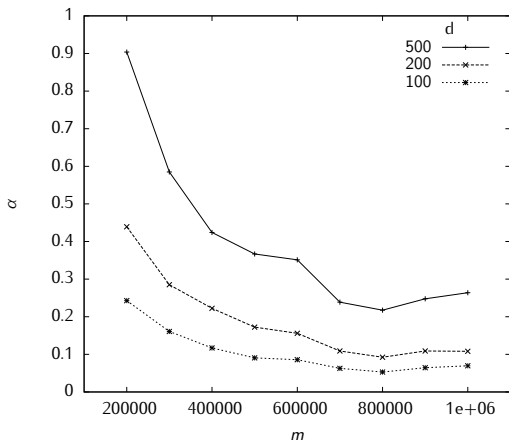
nombre total de destinations



Impact du nombre de destinations

Graphes Erdős–Rényi ( $n = 100\ 000$ ,  $m = 200\ 000$ ,  $sw = 100$ )

$$\alpha = f(\text{paramètres du modèle})$$

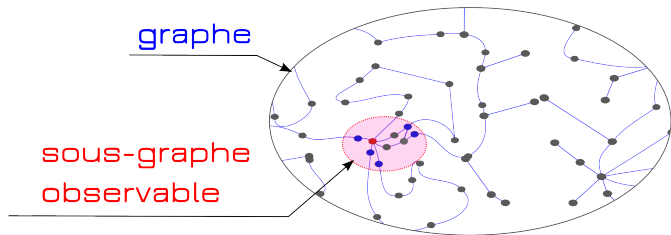


Un graphe aléatoire  $G(n, m)$   
avec  $n$  nœuds et  $m$  liens

$d$  nœuds comme destinations

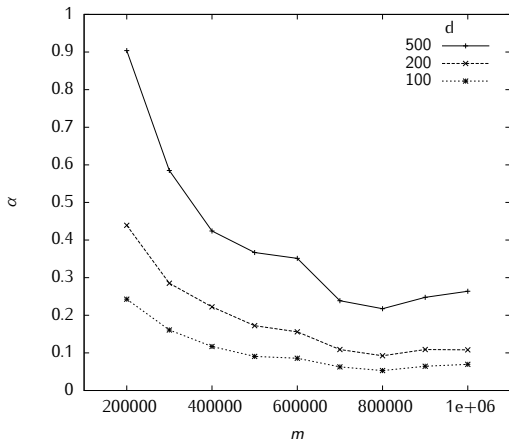
# La taille du sous-graphe des plus courts chemins

cette taille augmente avec le nombre de destinations



$$\alpha \propto \frac{|\text{sous-graphe}|}{|\text{graphe}|} = \frac{S_I}{m}$$

## Impact de la taille du sous-graphe des plus courts chemins

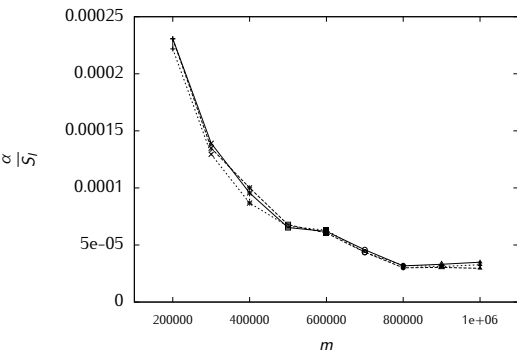


Un graphe aléatoire  $G(n, m)$   
avec  $n$  nœuds et  $m$  liens

$d$  nœuds aléatoires comme  
destinations

$$\alpha \approx K \frac{S_l}{m}$$

## Impact de la taille du sous-graphe des plus courts chemins



Un graphe aléatoire  $G(n, m)$   
avec  $n$  nœuds et  $m$  liens

$d$  nœuds aléatoires comme  
destinations

$$\alpha \approx K \frac{S_l}{m}$$

# Caractérisation de la dynamique

## Conclusion

---

$$\text{Erdős-Rényi : } \alpha \approx K \frac{S_l}{m} \quad \text{Power-law : } \alpha \approx K' \frac{S_l}{m}$$

Prévoir la pente sans exécuter les simulations coûteuses

Quelques comportements non linéaires et non-monotones

Comparison aux mesures réelles

Probabilité d'un changement à la place de la pente

# La taille du sous-graphe des plus courts chemins :

---

1 moniteur, 1 destination

[Kirgizov et Magnien, soumis]



# Graphes aléatoires

---

$G(n, p)$  est un graphe aléatoire à  $n$  sommets. Chacune des  $n(n - 1)/2$  arêtes est présente avec probabilité  $p$ .

1951–1952 *Solomonoff et Rapoport*

1959 *Gilbert, Erdős et Rényi*

**SPS** : sous-graphe des plus courts chemins

**S** : taille du sous-graphe des plus courts chemins

# Graphes aléatoires

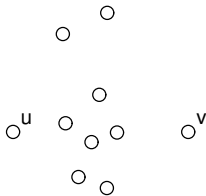
$G(n, p)$  est un graphe aléatoire à  $n$  sommets. Chacune des  $n(n - 1)/2$  arêtes est présente avec probabilité  $p$ .

1951–1952 *Solomonoff et Rapoport*

1959 *Gilbert, Erdős et Rényi*

**SPS** : sous-graphe des plus courts chemins

**S** : taille du sous-graphe des plus courts chemins



# Graphes aléatoires

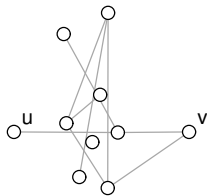
$G(n, p)$  est un graphe aléatoire à  $n$  sommets. Chacune des  $n(n - 1)/2$  arêtes est présente avec probabilité  $p$ .

1951–1952 *Solomonoff et Rapoport*

1959 *Gilbert, Erdős et Rényi*

**SPS** : sous-graphe des plus courts chemins

**S** : taille du sous-graphe des plus courts chemins



# Graphes aléatoires

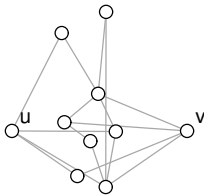
$G(n, p)$  est un graphe aléatoire à  $n$  sommets. Chacune des  $n(n - 1)/2$  arêtes est présente avec probabilité  $p$ .

1951–1952 *Solomonoff et Rapoport*

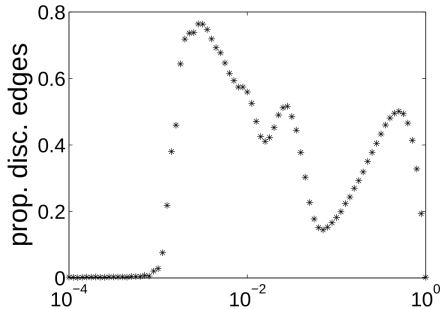
1959 *Gilbert, Erdős et Rényi*

**SPS** : sous-graphe des plus courts chemins

**S** : taille du sous-graphe des plus courts chemins



## Proportion des liens qui se trouvent sur les plus courts chemins



Graphe Erdős-Rényi ( $n = 1000$ ).

Proportion des liens sur les plus courts chemins.

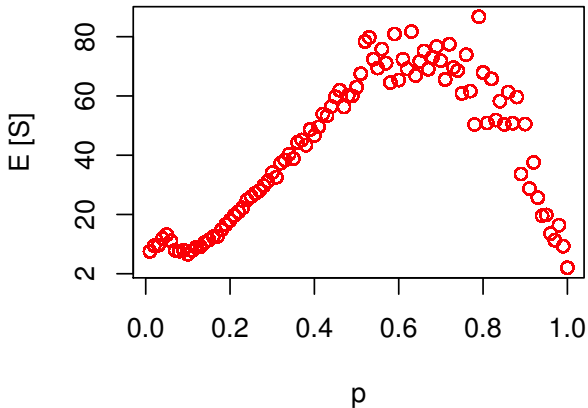
Tous les noeuds sont les destinations.

[Guillaume et Latapy, 2005]

[Blondel, Guillaume, Hendrickx et Jungers, 2007]

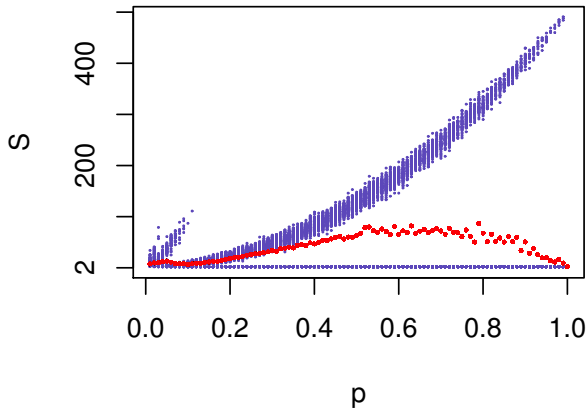
## Taille moyenne de SPS dans des graphes aléatoires

Fixons  $n$   
 $p \mapsto \mathbb{E}[S]$



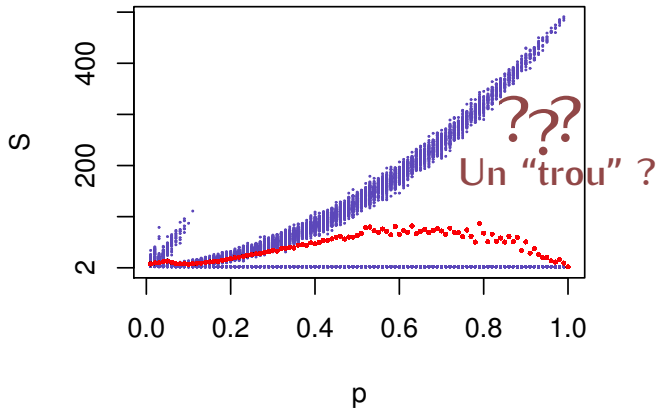
## Moyenne et tailles réelles de SPS

Fixons  $n$   
 $p \mapsto \{S\}$



## Moyenne et tailles réelles de SPS

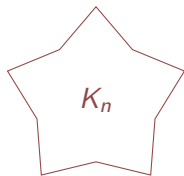
Fixons  $n$   
 $p \mapsto \{S\}$





# Graphe complet : $K_n$

$$p = 1$$

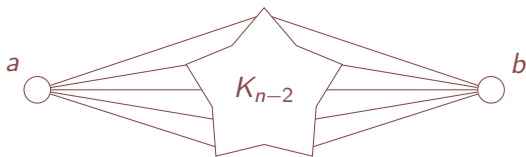


$$SPS(u, v) = \begin{array}{c} u \qquad v \\ \circ \text{---} \circ \end{array}$$

$$S(u, v) = 2$$

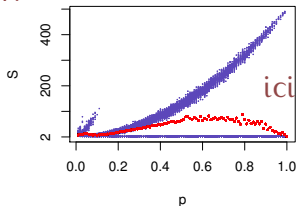
# Graphe quasi-complet : $K_n - ab$

$$p = 1 - \varepsilon$$

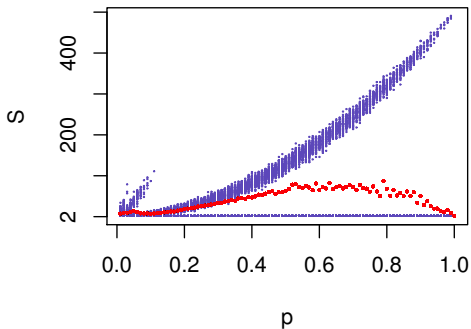


$$S(u, v) = \begin{cases} n & \text{si } \{u, v\} = \{a, b\}, \\ 2 & \text{sinon.} \end{cases}$$

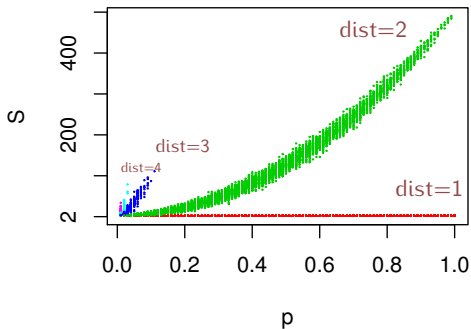
Si  $n > 3$ , alors il y a un "trou".



## Taille du SPS dans des graphes aléatoires

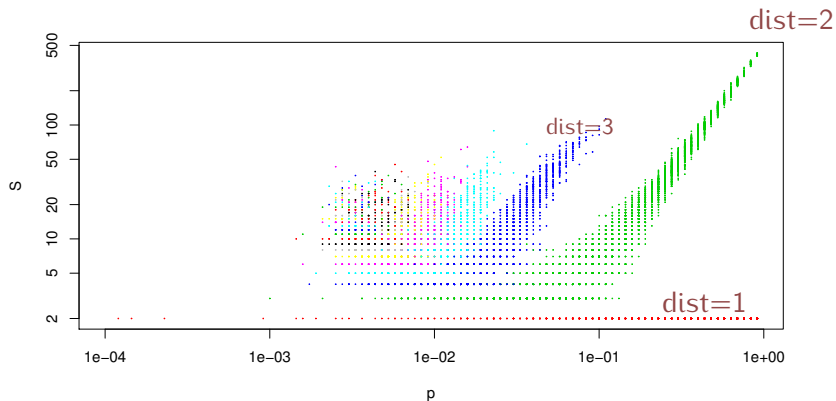


## Taille du SPS dans des graphes aléatoires



Y: nombre de nœuds qui sont directement liés à la fois à  $u$  et  $v$   
 $Y \sim B(n - 2, p^2)$

# Taille du SPS dans des graphes aléatoires échelle logarithmique



# Conclusion

---

Prévoir la taille du sous-graphe observable sans exécuter les simulations coûteuses

Oscillations ont été expliquées

$\text{dist}(u,v) = 2$  : la loi exacte

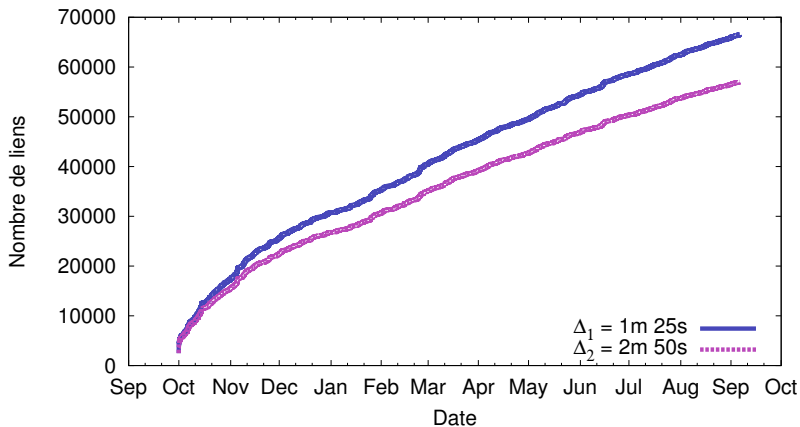
$\text{dist}(u,v) > 2$  : une approximation de la taille attendue

# Dynamique réelle et dynamique observée :

[Kirgizov, Magnien, Tarissan, Liu, 2013]

[Kirgizov et Queyroi, preprint, 2014]

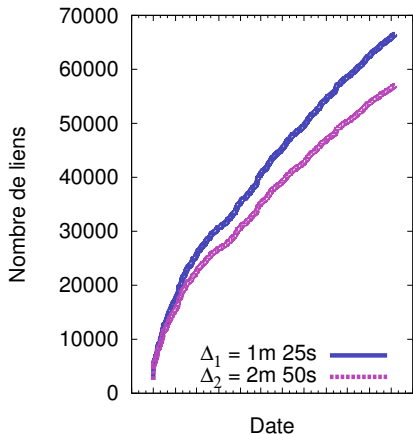
$$\alpha = f(\Delta)$$



$\alpha = f(\Delta)$ , où  $\Delta$  est le délai entre les mesures  
 $\alpha(\Delta_1) > \alpha(\Delta_2)$



$$\alpha = f(\Delta)$$



Réalité

$A \rightarrow \dots \rightarrow B \rightarrow \dots \rightarrow C$

$\Delta_1$

$A \rightarrow B \rightarrow C$

$\Delta_2$

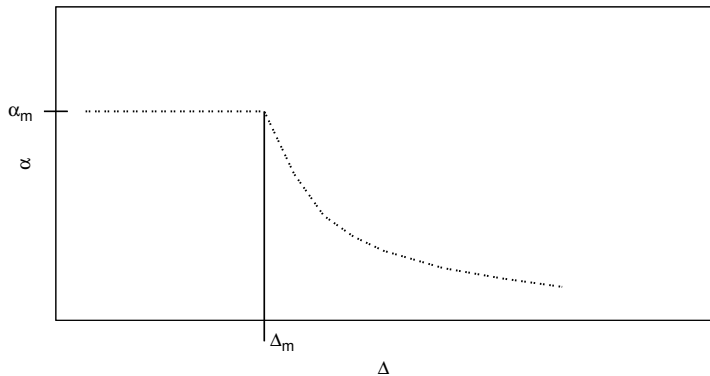
$A \rightarrow C$

$$|A \cup C| \leq |A \cup B \cup C|$$

$$\Delta_1 \leq \Delta_2 \Rightarrow$$

$$\Rightarrow \alpha(\Delta_1) \geq \alpha(\Delta_2)$$

# Situation attendue

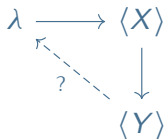


$\alpha_m$  est la pente maximale possible

# Autre approche. Estimation du taux de changements



Question : estimer le taux des changements réels  
Poisson paramétré par  $\lambda$



# Estimateur

---

$$\hat{\lambda} = -\Delta^{-1} \log \frac{W}{N},$$

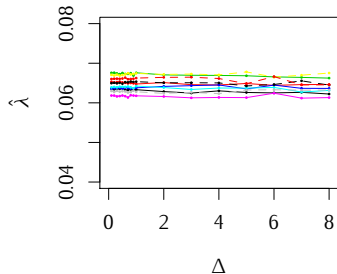
où

- $\Delta$  – taille de l'intervalle de temps entre les observations ;
- $W$  – nombre d'intervalles sans changement ;
- $N$  – nombre total d'intervalles .

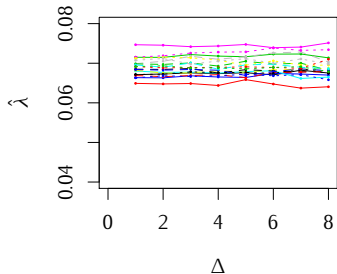
$$\mathbb{E} [\hat{\lambda} - \lambda] \approx \frac{1}{2N\Delta} (e^{\lambda\Delta} - 1)$$

$$\lim_{N \rightarrow \infty} \mathbb{E} [\hat{\lambda}] = \lambda$$

# $\lambda$ estimée en fonction de $\Delta$

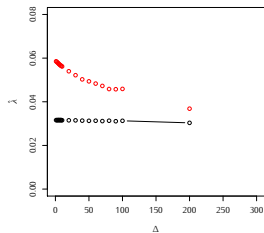


Processus de Poisson

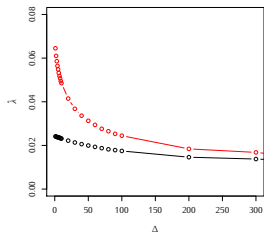


Modèle avec un graphe sous-jacent Erdős-Rényi, SPS est mesurée à chaque round.

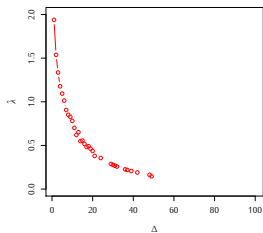
# $\lambda$ estimée en fonction de $\Delta$



Erdős-Rényi



Power-law



Mesures réelles

$\hat{\lambda}$  pour SPT (  ) et SPS (  ) séquences.

# Estimation de $\lambda$

## Conclusion

---

- ◆ Processus de Poisson  $\implies$  estimation possible
- ◆ Les autres cas ouvrent des directions de la recherche future.

# Conclusion

---

- ◆ Modèle de la dynamique de l'Internet
- ◆ Relations entre les paramètres du modèle et les caractéristiques de la dynamique
- ◆ La taille du sous-graphe des plus courts chemins
- ◆ Estimation du taux de changements



# Perspectives

---

- ◆ Améliorer le modèle
- ◆ Etude plus approfondie de la taille du sous-graphe observable
- ◆ Appliquer les méthodes développées à d'autres réseaux

# Merci beaucoup

<http://kirgizov.complexnetworks.fr/>