

# Structures et Dynamique des Réseaux

## Graphes dynamiques

Clémence Magnien, Lionel Tabourier, Fabien Tarissan

LIP6 – CNRS and Université Pierre et Marie Curie

`prenom.nom@lip6.fr`

- 1 « Mesure » de la dynamique
  - Contexte et approche
  - Tracetree
  - Mesures
- 2 Analyse de la dynamique
  - Propriétés élémentaires
  - Croissance
  - Présence et blocs de présence
- 3 Modélisation de la dynamique
  - Impact de la dynamique
  - Modélisation
  - Simulations

# Outline

- 1 « Mesure » de la dynamique
  - Contexte et approche
  - Tracetree
  - Mesures
  
- 2 Analyse de la dynamique
  - Propriétés élémentaires
  - Croissance
  - Présence et blocs de présence
  
- 3 Modélisation de la dynamique
  - Impact de la dynamique
  - Modélisation
  - Simulations

## Principaux travaux de référence

Clémence Magnien, Frédéric Ouédraogo, Guillaume Valadon  
et Matthieu Latapy

Fasts dynamics in Internet topology: observations and first explanations  
Fourth International Conference on Internet Monitoring and Protection, 2009.

Matthieu Latapy, Clémence Magnien et Frédéric Ouédraogo

A Radar for the internet  
Complex Systems, 2011.

Clémence Magnien, Amélie Medem, Sergey Kirgizov et Fabien Tarissan

Towards realistic modeling of IP-level routing topology dynamics.  
Journal of Networking Science, 2013.

# Outline

- 1 « Mesure » de la dynamique
  - Contexte et approche
    - Tracetree
    - Mesures
- 2 Analyse de la dynamique
  - Propriétés élémentaires
  - Croissance
  - Présence et blocs de présence
- 3 Modélisation de la dynamique
  - Impact de la dynamique
  - Modélisation
  - Simulations

# Propriétés inattendues

[Sigcomm 99, Infocom 00, Science 99, ...]

## Propriétés inattendues et importantes

Degrés des nœuds (nombre de liens) hétérogènes

## Conséquences sur

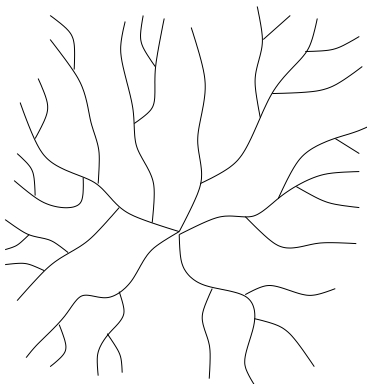
- La robustesse
- La propagation de virus
- ...

Actuellement : courant de travaux visant à cartographier l'internet

# Propriétés observées vs réelles

[Infocom 02, Infocom 03, STOC 05, ...]

Une source, beaucoup de destinations:



# Propriétés observées vs réelles

[Infocom 02, Infocom 03, STOC 05, ...]

- biais sur la **structure observée**
- Travaux théoriques et empiriques pour :
- Évaluer le biais
  - Évaluer de façon sûre certaines propriétés



# Résumé

## Cartographie au niveau IP

Long et coûteux

## Tendances actuelles

- Données massives  
(multiplier sources et destinations, mesures distribuées, ...)
- Améliorer la qualité

Dynamique ?

## Notre approche

**Ce qu'une machine voit de l'internet** est :

- intéressant en soi
- (plus) facile à mesurer
- (plus) facile à interpréter
- peut être mesuré efficacement (temps, charge)

notion de vision **égo-centrée**

Mesure efficace + simple  $\implies$  Répétition

étude de la dynamique

Radars

une source, des destinations, mesures périodiques

## Notre approche

**Ce qu'une machine voit de l'internet** est :

- intéressant en soi
- (plus) facile à mesurer
- (plus) facile à interpréter
- peut être mesuré efficacement (temps, charge)

notion de vision **égo-centrée**

Mesure efficace + simple  $\implies$  Répétition

étude de la dynamique

Radars

une source, des destinations, mesures périodiques

## Notre approche

**Ce qu'une machine voit de l'internet** est :

- intéressant en soi
- (plus) facile à mesurer
- (plus) facile à interpréter
- peut être mesuré efficacement (temps, charge)

notion de vision **égo-centrée**

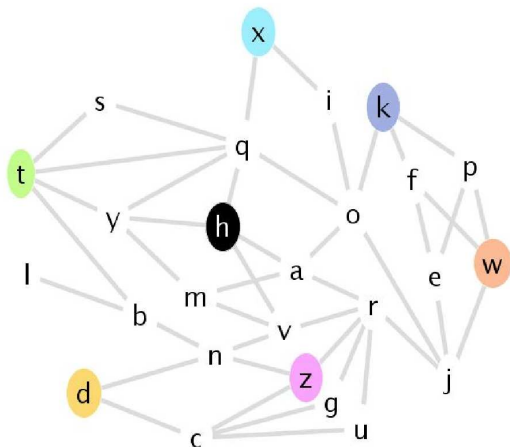
Mesure efficace + simple  $\implies$  Répétition

étude de la dynamique

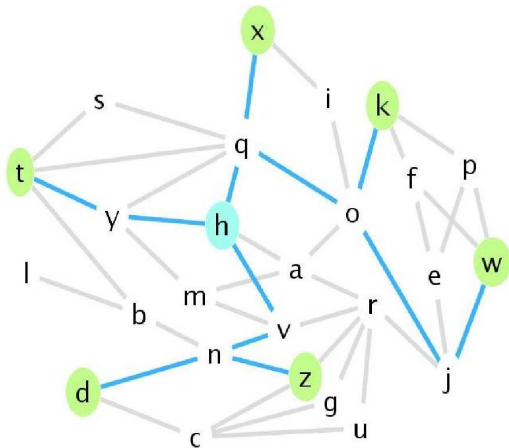
Radar

une source, des destinations, mesures périodiques

## Vue égo-centrée



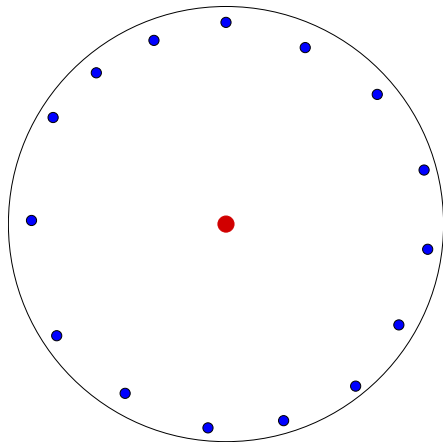
## Vue égo-centrée



# Outline

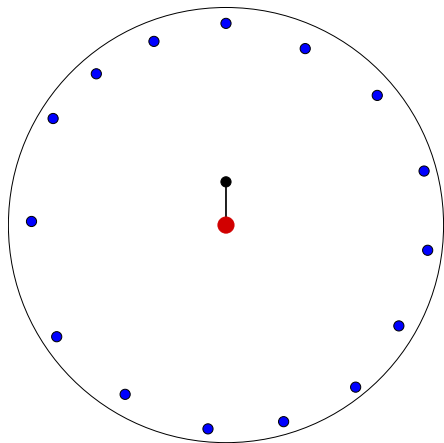
- 1 « Mesure » de la dynamique
  - Contexte et approche
  - **Tracetree**
  - Mesures
- 2 Analyse de la dynamique
  - Propriétés élémentaires
  - Croissance
  - Présence et blocs de présence
- 3 Modélisation de la dynamique
  - Impact de la dynamique
  - Modélisation
  - Simulations

## Mesure avec traceroute

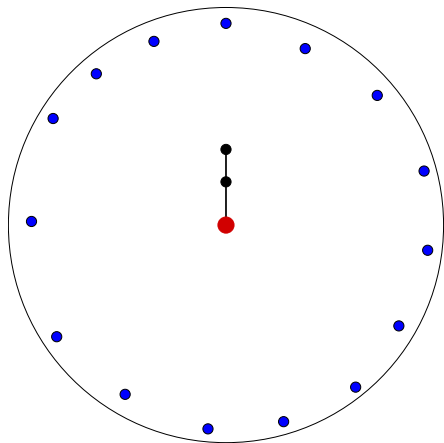




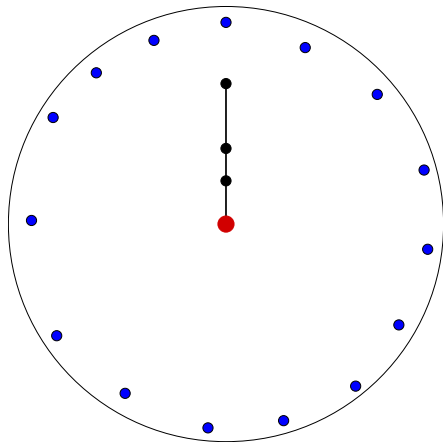
## Mesure avec traceroute



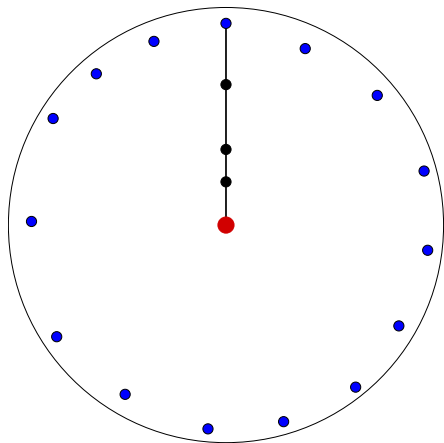
## Mesure avec traceroute



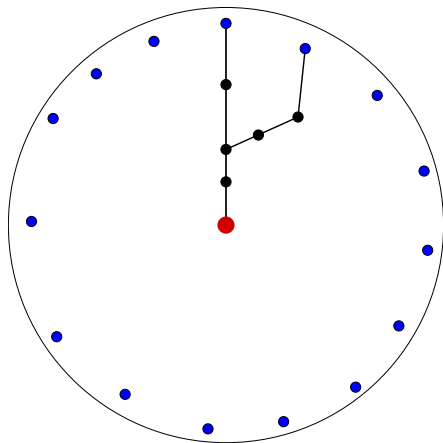
## Mesure avec traceroute



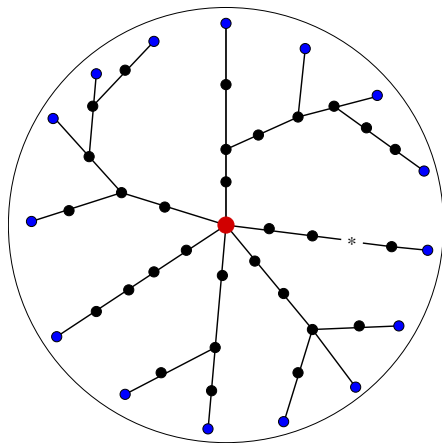
## Mesure avec traceroute



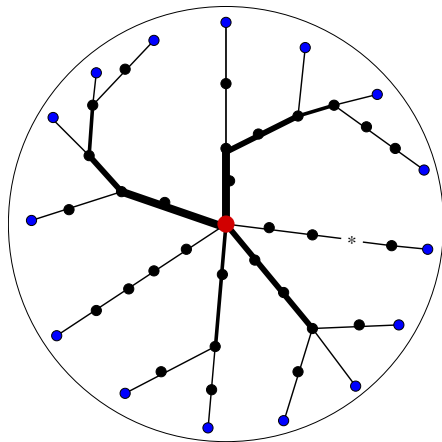
## Mesure avec traceroute



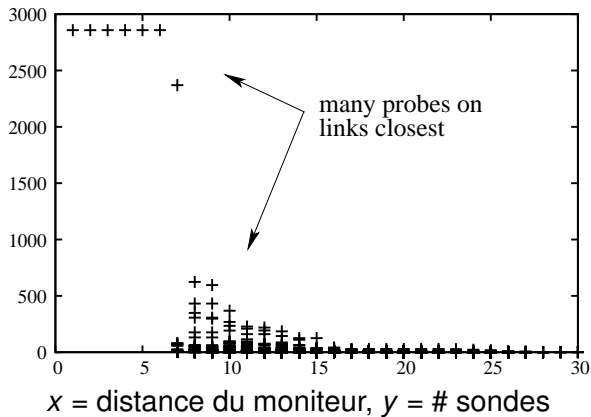
## Mesure avec traceroute



# Charge déséquilibrée



## Charge déséquilibrée





## Limites de traceroute

- **charge déséquilibrée**  
⇒ **information redondante**
- **pas un arbre**  
⇒ **complice l'analyse**

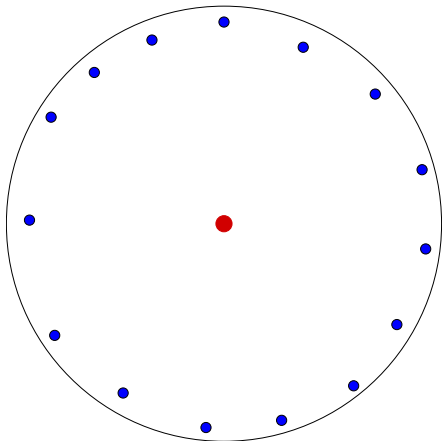
↔ **nécessité d'un outil dédié**  
**tracetree**

## Limites de traceroute

- **charge déséquilibrée**  
⇒ **information redondante**
- **pas un arbre**  
⇒ **complice l'analyse**

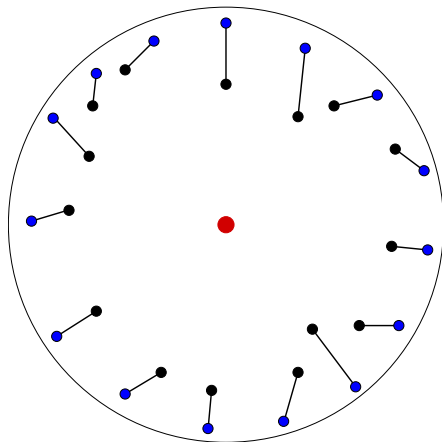
↪ **nécessité d'un outil dédié**  
**tracetree**

## Tracetree



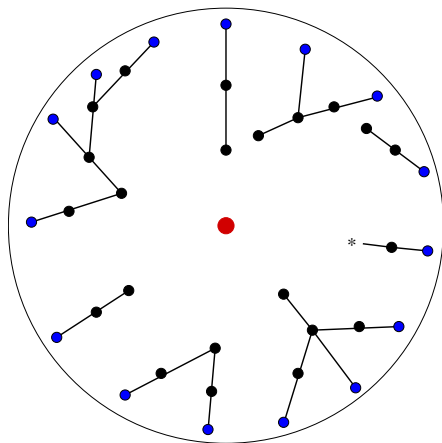
Mesures *en arrière*, en parallèle, arrêt lorsque les chemins se rencontrent

# Tracetree



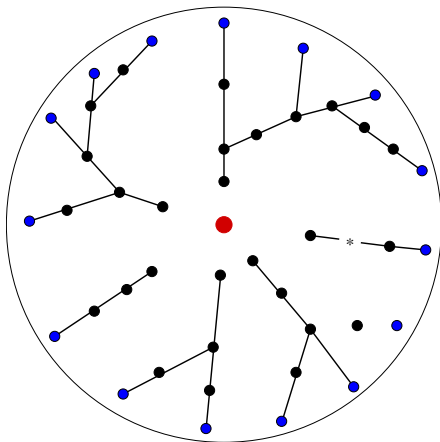
Mesures *en arrière*, en parallèle, arrêt lorsque les chemins se rencontrent

# Tracetree



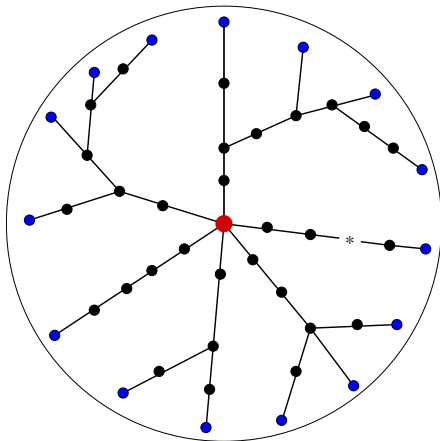
Mesures *en arrière*, en parallèle, arrêt lorsque les chemins se rencontrent

# Tracetree



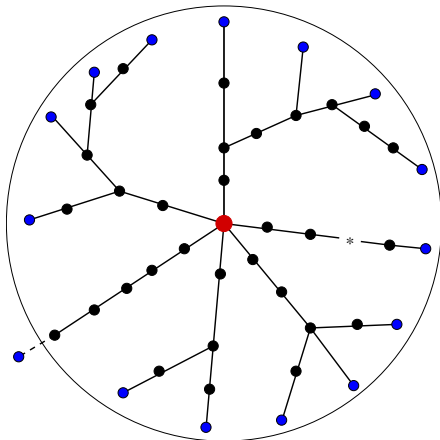
Mesures *en arrière*, en parallèle, arrêt lorsque les chemins se rencontrent

# Tracetree



Mesures *en arrière*, en parallèle, arrêt lorsque les chemins se rencontrent

# Tracetree



**Limite du rayon (TTL max):**  
Sonder jusqu'à une distance donnée du moniteur



# Tracetree et radar

## Tracetree

- 1 paquet par lien
- mesure homogène et rapide
- arbre

→ programme disponible, écrit en C.

## Radar

- Une source
- Un ensemble de destinations
- Mesures périodiques avec tracetree

Tracetree : mêmes problèmes que traceroute pour la qualité de la mesure

# Tracetree et radar

## Tracetree

- 1 paquet par lien
- mesure homogène et rapide
- arbre

→ programme disponible, écrit en C.

## Radar

- Une source
- Un ensemble de destinations
- Mesures périodiques avec tracetree

Tracetree : mêmes problèmes que traceroute pour la qualité de la mesure

# Tracetree et radar

## Tracetree

- 1 paquet par lien
- mesure homogène et rapide
- arbre

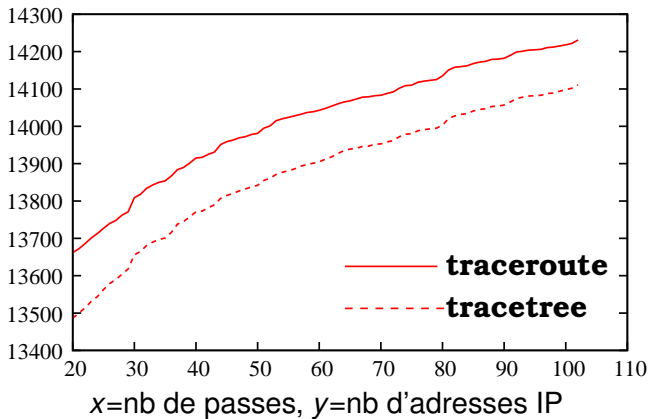
→ programme disponible, écrit en C.

## Radar

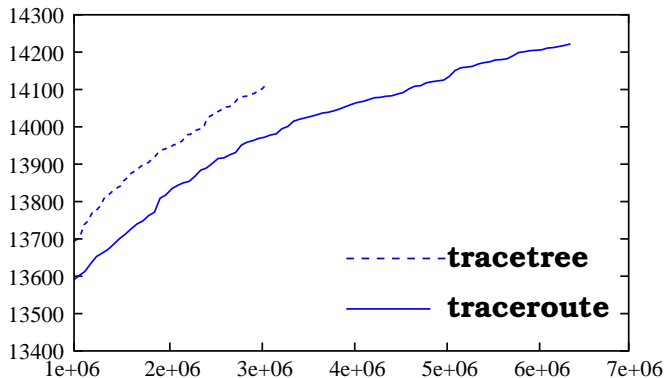
- Une source
- Un ensemble de destinations
- Mesures périodiques avec tracetree

Tracetree : mêmes **problèmes** que traceroute pour la qualité de la mesure

## tracetree vs traceroute: comparaison empirique

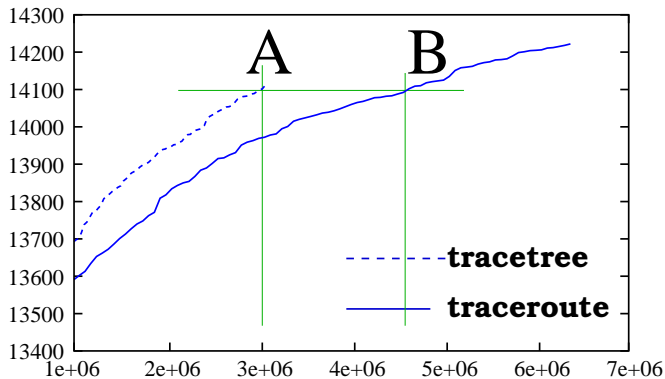


## tracetree vs traceroute: comparaison empirique



$x = \text{nb de paquets}$ ,  $y = \text{nb d'adresses IP}$

## tracetree vs traceroute: comparaison empirique



$x = \text{nb de paquets}$ ,  $y = \text{nb d'adresses IP}$

# Outline

- 1 « Mesure » de la dynamique
  - Contexte et approche
  - Tracetree
  - **Mesures**
- 2 Analyse de la dynamique
  - Propriétés élémentaires
  - Croissance
  - Présence et blocs de présence
- 3 Modélisation de la dynamique
  - Impact de la dynamique
  - Modélisation
  - Simulations

# Paramètres

## Plusieurs paramètres des mesures

Quelles sources / destinations?

Combien de destinations?

Délai entre passes?

Timeout?

...

## Compromis

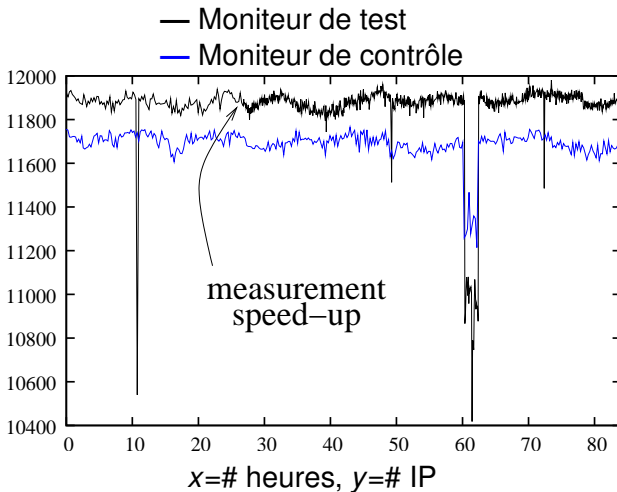
Haute fréquence

Grande quantité de données

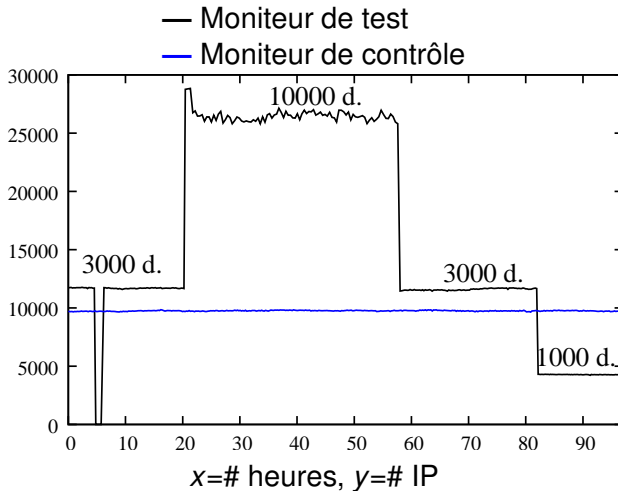
Faible charge sur le réseau



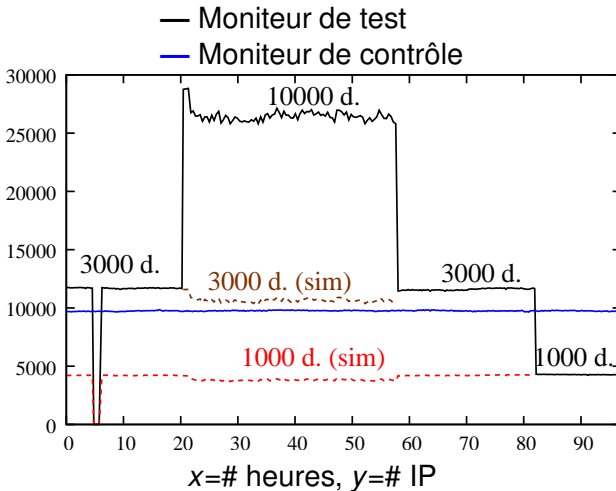
## Paramètres: la fréquence



## Paramètres: nombre de destinations



## Paramètres: nombre de destinations



# Nos mesures

## Deux **ensembles de paramètres**:

- *normal*: 3000 destinations, 10 min. délai entre passes,  
TTL max 30, ... ~ 100 passes / jour
- *rapide*: 1000 destinations, 1 min. délai entre passes,  
TTL max 15, ... > 800 passes / jour

**Sources**: PlanetLab et autres (> 100)

**Destinations**: Aléatoires, répondant au ping

Plusieurs mois de mesure ininterrompue  
Données disponibles pour étude

## Nos mesures

Deux **ensembles de paramètres**:

- *normal*: 3000 destinations, 10 min. délai entre passes,  
TTL max 30, ... ~ 100 passes / jour
- *rapide*: 1000 destinations, 1 min. délai entre passes,  
TTL max 15, ... > 800 passes / jour

**Sources**: PlanetLab et autres (> 100)

**Destinations**: Aléatoires, répondant au ping

**Plusieurs mois de mesure ininterrompue**  
**Données disponibles pour étude**

# Outline

- 1 « Mesure » de la dynamique
  - Contexte et approche
  - Tracetree
  - Mesures
- 2 Analyse de la dynamique
  - Propriétés élémentaires
  - Croissance
  - Présence et blocs de présence
- 3 Modélisation de la dynamique
  - Impact de la dynamique
  - Modélisation
  - Simulations

# Analyse

## Quelle dynamique ?

Pour :

- mieux comprendre
- modéliser, simuler
- détecter des événements
- ...

Comment répondre cette question ?

Pas de méthode toute faite

# Analyse

## Quelle dynamique ?

Pour :

- mieux comprendre
- modéliser, simuler
- détecter des événements
- ...

**Comment répondre cette question ?**

**Pas de méthode toute faite**



# Données

Questions valables pour **n'importe quel** graphe dynamique.

On se concentre sur les données radar

- 3000 destinations
- une passe toutes les 15 minutes (100 passes par jour)

# Outline

- 1 « Mesure » de la dynamique
  - Contexte et approche
  - Tracetree
  - Mesures
- 2 Analyse de la dynamique
  - **Propriétés élémentaires**
  - Croissance
  - Présence et blocs de présence
- 3 Modélisation de la dynamique
  - Impact de la dynamique
  - Modélisation
  - Simulations

# Étude d'une passe

Variations en fonction :

- de la source
- du moment

## Quelques tendances générales

- ~ 12 000 IP
- ~ 12 000 étoiles
- la plupart des nœuds à distance 13-18

On observe toujours le même **type** de comportements

## Étude d'une passe

Variations en fonction :

- de la source
- du moment

### Quelques tendances générales

- ~ 12 000 IP
- ~ 12 000 étoiles
- la plupart des nœuds à distance 13-18

On observe toujours le même **type** de comportements

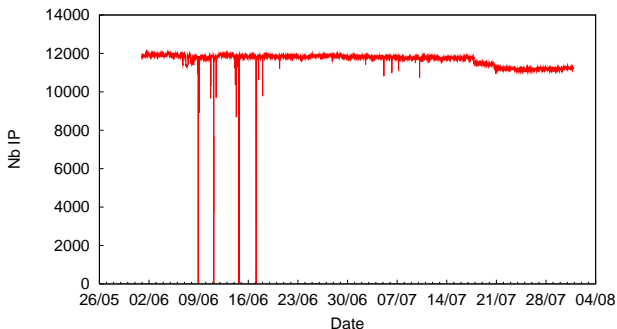
# Propriétés simples au fil du temps

Première approche :

Évolution de propriétés simples au fil du temps

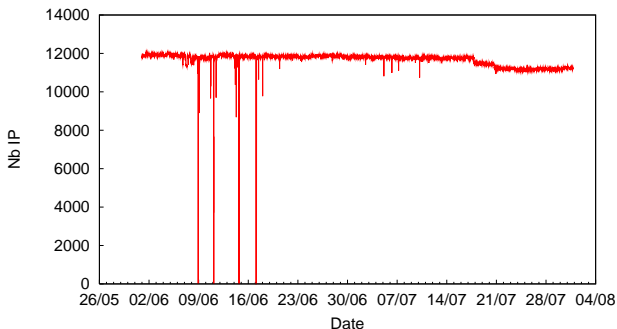
- nombre d'IP
- nombre d'étoiles
- durée des passes
- ...

## Propriétés simples au fil du temps



- source au Japon
- deux mois de mesures
- ~ 6 000 passes

## Propriétés simples au fil du temps



- Plus ou moins constant
- Pics vers le bas → déconnexions ?
- Pas de pics vers le haut

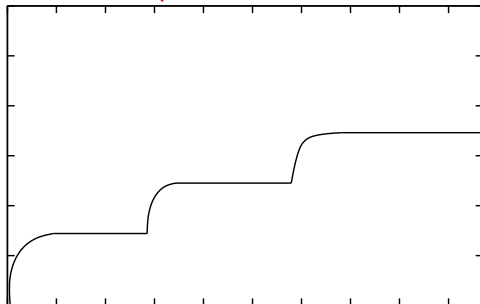
# Outline

- 1 « Mesure » de la dynamique
  - Contexte et approche
  - Tracetree
  - Mesures
- 2 Analyse de la dynamique
  - Propriétés élémentaires
  - **Croissance**
  - Présence et blocs de présence
- 3 Modélisation de la dynamique
  - Impact de la dynamique
  - Modélisation
  - Simulations



## Dynamique : stabilisation?

Comportement attendu

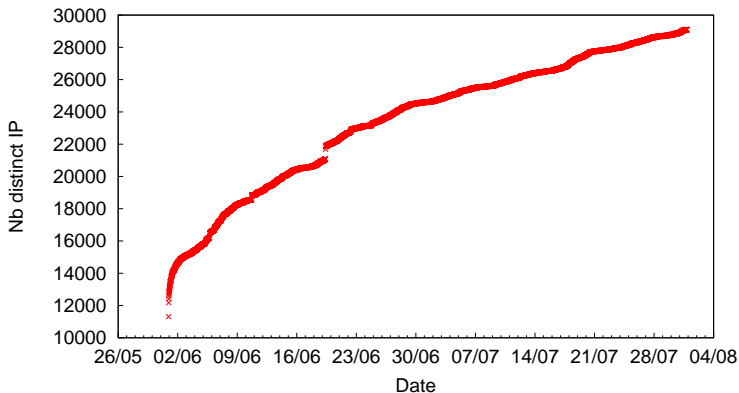


$x = \#$  de passes

$y = \#$  d'adresses IP distinctes depuis le début

## Dynamique : stabilisation ?

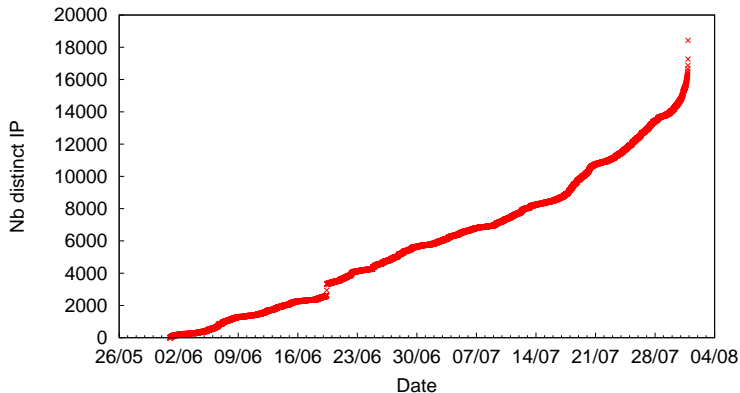
Nombre d'IP distinctes observées depuis le début



**nouvelles IP observées en permanence**

# Disparitions

Nombre d'IP distinctes qu'on ne verra plus jamais



**Renouvellement continu des IP observées**

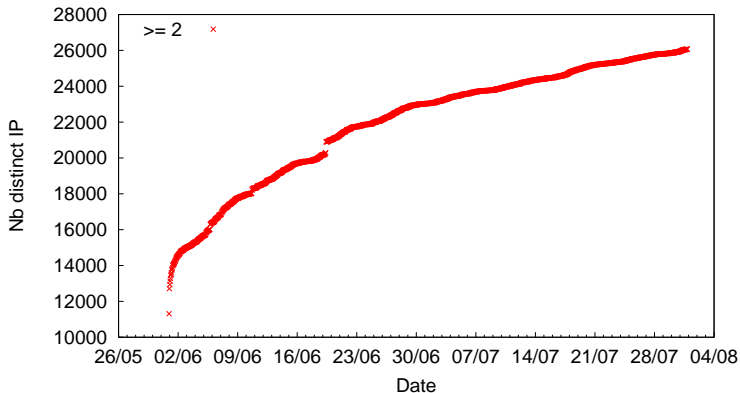
# Tentatives d'explication

## Croissance surprenante

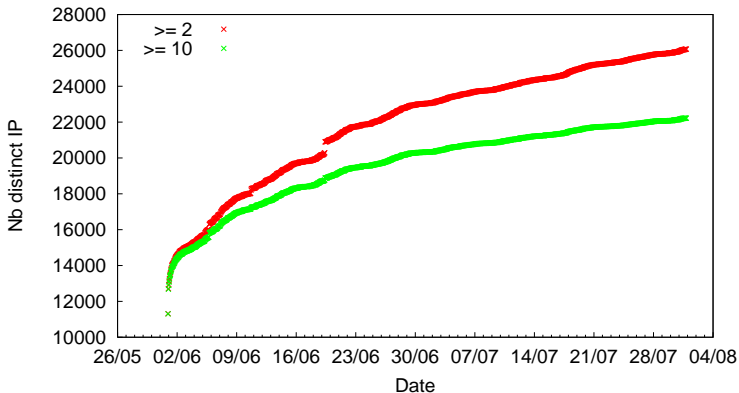
### Deux explications “naturelles”

- IP aléatoires
- IP dynamiques

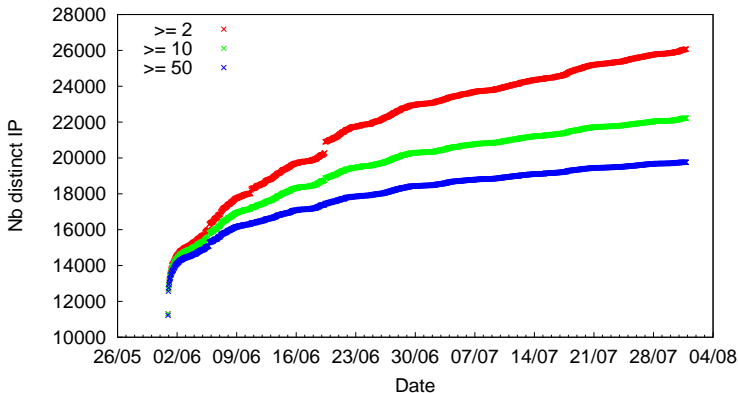
## IP vues une seule fois ?



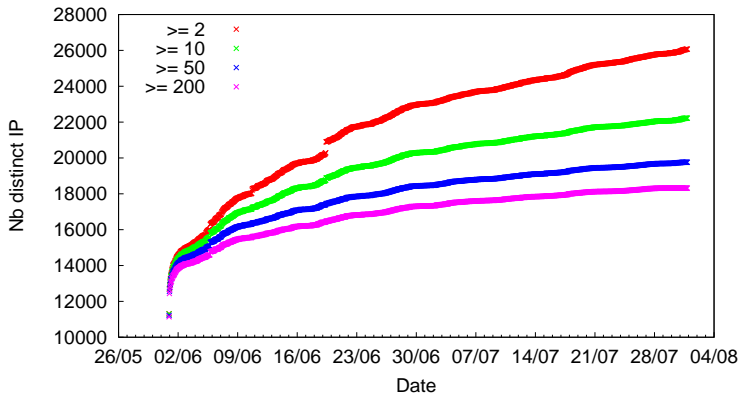
## IP vues une seule fois ?



## IP vues une seule fois ?

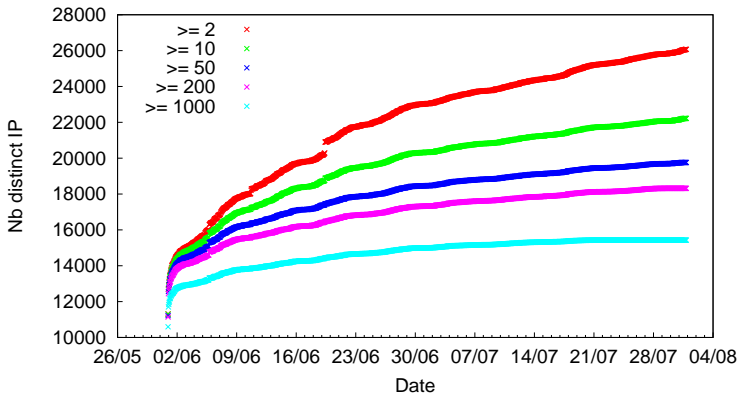


## IP vues une seule fois ?





## IP vues une seule fois ?

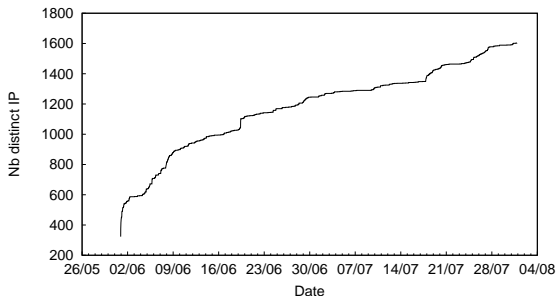


De nouvelles IP récurrentes...

Élimine l'hypothèse des IPs aléatoires

## Des IP dynamiques ?

Destinations “stables”... (un critère possible: le père dans l’arbre a toujours la même adresse)

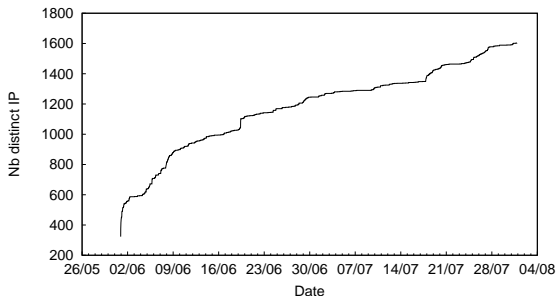


courbes de mesures sur ces adresses ont même allure.

Élimine l’hypothèse des IPs dynamiques

## Des IP dynamiques ?

Destinations “stables”... (un critère possible: le père dans l’arbre a toujours la même adresse)



courbes de mesures sur ces adresses ont même allure.

Élimine l’hypothèse des IPs dynamiques

# Outline

- 1 « Mesure » de la dynamique
  - Contexte et approche
  - Tracetree
  - Mesures
- 2 Analyse de la dynamique
  - Propriétés élémentaires
  - Croissance
  - **Présence et blocs de présence**
- 3 Modélisation de la dynamique
  - Impact de la dynamique
  - Modélisation
  - Simulations

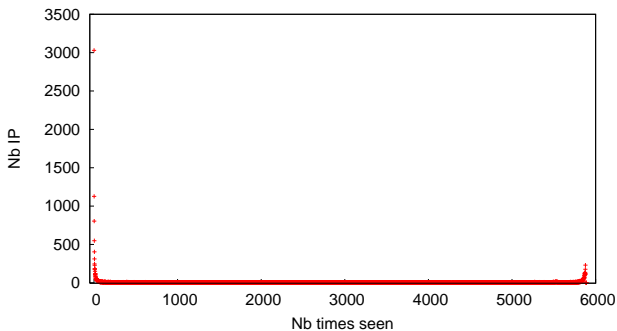
# Présence des IP

## Distribution du nombre de présences

Toutes les IP vues pendant la mesure ( $\sim 29\,000$ )  
Pour chaque IP : **nombre de passes** où on l'a vue  
→ Distribution

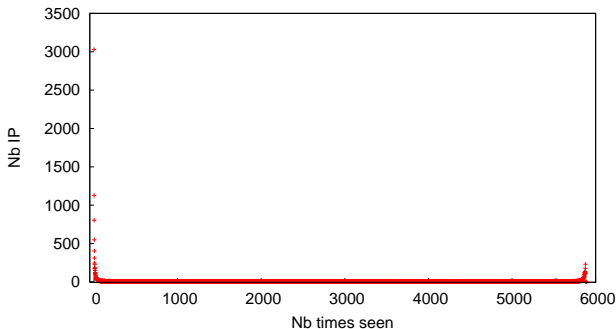
# Présence des IP

2 mois ~ 6000 passes



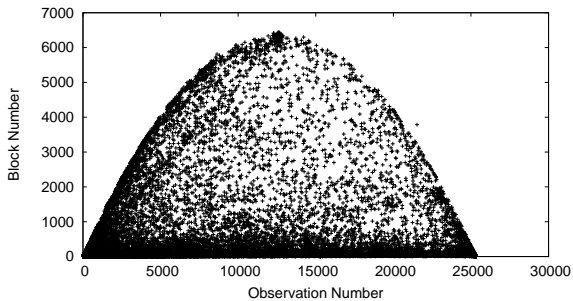
## Présence des IP

2 mois ~ 6000 passes



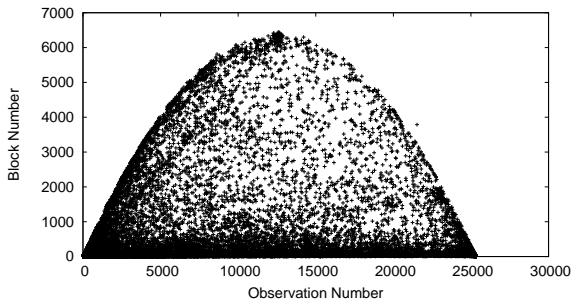
- Beaucoup d'IP **éphémères** (vues très peu de fois)
- Un nombre non négligeable d'IP **stables** (vues presque à chaque fois)

## Présences vs blocs de présence



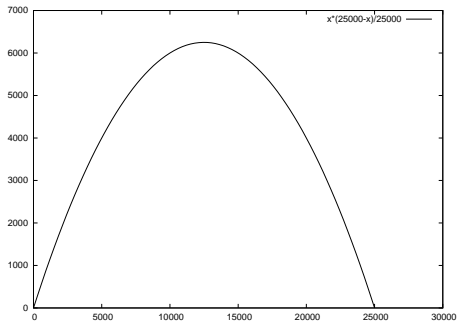


## Présences vs blocs de présence



Triangle : Nb blocs de présence  $\leq \max(\text{nb présences, nb absences})$

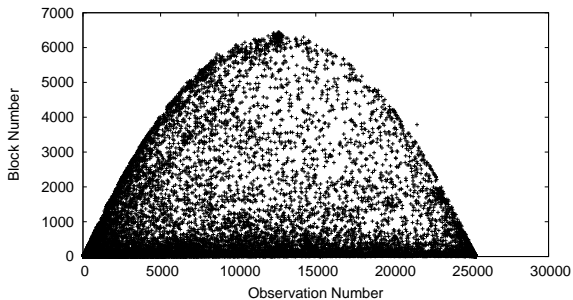
## Présences vs blocs de présence



Parabole : Nombre de blocs attendues en moyenne pour ce nombre de présences

$$\frac{x(25\,000-x)}{25\,000}$$

## Présences vs blocs de présence



Deux classes d'IP

aléatoires || majorité de stables

# Outline

- 1 « Mesure » de la dynamique
  - Contexte et approche
  - Tracetree
  - Mesures
- 2 Analyse de la dynamique
  - Propriétés élémentaires
  - Croissance
  - Présence et blocs de présence
- 3 Modélisation de la dynamique
  - Impact de la dynamique
  - Modélisation
  - Simulations

# Outline

- 1 « Mesure » de la dynamique
  - Contexte et approche
  - Tracetree
  - Mesures
- 2 Analyse de la dynamique
  - Propriétés élémentaires
  - Croissance
  - Présence et blocs de présence
- 3 Modélisation de la dynamique
  - Impact de la dynamique
  - Modélisation
  - Simulations

## Comment expliquer la dynamique

Une démarche possible :

- 1 Choix d'observables extraits de l'analyse (**fait !**)
- 2 Identifications de mécanismes sous-jacents qui sont liés aux observables
- 3 Modélisation des mécanismes sous-jacents
- 4 Validation via simulations

Plusieurs mécanismes peuvent être responsables des changements dans le routage :

- load-balancing
- modification de la topologie de routage

Ces facteurs ont-ils vraiment un impact sur nos mesures ?

## Comment expliquer la dynamique

Une démarche possible :

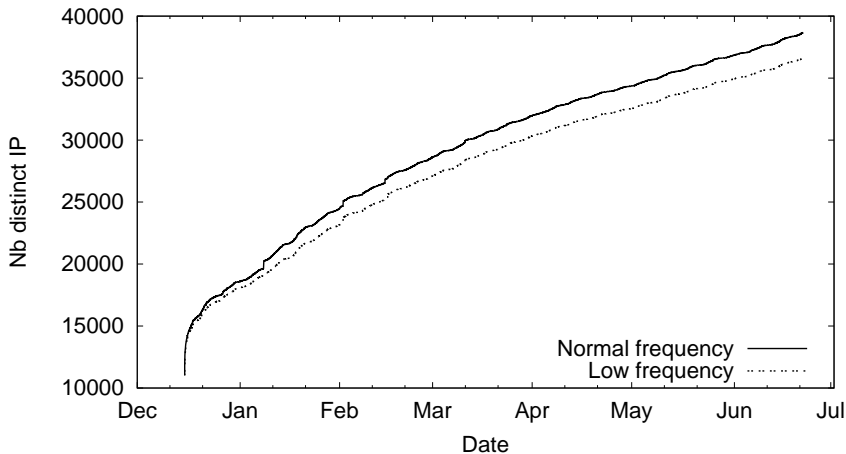
- 1 Choix d'observables extraits de l'analyse (**fait !**)
- 2 Identifications de mécanismes sous-jacents qui sont liés aux observables
- 3 Modélisation des mécanismes sous-jacents
- 4 Validation via simulations

Plusieurs mécanismes peut-être responsables des changements dans le routage :

- load-balancing
- modification de la topologie de routage

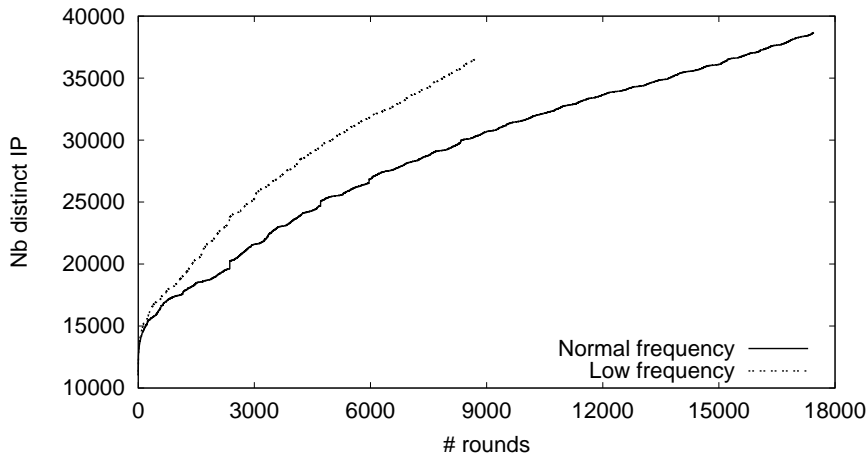
**Ces facteurs ont-ils vraiment un impact sur nos mesures ?**

## Impact de la fréquence en fonction du temps





## Impact de la fréquence en fonction du nb de passes



# Outline

- 1 « Mesure » de la dynamique
  - Contexte et approche
  - Tracetree
  - Mesures
- 2 Analyse de la dynamique
  - Propriétés élémentaires
  - Croissance
  - Présence et blocs de présence
- 3 **Modélisation de la dynamique**
  - Impact de la dynamique
  - **Modélisation**
  - Simulations

## Modéliser la dynamique

On recherche donc à la fois un modèle (simple) pour le load-balancing et pour les modifications de la topologie de routage.

**load-balancing** il est courant de modéliser le résultat d'un traceroute par un plus court chemin (i.e. BFS). Comment adapter le BFS pour tenir compte du load-balancing ?

**topologie** notion de ré-écriture de graphes. Mais quelles primitives choisir ? Nœuds aléatoirement ? Liens aléatoirement ? Et quelle topologie ?

## Modéliser traceroute et le load-balancing

Plusieurs possibilités pour choisir 1 plus court chemin. Étant donné un DAG des plus courts chemins partant du moniteur vers une destination, on peut :

- choisir 1 chemin parmi tous les chemins possibles
- pour chaque nœuds suivant dans le DAG, on choisit 1 nœud parmi ceux possibles

Intérêt du deuxième choix : choix local au nœud

Implémentation : BFS aléatoire

## Modéliser traceroute et le load-balancing

Plusieurs possibilités pour choisir 1 plus court chemin. Étant donné un DAG des plus courts chemins partant du moniteur vers une destination, on peut :

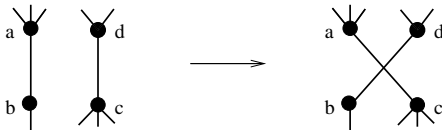
- choisir 1 chemin parmi tous les chemins possibles
- pour chaque nœuds suivant dans le DAG, on choisit 1 nœud parmi ceux possibles

**Intérêt du deuxième choix** : choix local au nœud

**Implémentation** : BFS aléatoire

## Modéliser les modifications de la topologie

On choisit ici de se concentrer sur les liens en considérant un rebranchement (rewiring) de deux paires de nœuds liés:

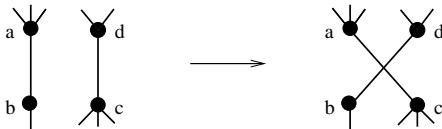


**Intérêt** : ne modifie pas la distribution des degrés du graphe !

**Inconvénient** : ne permet pas de traiter l'apparition/disparition des nœuds dans le graphe.

## Modéliser les modifications de la topologie

On choisit ici de se concentrer sur les liens en considérant un rebranchement (rewiring) de deux paires de nœuds liés:



**Intérêt** : ne modifie pas la distribution des degrés du graphe !

**Inconvénient** : ne permet pas de traiter l'apparition/disparition des nœuds dans le graphe.

# Outline

- 1 « Mesure » de la dynamique
  - Contexte et approche
  - Tracetree
  - Mesures
- 2 Analyse de la dynamique
  - Propriétés élémentaires
  - Croissance
  - Présence et blocs de présence
- 3 Modélisation de la dynamique
  - Impact de la dynamique
  - Modélisation
  - Simulations



## Valider par des simulations

Environnement des simulations :

- 1 On génère un graphe initial : **ER**, PL, CM, autres ...  $\Rightarrow G_1$
- 2 On simule une mesure : extraction d'un arbre de routage (BFS aléatoire)  $\Rightarrow T_1$
- 3 On simule une modification de la topologie : swap  $\Rightarrow G_2$
- 4 On retourne à l'étape 2 :  $\Rightarrow T_2, \dots, T_n$

Question :

- Observe-t-on les mêmes phénomènes dynamiques dans  $T_1, \dots, T_n$  ?
- Quel est l'impact des paramètres des modèles ?

## Valider par des simulations

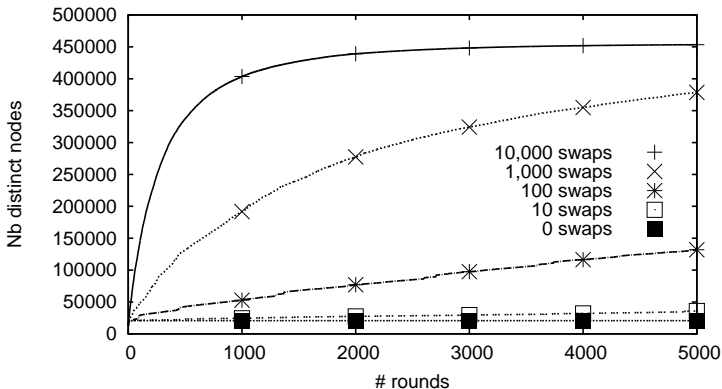
Environnement des simulations :

- 1 On génère un graphe initial : **ER**, PL, CM, autres ...  $\Rightarrow G_1$
- 2 On simule une mesure : extraction d'un arbre de routage (BFS aléatoire)  $\Rightarrow T_1$
- 3 On simule une modification de la topologie : swap  $\Rightarrow G_2$
- 4 On retourne à l'étape 2 :  $\Rightarrow T_2, \dots, T_n$

Question :

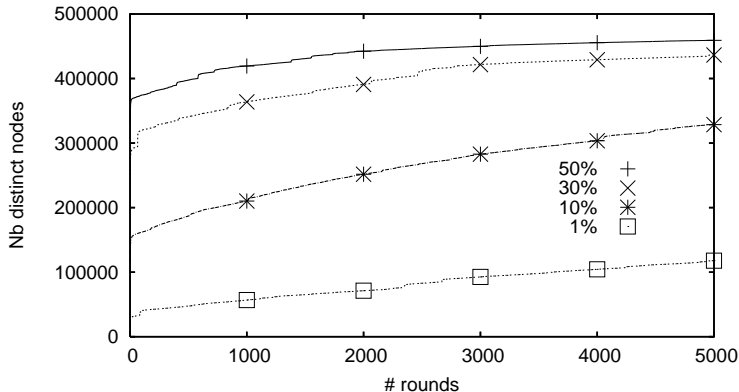
- Observe-t-on les mêmes phénomènes dynamiques dans  $T_1, \dots, T_n$  ?
- Quel est l'impact des paramètres des modèles ?

## Impact du nombre de swaps



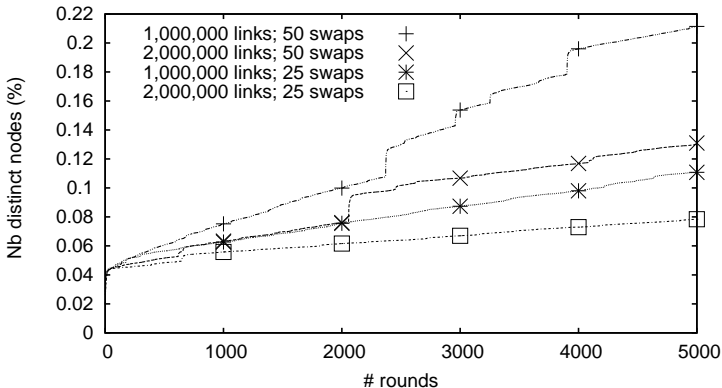
$$n = 500,000, m = 1,000,000, d = 3,000$$

## Impact du nombre de destinations



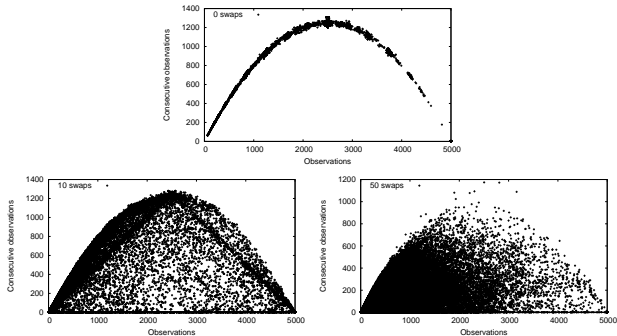
$$n = 500,000, m = 1,000,000, s = 50$$

## Relation entre le nombre de liens et de swaps



$$n = 500,000, d = 3,000$$

## Observation de la parabole



$$n = 500,000, m = 1,000,000, d = 3,000$$

## Conclusions sur les simulations

Conclusions globales :

- On retrouve bien **qualitativement** la dynamique observée
- Les modèles choisis réussissent à reproduire les observations
- Quelques contraintes sur les paramètres permettent d'identifier des invariants dans la dynamique

Une suite possible :

- Tester d'autres topologies
- Intégrer les phénomènes d'apparition/disparitions des nœuds
- Obtenir des résultats analytiques pour mieux comprendre les courbes observées