

Analyse des Grands Graphes de Terrain

Matthieu Latapy

LIP6 – CNRS et Université Pierre et Marie Curie (UPMC – Paris 6)

CORIA – exposé invité

Grands graphes de terrain

informatique : internet, web, pair-à-pair, etc

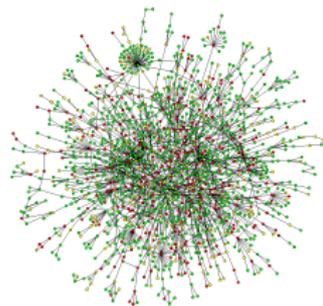
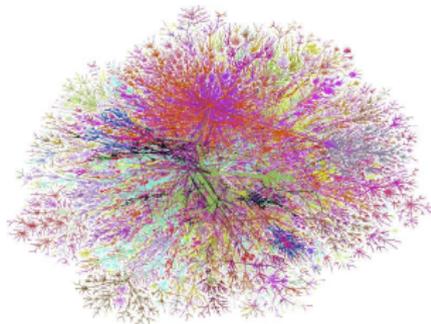
biologie : cerveau, gènes, protéines, écosystèmes, etc

sciences sociales : collaboration, amitié, contacts sexuels, échanges, économie, etc

linguistique : synonymie, co-occurrence, etc

transport : routier, aérien, électrique, etc

etc, etc



pas de contenu
graphes issus du *terrain*
contextes très différents

Des propriétés en commun

la plupart des graphes de terrain ont des propriétés non-triviales en commun [WS98]

propriétés en commun :

- distance moyenne et diamètre faibles (*petit-monde*)
- degrés hétérogènes (loi de puissance, *scale-free*)
- densité globale faible vs densité locale forte (*clustering*)

et des problématiques communes

- mesure et métrologie
- modélisation
- analyse
- algorithmique

Des propriétés en commun

la plupart des graphes de terrain ont des propriétés non-triviales en commun [WS98]

propriétés en commun :

- distance moyenne et diamètre faibles (*petit-monde*)
- degrés hétérogènes (loi de puissance, *scale-free*)
- densité globale faible vs densité locale forte (*clustering*)

et des problématiques communes

- mesure et métrologie
- modélisation
- analyse
- algorithmique

Des propriétés en commun

la plupart des graphes de terrain ont des propriétés non-triviales en commun [WS98]

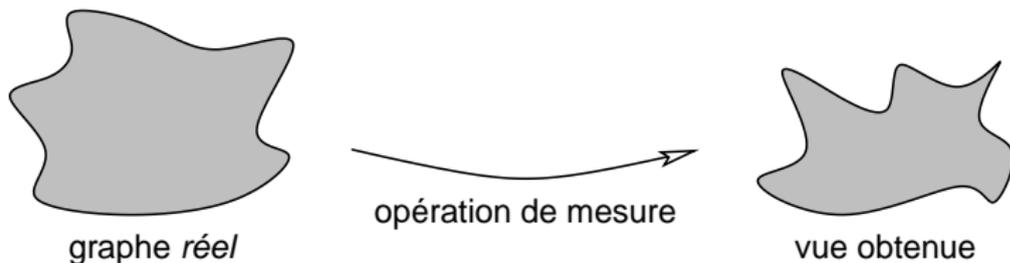
propriétés en commun :

- distance moyenne et diamètre faibles (*petit-monde*)
- degrés hétérogènes (loi de puissance, *scale-free*)
- densité globale faible vs densité locale forte (*clustering*)

et des problématiques communes

- mesure et métrologie
- analyse
- modélisation
- algorithmique

Mesure



que peut on dire sur l'objet réel à partir de la mesure ?

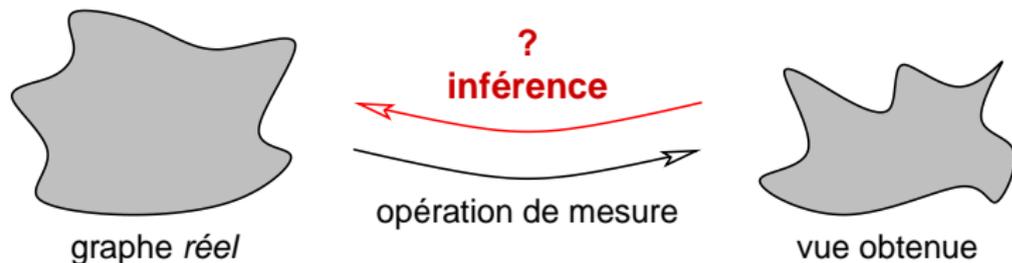
impact sur les propriétés observées ?

impact des propriétés sur la vue ?

mesures ciblant certaines propriétés ?

...

Mesure et Métrologie



que peut on dire sur l'objet réel à partir de la mesure ?

impact sur les propriétés observées ?

impact des propriétés sur la vue ?

mesures ciblant certaines propriétés ?

...

Analyse

décrire

extraire de l'information pertinente

statistiques

structure

densité
degrés
densité locale
corrélations

...

encore beaucoup de manques

décrire

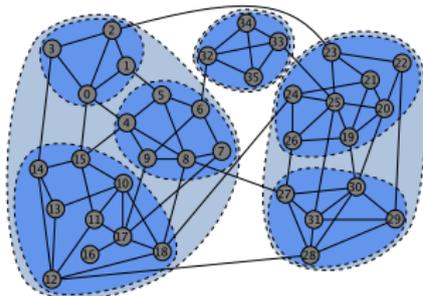
extraire de l'information pertinente

statistiques

densité
degrés
densité locale
corrélations

...

structure



encore beaucoup de manques

génération de graphes *réalistes* (*i.e.* ayant les propriétés observées)

motivations : approches formelles, simulation, explication

état de l'art :

- taille, densité, distances : facile
- degrés : consensus, modèles bien maîtrisés
- densité locale : pas de consensus

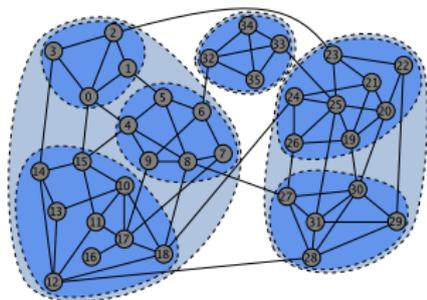
génération de graphes *réalistes* (i.e. ayant les propriétés observées)

motivations : approches formelles, simulation, explication

état de l'art :

- taille, densité, distances : facile
- degrés : consensus, modèles bien maîtrisés
- densité locale : pas de consensus

Algorithmique

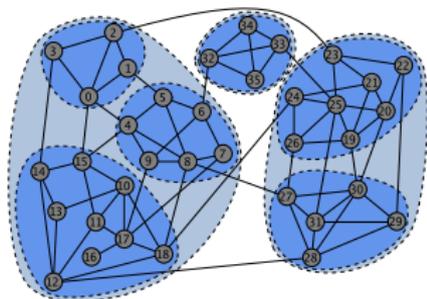


problèmes **spécifiques**
problèmes classiques à **revisiter**
restrictions en **espace**

- tirer parti des propriétés observées
- résultats approchés souvent suffisants
- algorithmique empirique / expérimentale

↪ **renouveau algorithmique**

Algorithmique

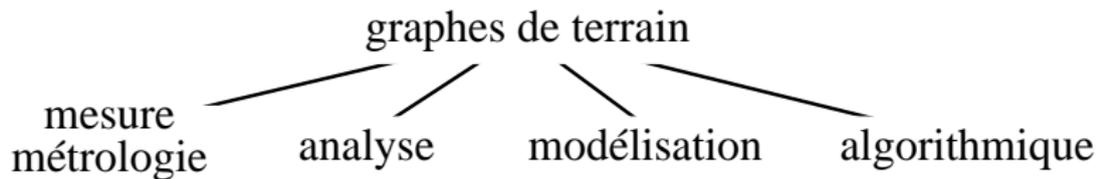


problèmes **spécifiques**
problèmes classiques à **revisiter**
restrictions en **espace**

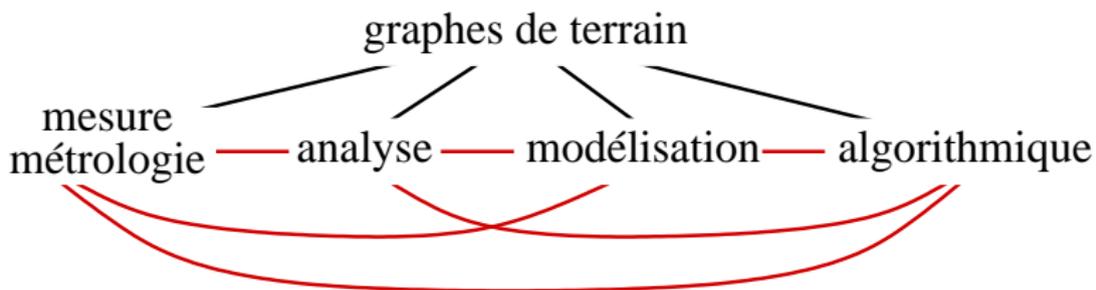
- tirer parti des propriétés observées
- résultats approchés souvent suffisants
- algorithmique empirique / expérimentale

↪ **renouveau algorithmique**

Vue d'ensemble



Vue d'ensemble



- mesure \rightarrow données \rightarrow analyse \rightarrow propriétés \rightarrow modélisation
- modélisation \rightarrow formalismes \rightarrow métrologie, algorithmique
- analyse \rightarrow calcul \rightarrow algorithmique \rightarrow calculables \rightarrow analyse
- métrologie \rightarrow propriétés robustes \rightarrow analyse
- ...

Un cas d'étude

Un radar pour l'internet

avec Clémence Magnien
et Frédéric Ouédraogo

+ Ludovic Denoyer
+ Patrick Gallinari

Contexte

Topologie de l'internet telle que vue par *traceroute*
propriétés inattendues et importantes
[Sigcomm 99, Infocom 00, Science 99, ...]

Problèmes de mesure / métrologie
biais sur certaines propriétés
[Infocom 02, Infocom 03, STOC 05, ...]

Actuellement : efforts pour **données massives**
augmenter le nombre de moniteurs
[Infocom 05, Sigcomm 05, ...]

Contexte

Topologie de l'internet telle que vue par *traceroute*
propriétés inattendues et importantes
[Sigcomm 99, Infocom 00, Science 99, ...]

Problèmes de mesure / métrologie
biais sur certaines propriétés
[Infocom 02, Infocom 03, STOC 05, ...]

Actuellement : efforts pour **données massives**
augmenter le nombre de moniteurs
[Infocom 05, Sigcomm 05, ...]

Contexte

Topologie de l'internet telle que vue par *traceroute*
propriétés inattendues et importantes
[Sigcomm 99, Infocom 00, Science 99, ...]

Problèmes de mesure / métrologie
biais sur certaines propriétés
[Infocom 02, Infocom 03, STOC 05, ...]

Actuellement : efforts pour **données massives**
augmenter le nombre de moniteurs
[Infocom 05, Sigcomm 05, ...]

Notre approche

[Clémence Magnien, Frédéric Ouédraogo]

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 **étude de la dynamique**

radar =
un moniteur, des destinations, mesures périodiques

Notre approche

[Clémence Magnien, Frédéric Ouédraogo]

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 **étude de la dynamique**

radar =
un moniteur, des destinations, mesures périodiques

Notre approche

[Clémence Magnien, Frédéric Ouédraogo]

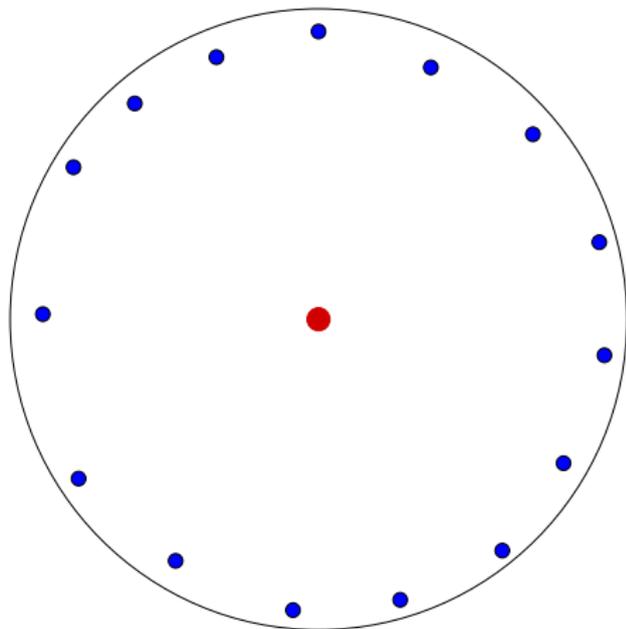
Ce qu'une machine voit de l'internet est :

- intéressant en soi
 - (plus) facile à mesurer
 - (plus) facile à interpréter
 - peut être mesuré efficacement (temps, charge)
- notion de vision **égo-centrée**

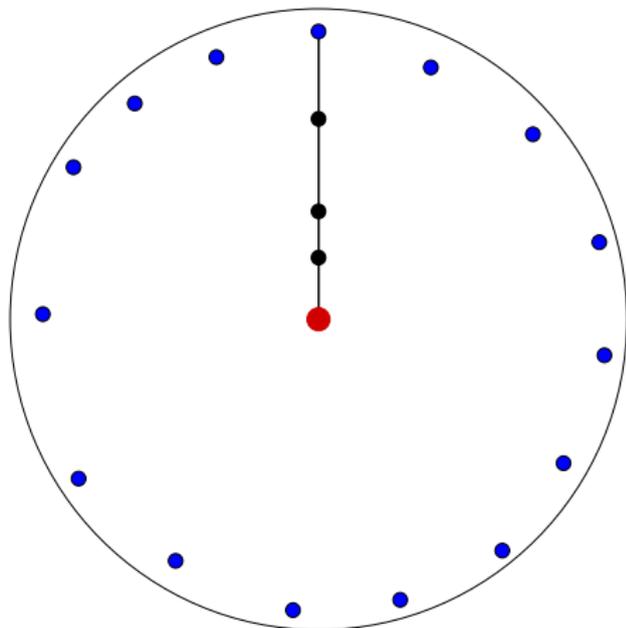
Mesure efficace + simple \implies **étude de la dynamique**

radar =
un moniteur, des destinations, mesures périodiques

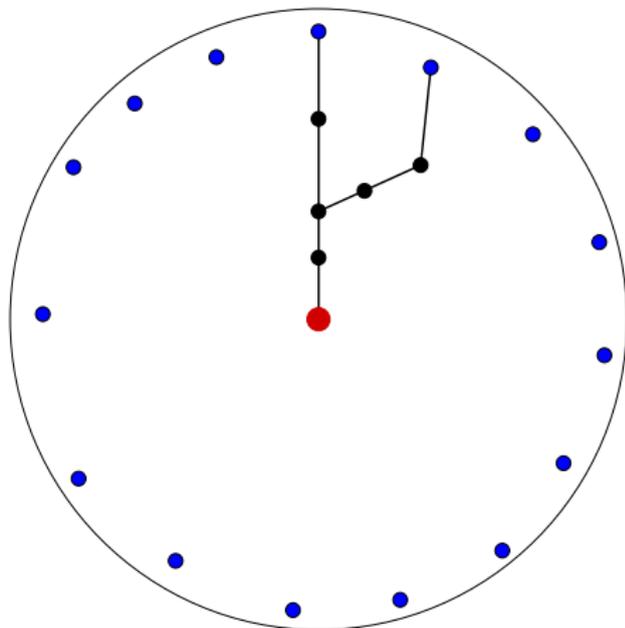
Visions égo-centrée



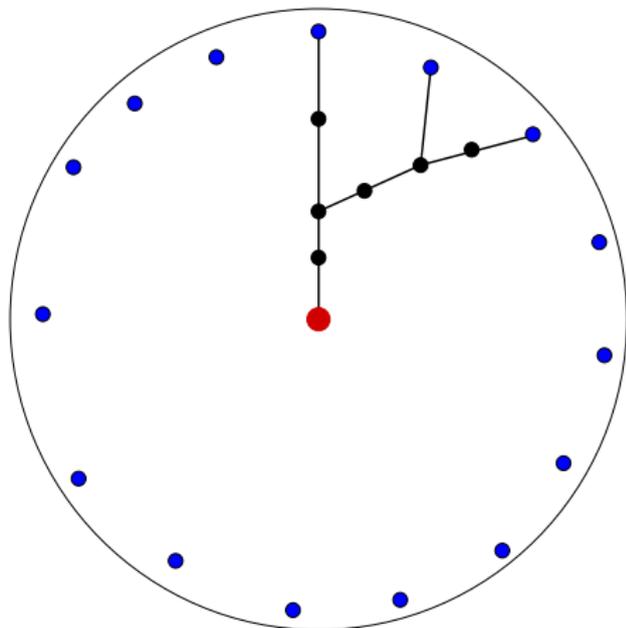
Visions égo-centrée



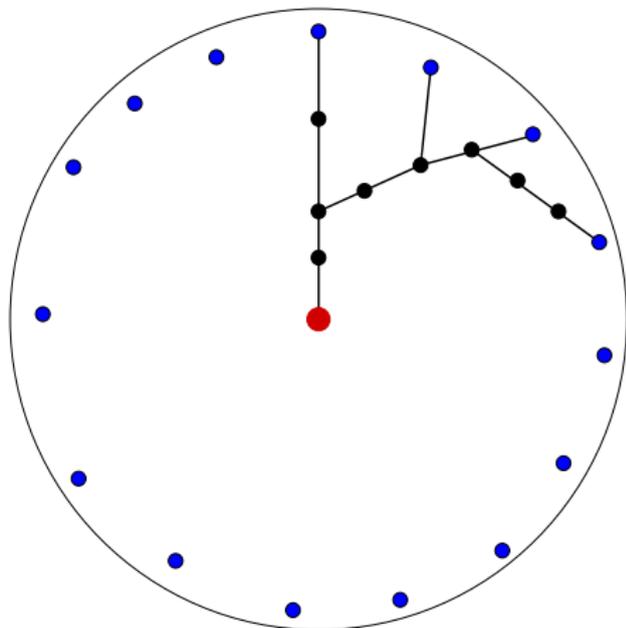
Visions égo-centrée



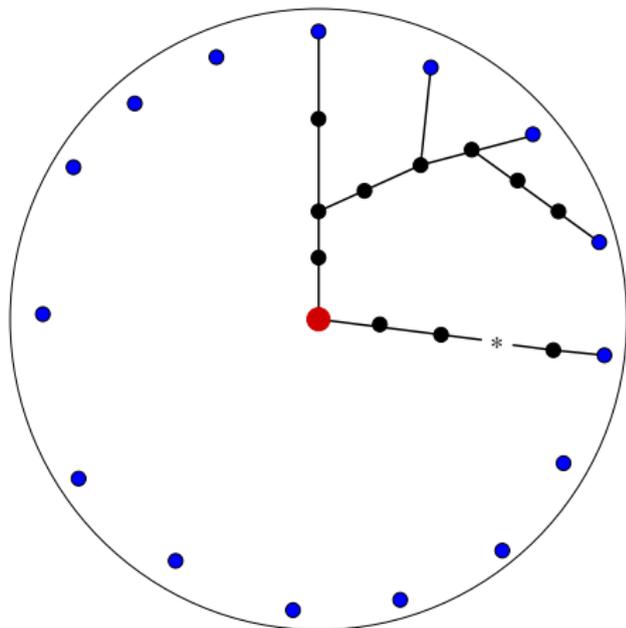
Visions égo-centrée



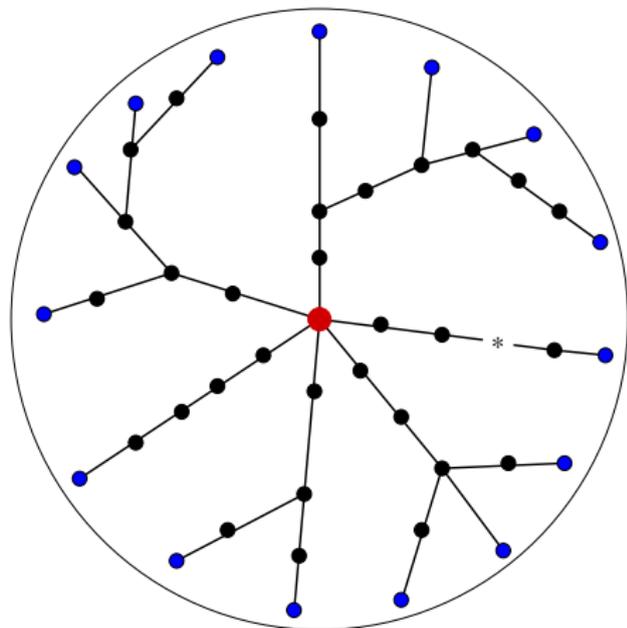
Visions égo-centrée



Visions égo-centrée



Visions égo-centrée



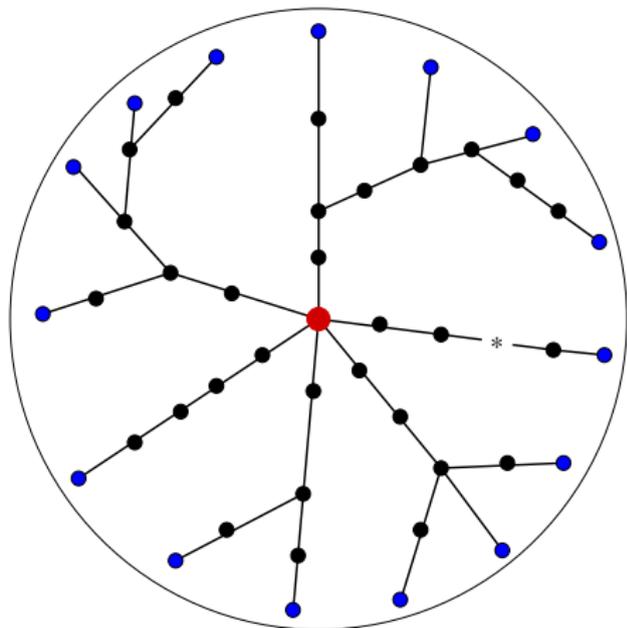
Contexte

Radar

Pub

Questions

Visions égo-centrée



radar = itération \implies suite d'arbres

Paramètres

Beaucoup de paramètres

quelles sources / destinations ?

combien de destinations ?

quel délai inter-passes ?

quel timeout ?

...

Compromis

haute fréquence

grandes mesures

charge réseau (biais)

Approche : paramètres 'raisonnables',
variation un à un,
mesures de contrôle

Paramètres

Beaucoup de paramètres

quelles sources / destinations ?

combien de destinations ?

quel délai inter-passes ?

quel timeout ?

...

Compromis

haute fréquence

grandes mesures

charge réseau (biais)

**Approche : paramètres 'raisonnables',
variation un à un,
mesures de contrôle**

Nos mesures

Sources : PlanetLab et autres (> 100)

Destinations : aléatoires 'pingables'

Deux jeux de **paramètres** :

- *normal* : 3000 destinations, délai entre passes 10mn,
TTL max 30, ... ~ 100 passes / jour
- *rapide* : 1000 destinations, délai entre passes 1mn,
TTL max 15, ... > 800 passes / jour, ~ 36 / h

plusieurs mois de mesure en continu
ensemble de données disponible

Nos mesures

Sources : PlanetLab et autres (> 100)

Destinations : aléatoires 'pingables'

Deux jeux de **paramètres** :

- *normal* : 3000 destinations, delai entre passes 10mn,
TTL max 30, ... ~ 100 passes / jour
- *rapide* : 1000 destinations, delai entre passes 1mn,
TTL max 15, ... > 800 passes / jour, ~ 36 / h

plusieurs mois de mesure en continu
ensemble de données disponible

Nos mesures

Sources : PlanetLab et autres (> 100)

Destinations : aléatoires 'pingables'

Deux jeux de **paramètres** :

- *normal* : 3000 destinations, delai entre passes 10mn,
TTL max 30, ... ~ 100 passes / jour
- *rapide* : 1000 destinations, delai entre passes 1mn,
TTL max 15, ... > 800 passes / jour, ~ 36 / h

plusieurs mois de mesure en continu
ensemble de données disponible

Dynamique : questions

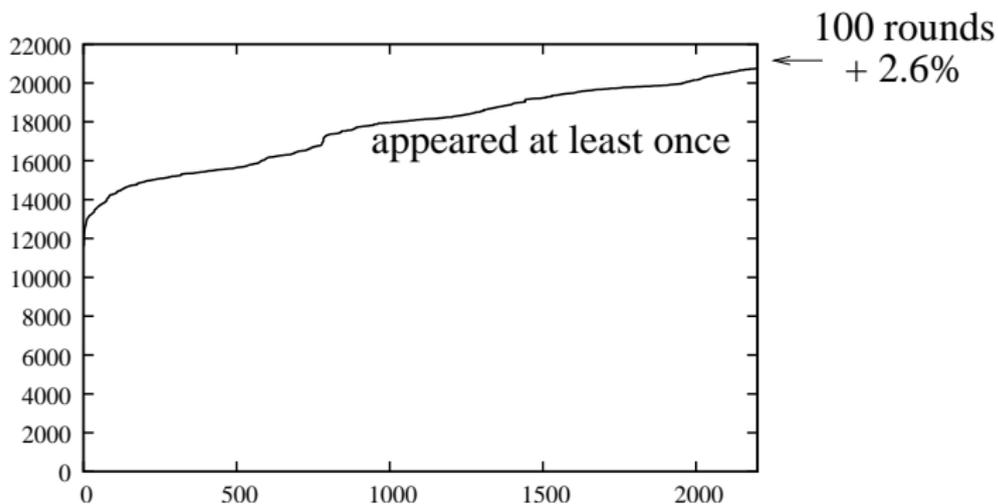
- événements ?
- notion de normalité ?
- zones particulières ? similaires ?
- classes de dynamiques d'objets ?
- différents régimes ?
- ...

Dynamique : visualisation

fichier externe `out3.gif`

Dynamique : stabilisation ?

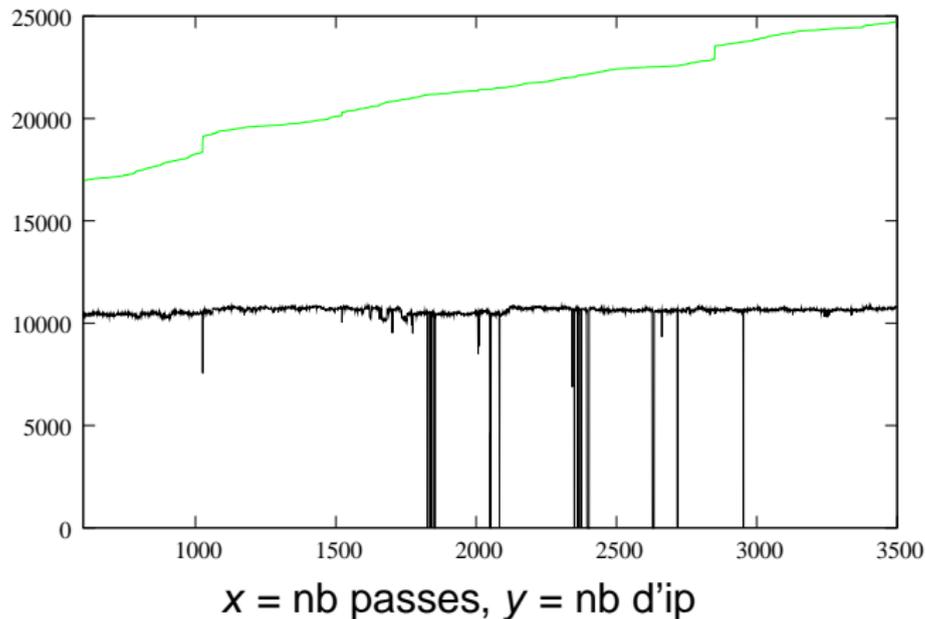
Dynamique : stabilisation ?



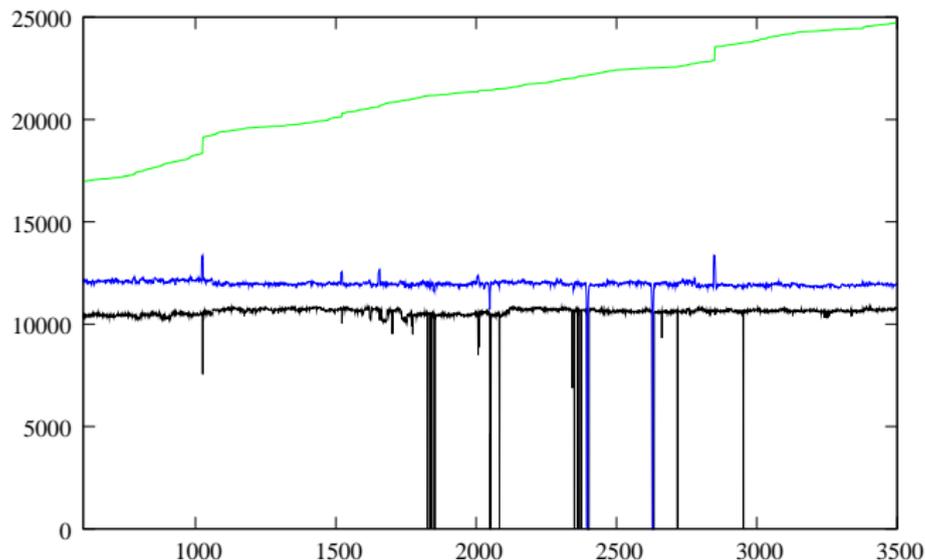
$x = \text{nb passes}$, $y = \text{nb IP distinctes}$
3 semaines ($3 \times 7 \times 100$ passes)

croissance continue et permanente

Dynamique : événements



Dynamique : événements



$x = \text{nb passes}, y = \text{nb d'ip}$

différents types d'événements

Dynamique : visualisation des événements

fichiers externes `output1c.pdf` et `outputjdir.pdf`

Conclusion sur radar

- **outil et méthode** de mesure égo-centrée
- **mesures massives** disponibles
- observation de la **dynamique**

- échelle(s) de temps pertinentes
- propriétés de la dynamique, analyse
- normal vs événement
- visualisation
- modélisation
- ...

Conclusion sur radar

- **outil et méthode** de mesure égo-centrée
- **mesures massives** disponibles
- observation de la **dynamique**

- échelle(s) de temps pertinentes
- propriétés de la dynamique, analyse
- normal vs événement
- visualisation
- modélisation
- ...

Conclusion – page de pub

Nombreux autres cas :

- traces P2P, noms de fichiers, lutte antipédo
- réseaux sociaux de l'internet
- échanges sur l'internet
- ...

Ouverts au LIP6 (CNRS – UPMC, Paris) :

- postdoc (deadline fin mars)
- maître de conférences (réseaux)

Matthieu.Latapy@lip6.fr

Conclusion – page de pub

Nombreux autres cas :

- traces P2P, noms de fichiers, lutte antipédo
- réseaux sociaux de l'internet
- échanges sur l'internet
- ...

Ouverts au LIP6 (CNRS – UPMC, Paris) :

- postdoc (deadline fin mars)
- maître de conférences (réseaux)

Matthieu.Latapy@lip6.fr

Questions

