Grands réseaux d'interactions et pair-à-pair

Clémence Magnien

magnien@shs.polytechnique.fr

CREA
École polytechnique

Plan

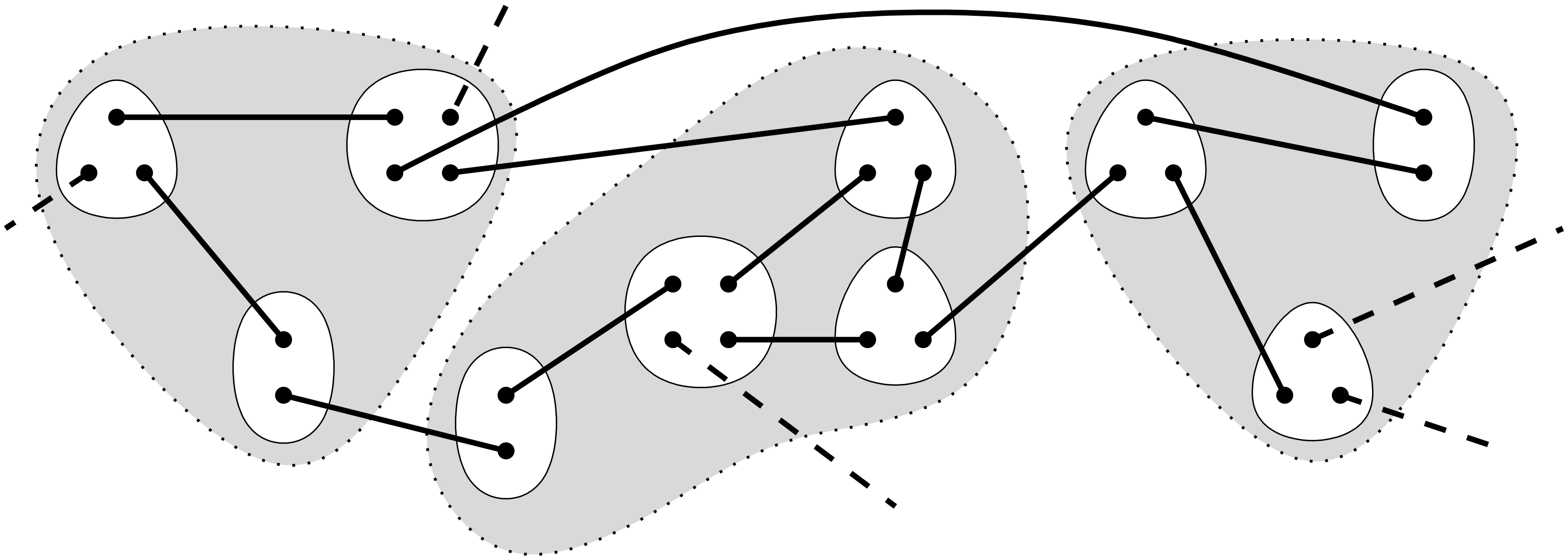
- Les Grands Réseaux d'Interactions (GRI)
- Les systèmes pair-à-pair
- Étude des GRI rencontrés dans le pair-à-pair

Plan

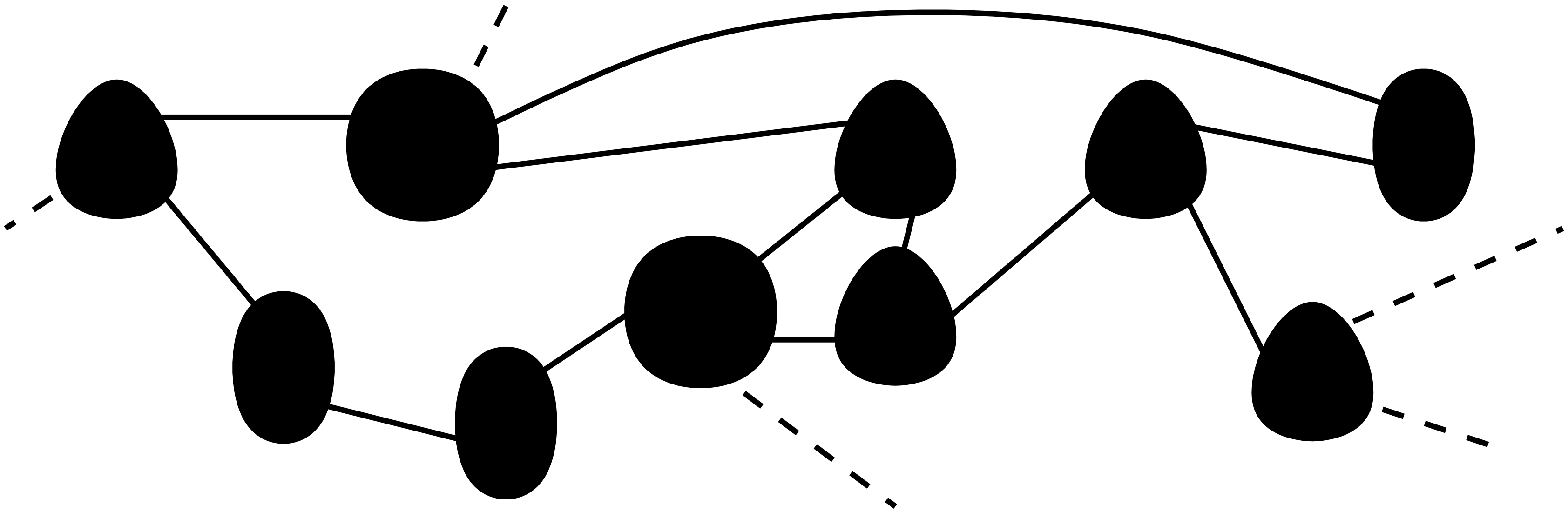
- Les Grands Réseaux d'Interactions (GRI)
- Les systèmes pair-à-pair
- Étude des GRI rencontrés dans le pair-à-pair
  - Spécificités du pair-à-pair
  - Inscription dans le contexte global

Grands Réseaux d'Interactions

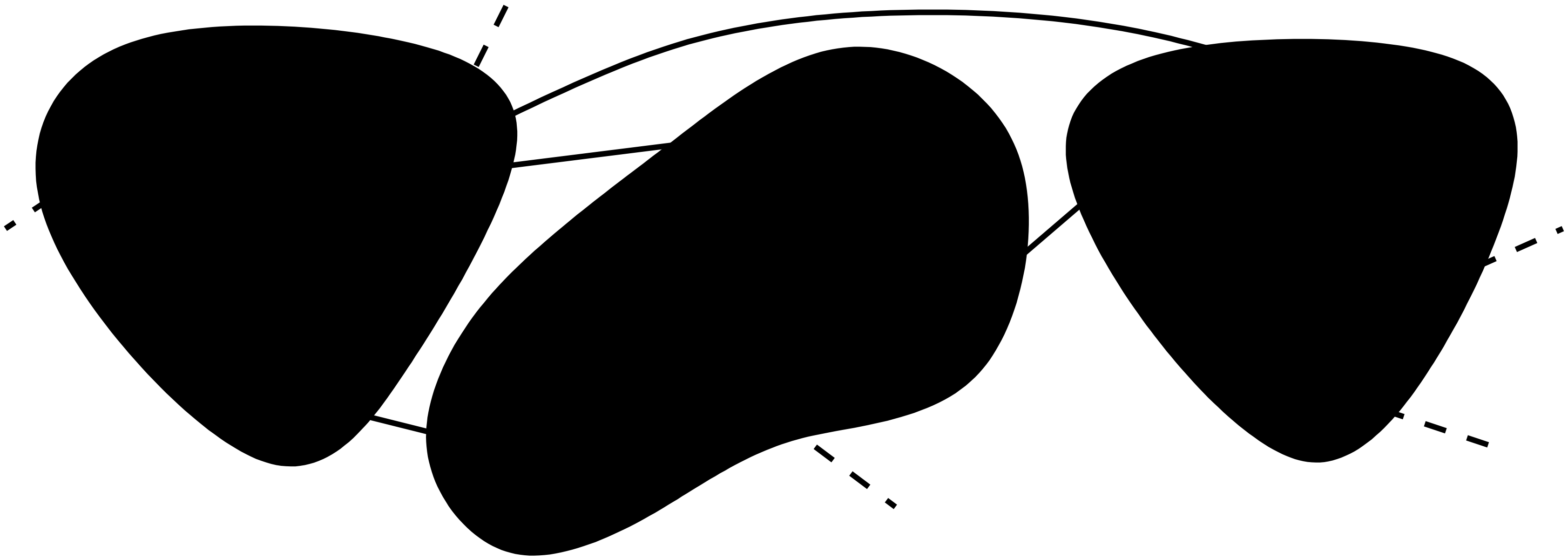
Un réseau physique : des machines et des câbles.



Un réseau physique : des machines et des câbles.



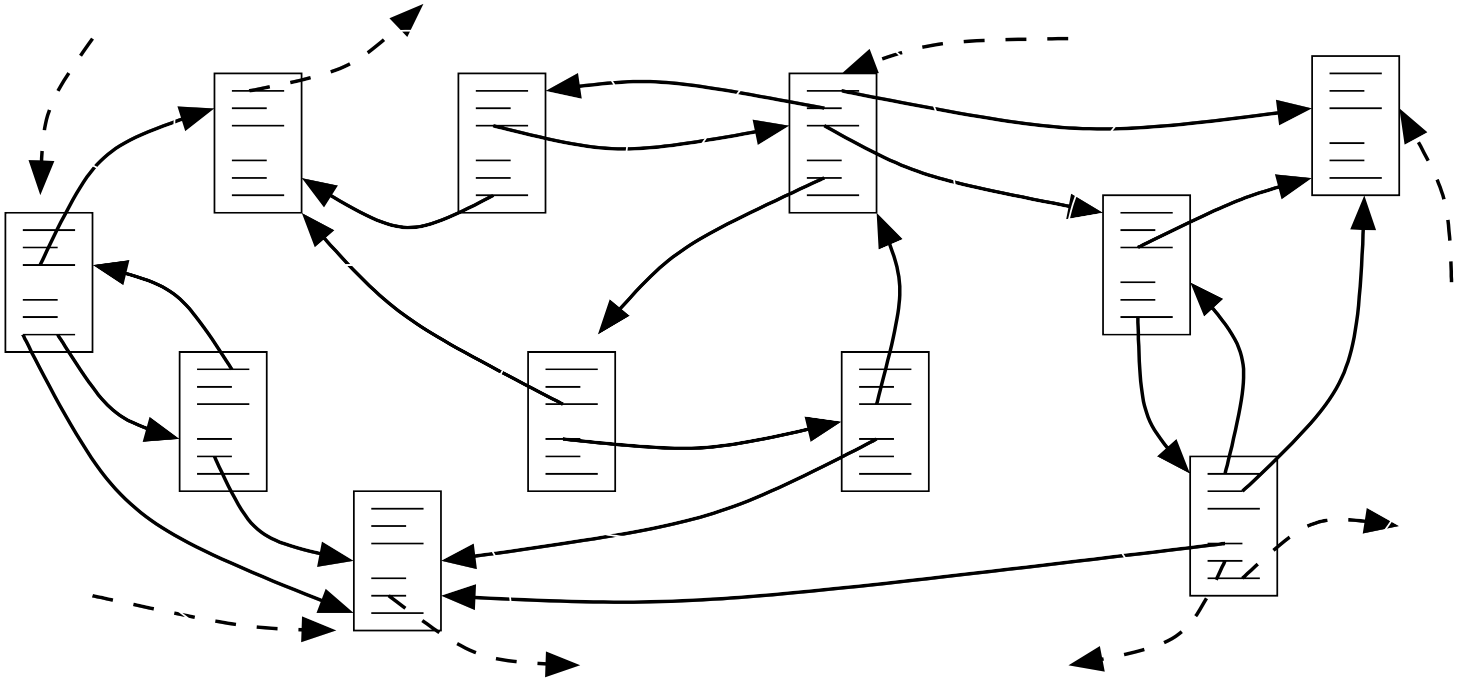
Un réseau social : des institutions et des contrats.



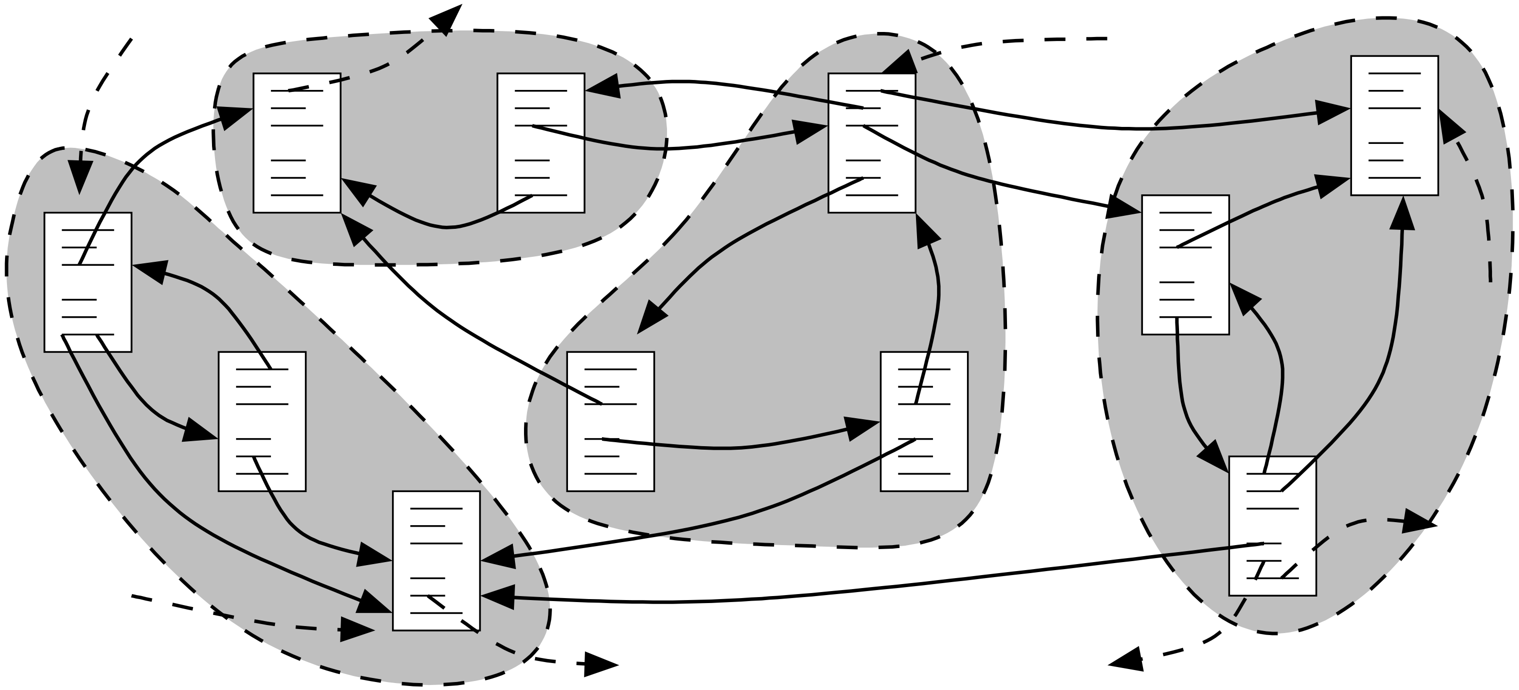
robustesse, efficacité, confidentialité, etc

centaines de milliers de sommets

Liens hypertextes (orientés) entre pages Web.



Liens hypertextes (orientés) entre sites Web.



recherche, communautés, etc

centaines de millions, milliards de sommets

Mais aussi…

En sciences de la vie : cerveau, gènes, protéines, réseaux trophiques, etc.

En sciences sociales : réseaux de collaborations professionnelles, d'amitié, de contacts sexuels, d'échanges, etc.

En linguistique : réseau de synonymie, de co-occurences, etc.

En transport : réseau routier, ferré ou aérien, distribution d'électricité, d'eau, etc.

etc, etc.

De (très) grands graphes dans des contextes (très) divers.

Un ensemble cohérent ?

distance moyenne

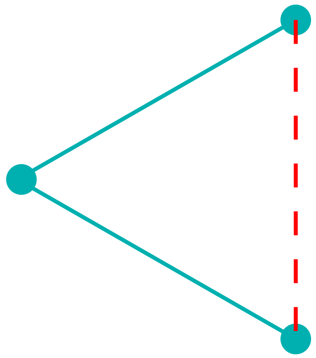
Moyenne pour tous couples de sommets de la distance entre eux (longueur d'un plus court chemin).

Un ensemble cohérent ?

distance moyenne

coefficient de clustering

Probabilité pour deux voisins d'un sommet d'être voisins entre-eux.



Un ensemble cohérent ?

distance moyenne

coefficient de clustering

distribution des degrés

Pour tout k, P_k est le nombre de sommets de degré k

- loi de Poisson : P_k décroît exponentiellement avec k.
- loi de puissance : P_k décroît polynomialement avec k.

Un ensemble cohérent ?

	en pratique
distance moyenne	courte
coefficient de clustering	grand
distribution des degrés	loi de puissance

Un ensemble cohérent ?

	en pratique	aléatoire
distance moyenne	courte	$\sim \log(n)$
coefficient de clustering	grand	$\sim \frac{1}{n}$
distribution des degrés	loi de puissance	Poisson

Un ensemble cohérent ?

	en pratique	aléatoire
distance moyenne	courte	$\sim \log(n)$
coefficient de clustering	grand	$\sim \frac{1}{n}$
distribution des degrés	loi de puissance	Poisson

$\hookrightarrow$ **graphes en pratique $\neq$ graphes aléatoires.**

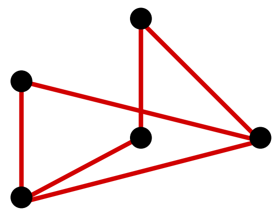
[Watts et Strogatz, 1998]

Les GRI apparaissent aussi dans le pair-à-pair

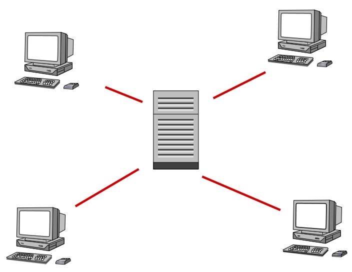
Les GRI apparaissent aussi dans le pair-à-pair

Description du contexte

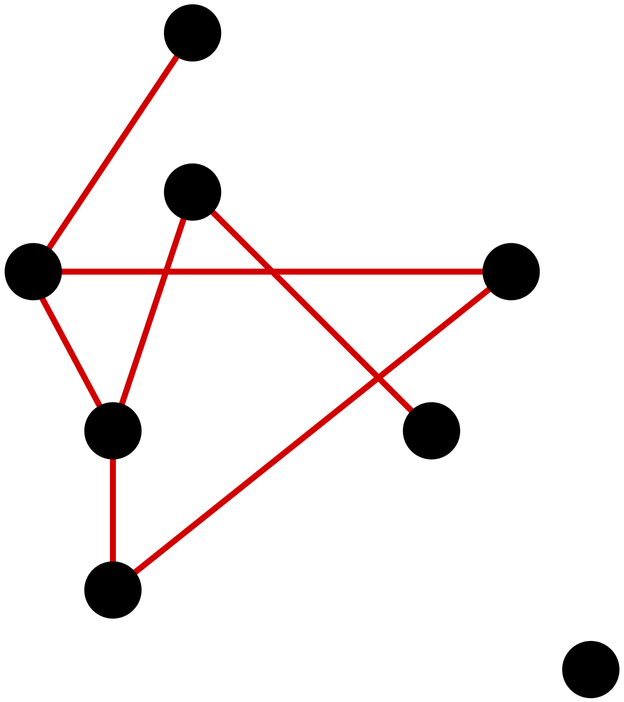
- Système décentralisé



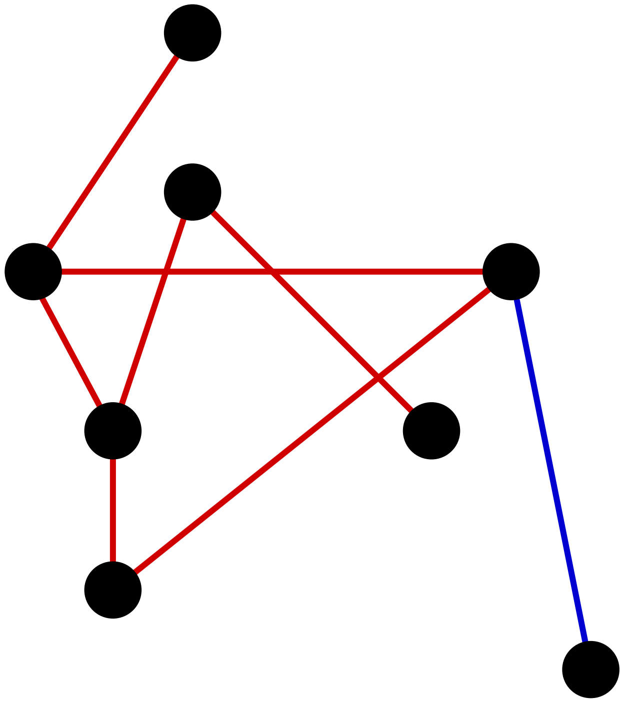
- Système avec serveurs



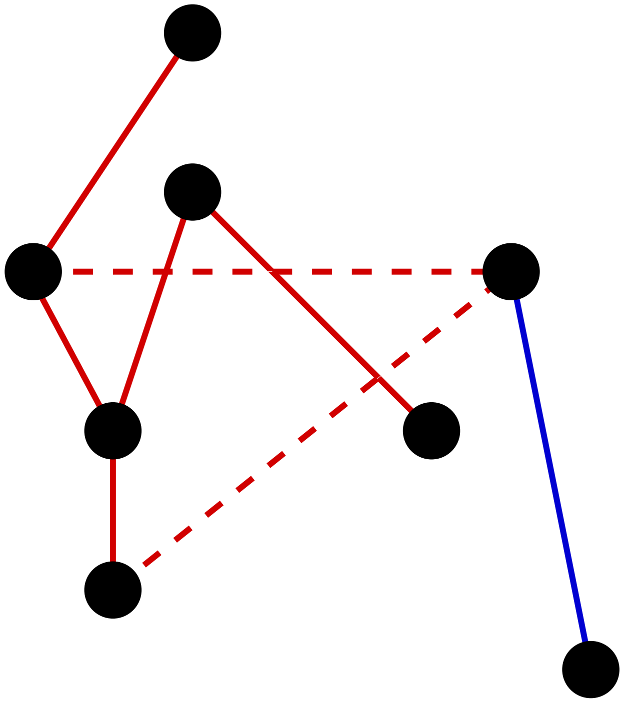
Système décentralisé



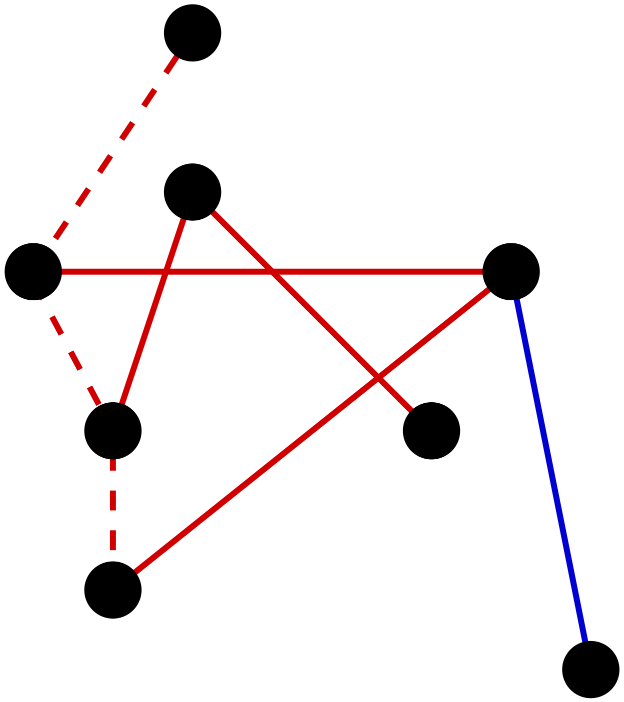
Système décentralisé



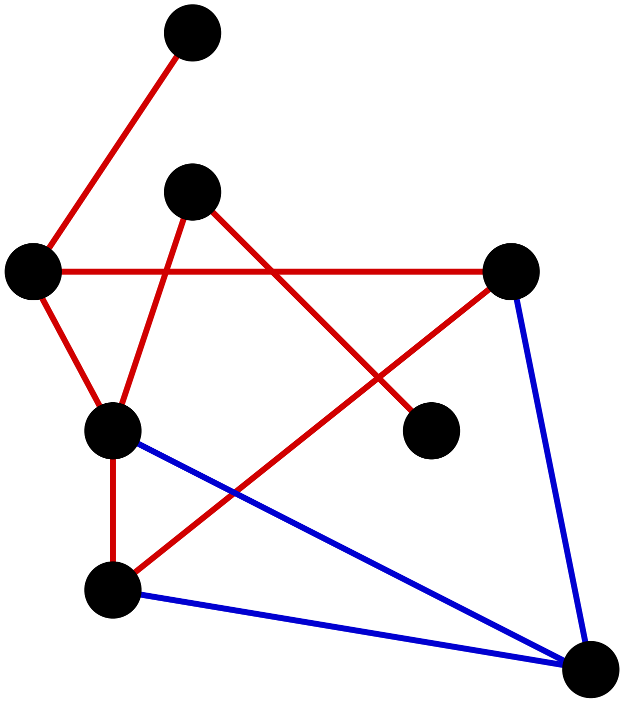
Système décentralisé



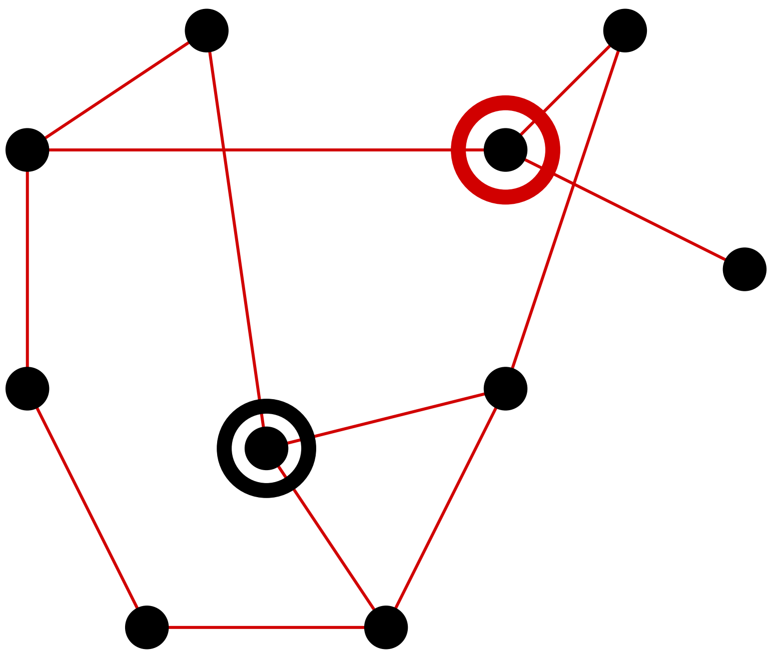
Système décentralisé



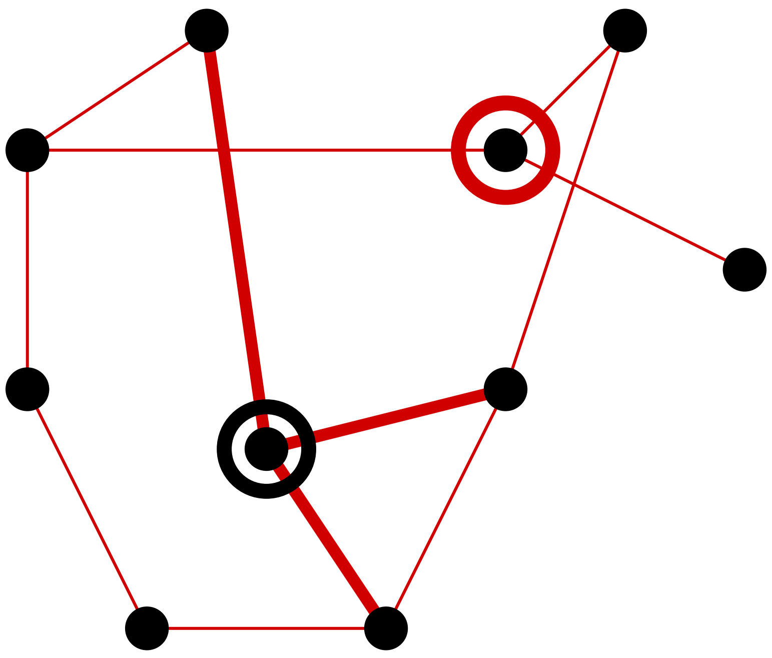
Système décentralisé



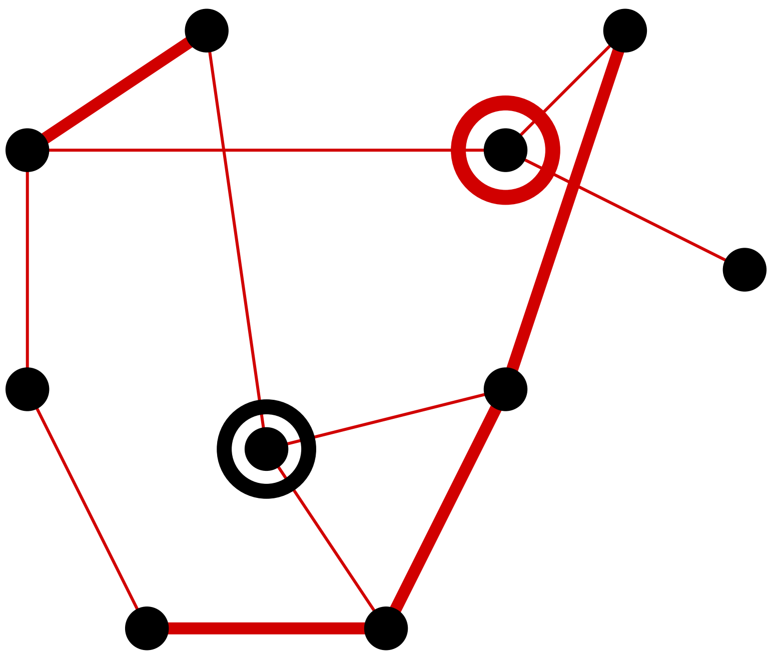
Recherche



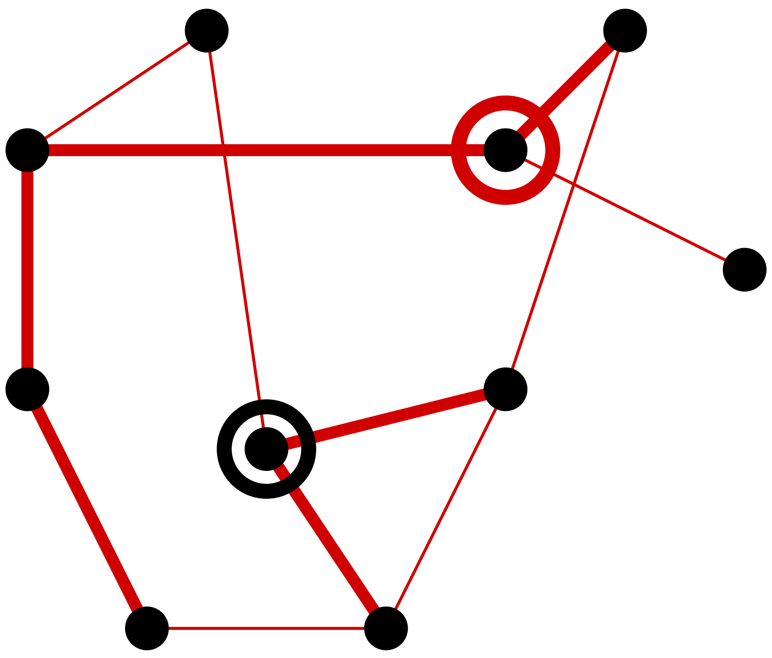
Recherche



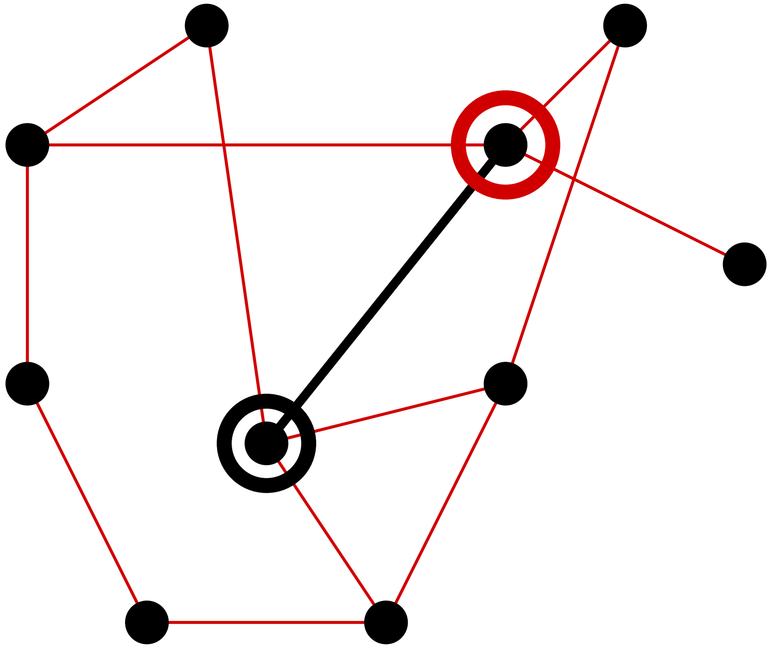
Recherche



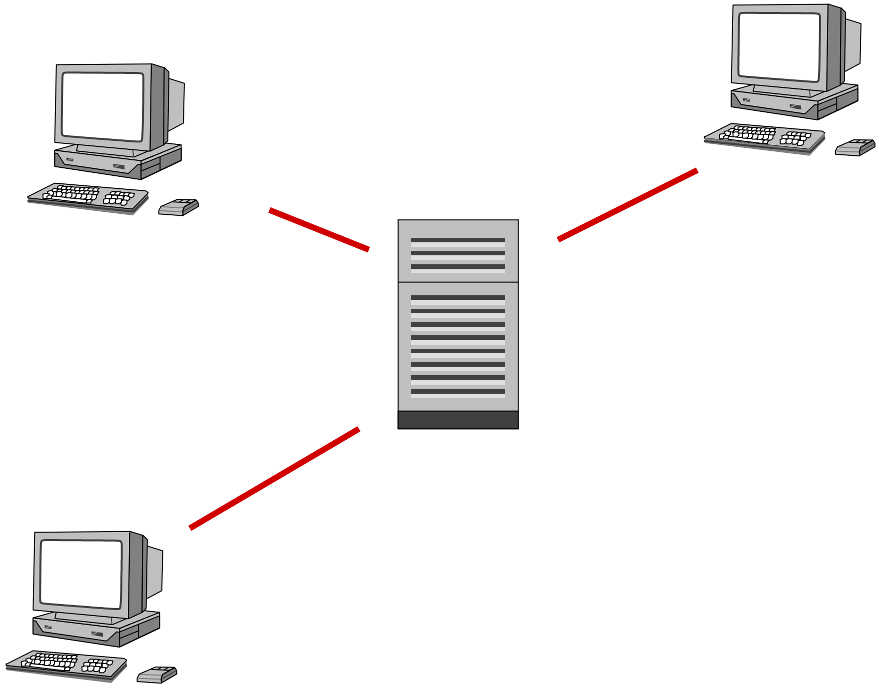
Recherche



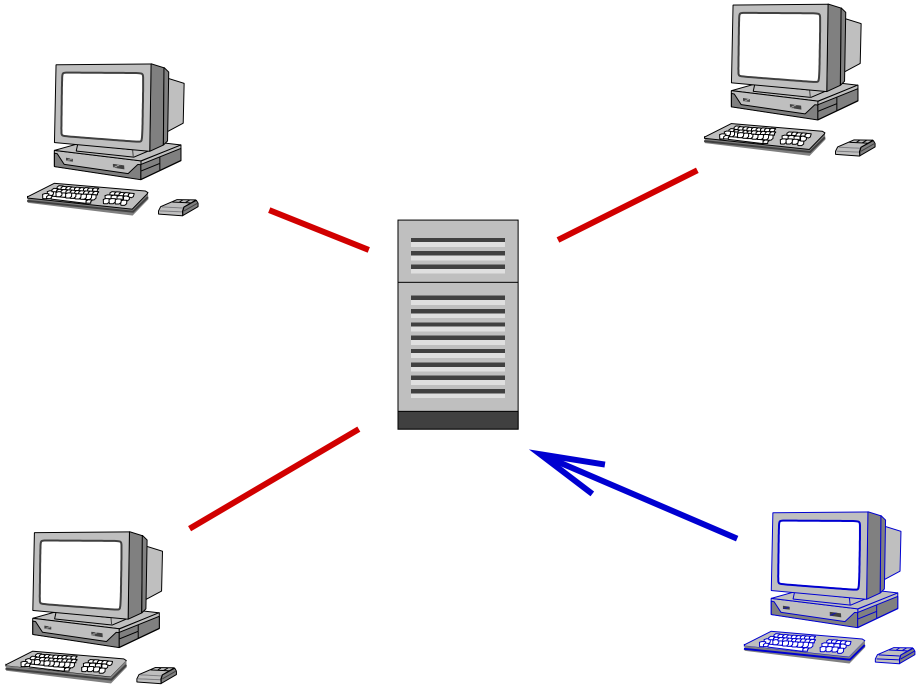
Recherche



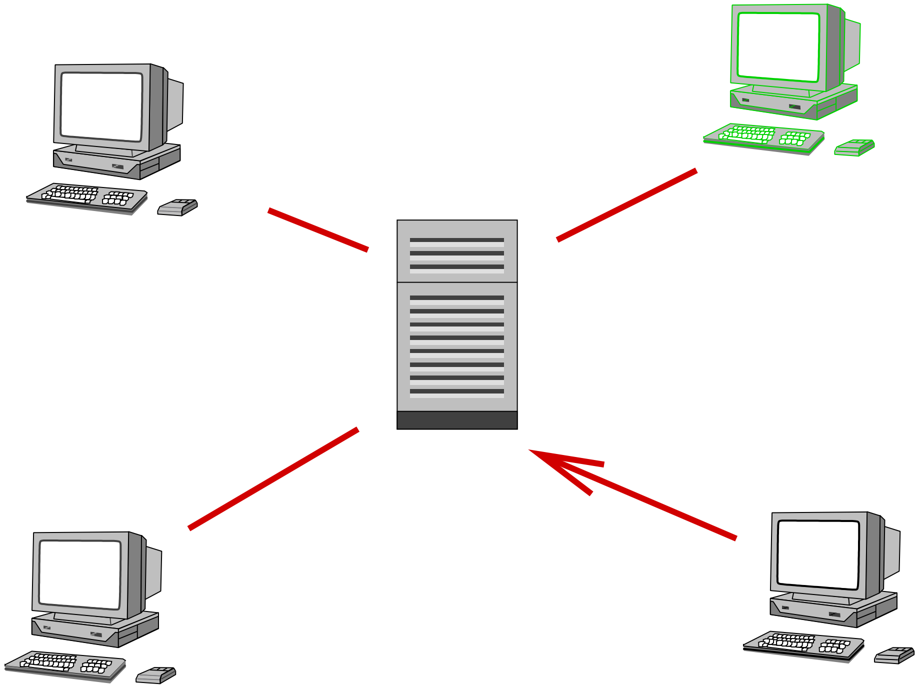
Fonctionnement d'un système centralisé



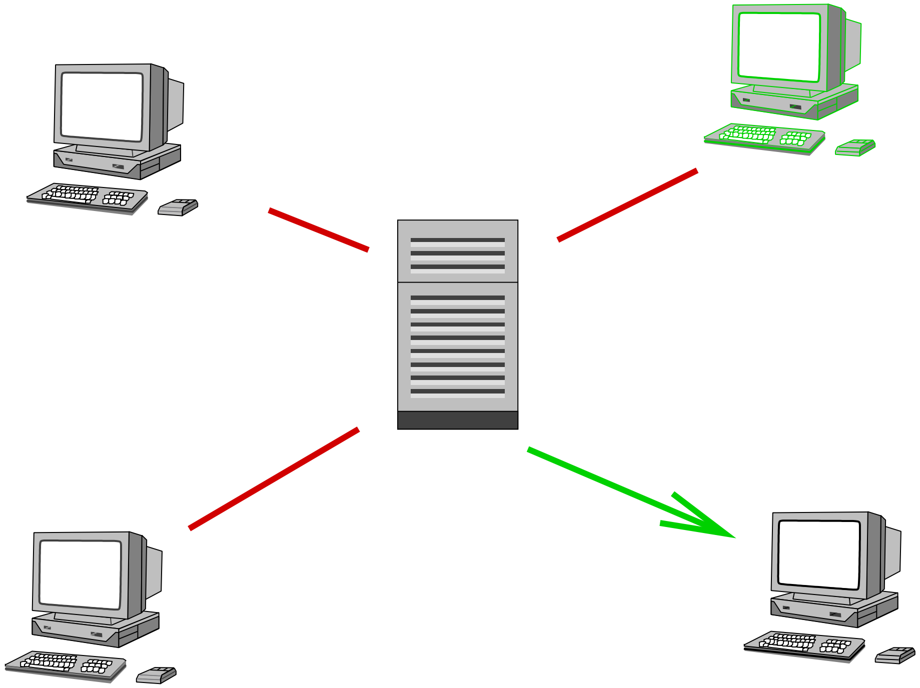
Fonctionnement d’un système centralisé



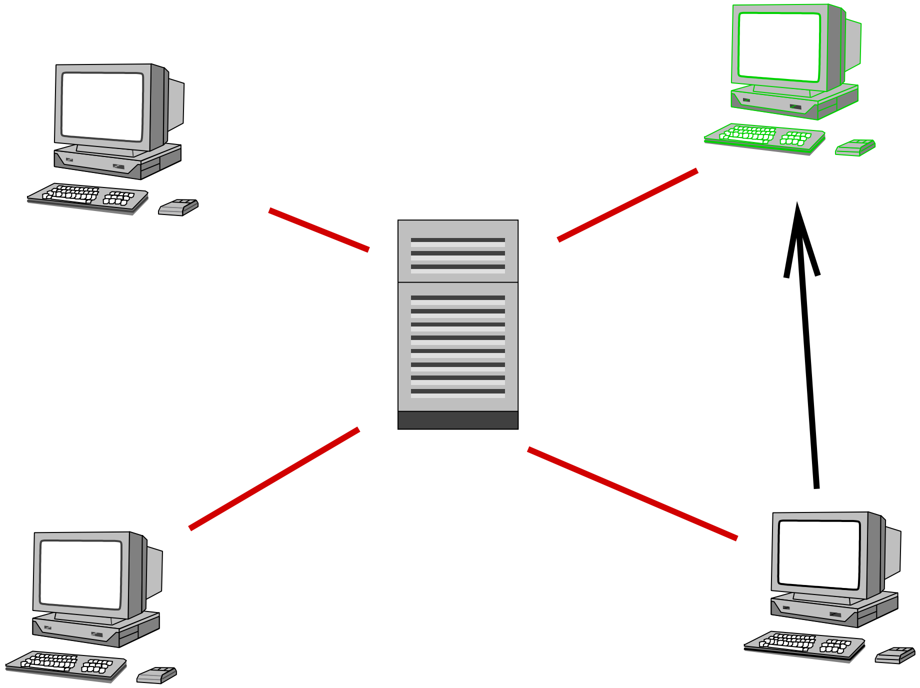
Requête



Requête

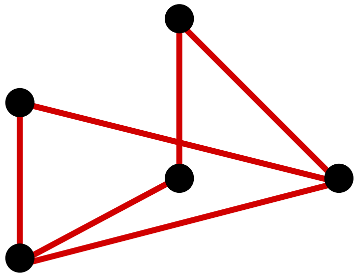


Requête



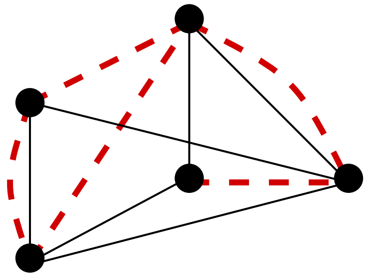
Deux types de GRI rencontrés :

Qui est connecté à qui (*overlay*)



Un réseau technique

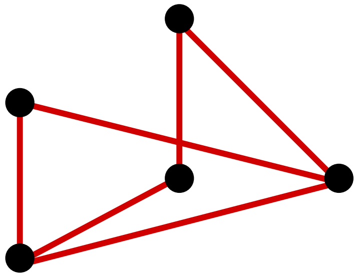
Qui échange des données avec qui



Un réseau social

Deux types de GRI rencontrés :

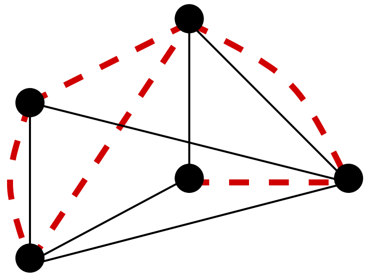
Qui est connecté à qui (*overlay*)



Un réseau technique

Quoique...

Qui échange des données avec qui



Un réseau social

Quoique...

GRI et pair-à-pair

Un axe d'étude possible

Simulation de protocoles

Modélisation

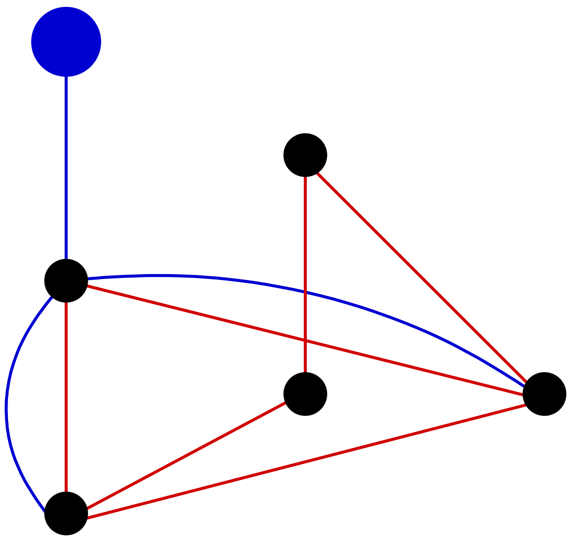
Analyse

Mesure

Mesure

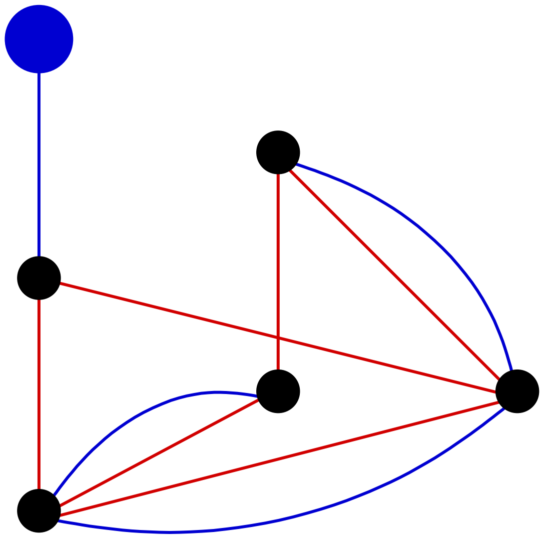
[Ripeanu et al 2002, . . .]

Crawl



[Ripeanu et al 2002, . . .]

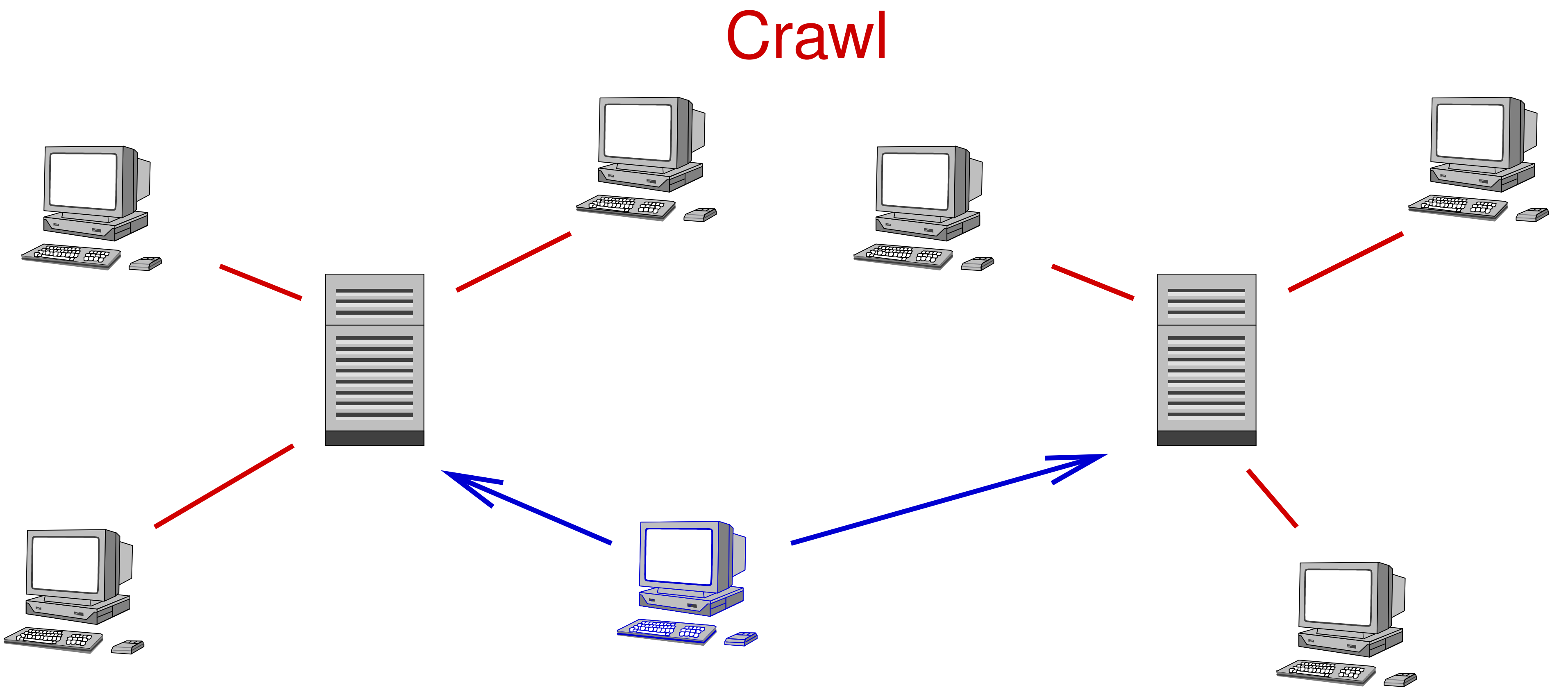
Crawl



Mesure sur les données

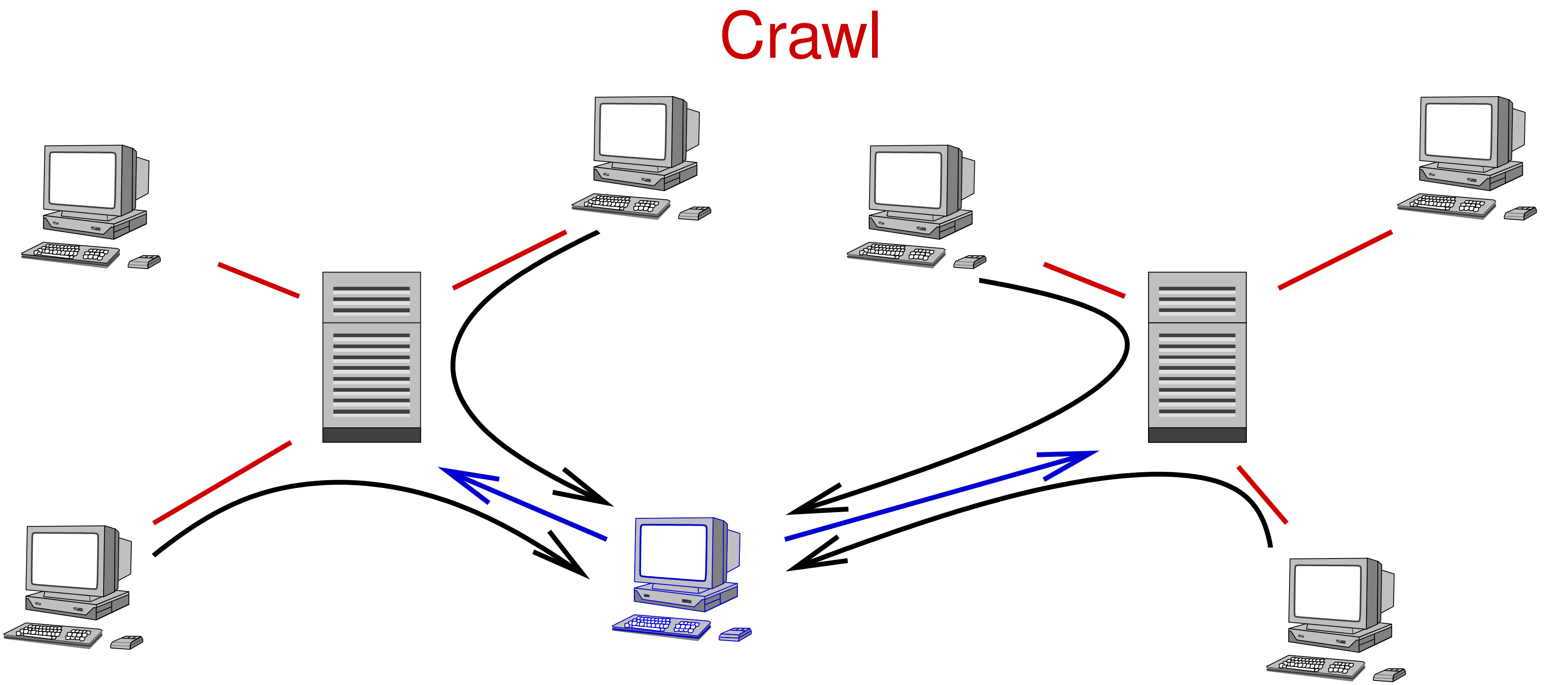
[Le Fessant et al 2004]

Qui a quelles données ?



[Le Fessant et al 2004]

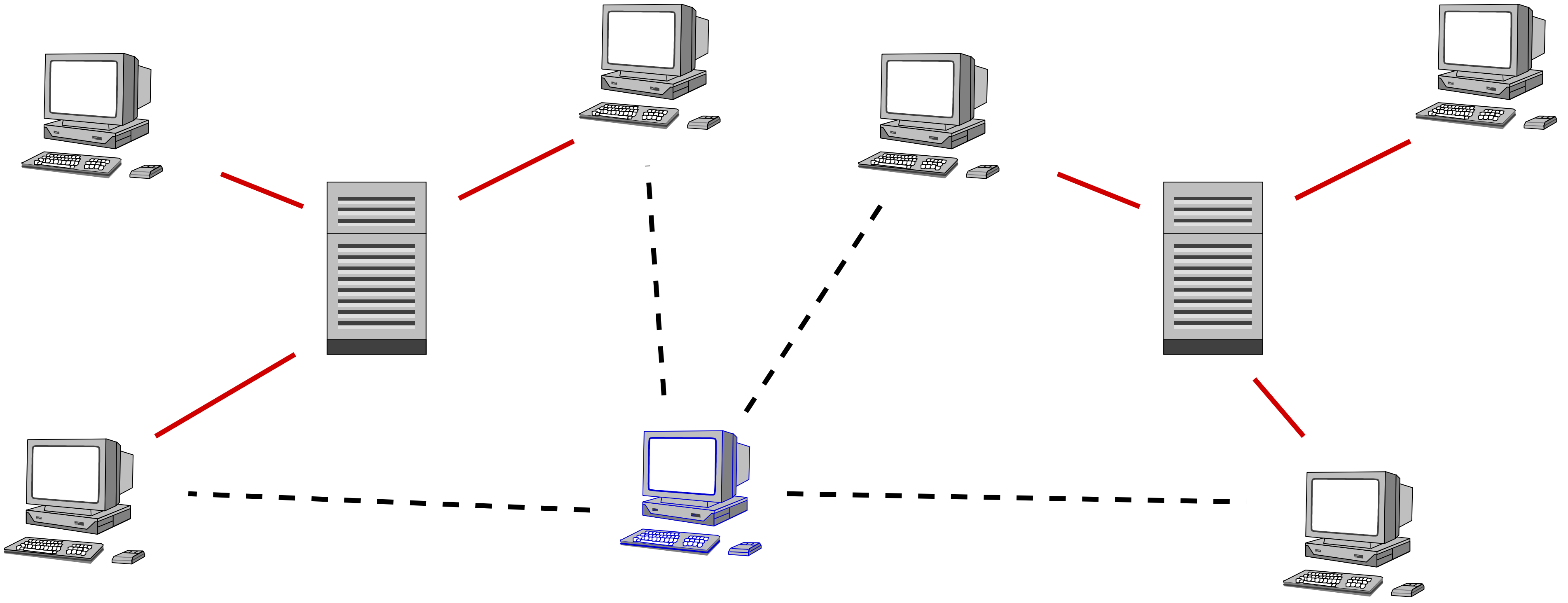
Qui a quelles données ?



[Le Fessant et al 2004]

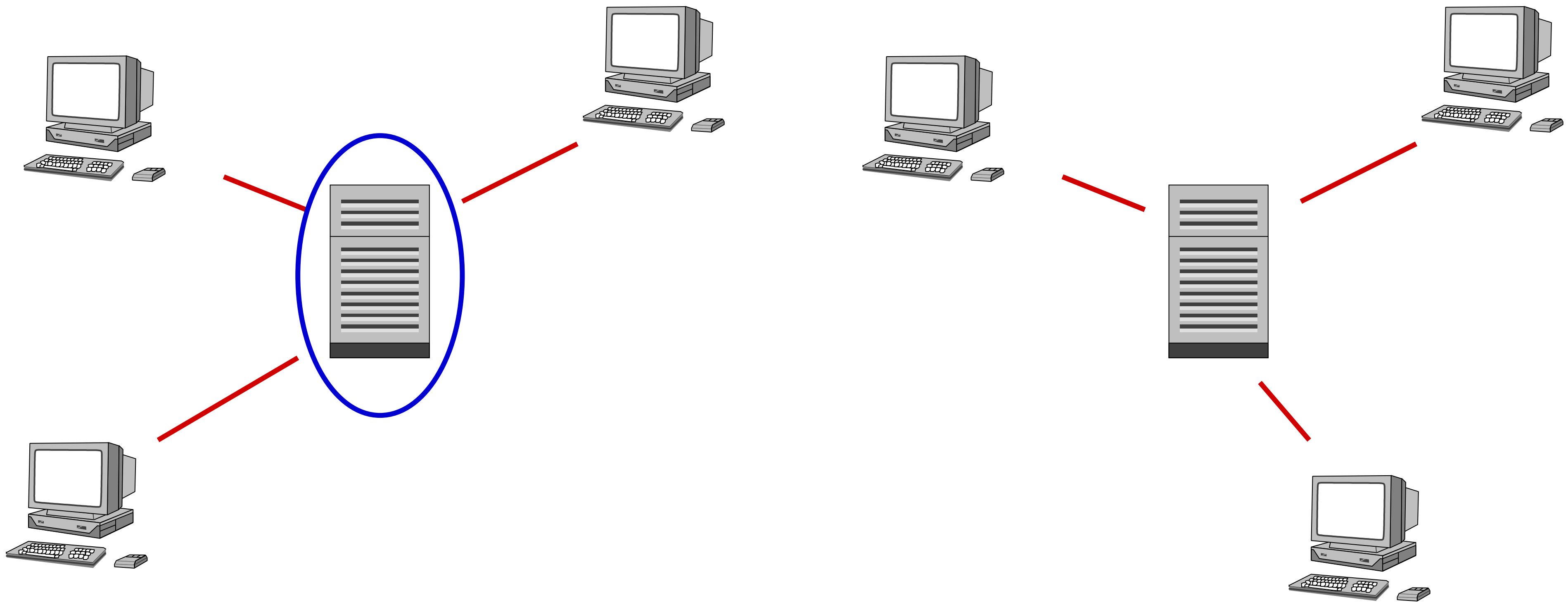
Qui a quelles données ?

Crawl



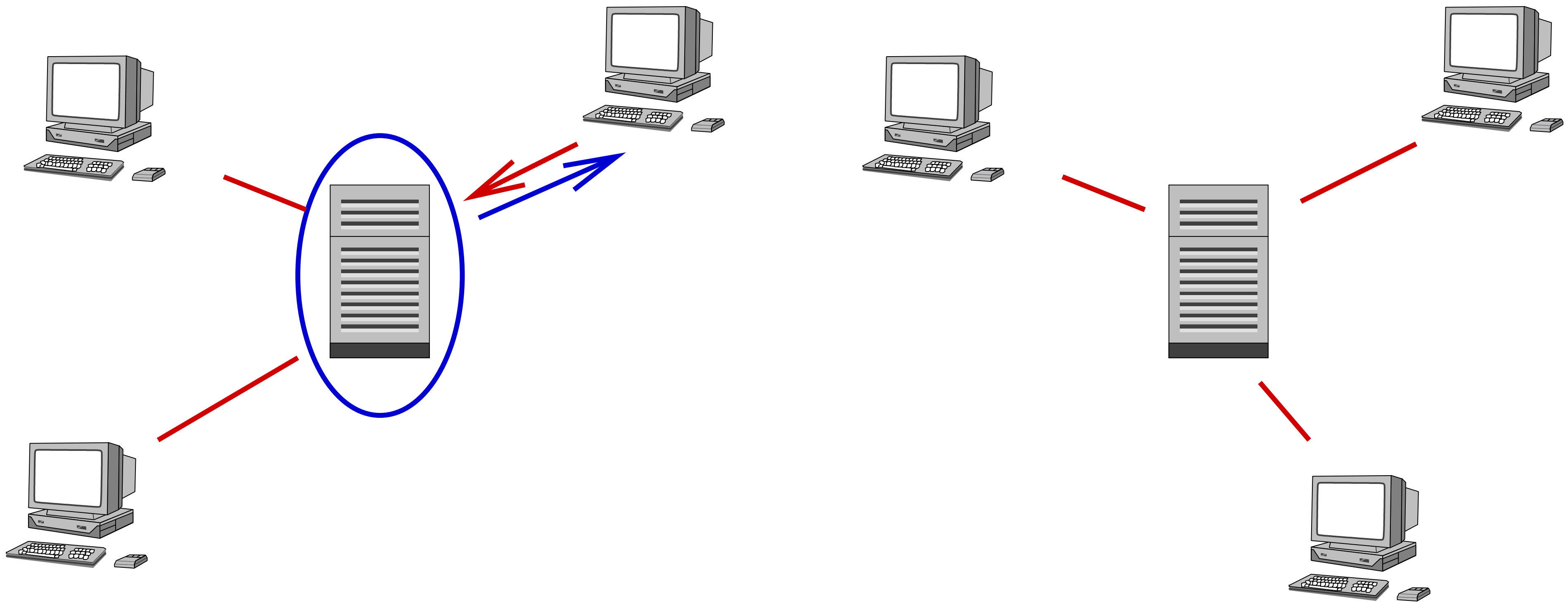
Mesure sur les données

[Guillaume et al 2004]



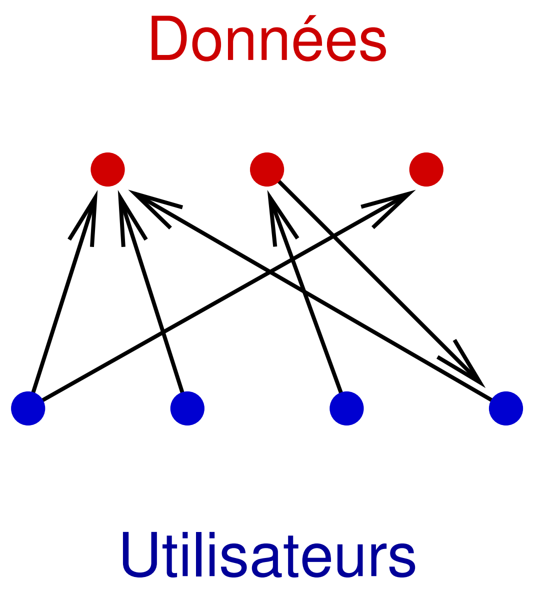
Mesure sur les données

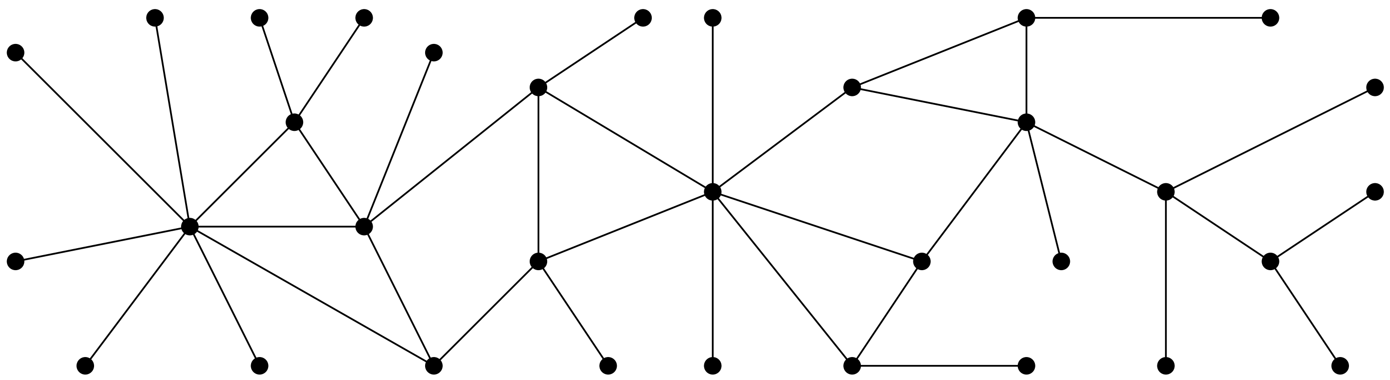
[Guillaume et al 2004]



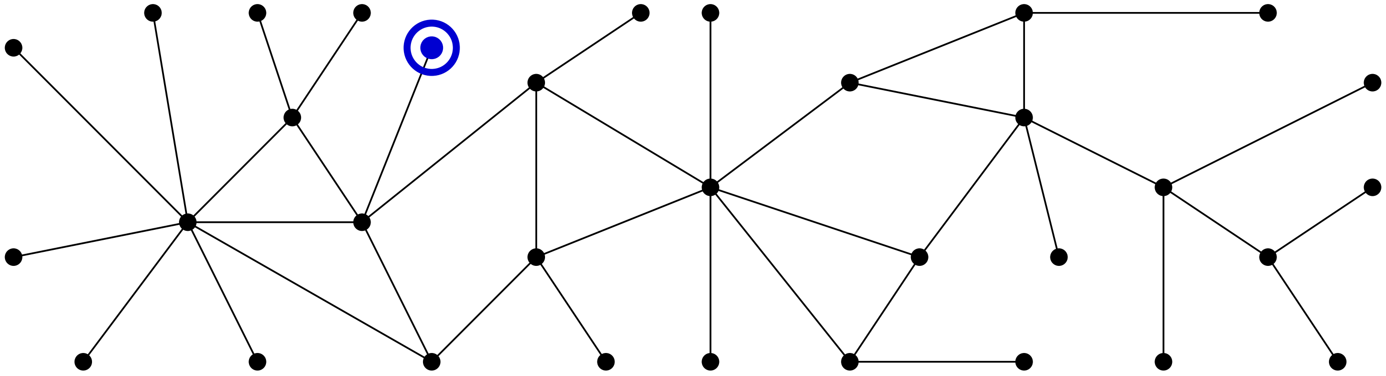
Mesure sur les données

[Guillaume et al 2004]

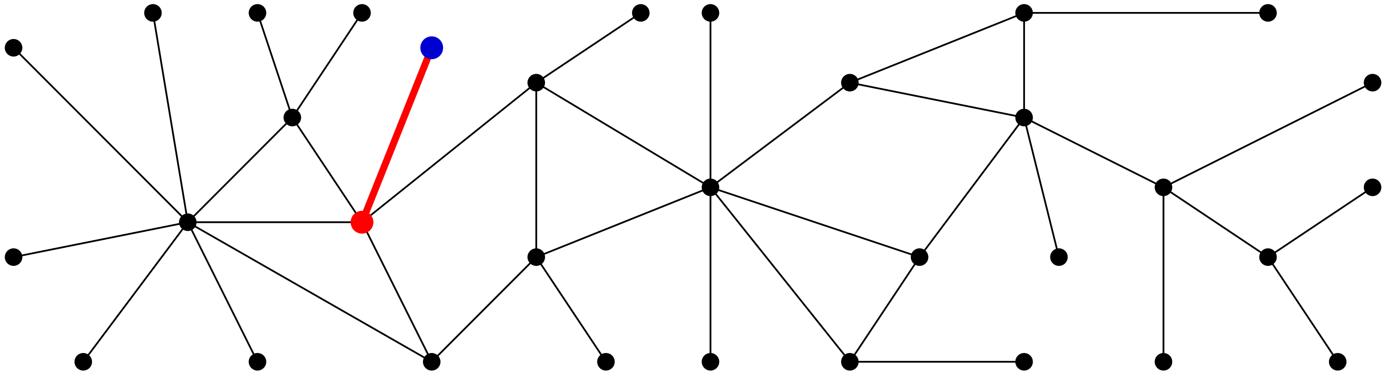




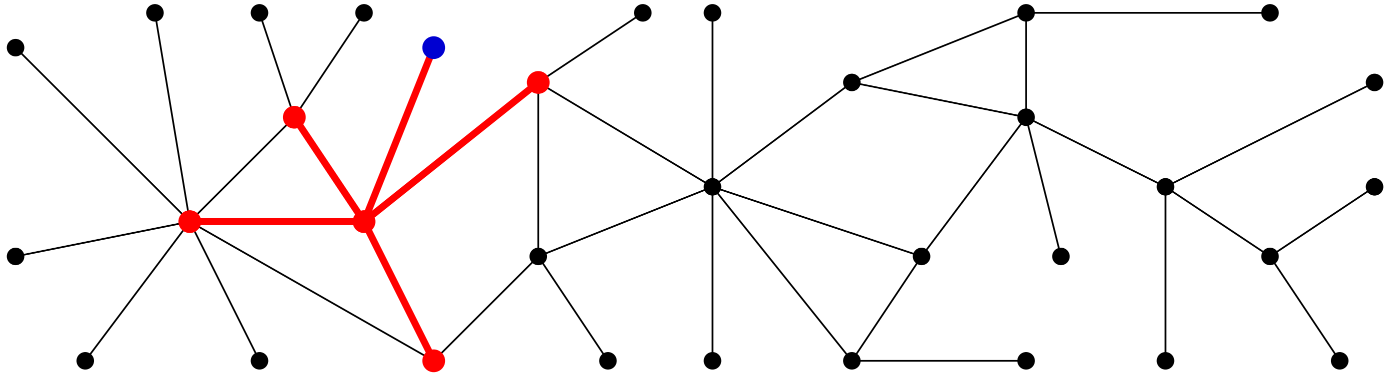
Opération de mesure = parcours en largeur



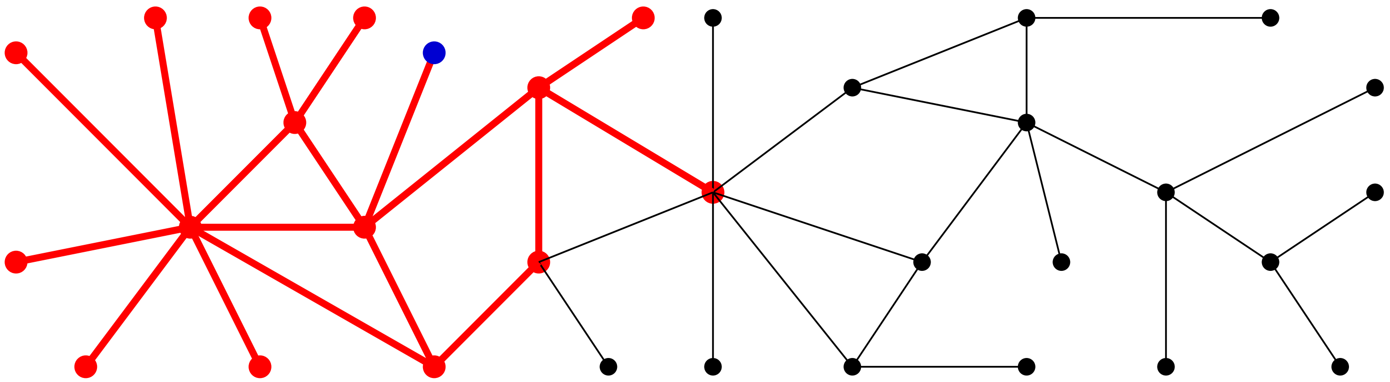
Opération de mesure = parcours en largeur



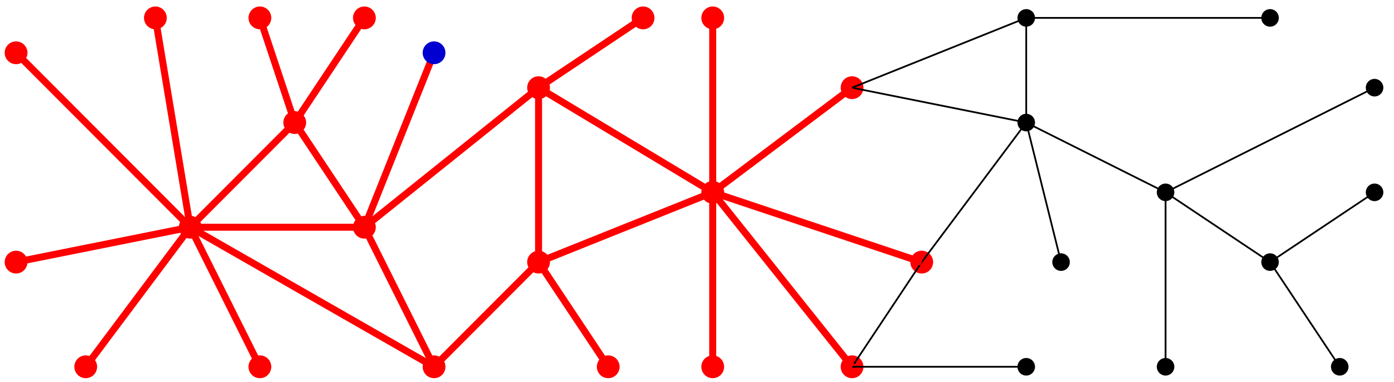
Opération de mesure = parcours en largeur



Opération de mesure = parcours en largeur

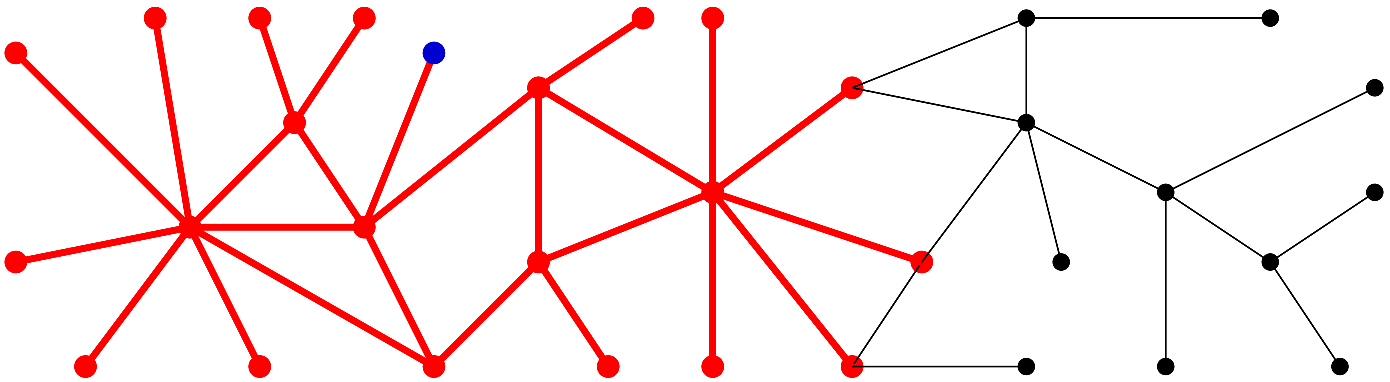


Opération de mesure = parcours en largeur



Opération de mesure = parcours en largeur

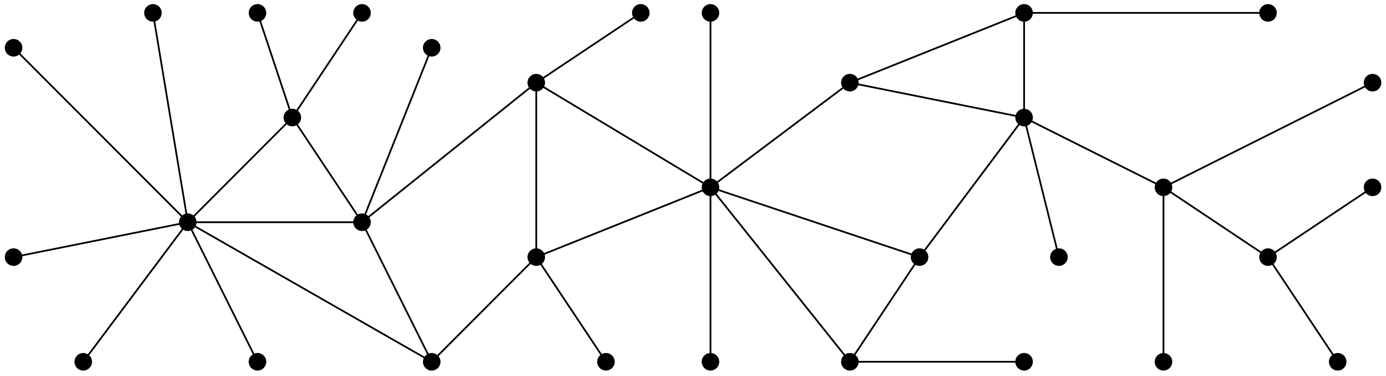
actuellement : 5 à 10 % du total



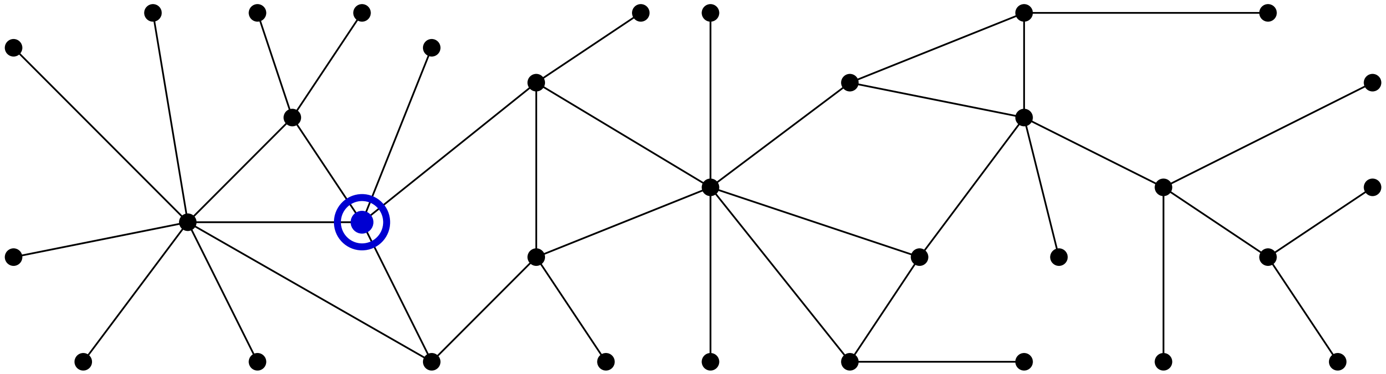
Opération de mesure = parcours en largeur

Représentativité ? Biais induit ? ...

Mesure de l’Internet

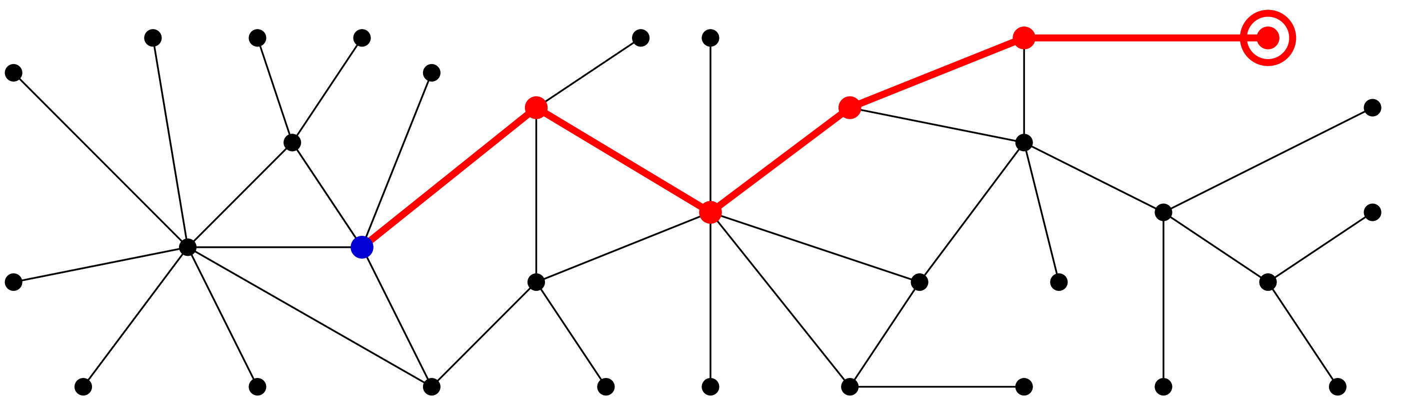


Opération de mesure = chemins



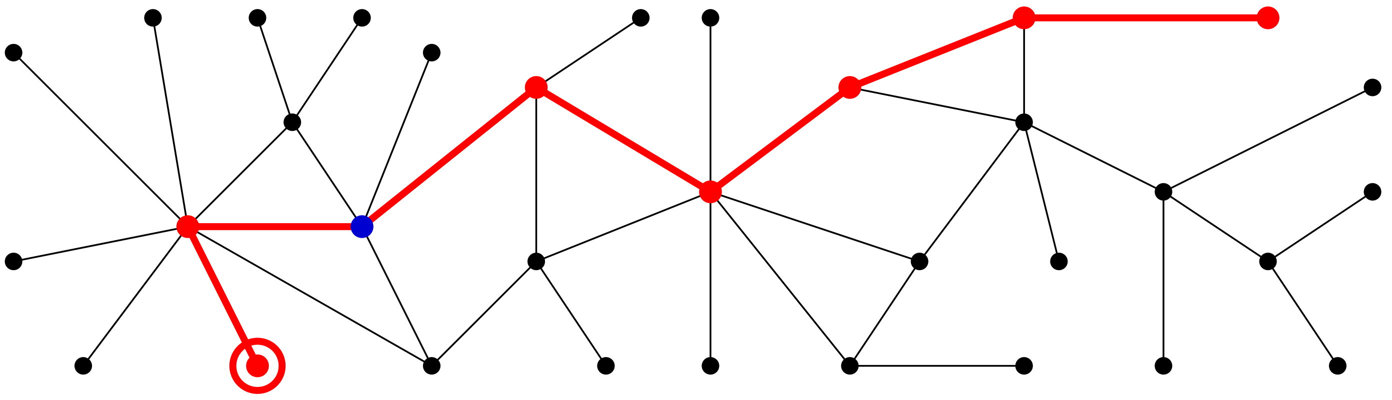
Opération de mesure = chemins

Mesure de l'Internet

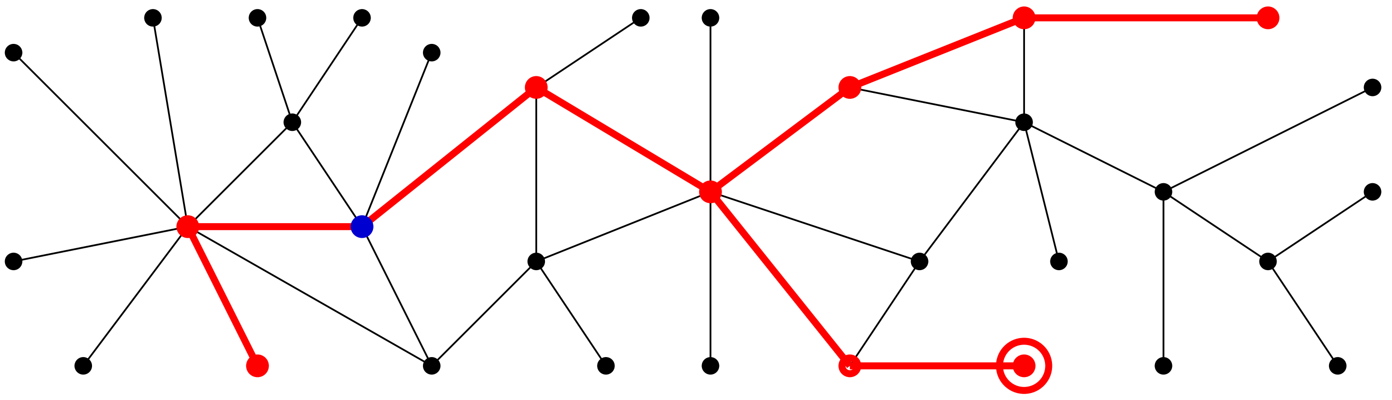


Opération de mesure = chemins

Mesure de l'Internet

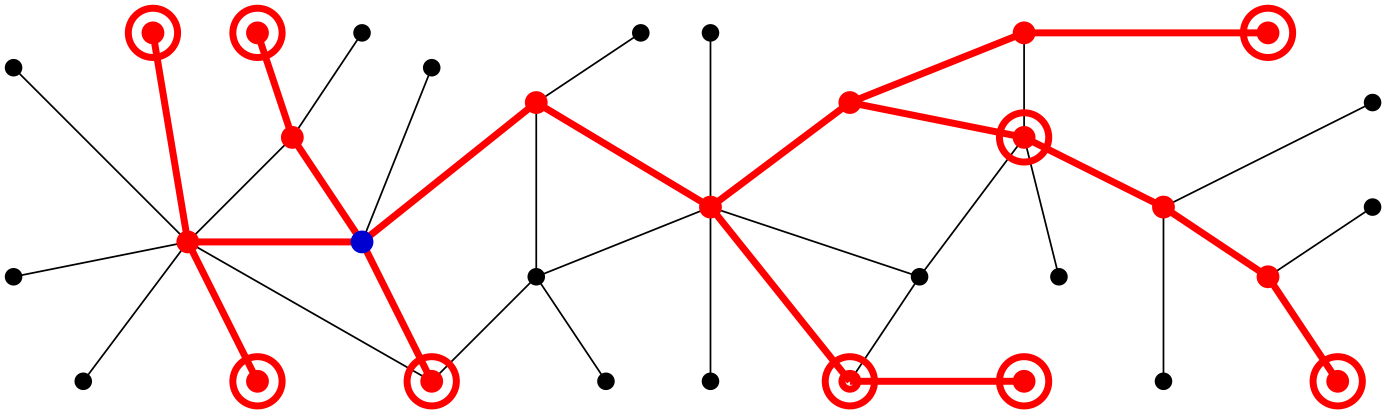


Opération de mesure = chemins



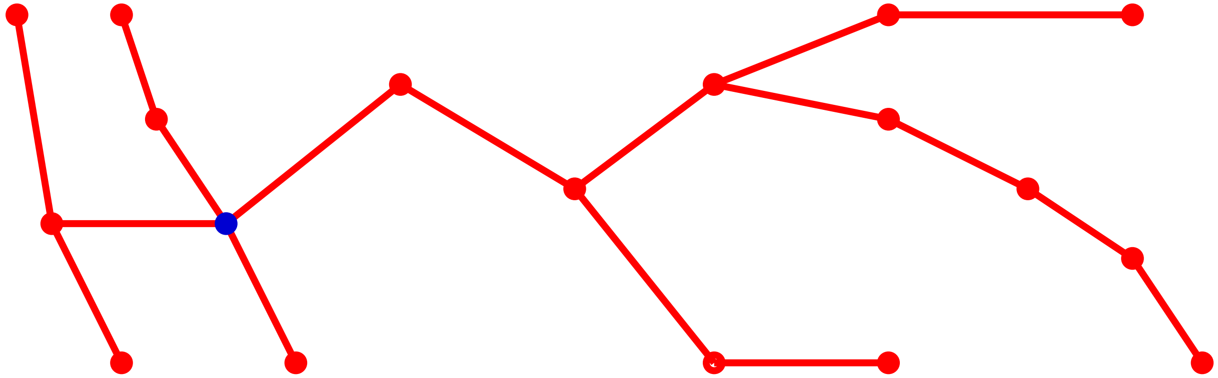
Opération de mesure = chemins

Mesure de l'Internet



Opération de mesure = chemins

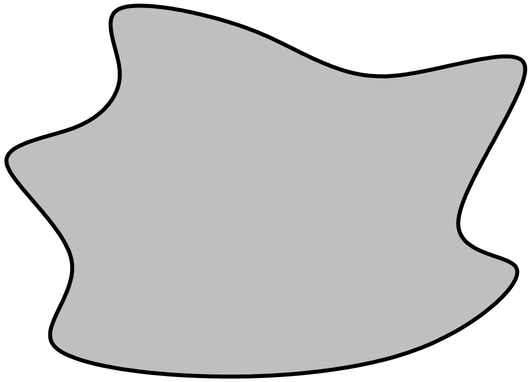
actuellement : ~ 10 sources, $\sim 10^5$ destinations



Opération de mesure = chemins

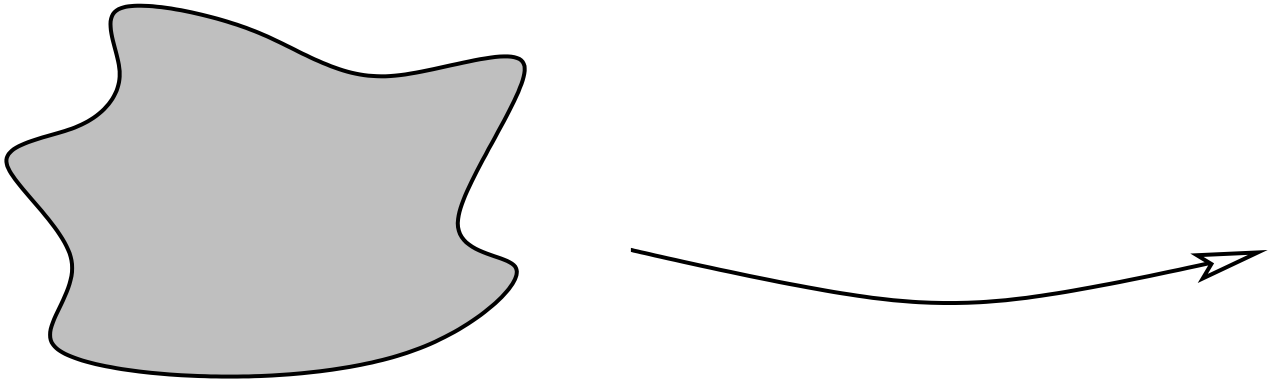
Représentativité ? Biais induit ? ...

Problématique générale : métrologie



Un graphe

Problématique générale : métrologie

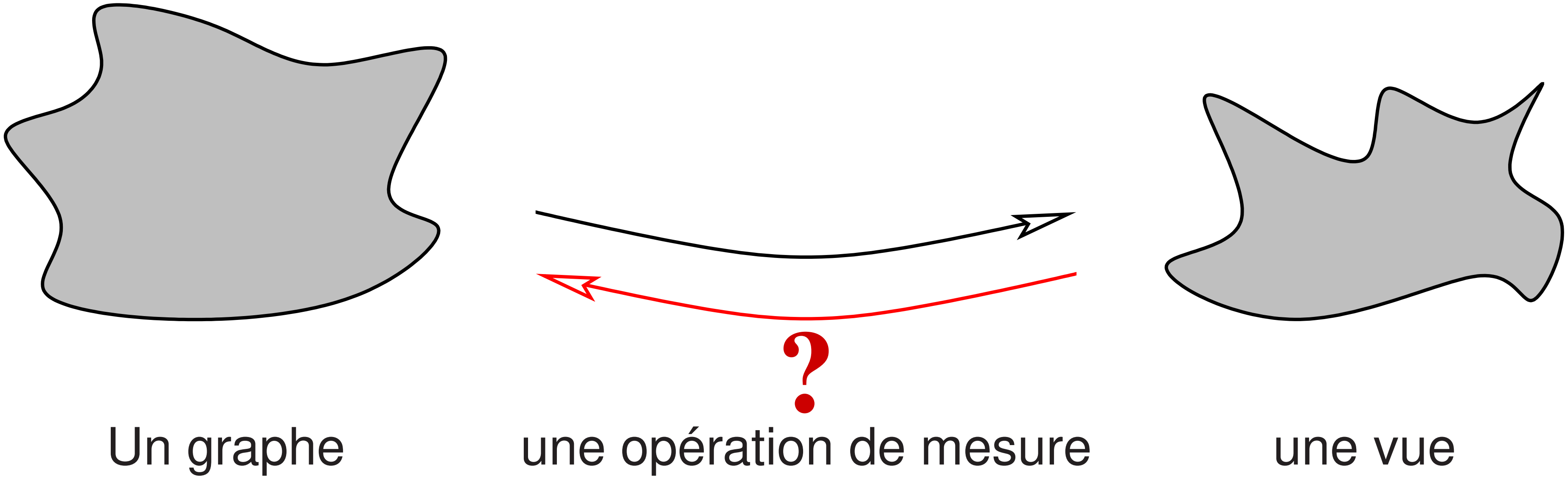


Un graphe une opération de mesure

Problématique générale : métrologie

Un graphe | une opération de mesure | une vue

Problématique générale : métrologie



Influence sur les propriétés observées ?

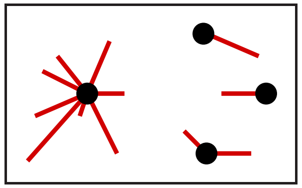
Influence des propriétés sur l'observation ?

Opérations de mesure ciblées ?

...

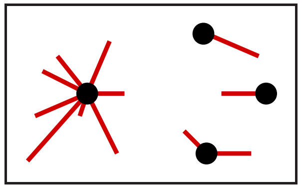
Analyse

Overlay

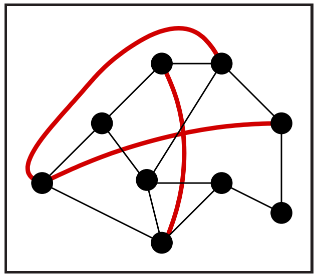


Charge,. . .

Overlay

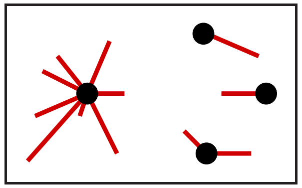


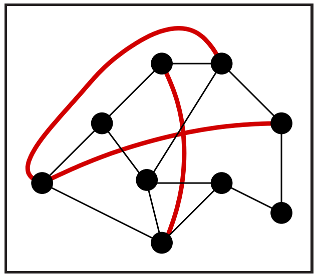
Charge,. . .



Efficacité

Overlay

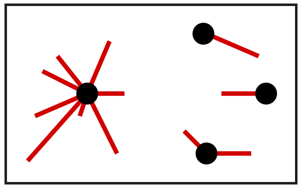

Charge,. . .


Efficacité

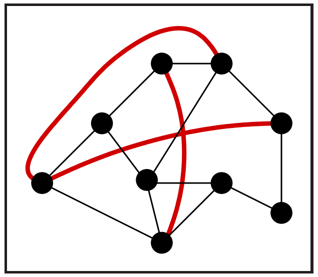
Données

Aspects sociaux

Overlay



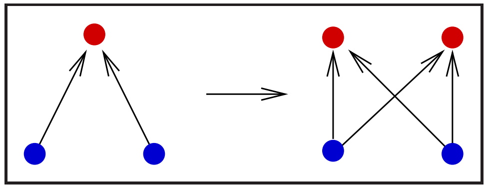
Charge,. . .



Efficacité

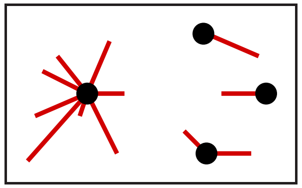
Données

$\boxed{\text{Aspects sociaux}}$

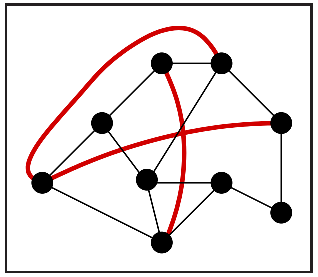


Clustering sémantique
[Le fessant et al, 2004]

Overlay



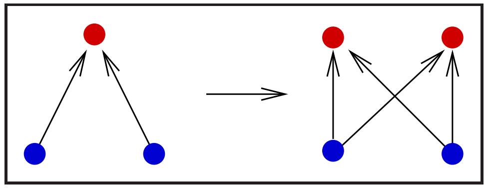
Charge,. . .



Efficacité

Données

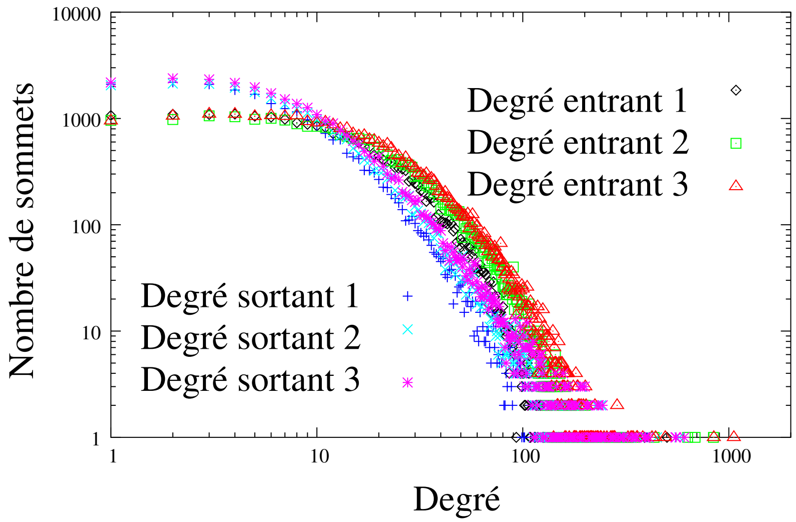
Aspects sociaux

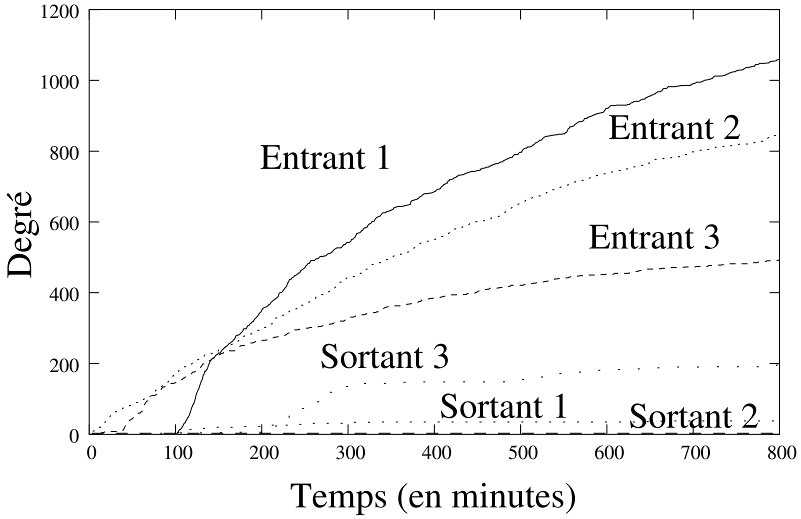


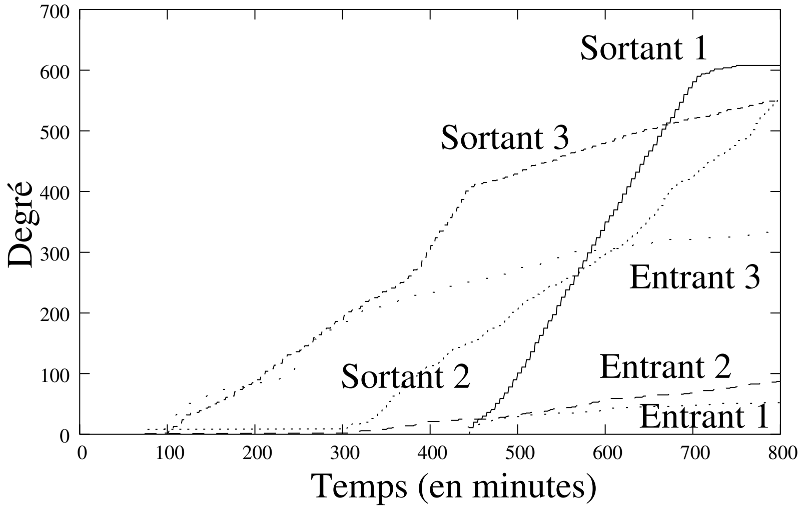
Clustering sémantique
[Le fessant et al, 2004]

Détection de comportements typiques

[Guillaume et al, 2004]







Plusieurs groupes de comportement

À quoi ressemblent ces graphes ?

Problématique générale

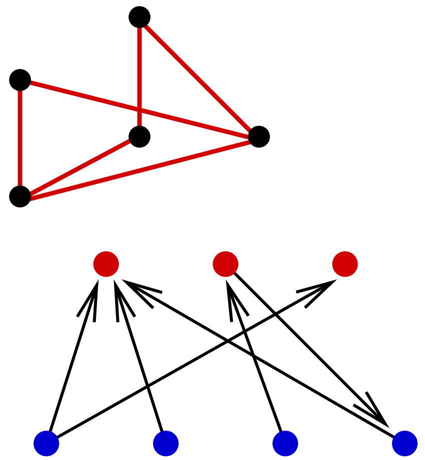
À quoi ressemblent ces graphes ?

Comment les distinguer d'autres graphes ?

Modélisation

Modélisation d'un système pair-à-pair

- overlay
- échanges
- ...



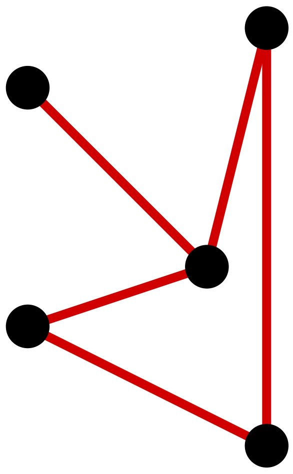
En général

Deux types de modèles

- Processus de construction
- Tirage aléatoire

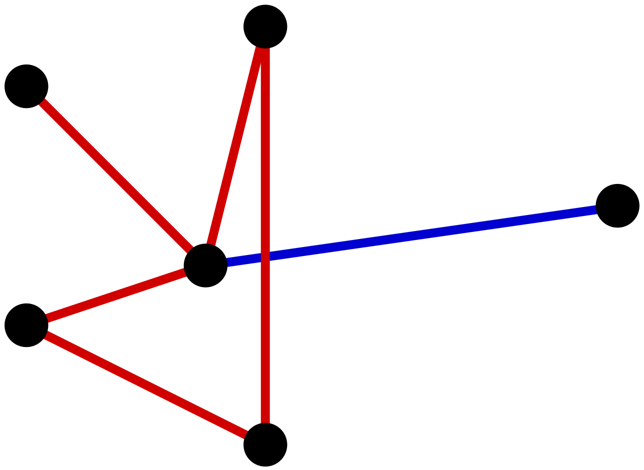
Attachement préférentiel

[Albert, Barabási, 99]



Attachement préférentiel

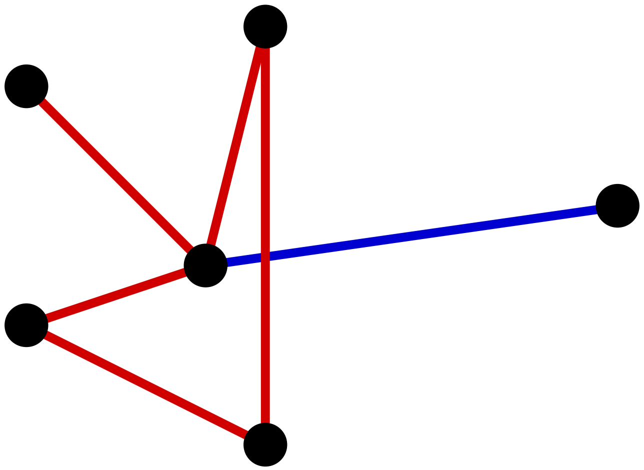
[Albert, Barabási, 99]



Loi de puissance pour les degrés

Attachement préférentiel

[Albert, Barabási, 99]



Loi de puissance pour les degrés

Nombreuses variantes

Graphes avec distribution des degrés donnée

[Molloy et Reed, 1995]

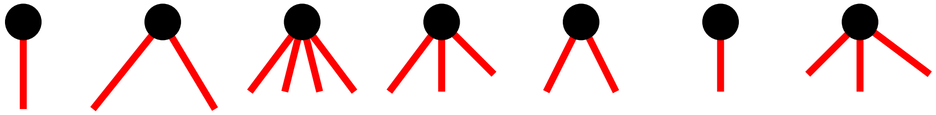
Distribution des degrés

$$p_1, p_2, p_3, \ldots$$

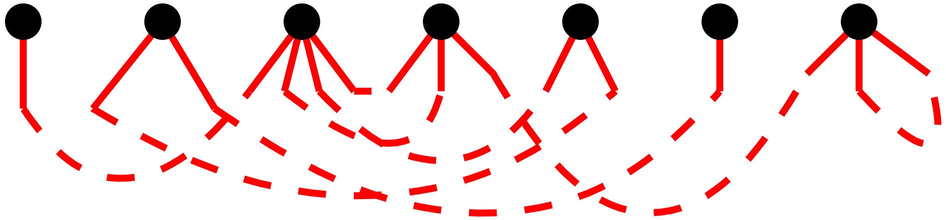
Tirer les degrés des sommets selon la distribution

$$1 \quad 2 \quad 4 \quad 3 \quad 2 \quad 1 \quad 3$$

Associer à chaque sommet des demi-arêtes

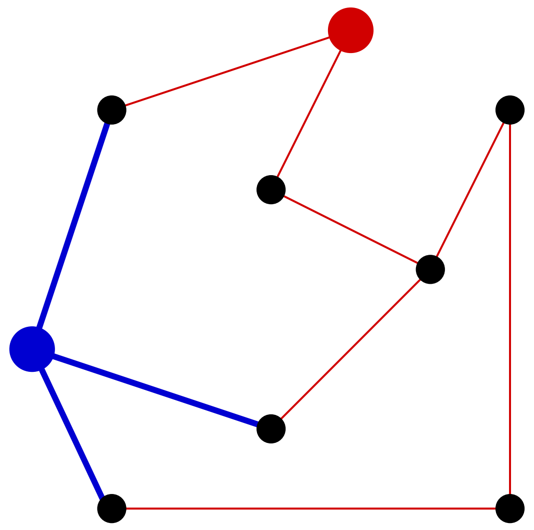


Tirer au hasard des paires de demi-arêtes



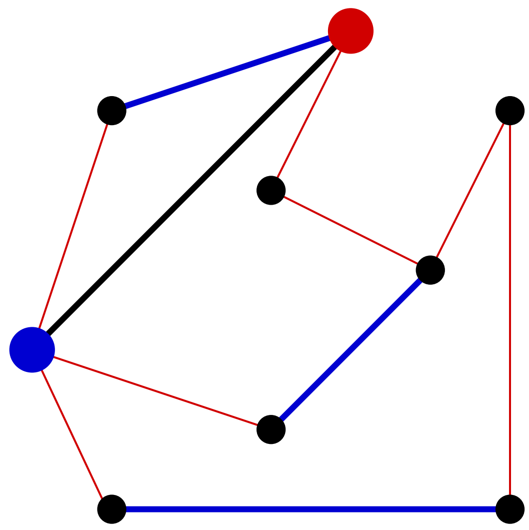
Application au protocole de requête

Requête par inondation



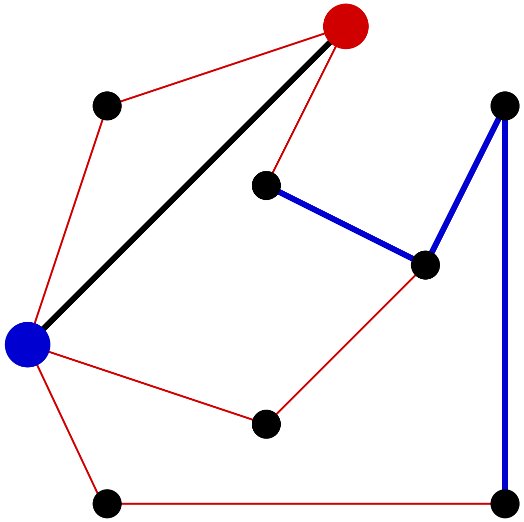
Application au protocole de requête

Requête par inondation



Application au protocole de requête

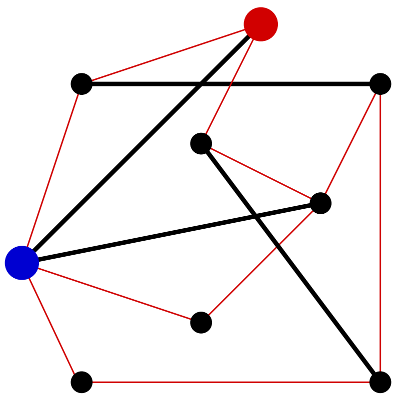
Requête par inondation



Application au protocole de requête

[Sripanidkulchai et al 2003, . . .]

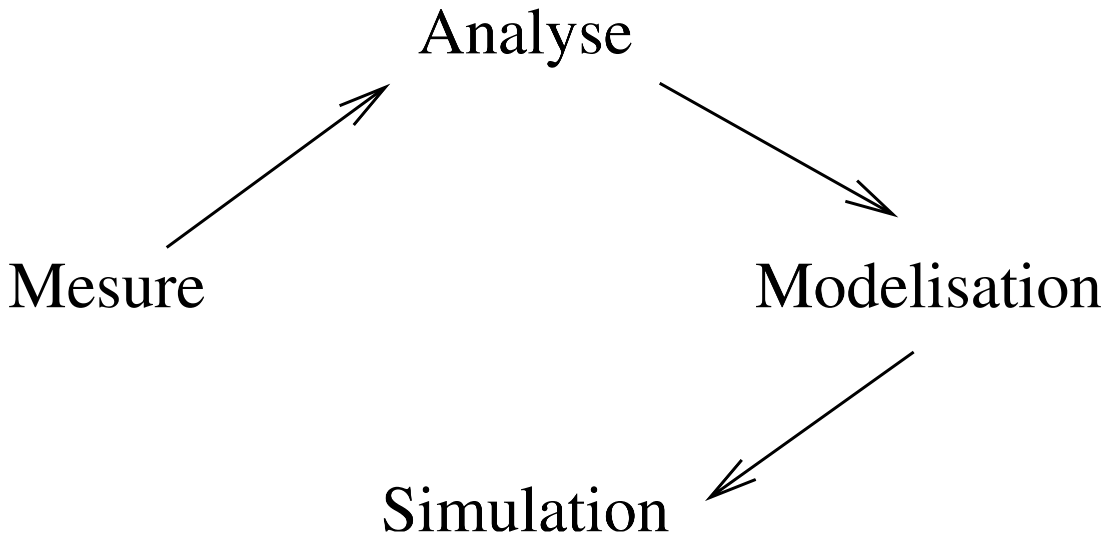
Utiliser la proximité sémantique



- Demander d'abord aux voisins sémantiques, puis
- Requête par inondation

Conclusion

Des problématiques communes à tous les GRI



Des problématiques communes à tous les GRI