

- RÉPUTATION
- INNOVATIONS
- PRÉDICTION DE LIENS

# Un exemple

- Ebay : pourcentage d'évaluation positive.

## Profil d'évaluation



**Iagroxa** ( 139 ★ )

Evaluations positives (12 derniers mois) : 99,1%

[Comment le pourcentage d'évaluations positives est-il calculé ?]

Membre depuis : 14-avr.-05 en France métropolitaine

### Mes évaluations récentes (12 derniers mois) ?

|          | 1 mois | 6 mois | 12 mois |
|----------|--------|--------|---------|
| Positive | 3      | 40     | 123     |
| Neutre   | 0      | 0      | 1       |
| Négative | 0      | 0      | 1       |

### Evaluations détaillées du vendeur (12 derniers mois) ?

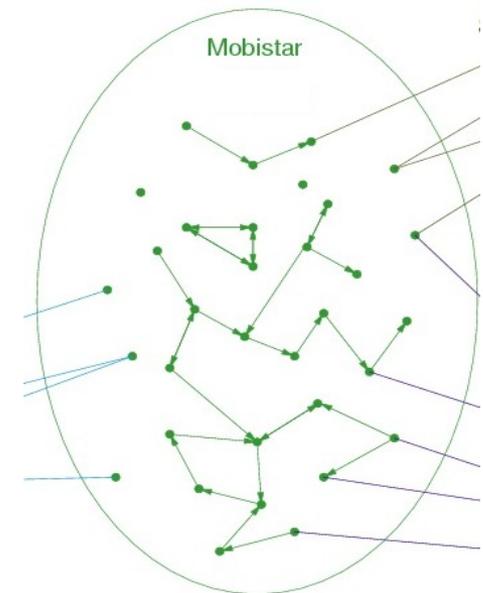
| Critères                           | Note moyenne | Nombre d'évaluations |
|------------------------------------|--------------|----------------------|
| Objet conforme à la description    | ★★★★★        | 77                   |
| Communication                      | ★★★★★        | 77                   |
| Délai de livraison                 | ★★★★★        | 77                   |
| Frais d'expédition et de livraison | ★★★★★        | 77                   |

# Pourquoi faire ?

- Détecter des participants « malhonnêtes » dans des systèmes d'enchères
  - Ebay
- Supprimer des spammers dans des systèmes de notation
  - Movielens
- Donner une réputation à des utilisateurs
  - Advogato
- Évaluer la confiance des sommets dans un système P2P
  - Kazaa
- Prédire quels sont les bons sommets dans un contexte donné :
  - Classification des pages web par pertinence
  - Classer des utilisateurs ebay par confiance
  - Noter des utilisateurs de téléphonie mobile par leur influence sur leurs contacts
- On a peu de chance qu'un techniques marche partout !

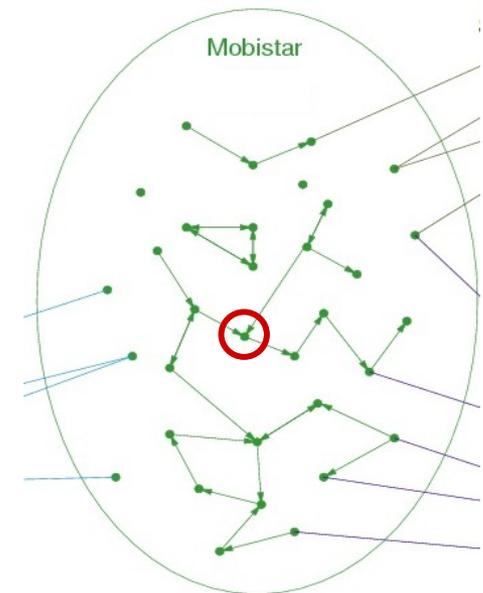
# Local vs. Global

- Mesure locale :
  - ▣ Ne dépend que du sommet et de ses voisins proches



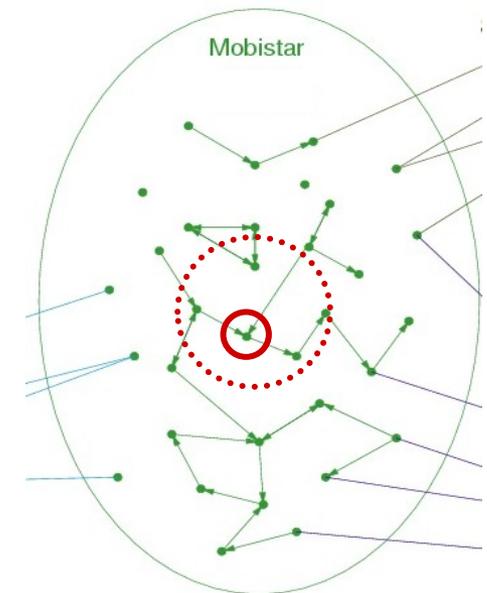
# Local vs. Global

- Mesure locale :
  - ▣ Ne dépend que du sommet et de ses voisins proches



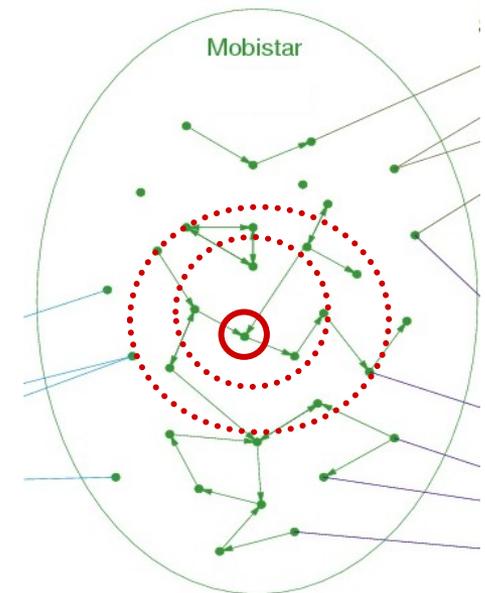
# Local vs. Global

- Mesure locale :
  - ▣ Ne dépend que du sommet et de ses voisins proches



# Local vs. Global

- **Mesure locale :**
  - ▣ Ne dépend que du sommet et de ses voisins proches
  - ▣ Défini un sous-graphe du graphe original



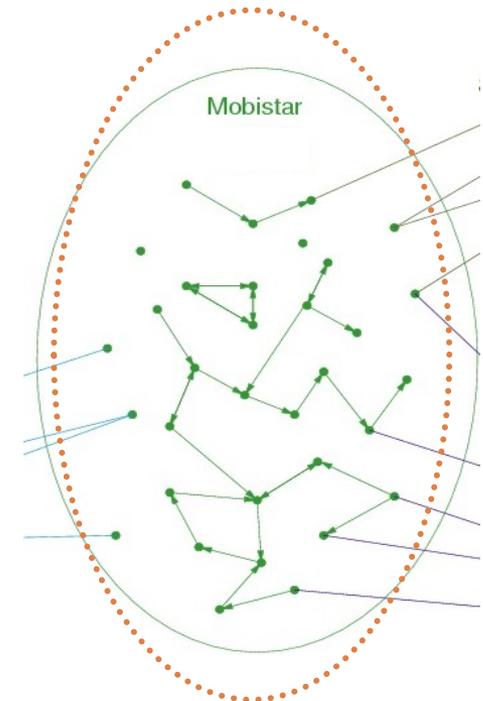
# Local vs. Global

## □ Mesure locale :

- Ne dépend que du sommet et de ses voisins proches
- Défini un sous-graphe du graphe original

## □ Mesure globale :

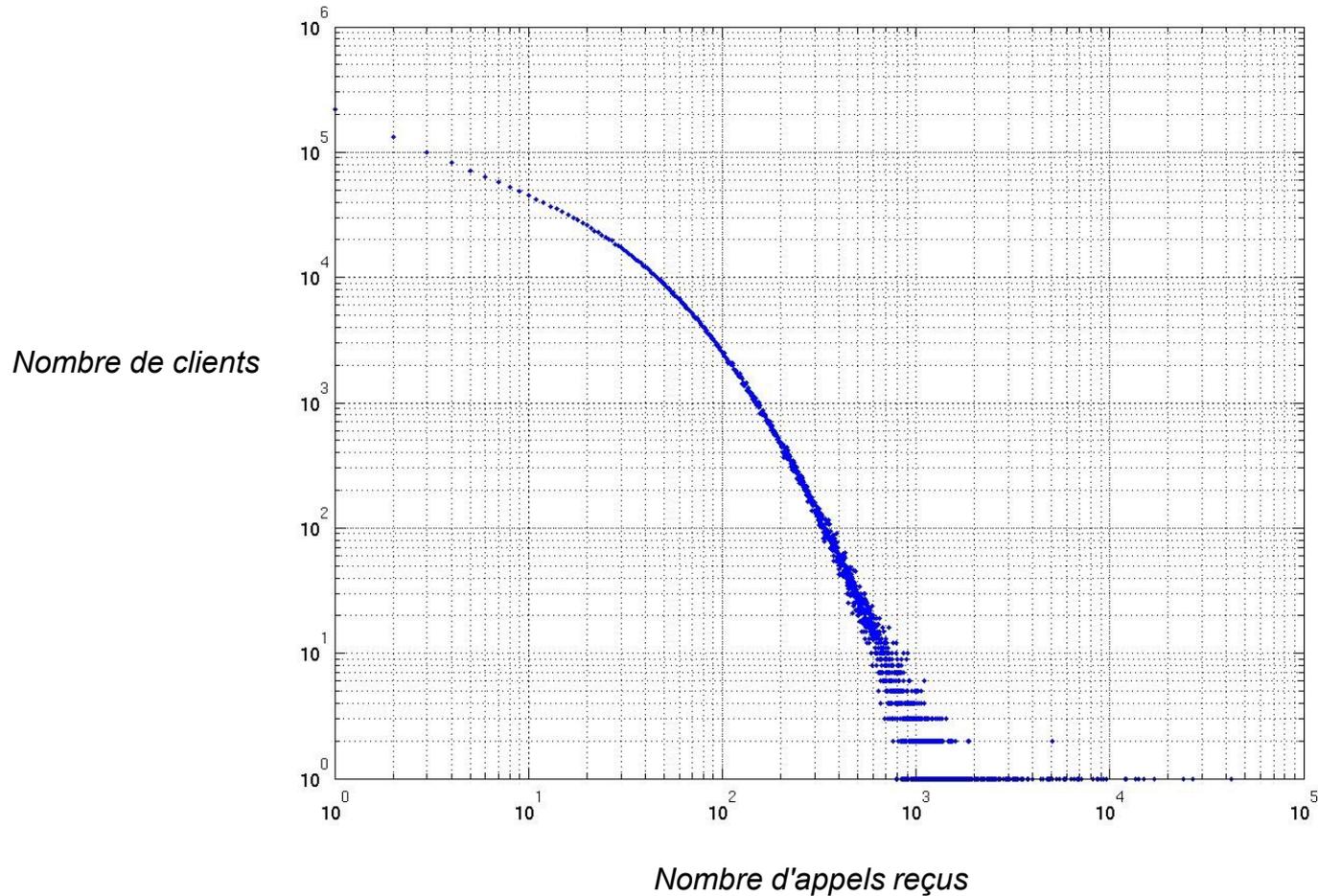
- Dépend de tout le graphe
- Toute modification a un impact



# Local vs. Global

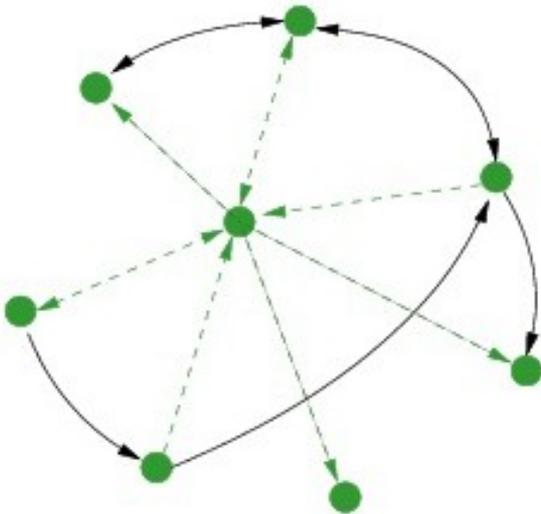
- **Mesure locale :**
  - Calcul plus simple en général
  - Interprétation plus aisée
  - Vision restreinte
  
- **Mesure globale :**
  - Vision globale
  - Interprétation complexe – boîte noire

# Propriété locale : degrés



# Propriété locale : degré social

- Combien de mes paires d'amis s'appellent ?

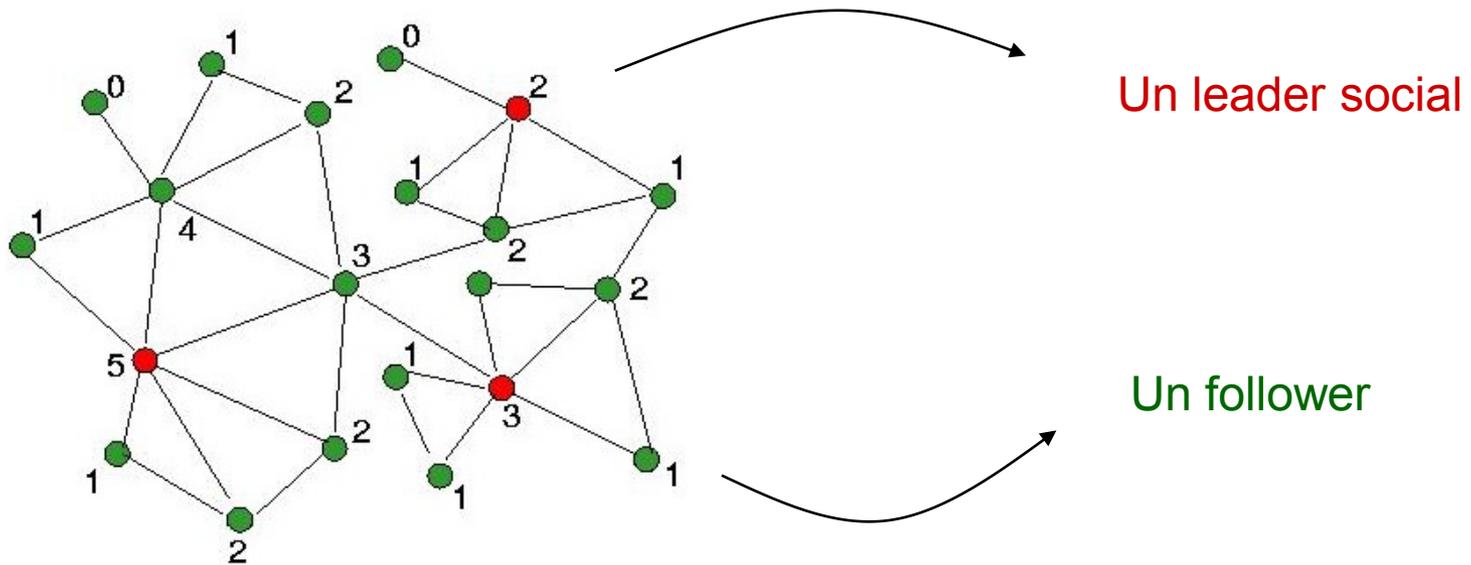


7 friends and 5 connections between them

| Connections | % customers |
|-------------|-------------|
| 0           | 41          |
| 1-5         | 39          |
| 6-10        | 10          |
| 10+         | 10          |

# Propriété locale : leaders sociaux

- Maximaux locaux du degré social :
  - ▣ Que faire des zones sans leaders ?

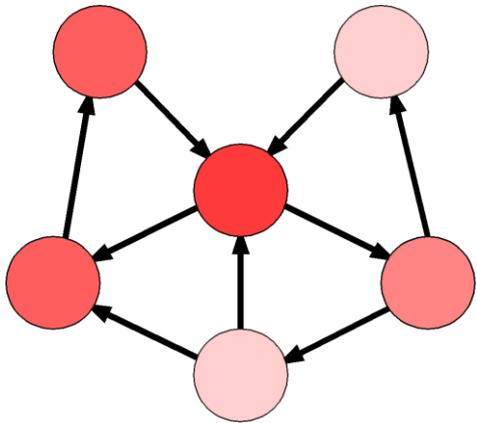


# Propriété locale d'ordre supérieur

- Est-ce que les (amis de) n mes amis sont amis ?
  - ▣ Permet d'étendre les définitions à des zones plus grandes
  - ▣ On peut être un leader dans un très petit domaine mais pas dans un plus grand

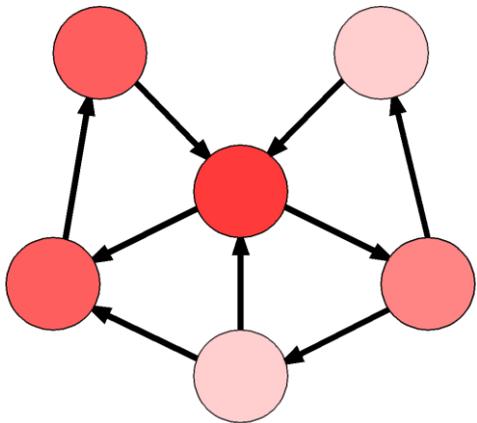
# Propriété globale : Markov

- Marche aléatoire de mobile en mobile :
  - Quel mobile sera le prochain appelé ?
  - Certains mobiles seront-ils plus souvent appelés ?

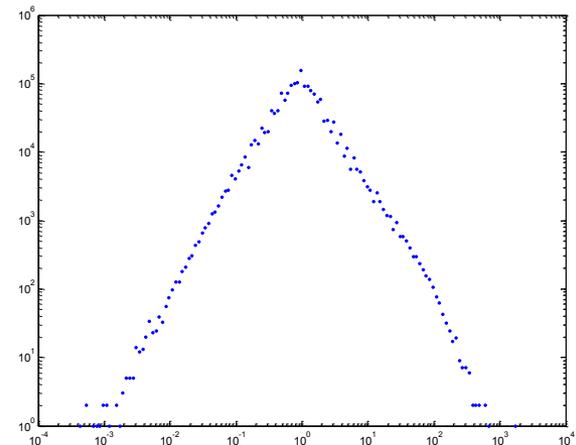


# Propriété globale : Markov

- Marche aléatoire de mobile en mobile :
  - ▣ Quel mobile sera le prochain appelé ?
  - ▣ Certains mobiles seront-ils plus souvent appelés ?
  - ▣ Ne fonctionne que sur les graphes orientés



*Nombre de clients*

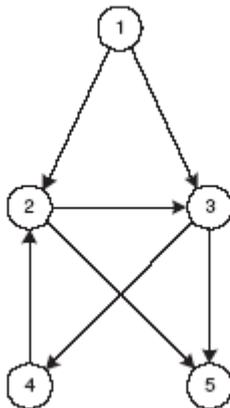
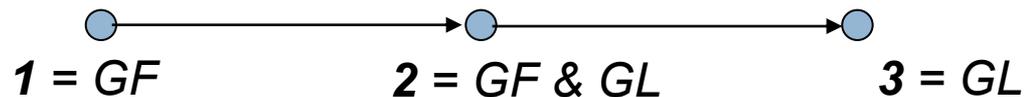


*Ratio appels emis/reçus*

*100.000 clients on un ratio de 1 et  
2/3 ont une balance inférieure à 2*

# Propriété globale : similarité

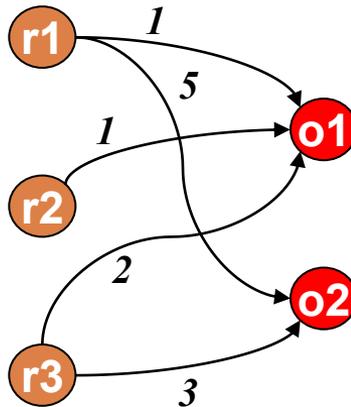
- Chaque sommet a un score pour chacune des trois classes 1 (GF), 2 (GF et GL) et 3 (GL) :
  - ▣ Un bon leader est pointé par plusieurs bons followers
  - ▣ Calcul de point fixe



|        |        |        |
|--------|--------|--------|
| 0.4433 | 0.1043 | 0      |
| 0.2801 | 0.3956 | 0.0858 |
| 0.0858 | 0.3956 | 0.2801 |
| 0.2216 | 0.0489 | 0.2216 |
| 0      | 0.1043 | 0.4433 |

# Propriété globale : réputation

- On considère  $n$  utilisateurs et  $m$  objets
  - ▣ Notes de chaque utilisateur pour chaque objet
  - ▣ La confiance accordée aux utilisateurs dépend de la normalité des notes données
  - ▣ La qualité d'un objet dépend des notes modifiées



# Remarques finale

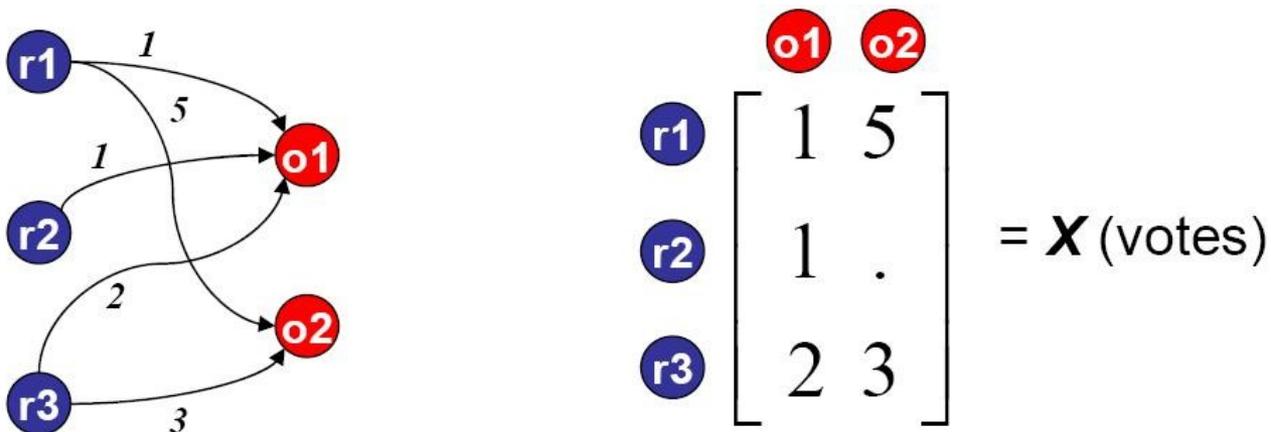
- Propriétés locales :
  - ▣ Degré : valeurs extremes
  - ▣ Social Degrees, leaders : 40% non impliqués
  - ▣ Ordre supérieur: interprétation, quel ordre ?
- Propriétés globales :
  - ▣ Markov, similarité : ne fonctionne que sur les graphes orienté
  - ▣ Réputation : notation de l'importance des liens
- Les propriétés ne donnent pas les mêmes indications
  - ▣ Une seule mesure ne permet pas de conclure

# RÉPUTATION

Jean-Loup Guillaume – travaux de C. Kerchove

# Objets / utilisateurs

- On considère un graphe biparti :
  - ▣ n objets (à noter).
  - ▣ m utilisateurs (donnant des notes).
  - ▣ Des notes (liens pondérés).
- Quelle est la réputation des objets et personnes ?



# Une solution

- Hypothèse :
  - ▣ Un utilisateur déviant est considéré comme moins fiable.
  
- Trois évaluateurs ( $m=3$ ), deux objets ( $n=2$ ) :
  - ▣ Évaluateur 1 : notes 3.3 et 4.2
  - ▣ Évaluateur 2 : notes 3.4 et 4.5
  - ▣ Évaluateur 3 : notes 4.9 et 2.8
  - ▣ L'évaluateur 3 est différent des 2 autres (moins fiable ?).
  
- Initialement, on fait confiance aux évaluateurs :
  - ▣  $c(i,0)=1$  : confiance de l'évaluateur  $i$  au temps 0.

# Calcul (par l'exemple)

## □ Ensuite on itère :

□  $s(i,1)$  = moyenne pondérée (par la confiance) des notes

■  $s(1,1) = (3.3*1 + 3.4*1 + 4.9*1) / (1+1+1) = 3.87$  et  
■  $s(2,1) = 3.83$

□  $c(i,1) = 1 - k/n * (\text{distance}(\text{notes}, s))^2$

■ Avec  $k$  (ici  $1/5$ ) à fixer pour  $n$  objets.

■ Distance = distance euclidienne entre les notes mises et les notes corrigées.

■  $c(1,1) = 1 - 0.1 * ((3.3 - 3.87)^2 + (4.2 - 3.83)^2) = 0.95$

■  $c(2,1) = 0.93$

■  $c(3,1) = 0.79$

# Calcul (par l'exemple)

- Deuxième étape :

- $s(2,1) = (3.3 \cdot .95 + 3.4 \cdot .93 + 4.9 \cdot .79) / (.95 + .93 + .79) = 3.81$

- $s(2,2) = (4.2 \cdot .95 + 4.5 \cdot .93 + 2.8 \cdot .79) / (.95 + .93 + .79) = 3.89$

- $c = \dots$

- Convergence :

- $s(1) = 3.79, s(2) = 3.91$  et  $c(1) = .97, c(2) = .95, c(3) = .75$

- Donne :

- Une note de confiance à chaque évaluateur.

- Un score corrigé à chaque objet.

# Graphe complet

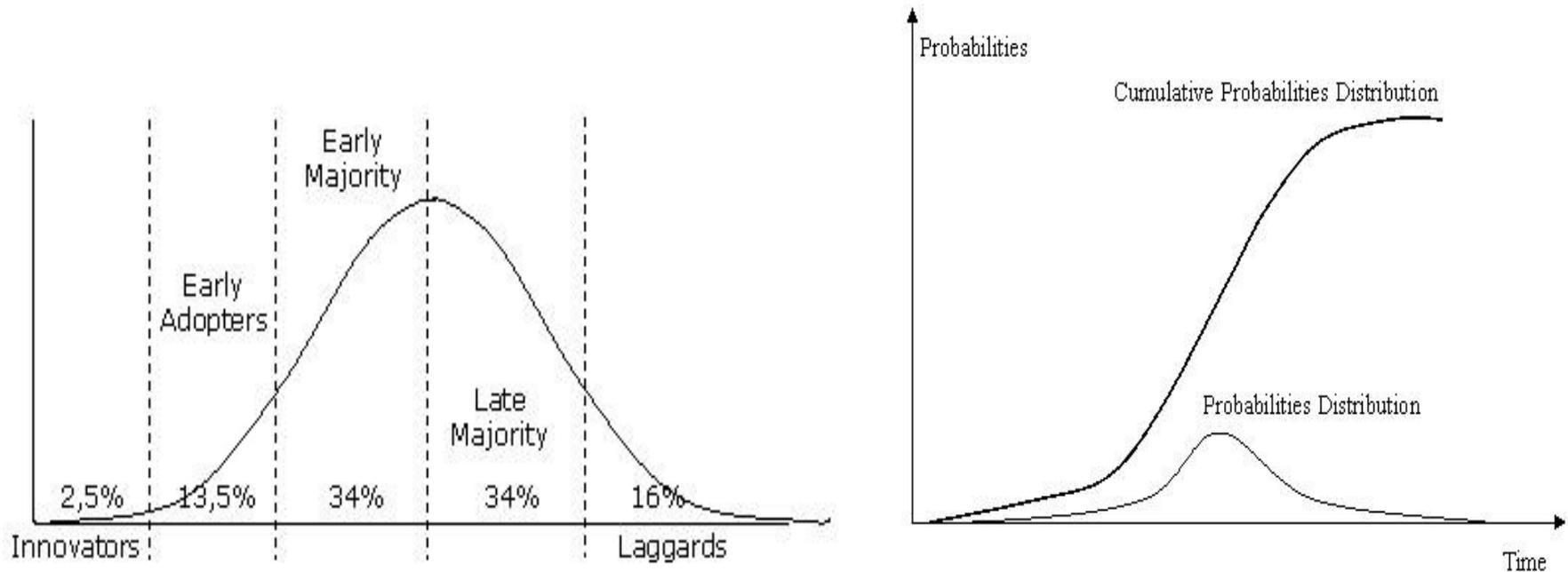
- On a supposé que tous les évaluateurs mettent une note à tous les objets, en pratique c'est faux.
- Idem mais en ne considérant que les notes mises :
  - ▣ Moyenne pondérée uniquement sur les notes mises.
  - ▣ Distance uniquement sur les notes.

# DIFFUSION D'INNOVATIONS

Jean-Loup Guillaume

# Diffusion d'innovations

- Étudier finement les liens entre diffusion et réseau
  - ▣ Pourquoi certaines innovations se diffusent mieux que d'autres ?

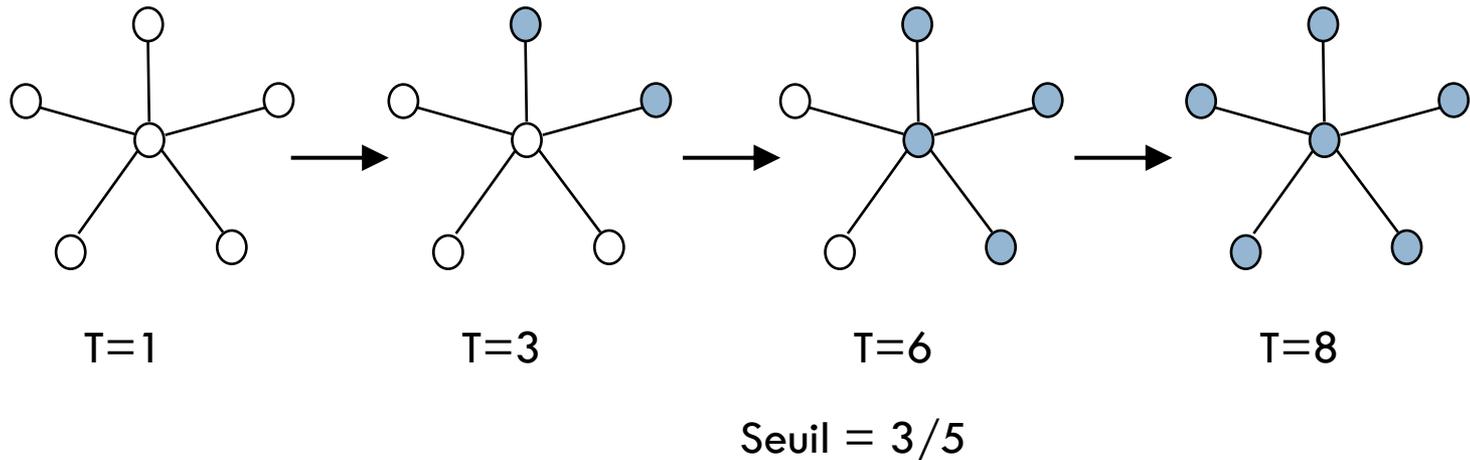


# Origines possibles de l'adoption

- Influences locales ou internes au réseau social : (discussions entre collègues)
  - ▣ conseils de proches, de personnes faisant autorité...
- Influences externes : (contacts avec des détaillants, lecture de journaux, ...)
  - ▣ Publicité.
  - ▣ Promotions ponctuelles.
  - ▣ Effets de mode, banalisation des technologies et des usages...
- Autres critères : (âge, propension à prescrire, "foi en la science")
  - ▣ Variables socio-démographiques.
  - ▣ Goût pour les nouvelles technologies et niveau de compétences dans leur utilisation...

# La notion de seuil

- Seuil de perméabilité d'un individu à une innovation
- Proportion de "connaissances" ayant adopté au moment de l'adoption.



- Seuil nul : influence externe uniquement (sauf adoption simultanée).
- Seuil non nul : les deux influences sont à l'œuvre.

# Influences externes

- Publicité, promotions ponctuelles, autres facteurs globaux
  - ▣ Mise en évidence : individus qui adoptent avec un seuil nul :
  - ▣ Nécessité de connaître les contacts entre individus.
  
- Influence externe et seuil non nul : comment distinguer les deux ?
  - ▣ Une approche possible : influence externe = facteur pouvant faire baisser le seuil d'un individu.
  
- Influence externe variable :
  - ▣ Constante (communication de l'entreprise).
  - ▣ Variable (promotions ponctuelles, campagnes ciblées sur un produit).
  - ▣ Plus le nombre d'individus ayant adopté est élevé, plus les gens vont être tentés d'adopter l'innovation.

# Influences locales

- Données nécessaires pour calculer les seuils :
  - ▣ Contacts de l'individu.
  - ▣ Utilisation des nouvelles technologies.
  
- Corrélations entre seuil et propriétés du réseau local :
  - ▣ Nombre de contacts de l'individu (degré) ?
  - ▣ Nombre de liens entre les contacts de l'individu (clustering) ?
  - ▣ Structure du réseau local (communautés) ?
  - ▣ ...
  
- Seuil = adoption rapide ou tardive **locale** :
  - ▣ Adoption globalement tardive mais localement rapide.
  - ▣ Adoption globalement rapide mais localement tardive.

# Pour aller plus loin

- Fréquence ou durée des contacts entre individus :
  - ▣ Distinguer les contacts « forts » des contacts « faibles ».
  - ▣ Mesurer précisément l'influence des contacts d'un individu dans son adoption.
  - ▣ Déterminer les individus qui influencent beaucoup ou peu leur entourage.
  
- Évolution des contacts au cours du temps :
  - ▣ Adoption grâce à un nouveau contact.
  - ▣ Densification du réseau local => diffusion plus rapide ?
  - ▣ ...

# Marketing viral

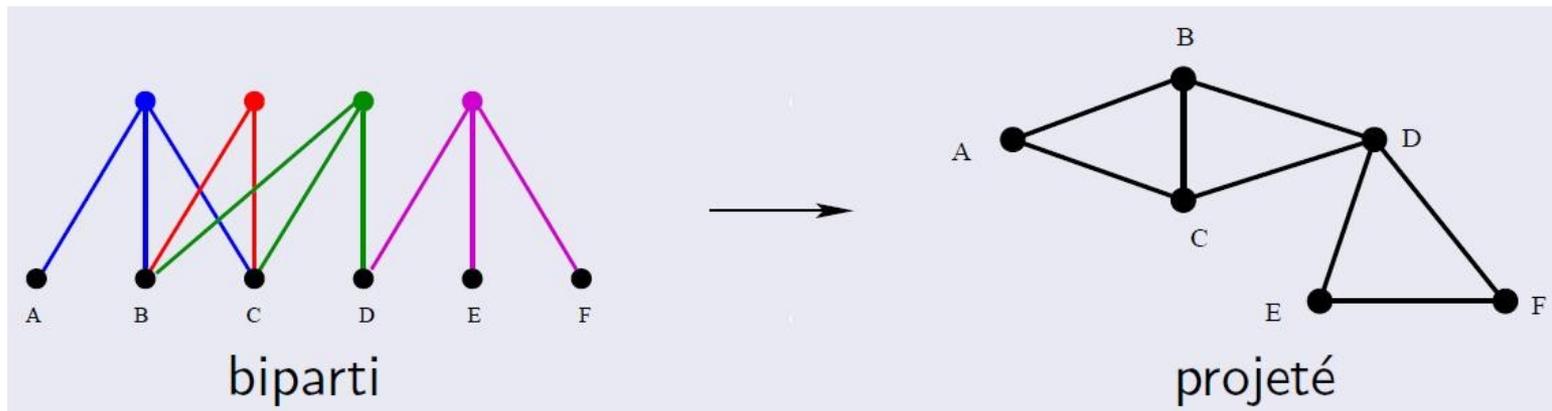
- Étudier la diffusion comme un acte de transmission et/ou de promotion de l'usage d'un produit ou service :
  - ▣ Ne plus se placer du côté de la personne qui adopte mais de celle qui fait adopter.
  
- À partir des informations sur l'adoption, retrouver l'origine de celles-ci.
  
- Influences locales : distinguer les individus qui :
  - ▣ Propagent beaucoup : leaders d'opinion.
  - ▣ Diffusent à des utilisateurs qui utilisent beaucoup ou peu.

# PRÉDICTION DE LIENS

Jean-Loup Guillaume – travaux de O. Allali

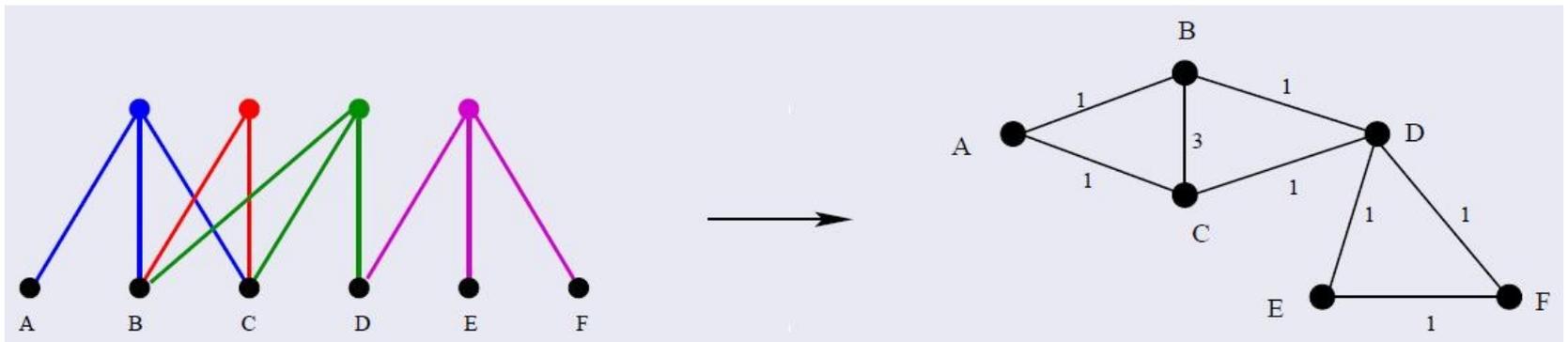
# Contexte

- Restriction aux graphes bipartis (collaboration) :
  - Acteurs/films
  - Chercheurs/publications
  - Utilisateurs/produits achetés
  - ...
- Relation entre graphe biparti et sa projection



# Projection valuée

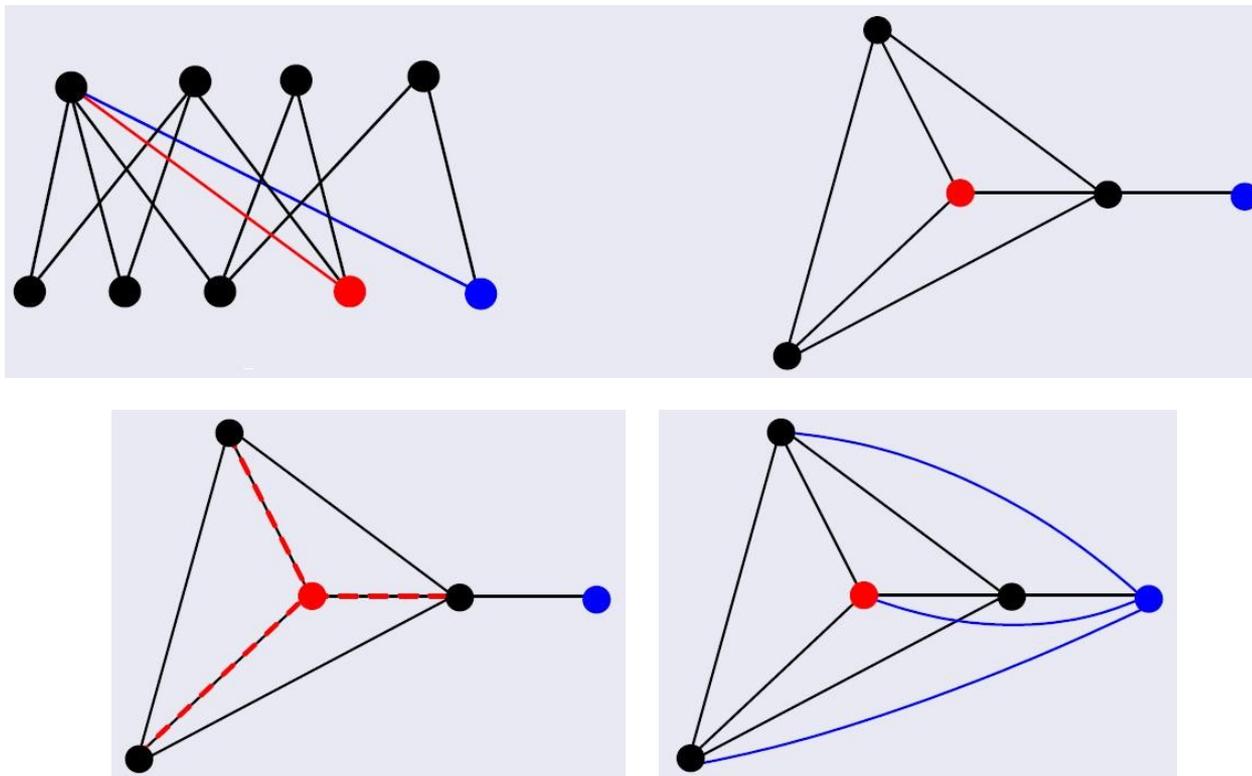
- Comment mettre des poids sur les liens :
  - Approche simple : nombre de créations du lien



- Fraction de voisins en commun dans le biparti (jaccard)
  - C-D : 4 voisins dont un seul en commun (vert) =  $1/4$
- Somme des votes des voisins dans le biparti (delta)
  - Chacun donne un vote de 1 partagé entre ses voisins
  - B-C :  $1/3$  des bleus, 100% des rouges et  $1/3$  des verts =  $5/3$

# Lien interne vs lien externe

- Interne : pas de création de lien dans le projeté
- Externe : création de lien dans le projeté

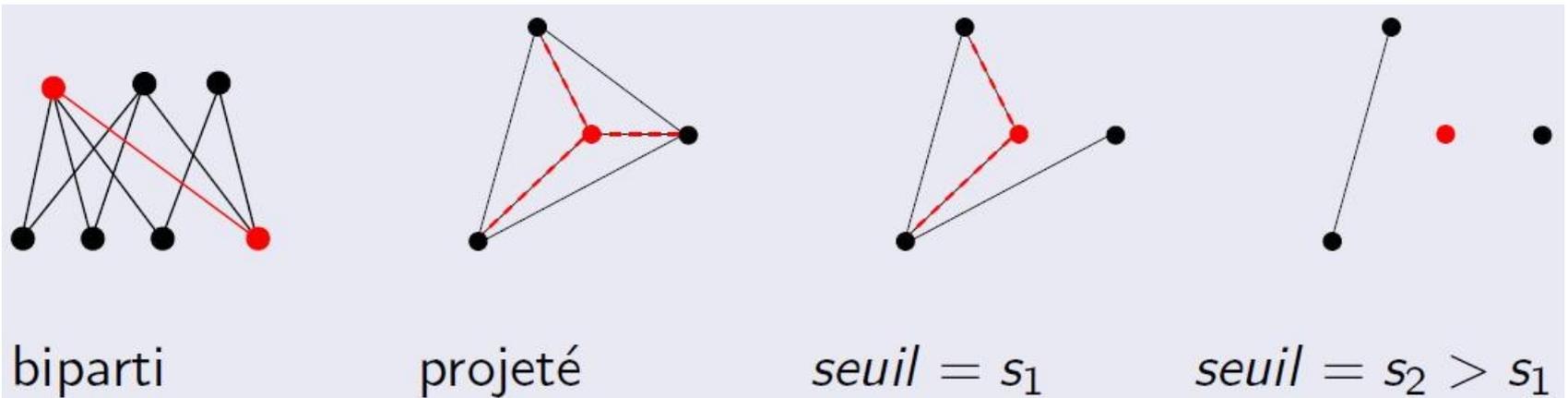


# Objectif

- Prédire les liens internes :
  - "Renforcement" de liens : intuitivement plus simple
  - On va se servir d'une période "d'apprentissage"
  - On valide en regardant la proportion de liens correctement/incorrectement prédits.

# Méthode utilisée

- Ensemble de liens  $L$  dans le projeté tels que
  - ▣ Poids  $>$  seuil (seuil variable)
  - ▣ Calculé sur la période d'apprentissage
- On prédit les liens du biparti qui induisent les liens de  $L$ 
  - ▣ Le lien rouge est prédit uniquement si on prend un seuil  $< s_2$
  - ▣ On commence qui renforcent avec un seuil élevé



# Résultats

- Deux critères (en fonction du # de liens prédits) :
  - Proportion de liens correctement prédits (précision)
    - Exactitude de la méthode
  - Proportion de liens prédits qui apparaissent (rappel)
    - Exhaustivité de la méthode

