



COLLÈGE
DE FRANCE
— 1530 —

Modélisation et réseaux sociaux

Claire Mathieu



Aujourd'hui

La Toile (World Wide Web)

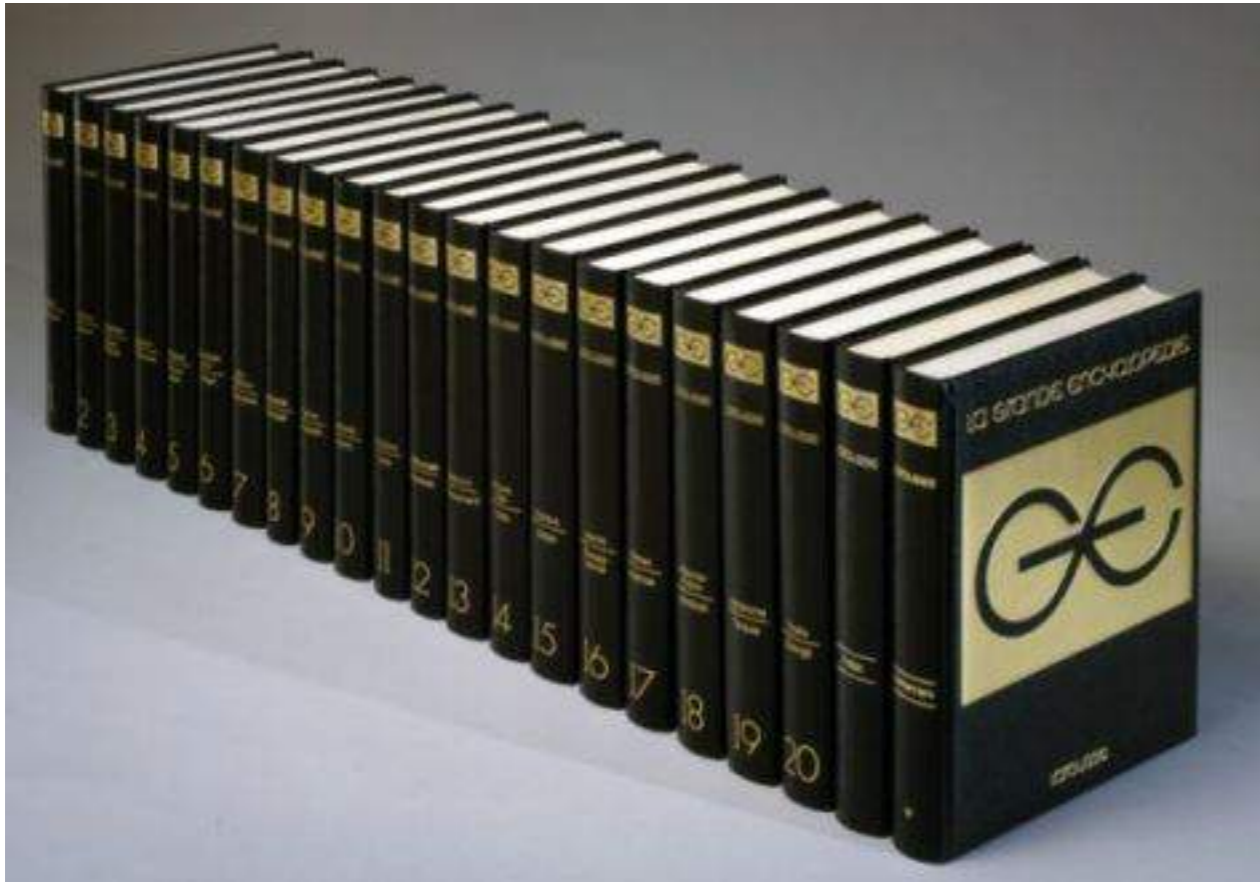
Structure de la Toile : les degrés

Structure de réseaux sociaux : le phénomène de petit monde

Les plafonds de verre dans les réseaux sociaux académiques

La Toile (le Web)

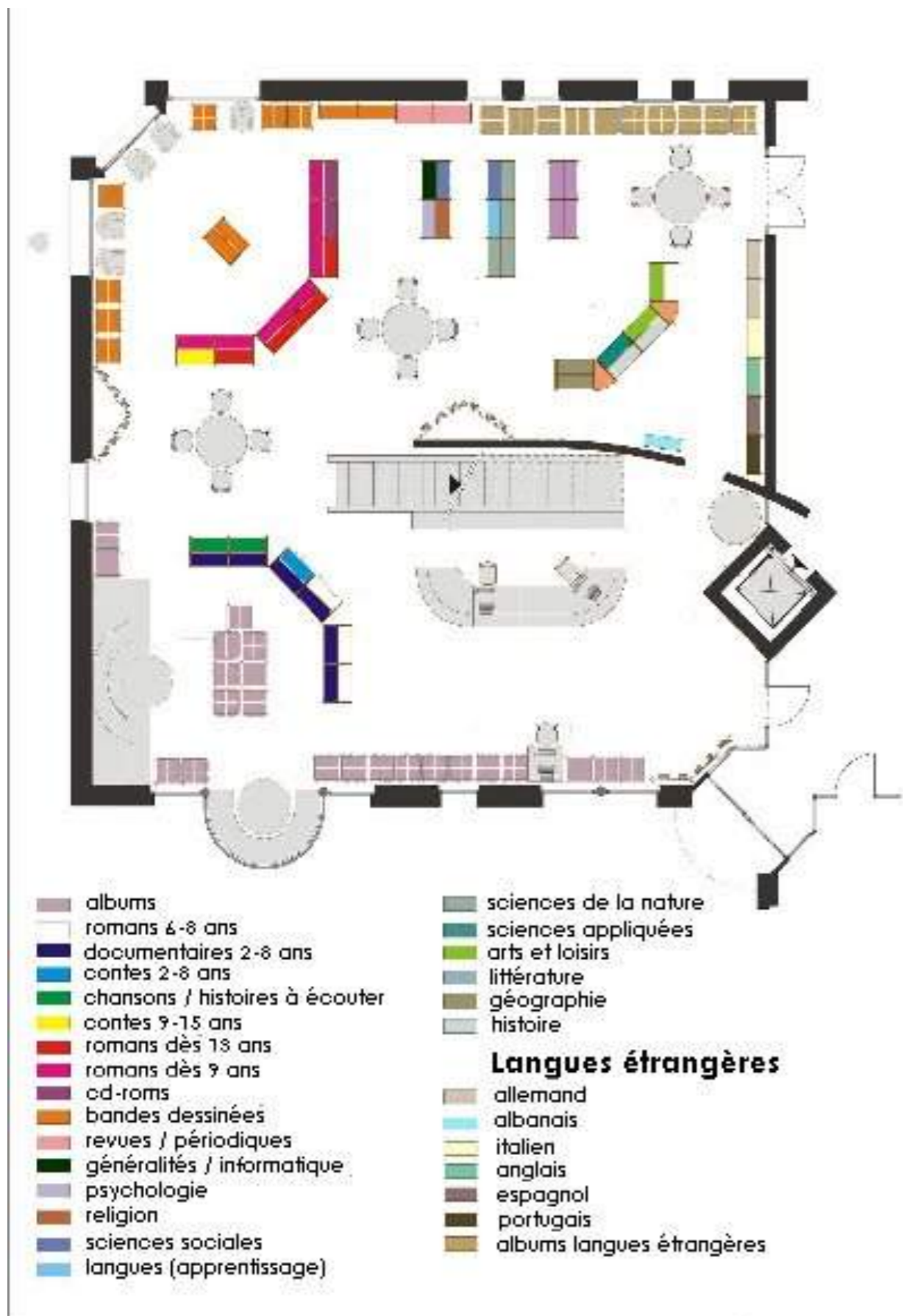
Comment trouver une information ?



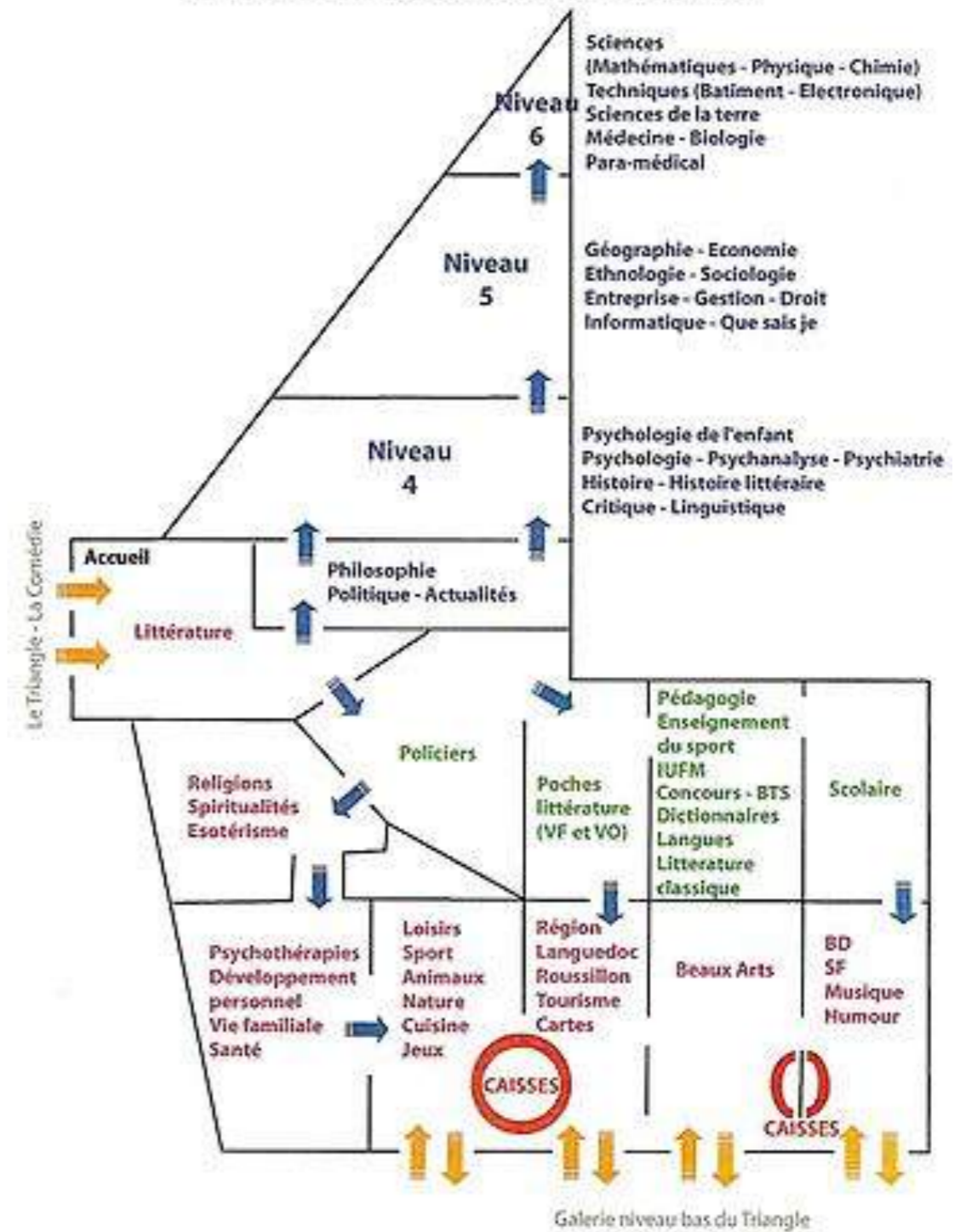
ne pay
♦ adv. Litt. Paris
3. FRANC, FRANQUE adj. Qui appartient
Francs.
FRANÇAIS, E adj. et n. 1. Habitant de la
France. 2. Qui appartient, qui est relatif à la
France, à ses habitants. Nationalité française.
3. Propre à la langue française. Grammaire
française. ♦ n.m. Langue romane parlée princi-
palement en France, au Canada, en Belgique,
en Suisse et en Afrique. ♦ En bon français : en
termes clairs et précis.
FRANC-ALLEU [frākalø] n.m. (pl. francs-alleux).
rÉOD. Alleu affranchi de toute servitude.
FRANC-BORD n.m. (pl. francs-bords). 1. MA
la mesurée au milieu d'un nav
et la pa

Ordre alphabétique

Comment trouver une information ?

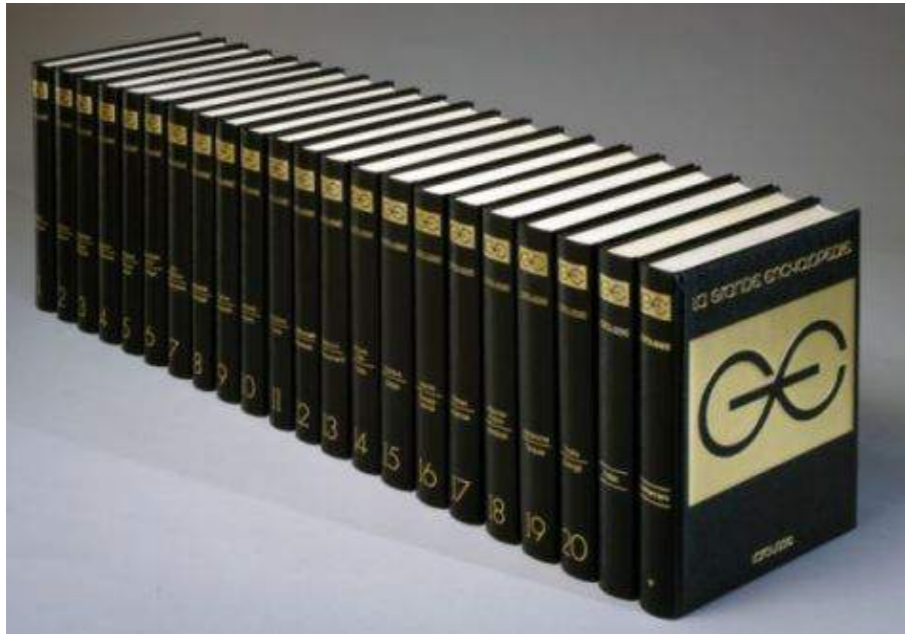


PLAN SOMMAIRE DE LA LIBRAIRIE



Par genre

Découvrir une information dont on n'aurait pas eu idée



Au hasard : lire l'encyclopédie

Homophily and the Glass Ceiling Effect in Social Networks

Chen Avin¹, Barbara Keller¹, Zvi Lotker¹
 Ben-Gurion University of the Negev, Israel
 Claire Mathieu², David Peleg¹, Yvonne-Arène Pignolet¹
 CNRS, Ecole Normale Supérieure, France
 WIZMANN INSTITUTE, ISRAEL
 IBM CORPORATION, SWITZERLAND

ABSTRACT
 The glass ceiling effect has been defined in a recent IJF and...
INTRODUCTION
 In many social networks, there are existing inequalities...

Categories and Subject Descriptors
 G.2.2 [Mathematics of Computing]: DISCRETE MATHEMATICS
 H.5.1 [Information Systems and Information Management]: SOCIAL AND BEHAVIORAL SCIENCES

Keywords
 social networks, homophily, glass ceiling

1. INTRODUCTION
 In many social networks, there are existing inequalities...
 The glass ceiling effect is a barrier that keeps minorities and women from attaining the upper range of the corporate ladder...
 The existence of the glass ceiling effect is well documented...
 The paper is organized as follows. Section 2 discusses the...
 Section 3 presents our model and its analysis. Section 4...
 Section 5 discusses the implications of our model and...
 Section 6 concludes.

Le plafond de verre expliqué par les mathématiques

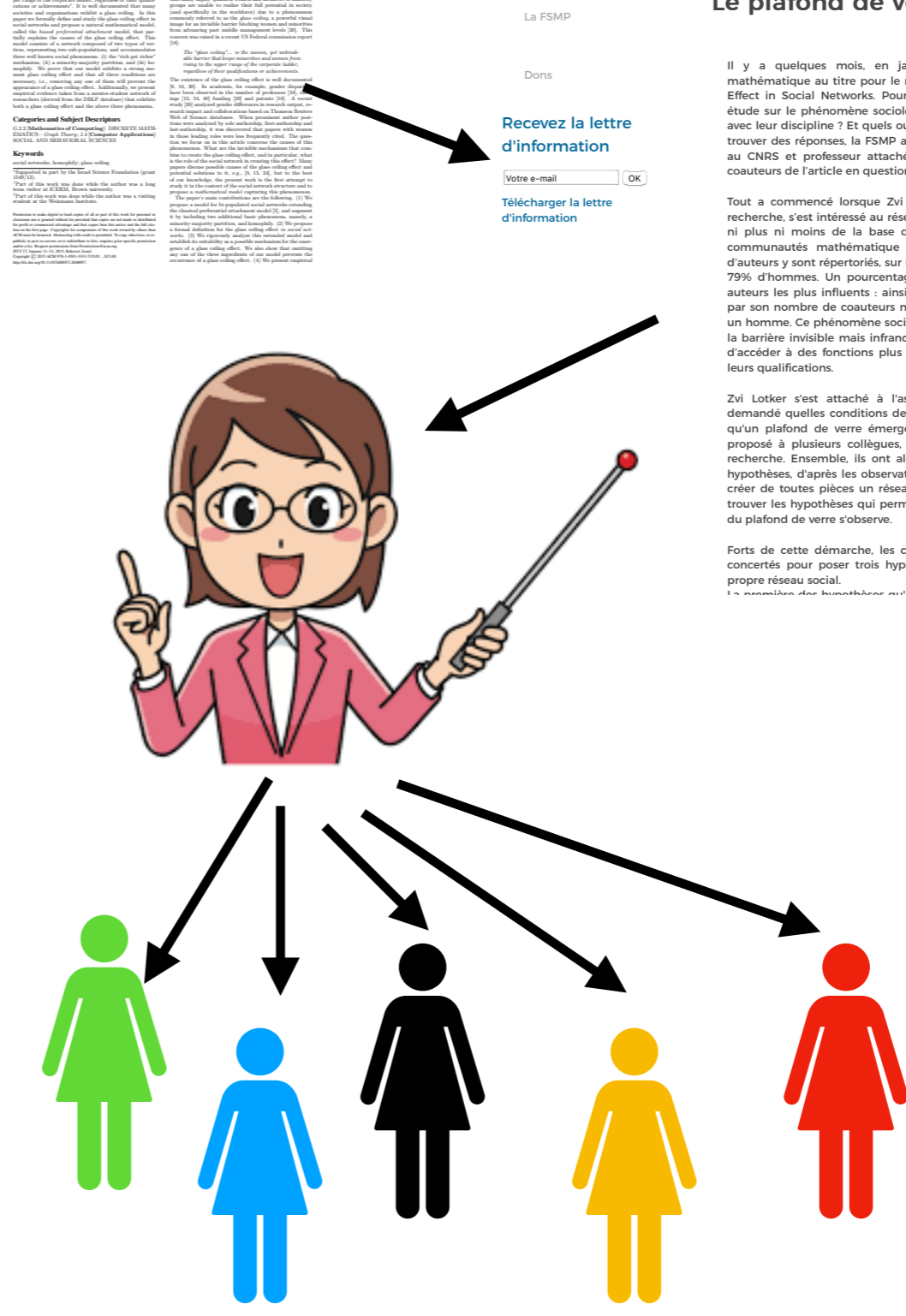
Le plafond de verre expliqué

Il y a quelques mois, en janvier 2015, paraissait u mathématique au titre pour le moins intrigant : Homophily and the Glass Ceiling Effect in Social Networks. Pourquoi donc des mathématiciens étudient-ils le phénomène sociologique du plafond de verre avec leur discipline ? Et quels outils mathématiques se cachent derrière ces réponses, la FSMP a rencontré Claire Mathieu, au CNRS et professeur attaché à l'ENS - Département coauteurs de l'article en question.

Tout a commencé lorsque Zvi Lotker, l'un des mathématiciens de la recherche, s'est intéressé au réseau DBLP. Un réseau bien plus ni plus ni moins de la base de données recensant toutes les communautés mathématique et informatique. Plusieurs auteurs y sont répertoriés, sur une période couvrant plus de 79% d'hommes. Un pourcentage qui augmente encore plus si l'on considère les auteurs les plus influents : ainsi, plus un chercheur est reconnu, plus il y a de femmes parmi ses coauteurs. Ce phénomène sociologique est couramment appelé la barrière invisible mais infranchissable qui empêche les femmes d'accéder à des fonctions plus élevées dans le monde de la recherche.

Zvi Lotker s'est attaché à l'aspect mathématique de la question : demandant quelles conditions devaient être réunies pour qu'un plafond de verre émerge. Dans le but de déterminer les conditions de ce phénomène, il a proposé à plusieurs collègues, parmi lesquels Claire Mathieu, de tester ses hypothèses. Ensemble, ils ont alors choisi l'approche suivante : créer de toutes pièces un réseau social à partir de ces hypothèses, et voir si elles permettent de créer un réseau où le plafond de verre s'observe.

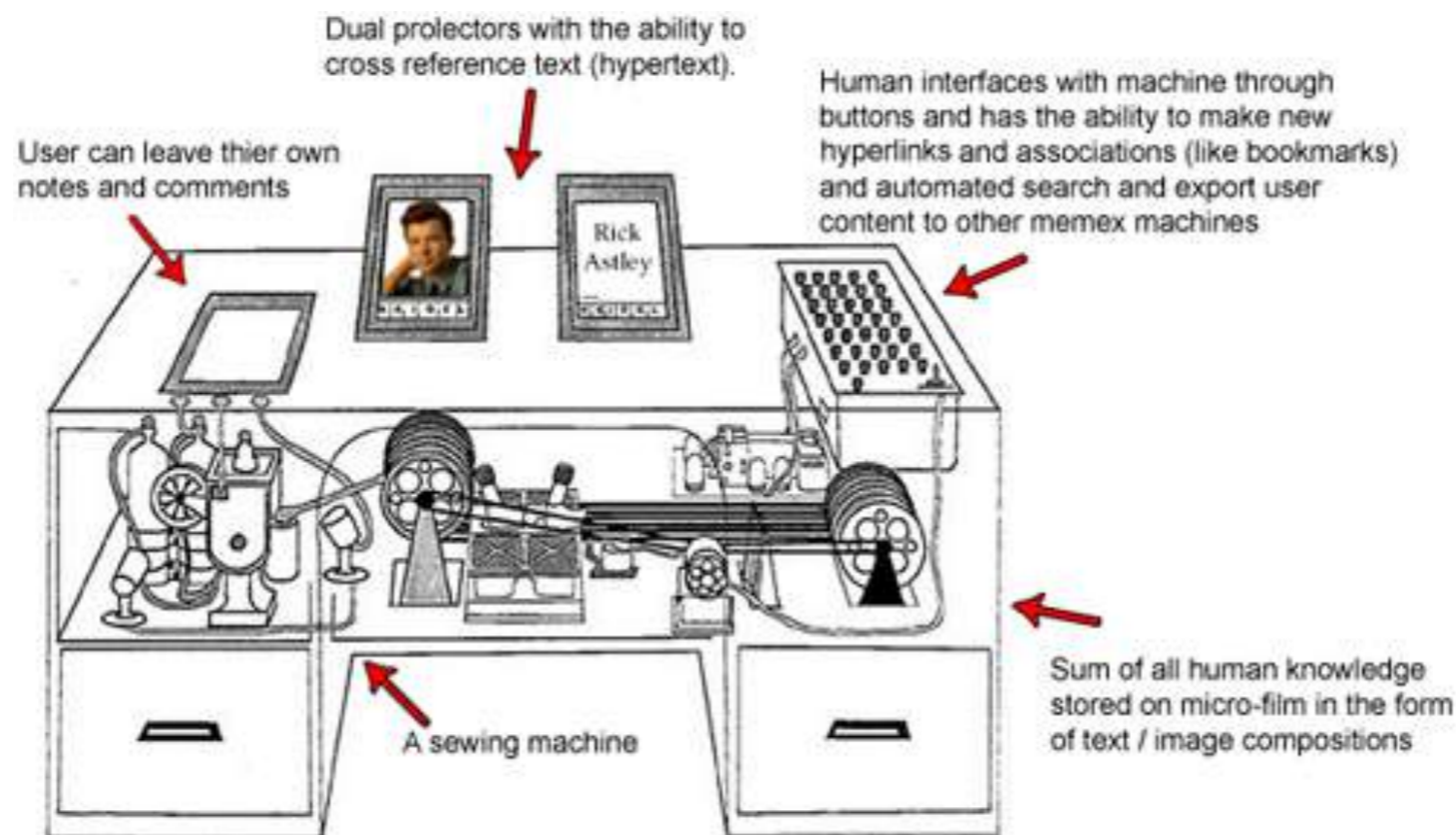
Fort de cette démarche, les chercheurs impliqués dans ce projet ont pu poser trois hypothèses de départ, sous la forme d'un réseau social. La première des hypothèses qu'ils ont ainsi définies est...



Diffusion hiérarchique de l'information

Un texte visionnaire

Consider a future device for individual use, which is a sort of mechanized private file and library. [...] A memex is a device in which an individual stores all his books, records, and communications, and which is mechanized so that it may be consulted with exceeding speed and flexibility. [...] It consists of a desk [...] On the top are [...] screens, on which material can be projected for convenient reading. There is a keyboard, and sets of buttons and levers. [...] In one end is the stored material. [...] Only a small part of the interior of the memex is devoted to storage, the rest to mechanism.



THE MEMEX order yours today!

The owner of the memex, let us say, is interested in the origin and properties of the bow and arrow. Specifically he is studying why the short Turkish bow was apparently superior to the English long bow in the skirmishes of the Crusades. He has dozens of possibly pertinent books and articles in his memex. First he runs through an encyclopedia, finds an interesting but sketchy article, leaves it projected. Next, in a history, he finds another pertinent item, and **ties the two together**. Thus he goes, **building a trail** of many items. Occasionally he inserts a comment of his own, either **linking it** into the main trail or joining it by a side trail to a particular item. When it becomes evident that the elastic properties of available materials had a great deal to do with the bow, he branches off on a side trail which takes him through textbooks on elasticity and tables of physical constants. He inserts a page of longhand analysis of his own. Thus he builds a trail of his interest through the maze of materials available to him.





And his trails do not fade. **Several years later**, his talk with a friend turns to the queer ways in which a people resist innovations, even of vital interest. He has an example, in the fact that the outraged Europeans still failed to adopt the Turkish bow. In fact he has a trail on it. A touch brings up the code book. **Tapping a few keys projects the head of the trail**. A lever runs through it at will, stopping at interesting items, going off on side excursions. It is an interesting trail, pertinent to the discussion.

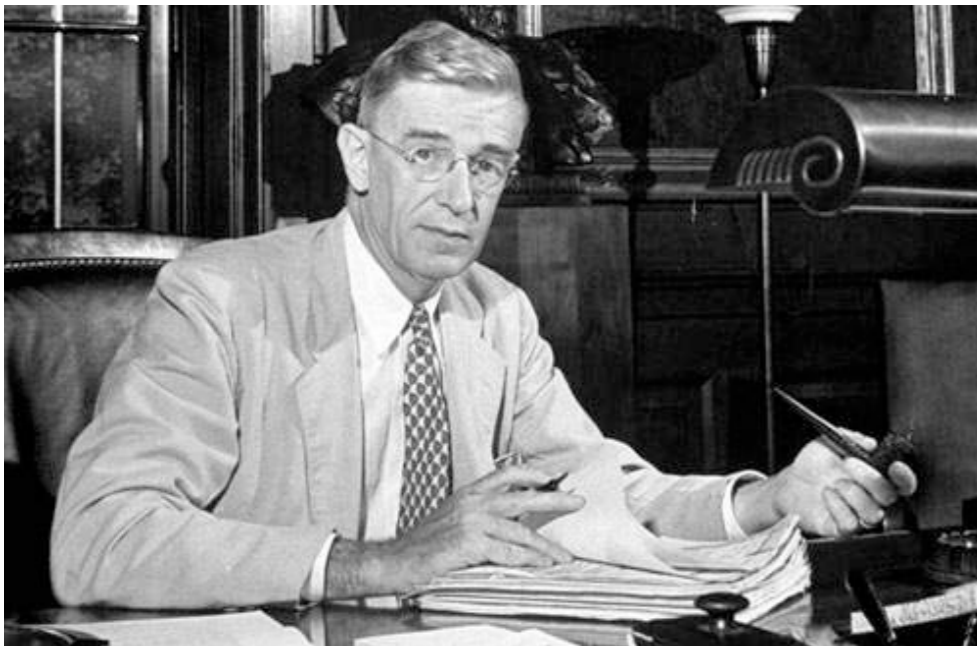


Wholly new forms of encyclopedias will appear, ready made with a mesh of associative trails running through them, ready to be dropped into the memex and there amplified. [...]

There is a new profession of trail blazers, those who find delight in the task of establishing useful trails through the enormous mass of the common record. The inheritance from the master becomes, not only his additions to the world's record, but for his disciples **the entire scaffolding by which they were erected.**

Wholly new forms of encyclopedias will appear, ready made with a mesh of associative trails running through them, ready to be dropped into the memex and there amplified. [...]

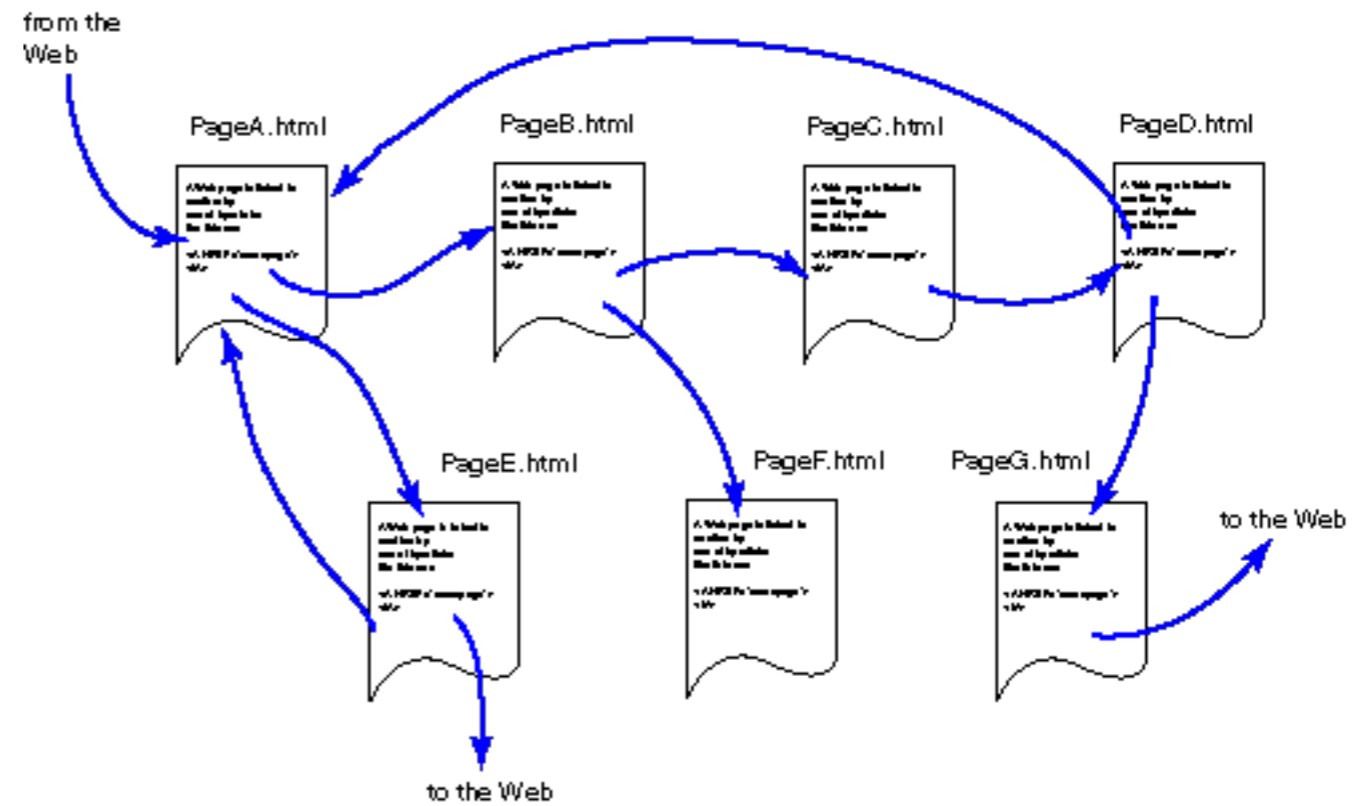
There is a new profession of trail blazers, those who find delight in the task of establishing useful trails through the enormous mass of the common record. The inheritance from the master becomes, not only his additions to the world's record, but for his disciples **the entire scaffolding by which they were erected.**



**“As We May Think”
Dr Vannevar Bush
The Atlantic
Juillet 1945**



Découvrir une information dont on n'aurait pas eu idée



Hyperliens

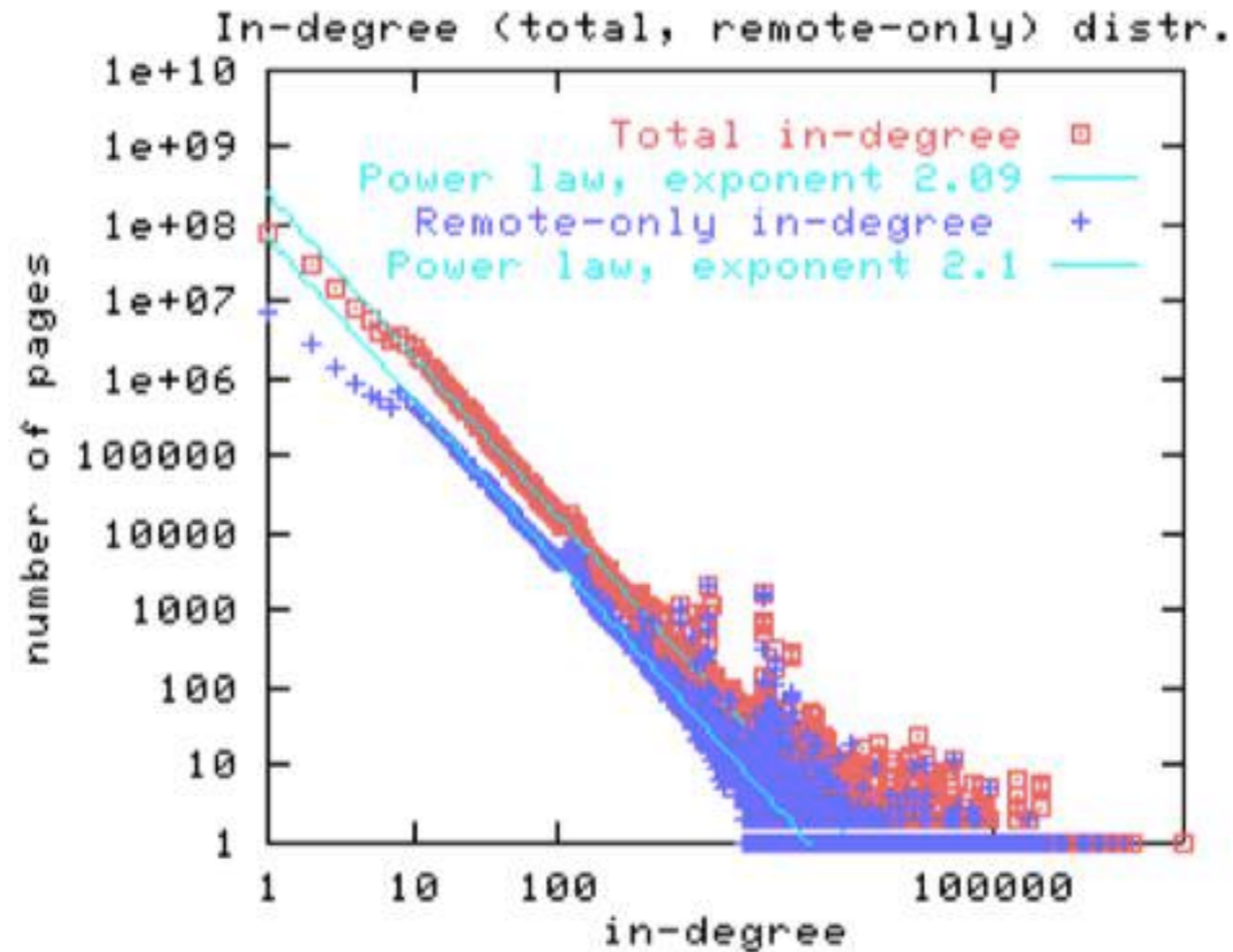
Structure non hiérarchique de diffusion de l'information
Tous contribuent à la créer

La Toile, source de sagesse ?

The **applications of science** have built man a well-supplied house, and are teaching him to live healthily therein. They have enabled him to throw masses of people against one another with cruel weapons. They may yet allow him truly to encompass the great record and **to grow in the wisdom of race experience**. He may perish in conflict before he learns to wield that record for his true good. Yet, in the application of science to the needs and desires of man, it would seem to be a singularly unfortunate stage at which to terminate the process, or to lose hope as to the outcome.

Structure de la Toile : les degrés des noeuds

Distribution du degré entrant



Loi de puissance

Nombre de liens vers une page

Qualitativement:

Beaucoup de pages ont peu de liens

Peu de pages ont beaucoup de liens

Fraction de pages vers lesquelles pointent k liens = a/k^c

$$\log(\text{fraction}) = \log a - c \log k$$

Comment expliquer cette distribution ?

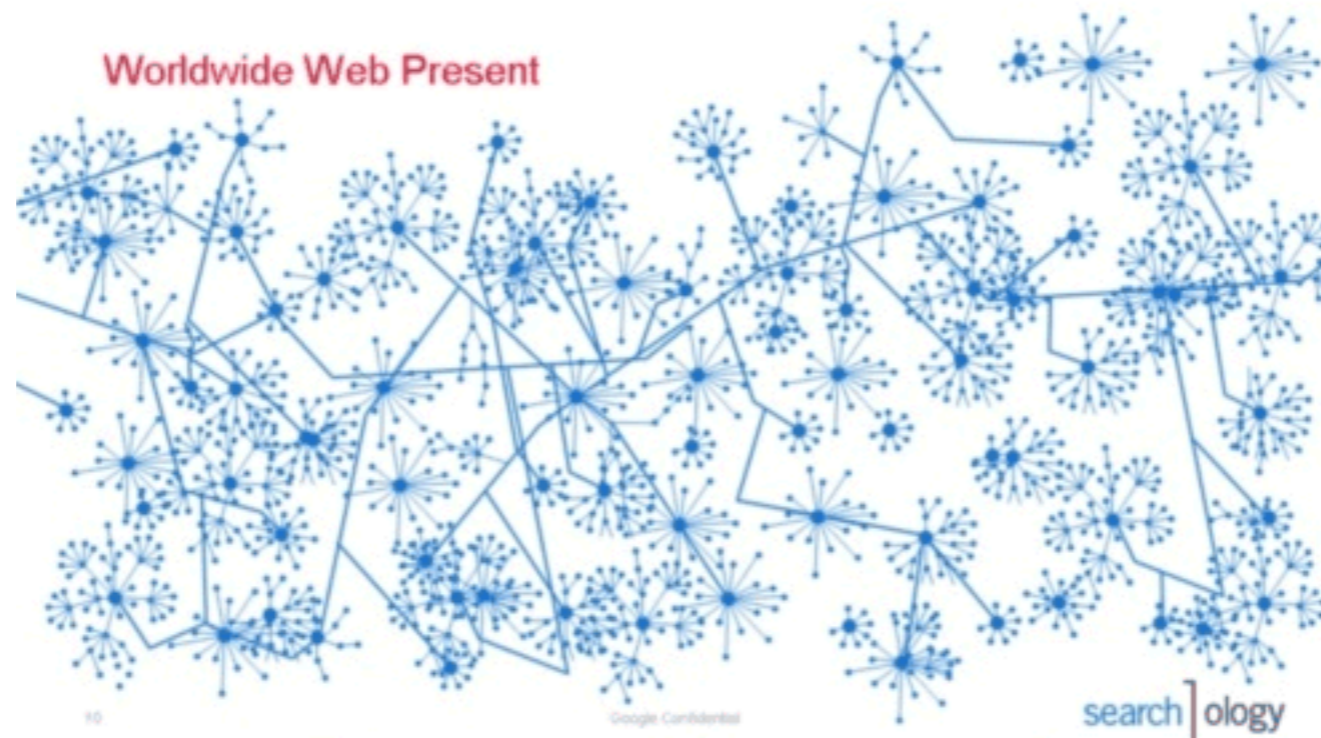
Nous avons vu

l'**explorateur** du web : partant d'une page au hasard, il clique sur un lien au hasard, puis sur un lien au hasard de la nouvelle page, etc., et de temps à autre redémarre d'une page prise au hasard...

la **construction** du web

Vers quoi mettre un lien à partir d'une nouvelle page ?

choisir une page au hasard et recopier un lien au hasard ← $1 - p$
ou de temps à autre mettre un lien vers une page au hasard ← p



Comment expliquer cette distribution ?

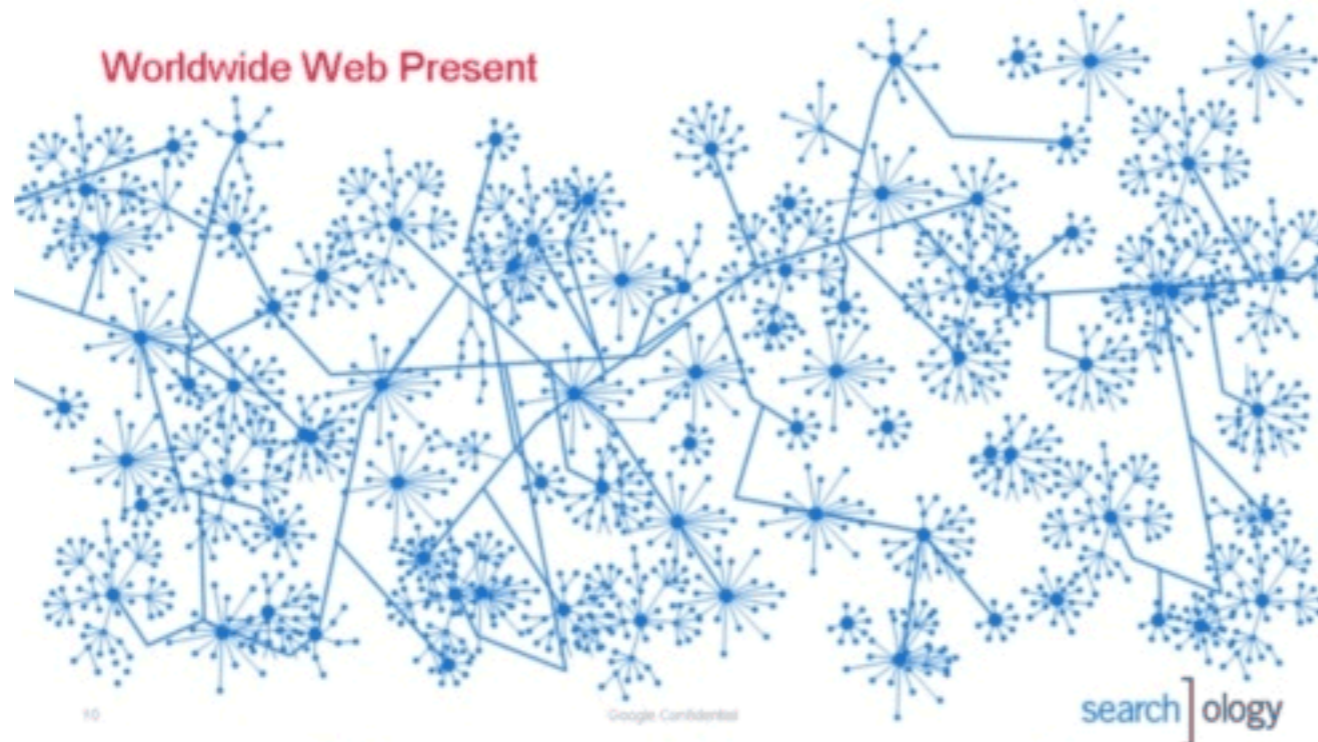
la **construction** du web

Vers quoi mettre un lien à partir d'une nouvelle page ?

choisir une page au hasard et recopier un lien au hasard

ou de temps à autre mettre un lien vers une page au hasard

$$\begin{array}{l} \leftarrow 1 - p \\ \leftarrow p \end{array}$$



Mettre un lien vers une page avec probabilité proportionnelle au nombre de liens vers elle

Plus une page a de liens vers elle,
plus elle en aura :
On ne prête qu'aux riches

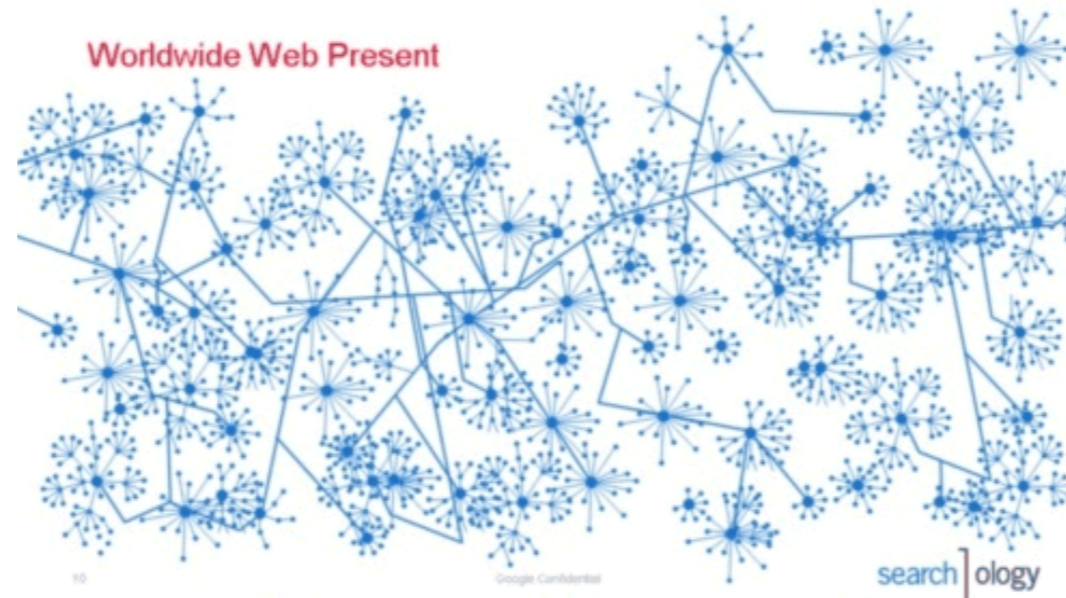


Comment expliquer cette distribution ? (1/3)

la **construction** du web

Vers quoi mettre un lien à partir d'une nouvelle page ?

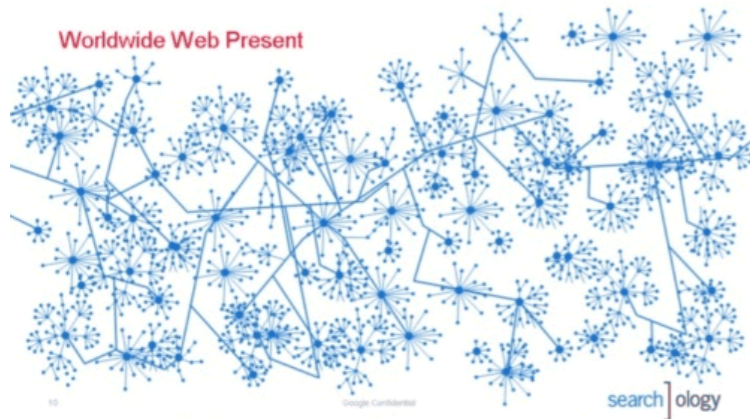
choisir une page au hasard et recopier un lien au hasard ← $1 - p$
ou de temps à autre mettre un lien vers une page au hasard ← p



$X_j(t)$ Nombre de liens vers la page j à l'instant t

$$X_j(j) = 0$$

$$X_j(t+1) = \begin{cases} X_j(t) + 1 & \text{avec proba. } (1-p) \frac{X_j(t)}{t} + p \frac{1}{t} \\ X_j(t) & \text{sinon} \end{cases}$$



(2/3)

$X_j(t)$ Nombre de liens vers la page j à l'instant t

Récurrance

$$X_j(j) = 0$$
$$X_j(t+1) = \begin{cases} X_j(t) + 1 & \text{avec proba. } (1-p) \frac{X_j(t)}{t} + p \frac{1}{t} \\ X_j(t) & \text{sinon} \end{cases}$$

Moyenne

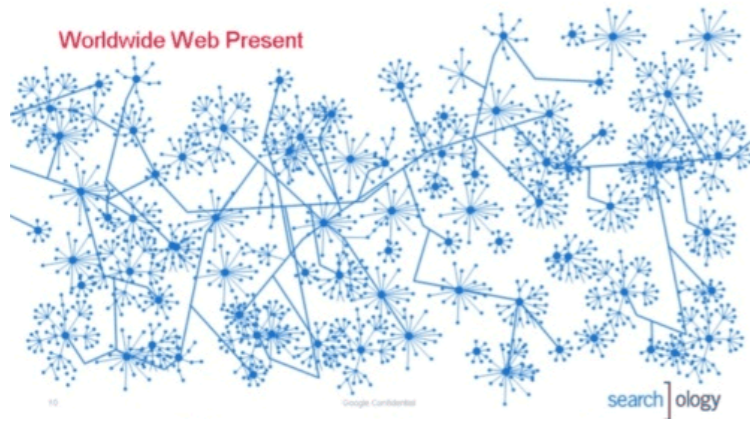
$$E[X_j(t+1) - X_j(t)] = (1-p) \frac{X_j(t)}{t} + p \frac{1}{t}$$

Équation différentielle

$$x_j(j) = 0 \qquad \frac{dx_j}{dt} = (1-p) \frac{x_j(t)}{t} + p \frac{1}{t}$$

Équation différentielle linéaire à coefficients constants

$$x_j(t) = \frac{p}{1-p} \left[\left(\frac{t}{j} \right)^{1-p} - 1 \right]$$



(3/3)

$X_j(t)$ Nombre de liens vers la page j à l'instant t

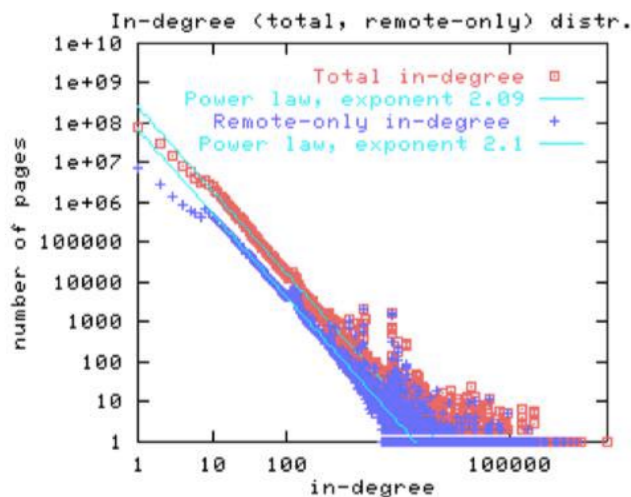
$$x_j(t) = \frac{p}{1-p} \left[\left(\frac{t}{j} \right)^{1-p} - 1 \right]$$

À quelle condition y a-t'il k liens vers la page j ?

$$\frac{p}{1-p} \left[\left(\frac{t}{j} \right)^{1-p} - 1 \right] \geq k \iff j \leq t \left[\frac{1-p}{p} k + 1 \right]^{-1/(1-p)}$$

Fraction de pages vers lesquelles pointent k liens

$$\left[\frac{1-p}{p} k + 1 \right]^{-1/(1-p)} \sim a k^{-1/(1-p)}$$



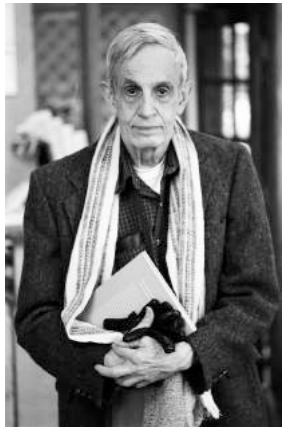
Loi de puissance



Le phénomène du petit monde dans la Toile et les réseaux sociaux

Le petit monde dans Wikipedia

équilibre de Nash (théorie des jeux)



John Nash

Un homme d'exception



Ron Howard

Apollo 13

NASA



| | | |
|---|------|------|
| | C | D |
| C | ☹️☹️ | ☹️☹️ |
| D | ☹️☹️ | 😊😊 |

équilibre de Nash (théorie des jeux)

John Nash

RAND

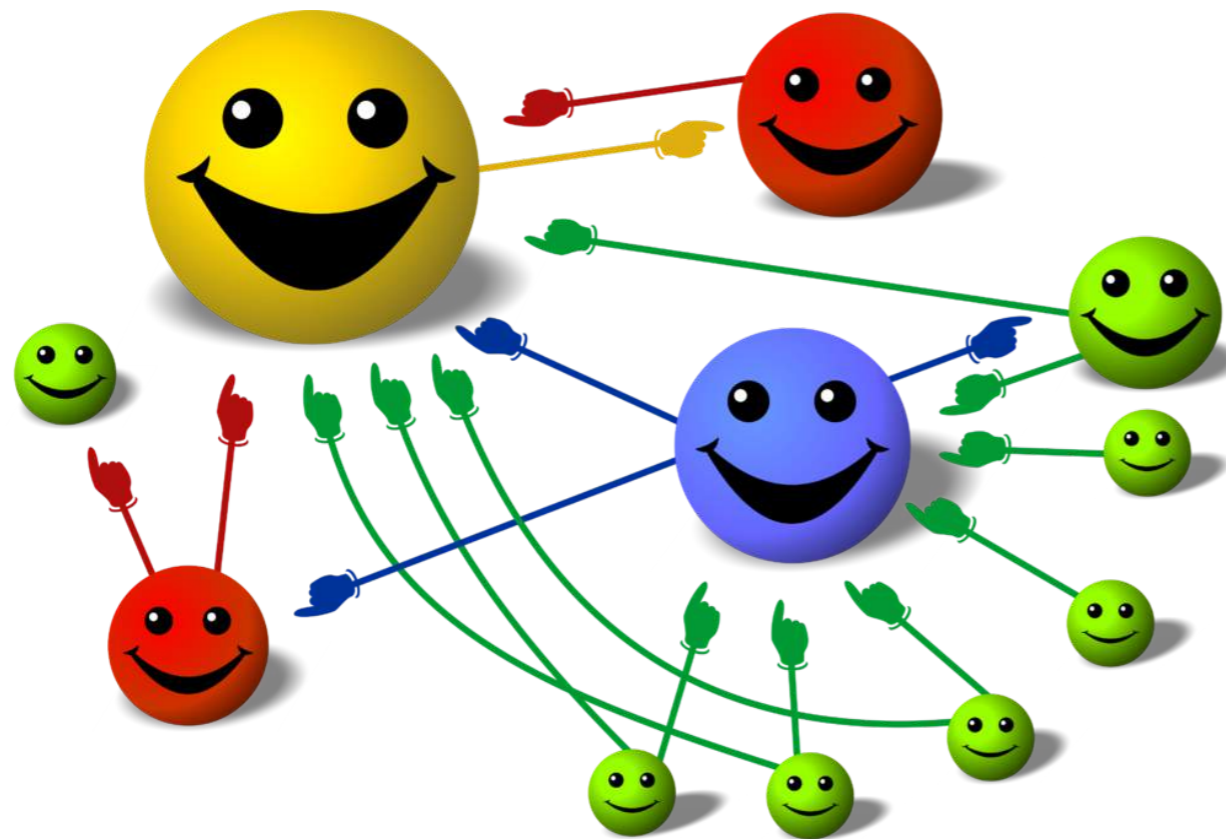
théorie du complot

NASA



C'est court !

Le petit monde dans la Toile



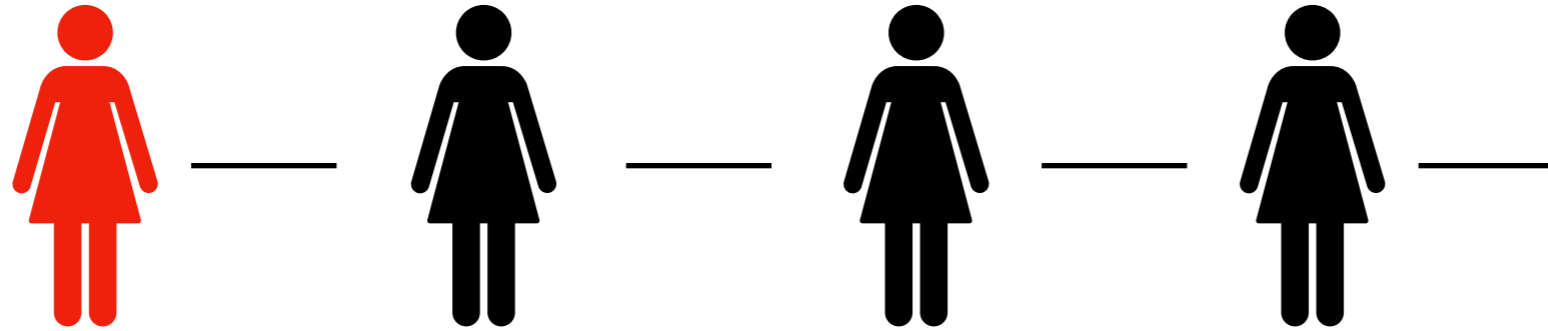
Très connectée :
La plupart des noeuds
ne sont qu'à
quelques liens de distance
les uns des autres

Le petit monde de don Camillo



Le petit monde dans les réseaux sociaux

Il n'y a que quelques degrés de séparation entre vous et Ivanka Trump.



La taille du réseau croît
mais pas les distances.

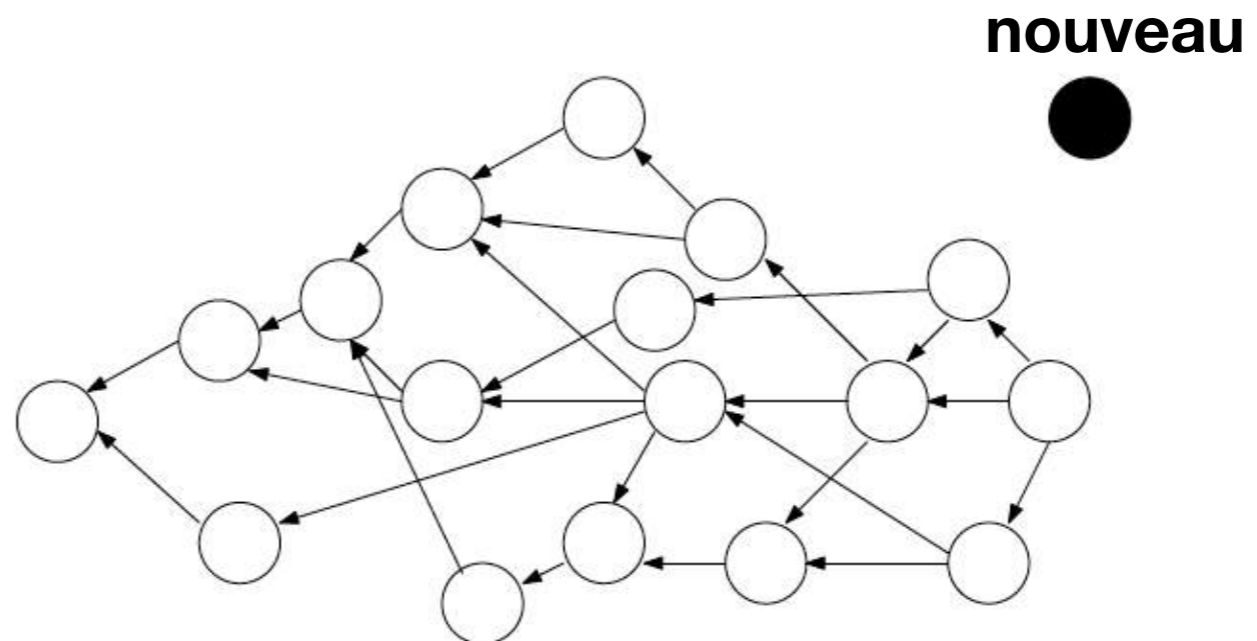
Comment l'expliquer ?

Modèle pour réseau social

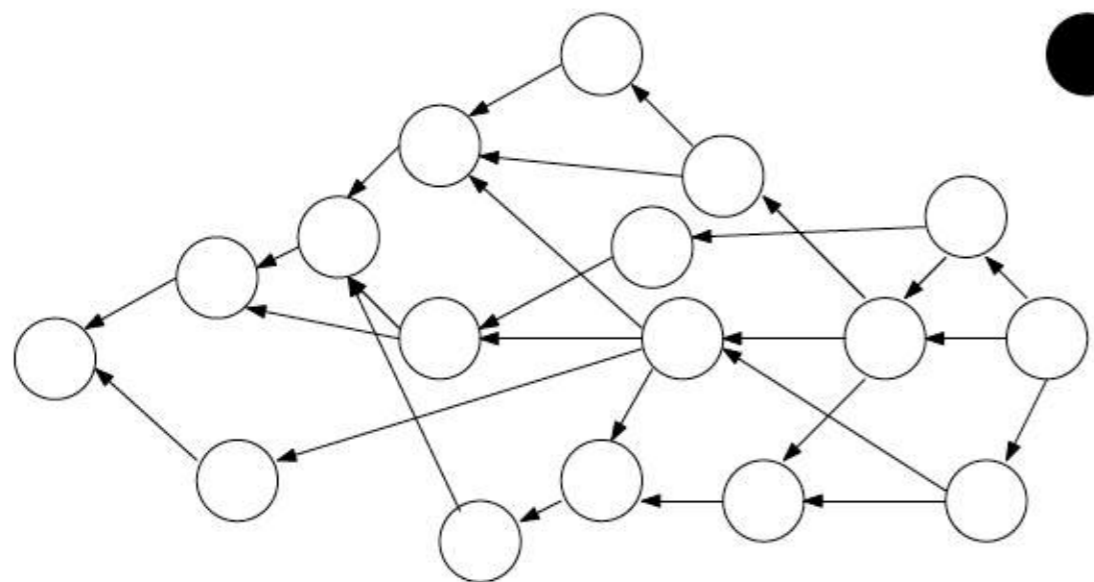


**Les amis
de mes amis
sont mes amis**

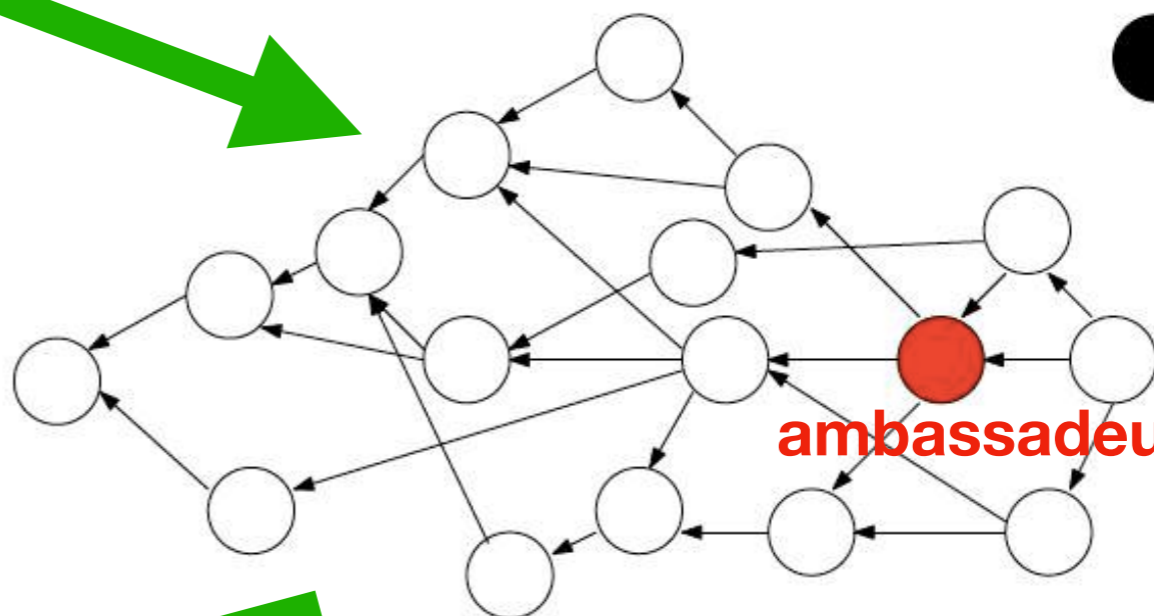
**Les nouveaux
sont introduits par une personne
("ambassadeur")
et se font un réseau
à partir des amis
de l'ambassadeur**



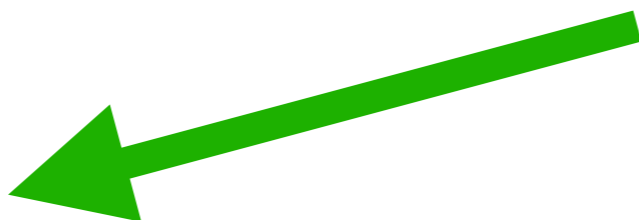
nouveau



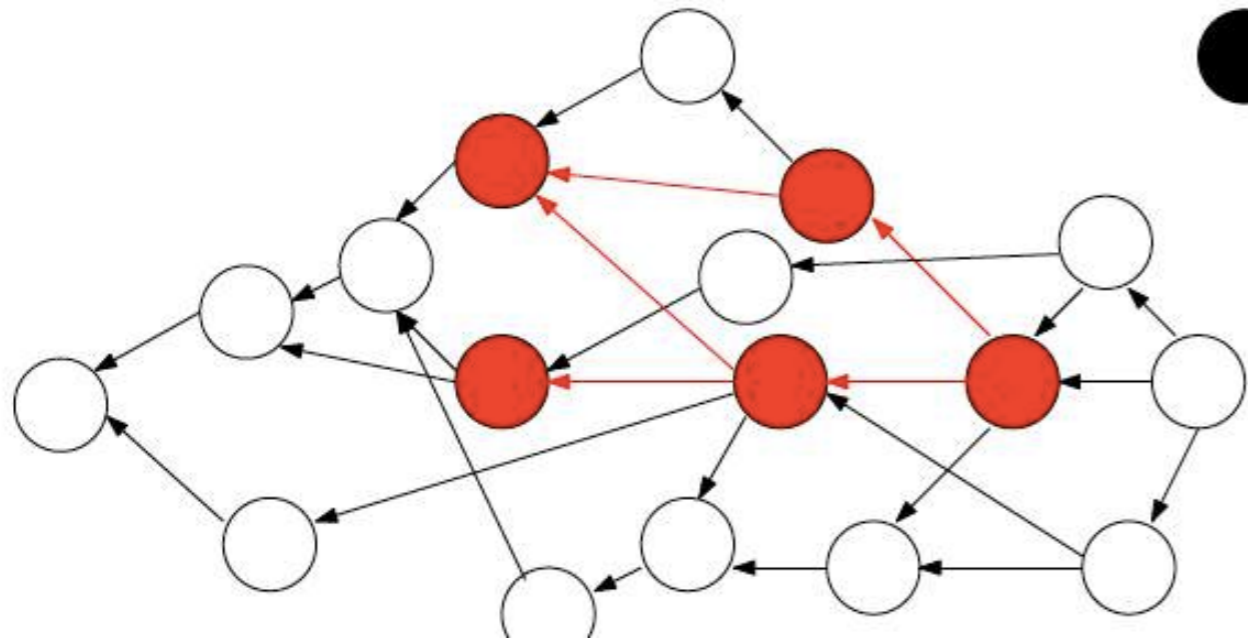
nouveau



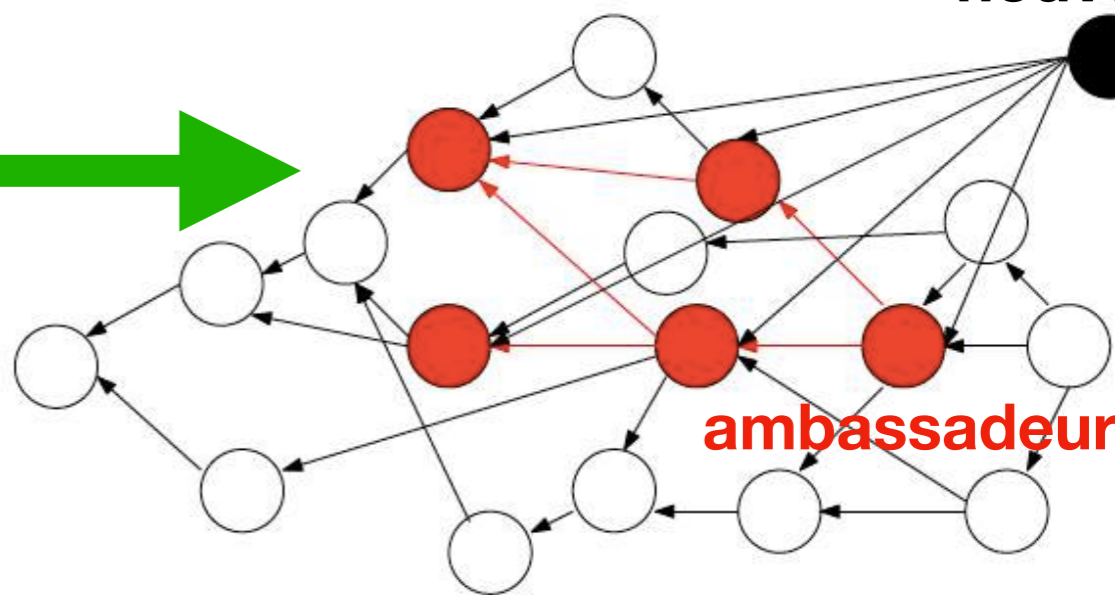
ambassadeur



$E(\text{Nb arêtes activées sortant de } u) = \alpha$



nouveau



ambassadeur

Analyse

**Si le graphe initial est assez gros
et si α est assez grand
alors la distance moyenne
entre deux sommets pris au hasard
est bornée**

Est-ce réaliste ?

Proche du modèle du feu de forêt pour les réseaux sociaux



Simulations

Nombre de liens : densification

Distribution des degrés : loi de puissance

Distance moyenne : bornée

Sur les réseaux sociaux, de citations,
d'acteurs, de courrier électronique, ...

Nombre de liens : densification

Distribution des degrés : loi de puissance

Distance moyenne : bornée

Le plafond de verre dans les réseaux sociaux académiques

Le plafond de verre



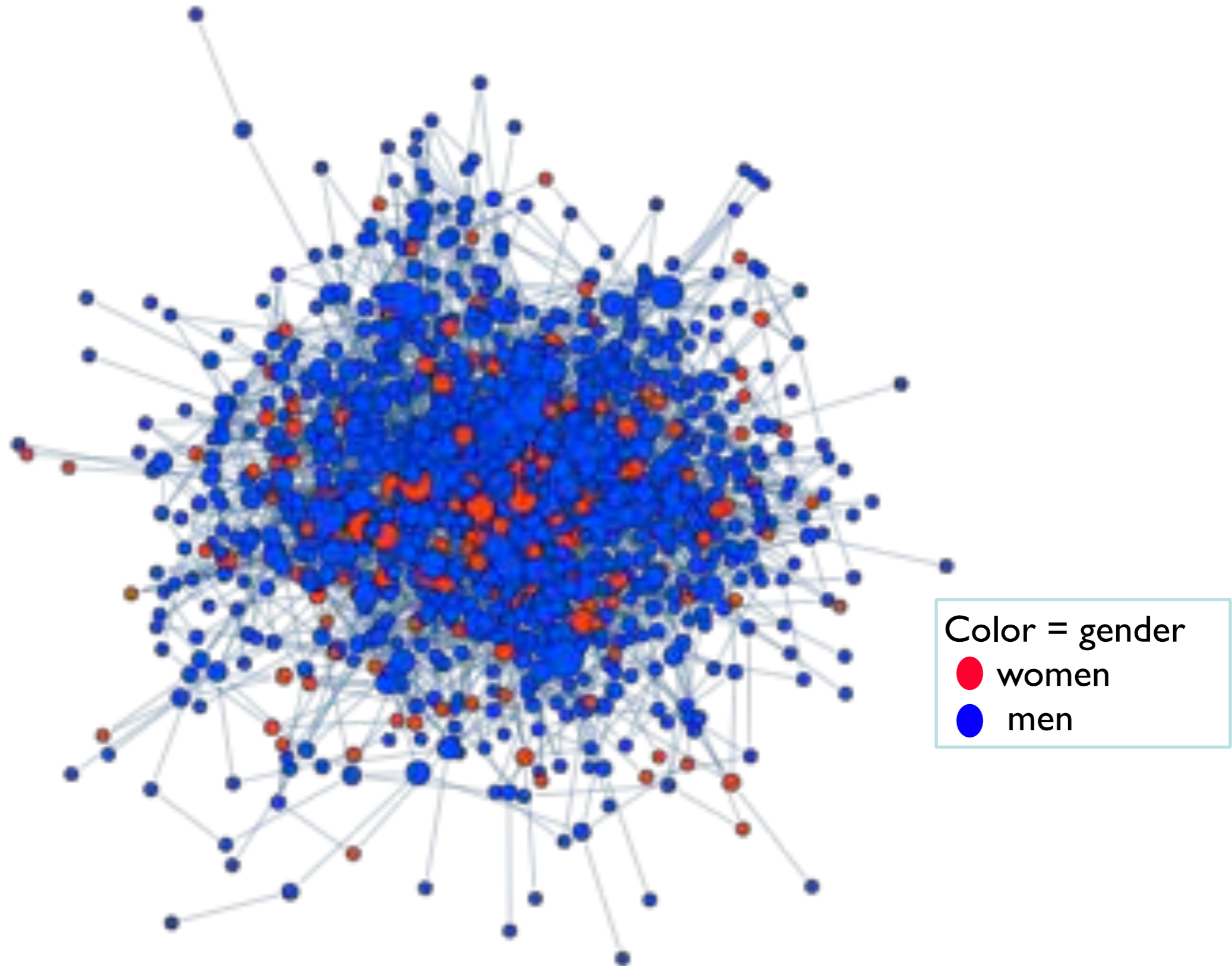
*La notion de « **plafond de verre** » renvoie au fait que les femmes peuvent progresser dans la hiérarchie de l'entreprise mais seulement jusqu'à un certain niveau. Résultat : elles sont en grande partie absentes du sommet de la hiérarchie.*

Modéliser plafond de verre dans réseau social académique

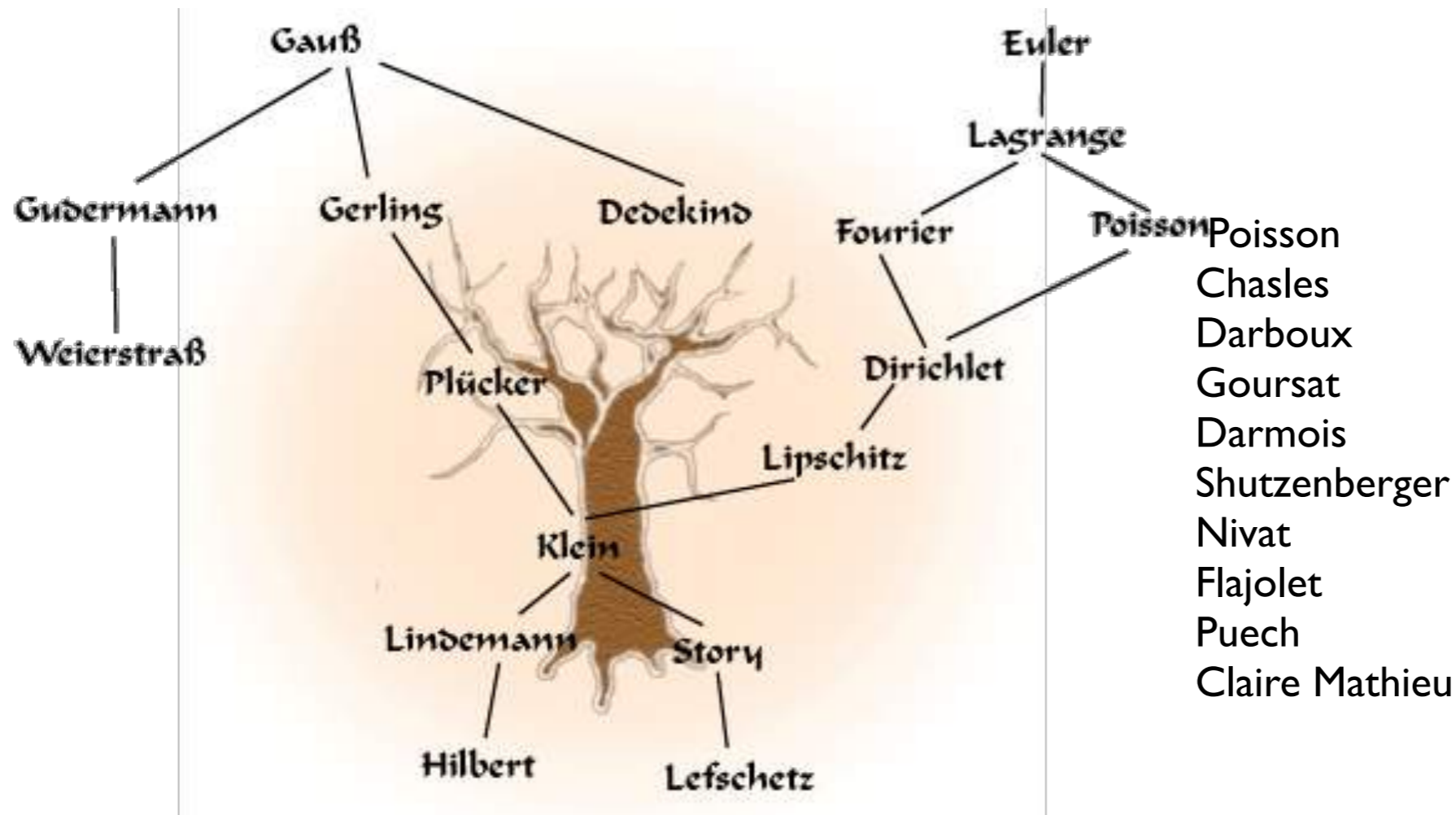
Analyser

Confronter à données empiriques

Réseau de publications DBLP



Réseau social académique : étudiants et directeurs de thèse



- sommets : chercheurs
- arêtes : étudiant et directeur de thèse

À partir de DBLP, on reconstruit le graphe en informatique

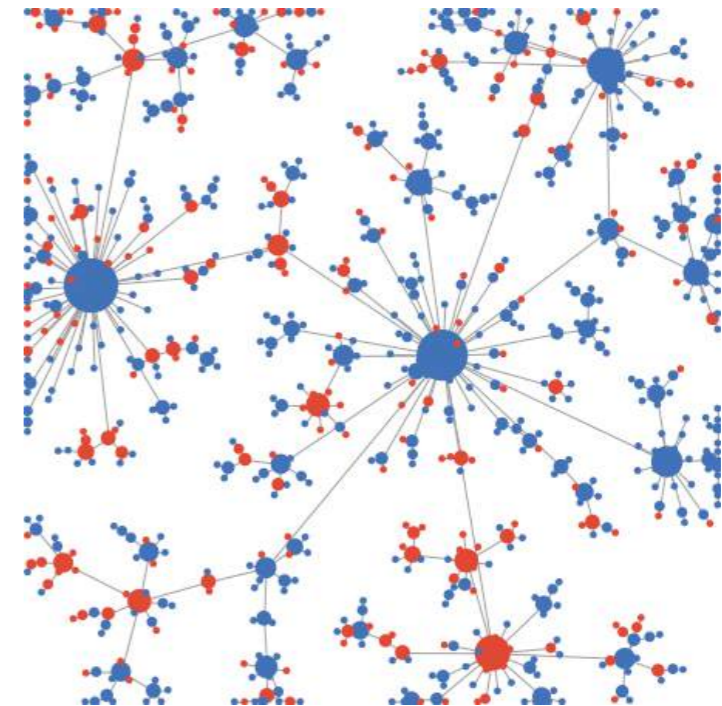
Trois observations, un modèle

1. En informatique, les femmes sont en minorité



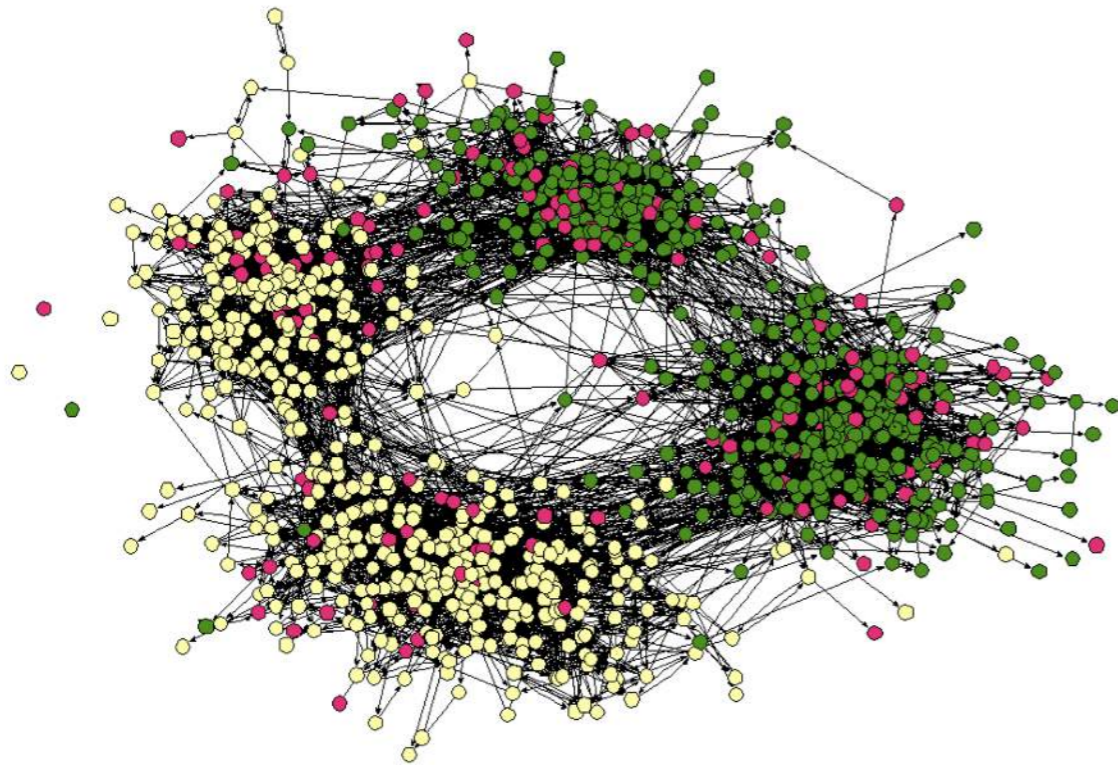
Trois observations, un modèle

1. En informatique, les femmes sont en minorité
2. Formation de liens : on ne prête qu'aux riches

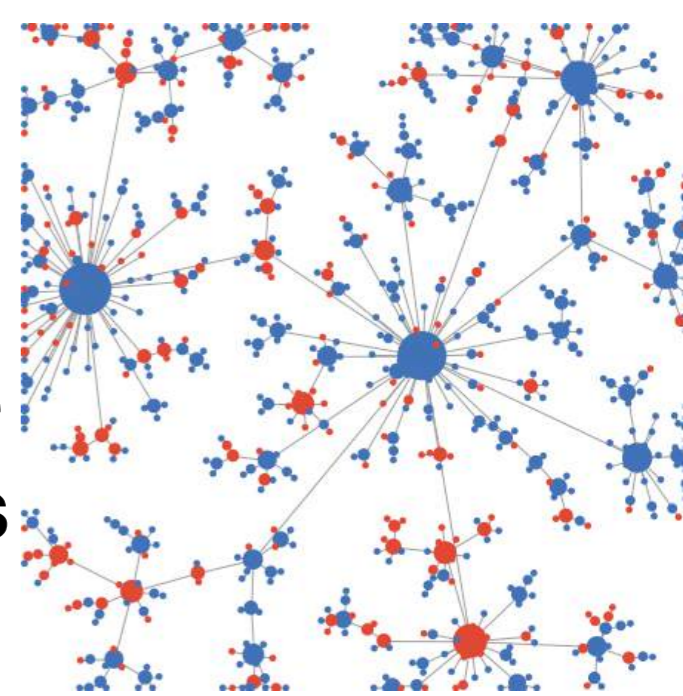


Trois observations, un modèle

1. En informatique, les femmes sont en minorité
2. Formation de liens : on ne prête qu'aux riches
3. Homophilie : qui se ressemble s'assemble



Trois observations, un modèle



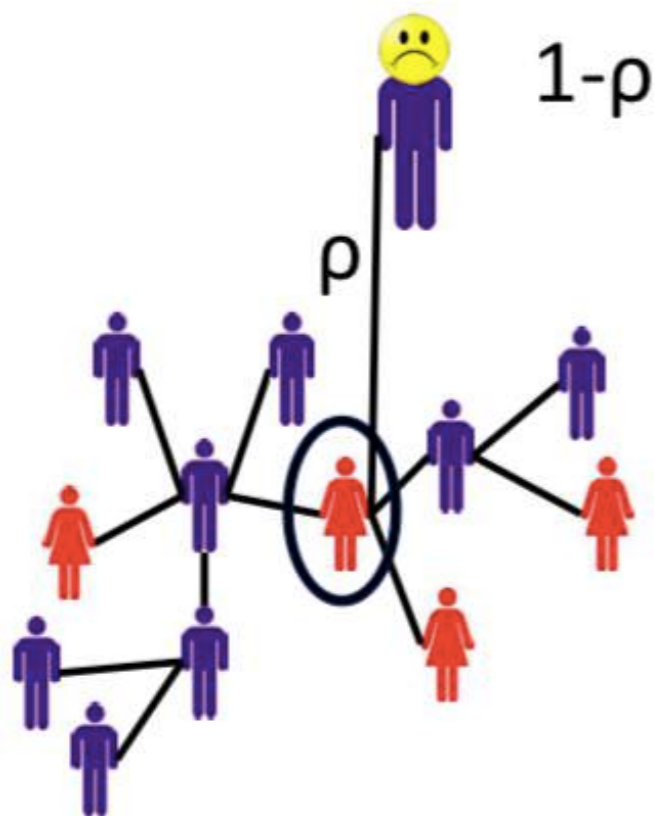
1. En informatique, les femmes sont en minorité
2. Formation de liens : on ne prête qu'aux riches
3. Homophilie : qui se ressemble s'assemble

Modèle

Nouveau noeud (étudiant) : **rouge** (r) ou **bleu** ($1-r$)

Son voisin (directeur de thèse) : probabilité proportionnelle au degré

Homophilie : si couleurs différentes, probabilité $1 - \rho$ de rejet et réessai



Plafond de verre

$$\frac{\text{Nb noeuds rouges de degre } \geq k}{\text{Nb noeuds de degre } \geq k} \rightarrow 0, \quad k \rightarrow \infty$$

Théorème

$$r < 1/2 \text{ et } \rho < 1 \implies \text{plafond de verre}$$

Idée de preuve

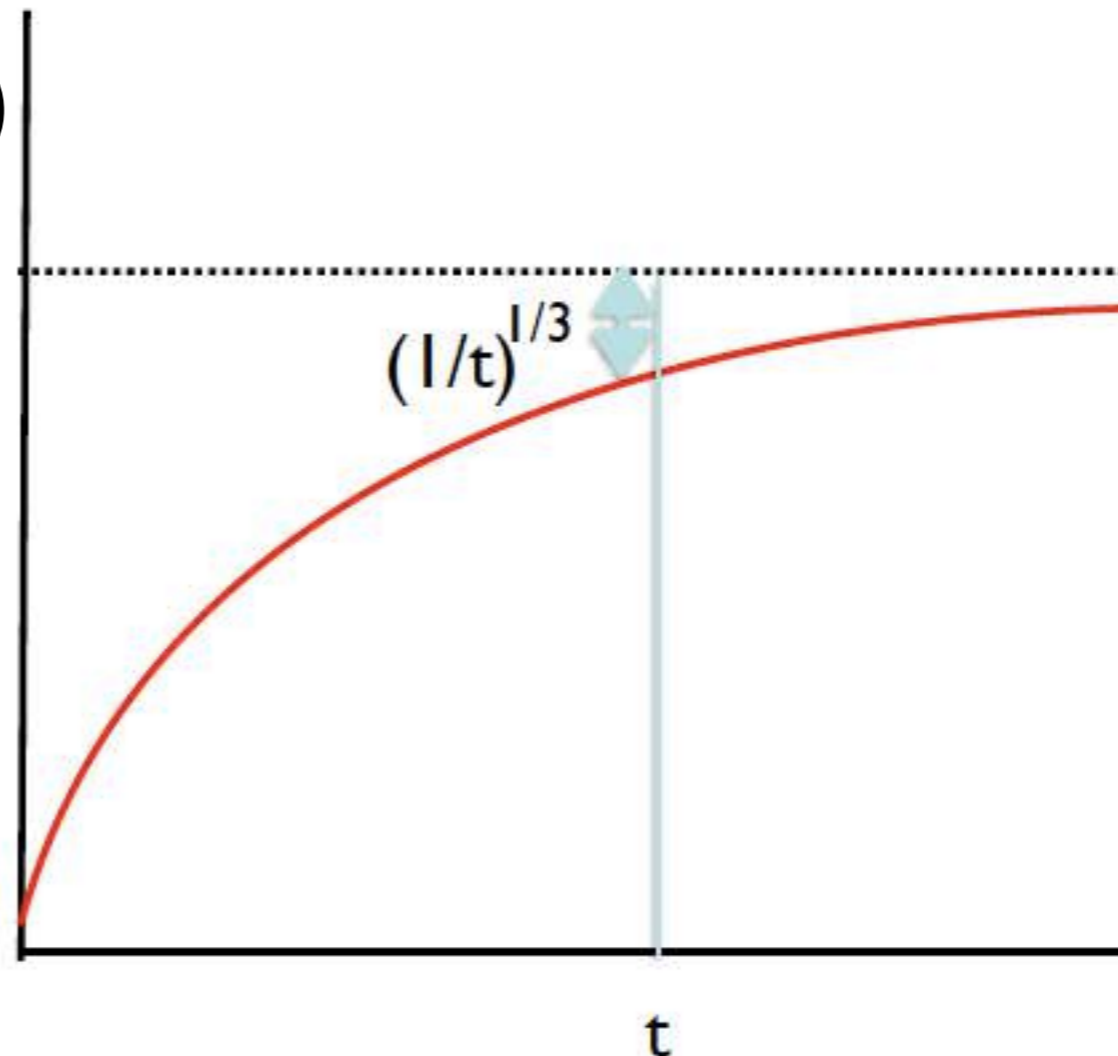
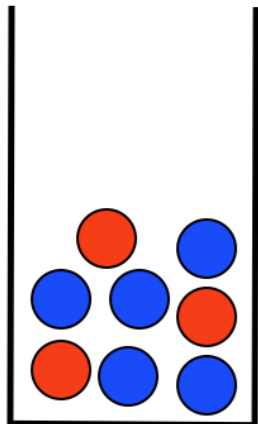
Analyser

$X_t =$ somme des degrés des noeuds rouges

$$X_{t+1} = \begin{cases} X_t & \text{avec proba. } \frac{1-r}{1-\rho} \\ X_t + 2 & \text{avec proba. } \frac{rX_t/(2t)}{1-(1-\rho)(1-X_t/(2t))} \\ X_t + 1 & \text{sinon} \end{cases}$$

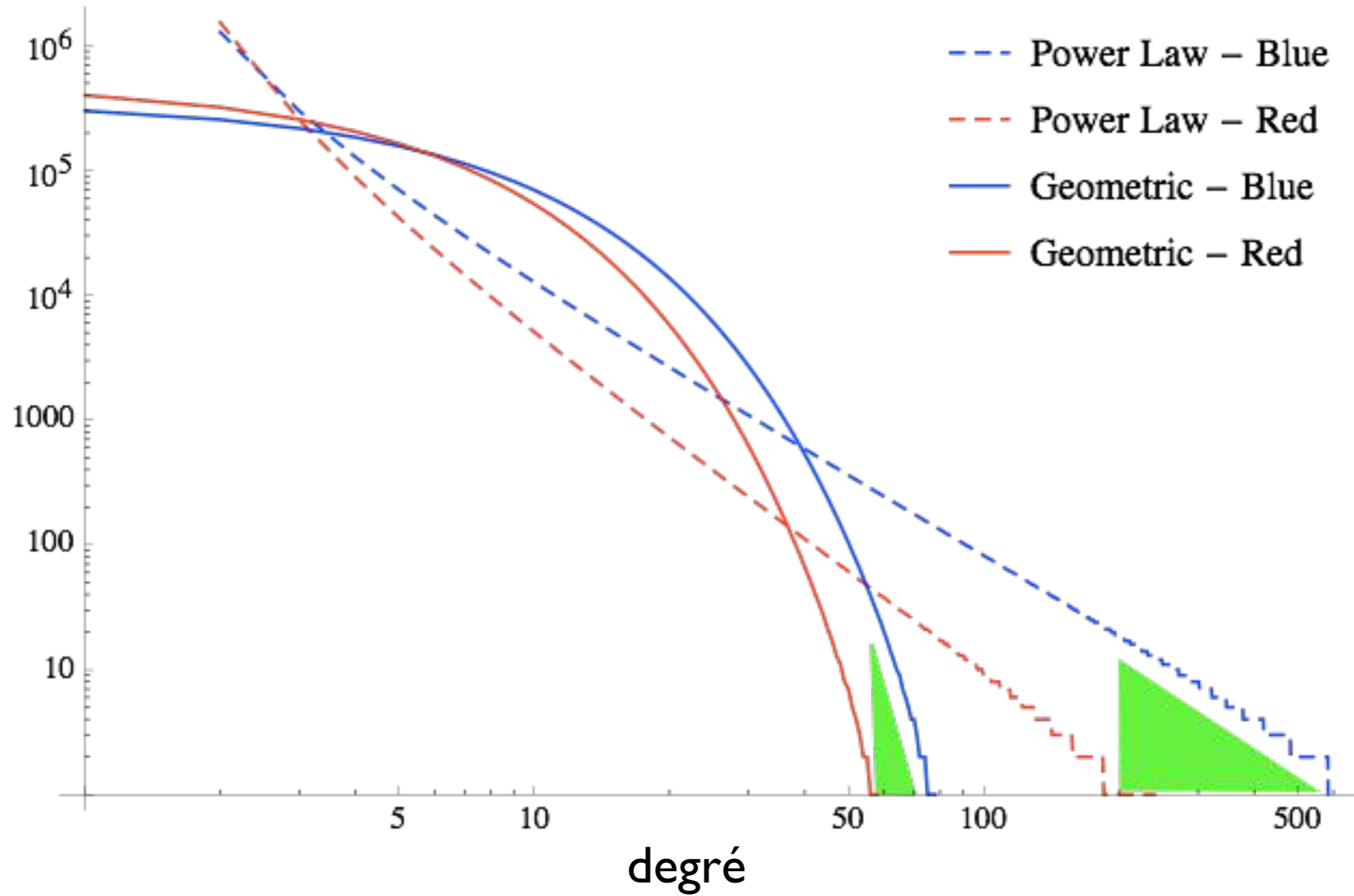
$$E(X_t/(2t))$$

Extension des techniques
pour analyser les
degrés entrants de la Toile



loi de puissance pour les degrés rouges
loi de puissance pour les degrés bleus

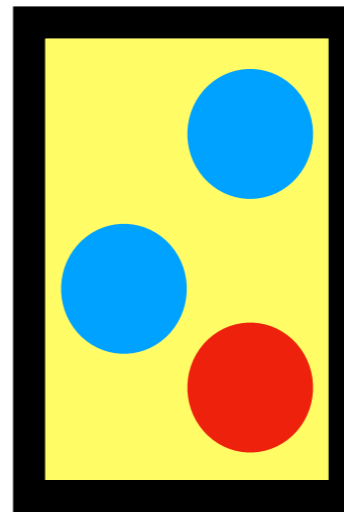
Nb. de noeuds



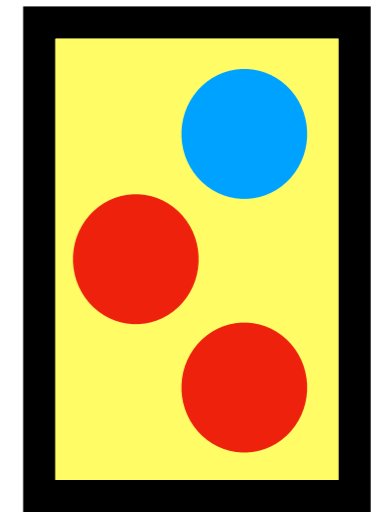
Bonus: la chambre d'échos

Une urne avec trois boules

100 personnes
chaque personne tire une boule de l'urne,
regarde sa couleur
la remet dans l'urne



ou



Chacun à son tour

chaque personne dit si elle pense que l'urne contient
une majorité de boules bleues, ou
une majorité de boules rouges.

Alice : dit la couleur de sa boule

Berthe : dit la couleur de sa boule

Charlotte : Si Alice et Berthe sont d'accord,
alors elle dit la même chose qu'Alice et Berthe,
sinon elle dit la couleur de sa boule

Danielle : si Alice et Berthe sont d'accord,
alors Danielle suit le mouvement
etc.

**Si Alice et Berthe ont tiré la même couleur de boule,
alors les 100 personnes disent pareil !!**

Conclusion : ce n'est pas parce que tout le monde dit pareil que c'est forcément vrai !

Conclusion

Quelques principes

Réseaux pour organiser l'information (Toile, Wikipedia)
Réseaux pour organiser les contacts humains (réseaux sociaux)
ont des structures communes

**Les modèles ne sont pas réalistes :
simplification extrême**

Mais une combinaison de :

- **choix de noeuds au hasard**
 - **établissement de nouveaux liens
par l'intermédiaire des voisins**
- suffit à faire émerger des phénomènes complexes**

Les lois de puissance sont omniprésentes

**Comprendre la structure des réseaux
pour profiter des avantages
sans de laisser induire en erreur par les phénomènes d'amplification**