

Enjeux numériques du monde contemporain

Algorithmes de classement & recommandations

Fabien Tarissan

CNRS – ENS Paris Saclay – ISP

Les AAls et API

Qu'est-ce que c'est ? Quel sont leurs rôle ?

- AAI : Autorité Administrative Indépendante
- API : Autorité Publique Indépendante

Elles sont créées par une loi.

Elles sont chargées de mission de **régulation**, de **contrôle** et d'**enquête**, de rendre des **décisions** (avis et injonctions).

Les AAls/APIs en lien avec le numérique ?

- ARCOM : Autorité de régulation de la communication audiovisuelle et numérique
(→ fusion du CSA et de HADOPI)
contenu, droit d'auteur, équilibre, protection
- ARCEP : Autorité de régulation des communications électroniques, des postes et de la distribution de la presse
économie, concurrence
- CNIL : Commission nationale de l'informatique et des libertés
libertés individuelles
- CADA : Commission d'accès aux documents administratifs

Autres agences ou comités en lien ?

- ANSII : Agence Nationale de la Sécurité des Systèmes d'Information
- CNNum : Conseil Nationale du Numérique
- CNPEN : Comité National Pilote d'Éthique du Numérique
- CESP : Comité Éthique et Scientifique Parcoursup
-

Internet et le web (rappel)

Internet

1966 projet **Arpanet**

29 oct. 1969 premier message

↳ *réseau de 4 sommets*

1971 projet Cyclade

1973 protocole **TCP/IP**
(Vint Cerf & Robert Kahn)

↳ *réseaux de ~ 50 de sommets*

1er jan. 1983 Arpanet adopte
TCP/IP

↳ *réseaux de ~ 1000 de sommets*

1987 Internet dépasse les
20 000 routeurs.

Le web

mars 1989 projet *World Wide Web* du CERN (Tim Berners-Lee & Robert Cailliau)

20 déc. 1990 1ère page web

30 avril 1993 CERN **renonce** aux droits

↳ *600 pages webs*

1994 *Netscape* et *Yahoo !*

1998 *Google*

2000 – ... *Facebook, Youtube, Twitter, ...*

2007 – ... *Deezer, Spotify, Netflix, ...*

En quête de visibilité

Tous les deux jours, nous créons autant d'information que l'humanité tout entière entre l'aube de son histoire et l'année 2003.

Eric Schmidt (PDG Google) en 2010

Le web en chiffres (2023)

<http://www.internetlivestats.com/>

- 5,5 milliards d'utilisateur
- 1,9 milliards de sites web
- 500 millions de tweets et 5 milliards de vidéos **par jours**
- 5 milliards de requêtes et 200 milliards de mails envoyés **par jours**

⇒ Nécessité d'**algorithmes** de classement

Quel impact sur l'information rendue visible ?

1. Les algorithmes de classement

2. L'impact des algorithmes sur notre manière de nous informer

Moteurs de recherche

Moteurs de recherche

Tâche d'un moteur de recherche :

- Sur réception d'une liste de mots-clefs (ex : "science" et "réseaux")
- Renvoyer une liste de pages web pertinentes *vis-à-vis* de ces mots-clefs.
 1. https://fr.wikipedia.org/wiki/Science_des_reseaux
 2. <https://www.pourlascience.fr/sd/informatique/le-monde-a-ses-reseaux-1082.php>
 3. ...

Moteurs de recherche

Tâche d'un moteur de recherche :

- Sur réception d'une liste de mots-clefs (ex : "science" et "réseaux")
- Renvoyer une liste de pages web pertinentes *vis-à-vis* de ces mots-clefs.
 1. https://fr.wikipedia.org/wiki/Science_des_reseaux
 2. <https://www.pourlascience.fr/sd/informatique/le-monde-a-ses-reseaux-1082.php>
 3. ...

Plus dur qu'il n'y paraît :

Collecte : Quelle est l'information disponible ?

Connaître le **contenu** et l'emplacement de l'**ensemble** des pages web

⇒ fait en continu par des programmes routiniers (*crawlers* ou *spider*)
(voir <https://fr.wikipedia.org/robots.txt>, <https://www.lemonde.fr/robots.txt>)

Indexation : Quelles pages correspondent aux requêtes ?

Génération d'une **indexation complète** associant à chaque mot-clef une liste de page web.
⇒ mis à jour en continu

Classement : Comment ordonner les pages ?

Choix d'un critère de classement : comment **quantifier la pertinence** ?
Algorithme pour systématiser le calcul

⇒ fait en continu

L'importance d'une page web

Question (1) : comment quantifier l'importance d'une page web ?

L'importance d'une page web

Question (1) : comment quantifier l'importance d'une page web ?

- Avant Google : médiamétrie (trafic, nb de visites, de clics, ...) $\Rightarrow$ popularité
- Idée Brin & Page (1998) : les références hypertextes constituent des gages d'autorité

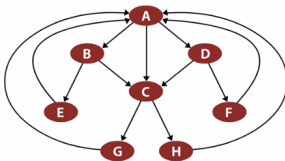
L'importance d'une page web

Question (1) : comment quantifier l'importance d'une page web ?

- Avant Google : médiamétrie (trafic, nb de visites, de clics, ...) $\Rightarrow$ popularité
- Idée Brin & Page (1998) : les références hypertextes constituent des gages d'autorité

Un des résultats de la phase de collecte : le réseau du web, ie. un graphe :

- Nœuds = pages web
- Liens = liens hypertextes



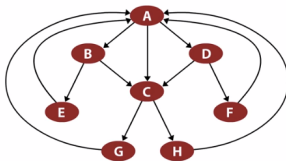
L'importance d'une page web

Question (1) : comment quantifier l'importance d'une page web ?

- Avant Google : médiamétrie (trafic, nb de visites, de clics, ...) $\Rightarrow$ popularité
- Idée Brin & Page (1998) : les références hypertextes constituent des gages d'autorité

Un des résultats de la phase de collecte : le réseau du web, ie. un graphe :

- Nœuds = pages web
- Liens = liens hypertextes



Question (2) : comment quantifier l'importance d'un nœud dans un réseau de citations ?

PageRank et marches aléatoires (1)

PageRank est basée sur la notion de marche aléatoire :

1. un marcheur part d'un nœud choisi aléatoirement
2. il choisit aléatoirement l'un des liens sortants du nœud
3. retourne en 2

PageRank et marches aléatoires (1)

Pagerank est basée sur la notion de marche aléatoire :

1. un marcheur part d'un nœud choisi aléatoirement
2. il choisit aléatoirement l'un des liens sortants du nœud
3. retourne en 2

Importance (score) d'un nœud v = probabilité d'observer le marcheur aléatoire sur le nœud v après une infinité d'étapes.

PageRank et marches aléatoires (1)

Pagerank est basée sur la notion de marche aléatoire :

1. un marcheur part d'un nœud choisi aléatoirement
2. il choisit aléatoirement l'un des liens sortants du nœud
3. retourne en 2

Importance (score) d'un nœud v = probabilité d'observer le marcheur aléatoire sur le nœud v après une infinité d'étapes.

Problème : et si le graphe a un (ou plusieurs) puits ? ou *spider trap* ?

PageRank et marches aléatoires (2)

Pagerank est basée sur la notion de marche aléatoire :

1. un marcheur part d'un nœud choisi aléatoirement
2. il se "téléporte" sur un nœud choisi aléatoirement avec probabilité α
3. S'il ne se téléporte pas alors :
 - il choisit aléatoirement l'un des liens sortant du nœud
 - si le nœud n'a pas de liens sortants, il se téléporte sur un nœud choisi aléatoirement
4. retourne en 2

PageRank(v) = probabilité d'observer le marcheur aléatoire sur le nœud v après une infinité d'étapes.

PageRank et marches aléatoires (2)

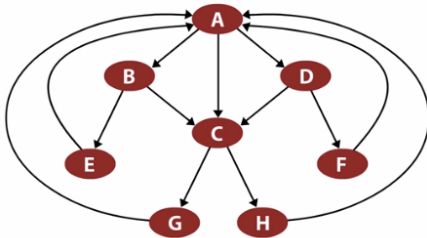
Pagerank est basée sur la notion de marche aléatoire :

1. un marcheur part d'un nœud choisi aléatoirement
2. il se "téléporte" sur un nœud choisi aléatoirement avec probabilité α
3. S'il ne se téléporte pas alors :
 - il choisit aléatoirement l'un des liens sortant du nœud
 - si le nœud n'a pas de liens sortants, il se téléporte sur un nœud choisi aléatoirement
4. retourne en 2

PageRank(v) = probabilité d'observer le marcheur aléatoire sur le nœud v après une infinité d'étapes.

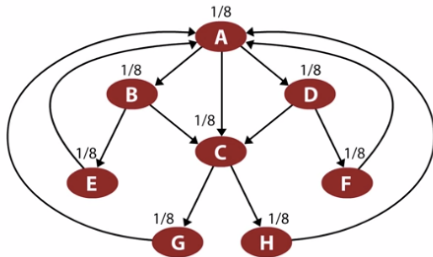
Question (3) : comment estimer cette probabilité ?

Algorithme du PageRank : intuition



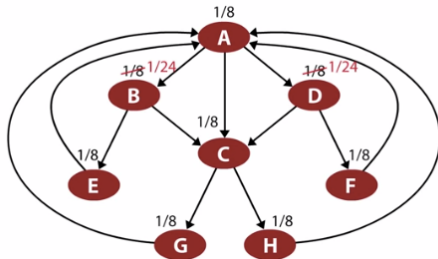
The PageRank Citation Ranking : Bringing Order to the Web, Lawrence Page, Sergey Brin, Rajeev Motwani, Terry Winograd, Technical Report. Stanford InfoLab, 1998.

Algorithme du PageRank : intuition



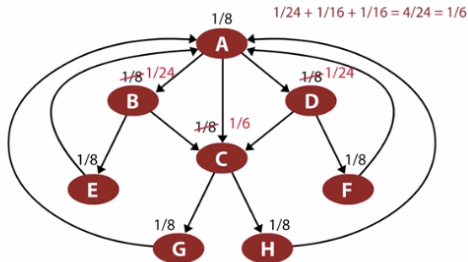
On donne un poids équivalent (normalisé) à l'ensemble des sommets du graphe : $1/N$

Algorithme du PageRank : intuition



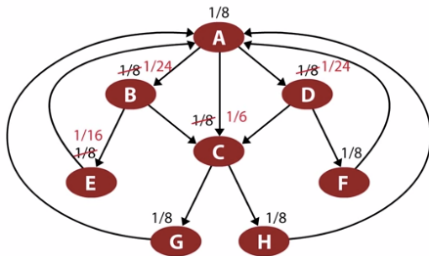
MAJ : On répartit **simultanément** le PR équitablement vers les nœuds sortants

Algorithme du PageRank : intuition



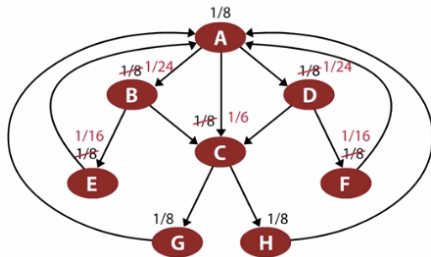
MAJ : On répartit **simultanément** le PR équitablement vers les nœuds sortants

Algorithme du PageRank : intuition



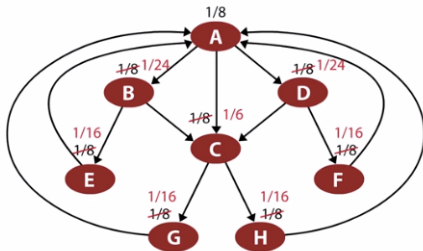
MAJ : On répartit **simultanément** le PR équitablement vers les nœuds sortants

Algorithme du PageRank : intuition



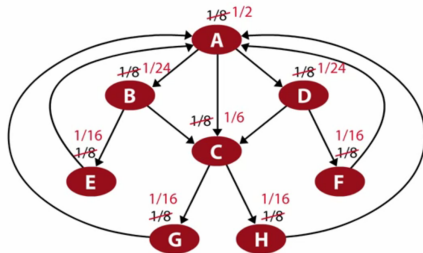
MAJ : On répartit **simultanément** le PR équitablement vers les nœuds sortants

Algorithme du PageRank : intuition



MAJ : On répartit **simultanément** le PR équitablement vers les nœuds sortants

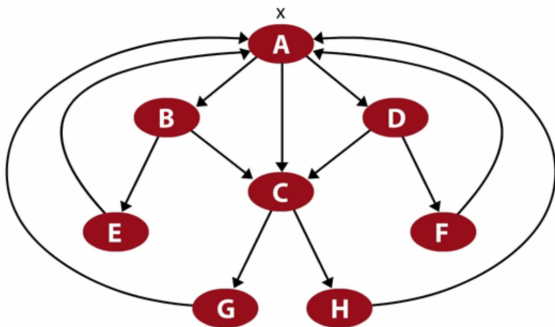
Algorithme du PageRank : intuition



MAJ : On répartit **simultanément** le PR équitablement vers les nœuds sortants

Algorithme du PageRank : principe

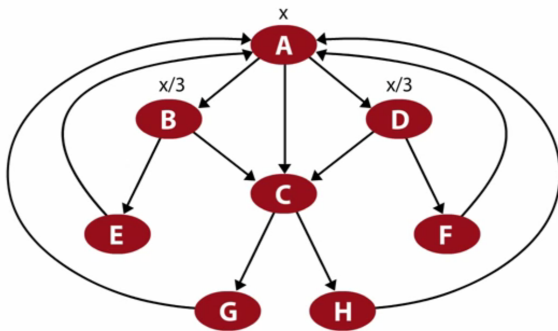
Quel est l'état final ?
[admis] un seul état d'équilibre



YouTube

Algorithme du PageRank : principe

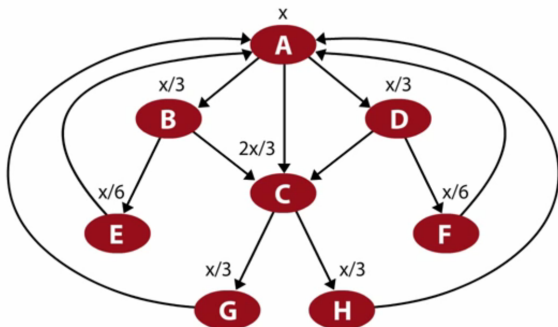
Quel est l'état final ?
[admis] un seul état d'équilibre



YouTube

Algorithme du PageRank : principe

Quel est l'état final ?
[admis] un seul état d'équilibre



YouTube

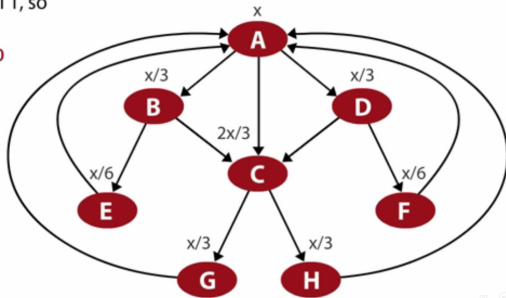
Algorithme du PageRank : principe

Quel est l'état final ?
[admis] un seul état d'équilibre

$$x + x/3 + x/3 + 2x/3 + x/6 + x/6 + x/3 + x/3 = 10x/3$$

This must equal 1, so

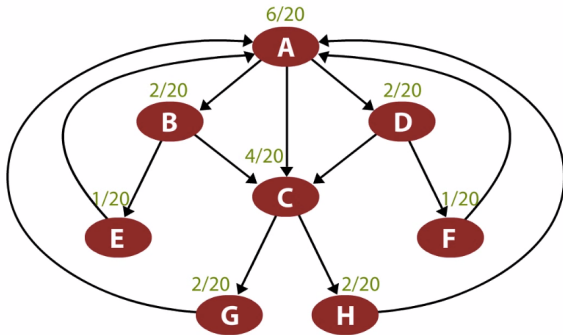
- $10x/3 = 1$
- $x = 3/10 = 6/20$



YouTube

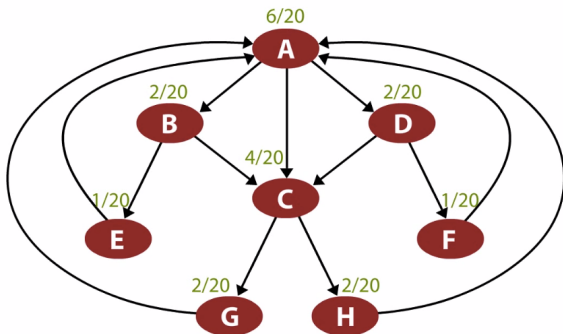
Algorithme du PageRank : principe

Quel est l'état final ?
[admis] un seul état d'équilibre



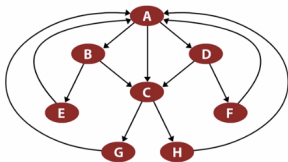
Algorithme du PageRank : principe

Quel est l'état final ?
[admis] un seul état d'équilibre



Question (4) : comment déterminer les valeurs stationnaires ?

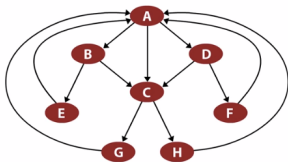
Algorithme du PageRank : matrices



On cherche à résoudre le système d'équations suivant :

- $PR_A = PR_F + PR_H + PR_G + PR_E$
- $PR_B = \frac{PR_A}{3}$
- $PR_C = \frac{PR_A}{3} + \frac{PR_B}{2} + \frac{PR_D}{2}$
- ...
- $PR_H = \frac{PR_C}{2}$

Algorithme du PageRank : matrices

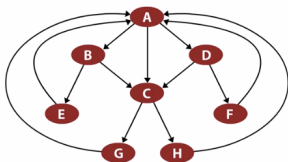


On cherche à résoudre le système d'équations suivant :

- $PR_A = PR_F + PR_H + PR_G + PR_E$
- $PR_B = \frac{PR_A}{3}$
- $PR_C = \frac{PR_A}{3} + \frac{PR_B}{2} + \frac{PR_D}{2}$
- ...
- $PR_H = \frac{PR_C}{2}$

⇒ C'est à dire à **déterminer le vecteur** $PR = [PR_A, PR_B, \dots, PR_H]$

Algorithme du PageRank : matrices

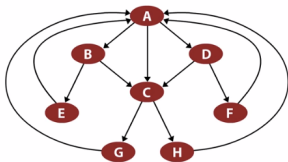


Matrice de transition T du graphe G :

- matrice carré $n \times n$
- chaque lien dirigé (u, v) de G donne la valeur $T[u][v] = \frac{1}{\deg(u)}$

0	1/3	1/3	1/3	0	0	0	0
0	0	1/2	0	1/2	0	0	0
0	0	0	0	0	0	1/2	1/2
0	0	1/2	0	0	1/2	0	0
1	0	0	0	0	0	0	0
1	0	0	0	0	0	0	0
1	0	0	0	0	0	0	0
1	0	0	0	0	0	0	0

Algorithme du PageRank : matrices



Matrice de transition T du graphe G :

- matrice carré $n \times n$
- chaque lien dirigé (u, v) de G donne la valeur $T[u][v] = \frac{1}{\deg(u)}$

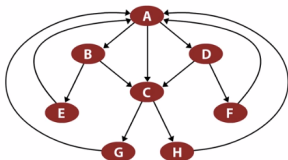
Problème des puits !

On travaille sur une variation de T qui tient compte du coefficient d'évaporation :

$$B = (1 - \alpha)T + \alpha \frac{1}{n} \cdot I$$

où I est la matrice ne contenant que des 1.

Algorithme du PageRank : matrices

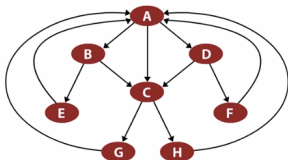


Si PR est le vecteur contenant le score du PageRank (inconnue), on retrouve le système d'équation :

$$B \times PR = PR$$

PR est un vecteur propre de la matrice de transition (variante) du graphe G .

Algorithme du PageRank : matrices



Si PR est le vecteur contenant le score du PageRank (inconnue), on retrouve le système d'équation :

$$B \times PR = PR$$

PR est un vecteur propre de la matrice de transition (variante) du graphe G .

Question (5) : comment calculer un vecteur propre d'une matrice ?

Algorithme du PageRank : implémentation

Pourquoi c'est difficile ? On doit :

- additionner des matrices
- multiplier des matrices par un réel
- multiplier une **matrice par un vecteur**

Rappel : le produit $B \times PR$ définit le vecteur X tel que :

$$\forall i \in \llbracket 1, n \rrbracket, X_i = \sum_{j=1}^n B_{ij} \times PR_j$$

Algorithme du PageRank : implémentation

Pourquoi c'est difficile ? On doit :

- additionner des matrices
- multiplier des matrices par un réel
- multiplier une **matrice par un vecteur**

Rappel : le produit $B \times PR$ définit le vecteur X tel que :

$$\forall i \in \llbracket 1, n \rrbracket, X_i = \sum_{j=1}^n B_{ij} \times PR_j$$

Algorithm 1 Classic matrix/vector multiplication

```
function PRODMATVECT( $B, PR$ )  
  for  $i$  from 1 to  $n$  do  
     $X[i] \leftarrow 0$   
    for  $j$  from 1 to  $n$  do  
       $X[i] += B[i][j] \times PR[j]$   
  return  $X$ 
```

Algorithme du PageRank : implémentation

Pourquoi c'est difficile ? On doit :

- additionner des matrices
- multiplier des matrices par un réel
- multiplier une **matrice par un vecteur**

Rappel : le produit $B \times PR$ définit le vecteur X tel que :

$$\forall i \in \llbracket 1, n \rrbracket, X_i = \sum_{j=1}^n B_{ij} \times PR_j$$

Algorithm 2 Classic sparse matrix/vector multiplication

```
function PRODMATVECT( $B, PR$ )  
  for  $i$  from 1 to  $n$  do  
     $X[i] \leftarrow 0$   
    for  $j$  from 1 to  $n$  such that  $B[i][j] \neq 0$  do  
       $X[i] += B[i][j] \times PR[j]$   
  return  $X$ 
```

Algorithme du PageRank : implémentation

Pourquoi c'est difficile ? On doit :

- additionner des matrices
- multiplier des matrices par un réel
- multiplier une **matrice par un vecteur**

Rappel : le produit $B \times PR$ définit le vecteur X tel que :

$$\forall i \in \llbracket 1, n \rrbracket, X_i = \sum_{j=1}^n B_{ij} \times PR_j$$

Algorithm 3 Edge iteration sparse matrix/vector multiplication

```
function PRODMATVECT2( $B, PR$ )  
  for  $i$  from 1 to  $n$  do  
     $X[i] \leftarrow 0$   
  for each  $(i, j)$  such that  $B[i][j] \neq 0$  do  
     $X[i] += B[i][j] \times PR[j]$   
  return  $X$ 
```


Algorithme du PageRank : implémentation

Pourquoi c'est difficile ? On doit :

- additionner des matrices
- multiplier des matrices par un réel
- multiplier une **matrice par un vecteur**

Rappel : le produit $B \times PR$ définit le vecteur X tel que :

$$\forall i \in \llbracket 1, n \rrbracket, X_i = \sum_{j=1}^n B_{ij} \times PR_j$$

Algorithm 4 Power iteration to compute PageRank

```
function POWERITERATION( $G, \alpha, t$ )  
   $T \leftarrow$  transition matrix of graph  $G$   
   $P \leftarrow \frac{1}{n} \times I$   
  for  $i$  from 1 to  $t$  do  
     $P \leftarrow \text{PROD MAT VECT2}(T, P)$   
     $P \leftarrow (1 - \alpha) \times P + \alpha \times I$   
     $P \leftarrow \text{NORMALIZE}(P)$   
return  $P$ 
```

$$\triangleright \forall i \in \llbracket 1, n \rrbracket, P[i] \leftarrow \frac{P[i]}{\|P\|_1}$$

Réseaux sociaux

Réseaux sociaux

Activités des utilisateurs variées :

- recherche
- lecture
- dépôts (*posts, tweets, ...*)
- marques d'intérêt (commentaires, *likes, ...*)
- relations ("amis" FB, *followers, ...*)
- propagation (*retweets, ...*)
- ...

Réseaux sociaux

Activités des utilisateurs variées :

- recherche
- lecture
- dépôts (*posts*, *tweets*, ...)
- marques d'intérêt (commentaires, *likes*, ...)
- relations ("amis" FB, *followers*, ...)
- propagation (*retweets*, ...)
- ...

Tâche des réseaux sociaux :

Identifier les informations pertinentes

- adaptées à chaque utilisateurs
- sans être guidé par des mot-clefs

Exemple de Facebook : le *Newsfeed*

Pari de Zuckerberg (2004) : c'est la régularité de l'activité des utilisateurs qui va guider l'algorithme pour identifier leurs centres d'intérêt.

EdgeRank

Comme pour un moteur de recherche, il y a plusieurs phases

Collecte : récolter les informations susceptibles d'intéresser un utilisateurs

↪ relations entre utilisateurs ($\sim$ facile)

Classement : ordonner les posts

↪ l'algorithme du EdgeRank

EdgeRank

Comme pour un moteur de recherche, il y a plusieurs phases

Collecte : récolter les informations susceptibles d'intéresser un utilisateurs

↪ relations entre utilisateurs ($\sim$ facile)

Classement : ordonner les posts

↪ l'algorithme du EdgeRank

Edgerank

Chaque relation entre un utilisateur B et un post p d'un "ami" A de B est évalué :

$$S(A, B, p) = \text{Aff}(B, A) \times W(p) \times D(p)$$

Affinité : $\text{Aff}(B, A)$ mesure l'affinité **déclarée** de B vers A

Poids : $W(p)$ mesure le poids du post p

Fraicheur : $D(p)$ mesure l'ancienneté du post p

EdgeRank

Comme pour un moteur de recherche, il y a plusieurs phases

Collecte : récolter les informations susceptibles d'intéresser un utilisateurs

↪ relations entre utilisateurs ($\sim$ facile)

Classement : ordonner les posts

↪ l'algorithme du EdgeRank

Edgerank

Chaque relation entre un utilisateur B et un post p d'un "ami" A de B est évalué :

$$S(A, B, p) = \text{Aff}(B, A) \times W(p) \times D(p)$$

Affinité : $\text{Aff}(B, A)$ mesure l'affinité **déclarée** de B vers A

Poids : $W(p)$ mesure le poids du post p

Fraicheur : $D(p)$ mesure l'ancienneté du post p

En fait, EdgeRank est une somme :

$$ER(p) = \sum_B S(A, B, p)$$

Quelle évolution ?

Dominique Cardon, À quoi rêvent les algorithmes. Nos vies à l'heure des big data, Paris, Seuil, La République des idées, 2015, 105 p.

Évolution des algorithmes classements de l'information :

Popularité (À côté du web) : **nombre** de références.
Clics, views, ..

Autorité (Au dessus du web) : **autorité** des références
Links $\Rightarrow$ PageRank

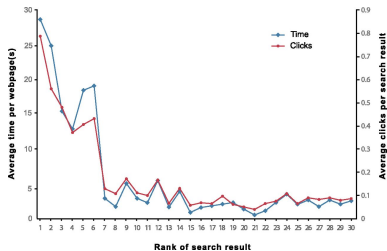
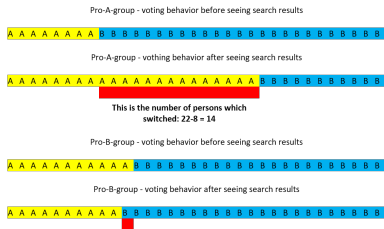
Notoriété (À l'intérieur du web) : **affinité** (et fraîcheur !) avec l'information issue de son **voisinage**.
Tweet, likes, ... $\Rightarrow$ EdgeRank

Prédiction (Sous le web) : information **inférée** par le comportement d'un utilisateur.
Traces $\Rightarrow$ Youtube / Spotify / Deezer / Netflix / ...

*Quel impact ont ces
algorithmes*

Search Engine Manipulation effect

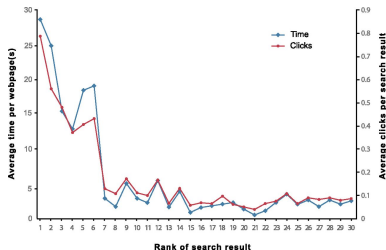
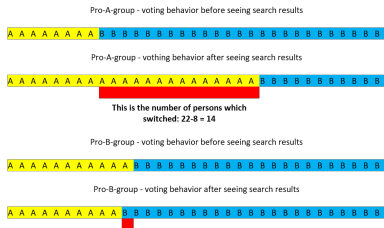
The search engine manipulation effect (SEME) and its possible impact on the outcomes of elections. Robert Epstein and Ronald E. Robertson. PNAS 112 (33), 2015 (doi :10.1073/pnas.1419828112).



1. Quel est l'objectif de cette étude ?
2. Quelle méthode est mise en place ?
3. Quels sont les résultats obtenus ?
4. D'où vient (qu'est-ce qui explique) le phénomène mis en évidence ?
5. En quoi la manipulation est-elle particulièrement dangereuse ?
6. Quelles critiques peut-on faire de l'article ?

Search Engine Manipulation effect

The search engine manipulation effect (SEME) and its possible impact on the outcomes of elections. Robert Epstein and Ronald E. Robertson. PNAS 112 (33), 2015 (doi :10.1073/pnas.1419828112).



Résultats

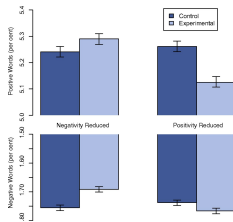
- Classements biaisés ont un impact sur les votants (indécis) : $\geq 25\%$
- Les moteurs de recherche biaisés sont indétectés

À mettre en perspective avec :

- 95% de la navigation web se fait sur 0.03% des pages existantes ...
- **The Fake News Machine. How Propagandists Abuse the Internet and Manipulate the Public**, Gu, Kropotov and Yarochkin, Trendlabs research paper, Trend Micro, 2015.

Emotional contagion

Experimental evidence of massive-scale emotional contagion through social networks. Adam D. I. Kramer, Jamie E. Guillory, and Jeffrey T. Hancock . PNAS 111 (24), 2014 (doi :10.1073/pnas.1320040111).



Résultats (N = 689 003)

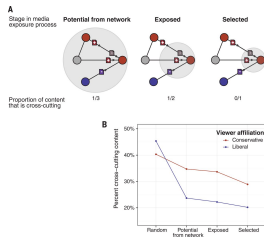
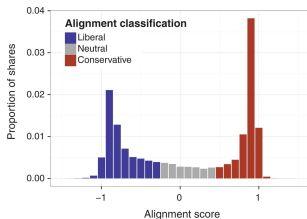
Les utilisateurs exposés à des contenus véhiculant une émotion "positive" (resp. "négative") se mette à poster des contenus positifs (resp. "négatifs").

À mettre en perspective avec **New Beauty Study Reveals Days, Times And Occasions When U.S. Women Feel Least Attractive**, PRNews Media, 2014.

Women
normale
supérieure
paris-saclay

Chambres d'écho & bulles filtrantes

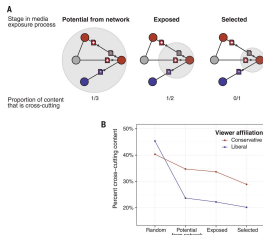
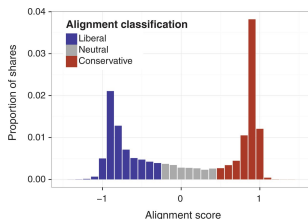
Exposure to ideologically diverse news and opinion on Facebook. Eytan Bakshy, Solomon Messing, Lada A. Adamic. Science, 348 (6239), 2015 (doi :10.1126/science.aaa1160).



1. Quel est l'objectif de cette étude ?
2. Quelle méthode est mise en place ?
3. Quels sont les résultats obtenus ?
4. D'où provient le phénomène de confinement ?
5. Quelles critiques peut-on faire de l'article ?

Chambres d'écho & bulles filtrantes

Exposure to ideologically diverse news and opinion on Facebook. Eytan Bakshy, Solomon Messing, Lada A. Adamic. Science, 348 (6239), 2015 (doi :10.1126/science.aaa1160).



Résultats (N = 10.1 M)

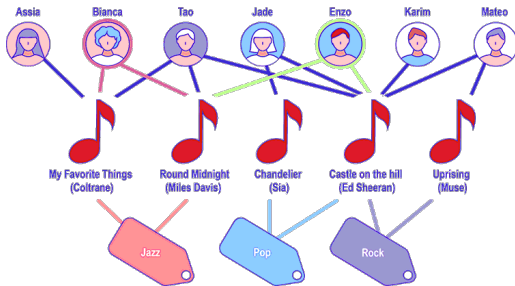
- Les chambres d'échos sont dues principalement aux relations sociales
- Petit effet du filtrage algorithmique

Mesurer la diversité

Proposition : brancher le résultat de marches aléatoires sur des indices classiques de dispersion :

Soit $P = \text{RandWalk}(\mathbb{T}, u) = (p_i)_i$:

$$\alpha\text{-diveristy} : \text{div}_\alpha(P) = \left(\sum_i p_i^\alpha \right)^{\frac{1}{1-\alpha}}$$



$$\mathcal{H}(\mathbb{T}, \text{Bianca}) = 1 \text{ mais } \mathcal{H}(\mathbb{T}, \text{Enzo}) = \frac{8}{3}$$
$$\mathcal{H}(\mathbb{T}, \text{Jazz}) = 3.6 \text{ mais } \mathcal{H}(\mathbb{T}, \text{Rock}) = 2.5$$

Une expérience

Filtrage collaboratif pour données implicites

- Utilisé par bcp de plateformes (*Netflix, Youtube, Spotify, Amazon, ...*)
- Fonder une recommandation sur ce que des utilisateurs *similaires* ont aimé
- S'appuie sur la *Factorisation Matricielle* (Matrice = Utilisateurs $\times$ Produits)

Expérience

Après entraînement du modèle sur une partie du jeu de données :

1. **Prédiction** : inférence de la *force* des liens manquants entre utilisateurs et produits
2. **Sélection** : pour chaque utilisateur, classement puis sélection des *k*-top produits ($k = 10, 50, 500$)

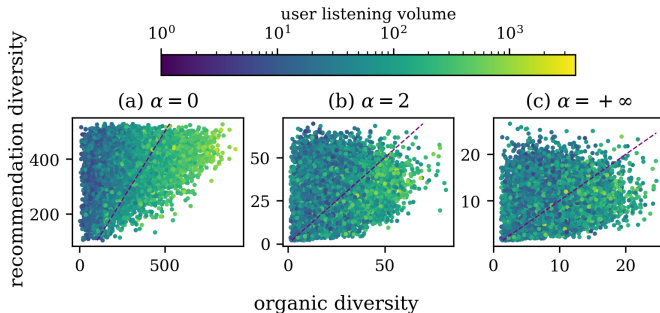
Ce qui permet de mesurer :

- la diversité des utilisateurs avant recommandation
- la diversité des recommandations
- la diversité des utilisateurs **après avoir été exposé à *k* recommandations** (décroissance linéaire des poids)

Diversité des recommandations

Quelle est la diversité des recommandations ?

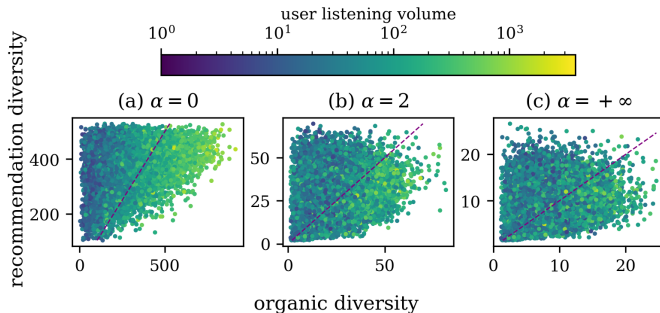
Est-elle liée à celle des utilisateurs ?



Diversité des recommandations

Quelle est la diversité des recommandations ?

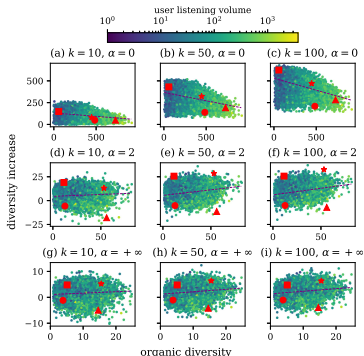
Est-elle liée à celle des utilisateurs ?



- pas de corrélation forte entre la diversité organique et celle des recommandations
- les recommandations tendent à être plus diversifiées que le profil d'écoute

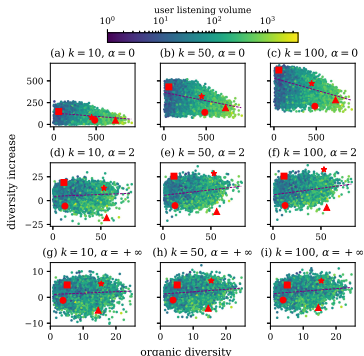
Effets des recommandations

Quelle diversité des utilisateurs
après avoir été exposés aux recommandations ?



Effets des recommandations

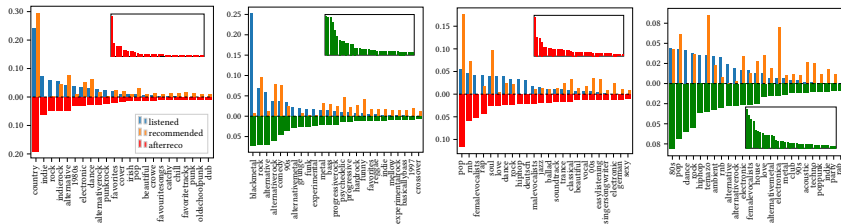
Quelle diversité des utilisateurs
après avoir été exposés aux recommandations ?



- La majorité des utilisateurs voit sa **diversité augmentée** par les recommandations
- La diversité **augmente avec le nombre de produits** recommandés
- L'effet sur la diversité **dépend du profil de l'utilisateur**

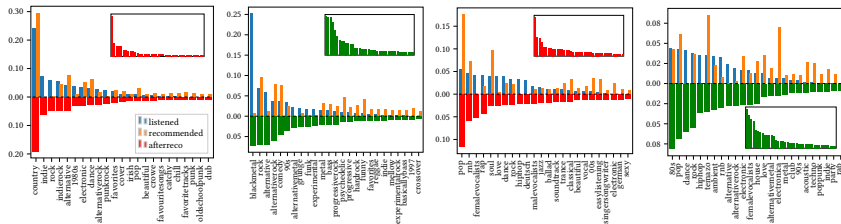
Différents effets

Comment expliquer les **différences** dans la manière dont les recommandations affectent les utilisateurs ?



Différents effets

Comment expliquer les **différences** dans la manière dont les recommandations affectent les utilisateurs ?



- les recommandations **augmentent la variété** des produits
- mais **échouent à fournir une exposition équilibrée**

Questions ?

<http://tarissan.complexnetworks.fr/>