Comment fonctionne Google ?

Michael Eisermann

Universität Stuttgart

Conférence le 26 octobre 2009
Document mis à jour le 6 novembre 2009

Journées nationales de l'APMEP à Rouen, 24–27 octobre 2009
« Explorer les mathématiques, les mathématiques pour explorer »

www.igt.uni-stuttgart.de/eiserm/popularisation/#google

Plan de l'exposé

Plan de l'exposé

L’entreprise Google

D’un projet d’étudiants à une entreprise mondiale :

L’entreprise Google

D’un projet d’étudiants à une entreprise mondiale :

- Fondée en 1998 par Sergey Brin et Larry Page.

L’entreprise Google

D’un projet d’étudiants à une entreprise mondiale :

- Fondée en 1998 par Sergey Brin et Larry Page.
- Depuis 2000 vente de publicités.

L’entreprise Google

D’un projet d’étudiants à une entreprise mondiale :

- Fondée en 1998 par Sergey Brin et Larry Page.
- Depuis 2000 vente de publicités.
- Août 2004 lancement en bourse.

L'entreprise Google

D'un projet d'étudiants à une entreprise mondiale :

- Fondée en 1998 par Sergey Brin et Larry Page.
- Depuis 2000 vente de publicités.
- Août 2004 lancement en bourse.

Valeur en bourse (octobre 2009) plus de 100 milliards USD

L’entreprise Google

D’un projet d’étudiants à une entreprise mondiale :

- Fondée en 1998 par Sergey Brin et Larry Page.
- Depuis 2000 vente de publicités.
- Août 2004 lancement en bourse.

Valeur en bourse	(octobre 2009)	plus de 100 milliards USD
Chiffres d’affaires	(année 2008)	21.8 milliards USD

L'entreprise Google

D'un projet d'étudiants à une entreprise mondiale :

- Fondée en 1998 par Sergey Brin et Larry Page.
- Depuis 2000 vente de publicités.
- Août 2004 lancement en bourse.

Valeur en bourse	(octobre 2009)	plus de 100 milliards USD
Chiffres d'affaires	(année 2008)	21.8 milliards USD
Résultat net	(année 2008)	4.2 milliards USD

L’entreprise Google

D’un projet d’étudiants à une entreprise mondiale :

- Fondée en 1998 par Sergey Brin et Larry Page.
- Depuis 2000 vente de publicités.
- Août 2004 lancement en bourse.

Valeur en bourse	(octobre 2009)	plus de 100 milliards USD
Chiffres d’affaires	(année 2008)	21.8 milliards USD
Résultat net	(année 2008)	4.2 milliards USD
Employés	(sept. 2009)	environ 20 000

L’entreprise Google

D’un projet d’étudiants à une entreprise mondiale :

- Fondée en 1998 par Sergey Brin et Larry Page.
- Depuis 2000 vente de publicités.
- Août 2004 lancement en bourse.

Valeur en bourse	(octobre 2009)	plus de 100 milliards USD
Chiffres d’affaires	(année 2008)	21.8 milliards USD
Résultat net	(année 2008)	4.2 milliards USD
Employés	(sept. 2009)	environ 20 000
Serveurs/PC	(estimation)	plus de 500 000

L'entreprise Google

D'un projet d'étudiants à une entreprise mondiale :

- Fondée en 1998 par Sergey Brin et Larry Page.
- Depuis 2000 vente de publicités.
- Août 2004 lancement en bourse.

Valeur en bourse	(octobre 2009)	plus de 100 milliards USD
Chiffres d'affaires	(année 2008)	21.8 milliards USD
Résultat net	(année 2008)	4.2 milliards USD
Employés	(sept. 2009)	environ 20 000
Serveurs/PC	(estimation)	plus de 500 000

D. Vise, M. Malseed : *Google story*, Dunod, Paris, 2006

L'entreprise Google

D'un projet d'étudiants à une entreprise mondiale :

- Fondée en 1998 par Sergey Brin et Larry Page.
- Depuis 2000 vente de publicités.
- Août 2004 lancement en bourse.

Valeur en bourse	(octobre 2009)	plus de 100 milliards USD
Chiffres d'affaires	(année 2008)	21.8 milliards USD
Résultat net	(année 2008)	4.2 milliards USD
Employés	(sept. 2009)	environ 20 000
Serveurs/PC	(estimation)	plus de 500 000

📖 D. Vise, M. Malseed : *Google story*, Dunod, Paris, 2006
📖 Wikipédia, `http://fr.wikipedia.org/wiki/Google`

Que fait un moteur de recherche ?

Utilisateur

Moteur de recherche

Que fait un moteur de recherche ?

Utilisateur	Moteur de recherche
Requête : mots clés cherchés	

Que fait un moteur de recherche ?

Utilisateur		Moteur de recherche
Requête : mots clés cherchés	$\longrightarrow$	Liste de pages contenant les mots clés

Que fait un moteur de recherche ?

Utilisateur		Moteur de recherche
Requête : mots clés cherchés	$\longrightarrow$	Liste de pages contenant les mots clés
		$\downarrow$
Affichage de la liste des résultats	$\longleftarrow$	

Que fait un moteur de recherche ?

Utilisateur		Moteur de recherche
Requête : mots clés cherchés	$\longrightarrow$	Liste de pages contenant les mots clés
		$\downarrow$
Affichage de la liste des résultats	$\longleftarrow$	Tri par ordre de pertinence

Que fait un moteur de recherche ?

Utilisateur		Moteur de recherche
Requête : mots clés cherchés	⟶	Liste de pages contenant les mots clés
		↓
Affichage de la liste des résultats	⟵	Tri par ordre de pertinence

Ingrédients cruciaux :

Que fait un moteur de recherche ?

Utilisateur		Moteur de recherche
Requête : mots clés cherchés	⟶	Liste de pages contenant les mots clés
		↓
Affichage de la liste des résultats	⟵	Tri par ordre de pertinence

Ingrédients cruciaux :

1. Modélisation mathématique :
   Comment définir/calculer la pertinence d'une page web ?

Que fait un moteur de recherche ?

Utilisateur		Moteur de recherche
Requête : mots clés cherchés	$\longrightarrow$	Liste de pages contenant les mots clés
		$\downarrow$
Affichage de la liste des résultats	$\longleftarrow$	Tri par ordre de pertinence

Ingrédients cruciaux :

1. Modélisation mathématique :
   Comment définir/calculer la pertinence d'une page web ?
2. Traitement informatique :
   Comment stocker/traiter d'énormes quantités de données ?

Que fait un moteur de recherche ?

Utilisateur		Moteur de recherche
Requête : mots clés cherchés	$\longrightarrow$	Liste de pages contenant les mots clés
		$\downarrow$
Affichage de la liste des résultats	$\longleftarrow$	Tri par ordre de pertinence

Ingrédients cruciaux :

1. Modélisation mathématique :
   Comment définir/calculer la pertinence d’une page web ?
2. Traitement informatique :
   Comment stocker/traiter d’énormes quantités de données ?
3. Stratégie financière :
   Comment générer des bénéfices à partir d’un service gratuit ?

L’anarchie du web

Contenu hétérogène : toute sorte d’information,
mais peu structurée et peu hiérarchisée.

L’anarchie du web

Contenu hétérogène : toute sorte d’information,
mais peu structurée et peu hiérarchisée.

Contributions dispersées : Une multitude d’auteurs ajoutent constamment de nouvelles pages et modifient les pages existantes.

L’anarchie du web

Contenu hétérogène : toute sorte d’information,
mais peu structurée et peu hiérarchisée.

Contributions dispersées : Une multitude d’auteurs ajoutent constamment de nouvelles pages et modifient les pages existantes.

Syntaxe commune : hypertext markup language (HTML)

L’anarchie du web

Contenu hétérogène : toute sorte d’information,
mais peu structurée et peu hiérarchisée.

Contributions dispersées : Une multitude d’auteurs ajoutent constamment de nouvelles pages et modifient les pages existantes.

Syntaxe commune : hypertext markup language (HTML)

- Structuration logique (titres, sous-titres, paragraphes, . . .)

```
<h1> Le Titre </h1>
<p> Ceci est un paragraphe. </p>
```

L'anarchie du web

Contenu hétérogène : toute sorte d'information,
mais peu structurée et peu hiérarchisée.

Contributions dispersées : Une multitude d'auteurs ajoutent constamment de nouvelles pages et modifient les pages existantes.

Syntaxe commune : hypertext markup language (HTML)

- Structuration logique (titres, sous-titres, paragraphes, ...)

  ```
  <h1> Le Titre </h1>
  <p> Ceci est un paragraphe. </p>
  ```

- Apparence graphique (police, gras, cursif, couleur, ...)

  ```
  <b> Ceci est un texte en gras. </b>
  <i> Ceci est un texte cursif. </i>
  ```

L’anarchie du web

Contenu hétérogène : toute sorte d’information,
mais peu structurée et peu hiérarchisée.

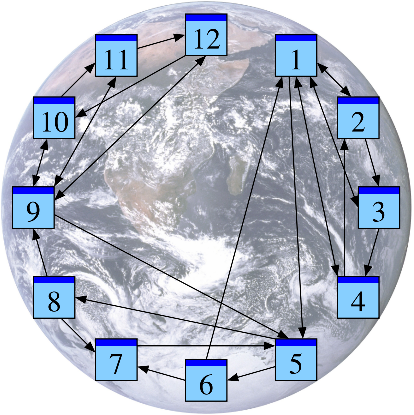
Contributions dispersées : Une multitude d’auteurs ajoutent constamment de nouvelles pages et modifient les pages existantes.

Syntaxe commune : hypertext markup language (HTML)

- Structuration logique (titres, sous-titres, paragraphes, ...)
```
<h1> Le Titre </h1>
<p> Ceci est un paragraphe. </p>
```
- Apparence graphique (police, gras, cursif, couleur, ...)
```
<b> Ceci est un texte en gras. </b>
<i> Ceci est un texte cursif. </i>
```
- Liens (références, citations entre les pages)
```
<a href="http://www.igt.uni-stuttgart.de/eiserm/">
Cliquer ici pour aller sur ma page web. </a>
```

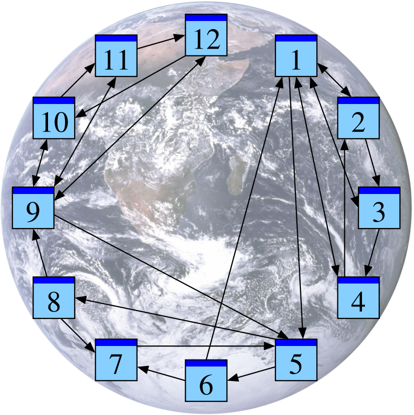
Le web est un graphe !

Exemple en miniature :



Le web est un graphe !

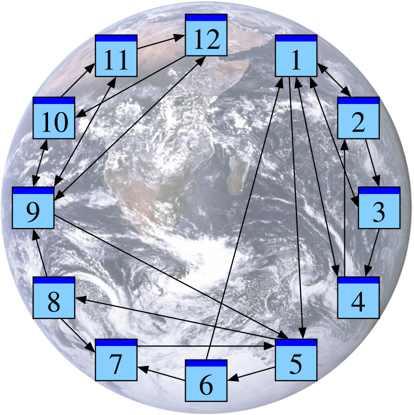
Exemple en miniature :



Notation : J'écris $j \to i$ pour un lien de la page P_j vers la page P_i.

Le web est un graphe !

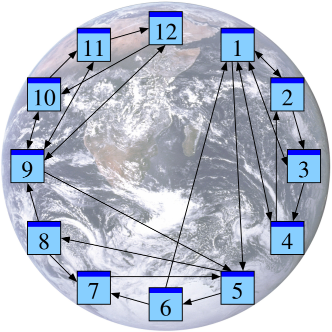
Exemple en miniature :



Ainsi notre graphe peut s'écrire comme $1 \to 2, 3, 4, 5;$ $2 \to 1, 3;$ $3 \to 1, 4;$ $4 \to 1, 2$; $5 \to 6, 8;$ $6 \to 1, 7;$ $7 \to 5;$ $8 \to 7, 9;$ $9 \to 5, 10, 11, 12$; $10 \to 9, 11;$ $11 \to 9, 12;$ $12 \to 9, 10$.

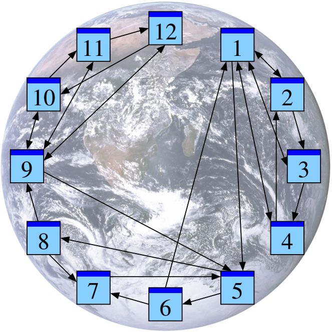
Comment exploiter ce graphe ?

Comment hiérarchiser notre graphe ?

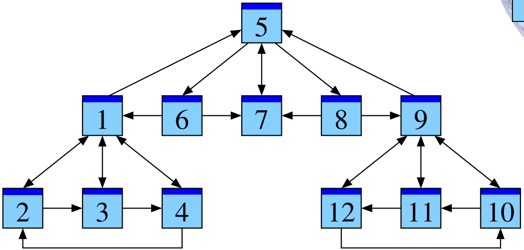


Comment exploiter ce graphe ?

Comment hiérarchiser notre graphe ?

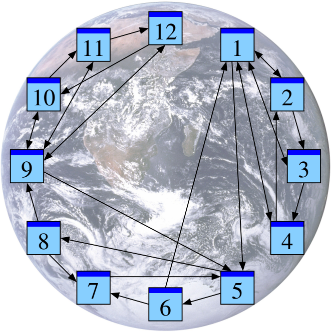


En voici une proposition ad hoc :

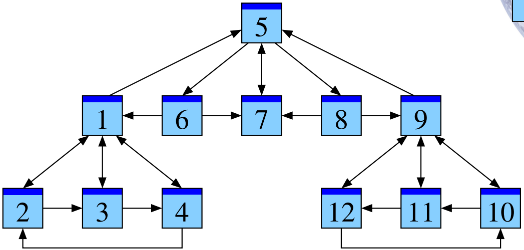


Comment exploiter ce graphe ?

Comment hiérarchiser notre graphe ?



En voici une proposition ad hoc :



Comment le faire en général ? Suivant quelles heuristiques ?

Plan de l'exposé

Premier modèle : comptage naïf

Heuristique : *Une page importante reçoit beaucoup de liens.*

Premier modèle : comptage naïf

Heuristique : *Une page importante reçoit beaucoup de liens.*

Avec un peu de naïveté on croira aussi la réciproque :
Si une page reçoit beaucoup de liens, alors elle est importante.

Premier modèle : comptage naïf

Heuristique : *Une page importante reçoit beaucoup de liens.*

Avec un peu de naïveté on croira aussi la réciproque :
Si une page reçoit beaucoup de liens, alors elle est importante.

Première tentative d'une définition mathématique :

$$m_i := \sum_{j \to i} 1.$$

Premier modèle : comptage naïf

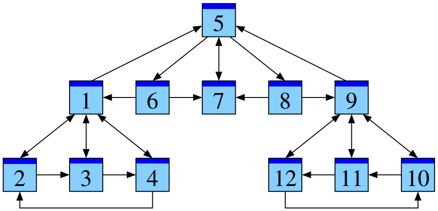
Heuristique : *Une page importante reçoit beaucoup de liens.*

Avec un peu de naïveté on croira aussi la réciproque :
Si une page reçoit beaucoup de liens, alors elle est importante.

Première tentative d'une définition mathématique :

$$m_i := \sum_{j \to i} 1.$$

Appliquons-la à notre exemple :



Premier modèle : comptage naïf

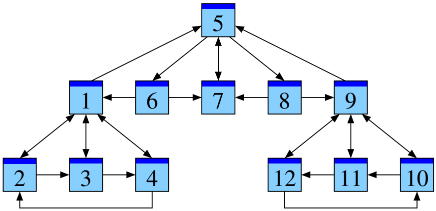
Heuristique : *Une page importante reçoit beaucoup de liens.*

Avec un peu de naïveté on croira aussi la réciproque :
Si une page reçoit beaucoup de liens, alors elle est importante.

Première tentative d'une définition mathématique :

$$m_i := \sum_{j \to i} 1.$$

Appliquons-la à notre exemple :



Ici on trouve $m_1 = m_9 = 4$ devant $m_5 = m_7 = 3$.

Second modèle : comptage pondéré

Heuristique : *Un lien* $j \to i$ *est un vote de la page* P_j *en faveur de* P_i*.*

Second modèle : comptage pondéré

Heuristique : *Un lien* $j \to i$ *est un vote de la page* P_j *en faveur de* P_i*.*

Supposons d'abord que toutes les pages ont un poids égal.

Second modèle : comptage pondéré

Heuristique : *Un lien* $j \to i$ *est un vote de la page* P_j *en faveur de* P_i*.*

Supposons d'abord que toutes les pages ont un poids égal.
Nous partageons le vote de la page P_j en ℓ_j parts égales.

Second modèle : comptage pondéré

Heuristique : *Un lien $j \to i$ est un vote de la page P_j en faveur de P_i.*

Supposons d'abord que toutes les pages ont un poids égal.
Nous partageons le vote de la page P_j en ℓ_j parts égales.

Seconde tentative d'une définition mathématique :

$$m_i := \sum_{j \to i} \frac{1}{\ell_j}.$$

Second modèle : comptage pondéré

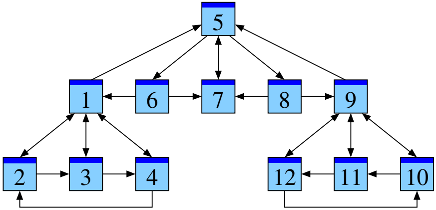
Heuristique : *Un lien $j \to i$ est un vote de la page P_j en faveur de P_i.*

Supposons d'abord que toutes les pages ont un poids égal.
Nous partageons le vote de la page P_j en ℓ_j parts égales.

Seconde tentative d'une définition mathématique :

$$m_i := \sum_{j \to i} \frac{1}{\ell_j}.$$

Appliquons-la à notre exemple :



Second modèle : comptage pondéré

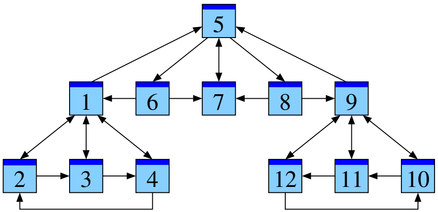
Heuristique : *Un lien $j \to i$ est un vote de la page P_j en faveur de P_i.*

Supposons d'abord que toutes les pages ont un poids égal.
Nous partageons le vote de la page P_j en ℓ_j parts égales.

Seconde tentative d'une définition mathématique :

$$m_i := \sum_{j \to i} \frac{1}{\ell_j}.$$

Appliquons-la à notre exemple :



Ici on trouve $m_1 = m_9 = 2$ devant $m_5 = {}^3\!/_2$ et $m_7 = {}^4\!/_3$.

Troisième modèle : comptage récursif

Heuristique :
Une page est importante si beaucoup de pages importantes la citent.

Troisième modèle : comptage récursif

Heuristique :
Une page est importante si beaucoup de pages importantes la citent.

Troisième tentative d'une définition mathématique :

$$m_i = \sum_{j \to i} \frac{1}{\ell_j} m_j.$$

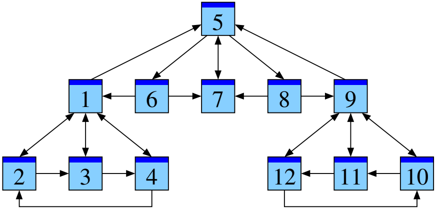
Troisième modèle : comptage récursif

Heuristique :
Une page est importante si beaucoup de pages importantes la citent.

Troisième tentative d'une définition mathématique :

$$m_i = \sum_{j \to i} \frac{1}{\ell_j} m_j.$$

Appliquons-la à notre exemple :



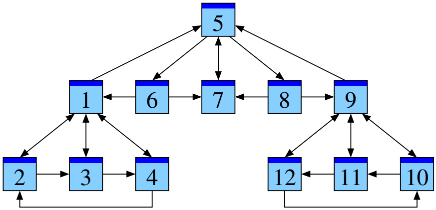
Troisième modèle : comptage récursif

Heuristique :
Une page est importante si beaucoup de pages importantes la citent.

Troisième tentative d'une définition mathématique :

$$m_i = \sum_{j \to i} \frac{1}{\ell_j} m_j.$$

Appliquons-la à notre exemple :



Ici on trouve $\quad m = \begin{array}{cccccccccccc} P_1 & P_2 & P_3 & P_4 & P_5 & P_6 & P_7 & P_8 & P_9 & P_{10} & P_{11} & P_{12} \\ (\ 2, & 1, & 1, & 1, & 3, & 1, & 2, & 1, & 2, & 1, & 1, & 1\ ). \end{array}$

Plan de l'exposé

1. Origines et motivations
2. Comment définir la pertinence d'une page web ?
3. Développement mathématique
   - Reformulations matricielle et probabiliste
   - Le modèle PageRank utilisé par Google
   - Le théorème du point fixe

Reformulation matricielle

Nous avons dégagé un système d'équations linéaires :

$$m_i = \sum_{j \to i} \frac{1}{\ell_j} m_j.$$

Reformulation matricielle

Nous avons dégagé un système d'équations linéaires :

$$m_i = \sum_{j \to i} \frac{1}{\ell_j} m_j.$$

Définissons alors la matrice $A = (a_{ij})$ par

$$a_{ij} := \begin{cases} \frac{1}{\ell_j} & \text{si } j \to i, \\ 0 & \text{sinon.} \end{cases}$$

Reformulation matricielle

Nous avons dégagé un système d'équations linéaires :

$$m_i = \sum_{j \to i} \frac{1}{\ell_j} m_j.$$

Définissons alors la matrice $A = (a_{ij})$ par

$$a_{ij} := \begin{cases} \frac{1}{\ell_j} & \text{si } j \to i, \\ 0 & \text{sinon.} \end{cases}$$

Ainsi notre équation s'écrit

$$Am = m$$

Reformulation matricielle

Nous avons dégagé un système d'équations linéaires :

$$m_i = \sum_{j \to i} \frac{1}{\ell_j} m_j.$$

Définissons alors la matrice $A = (a_{ij})$ par

$$a_{ij} := \begin{cases} \frac{1}{\ell_j} & \text{si } j \to i, \\ 0 & \text{sinon.} \end{cases}$$

Ainsi notre équation s'écrit

$$Am = m$$

ou encore

$$(A - \mathrm{Id})m = 0.$$

Reformulation matricielle

Nous avons dégagé un système d'équations linéaires :

$$m_i = \sum_{j \to i} \frac{1}{\ell_j} m_j.$$

Définissons alors la matrice $A = (a_{ij})$ par

$$a_{ij} := \begin{cases} \frac{1}{\ell_j} & \text{si } j \to i, \\ 0 & \text{sinon.} \end{cases}$$

Ainsi notre équation s'écrit

$$Am = m$$

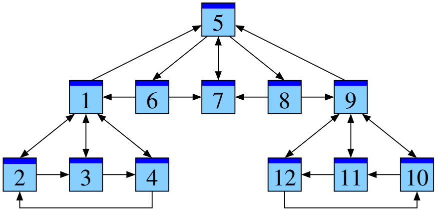
ou encore

$$(A - \mathrm{Id})m = 0.$$

Vive l'algèbre linéaire !

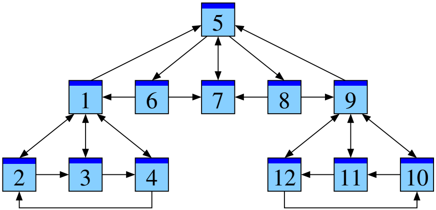
Reformulation matricielle de notre exemple

Notre graphe :



Reformulation matricielle de notre exemple

Notre graphe :

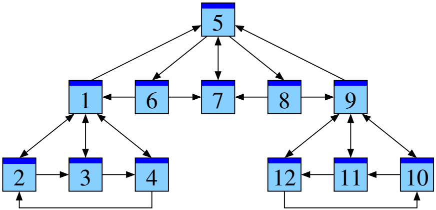


Et sa matrice :

$$A = \begin{pmatrix}
0 & 1/2 & 1/2 & 1/2 & 0 & 1/2 & 0 & 0 & 0 & 0 & 0 & 0 \\
1/4 & 0 & 0 & 1/2 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
1/4 & 1/2 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
1/4 & 0 & 1/2 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
1/4 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 1/4 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 1/3 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 1/3 & 1/2 & 0 & 1/2 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 1/3 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 1/2 & 0 & 1/2 & 1/2 & 1/2 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1/4 & 0 & 0 & 1/2 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1/4 & 1/2 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1/4 & 0 & 1/2 & 0
\end{pmatrix}.$$

Reformulation matricielle de notre exemple

Notre graphe :



Et sa matrice :

$$A = \begin{pmatrix}
0 & 1/2 & 1/2 & 1/2 & 0 & 1/2 & 0 & 0 & 0 & 0 & 0 & 0 \\
1/4 & 0 & 0 & 1/2 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
1/4 & 1/2 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
1/4 & 0 & 1/2 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
1/4 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 1/4 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 1/3 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 1/3 & 1/2 & 0 & 1/2 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 1/3 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 1/2 & 0 & 1/2 & 1/2 & 1/2 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1/4 & 0 & 0 & 1/2 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1/4 & 1/2 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1/4 & 0 & 1/2 & 0
\end{pmatrix}.$$

L'équation $Am = m$ admet comme solution

$$m = (\ 2,\ 1,\ 1,\ 1,\ 3,\ 1,\ 2,\ 1,\ 2,\ 1,\ 1,\ 1\ )^t.$$

Promenade aléatoire sur le web

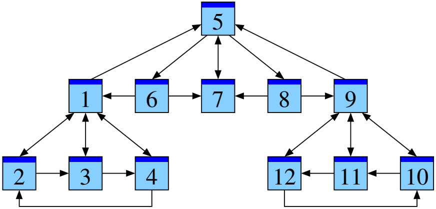
Promenade aléatoire sur le web

Imaginons un « surfeur aléatoire » qui se balade sur internet en cliquant sur les liens au hasard. Comment évolue sa position ?

Promenade aléatoire sur le web

Imaginons un « surfeur aléatoire » qui se balade sur internet en cliquant sur les liens au hasard. Comment évolue sa position ?

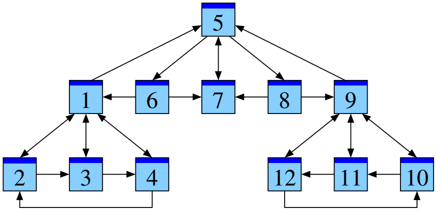
Notre graphe :



Promenade aléatoire sur le web

Imaginons un « surfeur aléatoire » qui se balade sur internet en cliquant sur les liens au hasard. Comment évolue sa position ?

Notre graphe :

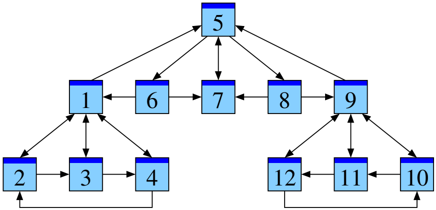


Évolution des probabilités en partant de la page P_7 :

Promenade aléatoire sur le web

Imaginons un « surfeur aléatoire » qui se balade sur internet en cliquant sur les liens au hasard. Comment évolue sa position ?

Notre graphe :



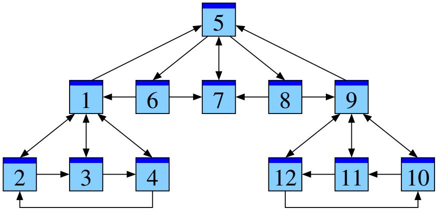
Évolution des probabilités en partant de la page P_7 :

	P_1	P_2	P_3	P_4	P_5	P_6	P_7	P_8	P_9	P_{10}	P_{11}	P_{12}
$t=0$	.000	.000	.000	.000	.000	.000	1.00	.000	.000	.000	.000	.000

Promenade aléatoire sur le web

Imaginons un « surfeur aléatoire » qui se balade sur internet en cliquant sur les liens au hasard. Comment évolue sa position ?

Notre graphe :



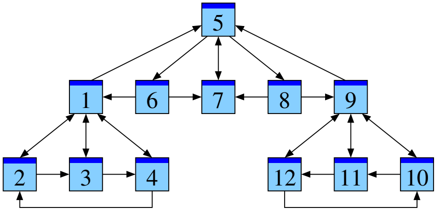
Évolution des probabilités en partant de la page P_7 :

	P_1	P_2	P_3	P_4	P_5	P_6	P_7	P_8	P_9	P_{10}	P_{11}	P_{12}
$t=0$	.000	.000	.000	.000	.000	.000	1.00	.000	.000	.000	.000	.000
$t=1$	.000	.000	.000	.000	1.00	.000	.000	.000	.000	.000	.000	.000

Promenade aléatoire sur le web

Imaginons un « surfeur aléatoire » qui se balade sur internet en cliquant sur les liens au hasard. Comment évolue sa position ?

Notre graphe :



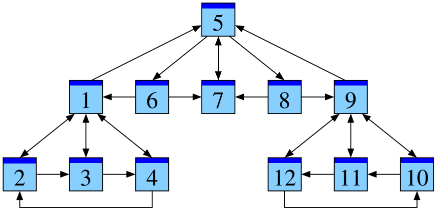
Évolution des probabilités en partant de la page P_7 :

	P_1	P_2	P_3	P_4	P_5	P_6	P_7	P_8	P_9	P_{10}	P_{11}	P_{12}
$t=0$	.000	.000	.000	.000	.000	.000	1.00	.000	.000	.000	.000	.000
$t=1$	.000	.000	.000	.000	1.00	.000	.000	.000	.000	.000	.000	.000
$t=2$	.000	.000	.000	.000	.000	.333	.333	.333	.000	.000	.000	.000

Promenade aléatoire sur le web

Imaginons un « surfeur aléatoire » qui se balade sur internet en cliquant sur les liens au hasard. Comment évolue sa position ?

Notre graphe :



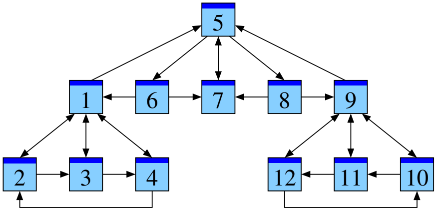
Évolution des probabilités en partant de la page P_7 :

	P_1	P_2	P_3	P_4	P_5	P_6	P_7	P_8	P_9	P_{10}	P_{11}	P_{12}
$t=0$	.000	.000	.000	.000	.000	.000	1.00	.000	.000	.000	.000	.000
$t=1$	.000	.000	.000	.000	1.00	.000	.000	.000	.000	.000	.000	.000
$t=2$	.000	.000	.000	.000	.000	.333	.333	.333	.000	.000	.000	.000
$t=3$	.167	.000	.000	.000	.333	.000	.333	.000	.167	.000	.000	.000

Promenade aléatoire sur le web

Imaginons un « surfeur aléatoire » qui se balade sur internet en cliquant sur les liens au hasard. Comment évolue sa position ?

Notre graphe :



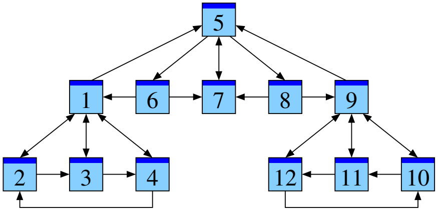
Évolution des probabilités en partant de la page P_7 :

	P_1	P_2	P_3	P_4	P_5	P_6	P_7	P_8	P_9	P_{10}	P_{11}	P_{12}
$t=0$	.000	.000	.000	.000	.000	.000	1.00	.000	.000	.000	.000	.000
$t=1$	.000	.000	.000	.000	1.00	.000	.000	.000	.000	.000	.000	.000
$t=2$	.000	.000	.000	.000	.000	.333	.333	.333	.000	.000	.000	.000
$t=3$	.167	.000	.000	.000	.333	.000	.333	.000	.167	.000	.000	.000
$t=4$	.000	.042	.042	.042	.417	.111	.111	.111	.000	.042	.042	.042

Promenade aléatoire sur le web

Imaginons un « surfeur aléatoire » qui se balade sur internet en cliquant sur les liens au hasard. Comment évolue sa position ?

Notre graphe :



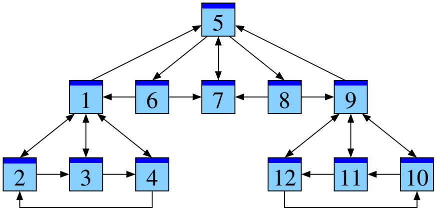
Évolution des probabilités en partant de la page P_7 :

	P_1	P_2	P_3	P_4	P_5	P_6	P_7	P_8	P_9	P_{10}	P_{11}	P_{12}
$t=0$	.000	.000	.000	.000	.000	.000	1.00	.000	.000	.000	.000	.000
$t=1$	.000	.000	.000	.000	1.00	.000	.000	.000	.000	.000	.000	.000
$t=2$	.000	.000	.000	.000	.000	.333	.333	.333	.000	.000	.000	.000
$t=3$	.167	.000	.000	.000	.333	.000	.333	.000	.167	.000	.000	.000
$t=4$	.000	.042	.042	.042	.417	.111	.111	.111	.000	.042	.042	.042
$t=5$	.118	.021	.021	.021	.111	.139	.250	.139	.118	.021	.021	.021
...												

Promenade aléatoire sur le web

Imaginons un « surfeur aléatoire » qui se balade sur internet en cliquant sur les liens au hasard. Comment évolue sa position ?

Notre graphe :



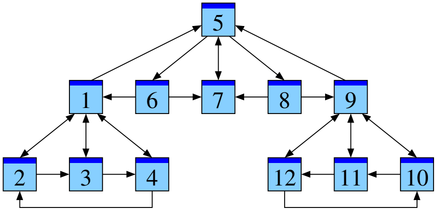
Évolution des probabilités en partant de la page P_7 :

	P_1	P_2	P_3	P_4	P_5	P_6	P_7	P_8	P_9	P_{10}	P_{11}	P_{12}
t=0	.000	.000	.000	.000	.000	.000	1.00	.000	.000	.000	.000	.000
t=1	.000	.000	.000	.000	1.00	.000	.000	.000	.000	.000	.000	.000
t=2	.000	.000	.000	.000	.000	.333	.333	.333	.000	.000	.000	.000
t=3	.167	.000	.000	.000	.333	.000	.333	.000	.167	.000	.000	.000
t=4	.000	.042	.042	.042	.417	.111	.111	.111	.000	.042	.042	.042
t=5	.118	.021	.021	.021	.111	.139	.250	.139	.118	.021	.021	.021
...												
t=29	.117	.059	.059	.059	.177	.059	.117	.059	.117	.059	.059	.059

Promenade aléatoire sur le web

Imaginons un « surfeur aléatoire » qui se balade sur internet en cliquant sur les liens au hasard. Comment évolue sa position ?

Notre graphe :



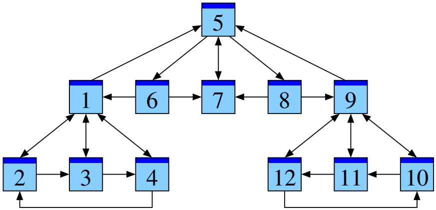
Évolution des probabilités en partant de la page P_7 :

	P_1	P_2	P_3	P_4	P_5	P_6	P_7	P_8	P_9	P_{10}	P_{11}	P_{12}
$t=0$	.000	.000	.000	.000	.000	.000	1.00	.000	.000	.000	.000	.000
$t=1$	.000	.000	.000	.000	1.00	.000	.000	.000	.000	.000	.000	.000
$t=2$	.000	.000	.000	.000	.000	.333	.333	.333	.000	.000	.000	.000
$t=3$	.167	.000	.000	.000	.333	.000	.333	.000	.167	.000	.000	.000
$t=4$	.000	.042	.042	.042	.417	.111	.111	.111	.000	.042	.042	.042
$t=5$	.118	.021	.021	.021	.111	.139	.250	.139	.118	.021	.021	.021
...												
$t=29$	.117	.059	.059	.059	.177	.059	.117	.059	.117	.059	.059	.059
$t=30$	.117	.059	.059	.059	.177	.059	.117	.059	.117	.059	.059	.059

Promenade aléatoire sur le web

Imaginons un « surfeur aléatoire » qui se balade sur internet en cliquant sur les liens au hasard. Comment évolue sa position ?

Notre graphe :



Évolution des probabilités en partant de la page P_7 :

	P_1	P_2	P_3	P_4	P_5	P_6	P_7	P_8	P_9	P_{10}	P_{11}	P_{12}
$t=0$	.000	.000	.000	.000	.000	.000	1.00	.000	.000	.000	.000	.000
$t=1$	.000	.000	.000	.000	1.00	.000	.000	.000	.000	.000	.000	.000
$t=2$	.000	.000	.000	.000	.000	.333	.333	.333	.000	.000	.000	.000
$t=3$	.167	.000	.000	.000	.333	.000	.333	.000	.167	.000	.000	.000
$t=4$	.000	.042	.042	.042	.417	.111	.111	.111	.000	.042	.042	.042
$t=5$	.118	.021	.021	.021	.111	.139	.250	.139	.118	.021	.021	.021
...												
$t=29$	.117	.059	.059	.059	.177	.059	.117	.059	.117	.059	.059	.059
$t=30$	.117	.059	.059	.059	.177	.059	.117	.059	.117	.059	.059	.059

Cette diffusion converge vers une distribution stationnaire !

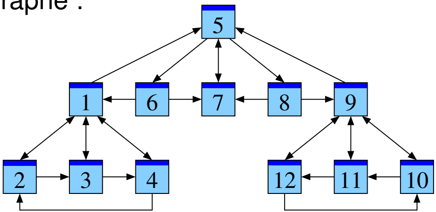
Promenade aléatoire sur le web

Vérifions notre observation par un second exemple, partant de P_1.

Promenade aléatoire sur le web

Vérifions notre observation par un second exemple, partant de P_1.

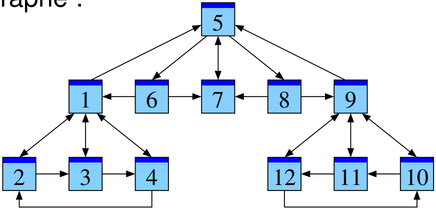
Toujours notre graphe :



Promenade aléatoire sur le web

Vérifions notre observation par un second exemple, partant de P_1.

Toujours notre graphe :

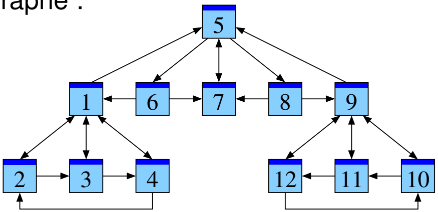


Évolution des probabilités en partant de la page P_1 :

Promenade aléatoire sur le web

Vérifions notre observation par un second exemple, partant de P_1.

Toujours notre graphe :



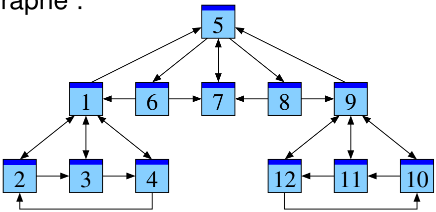
Évolution des probabilités en partant de la page P_1 :

	P_1	P_2	P_3	P_4	P_5	P_6	P_7	P_8	P_9	P_{10}	P_{11}	P_{12}
$t{=}0$	1.00	$.000$	$.000$	$.000$	$.000$	$.000$	$.000$	$.000$	$.000$	$.000$	$.000$	$.000$

Promenade aléatoire sur le web

Vérifions notre observation par un second exemple, partant de P_1.

Toujours notre graphe :



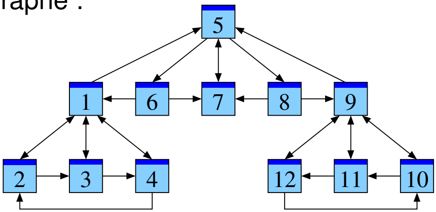
Évolution des probabilités en partant de la page P_1 :

	P_1	P_2	P_3	P_4	P_5	P_6	P_7	P_8	P_9	P_{10}	P_{11}	P_{12}
t=0	1.00	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000
t=1	.000	.250	.250	.250	.250	.000	.000	.000	.000	.000	.000	.000

Promenade aléatoire sur le web

Vérifions notre observation par un second exemple, partant de P_1.

Toujours notre graphe :



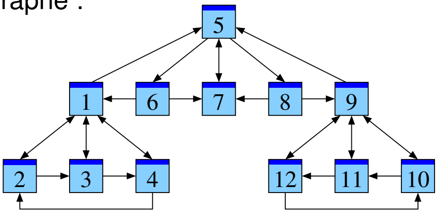
Évolution des probabilités en partant de la page P_1 :

	P_1	P_2	P_3	P_4	P_5	P_6	P_7	P_8	P_9	P_{10}	P_{11}	P_{12}
$t{=}0$	1.00	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000
$t{=}1$	.000	.250	.250	.250	.250	.000	.000	.000	.000	.000	.000	.000
$t{=}2$	.375	.125	.125	.125	.000	.083	.083	.083	.000	.000	.000	.000

Promenade aléatoire sur le web

Vérifions notre observation par un second exemple, partant de P_1.

Toujours notre graphe :



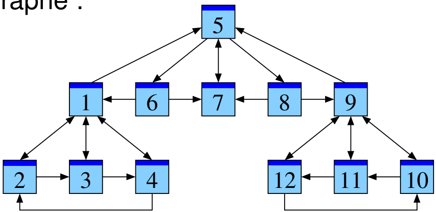
Évolution des probabilités en partant de la page P_1 :

	P_1	P_2	P_3	P_4	P_5	P_6	P_7	P_8	P_9	P_{10}	P_{11}	P_{12}
$t=0$	1.00	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000
$t=1$	.000	.250	.250	.250	.250	.000	.000	.000	.000	.000	.000	.000
$t=2$	.375	.125	.125	.125	.000	.083	.083	.083	.000	.000	.000	.000
$t=3$	.229	.156	.156	.156	.177	.000	.083	.000	.042	.000	.000	.000

Promenade aléatoire sur le web

Vérifions notre observation par un second exemple, partant de P_1.

Toujours notre graphe :



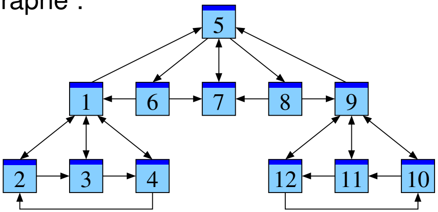
Évolution des probabilités en partant de la page P_1 :

	P_1	P_2	P_3	P_4	P_5	P_6	P_7	P_8	P_9	P_{10}	P_{11}	P_{12}
$t=0$	1.00	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000
$t=1$	.000	.250	.250	.250	.250	.000	.000	.000	.000	.000	.000	.000
$t=2$	.375	.125	.125	.125	.000	.083	.083	.083	.000	.000	.000	.000
$t=3$	.229	.156	.156	.156	.177	.000	.083	.000	.042	.000	.000	.000
$t=4$	.234	.135	.135	.135	.151	.059	.059	.059	.000	.010	.010	.010

Promenade aléatoire sur le web

Vérifions notre observation par un second exemple, partant de P_1.

Toujours notre graphe :



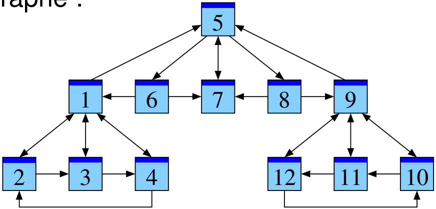
Évolution des probabilités en partant de la page P_1 :

	P_1	P_2	P_3	P_4	P_5	P_6	P_7	P_8	P_9	P_{10}	P_{11}	P_{12}
$t{=}0$	1.00	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000
$t{=}1$	.000	.250	.250	.250	.250	.000	.000	.000	.000	.000	.000	.000
$t{=}2$	.375	.125	.125	.125	.000	.083	.083	.083	.000	.000	.000	.000
$t{=}3$	.229	.156	.156	.156	.177	.000	.083	.000	.042	.000	.000	.000
$t{=}4$	.234	.135	.135	.135	.151	.059	.059	.059	.000	.010	.010	.010
$t{=}5$	.233	.126	.126	.126	.118	.050	.109	.050	.045	.005	.005	.005
...												

Promenade aléatoire sur le web

Vérifions notre observation par un second exemple, partant de P_1.

Toujours notre graphe :



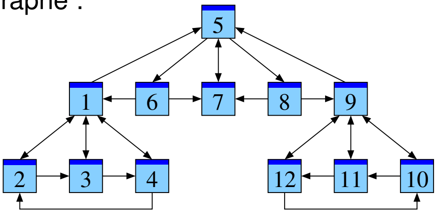
Évolution des probabilités en partant de la page P_1 :

	P_1	P_2	P_3	P_4	P_5	P_6	P_7	P_8	P_9	P_{10}	P_{11}	P_{12}
t=0	1.00	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000
t=1	.000	.250	.250	.250	.250	.000	.000	.000	.000	.000	.000	.000
t=2	.375	.125	.125	.125	.000	.083	.083	.083	.000	.000	.000	.000
t=3	.229	.156	.156	.156	.177	.000	.083	.000	.042	.000	.000	.000
t=4	.234	.135	.135	.135	.151	.059	.059	.059	.000	.010	.010	.010
t=5	.233	.126	.126	.126	.118	.050	.109	.050	.045	.005	.005	.005
...												
t=69	.117	.059	.059	.059	.177	.059	.117	.059	.117	.059	.059	.059

Promenade aléatoire sur le web

Vérifions notre observation par un second exemple, partant de P_1.

Toujours notre graphe :



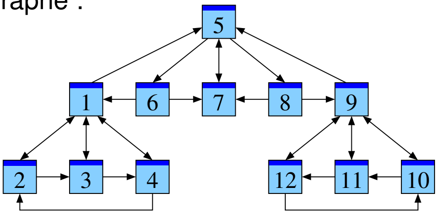
Évolution des probabilités en partant de la page P_1 :

	P_1	P_2	P_3	P_4	P_5	P_6	P_7	P_8	P_9	P_{10}	P_{11}	P_{12}
t=0	1.00	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000
t=1	.000	.250	.250	.250	.250	.000	.000	.000	.000	.000	.000	.000
t=2	.375	.125	.125	.125	.000	.083	.083	.083	.000	.000	.000	.000
t=3	.229	.156	.156	.156	.177	.000	.083	.000	.042	.000	.000	.000
t=4	.234	.135	.135	.135	.151	.059	.059	.059	.000	.010	.010	.010
t=5	.233	.126	.126	.126	.118	.050	.109	.050	.045	.005	.005	.005
...												
t=69	.117	.059	.059	.059	.177	.059	.117	.059	.117	.059	.059	.059
t=70	.117	.059	.059	.059	.177	.059	.117	.059	.117	.059	.059	.059

Promenade aléatoire sur le web

Vérifions notre observation par un second exemple, partant de P_1.

Toujours notre graphe :



Évolution des probabilités en partant de la page P_1 :

	P_1	P_2	P_3	P_4	P_5	P_6	P_7	P_8	P_9	P_{10}	P_{11}	P_{12}
$t=0$	1.00	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000
$t=1$	.000	.250	.250	.250	.250	.000	.000	.000	.000	.000	.000	.000
$t=2$	.375	.125	.125	.125	.000	.083	.083	.083	.000	.000	.000	.000
$t=3$	.229	.156	.156	.156	.177	.000	.083	.000	.042	.000	.000	.000
$t=4$	.234	.135	.135	.135	.151	.059	.059	.059	.000	.010	.010	.010
$t=5$	.233	.126	.126	.126	.118	.050	.109	.050	.045	.005	.005	.005
...												
$t=69$	.117	.059	.059	.059	.177	.059	.117	.059	.117	.059	.059	.059
$t=70$	.117	.059	.059	.059	.177	.059	.117	.059	.117	.059	.059	.059

La mesure stationnaire est la même !

La loi de transition

Comment formaliser cette diffusion des probabilités ?

La loi de transition

Comment formaliser cette diffusion des probabilités ?

Au temps t notre surfeur se trouve sur la page P_j avec probabilité p_j.

La loi de transition

Comment formaliser cette diffusion des probabilités ?

Au temps t notre surfeur se trouve sur la page P_j avec probabilité p_j.
La probabilité de partir de P_j et de suivre le lien $j \to i$ est alors $\frac{1}{\ell_j} p_j$.

La loi de transition

Comment formaliser cette diffusion des probabilités ?

Au temps t notre surfeur se trouve sur la page P_j avec probabilité p_j.
La probabilité de partir de P_j et de suivre le lien $j \to i$ est alors $\frac{1}{\ell_j} p_j$.
La probabilité d'arriver au temps $t+1$ sur la page P_i est donc

$$p'_i := \sum_{j \to i} \frac{1}{\ell_j} p_j.$$

La loi de transition

Comment formaliser cette diffusion des probabilités ?

Au temps t notre surfeur se trouve sur la page P_j avec probabilité p_j.
La probabilité de partir de P_j et de suivre le lien $j \to i$ est alors $\frac{1}{\ell_j} p_j$.
La probabilité d'arriver au temps $t+1$ sur la page P_i est donc

$$p'_i := \sum_{j \to i} \frac{1}{\ell_j} p_j .$$

Cette loi de transition définit la distribution suivante, notée $p' = T(p)$.

La loi de transition

Comment formaliser cette diffusion des probabilités ?

Au temps t notre surfeur se trouve sur la page P_j avec probabilité p_j.
La probabilité de partir de P_j et de suivre le lien $j \to i$ est alors $\frac{1}{\ell_j} p_j$.
La probabilité d'arriver au temps $t+1$ sur la page P_i est donc

$$p'_i := \sum_{j \to i} \frac{1}{\ell_j} p_j .$$

Cette loi de transition définit la distribution suivante, notée $p' = T(p)$.
Une mesure stationnaire est caractérisée par l'équation d'équilibre

$$m = T(m) \quad \text{c'est-à-dire} \quad m_i = \sum_{j \to i} \frac{1}{\ell_j} m_j .$$

La loi de transition

Comment formaliser cette diffusion des probabilités ?

Au temps t notre surfeur se trouve sur la page P_j avec probabilité p_j.
La probabilité de partir de P_j et de suivre le lien $j \to i$ est alors $\frac{1}{\ell_j} p_j$.
La probabilité d'arriver au temps $t+1$ sur la page P_i est donc

$$p'_i := \sum_{j \to i} \frac{1}{\ell_j} p_j .$$

Cette loi de transition définit la distribution suivante, notée $p' = T(p)$.
Une mesure stationnaire est caractérisée par l'équation d'équilibre

$$m = T(m) \quad \text{c'est-à-dire} \quad m_i = \sum_{j \to i} \frac{1}{\ell_j} m_j .$$

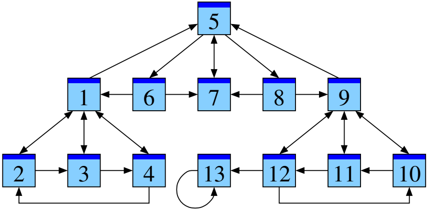
Vive la théorie des probabilités !

Attention aux trous noirs !

Que se passe-t-il quand notre graphe contient une page sans issue ?

Attention aux trous noirs !

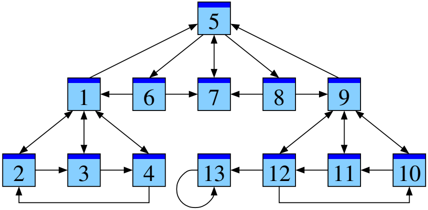
Que se passe-t-il quand notre graphe contient une page sans issue ?



Notre surfeur aléatoire tombera tôt ou tard sur la page P_{13},
où il demeure pour le reste de sa vie.

Attention aux trous noirs !

Que se passe-t-il quand notre graphe contient une page sans issue ?



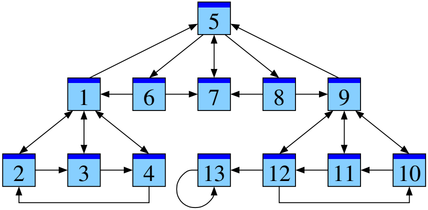
Notre surfeur aléatoire tombera tôt ou tard sur la page P_{13},
où il demeure pour le reste de sa vie.

La seule mesure stationnaire est

$$m = (0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 1).$$

Attention aux trous noirs !

Que se passe-t-il quand notre graphe contient une page sans issue ?



Notre surfeur aléatoire tombera tôt ou tard sur la page P_{13}, où il demeure pour le reste de sa vie.

La seule mesure stationnaire est

$$m = (0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 1).$$

Dans ce cas notre modèle n'est pas très réaliste !

Le modèle PageRank utilisé par Google

Pour échapper aux trous noirs, Google utilise un modèle plus raffiné :

Le modèle PageRank utilisé par Google

Pour échapper aux trous noirs, Google utilise un modèle plus raffiné :

1. Téléportation :

   Avec une probabilité fixée $c \in [0, 1]$ le surfeur abandonne sa page actuelle P_j et recommence sur une des n pages du web.

Le modèle PageRank utilisé par Google

Pour échapper aux trous noirs, Google utilise un modèle plus raffiné :

1. Téléportation :

   Avec une probabilité fixée $c \in [0,1]$ le surfeur abandonne sa page actuelle P_j et recommence sur une des n pages du web.

2. Promenade aléatoire :

   Sinon, avec probabilité $1-c$, le surfeur suit un des liens de la page P_j, choisi de manière équiprobable (comme avant).

Le modèle PageRank utilisé par Google

Pour échapper aux trous noirs, Google utilise un modèle plus raffiné :

1. Téléportation :

   Avec une probabilité fixée $c \in [0,1]$ le surfeur abandonne sa page actuelle P_j et recommence sur une des n pages du web.

2. Promenade aléatoire :

   Sinon, avec probabilité $1-c$, le surfeur suit un des liens de la page P_j, choisi de manière équiprobable (comme avant).

Dans ce modèle la transition est donnée par

$$p'_i := \frac{c}{n} + \sum_{j \to i} \frac{1-c}{\ell_j} p_j .$$

Le modèle PageRank utilisé par Google

Pour échapper aux trous noirs, Google utilise un modèle plus raffiné :

1. Téléportation :

   Avec une probabilité fixée $c \in [0, 1]$ le surfeur abandonne sa page actuelle P_j et recommence sur une des n pages du web.

2. Promenade aléatoire :

   Sinon, avec probabilité $1 - c$, le surfeur suit un des liens de la page P_j, choisi de manière équiprobable (comme avant).

Dans ce modèle la transition est donnée par

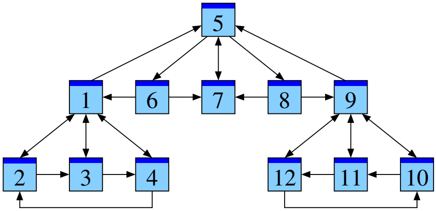
$$p'_i := \frac{c}{n} + \sum_{j \to i} \frac{1-c}{\ell_j} p_j.$$

La mesure d'équilibre vérifie donc

$$m_i = \frac{c}{n} + \sum_{j \to i} \frac{1-c}{\ell_j} m_j.$$

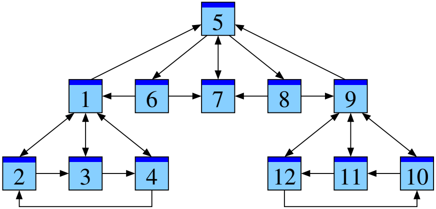
Application à notre exemple

Notre graphe :



Application à notre exemple

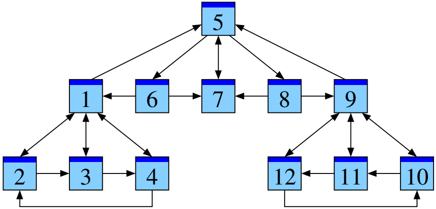
Notre graphe :



Évolution des probabilités en partant de la page P_1, avec $c = 0.15$:

Application à notre exemple

Notre graphe :

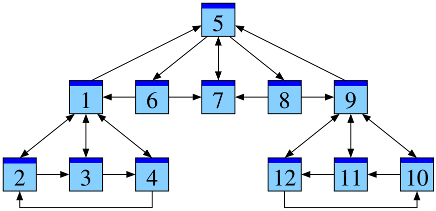


Évolution des probabilités en partant de la page P_1, avec $c = 0.15$:

	P_1	P_2	P_3	P_4	P_5	P_6	P_7	P_8	P_9	P_{10}	P_{11}	P_{12}
$t{=}0$	1.00	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000

Application à notre exemple

Notre graphe :

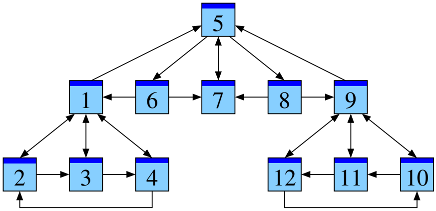


Évolution des probabilités en partant de la page P_1, avec $c = 0.15$:

	P_1	P_2	P_3	P_4	P_5	P_6	P_7	P_8	P_9	P_{10}	P_{11}	P_{12}
$t{=}0$	1.00	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000
$t{=}1$	.013	.225	.225	.225	.225	.013	.013	.013	.013	.013	.013	.013

Application à notre exemple

Notre graphe :

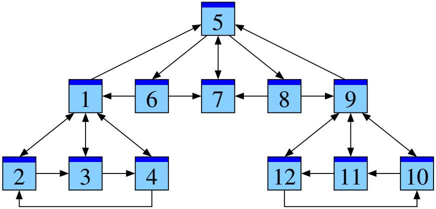


Évolution des probabilités en partant de la page P_1, avec $c = 0.15$:

	P_1	P_2	P_3	P_4	P_5	P_6	P_7	P_8	P_9	P_{10}	P_{11}	P_{12}
t=0	1.00	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000
t=1	.013	.225	.225	.225	.225	.013	.013	.013	.013	.013	.013	.013
t=2	.305	.111	.111	.111	.028	.076	.087	.076	.034	.020	.020	.020

Application à notre exemple

Notre graphe :

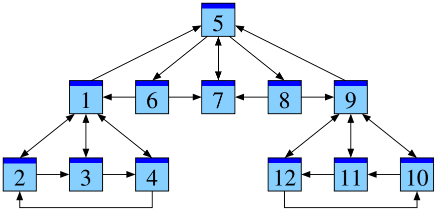


Évolution des probabilités en partant de la page P_1, avec $c = 0.15$:

	P_1	P_2	P_3	P_4	P_5	P_6	P_7	P_8	P_9	P_{10}	P_{11}	P_{12}
$t=0$	1.00	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000
$t=1$	.013	.225	.225	.225	.225	.013	.013	.013	.013	.013	.013	.013
$t=2$	.305	.111	.111	.111	.028	.076	.087	.076	.034	.020	.020	.020
$t=3$	.186	.124	.124	.124	.158	.021	.085	.021	.071	.028	.028	.028

Application à notre exemple

Notre graphe :

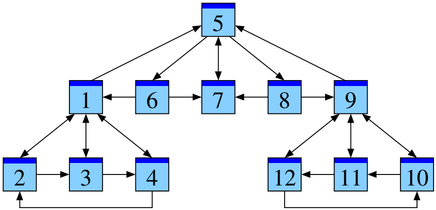


Évolution des probabilités en partant de la page P_1, avec $c = 0.15$:

	P_1	P_2	P_3	P_4	P_5	P_6	P_7	P_8	P_9	P_{10}	P_{11}	P_{12}
t=0	1.00	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000
t=1	.013	.225	.225	.225	.225	.013	.013	.013	.013	.013	.013	.013
t=2	.305	.111	.111	.111	.028	.076	.087	.076	.034	.020	.020	.020
t=3	.186	.124	.124	.124	.158	.021	.085	.021	.071	.028	.028	.028
t=4	.180	.105	.105	.105	.140	.057	.075	.057	.057	.040	.040	.040

Application à notre exemple

Notre graphe :

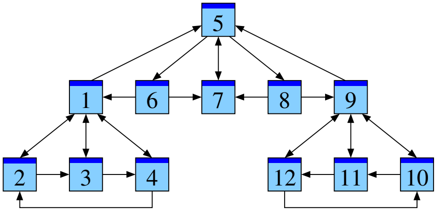


Évolution des probabilités en partant de la page P_1, avec $c = 0.15$:

	P_1	P_2	P_3	P_4	P_5	P_6	P_7	P_8	P_9	P_{10}	P_{11}	P_{12}
$t{=}0$	1.00	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000
$t{=}1$	.013	.225	.225	.225	.225	.013	.013	.013	.013	.013	.013	.013
$t{=}2$	.305	.111	.111	.111	.028	.076	.087	.076	.034	.020	.020	.020
$t{=}3$	.186	.124	.124	.124	.158	.021	.085	.021	.071	.028	.028	.028
$t{=}4$	.180	.105	.105	.105	.140	.057	.075	.057	.057	.040	.040	.040
$t{=}5$	.171	.095	.095	.095	.126	.052	.101	.052	.087	.042	.042	.042
...												

Application à notre exemple

Notre graphe :

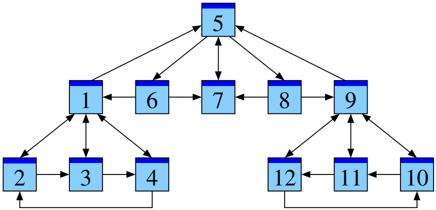


Évolution des probabilités en partant de la page P_1, avec $c = 0.15$:

	P_1	P_2	P_3	P_4	P_5	P_6	P_7	P_8	P_9	P_{10}	P_{11}	P_{12}
$t{=}0$	1.00	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000
$t{=}1$	.013	.225	.225	.225	.225	.013	.013	.013	.013	.013	.013	.013
$t{=}2$	.305	.111	.111	.111	.028	.076	.087	.076	.034	.020	.020	.020
$t{=}3$	.186	.124	.124	.124	.158	.021	.085	.021	.071	.028	.028	.028
$t{=}4$	.180	.105	.105	.105	.140	.057	.075	.057	.057	.040	.040	.040
$t{=}5$	.171	.095	.095	.095	.126	.052	.101	.052	.087	.042	.042	.042
...												
$t{=}29$	.120	.066	.066	.066	.150	.055	.102	.055	.120	.066	.066	.066

Application à notre exemple

Notre graphe :



Évolution des probabilités en partant de la page P_1, avec $c = 0.15$:

	P_1	P_2	P_3	P_4	P_5	P_6	P_7	P_8	P_9	P_{10}	P_{11}	P_{12}
$t=0$	1.00	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000
$t=1$	.013	.225	.225	.225	.225	.013	.013	.013	.013	.013	.013	.013
$t=2$	.305	.111	.111	.111	.028	.076	.087	.076	.034	.020	.020	.020
$t=3$	.186	.124	.124	.124	.158	.021	.085	.021	.071	.028	.028	.028
$t=4$	.180	.105	.105	.105	.140	.057	.075	.057	.057	.040	.040	.040
$t=5$	.171	.095	.095	.095	.126	.052	.101	.052	.087	.042	.042	.042
...												
$t=29$	.120	.066	.066	.066	.150	.055	.102	.055	.120	.066	.066	.066
$t=30$	.120	.066	.066	.066	.150	.055	.102	.055	.120	.066	.066	.066

Quel rôle jouent les mathématiques dans tout cela ?

Question initiale : Comment définir la « pertinence » des pages web ?

Quel rôle jouent les mathématiques dans tout cela ?

Question initiale : Comment définir la « pertinence » des pages web ?

C'est d'abord un défi de modélisation :

Quel rôle jouent les mathématiques dans tout cela ?

Question initiale : Comment définir la « pertinence » des pages web ?

C'est d'abord un défi de modélisation :

- Quelles sont les données initiales ? Où veut-on aboutir ?

Quel rôle jouent les mathématiques dans tout cela ?

Question initiale : Comment définir la « pertinence » des pages web ?

C'est d'abord un défi de modélisation :

- Quelles sont les données initiales ? Où veut-on aboutir ?
- Les maths fournissent un langage pour formuler nos modèles.

Quel rôle jouent les mathématiques dans tout cela ?

Question initiale : Comment définir la « pertinence » des pages web ?

C'est d'abord un défi de modélisation :

- Quelles sont les données initiales ? Où veut-on aboutir ?
- Les maths fournissent un langage pour formuler nos modèles.
- Des calculs permettent de tester puis de raffiner nos modèles.

Quel rôle jouent les mathématiques dans tout cela ?

Question initiale : Comment définir la « pertinence » des pages web ?

C'est d'abord un défi de modélisation :

- Quelles sont les données initiales ? Où veut-on aboutir ?
- Les maths fournissent un langage pour formuler nos modèles.
- Des calculs permettent de tester puis de raffiner nos modèles.

Une fois le modèle fixé, les maths permettent de l'analyser :

Quel rôle jouent les mathématiques dans tout cela ?

Question initiale : Comment définir la « pertinence » des pages web ?

C'est d'abord un défi de modélisation :

- Quelles sont les données initiales ? Où veut-on aboutir ?
- Les maths fournissent un langage pour formuler nos modèles.
- Des calculs permettent de tester puis de raffiner nos modèles.

Une fois le modèle fixé, les maths permettent de l'analyser :

1. Existe-t-il toujours une solution à notre équation ?

Quel rôle jouent les mathématiques dans tout cela ?

Question initiale : Comment définir la « pertinence » des pages web ?

C'est d'abord un défi de modélisation :

- Quelles sont les données initiales ? Où veut-on aboutir ?
- Les maths fournissent un langage pour formuler nos modèles.
- Des calculs permettent de tester puis de raffiner nos modèles.

Une fois le modèle fixé, les maths permettent de l'analyser :

1. Existe-t-il toujours une solution à notre équation ?
2. Y en a-t-il plusieurs ? Ou une seule ?

Quel rôle jouent les mathématiques dans tout cela ?

Question initiale : Comment définir la « pertinence » des pages web ?

C'est d'abord un défi de modélisation :

- Quelles sont les données initiales ? Où veut-on aboutir ?
- Les maths fournissent un langage pour formuler nos modèles.
- Des calculs permettent de tester puis de raffiner nos modèles.

Une fois le modèle fixé, les maths permettent de l'analyser :

1. Existe-t-il toujours une solution à notre équation ?
2. Y en a-t-il plusieurs ? Ou une seule ?
3. Comment la calculer ? Efficacement ?

Le théorème du point fixe

Rappel : Dans notre modèle la loi de transition est donnée par

$$p'_i := \frac{c}{n} + \sum_{j \to i} \frac{1-c}{\ell_j} p_j. \tag{1}$$

Le théorème du point fixe

Rappel : Dans notre modèle la loi de transition est donnée par

$$p_i' := \frac{c}{n} + \sum_{j \to i} \frac{1-c}{\ell_j} p_j. \tag{1}$$

La mesure d'équilibre vérifie donc

$$m_i = \frac{c}{n} + \sum_{j \to i} \frac{1-c}{\ell_j} m_j. \tag{2}$$

Le théorème du point fixe

Rappel : Dans notre modèle la loi de transition est donnée par

$$p'_i := \frac{c}{n} + \sum_{j \to i} \frac{1-c}{\ell_j} p_j. \tag{1}$$

La mesure d'équilibre vérifie donc

$$m_i = \frac{c}{n} + \sum_{j \to i} \frac{1-c}{\ell_j} m_j. \tag{2}$$

Théorème ($\Leftarrow$ Théorème de point fixe de Banach)

Le théorème du point fixe

Rappel : Dans notre modèle la loi de transition est donnée par

$$p'_i := \frac{c}{n} + \sum_{j \to i} \frac{1-c}{\ell_j} p_j. \tag{1}$$

La mesure d'équilibre vérifie donc

$$m_i = \frac{c}{n} + \sum_{j \to i} \frac{1-c}{\ell_j} m_j. \tag{2}$$

Théorème ($\Leftarrow$ Théorème de point fixe de Banach)

Considérons un graphe fini et fixons le paramètre $c \in \,]0,1]$*.*

Le théorème du point fixe

Rappel : Dans notre modèle la loi de transition est donnée par

$$p'_i := \frac{c}{n} + \sum_{j \to i} \frac{1-c}{\ell_j} p_j. \tag{1}$$

La mesure d'équilibre vérifie donc

$$m_i = \frac{c}{n} + \sum_{j \to i} \frac{1-c}{\ell_j} m_j. \tag{2}$$

Théorème ($\Leftarrow$ Théorème de point fixe de Banach)

Considérons un graphe fini et fixons le paramètre $c \in \,]0,1]$*. Alors :*

1. *L'équation* (2) *admet une unique solution.*

Le théorème du point fixe

Rappel : Dans notre modèle la loi de transition est donnée par

$$p'_i := \frac{c}{n} + \sum_{j \to i} \frac{1-c}{\ell_j} p_j. \tag{1}$$

La mesure d'équilibre vérifie donc

$$m_i = \frac{c}{n} + \sum_{j \to i} \frac{1-c}{\ell_j} m_j. \tag{2}$$

Théorème ($\Leftarrow$ Théorème de point fixe de Banach)

Considérons un graphe fini et fixons le paramètre $c \in \,]0,1]$*. Alors :*

1. *L'équation* (2) *admet une unique solution.*
   Elle vérifie $m_1, \dots, m_n > 0$ *et* $m_1 + \dots + m_n = 1$*.*

Le théorème du point fixe

Rappel : Dans notre modèle la loi de transition est donnée par

$$p'_i := \frac{c}{n} + \sum_{j \to i} \frac{1-c}{\ell_j} p_j. \tag{1}$$

La mesure d'équilibre vérifie donc

$$m_i = \frac{c}{n} + \sum_{j \to i} \frac{1-c}{\ell_j} m_j. \tag{2}$$

Théorème ($\Leftarrow$ Théorème de point fixe de Banach)

Considérons un graphe fini et fixons le paramètre $c \in \,]0,1]$. *Alors :*

1. *L'équation* (2) *admet une unique solution.*
   Elle vérifie $m_1, \dots, m_n > 0$ *et* $m_1 + \dots + m_n = 1$.
2. *Pour toute distribution de probabilité initiale le processus de diffusion* (1) *converge vers cette unique mesure stationnaire* m.

Le théorème du point fixe

Rappel : Dans notre modèle la loi de transition est donnée par

$$p'_i := \frac{c}{n} + \sum_{j \to i} \frac{1-c}{\ell_j} p_j. \tag{1}$$

La mesure d'équilibre vérifie donc

$$m_i = \frac{c}{n} + \sum_{j \to i} \frac{1-c}{\ell_j} m_j. \tag{2}$$

Théorème ($\Leftarrow$ Théorème de point fixe de Banach)

Considérons un graphe fini et fixons le paramètre $c \in \,]0,1]$. *Alors :*

1. *L'équation* (2) *admet une unique solution.*
   Elle vérifie $m_1, \dots, m_n > 0$ *et* $m_1 + \cdots + m_n = 1$.
2. *Pour toute distribution de probabilité initiale le processus de diffusion* (1) *converge vers cette unique mesure stationnaire* m.
3. *La convergence est au moins aussi rapide que celle de la suite géométrique* $(1-c)^n$ *vers* 0.

Le théorème du point fixe

Rappel : Dans notre modèle la loi de transition est donnée par

$$p'_i := \frac{c}{n} + \sum_{j \to i} \frac{1-c}{\ell_j} p_j. \tag{1}$$

La mesure d'équilibre vérifie donc

$$m_i = \frac{c}{n} + \sum_{j \to i} \frac{1-c}{\ell_j} m_j. \tag{2}$$

Théorème ($\Leftarrow$ Théorème de point fixe de Banach)

Considérons un graphe fini et fixons le paramètre $c \in \,]0,1]$. Alors :

1. *L'équation* (2) *admet une unique solution.*
   Elle vérifie $m_1, \dots, m_n > 0$ *et* $m_1 + \dots + m_n = 1$.
2. *Pour toute distribution de probabilité initiale le processus de diffusion* (1) *converge vers cette unique mesure stationnaire* m.
3. *La convergence est au moins aussi rapide que celle de la suite géométrique* $(1-c)^n$ *vers* 0.

Vive l'analyse !

Conclusion

Pour être utile, un moteur de recherche doit non seulement *énumérer* les résultats d'une requête mais les *classer* par ordre de pertinence.

Conclusion

Pour être utile, un moteur de recherche doit non seulement *énumérer* les résultats d'une requête mais les *classer* par ordre de pertinence.

- En première approximation Google analyse le graphe formé par les pages web et les liens entre elles.

Conclusion

Pour être utile, un moteur de recherche doit non seulement *énumérer* les résultats d'une requête mais les *classer* par ordre de pertinence.

- En première approximation Google analyse le graphe formé par les pages web et les liens entre elles.
- Interprétant un lien $j \to i$ comme « vote » de la page P_j en faveur de P_i, le modèle PageRank définit une mesure de « popularité ».

Conclusion

Pour être utile, un moteur de recherche doit non seulement *énumérer* les résultats d'une requête mais les *classer* par ordre de pertinence.

- En première approximation Google analyse le graphe formé par les pages web et les liens entre elles.
- Interprétant un lien $j \to i$ comme « vote » de la page P_j en faveur de P_i, le modèle PageRank définit une mesure de « popularité ».
- Le théorème du point fixe assure que cette équation admet une unique solution, et justifie l'algorithme itératif pour l'approcher.

Conclusion

Pour être utile, un moteur de recherche doit non seulement *énumérer* les résultats d'une requête mais les *classer* par ordre de pertinence.

- En première approximation Google analyse le graphe formé par les pages web et les liens entre elles.
- Interprétant un lien $j \to i$ comme « vote » de la page P_j en faveur de P_i, le modèle PageRank définit une mesure de « popularité ».
- Le théorème du point fixe assure que cette équation admet une unique solution, et justifie l'algorithme itératif pour l'approcher.

Muni de ces outils mathématiques et d'une habile stratégie d'entreprise, Google gagne des milliards de dollars.

Conclusion

Pour être utile, un moteur de recherche doit non seulement *énumérer* les résultats d'une requête mais les *classer* par ordre de pertinence.

- En première approximation Google analyse le graphe formé par les pages web et les liens entre elles.
- Interprétant un lien $j \to i$ comme « vote » de la page P_j en faveur de P_i, le modèle PageRank définit une mesure de « popularité ».
- Le théorème du point fixe assure que cette équation admet une unique solution, et justifie l'algorithme itératif pour l'approcher.

Muni de ces outils mathématiques et d'une habile stratégie d'entreprise, Google gagne des milliards de dollars.

Il fallait y penser.

Conclusion

Pour être utile, un moteur de recherche doit non seulement *énumérer* les résultats d'une requête mais les *classer* par ordre de pertinence.

- En première approximation Google analyse le graphe formé par les pages web et les liens entre elles.
- Interprétant un lien $j \to i$ comme « vote » de la page P_j en faveur de P_i, le modèle PageRank définit une mesure de « popularité ».
- Le théorème du point fixe assure que cette équation admet une unique solution, et justifie l'algorithme itératif pour l'approcher.

Muni de ces outils mathématiques et d'une habile stratégie d'entreprise, Google gagne des milliards de dollars.

Il fallait y penser.

Je vous remercie de votre attention !

Littérature

S. Brin, L. Page : *The Anatomy of a Large-Scale Hypertextual Web Search Engine*, Stanford University 1998

Littérature

S. Brin, L. Page : *The Anatomy of a Large-Scale Hypertextual Web Search Engine*, Stanford University 1998

K. Bryan, T. Leise : *The $825,000,000,000 eigenvector : the linear algebra behind Google*, SIAM Review 48 (2006) 569-581

Littérature

S. Brin, L. Page : *The Anatomy of a Large-Scale Hypertextual Web Search Engine*, Stanford University 1998

K. Bryan, T. Leise : *The $825,000,000,000 eigenvector : the linear algebra behind Google*, SIAM Review 48 (2006) 569-581

(Ces articles sont disponibles en ligne, cherchez-les avec Google. ;-)

Littérature

📖 S. Brin, L. Page : *The Anatomy of a Large-Scale Hypertextual Web Search Engine*, Stanford University 1998

📖 K. Bryan, T. Leise : *The $825,000,000,000 eigenvector : the linear algebra behind Google*, SIAM Review 48 (2006) 569-581

(Ces articles sont disponibles en ligne, cherchez-les avec Google. ;-)

📖 M. Eisermann : *Comment Google classe les pages web*, Images des Mathématiques, La Tangente, Quadrature
`www.igt.uni-stuttgart.de/eiserm/popularisation/#google`

Littérature

📖 S. Brin, L. Page : *The Anatomy of a Large-Scale Hypertextual Web Search Engine*, Stanford University 1998

📖 K. Bryan, T. Leise : *The $825,000,000,000 eigenvector : the linear algebra behind Google*, SIAM Review 48 (2006) 569-581

(Ces articles sont disponibles en ligne, cherchez-les avec Google. ;-)

📖 M. Eisermann : *Comment Google classe les pages web*, Images des Mathématiques, La Tangente, Quadrature
`www.igt.uni-stuttgart.de/eiserm/popularisation/#google`

(Expérience pratique : si vous pensez que ce document le mérite, faites-y pointer vos liens pour augmenter son PageRank. ;-)

Le modèle PageRank est-il plausible ?

La structure caractéristique des documents hypertextes sont les citations mutuelles.

Le modèle PageRank est-il plausible ?

La structure caractéristique des documents hypertextes sont les citations mutuelles.

- L’hypothèse à la base du modèle PageRank est que l’auteur d’une page ajoute des liens vers les pages qu’il considère utiles.

Le modèle PageRank est-il plausible ?

La structure caractéristique des documents hypertextes sont les citations mutuelles.

- L'hypothèse à la base du modèle PageRank est que l'auteur d'une page ajoute des liens vers les pages qu'il considère utiles.
- Ainsi des millions d'auteurs lisent et jugent mutuellement leurs pages, et leurs jugements s'expriment par leurs liens.

Le modèle PageRank est-il plausible ?

La structure caractéristique des documents hypertextes sont les citations mutuelles.

- L'hypothèse à la base du modèle PageRank est que l'auteur d'une page ajoute des liens vers les pages qu'il considère utiles.
- Ainsi des millions d'auteurs lisent et jugent mutuellement leurs pages, et leurs jugements s'expriment par leurs liens.
- Le modèle de la marche aléatoire en profite en transformant l'évaluation mutuelle en une mesure globale de popularité.

Le modèle PageRank est-il plausible ?

La structure caractéristique des documents hypertextes sont les citations mutuelles.

- L'hypothèse à la base du modèle PageRank est que l'auteur d'une page ajoute des liens vers les pages qu'il considère utiles.
- Ainsi des millions d'auteurs lisent et jugent mutuellement leurs pages, et leurs jugements s'expriment par leurs liens.
- Le modèle de la marche aléatoire en profite en transformant l'évaluation mutuelle en une mesure globale de popularité.

Cet argument de plausibilité est à débattre et à expérimenter. . .

Le modèle PageRank est-il plausible ?

La structure caractéristique des documents hypertextes sont les citations mutuelles.

- L'hypothèse à la base du modèle PageRank est que l'auteur d'une page ajoute des liens vers les pages qu'il considère utiles.
- Ainsi des millions d'auteurs lisent et jugent mutuellement leurs pages, et leurs jugements s'expriment par leurs liens.
- Le modèle de la marche aléatoire en profite en transformant l'évaluation mutuelle en une mesure globale de popularité.

Cet argument de plausibilité est à débattre et à expérimenter. . .

L'ultime argument en faveur du modèle PageRank est son succès : le classement semble bien refléter les attentes des utilisateurs.

Le modèle PageRank est-il descriptif ou normatif ?

Au début de son existence, Google se voulait un outil *descriptif* :
si une page est importante, alors elle figure en tête du classement.

Le modèle PageRank est-il descriptif ou normatif ?

Au début de son existence, Google se voulait un outil *descriptif* :
si une page est importante, alors elle figure en tête du classement.

Son écrasant succès a fait de Google une référence *normative* :
si une page figure en tête du classement, alors elle est importante.

Le modèle PageRank est-il descriptif ou normatif ?

Au début de son existence, Google se voulait un outil *descriptif* :
si une page est importante, alors elle figure en tête du classement.

Son écrasant succès a fait de Google une référence *normative* :
si une page figure en tête du classement, alors elle est importante.

Pour des sites web commerciaux, l'optimisation de leur classement PageRank est ainsi devenue un enjeu vital.

Le modèle PageRank est-il descriptif ou normatif ?

Au début de son existence, Google se voulait un outil *descriptif* :
si une page est importante, alors elle figure en tête du classement.

Son écrasant succès a fait de Google une référence *normative* :
si une page figure en tête du classement, alors elle est importante.

Pour des sites web commerciaux, l'optimisation de leur classement PageRank est ainsi devenue un enjeu vital.

Stratégie évidente : il suffit d'attirer des liens, de préférence ceux émis des pages importantes, et il vaut mieux en émettre très peu.

Le modèle PageRank est-il descriptif ou normatif ?

Au début de son existence, Google se voulait un outil *descriptif* :
si une page est importante, alors elle figure en tête du classement.

Son écrasant succès a fait de Google une référence *normative* :
si une page figure en tête du classement, alors elle est importante.

Pour des sites web commerciaux, l'optimisation de leur classement PageRank est ainsi devenue un enjeu vital.

Stratégie évidente : il suffit d'attirer des liens, de préférence ceux émis des pages importantes, et il vaut mieux en émettre très peu.

Ainsi l'omniprésence de Google change l'utilisation des liens par les auteurs des pages web. . . ce qui remet en question l'hypothèse à la base même du modèle PageRank.