

# L'algorithme du PageRank expliqué

## Le PageRank, c'est quoi ?

La base du PageRank - que nous noterons parfois PR dans la suite de ce document - est une formule mathématique, à l'allure rébarbative, mais en définitive assez simple à comprendre. Cette méthode est utilisée par Google pour déterminer **l'importance d'une page Web**.

**Elle se base sur un concept très simple** : un lien émis par une page A vers une page B est assimilé à un « vote » de A pour B. Au plus une page reçoit de « votes », au plus cette page est considérée comme importante par Google, exactement comme le principe des élections que nous connaissons tous.

Nous reviendrons plus en détail sur ce point, mais retenez dès à présent qu'un vote émis par la page d'accueil d'un site majeur tel que Microsoft ou CNN pèse beaucoup plus lourd qu'un vote émis par la page perso de votre cousine, si mignonne soit-elle.

## Combattons les idées fausses

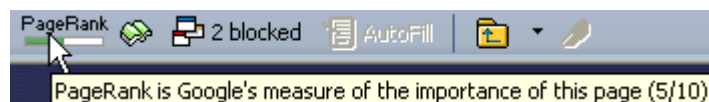
Retenons aussi que le PageRank est une mesure de l'importance d'une page, et non d'un site entier. Vous entendrez souvent parler de « site de rang n », il s'agit d'un abus de langage décrivant le rang de la page d'accueil du site.

Il n'y a pas, nous le verrons plus bas, de notion d'importance de site dans l'algorithme du PageRank.

De même, l'**importance** d'une page est sans rapport aucun avec l'intérêt ou la pertinence de celle-ci, ces deux dernières notions étant totalement absentes de l'algorithme du PageRank. Elles interviennent néanmoins dans les pages de résultat de recherche.

Ce PageRank peut être visualisé par les utilisateurs de la « toolbar » Google.

La représentation graphique se fait sur une échelle de 1 à 10. L'exemple ci-dessus montre l'affichage d'une page ayant un PageRank égal à 5 (noté PR5).



### La version 2.0 de la toolbar Google

Elle permet aussi de bloquer les popups indésirables

## Et cette fameuse formule, alors ?

En reprenant - *après traduction* - la publication originale de Google, voici les explications données :

Nous assumons qu'une page A reçoit des liens (ou "votes") émis par les pages T1...Tn.

Le paramètre d est un facteur d'amortissement pouvant être ajusté entre 0 et 1.

Nous donnons généralement à d la valeur 0.85.

De même, C(A) est défini comme le nombre de liens émis par la page A (liens sortants). *Le PageRank de la page A est défini comme suit :*

$$\mathbf{PR(A) = (1-d) + d (PR(T1)/C(T1) + \dots + PR(Tn)/C(Tn))}$$

Le PageRank peut être calculé en utilisant un simple algorithme itératif, et correspond au vecteur propre principal de la matrice normalisée des liens du Web.

Tout cela est bien moins compliqué qu'il n'y paraît, essayons de disséquer l'expression.

Pour ce faire, voici l'explication de la notation utilisée :

PR(A) le PageRank de la page A

PR(Tn) le PageRank de la page Tn

C(Tn) le nombre de liens émis sur la page Tn

d tous les « votes » sont additionnés, mais pour en limiter l'importance, le total est multiplié par ce coefficient d'amortissement (0.85)

1 - d Un petit peu de « magie mathématique » qui permet de garantir que la moyenne des PageRank de l'ensemble des pages du Web sera de 1.

L'examen de cette formule permet de voir que le PageRank d'une page n'ayant aucun lien entrant sera de 0.15 , **soit** :  $(1 - 0.85) + 0.85*(0) = 0.15$

Et là apparaît la cause de la confusion la plus répandue au sujet du PageRank :

**Que vient faire ici cette valeur fractionnaire alors que la toolbar n'affiche que des valeurs entières ?**

## Oublions la toolbar quelques instants !

Il est généralement admis que l'échelle du PageRank est logarithmique, sans que ceci ne soit officiellement confirmé. Pour cette raison, la base utilisée ne peut qu'être estimée. Il est de même raisonnable de penser que cette base évolue dans le temps.

Prenons une échelle logarithmique de base 10 pour simplifier nos calculs, le raisonnement restant valable quelle que soit la base choisie.

### **PageRank Affiché**   **PageRank réel**

(log base 10)

(calculé)

PR0

$0 \leq PR < 1$

PR1

$1 \leq PR < 10$

PR2

$10 \leq PR < 100$

PR3

$100 \leq PR < 1000$

PR4

$1000 \leq PR < 10000$

et ainsi de suite jusqu'au PR10 pour les plus heureux.

On voit ici, que chaque niveau de PageRank est 10 fois plus élevé que le niveau précédent. Ce qui signifie en clair qu'il est 10 fois plus ardu de passer de PR4 à PR5 que de passer de PR3 à PR4 (pour mémoire, la base 10 a été choisie arbitrairement dans notre exemple).

Une des raisons pour lesquelles on estime que l'échelle évolue dans le temps, est que le PageRank maximum n'est calculé que lorsque Google fait sa mise à jour de l'index, et que le nombre de pages indexées est en constante augmentation.

Cette évolution de l'échelle expliquerait pourquoi certaines pages voient leur PageRank diminuer au fil des indexations, alors que le nombre de liens entrant reste inchangé.

En reprenant l'exemple de la page sans lien entrant donné précédemment (PR=0.15), nous voyons que la toolbar nous affichera bien la valeur 0.

## Comment le PageRank est-il calculé ?

C'est ici que les choses se compliquent un petit peu.

Nous avons vu que le PageRank d'une page A dépend du PageRank des pages T1...Tn qui émettent un lien vers A, et ne peut donc pas être déterminé sans connaître le PR de ces dernières, et de toutes celles qui émettent un lien vers elles, et ainsi de suite...

Lorsqu'on réalise que les liens inter pages peuvent boucler, cela ressemble bien à « *mission impossible* ».

Reprenons la publication de Google décrivant le PageRank :

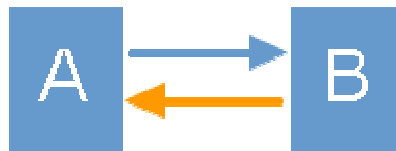
Le PageRank peut être calculé en utilisant un simple algorithme itératif, et correspond au vecteur propre principal de la matrice normalisée des liens du Web

Ceci signifie que le calcul du PageRank d'une page peut être effectué sans connaître le PR final des pages émettant un lien vers elle.

Cela peut sembler paradoxal, mais chaque itération fait converger les résultats vers une valeur de plus en plus précise. La seule chose à faire, est de retenir la valeur obtenue pour pouvoir démarrer l'itération suivante avec cette dernière.

### Ce sera plus simple avec quelques exemples :

Réinventons le Web dans sa forme la plus simple : 2 pages A et B pointant l'une vers l'autre. Chaque page a un lien sortant, donc  $C(A) = C(B) = 1$



#### Première estimation :

Nous ne connaissons pas le PR des deux pages, donc il nous faut une valeur de départ : 1.0 par exemple.

$$PR(A) = (1 - d) + d(PR(B)/1)$$

$$PR(B) = (1 - d) + d(PR(A)/1)$$

#### Soit, avec un facteur d'amortissement de 0.85 :

$$PR(A) = 0.15 + 0.85 * 1 = 1$$

$$PR(B) = 0.15 + 0.85 * 1 = 1$$

Bon, les valeurs ne changent pas... nous avons peut-être eu trop de chance avec notre estimation.

Prenons une valeur de départ différente : 0

### **Première itération**

$$PR(A) = 0.15 + 0.85 * 0 = \mathbf{0.15}$$

$$PR(B) = 0.15 + 0.85 * 0.15 = \mathbf{0.2775}$$

### **Deuxième itération**

$$PR(A) = 0.15 + 0.85 * 0.2775 = \mathbf{0.385875}$$

$$PR(B) = 0.15 + 0.85 * 0.385875 = \mathbf{0.47799375}$$

### **Troisième itération**

$$PR(A) = 0.15 + 0.85 * 0.47799375 = \mathbf{0.5562946875}$$

$$PR(B) = 0.15 + 0.85 * 0.5562946875 = \mathbf{0.622850484375}$$

Nous remarquons que les valeurs augmentent à chaque itération.

Dans notre exemple, avec nos deux pages A et B, nous savons que le PR doit être égal à un, l'algorithme nous précisant que le PR moyen de toutes les pages du Web est égal à 1. Est-ce que nos valeurs de PR calculées ne peuvent pas augmenter indéfiniment et dépasser 1, ce qui invaliderait la formule ?

Essayons avec une valeur supérieure pour voir ce qui se passe : prenons une valeur 2.0 pour redémarrer notre expérience.

$$PR(A) = 0.15 + 0.85 * 2 = \mathbf{1.85}$$

$$PR(B) = 0.15 + 0.85 * 1.85 = \mathbf{1.7225}$$

**Bon, cela baisse ! Essayons une fois de plus :**

$$PR(A) = 0.15 + 0.85 * 1.7225 = \mathbf{1.614125}$$

$$PR(B) = 0.15 + 0.85 * 1.614125 = \mathbf{1.52200625}$$

**Une troisième fois :**

$$PR(A) = 0.15 + 0.85 * 1.52200625 = \mathbf{1.4437053125}$$

$$PR(B) = 0.15 + 0.85 * 1.4437053125 = \mathbf{1.377149515625}$$

Nos valeurs continuent à converger vers 1, c'est ce que nous attendions.

### **Premier enseignement :**

Quelle que soit la valeur de départ prise pour le calcul du PR, la moyenne normalisée tendra vers 1.

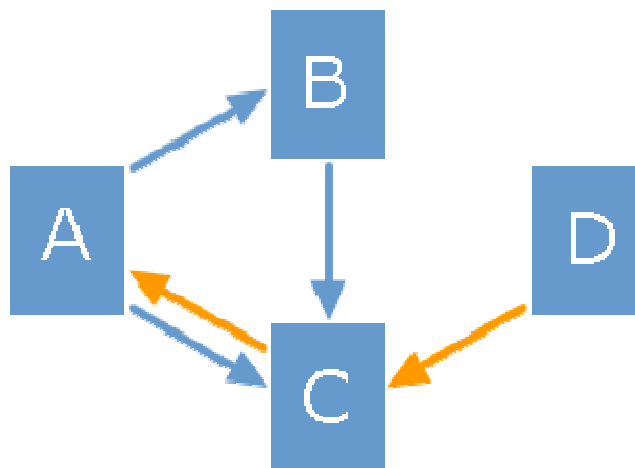
## Accélérer les calculs grâce au facteur d'amortissement

L'exemple qui a précédé nous montre un Web simplissime, seulement 2 pages. Combien d'itérations faut-il pour voir les résultats converger pour un grand nombre de pages ? A l'heure actuelle, Google a près de 4 milliards de pages dans sa base, ce qui pourrait nécessiter plusieurs milliards d'itérations.

C'est ici que le facteur d'amortissement joue son rôle. S'il est choisi trop élevé, le calcul demandera un nombre d'itérations énorme, alors que s'il est trop bas les valeurs ne convergeront pas véritablement, mais finiront par osciller autour de la valeur théorique vraie, un peu à la manière d'un pendule.

Avec un facteur d'amortissement de 0.85, il nous faut une quarantaine d'itérations pour affiner le calcul du PageRank.

### Deuxième exemple : quatre pages liées



Dans cet exemple, nous avons un site comprenant quatre pages, dont une ne recevant aucun lien (la page D). Le PR de cette page sera donc de 0.15, grâce au premier terme de la formule du PageRank ( $1 - d$ )

Bien qu'ayant un PR calculé, il est vraisemblable que cette page disparaîtra de l'index Google très vite, n'ayant aucun lien entrant.

Au bout d'une vingtaine d'itérations, les valeurs de PR pour nos pages convergent vers les valeurs suivantes :

Page A	1.49
Page B	0.78
Page C	1.58
Page D	0.15
<b>Somme des PageRank</b>	<b>4.0</b>
<b>Moyenne</b>	<b>1.0</b>

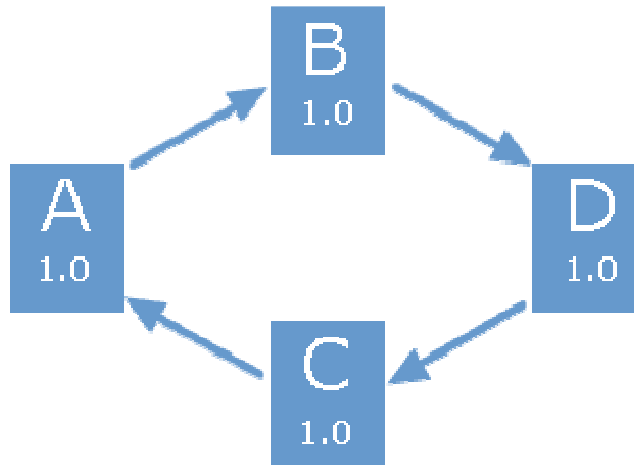
Nous voyons que dans notre exemple, la page C a le PR le plus élevé. C'était prévisible dès l'examen du graphique, comme elle reçoit un lien entrant des pages A, B et D, et n'en émet qu'un seul vers la page A.

### Calculez vous-même les exercices suivants...

Pour les exemples suivants, nous intégrerons les résultats directement dans le graphique, pour ne pas alourdir l'article avec les tableaux comprenant les valeurs de PageRank intermédiaires. Sauf indication contraire, la valeur 1.0 est prise pour la première itération, et le nombre d'itérations sera de 40.

Le lecteur qui souhaite expérimenter par lui-même pourra utiliser l'excellent calculateur de PageRank disponible sur le site de [WebWorkShop](#).

### Troisième exemple : liens circulaires



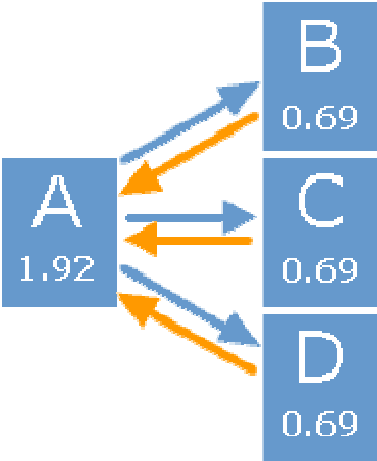
	A	B	C	D	E	F	PR
A	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	1.0
B	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	1.0
C	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	1.0
D	<input type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	1.0

Cases cochées pour l'exemple 3

Comme on pouvait s'y attendre dans ce cas, les liens circulaires ne favorisent aucune page du site, chaque page ayant exactement un lien entrant et un lien sortant.

Le PageRank de chaque page s'établira donc à 1.0

**Quatrième exemple : structure hiérarchique simple**

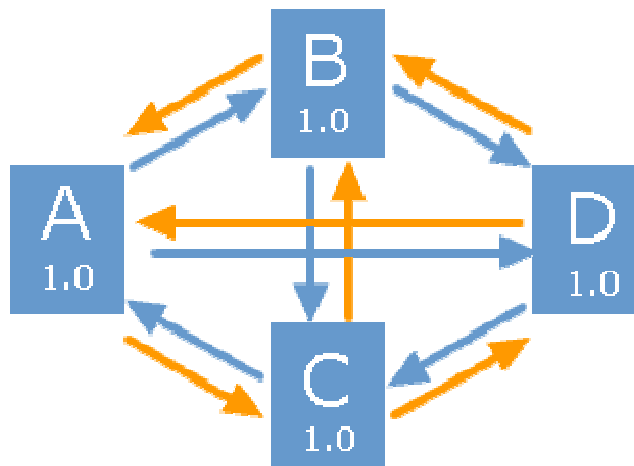


	A	B	C	D	E	F	PR
A	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	1.9189214
B	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	0.6936944
C	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	0.6936944
D	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	0.6936944

**Cases cochées pour l'exemple 4**

Voilà qui est mieux ! Le PageRank de la page A est optimisé grâce à la structure de liens hiérarchique.

### Cinquième exemple : on lie à tout va !

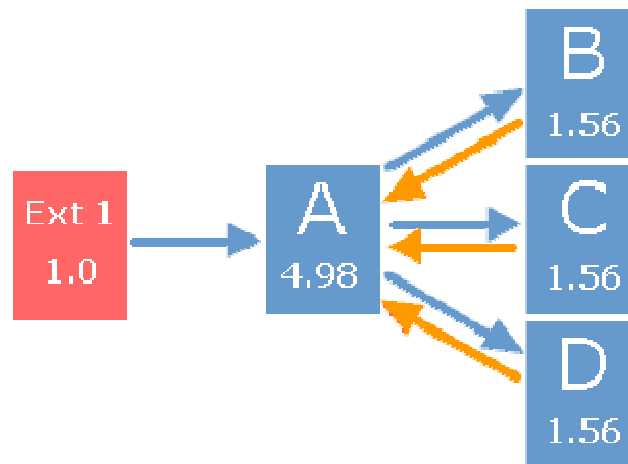


	A	B	C	D	E	F	PR
A	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	1.0
B	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	1.0
C	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	1.0
D	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	1.0

Cases cochées pour l'exemple 5

Ici, on a voulu lier toutes les pages entre-elles, ce qui fait qu'aucune page n'est prépondérante. La page d'accueil du site hérite d'un PR1, au même titre que toutes les autres pages. On obtient le même PR qu'avec les liaisons circulaires, tout en gagnant en facilité de navigation pour les visiteurs. Ce type de chaînage devient très vite difficile à réaliser dès que le nombre de pages du site augmente.

## Sixième exemple : structure hiérarchique avec lien entrant



	A	B	C	D	E	F	PR	s1	s2	s3	s4
A	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	4.9819775	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
B	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	1.5615603	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
C	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	1.5615603	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
D	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	1.5615603	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
E	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	0.15	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
F	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	0.15	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

Liens entrants

e1       PR 1

Cases cochées pour l'exemple 6

Nous avons estimé à 1.0 le PR de la page externe (backlink) liant vers notre page A. Dans notre exemple, comme nous faisons abstraction du reste du Web, nous imaginons que le Webmaster du site extérieur nous aime vraiment beaucoup et que le seul lien émis par sa page pointe vers la nôtre. Ceci a peu de chances de se produire dans la réalité.

C'est tout bénéfique pour la page d'accueil qui, non contente d'hériter de 0.85 point de PR de la page externe, répercute cet accroissement de PR sur les pages internes du site et le récupère grâce aux liens en retour.

Cela fait **pas moins de trois points de PR** gagnés par la page d'accueil par rapport à la structure hiérarchique de notre quatrième exemple.

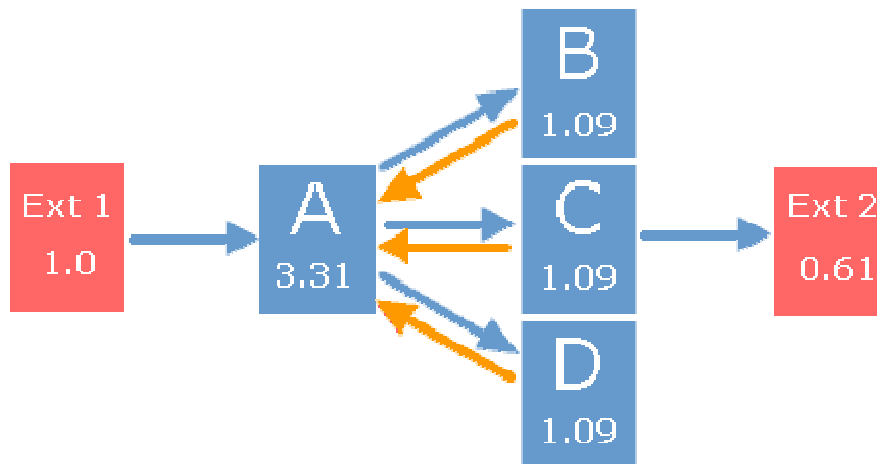
### Gardons à l'esprit un point important !

Les valeurs de PageRank données ici sont celles reprises dans notre tableau en début d'article, dans la colonne "**Pagerank Réel**" et non dans celle du Pagerank affiché par le Toolbar.

Ne rêvons tout de même pas, il ne suffit pas d'un petit lien à PR1 pour obtenir un PR4 affiché par la toolbar !

En reprenant ce tableau et dans l'hypothèse d'une échelle logarithmique de base 10, cette page aurait un PR1 affiché, son PR réel se situant dans l'intervalle 1...10

## Septième exemple : structure hiérarchique avec lien entrant et lien sortant



	A	B	C	D	E	F	PR	s1	s2	s3	s4
A	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	3.3141361	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
B	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	1.0890052	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
C	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	1.0890052	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
D	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	1.0890052	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
E	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	0.15	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
F	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	0.15	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
<b>Liens entrants</b> e1 <input checked="" type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> PR <input type="text" value="1"/>											

### Cases cochées pour l'exemple 7

Le calculateur démontrant un site de 6 pages, les pages non liées E-F ont un PR de (1-0.85) suivant l'algorithme.

Le simple fait d'émettre un lien vers une page extérieure au départ de la page C fait chuter le PR de manière conséquente sur toutes les pages du site.

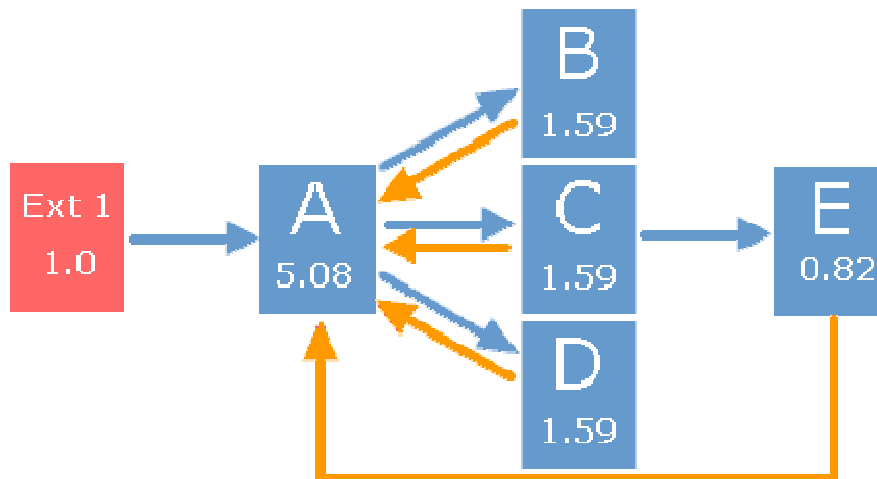
Tous les autres liens sont identiques à l'exemple précédent.

Cet exemple en fera méditer plus d'un, nous en sommes certains !

La seule cause de la chute de PR vient du fait que la page C qui redistribuait dans l'exemple 6 tout son PR à la page A, ne lui renvoie plus que la moitié de celui-ci. Que l'autre moitié du PR soit redistribuée à une page interne ou externe ne fait pas de différence s'il n'y a pas de lien en retour.

Mais, en tant que webmaster avisé, nous ferons quelque chose comme... (Voir exemple 8).

## Huitième exemple : variation du précédent



	A	B	C	D	E	F	PR	s1	s2	s3	s4
A	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	5.0765477	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
B	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	1.5883552	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
C	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	1.5883552	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
D	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	1.5883552	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
E	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	0.8250509	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
F	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>	0.15	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

▲ ▲ ▲ ▲ ▲ ▲

**Liens entrants**

e1       PR

Cases cochées pour l'exemple 8

Changeons la page externe en page interne, pour pouvoir s'assurer d'un lien en retour vers la page d'accueil.

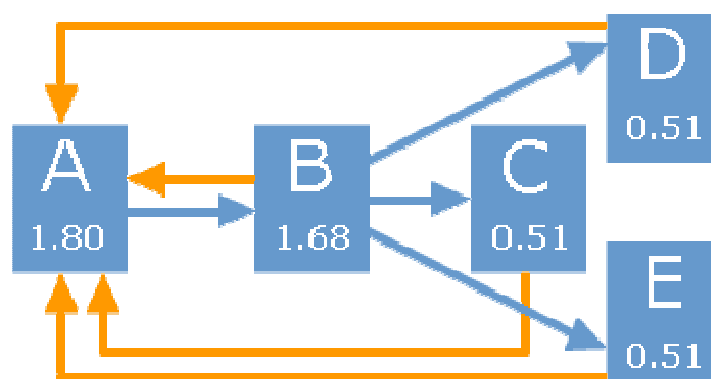
Grâce à ce lien, le PR de tout le site est augmenté !

Retenons que l'algorithme du PageRank ne fait aucune différence entre les pages internes et les pages externes.

L'avantage des pages internes est surtout dû au meilleur contrôle sur les liens émis, mais un bon échange de lien est toujours souhaitable.

Seules les pages extérieures au site peuvent augmenter le capital PR global. Sans ces liens externes le PR global ne pourrait dépasser le nombre de pages du site.

## Neuvième exemple : un plan de site



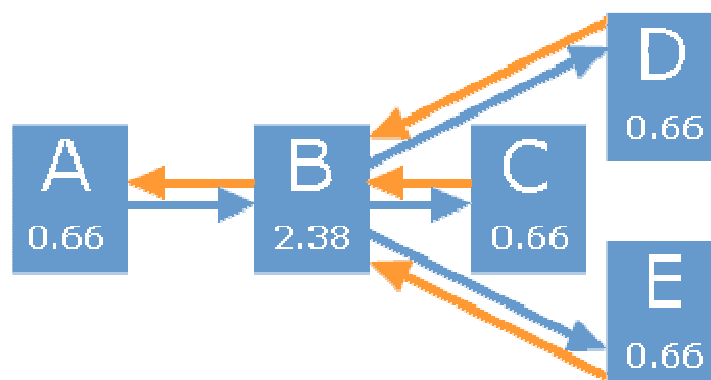
	A	B	C	D	E	F	PR
A	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	1.799582
B	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	1.6796447
C	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	0.5069245
D	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	0.5069245
E	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	0.5069245

Cases cochées pour l'exemple 9

Dans cet exemple, la page d'accueil (A) émet un lien vers le plan du site (B). Celui-ci, en plus du lien retour vers la page d'accueil émet un lien vers chacune des pages du site (C, D et E). Pour éviter les "fuites" de PageRank, celles-ci émettent un lien en retour vers la page d'accueil.

La somme des PR de nos pages étant de 5, nous n'avons donc aucune perte due à un mauvais chaînage.

## Dixième exemple : un plan de site lié différemment



	A	B	C	D	E	F	PR
A	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	0.655406
B	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	2.3783803
C	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	0.6554058
D	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	0.6554058
E	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	0.6554058

Cases cochées pour l'exemple 10

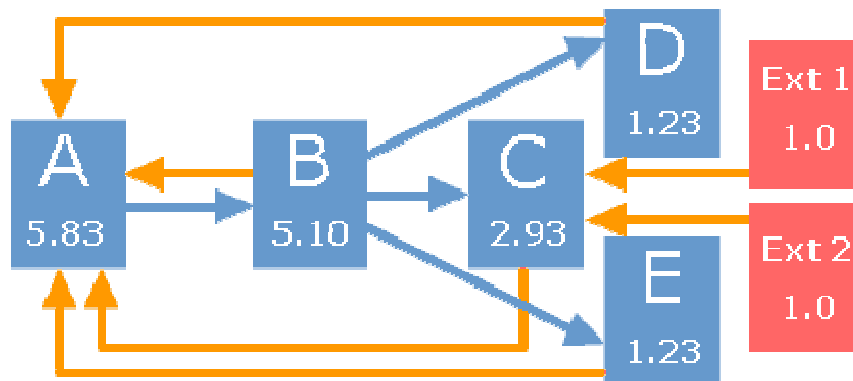
La seule différence avec l'exemple précédent vient du lien en retour des pages internes du site. Plutôt que d'émettre un lien vers la page d'accueil, elles émettent le lien en retour vers le plan du site (B).

Ceci a pour effet de favoriser le PageRank de notre page B, ce qui peut être souhaité si elle est riche en mots clés.

Par contre, notre page d'accueil voit son PageRank diminuer fortement comme elle n'a plus qu'un lien entrant.

La somme des PR des pages reste à 5 - aux erreurs d'arrondi près - ce qui démontre bien l'absence de perte de PageRank.

## Onzième exemple : un plan avec liens entrants sur page interne



	A	B	C	D	E	F	PR	s1	s2	s3	s4
A	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	5.8271056	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
B	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	5.1030397	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
C	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	2.9343959	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
D	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	1.2343959	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
E	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	1.2343959	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
F	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>		<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
▲ ▲ ▲ ▲ ▲ ▲											
<b>Liens entrants</b>											
e1	<input type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	PR 1	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
e2	<input type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	PR 1	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

Cases cochées pour l'exemple 11

### Rêvons un peu...

Le contenu intéressant de la page C lui vaut des liens spontanés de la part de deux sites extérieurs, sans lien en retour.

C'est ce qui va arriver à notre article, non ? 😊

Comme dans les exemples précédents, nous estimerons le PR des pages liantes à 1.0.

La page C voit tout naturellement l'accroissement de PR le plus important (multiplié par 5+), du fait des liens directs. Mais, grâce au chaînage interne, la page d'accueil et le plan du site font toutes deux un bond vers le haut.

Dans une moindre mesure, les pages D et E profitent du gain en PR du plan du site (B).

## **Question brûlante : Comment accroître son PageRank ?**

Nous ne pouvons pas passer sous silence cette interrogation qui nous concerne tous.

Nous avons vu précédemment que, selon la formule utilisée, la moyenne des PageRank des pages du Web s'établit à 1.0

Chaque page ajoutée à un site fait donc monter de un point et un seul son "capital PageRank" et un chaînage interne approprié des pages du site aura pour fonction de redistribuer ce capital. Cette accroissement de capital ne sera réalisée que si les nouvelles pages sont liées avec le reste de votre site, et sont de ce fait prises en compte par l'algorithme.

Sans lien extérieur, le "capital PageRank" d'un site ne pourra jamais excéder le nombre de pages.

Il vous reste toujours la possibilité de vous transformer en "Mega-Site", un peu à la mode de CNN. Ces sites sont régulièrement enrichis de nouvelles pages, créées par une véritable armée de rédacteurs, toutes pourvues d'au moins un lien vers la page d'accueil. Ces Mega-Sites sont le plus souvent abonnés au PageRank élevé, en tout cas pour la page d'accueil. Le chaînage interne qui privilégie celle-ci fait s'effondrer très rapidement le PR des pages internes du site.

### **Vous n'avez pas d'armée de rédacteurs à disposition ?**

Une conclusion bien connue de tous les Webmasters s'impose : gagnez des liens en provenance d'autres sites. Cette chasse aux "backlinks" est, sous certains aspects, la forme moderne de la "conquête du Graal".

### **Comment gagner ces fameux backlinks ?**

De plusieurs manières, la plus pratiquée restant l'échange de liens avec des sites amis.

Ce n'est malheureusement pas la plus "rentable", les pages de liens étant souvent chargées d'une multitude de liens, ce qui réduit sensiblement le PR distribué. Souvenons nous de la partie de la formule qui nous intéresse dans ce cas :  $PR(Tn)/C(Tn)$ , soit le PR de la page divisé par le nombre de liens sortants.

De plus, les webmasters qui ont compris le fonctionnement de l'algorithme ne feront vraisemblablement pas beaucoup d'efforts pour optimiser le PageRank de leur(s) page(s) de liens... c'est la vie !

La meilleure manière d'obtenir ces fameux backlinks dépendra entièrement de vous. Il vous faudra donner envie à vos visiteurs webmasters d'émettre un lien vers votre site, et seul le contenu de votre site ou d'une page spécifique peut susciter cet élan.

Gardez à l'esprit que "le contenu est Roi", et qu'un contenu riche reste la meilleure manière d'obtenir des liens spontanés. C'est d'ailleurs un des buts - avoué - poursuivi par cet article.

Après 24 heures de mise en ligne, nous avons observé dans nos logs une diversification des référents, qui sont autant de backlinks supplémentaires. Sur ce plan au moins, nous pouvons dire : "*mission accomplie*".

Une dernière manière, non sans intérêt, est de proposer un programme utile en téléchargement gratuit. Un site comme phpBB a droit à un PR8 parce que plusieurs milliers - ou millions - de sites utilisant ce système de forum ont au bas de chaque page un lien vers le site phpBB.com. Cette multitude de liens unidirectionnels contribuent énormément à la notoriété du site.

## Les liens sortants perdent du PageRank

Cette affirmation est peut-être un peu brutale et mathématiquement fausse. La formule du PageRank ne fait pas la différence entre un lien interne et un lien externe. Il n'y a donc pas de coefficient négatif pour ces derniers.

Une chose est certaine toutefois, chaque lien émis distribue une fraction du pouvoir de vote de la page qui l'héberge. Pour chaque vote émis vers le monde extérieur, c'est donc un petit peu de votre capital qui s'en va... c'est le prix à payer pour ne pas vivre en autarcie.

### Pourquoi pas des liens en JavaScript ?

C'est un fait avéré, les moteurs de recherche n'interprètent pas JavaScript, donc ne verront pas ce type de liens et ne distribueront pas le PR.

Il est donc tentant d'émettre tous ses liens sortants en JavaScript, non ? ... C'est à notre avis une **très mauvaise idée** pour les raisons suivantes :

- ▶ Cette façon de faire est contraire à "l'esprit du Web" et, reconnaissons-le, assez peu fair-play vis à vis du webmaster du site externe. S'il découvre la supercherie, il y a gros à parier qu'il ne sera pas enclin à vous faire une bonne publicité et qu'il supprimera illico tout lien vers votre site. A vouloir trop gagner... le beurre et l'argent du beurre... on peut aussi tout perdre.
- ▶ Les robots Google et l'algorithme détectent très facilement les sites qui n'ont aucun lien sortant, ou ceux dont le déséquilibre est flagrant. Ces sites sont perçus comme des anomalies de l'algorithme et se voient souvent réserver un "sort particulier" - pas toujours au goût du webmaster.
- ▶ Les gros annuaires, pensez à Yahoo ou Dmoz, sont des sites à fort PageRank (PR10) bien qu'ils aient énormément de liens sortants. Google crédite un annuaire tel que Yahoo d'environ 700,000 backlinks, alors que Yahoo comporte plusieurs millions de liens sortants et devrait donc se retrouver avec un PageRank très bas...