

EdgeRank et GraphRank : des algorithmes au cœur du fonctionnement de Facebook

[Retour au sommaire de la lettre](#)

Domaine :	Recherche	Référencement
Niveau :	Pour tous	Avancé

Pour classer leurs résultats, les moteurs de recherche tels que Google utilisent aujourd'hui des algorithmes extrêmement sophistiqués. Mais les réseaux sociaux comme Facebook vont également, et jour après jour, dans cette voie pour choisir et classer les informations qu'ils vont afficher sur le mur d'un utilisateur. Une connaissance approfondie de ces mécanismes devient donc essentielle dans le succès - ou l'échec - d'une campagne de promotion sur ce réseau social. Plongeons donc dans cet article pour une exploration de l'EdgeRank et du GraphRank, deux générations de systèmes d'évaluation utilisées par Facebook à cette fin...


$$\text{EdgeRank Facebook} = \sum_{\text{edges } e} u_e w_e d_e$$

Facebook est devenu un support incontournable pour ceux qui veulent développer la visibilité de leur marque, interagir avec leurs communautés ou faire connaître leur site. Mais le Net regorge d'histoires de campagnes ratées sur le leader des réseaux sociaux, certains allant jusqu'à renoncer à ce canal, jugé décevant.

En réalité, l'explication de ces échecs réside souvent dans une méconnaissance ou une mauvaise compréhension des algorithmes utilisés par Facebook pour sélectionner les items qui seront sélectionnés dans les flux d'actualités de chaque utilisateur. L'algorithme central dans ce processus s'appelle l'**EdgeRank**.

Qu'est-ce que l'EdgeRank ?

Un utilisateur moyen de Facebook est relié à des dizaines, quand ce ne sont pas des centaines de profils d'amis ou de connaissances. Il suit également plusieurs pages fans. Si chaque action produite sur toutes ces pages était montrée dans le flux d'actualités de l'utilisateur, ce dernier serait noyé sous un déluge d'informations.

Facebook cherche donc à filtrer et à classer ces informations. Pour y parvenir, il utilise une fonction d'évaluation : l'EdgeRank. Notons que "fonction d'évaluation" est un terme plus exact qu'algorithme dans ce contexte, même si désormais il est devenu habituel de l'appeler l'"algorithme Edgerank".

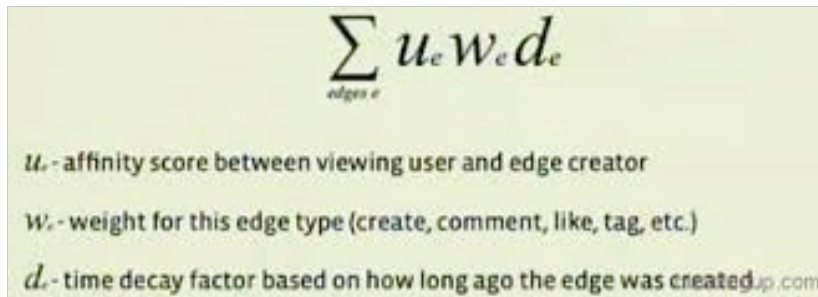
Ce qui est évalué dans l'EdgeRank, c'est le niveau d'importance de l'item à afficher dans les flux d'actualité. Notons que Facebook préfère appeler ces items des Objets.

Les Objets dont l'importance est jugée insuffisante sont éliminés et ne seront donc pas montrés à l'utilisateur.

Les critères utilisés dans l'EdgeRank

Il existe trois principaux critères utilisés dans la fonction d'évaluation :

- le niveau d'affinité (**Affinity**) ;
- le poids (**Weight**) ;
- l'érosion temporelle (**Time Decay**).



The image shows a screenshot of a presentation slide. At the top, the formula $\sum_{\text{edges } e} u_e w_e d_e$ is displayed. Below the formula, three lines of text define the variables: *u* - affinity score between viewing user and edge creator; *w* - weight for this edge type (create, comment, like, tag, etc.); *d* - time decay factor based on how long ago the edge was created. The source is cited as 'La Formule de l'Edgerank, tirée de la présentation au forum développeurs de Facebook F8 en 2010'.

La Formule de l'Edgerank, tirée de la présentation au forum développeurs de Facebook F8 en 2010

Le niveau d'affinité

Le niveau d'affinité mesure (comme son nom l'indique) l'étroitesse des relations entre l'utilisateur et l'Objet, ou la correspondance entre les profils. Pour un item émis par un ami dont vous n'avez pas consulté le profil depuis deux ans, et avec lequel vous n'avez rien échangé depuis longtemps, ni interagi d'aucune manière, le niveau d'affinité sera évalué comme très bas.

Présenté comme cela, on pourrait penser que le niveau d'affinité est considéré comme symétrique dans l'algorithme : si A est en forte affinité avec B, B est forcément en forte affinité avec A. Mais pas du tout. Le niveau d'affinité mesure en fait une "probabilité d'intérêt" de A pour B, ce qui ne garantit pas que B s'intéresse à A.

Le poids

Le poids d'un item combine la prise en compte de la nature de l'item et le mode et l'importance des interactions avec cet item. Facebook ne communique pas sur les poids réels attribués à certaines classes d'items, mais on considère généralement que les photos, les vidéos et les liens se voient attribués les niveaux de poids les plus importants. En ce qui concerne les modes d'interaction, un "Like" se voit attribuer un poids inférieur à celui d'un commentaire.

L'érosion temporelle

L'érosion temporelle est un facteur qui dépend de l'ancienneté de l'item. Sur Facebook, l'obsolescence des items est très rapide, les "news" disparaissent en quelques heures des flux d'actualités...

POSTS ON NEWS FEED (in order)	EDGE	AFFINITY	WEIGHT	TIME DECAY	EDGE VALUE
Brother Created recently NO ENGAGEMENT	Photo				HIGH
High School Friend Created 12 hours ago HIGH ENGAGEMENT	Status				LOW
	Mutual Friend Comment				MED
	Mutual Friend Comment				HIGH
	Mutual Friend Comment				HIGH
Brand Page Created 4 hours ago HIGH ENGAGEMENT	Photo				HIGH
	Fan Like				LOW
	Fan Share				HIGH
	Fan Comment				HIGH
Acquaintance Created 2 hours ago LOW ENGAGEMENT	Status				LOW

L'edgeRank en action : pourquoi certains Objets sont affichés et d'autres pas (illustration de Edgerankchecker)

Euh ... pourquoi cela s'appelle l'EdgeRank, au fait ?

Edge, dans ce contexte, se traduit par arête"... Il s'agit d'un terme typique de la théorie des graphes, une théorie mathématique qui connaît bien sûr de nombreuses applications sur le graphe social.

Un "edge" est ce qui relie deux points du graphe social. Toute relation entre deux points du graphe, quelle que soit sa nature, est qualifiée du terme générique "edge". Et toute interaction entre deux points du graphe produit une nouvelle relation, et une nouvelle "arête".

Sauf que Facebook n'appelle plus cet algorithme "Edgerank". Même si par habitude, Edgerank est un terme encore communément employé par les experts du réseau social et même en interne chez Facebook, un ingénieur a révélé en juillet dernier que ce terme avait été officiellement abandonné.

Comment connaît-on cet algorithme ?

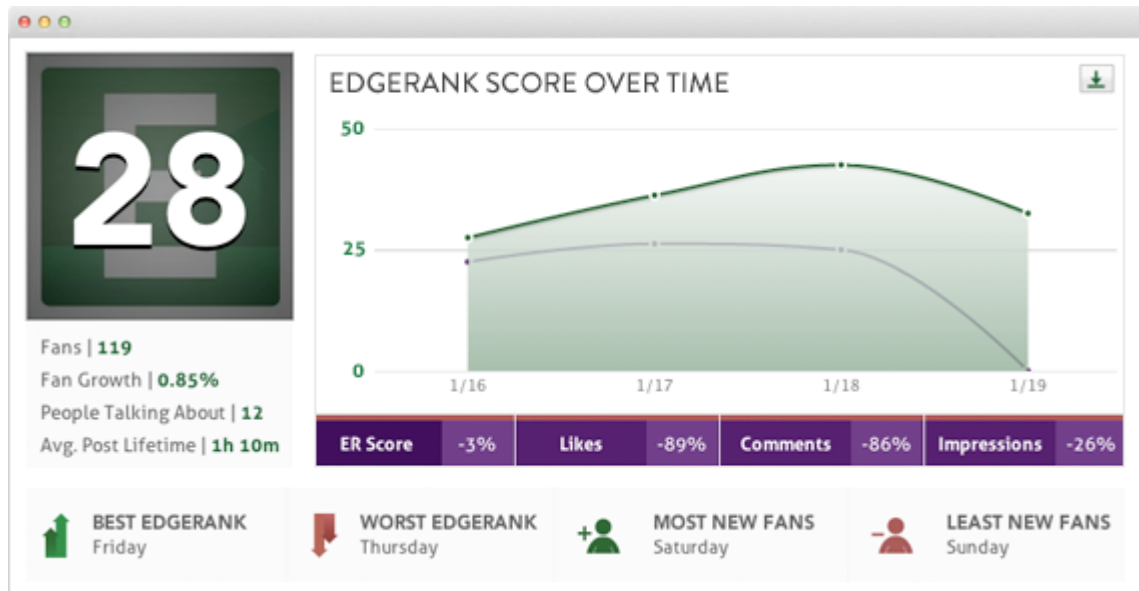
Ce sont deux ingénieurs de Facebook, Ruchi Sanghvi et Ari Steinberg qui ont présenté pour la première fois l'algorithme au cours de la conférence développeurs Facebook F8, le 21 avril 2010.

La vidéo de cette conférence n'est hélas plus accessible dans la zone développeurs de Facebook.

Peut-on estimer son Edgerank ?

La formule exacte de calcul de l'Edgerank est inconnue. Et Facebook ne communique pas les valeurs d'EdgeRank. On peut donc uniquement créer une fonction d'évaluation similaire et tenter de mesurer soi-même des critères comme le niveau d'affinité, le poids, et l'érosion temporelle des objets. Mais l'exercice a ses limites, et elles sont vite atteintes. Même si l'Edgerank n'est pas un algorithme sophistiqué, et même si la plupart des réseaux sociaux ont développé leur propre version de l'Edgerank (LinkedIn par exemple), on sait trop peu de choses sur la façon dont il est réellement calculé pour parvenir à prédire exactement ce qui va réellement apparaître dans les "newsfeeds".

Un outil (payant) disponible sur le Net fournit une évaluation des EdgeRanks aux propriétaires de pages Fan (par exemple) qui veulent suivre l'évolution de l'EdgeRank estimé. Cet outil (Edgerankchecker : <http://edgerankchecker.com/>) fournit une estimation intéressante qui permet de voir si ses efforts d'optimisation de l'Edgerank portent leurs fruits, mais il ne faut surtout pas prendre les valeurs produites par l'outil pour argent comptant.



Edgerankchecker, un outil qui sert à estimer l'Edgerank de ses objets

Le GraphRank

Au cours de la conférence développeurs F8 de 2011, Facebook a présenté un nouvel algorithme similaire au Edgerank, destiné à "sélectionner" les Objets à afficher produit par un nouveau type d'applications alimentant le graphe social en actions (comme le fait d'écouter un morceau de musique). Cet algorithme s'appelle le **GraphRank**. On sait peu de choses sur son mode de calcul réel, mais il semble reprendre les principes de l'Edgerank.

Un algorithme qui a été récemment fortement modifié

Au cours de l'automne 2012, un grand nombre d'observateurs ont constaté que le "reach" (la "couverture", le nombre d'utilisateurs qui voient s'afficher un post donné) des Objets qu'ils postaient sur Facebook avait brusquement chuté. Des rumeurs à propos d'un nouveau score de qualité ou d'autorité des pages ont commencé à se répandre.

En réalité, Facebook avait commencé à être confronté à des méthodes de spam "industrielles", et a décidé d'introduire un nouvel élément dans l'algorithme pour

minimiser la visibilité de ce spam. Ce nouveau score tient compte du "negative feedback" sur les pages. Une page qui a fait l'objet de nombreuses "plaintes" sera pénalisée, alors qu'une page qui a un nombre de plaintes plus bas que la moyenne sera favorisé.

Ce changement dans l'algorithme a été confirmé par Will Cathcart, product manager des "newsfeeds" chez Facebook, dans un entretien accordé au magazine en ligne

Techcrunch :

<http://techcrunch.com/2012/11/07/killing-rumors-with-facts-no-facebook-didnt-decrease-page-news-feed-reach-to-sell-more-promoted-posts/>

Comment la connaissance de cet algorithme permet-elle de réussir ses campagnes d'action sur Facebook ?

La formule de l'EdgeRank nous apprend pourquoi certaines marques font des « flops » sur Facebook même en dépensant des millions de dollars en achat de produits publicitaires sur Facebook.

Ce n'est pas tout d'avoir des fans ou des amis, il faut qu'ils soient « engagés »

Si vous n'avez pas réussi à créer un « engagement » de vos fans sur une page Facebook, c'est-à-dire si vos précédents posts et objets n'ont engendré aucune réaction et aucune interaction avec votre communauté, alors toute nouvelle publication risque de n'être visible que par un tout petit nombre d'utilisateurs, y compris si vous lâchez une information qui est une véritable « bombe » ou si vous publiez une vidéo virale particulièrement réussie.

Postez les bons objets, et des objets populaires

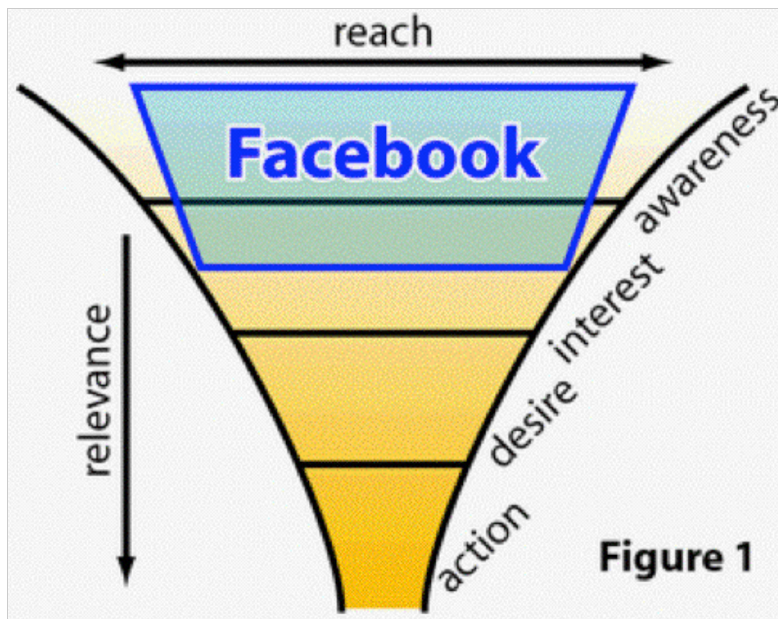
Images, vidéos et liens ont un poids important, c'est ce qu'il faut privilégier. Communiquer *via* un changement de statut vous permettra de toucher un nombre plus réduit de membres.

Avant l'heure, c'est pas l'heure, après l'heure c'est plus l'heure

Le facteur « time decay » a une implication évidente : si vous publiez vos posts ou si vous déclenchez une action quelconque un jour ou à une heure pendant lesquels un faible nombre des utilisateurs sont connectés, il y'a peu de chances que l'Objet Facebook en question soit vu.

Et ne faites pas n'importe quoi

Attention aux feedback négatifs : le temps où l'on pouvait faire n'importe quoi sur Facebook en passant sous le radar parce qu'il n'y avait pas assez d'évaluateurs humains pour éliminer les spammeurs est révolu. Même les comportements « limites » tolérés par Facebook sont maintenant pénalisés. Si vous faites l'objet de plaintes d'utilisateurs, le « reach » de vos campagnes risque de tomber à la portion congrue.



Source Michael WU : attention, le « reach » d'une campagne n'est pas tout, il faut aussi faire attention à ce que la publication soit « pertinente » pour les personnes touchées. En fait, plus le « reach » est important, plus la pertinence risque d'être dégradée. Il faut donc trouver un juste milieu...

Conclusion : le SMO s'est sophistiqué et demande aussi qu'on s'intéresse aux algorithmes du graphe social

Les applications du graphe social ont atteint un niveau de maturité et d'usage suffisant chez les utilisateurs pour obliger les principaux leaders du secteur à déployer des algorithmes destinés à trier, classer et filtrer ce qui est montré aux utilisateurs.

Le résultat est qu'il n'est plus possible de se contenter de connaître par cœur les fonctionnalités et les produits proposés par les réseaux sociaux pour savoir mener des campagnes efficaces sur ces supports. Il faut maintenant s'intéresser aux mécanismes au cœur du fonctionnement de ces outils. Pour le moment, nous ne sommes pas encore confrontés à des algorithmes d'une complexité comparable à ceux que l'on rencontre dans les moteurs de recherche, mais on s'en rapproche à grand pas...

Car les volumes d'information à traiter sont considérables, et les mathématiques décrivant des graphes aussi complexes que le graphe social très sophistiquées. Gageons qu'avec la recherche effrénée lancée sur le sujet au sein des outils sociaux et des universités, de nouvelles découvertes vont entraîner le monde du SMO vers des sphères inexplorées...

Bibliographie :

La page de facebook sur le sujet :
(plutôt laconique, voire elliptique)
<http://www.facebook.com/help/search/?q=edgerank>

L'outil Edgerankchecker :
<http://www.edgerankchecker.com>

L'article de Techcrunch sur l'entretien avec Will Cathcart :
<http://techcrunch.com/2012/11/07/killing-rumors-with-facts-no-facebook-didnt-decrease-page-news-feed-reach-to-sell-more-promoted-posts/>

Un article scientifique passionnant sur le sujet :

All Friends are Not Equal: Using Weights in Social Graphs to Improve Search

Sudheendra Hangal Diana MacLean Monica S. Lam Jeffrey Heer
Computer Science Department
Stanford University, Stanford, CA 94305
<http://xenon.stanford.edu/~hangal/weighted-social-graphs.pdf>

Le brevet sur l'Edgerank ?

Dynamically providing a news feed about a user of a social network
<http://www.google.com/patents/US7669123>
Brevet attribué à Facebook

Philippe YONNET , *Directeur de l'agence Search-Foresight / Groupe MyMedia.*
Président de l'association SEO Camp (<http://www.seo-camp.org/>)