

Moteurs de Recherche,  
Algorithmes de Recommandation II  
Cours Master 2, 2011

Michel Habib  
habib@liafa.jussieu.fr  
<http://www.liafa.jussieu.fr/~habib>

Février 2011

# Plan

Algorithmes de recommandation

Chasseurs Algorithmiques de têtes

Le mur de Facebook

On part du graphe biparti personnes – films les arêtes étant valuées par les notes.

Sur ce graphe on étudie les corrélations entre les notes, ceci pour chaque paire de personnes.

## Types de corrélations

- ▶ Profils inclus avec notes voisines
- ▶ Profils calculables à partir d'un autre à l'aide d'une formule (linéaire)
- ▶ Avec pour cas particulier, celui des Profils antinomiques (l'un adore ce que l'autre déteste)

A partir de ce graphe on peut calculer des un graphe orienté entre les personnes, les arcs étant typés par les correlations.  
On peut alors sur ce graphe calculer un ensemble de recommandations en suivant des chemins.

# Une thèse CIFRE



## PROPOSITION DE THESE CIFRE

### LE PROJET DE THESE

L'objectif de la thèse est le développement d'un algorithme de matching entre les offres d'emploi et les données utilisateurs contenues sur les réseaux sociaux LinkedIn et Facebook.

Plus particulièrement l'algorithme a pour but de :

- Proposer les offres d'emploi les plus pertinentes en fonction du profil d'un utilisateur,
- Suggérer des contacts susceptibles d'être intéressés par une offre d'emploi donnée,
- Optimiser la création de campagnes publicitaires ciblées sur le réseau Facebook.

Cette thèse est proposée dans le cadre d'une convention CIFRE.

### LA SOCIETE

**Work4 Labs** ([www.work4labs.com](http://www.work4labs.com)) est une société innovante créée en 2010 dont les locaux sont basés à Paris et à San Francisco.

Nous développons des applications Facebook pour aider les recruteurs à trouver des candidats sur ce réseau social. Notre application phare s'appelle **Work for Us** et permet aux entreprises de diffuser directement leurs offres d'emploi sur leur page entreprise Facebook et d'exploiter ainsi la plus grande base de données au monde.

En moins de 6 mois, plus de 4 500 clients utilisent notre application et pas moins de 1 000 nouveaux souscrivent chaque mois. Parmi eux, on compte L'Oréal, Sodexo, PwC, State Street, Accenture, Areva, Deloitte, KPMG, etc.

### PROFIL RECHERCHE

Master en Informatique

Des connaissances en statistique et apprentissage automatique sont un plus.

Le pôle emploi est confronté à une problématique du même genre :  
appairer demandeurs d'emploi et offres d'embauche

## Comment attaquer un tel problème

### Première étape

Identifier les personnes et les postes ayant le même profil (bien que les rubriques soient différentes)

Calcul symbolique ou numérique à l'aide d'une distance chiffrée de 0 à 1

distance à l'aide de mots clés ou vectorielle

Construire un graphe biparti  $G = (P, J, E)$  : personnes – jobs

On peut se servir de l'étape précédente pour construire le graphe  $G$   
 $xy \in E$  ssi  $dist(x, y) \leq k$  où  $k$  est un paramètre fixé.

## Formalisation du problème

- ▶ Chercher un couplage (un ensemble d'arêtes deux à deux disjointes) de taille maximum dans le graphe  $G$ .
- ▶ Il existe un algorithme en  $O((|P| + |J|)^{5/2})$  qui procède par améliorations successives (échanges sur des chaînes alternées impaires) à partir du couplage vide.

## Optimalité

- ▶ Cela dépend de la fonction de coût d'une affectation imposée par le problème,
- ▶ du modèle économique choisi (il se peut que le gain se fasse par la publicité et soit indépendant du nombre d'appariements)
- ▶ Ou que le gain soit proportionnel au nombre d'appariements (éventuellement pondéré par un % du salaire)
- ▶ Le modèle économique doit être différent pour Work4Labs et le pôle emploi !

## Difficultés

- ▶ Il est possible d'avoir à résoudre un problème de couplage maximum de poids maximum.
- ▶ Cela reste polynomial mais l'algorithme est plus compliqué.
- ▶ Le graphe peut être gigantesque et un algorithme non linéaire impossible à faire tourner.  
Dans ce cas, il va falloir inventer des heuristiques par exemple des algorithmes de couplages gloutons.

## Le mur de Facebook

Currently on Facebook, when you log in as a user, you are brought to a home page that has your Top News Feed. The Top News Feed is a stream of information from your social connections (Facebook Friends, Groups you have joined, and Pages that you have "liked") that Facebook thinks is most important to you. It is not every piece of information being posted by every one of your friends. You have access to that, of course, through the "Recent News" feed - but the default for users who log into Facebook is the "Top News" in their News Feed.

Il s'agit pour chaque individu  $F$  de Facebook de classer les événements qui se sont déroulés chez ses amis.

On ne considère que les amis qui ont fait qq chose. Il suffit d'ordonner ces événements, cela revient à ordonner le voisinage de  $F$ .

Pour ce faire, on utilise un algorithme qui construit un ordre total sur des arêtes issues de  $F$  : **EdgeRank**

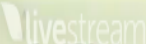
Cet algorithme est peu documenté.

## EdgeRank

Each Edge has three components important to Facebook's algorithm :

1. First, there is an affinity score between the viewing user and the item's creator. If you send your friend a lot of Facebook messages and check their profile often, then you'll have a higher affinity score for that user than you would, say, an old acquaintance you haven't spoken to in years.
2. Second, there is a weight given to each type of Edge. A comment probably has more importance than a Like, for example.
3. And finally there is the most obvious factor time. The older an Edge is, the less important it becomes.

## La version officielle



### 6. NFO: News Feed Optimization

#### EdgeRank

$$\sum_{\text{edges } e} u_e w_e d_e$$

$u_e$  - affinity score between viewing user and edge creator

$w_e$  - weight for this edge type (create, comment, like, tag, etc.)

$d_e$  - time decay factor based on how long ago the edge was created