

Devoir de programmation
À rendre le 17/4/2016

On propose de réaliser dans ce projet une étude *simplifiée* d'un réseau social *simplifié* de type Facebook© ou Twitter©. Pour mettre en place cette étude, nous utilisons quelques notions de la théorie des graphes qui nous permettront de représenter ces réseaux et de les analyser efficacement.

Comment représenter un réseau social ?

Définition 1. Graphe orienté

Un graphe orienté G est un schéma constitué d'un ensemble de points et d'un ensemble de flèches chacune reliant deux points. Les points sont appelés *sommets* du graphe. L'ensemble des sommets est noté \mathcal{V} . Les flèches sont appelées *arcs* du graphe. L'ensemble des arcs est noté \mathcal{A} . Soient $x, y \in \mathcal{V}$, on a $(x, y) \in \mathcal{A}$ si et seulement s'il existe un arc ayant pour origine x et pour extrémité y .

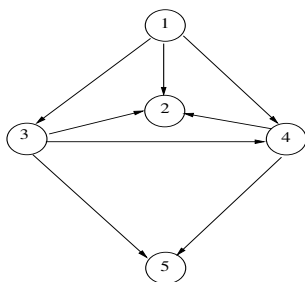


Figure 1.

Dans Figure 1, on observe que $\mathcal{V} = \{1, 2, 3, 4, 5\}$ et $\mathcal{A} = \{(1, 2), (1, 3), (1, 4), (2, 3), (2, 4), (3, 5), (4, 5), (3, 4), (4, 3)\}$

Définition 2. Degré entrant et degré sortant

Le degré entrant (respec. sortant) d'un sommet $v \in \mathcal{V}$ est le nombre d'arcs dont l'extrémité (respec. l'origine) est v . On le note $d^-(v)$ (respec. $d^+(v)$).

Dans Figure 1, on a $d^+(1) = 3$, $d^-(1) = 0$, $d^+(5) = 0$, $d^-(5) = 2$, $d^+(3) = 3$, $d^-(3) = 1$, etc...

Définition 3. Liste d'adjacence

$v \in \mathcal{V}$ est un voisin sortant (respec. entrant) de $u \in \mathcal{V}$ s'il existe un arc (u, v) (respec. $(v, u) \in \mathcal{A}$).

Pour $v \in \mathcal{V}$, on note $\mathcal{N}^+(v)$ (respec. $\mathcal{N}^-(v)$) l'ensemble des voisins sortants (respec. entrants) de v .

Dans Figure 1, $\mathcal{N}^+(1) = \{2, 3, 4\}$ et $\mathcal{N}^-(1) = \emptyset$, $\mathcal{N}^+(5) = \emptyset$ et $\mathcal{N}^-(5) = \{3, 4\}$, $\mathcal{N}^+(3) = \{2, 4, 5\}$ et $\mathcal{N}^-(3) = \{1\}$.

On peut choisir de représenter le graphe sous forme de liste d'adjacence. Cette liste donne pour chaque sommet $v_1, \dots, v_n \in \mathcal{V}$ la liste de ses voisins sortants $\mathcal{N}^+(v_1), \dots, \mathcal{N}^+(v_n)$ sous la forme $[(v_1, \mathcal{N}^+(v_1)); \dots; (v_n, \mathcal{N}^+(v_n))]$. La liste d'adjacence du graphe représenté dans la Figure 1 est comme suit :

$$[(1, \{2, 3, 4\}); (2, \emptyset); (3, \{2, 4, 5\}); (4, \{2, 5\}); (5, \emptyset)]$$

Comment modéliser le réseau social ?

Les utilisateurs du réseau social créent des comptes de deux types : *Utilisateur* et *Page*. Les comptes sont représentés par des sommets dans ce graphe.

- Chaque compte de type *Utilisateur* peut *suivre* d'autres comptes de type *Utilisateur* pour voir leurs notifications.
- Chaque compte de type *Utilisateur* peut *aimer* des comptes de type *Page* et ainsi devenir capable d'en consulter le contenu.
- En revanche, les comptes de type *Page* n'ont aucune action possible.

La relation x *suit* y où x et y sont des comptes *Utilisateur* est représentée par un arc du sommet x vers le sommet y dans le graphe.

La relation x *aime* y où x est un compte *Utilisateur* et y un compte *Page* est représentée par un arc du sommet x vers le sommet y dans le graphe.

Définition du graphe

Les comptes *Utilisateur* sont identifiés par un nom et ont des informations supplémentaires comme le prénom et l'âge.

Les comptes *Page* sont aussi identifiés par un nom et ont une liste d'administrateurs de type *Utilisateur*.

On demande de représenter le graphe par sa liste d'adjacence.

Opérations de gestion du graphe

Pour étudier les propriétés du graphe d'un réseau social, il faut d'abord être capable d'en explorer facilement le contenu et de disposer de quelques mesures simples dont les suivantes :

1. connaître le nombre de sommets, d'arcs,
 2. obtenir l'ensemble des sommets,
 3. obtenir l'ensemble des sommets trié par nom,
 4. obtenir l'ensemble des sommets trié par degré sortant,
 5. obtenir l'ensemble des arcs,
 6. ajouter/supprimer un sommet (et les arcs qui lui sont liés),
 7. ajouter/supprimer un arc,
 8. obtenir les informations d'un sommet en connaissant son nom,
 9. calculer le nombre de comptes de type *Utilisateur*, *Page*
 10. connaître l'âge moyen des *Utilisateur*,
 11. connaître tous les comptes *Utilisateur* qui sont des administrateurs de *Page*,
 12. afficher le graphe sous forme de liste d'adjacence,
 13. lire un graphe sous forme de liste d'adjacence.
- On pourra penser à coder une interface permettant de faire appel à l'une de ces fonctionnalités au choix.

Qui est le plus influent ?

De nombreuses méthodes sont utilisées pour mesurer l'influence ou la pertinence d'un utilisateur sur un réseau social, notamment sur Twitter®. Une méthode possible consiste à appliquer un algorithme de *PageRank*, également utilisé par Google® pour réaliser le classement des résultats

sur son moteur de recherche.

L'intuition de cet algorithme est celle-ci :

- plus le nombre de comptes qui me suivent est élevé, plus mon compte est influent et pertinent ;
- plus les comptes qui me suivent sont influents et pertinents, plus mon compte est influent et pertinent.

Calcul du PageRank

```
// Initialisation du PageRank à 1 pour tous les sommets
 $\forall v \in \mathcal{V}, PR(v) = 1$ 
// Calcul du PageRank
i = 0
TantQue  $i \leq 100$  faire
    // pour tous les sommets du graphe
    pour tous les  $v \in \mathcal{V}$  faire
        // on calcule le pageRank du sommet
 $PR(v) = \frac{0.15}{|V|} + 0.85 \sum_{u \in N^-(v)} \frac{PR(u)}{d^+(u)}$  //  $|V|$  = nombre de sommets du graphe
    i = i+1
FinTQ
```

Afin de déterminer les comptes *Utilisateur* et *Page* qui sont les plus influents au sein du graphe, cet algorithme devra être codé. Vous devrez donner la possibilité d'afficher les résultats obtenus.

Travail demandé :

On demande d'implanter le graphe et ses différentes fonctionnalités, ainsi que l'algorithme de calcul du PageRank proposé. L'interface utilisateur pourra être simple mais devra être suffisamment agéable.

Le projet est à réaliser par groupe de 3 étudiants maximum. On demande de rédiger un document comportant une analyse claire et détaillée du travail effectué, ainsi qu'un mode d'emploi de l'outil développé sous forme électronique en pdf. Ce rapport doit indiquer la répartition du travail dans le groupe et comporter quelques jeux d'essai. Le tout est à envoyer sous forme d'archive (.zip) par courrier électronique à votre chargé de TD :

Wadoud BOUSDIRA wadoud.bousdira@univ-orleans.fr
Frédéric DABROWSKI frederic.dabrowski@univ-orleans.fr