Analyse d'un réseau social

Résumé

Les réseaux sociaux tels que Twitter et Facebook sont des moyens de communication de plus en plus importants au sein de nos sociétés. Ils sont utilisés par des millions de gens et leur essor fait qu'ils suscitent beaucoup d'intérêt. Ils sont par exemple analysés d'un point de vue sociologique : études sur la diffusion de l'information dans ces réseaux, sur l'influence de ces réseaux virtuels sur les gens dans le monde réel [1], ou d'un point de vue marketing : comment diffuser l'information sur ses produits au mieux dans ces réseaux.

Nous proposons pour ce projet de réaliser une étude (simplifiée) d'un réseau social (simplifié) de type Facebook. Pour mettre en place cette étude, nous utiliserons quelques notions de théorie des graphes qui vont nous permettre de représenter ces réseaux et de les analyser efficacement.

1 Représenter les réseaux sociaux

1.1 Rappel de théorie des graphes

Definition 1 *(Intuitive) Un graphe orienté G est un schéma constitué par un ensemble de points et par un ensemble de flèches reliant chacune deux points. Les points sont appelés les sommets du graphe. L'ensemble des sommets est noté V. Les flèches sont appelées les arcs du graphe. L'ensemble des arcs est noté A. Soient $x, y \in V$, on a $(x, y) \in A$ si et seulement si il existe un arc avec pour origine x et pour extrémité y.*

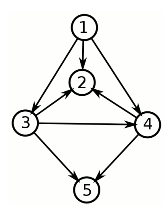

FIGURE 1 – Un graphe orienté

Dans la Figure 1, on observe l'ensemble des sommets $V = (1, 2, 3, 4, 5)$
et l'ensemble des arcs $A = ((1, 2), (1, 3), (1, 4), (3, 2), (3, 4), (3, 5), (4, 2), (4, 5))$

Definition 2 *On appelle degré entrant (resp. sortant) d'un sommet $v \in V$ le nombre d'arcs dont l'extrémité (resp. l'origine) est v. On le note $d^-(v)$ (resp. $d^+(v)$).*

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

1.2 La liste d'adjacence

Definition 3 *On dit que $v \in V$ est un voisin sortant (resp. entrant) de $u \in V$ s'il existe un arc (u, v) (resp. (v, u)) $\in A$.*

Definition 4 *Pour $v \in V$, nous définissons $N^+(v)$ (resp. $N^-(v)$) comme l'ensemble des voisins* sortants *(resp.* entrants*) de v.*

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

On peut choisir de représenter les graphes sous forme de liste d'adjacence. La liste d'adjacence est une liste qui donne pour chaque sommet $v1, \ldots, vn \in V$, la liste de ses voisins sortants $N^+(v1), \ldots, N^+(vn)$ sous la forme $((v1, N^+(v1)), \ldots, (vn, N^+(vn)))$. La liste d'adjacence du graphe représenté dans la Figure 1 est :

$$((1, (2, 3, 4)), (2, \emptyset), (3, (2, 5)), (4, (2, 5)), (5, \emptyset))$$

1.3 Le réseau à modéliser

On considère un réseau social. Les utilisateurs de ce réseau créent des comptes de deux types : *Utilisateur*, *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, avec x et y des comptes *Utilisateur* est représentée par un arc du sommet x vers le sommet y dans le graphe. La relation x aime y', avec x un compte *Utilisateur* et y' un compte *Page* est représentée par un arc du sommet x vers le sommet y' dans le graphe.

2 Gestion du graphe en Java

Les comptes *Utilisateur* sont identifiés par un nom et ont des informations supplémentaires comme le prénom et l'âge. Les comptes *Page* sont également identifiés par un nom et ont une liste d'administrateurs de type *Utilisateur*.

- Créer une classe abstraite *Sommet*, définir un constructeur ainsi que des accesseurs publics. Chaque *Sommet* contiendra la liste de ses voisins sortants.
- Ecrire les classes *Utilisateur* et *Page* qui héritent de la classe *Sommet*.
- Coder la classe *Graphe* qui devra représenter un graphe sous forme de liste d'adjacence et donc contenir un ensemble de *Sommet*.

3 Opérations de base sur le graphe

Pour commencer à étudier les propriétés du graphe d'un réseau social, il s'agit tout d'abord d'être capable d'en explorer le contenu facilement et de disposer de quelques mesures simples mais très utilisées :

- Connaître le nombre de sommets, d'arcs
- Obtenir l'ensemble des sommets
- Obtenir l'ensemble des sommets trié par nom
- Obtenir l'ensemble des sommets trié par degré sortant
- Obtenir l'ensemble des arcs
- Ajouter/Supprimer un sommet (et les arcs qui y sont liés)
- Ajouter/Supprimer un arc
- Obtenir les informations sur un sommet via son nom
- Connaître le nombre de comptes de type *Page*, *Utilisateur*
- Connaître l'âge moyen des *Utilisateur*
- Connaître tous les comptes *Utilisateur* qui sont des administrateurs de *Page*
- Ecrire le graphe dans un fichier sous forme de liste d'adjacence
- Lire un graphe dans un fichier sous forme de liste d'adjacence

Une interface devra être proposée dans votre application pour qu'un utilisateur puisse appeler toutes ces fonctionnalités, en visualiser les résultats et ainsi commencer à fouiller les données de son graphe.

4 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 est d'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 ;

Algorithm 1: Calcul du PageRank

```
// Initialization du PageRank à 1 pour tous les sommets
∀ v ∈ V, PR(v) = 1
// Calcul du pageRank
i := 0;
tant que i ≤ 100 faire
    // Pour tous les sommets du graphe
    pour tous les v ∈ V faire
        // On calcule le PageRank du sommet
        PR(v) = 0.15/|V| + 0.85 Σ_{u∈N⁻(v)} PR(u)/d⁺(u)   // | V |= nombre de sommets du graphe
        i := i + 1 ;
```

$$PR(v) = \frac{0.15}{|V|} + 0.85 \sum_{u \in N^-(v)} \frac{PR(u)}{d^+(u)}$$

Afin de pouvoir déterminer les comptes *Utilisateur* et *Page* qui sont les plus influents au sein du graphe, cet algorithme devra être implémenté dans votre application. Par ailleurs, l'interface utilisateur sera enrichie de la possibilité de lancer cet algorithme et d'en visualiser les résultats.

5 En pratique

Rapport Le projet est à réaliser par groupes de 3 étudiants. Il est demandé un rapport sous forme électronique en pdf. Ce rapport doit présenter le projet, son niveau d'avancement et son architecture (diagrammes de classe). Il doit aussi indiquer la répartition du travail dans le groupe. Le rapport est également l'occasion d'aborder les difficultés rencontrées durant le projet et de justifier les choix techniques mis en oeuvre pour y faire face.

Jeu de test Le projet devra être accompagné d'un fichier décrivant un graphe et servant de jeu de test. Ce fichier devra pouvoir être chargé au sein de votre application. Ainsi, les fonctionnalités de lecture et d'écriture du graphe dans un fichier sont prioritaires. Elles nous permettront en effet de charger le graphe qui nous servira à évaluer votre projet.

Rendu du projet Enfin, n'oubliez pas que toutes vos classes doivent être dûment commentées ! Le tout est à envoyer sous forme d'archive (.zip) par courrier électronique à votre chargé de TD avant le **6 janvier, minuit** :

Wadoud BOUSDIRA wadoud.bousdira@univ-orleans.fr
Nicolas DUGUÉ nicolas.dugue@univ-orleans.fr

6 Bonus : Des réseaux "Small world"

"Ah ! C'Que'l monde est p'tit". Cette affirmation est également connue comme le "paradoxe de Milgram" [2] du nom du sociologue qui l'a formalisée en 1967. Stanley Milgram clamait à l'époque que chaque être humain sur terre était relié à n'importe quel autre être humain via quelques (six) degrés de connaissance seulement.
Il est tentant de vouloir vérifier cette hypothèse au sein de réseaux sociaux planétaires réunissant des millions d'utilisateurs comme facebook. C'est ce que nous allons calculer avec l'algorithme qui suit.

Cet algorithme prend en paramètre un sommet du graphe (qu'on appelle source) et permet de calculer la plus petite distance entre ce sommet source et chacun des autres sommets du graphe. En exécutant cet algorithme avec pour source chacun des sommets du graphe, on peut connaître la plus petite distance entre chaque paire de sommets du graphe.

Afin de valider que le monde est petit au sein du graphe, implémenter cet algorithme et vérifier que la plus courte distance entre chaque paire de sommets est au plus égale à 6.

Algorithm 2: Plus courtes distances entre la source s et les sommets de V

// Initialisation des distances entre chaque sommet et la source s à 10000000
$\forall v \in V, dist(v) = 10000000$;
// Initialisation de la distance du sommet source à 0
$dist(s) = 0$;
// P prend l'ensemble des sommets du graphe
$P := V$;
// Calcul de la plus petite distance entre chaque sommet et la source
tant que $P \neq Vide$ **faire**
    // On prend le sommet u de P le plus proche de la source
    $u :=$ sommet de P avec dist minimale;
    Supprimer u de P;
    // Pour tous les voisins de ce sommet u
    pour tous les $v \in N^+(u)$ **faire**
        $alt := dist(u, s) + 1$;
        si $alt \leq dist(v, s)$ **alors**
            $dist(v, s) := alt$;

Références

[1] Sciences et Avenir. Facebook envoie les citoyens aux urnes, Septembre 2012. `http://sciencesetavenir.nouvelobs.com/fondamental/20120917.OBS2643/facebook-envoie-les-citoyens-aux-urnes.html`.

[2] Wikipedia. Etude du petit monde. `http://fr.wikipedia.org/wiki/Etude_du_petit_monde`.