Lif10 – Fondements des bases de données
TP7 – Graphes augmentés

Licence informatique – Automne 2014–2015

4-11 Décembre 2014

Résumé

Ce TP doit être réalisé seul ou en binôme sur deux séances [1]. Un compte rendu devra être déposé sur Spiral avant le vendredi 12 décembre 2014, 23h59.

Les graphes sont un puissant outil mathématique permettant de décrire de nombreux phénomènes réels, les sommets décrivant les entités, et les arêtes, les interactions entre ces entités. Par exemple, un réseau social peut être décrit sous forme de graphe où les sommets représentent les utilisateurs, et une arête entre deux utilisateur décrit une interaction. Un graphe augmenté est une structure enrichie de graphe. Outre, l'ensemble de sommets, et d'arêtes, nous disposons d'informations propres aux sommets, aux arêtes ainsi que des informations temporelles sur la dynamique du graphe :

– Les nœuds peuvent être décrits par des informations additionnelles. Par exemple, si un sommet décrit un utilisateur d'un réseau social, nous pouvons connaître son âge, sa catégorie socio-professionnelle, son lieu d'habitation, etc.
– De même, nous disposons d'informations supplémentaires pour caractériser l'interaction (arête) entre deux individus (sommets) : nature de l'interaction (amitié, relation professionnelle, familiale, etc.).
– Enfin, toutes ces informations ainsi que la structure du graphes peuvent évoluer au cours du temps. Par exemple, un sommet/arête peut apparaître/disparaitre, et les attributs décrivant les sommets et les arêtes peuvent évoluer au cours du temps.

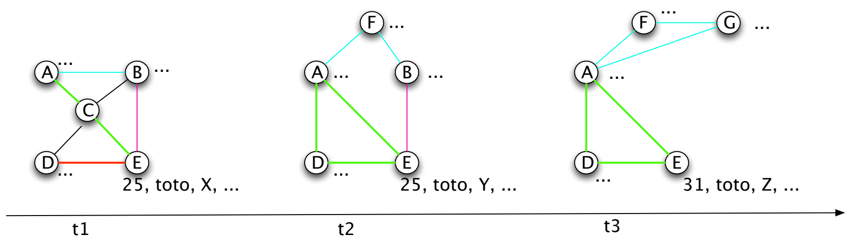

Figure 1 – Un exemple de graphe augmenté

Exercice 1 : Modélisation

1. Modélisez puis implantez un graphe augmenté en relationnel. On considérera qu'un nœud a au moins deux attributs de type chaine de caractères et deux autres de type numérique. Pour les arêtes, on supposera qu'elles sont décrites a minima par un attribut numérique et une chaine de caractères. Enfin, pour le temps, on considérera des estampilles temporelles (les t_i sont des entiers).
2. Peuplez vos relations afin de pouvoir faire les questions suivantes.

Exercice 2 : Opérations

Attention, pour les questions suivantes, il n'est pas toujours nécessaire d'utiliser du PL/SQL, une simple requête SQL peut parfois suffire.

1. Pour chaque pas de temps, identifiez les sommets isolés (connectés à aucun autre sommet).
2. Pour chaque sommet, déterminer le nombre de pas de temps dans lesquels il apparait ainsi que degré moyen.

1. Il se peut que des questions/données supplémentaires vous soient données en début de deuxième séance.

3. Degré de Centralité : cette mesure nous permet d'obtenir l'activité du nœud étudié à un pas de temps donné. En effet il constitue le rapport entre le nombre de liens sortant du nœud et le degré maximal possible. On obtient ainsi, pour un nœud appelé i et un nombre total de nœuds n dans le graphe à l'instant t : $Dc(i,t) = \frac{degree(i,t)}{(n-1)}$. Par exemple dans le graphe présenté précédemment, $Dc(C, t_1) = \frac{4}{4}$.
   1. Pour chaque nœud, retournez son degré de centralité pour chaque pas de temps (on considérera que sa centralité est -1 quand il n'apparait pas dans un graphe associé à un pas de temps.
   2. On souhaite affiner la mesure de centralité, et ne compter que les interactions (arêtes) d'un certain type : $Dc(i,t,type) = \frac{degree(i,t,type)}{(n-1)}$ où type est une condition sur les arêtes. Retournez pour chaque nœud, chaque type d'interaction (i.e., par rapport à l'attribut de type chaine de caractères) et chaque pas de temps, le degré de centralité associé.
4. Pour un pas de temps fixé, calculez le plus court chemin entre deux sommets.
5. Pour un pas de temps fixé et un sommet, retournez la composante connexe dont fait partie ce sommet.
6. Pour un pas de temps fixé et un sommet n_i, retournez tous les sommets accessibles par un chemin de longueur 2 à partir de n_i.
7. Même question en ne considérant qu'un type d'interaction spécifié.
8. Le coefficient de clustering [2] d'un graphe est une mesure du regroupement des nœuds dans un réseau. Plus précisément ce coefficient mesure à quel point le voisinage d'un sommet est connecté. Calculez le coefficient de clustering global (du graphe) et local (de chaque sommet) pour chaque pas de temps.
9. Le tau de Kendall (ou τ de Kendall) est une statistique qui mesure l'association entre deux variables. Plus spécifiquement, le tau de Kendall [3] mesure la corrélation de rang entre deux variables.
   1. Pour chaque pas de temps, calculez le tau de Kendall entre les deux attributs numériques associés aux sommets.
   2. Pour chaque pas de temps, calculez le tau de Kendall entre chacun des attributs numériques associés aux sommets et le degré de centralité des sommets.
10. Etant donné un ensemble de sommet X et un pas de temps t, déterminez si X est une clique à l'instant t.
11. **Bonus :** définissez et implantez d'autres opérations sur de tels graphes qui vous semblent pertinentes.

2. `http://fr.wikipedia.org/wiki/Coefficient_de_clustering`
3. `http://fr.wikipedia.org/wiki/Tau_de_Kendall`