

Jeux de cubes pour les graphes...

Ou comment des graphes enrichis par des cubes (GreC) peuvent contribuer à l'analyse de données textuelles ?

Cécile Favre*, Wararat Jakawat**, Sabine Loudcher*

*Université de Lyon, Université Lyon 2, ERIC EA 3083, France
{cecile.favre|sabine.loudcher}@univ-lyon2.fr

**Computer Science Department, Prince of Songkhla University, Thailand
wararat.j@psu.ac.th

Les données textuelles constituent une part importante des données porteuses de connaissances à découvrir. Ces données présentent souvent deux caractéristiques : d'une part, le fait que différents axes d'analyse peuvent être pertinents pour résumer l'information ; d'autre part la possibilité de représenter en réseau certaines informations. La première caractéristique sous-tend une analyse multidimensionnelle de type OLAP (OnLine Analytical Processing) très liée aux entrepôts de données avec notamment une représentation sous forme de cube de données, dont les cellules contiennent une mesure (indicateur) permettant l'analyse d'un fait, en fonctions de dimensions (axes d'analyse) qui définissent le cube. Ce type d'analyse se base sur des opérateurs de navigation dans les données. La deuxième caractéristique relève d'une approche d'analyse/de visualisation de graphe.

Pour illustrer cela, un premier exemple se trouve dans l'analyse de données textuelles issues de messages Twitter : d'une part le nombre de messages peut être analysé en fonction de leur auteur, de leur temporalité, du sujet traité, etc. ; d'autre part, il peut être utile de s'intéresser au réseau des Twittos où les arêtes représentent par exemple les liens d'abonnement. Un second exemple se situe dans le contexte de l'analyse de la production scientifique. Il peut être pertinent d'analyser cette production au travers du nombre de papiers publiés selon différents axes que seraient leurs auteurs, l'année de publication, les thématiques, etc. De plus, il est intéressant d'analyser le graphe des auteurs pour s'intéresser au réseau de collaborations (co-publications), ou au réseau de citations. Dans cette communication, nous utiliserons ce second exemple de données bibliographiques pour illustrer nos propos.

Considérant la double caractéristique des données textuelles évoquée précédemment, et afin de tirer ainsi parti de ces deux visualisations (graphe et cube), un nouveau type d'analyse est apparu : *Graph OLAP* (Chen et al., 2008). L'idée, sur laquelle elle repose, consiste à construire un cube de graphes dans lequel il est possible de naviguer. Plus précisément, dans cette approche de *Graph OLAP*, il s'agit de considérer des cubes définis selon des dimensions dites informationnelles, et la mesure contenue dans les cellules correspond à des graphes ou plus exactement à des sous-graphes, qui peuvent présenter des dimensions dites topologiques. Ainsi, les différentes approches relevant de *Graph OLAP* permettent de visualiser des "instantanés" de graphes en fonction des dimensions d'analyse choisies. Différents opérateurs ont été

Jeux de cubes pour les graphes

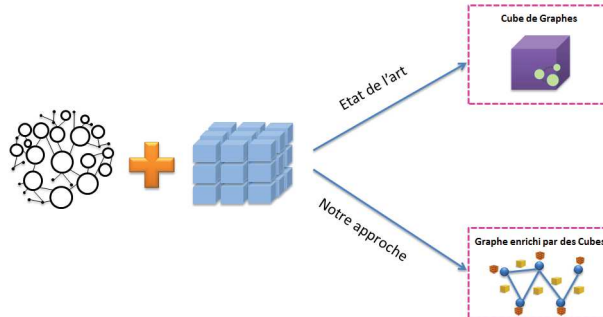


FIG. 1 – *Combinaison des graphes et des cubes OLAP*

proposés pour naviguer dans le cube de graphes : des opérations informationnelles ou topologiques, selon si les opérations s’appliquent selon les dimensions du cube ou les dimensions des graphes.

Cependant, dans cette combinaison de l’OLAP et des graphes basée sur des cubes de graphes, la visualisation plus globale du graphe est perdue, alors même que celle-ci est intéressante d’un point de vue analytique. Parallèlement, la dynamique des données est importante pour l’analyse du graphe, et ceci n’est pas toujours bien perceptible dans la visualisation des parties de graphe.

Par conséquent, nous proposons l’approche *GreC* (Graphes enrichis par des Cubes) qui est une nouvelle façon de considérer la combinaison de l’OLAP et des graphes pour l’analyse de réseaux d’information (Jakawat et al., 2016). *GreC* est une approche originale et complémentaire des approches basées sur une construction d’un cube de graphes (cf. figure 1). Elle permet de construire un graphe qui réponde aux besoins d’analyse de l’utilisateur et de l’enrichir avec des cubes de données qui vont décrire et valuer les nœuds et/ou les arêtes selon les besoins d’analyse. L’utilisateur peut ainsi avoir une vue globale du réseau avec des informations multi-dimensionnelles et faire des analyses intéressantes en naviguant au sein du graphe enrichi avec des opérateurs dédiés informationnels et topologiques.

Cette communication vise alors à montrer en quoi l’approche *GreC* permet une analyse pertinente des données textuelles, en explicitant cette approche avec une illustration sur le cas des données bibliographiques.

Références

- Chen, C., X. Yan, F. Zhu, J. Han, et P. S. Yu (2008). Graph OLAP : Towards online analytical processing on graphs. In *8th IEEE International Conference on Data Mining (ICDM’08)*, pp. 103–112.
- Jakawat, W., C. Favre, et S. Loudcher (2016). Graphs enriched by cubes for OLAP on bibliographic networks. *International Journal of Business Intelligence and Data Mining (IJ-BIDM’16)* 11(1), 85–107.