

eDOI : exploration itérative de grands graphes multi-couches basée sur une mesure de l'intérêt de l'utilisateur

Antoine Laumond*, Norbert Feron*, Guy Melançon*, Bruno Pinaud*

*Université de Bordeaux, CNRS UMR 5800 LaBRI
{prenom.nom}@u-bordeaux.fr

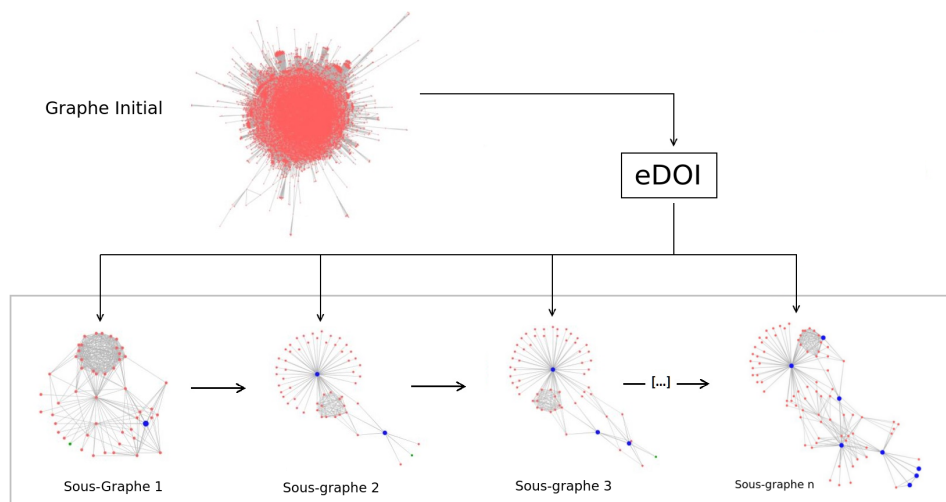


FIG. 1 – *eDOI* : création d'une série de sous-graphes basés sur une mesure d'intérêt. Les sommets bleus sont sélectionnés par l'utilisateur. Chaque sélection utilisateur permet la création d'un nouveau sous-graphe d'intérêt supérieur. Ce dernier est issu du graphe initial et prend en compte les informations sélectionnées dans les sous-graphes précédents.

Bien souvent les représentations nœuds-liens d'un réseau sont difficilement lisibles à cause d'une grande taille ou d'une topologie dense ou complexe. Il s'avère dès lors important de s'écarter de la stratégie classique proposant en première intention une vue d'ensemble du graphe. Dans cette optique, nous proposons une stratégie d'exploration appelée *eDOI* basée sur un calcul de sous-ensembles de nœuds intéressants pour l'utilisateur intégrant à la fois un aspect sémantique et prenant en compte la structure multi-couches potentielle de ces réseaux (i.e. des nœuds d'un type donné pouvant être regroupés en sous-réseaux (Kivelä et al., 2014)).

Dans le scénario que nous considérons, l'utilisateur mène une activité d'exploration en ayant une vue partielle des données d'une manière comparable aux moteurs de recherche qui déterminent les éléments les plus pertinents au vu d'une requête utilisateur. Notre méthode

eDOI permet cependant d'améliorer l'interactivité en autorisant l'utilisateur à affiner les recommandations obtenues. L'utilisateur agit pour cela sur un sous-ensemble du graphe initial. Ce sous-graphe donne un contexte aux réponses de la requête tout en étant un dispositif pour affiner le résultat obtenu. En y sélectionnant de nouveaux éléments (appelés focus), l'utilisateur permet d'affiner la définition de ses objectifs afin de créer un nouveau sous-graphe d'intérêt supérieur. Doté de cette visualisation et de ce moyen d'action, l'utilisateur se trouve donc lui-même aux commandes du moteur de recommandations.

La création itérative de sous-graphes (Fig. 1) est basée sur un calcul d'intérêt dérivé des travaux de Van Ham et Perer (2009). Avant chaque création d'un nouveau sous-graphe, un score est calculé pour chaque nœud du graphe initial en fonction de ses informations topologiques et sémantiques (attributs, couches du nœud et sélection utilisateur). À partir de ce score, un algorithme glouton permet l'extraction du nouveau sous-graphe à partir des nœuds focus. La formule utilisée est la suivante :

$$\mathbf{eDOI}(x|Y, L, z) = \alpha \cdot API(x) + \beta \cdot UI_{multi}(x, z(Y, L(x))) + \gamma \cdot D_{exp}(x, Y)$$

$API(x)$ ("A Priori") donne un score d'intérêt basé sur la topologie du graphe (degré, centralité, etc.) et peut être pré-calculé à priori avant même la sélection utilisateur.

UI ("User Interest") indique l'importance sémantique de l'élément. Il se base sur une requête utilisateur z reflétant la volonté utilisateur en fonction de sa sélection (les sommets focus Y) et des couches L auxquelles appartient x . Le DOI original de Van Ham et Perer (2009) n'avait pas pour vocation de faire évoluer l'objectif de l'utilisateur. Avec notre approche, il est nécessaire que la fonction qui capture l'intérêt sémantique de l'utilisateur z évolue en fonction de ses choix. L'ensemble Y incarne le choix de l'utilisateur et doit donc être utilisé par z afin que le score d'intérêt fluctue en fonction de Y . Il est important de différencier les traitements en fonction des couches car celles-ci n'auront pas nécessairement le même apport pour l'utilisateur. Si le graphe admet des sommets ayant plusieurs types (et donc appartenant à plusieurs couches), z attribue une fonction unique pour chaque couche mais peut aussi attribuer une fonction unique à un n-uplet de couches de x .

Enfin, D ("Distance") est un score basé sur la distance moyenne entre x et chaque sommet focus y de l'ensemble Y . Il est possible d'utiliser un calcul de centroïde sur les sommets de Y auquel cas D est basée sur la distance entre x et ce centroïde. L'ensemble Y permet ainsi de définir une zone dite focale. Plus un sommet est proche du centre de cette zone, plus son score sera élevé. Avec la modification incrémentale de Y , la zone focale se déplace dans le graphe et fait ainsi varier D en fonction des choix de l'utilisateur. Finalement, l'utilisateur peut naviguer au sein du graphe simplement par l'action de ses sélections de sommets dans les sous-graphes qui lui sont proposés.

Références

- Kivelä, M., A. Arenas, M. Barthelemy, J. P. Gleeson, Y. Moreno, et M. A. Porter (2014). Multilayer networks. *J. of Complex Networks* (2), 203–271.
- Van Ham, F. et A. Perer (2009). "search, show context, expand on demand" : Supporting large graph exploration with degree-of-interest. *IEEE Trans. on Vis. and Comp. Graph.* 15(6), 953–960.