

Énumération Randomisée des Triangles dans des Graphes à Grande Echelle à base de SQL

Abir Farouzi^{*,***}, Ladjel Bellatreche^{*}, Carlos Ordonez^{**}
Gopal Pandurangan^{**}, Mimoun Malki^{***}

^{*}ISAE-ENSMA, Poitiers, France
(abir.farouzi, bellatreche)@ensma.fr

^{**} University of Houston, Texas, USA
carlos@central.uh.edu, gopal@cs.uh.edu

^{***}ESI-SBA, Sidi Bel Abbès, Algérie
(a.farouzi, m.malki)@esi-sba.dz

Résumé. Chercher tous les triangles dans un graphe donné est un problème théorique classique avec un nombre important d'applications pratiques. Plusieurs algorithmes d'énumération de triangles dans des graphes à grande échelle ont été proposés et implémentés dans des environnements centralisés/distribués avec des langages comme C/Python. Dans le contexte des bases de données, il existe une grande masse de données qui peuvent être modélisées et analysées sous forme de graphes. Offrir des solutions d'énumération de triangles avec des requêtes SQL pour ces graphes représente un enjeu important pour la communauté. Dans cet article, nous proposons un algorithme distribué randomisé implémenté à l'aide de requêtes SQL pour l'énumération des triangles dans des graphes à grande échelle. Cet algorithme assure l'équilibrage de charge entre les processeurs grâce à sa stratégie de partitionnement visant à éliminer les échanges coûteux de données entre les hôtes lors de l'énumération de triangles. Nos expériences ont montré le passage à l'échelle de notre approche déployée sur un cluster de 8 hôtes et hébergeant le SGBD Vertica.

1 Introduction

Les graphes deviennent omniprésents du fait que notre monde est de plus en plus interconnecté. L'un des problèmes fondamentaux des graphes est l'énumération des triangles. Il consiste à lister tous les triangles d'un graphe. Ce problème a suscité beaucoup d'intérêt en raison de ses nombreuses applications pratiques : l'analyse des processus sociaux dans les réseaux (Watts et Strogatz, 1998), l'extraction de sous-graphes denses (Wang et al., 2010), les jointures dans les bases de données (Ngo et al., 2013). Ainsi, d'autres applications sont largement commentées dans (Chu et Cheng, 2012; Berry et al., 2015). En outre, il est important de souligner que les processus de détection et de comptage de triangles sont plus faciles que leur énumération qui est une tâche cruciale pour plusieurs applications. En effet, l'énumération teste la possibilité de former un triangle à partir des triplets d'arêtes (Farouzi et al., 2020). Donc, en utilisant le résultat de l'énumération nous pouvons facilement obtenir le nombre des triangles dans le graphe. En revanche, un simple comptage ne fournit pas forcément la liste des