

**CẢI TIẾN BÀI TOÁN ĐẾM TAM GIÁC TRÊN HADOOP BẰNG PHÂN VÙNG
DỰA TRÊN ĐỒ HÌNH THẬT SỰ**

Trần Duy Quang

Khoa Công nghệ Thông tin, Trường ĐH KHTN, ĐHQG-HCM

Tóm tắt

Bài toán đếm tam giác không chỉ là cơ sở cho việc phân tích đồ thị và đồ thị con mà còn đóng một vai trò quan trọng trong bài toán phân tích các mạng phức tạp. Khi đồ thị có kích thước lớn, Hadoop MapReduce thường là giải pháp phân tán được sử dụng. Tuy nhiên, vẫn còn có nhiều hạn chế về hiệu năng khi xem xét đến việc phân chia dữ liệu đầu vào, tuần tự hóa và phân bố đồ thị. Để giải quyết vấn đề này, chúng tôi sử dụng việc phân vùng dữ liệu đồ thị đầu vào dựa trên đồ hình thực tế. Các thí nghiệm cho thấy bằng cách phân vùng dữ liệu đầu vào dựa trên đồ hình thực tế có thể cải tiến hiệu năng của bài toán đếm tam giác.

**TRIANGLE COUNTING ON HADOOP IMPROVEMENT USING PARTITIONING
BASED ON ACTUAL TOPOLOGY**

Abstract

Triangle counting is not only the basis for graph and subgraph analysis, but it also plays a significant role in complex network analysis and metrics computing such as clustering coefficient and transitivity ratio. As the graph grows in size, Hadoop MapReduce is commonly used as a popular distributed solution. However, the performance is still limited in consideration to partitioning, serializing and distributing the graph. To address these issues, we apply the partitioning of graph input based on actual topology. Experiments show that using input partitioning based on topology can improve the performance of the Triangle counting problem.