

NÂNG CAO TỐC ĐỘ TRUY XUẤT DỮ LIỆU CHO ỨNG DỤNG DỰA TRÊN CƠ SỞ DỮ LIỆU ĐỒ THỊ

Phạm Thành Đạt¹, Nguyễn Trần Minh Thu²

¹Khoa Công nghệ thông tin, Trường Đại học Khoa học Tự Nhiên, ĐHQG-HCM

²Trường Đại học Khoa học Tự Nhiên, ĐHQG-HCM

phamthanhdatt.cdsp@longan.edu.vn, ntmthu@fit.hcmus.edu.vn

Tóm tắt

Cơ sở dữ liệu đồ thị ngày càng được sử dụng phổ biến để lưu trữ dữ liệu cho các ứng dụng hệ thống thông tin. Việc áp dụng cơ sở dữ liệu đồ thị có thể làm tăng khả năng trực quan hóa dữ liệu, giải quyết các vấn đề về truy xuất dữ liệu có mức liên kết phức tạp và tăng hiệu suất truy xuất. Trong bài báo này, chúng tôi trình bày kỹ thuật áp dụng cơ sở dữ liệu đồ thị để quản lý và kiểm tra hiệu suất truy xuất trong hệ thống gợi ý bán hàng. Thực nghiệm được xây dựng dựa trên cơ sở dữ liệu quan hệ và cơ sở dữ liệu đồ thị. Kết quả cho thấy việc lưu trữ trên cơ sở dữ liệu đồ thị giảm khoảng 15% số lượng bảng lưu trữ, 65% số bộ dữ liệu và đặc biệt là hiệu suất xử lý dữ liệu cao hơn khá nhiều so với cơ sở dữ liệu quan hệ. Kết quả này đã minh chứng được những lợi thế của việc áp dụng cơ sở dữ liệu đồ thị vào các ứng dụng hiện nay.

Từ khóa: NoSQL, cơ sở dữ liệu đồ thị, mô hình đồ thị, hệ thống gợi ý.

1. GIỚI THIỆU

Ngày nay, Internet là một phần không thể thiếu trong cuộc sống hằng ngày bởi những lợi ích mà nó mang lại cho chúng ta như: tìm kiếm thông tin, nghe nhạc, xem phim, mua sắm trực tuyến, trò chuyện hoặc kết nối với nhiều người,... Tuy nhiên, lượng thông tin trên Internet là rất lớn và đang tăng lên một cách nhanh chóng, do đó việc tìm kiếm, lọc và truy xuất thông tin, tài liệu, sách báo, phim, nhạc,... (từ đây gọi chung là “sản phẩm”) sao cho phù hợp với mỗi cá nhân người dùng như việc “mò kim đáy bể”, rất khó khăn, mất nhiều thời gian và đôi khi không đạt được kết quả như mong muốn. Để khắc phục tình trạng này, các ứng dụng hệ thống thông tin đã ra đời, đặc biệt là sự ra đời của các hệ thống gợi ý – một loại ứng dụng hệ thống thông tin thuộc lĩnh vực “Lọc thông tin”, không chỉ hỗ trợ trong việc tìm kiếm thông thường mà còn tìm kiếm và chủ động gợi ý danh sách sản phẩm sao cho phù hợp với nhu cầu của mỗi cá nhân người dùng [11].

Hiện nay, đa số các hệ thống gợi ý đều được xây dựng dựa trên cơ sở dữ liệu quan hệ và đang hoạt động rất tốt. Tuy nhiên, kích thước dữ liệu đang tăng lên một cách nhanh chóng, đặc biệt là sự kết nối của dữ liệu ngày càng nhiều và phức tạp, điều này làm cho hiệu suất truy xuất dữ liệu bị giảm đi đáng kể. Bên cạnh đó, nhu cầu của người dùng cũng phức tạp hơn và đòi hỏi tốc độ truy vấn phải nhanh hơn, đặc biệt là kết quả truy vấn phải mang tính cá nhân. Do đó, cơ sở dữ liệu quan hệ phải mất nhiều bộ nhớ và thời gian hơn để lưu trữ và xử lý dữ liệu, đồng nghĩa chi phí phải trả sẽ tăng lên và hiệu suất truy xuất sẽ giảm xuống. Chính vì vậy, việc tìm ra một loại cơ sở dữ liệu mới để lưu trữ dữ liệu, giải quyết các vấn đề về truy xuất dữ liệu có mức liên kết dữ liệu phức tạp và tăng hiệu suất truy xuất trong việc xây dựng các hệ thống gợi ý là rất cấp thiết. Một trong những loại cơ sở dữ liệu mới có thể giải quyết được những vấn đề này là cơ sở dữ liệu đồ thị.

Trong bài báo này, chúng tôi trình bày kỹ thuật quản lý dữ liệu trong một hệ thống gợi ý bán hàng và xây dựng các câu truy vấn trên 2 mô hình: cơ sở dữ liệu quan hệ và cơ sở dữ liệu đồ thị, sau đó là thực thi truy vấn để tiến hành đánh giá và so sánh hiệu suất truy xuất của chúng. Cuối cùng, với kết quả thực nghiệm, chúng tôi minh chứng việc xây dựng các hệ thống gợi ý, cũng như các ứng dụng hệ thống thông tin dựa trên cơ sở dữ liệu đồ thị sẽ mang lại nhiều ưu thế hơn, đặc biệt là hiệu suất truy xuất dữ liệu sẽ cao hơn so với cơ sở dữ liệu quan hệ.

Cấu trúc bài báo được trình bày như sau: đầu tiên, trong phần 2, chúng tôi giới thiệu những nghiên cứu có liên quan và mô tả về bài toán mà chúng tôi dự kiến thực nghiệm; tiếp theo là trình bày kết quả của thực nghiệm trong phần 3; và cuối cùng, từ kết quả thực nghiệm, chúng tôi đưa ra kết luận cùng một số hướng phát triển trong tương lai ở phần 4.

2. NHỮNG NGHIÊN CỨU LIÊN QUAN VÀ MÔ TẢ BÀI TOÁN

Cơ sở dữ liệu đồ thị (Graph Database) là một loại cơ sở dữ liệu phi quan hệ (hay còn gọi là NoSQL) được xây dựng dựa trên nền tảng lý thuyết đồ thị và có 4 thành phần chính: nút (đại diện cho các đối tượng khác nhau), mối quan hệ (kết nối các nút với nhau), thuộc tính (mô tả đặc tính cho nút hoặc mối quan hệ) và nhãn (gom nhóm các nút tương đồng) [26]. Cơ sở dữ liệu đồ thị có thể phản ánh và mô tả được những

mối quan hệ giữa các đối tượng một cách tự nhiên, đơn giản và dễ hiểu nhất. Đặc biệt, bằng cách áp dụng các phương pháp duyệt đồ thị, cơ sở dữ liệu đồ thị có thể xác định hoặc loại bỏ các nút, các mối quan hệ, cũng như các nút liền kề rất dễ dàng và nhanh chóng, chính vì vậy mà hiệu suất truy xuất dữ liệu cũng được cải thiện khá nhiều.

Với tất cả những ưu điểm đó, cơ sở dữ liệu đồ thị đã trở thành một lựa chọn hàng đầu của nhiều nhà phát triển trong các lĩnh vực có mức độ liên kết dữ liệu phức tạp như: mạng xã hội, thương mại điện tử, phát hiện gian lận, quản lý tài sản, trí tuệ nhân tạo, IOT,... đặc biệt là lĩnh vực hệ thống gợi ý – một lĩnh vực có nhu cầu sử dụng rất cao trong thời đại công nghệ số như hiện nay, cụ thể có: hệ thống gợi ý hàng hóa của tập đoàn Walmart, hệ thống gợi ý việc làm của CareerLink, hệ thống gợi ý việc làm của InfoJobs,... Bên cạnh đó, cũng có rất nhiều nghiên cứu về việc xây dựng các hệ thống gợi ý dựa trên cơ sở dữ liệu đồ thị này, ví dụ như: “Kiến trúc và tập dữ liệu của hệ thống gợi ý bài báo khoa học Docear (The Architecture and Datasets of Docear's Research Paper Recommender System)” (B. Joeran, L. Stefan, G. Bela và N. Andreas, 2014) [15]; “Hệ thống gợi ý bạn bè trên mạng xã hội sử dụng cơ sở dữ liệu đồ thị (Social Network Recommendation using Graph Databases)” (Š. Jan, 2015) [14]; “Phân tích dữ liệu bán lẻ bằng cơ sở dữ liệu đồ thị (Retail data analytics using Graph Database)” (P. Rashmi, 2018) [23],...

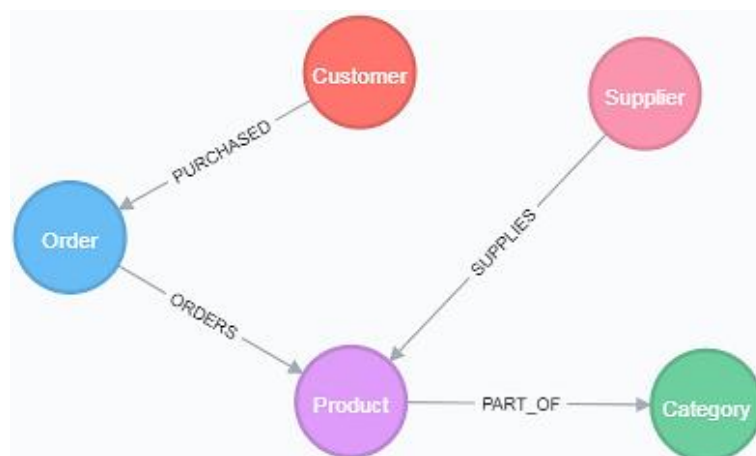
Nhìn chung, thông qua việc tìm hiểu những nghiên cứu trên, chúng tôi nhận thấy cơ sở dữ liệu đồ thị mang lại khá nhiều lợi ích trong việc xây dựng các hệ thống gợi ý, nổi trội nhất là khả năng trực quan hóa và khả năng làm tăng hiệu suất truy xuất dữ liệu. Song song đó, chúng tôi cũng băn khoăn việc xây dựng các hệ thống gợi ý dựa trên cơ sở dữ liệu đồ thị có thật sự mang lại nhiều ưu thế hơn so với cơ sở dữ liệu quan hệ truyền thống hay không? Chính vì vậy, chúng tôi đã dự kiến xây dựng một hệ thống gợi ý bán hàng trên cả 2 mô hình cơ sở dữ liệu này để tiến hành thực nghiệm và tìm ra câu trả lời với các bước như sau: trước tiên, chúng tôi lựa chọn một tập dữ liệu mẫu để xây dựng cơ sở dữ liệu thực nghiệm trên cả 2 mô hình; tiếp theo, chúng tôi dựa vào những yêu cầu của một hệ thống gợi ý bán hàng để xây dựng các câu truy vấn; và sau đó là thực thi chúng để kiểm tra, đánh giá và so sánh những ưu khuyết điểm, đặc biệt là hiệu suất truy xuất dữ liệu của 2 mô hình cơ sở dữ liệu; cuối cùng, kết quả nhận được là việc lưu trữ trên cơ sở dữ liệu đồ thị giảm khoảng 15% số lượng bảng lưu trữ,

65% số bộ dữ liệu và đặc biệt là hiệu suất xử lý dữ liệu cao hơn khá nhiều so với cơ sở dữ liệu quan hệ trong việc xây dựng các hệ thống gợi ý. Như vậy, để đạt được kết quả này, chúng tôi đã thực hiện một số bước như trên và trình tự thực hiện các bước đó cũng chính là trình tự của nội dung mà chúng tôi sẽ trình bày tiếp theo.

Đầu tiên là việc lựa chọn tập dữ liệu và tập dữ liệu mà chúng tôi lựa chọn để thực nghiệm trong bài báo này là NorthWind, một tập dữ liệu mẫu về bán hàng gồm 13 bảng dữ liệu và thường được sử dụng trong các nghiên cứu về vấn đề dữ liệu. Như dự kiến, chúng tôi sẽ xây dựng tập dữ liệu này trên 2 mô hình: cơ sở dữ liệu quan hệ và cơ sở dữ liệu đồ thị.

Đối với việc tạo tập dữ liệu trên cơ sở dữ liệu quan hệ, chúng tôi sẽ không đề cập đến vì đây là một việc rất quen thuộc. Tuy nhiên, trong nghiên cứu này, chúng tôi chỉ tập trung vào những bảng dữ liệu có liên quan đến thực nghiệm. Do đó, tập dữ liệu thực nghiệm trên cơ sở dữ liệu quan hệ của chúng tôi chỉ bao gồm 6 bảng dữ liệu: Products (có 77 bộ dữ liệu), Categories (có 8 bộ dữ liệu), Suppliers (có 29 bộ dữ liệu), Customers (có 91 bộ dữ liệu), Orders (có 830 bộ dữ liệu) và Order Details (có 2155 bộ dữ liệu).

Song song đó, để tạo sự tương đồng với tập dữ liệu thực nghiệm trên cơ sở dữ liệu quan hệ, chúng tôi đã xây dựng tập dữ liệu thực nghiệm trên cơ sở dữ liệu đồ thị bằng cách chuyển đổi mỗi bảng dữ liệu trong tập dữ liệu thực nghiệm trên cơ sở dữ liệu quan hệ sang những tập tin có đuôi “.CSV”, sau đó tiến hành chuyển đổi những tập tin này thành cơ sở dữ liệu đồ thị (hầu như tất cả hệ quản trị cơ sở dữ liệu đồ thị hiện nay đều có hỗ trợ cho việc chuyển đổi này). Kết quả nhận được là một tập dữ liệu thực nghiệm trên cơ sở dữ liệu đồ thị gồm 5 loại nhãn: Product (có 77 nút), Category (có 8 nút), Supplier (có 29 nút), Customer (có 91 nút), Order (có 830 nút) và có lược đồ như Hình 1.



Hình 1: Lược đồ cơ sở dữ liệu đồ thị NorthWind trong thực nghiệm.

Như vậy, qua việc chuyển đổi giữa 2 mô hình cơ sở dữ liệu, chúng tôi nhận thấy bảng dữ liệu và bộ dữ liệu (trong cơ sở dữ liệu quan hệ) sẽ tương đồng với nhãn và nút (trong cơ sở dữ liệu đồ thị).

Bên cạnh đó, chúng tôi còn thực hiện việc lập chỉ mục trên cả 2 mô hình cơ sở dữ liệu để có thể tối ưu hóa việc truy vấn, cụ thể xem ở Bảng 1.

Bảng 1: Mô tả việc lập chỉ mục trên 2 mô hình cơ sở dữ liệu.

Tên bảng/nhãn	Thuộc tính lập chỉ mục	
	Trong cơ sở dữ liệu quan hệ	Trong cơ sở dữ liệu đồ thị
Products/Product	ProductID	productID
Categories/Category	CategoryID	categoryID
Suppliers/Supplier	SupplierID	supplierID
Customers/Customer	CustomerID	customerID
Orders/Order	OrderID	orderID
Order Details	OrderID, ProductID	Không có nhãn này

Để việc đánh giá thực nghiệm được khách quan hơn, chúng tôi cũng dự kiến thực nghiệm trên 2 trường hợp, mỗi trường hợp sẽ có khối lượng dữ liệu khác nhau nhưng vẫn đảm bảo được sự tương đồng giữa 2 mô hình cơ sở dữ liệu, cụ thể xem ở Bảng 2.

Bảng 2: Mô tả các trường hợp trong thực nghiệm.

Tên bảng/nhãn	TRƯỜNG HỢP 1		TRƯỜNG HỢP 2	
	Số lượng bộ dữ liệu trong cơ sở dữ liệu quan hệ	Số lượng nút trong cơ sở dữ liệu đồ thị	Số lượng bộ dữ liệu trong cơ sở dữ liệu quan hệ	Số lượng nút trong cơ sở dữ liệu đồ thị
Products/Product	77	77	77	77
Categories/Category	8	8	8	8
Suppliers/Supplier	29	29	29	29
Customers/Customer	43	43	91	91
Orders/Order	393	393	830	830
Order Details	1063	Không có nhãn này	2155	Không có nhãn này
TỔNG SỐ BỘ/NÚT	1613	550	3190	1035

Tiếp theo, để chuẩn bị cho việc đánh giá và so sánh hiệu suất truy xuất dữ liệu, chúng tôi đã xây dựng các câu truy vấn cho cả 2 mô hình cơ sở dữ liệu dựa trên những yêu cầu của hệ thống gợi ý bán hàng theo phương pháp lọc dựa trên nội dung – phương pháp sử dụng hồ sơ sản phẩm (các thuộc tính liên quan đến đặc tính của sản phẩm như: tên hàng, loại hàng, giá,...) và hồ sơ người dùng (những sản phẩm mà người dùng đã thích, đã tìm kiếm, hay đã mua,...) để tìm và gợi ý danh sách sản phẩm tương đồng với những sản phẩm mà người dùng đã quan tâm [12]. Cụ thể trong thực nghiệm này, chúng tôi đã xây dựng 3 câu truy vấn (tương ứng với 3 chức năng gợi ý):

- Gợi ý danh sách sản phẩm “**cùng loại**” với những sản phẩm mà khách hàng đã mua (gọi tắt là “**Gợi ý 1**”).
- Gợi ý danh sách sản phẩm “**cùng nhà cung cấp**” với những sản phẩm mà khách hàng đã mua (gọi tắt là “**Gợi ý 2**”).
- Gợi ý danh sách sản phẩm “**cùng xuất xứ**” với những sản phẩm mà khách hàng đã mua (gọi tắt là “**Gợi ý 3**”).

3. THỰC NGHIỆM

3.1. Môi trường thực nghiệm

Việc thực nghiệm trong bài báo này được tiến hành trên máy tính cá nhân có cấu hình như sau:

- Hệ điều hành: Windows 7 Professional 64-bit.
- Nhà sản xuất: Dell Inspiron 14 5000 Series.
- RAM: 8 GB.
- Chip xử lý: Intel[®] Core™ i7-4510U CPU @ 2.00GHz 2.40GHz.

Ngoài ra, vì hệ thống gợi ý bán hàng dự kiến được xây dựng trên 2 mô hình cơ sở dữ liệu quan hệ và cơ sở dữ liệu đồ thị, nên cần phải có 2 hệ quản trị và 2 ngôn ngữ truy vấn dữ liệu tương ứng để tiến hành thực nghiệm, cụ thể xem ở Bảng 3.

Bảng 3: Hệ quản trị và ngôn ngữ truy vấn của 2 mô hình cơ sở dữ liệu.

Mô hình	Hệ quản trị cơ sở dữ liệu	Ngôn ngữ truy vấn dữ liệu
Cơ sở dữ liệu quan hệ	Microsoft [®] SQL Server [®] 2012	SQL
Cơ sở dữ liệu đồ thị	Neo4j	Cypher

Kết quả so sánh thực thi giữa hai loại cơ sở dữ liệu được đo trên thời gian thực thi của từng yêu cầu truy xuất dữ liệu. Với mỗi câu truy vấn, chúng tôi thực thi 10 lần và lấy giá trị trung bình của thời gian truy xuất dữ liệu với đơn vị tính là ms (mili giây).

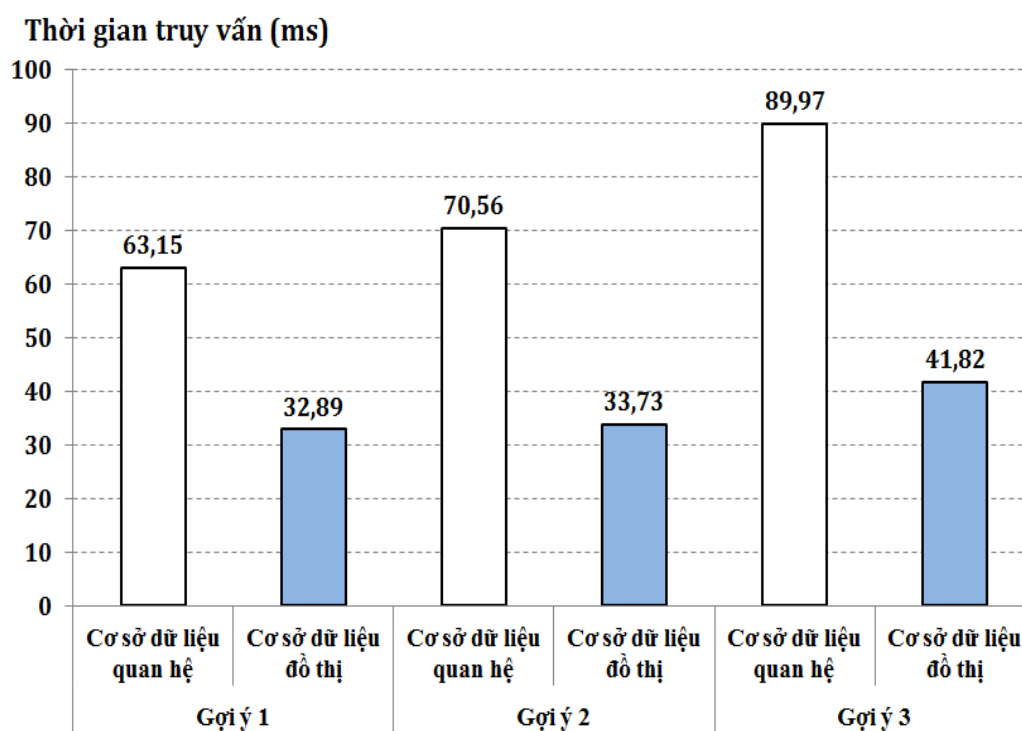
3.2. Đánh giá hiệu suất truy xuất dữ liệu

3.2.1. Hiệu suất truy xuất dữ liệu ở trường hợp 1

Kết quả chạy thực nghiệm ở trường hợp 1 được chúng tôi thể hiện thông qua Bảng 4 và Hình 2.

Bảng 4: Kết quả chạy thực nghiệm của trường hợp 1.

Loại truy vấn	Loại cơ sở dữ liệu	Tổng số bảng/nhãn kết nối	Tổng số lần bảng/nhãn lặp lại	Tổng số điều kiện trong câu truy vấn	Thời gian truy vấn (ms)
Gợi ý 1	Quan hệ	7	2	2	63,15
	Đồ thị	5	1	1	32,89
Gợi ý 2	Quan hệ	7	2	2	70,56
	Đồ thị	5	1	1	33,73
Gợi ý 3	Quan hệ	11	6	3	89,97
	Đồ thị	9	5	2	41,82



Hình 2: Biểu đồ thể hiện kết quả chạy thực nghiệm của trường hợp 1.

Như vậy, trong trường hợp 1 của thực nghiệm này, đối với cả 3 câu truy vấn, hiệu suất truy xuất dữ liệu của cơ sở dữ liệu đồ thị đều cao hơn cơ sở dữ liệu quan hệ, cụ thể:

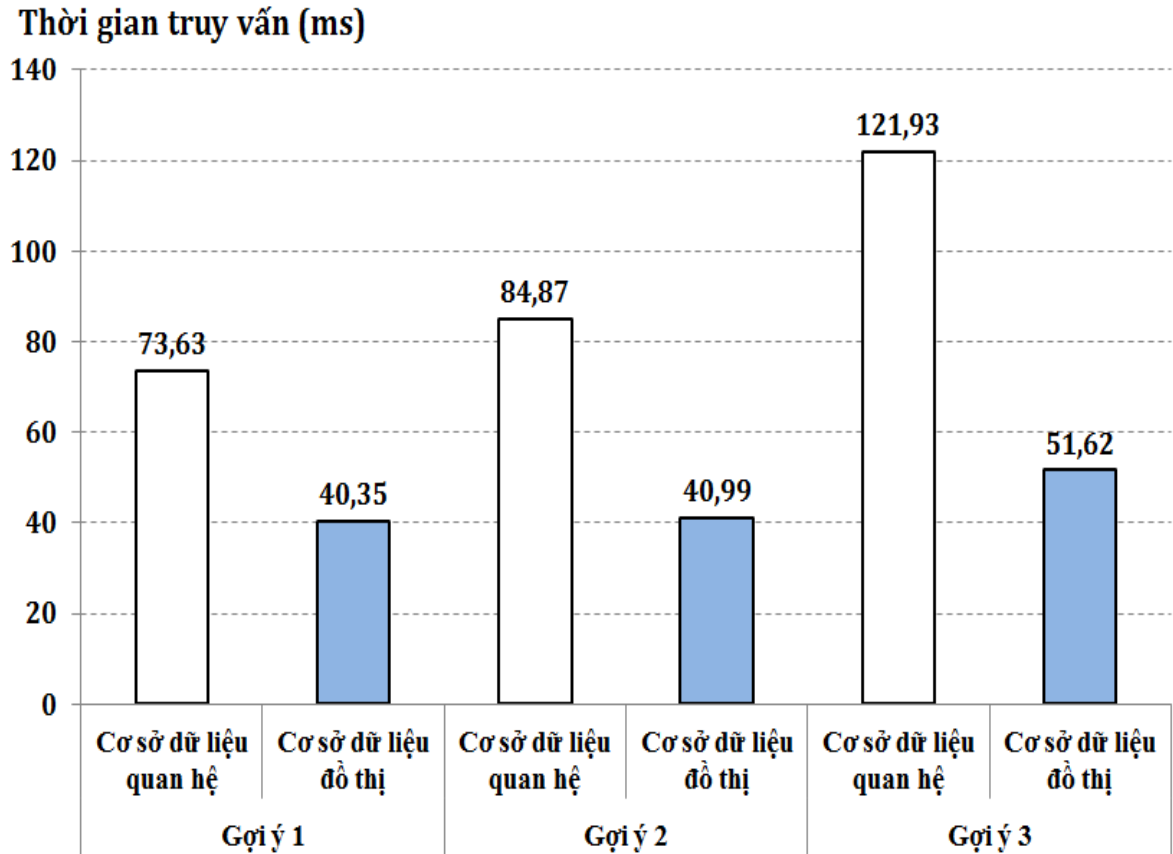
- Đối với truy vấn gợi ý 1: cao hơn 47,92%.
- Đối với truy vấn gợi ý 2: cao hơn 52,2%.
- Đối với truy vấn gợi ý 3: cao hơn 53,52%.

3.2.2. Hiệu suất truy xuất dữ liệu ở trường hợp 2

Kết quả chạy thực nghiệm ở trường hợp 2 được chúng tôi thể hiện thông qua Bảng 5 và Hình 3.

Bảng 5: Kết quả chạy thực nghiệm của trường hợp 2.

Loại truy vấn	Loại cơ sở dữ liệu	Tổng số bảng/nhãn kết nối	Tổng số lần bảng/nhãn lặp lại	Tổng số điều kiện trong câu truy vấn	Thời gian truy vấn (ms)
Gợi ý 1	Quan hệ	7	2	2	73,63
	Đồ thị	5	1	1	40,35
Gợi ý 2	Quan hệ	7	2	2	84,87
	Đồ thị	5	1	1	40,99
Gợi ý 3	Quan hệ	11	6	3	121,93
	Đồ thị	9	5	2	51,62



Hình 3: Biểu đồ thể hiện kết quả chạy thực nghiệm của trường hợp 2.

Như vậy, trong trường hợp 2 của thực nghiệm này, đối với cả 3 câu truy vấn, hiệu suất truy xuất dữ liệu của cơ sở dữ liệu đồ thị cũng cao hơn cơ sở dữ liệu quan hệ, cụ thể:

- Đối với truy vấn gợi ý 1: cao hơn 45,2%.
- Đối với truy vấn gợi ý 2: cao hơn 51,7%.
- Đối với truy vấn gợi ý 3: cao hơn 57,66%.

3.3. Một số đánh giá khác

Trong thực nghiệm này, bên cạnh việc so sánh hiệu suất truy xuất dữ liệu, chúng tôi còn so sánh giữa cơ sở dữ liệu đồ thị và cơ sở dữ liệu quan hệ ở một số phương diện khác, cụ thể như Bảng 6.

Bảng 6: So sánh số lượng bảng dữ liệu, bộ dữ liệu (trong cơ sở dữ liệu quan hệ) với số lượng nhân, nút (trong cơ sở dữ liệu đồ thị).

	TRƯỜNG HỢP 1		TRƯỜNG HỢP 2	
	Trong cơ sở dữ liệu quan hệ	Trong cơ sở dữ liệu đồ thị	Trong cơ sở dữ liệu quan hệ	Trong cơ sở dữ liệu đồ thị
Tổng số lượng bảng/nhân	6	5	6	5
Tổng số lượng bộ/nút	1613	550	3190	1035

Dựa vào Bảng 6, chúng tôi có được một số kết luận như sau:

- Số lượng nhân (trong cơ sở dữ liệu đồ thị) ít hơn số lượng bảng dữ liệu (trong cơ sở dữ liệu quan hệ), cụ thể ở cả 2 trường hợp là ít hơn 16,67%.
- Số lượng nút (trong cơ sở dữ liệu đồ thị) cũng ít hơn số lượng bộ dữ liệu (trong cơ sở dữ liệu quan hệ), cụ thể ở trường hợp 1 là ít hơn 65,9% và ở trường hợp 2 là ít hơn 67,55%.

Như vậy, cơ sở dữ liệu đồ thị đã đơn giản hóa dữ liệu hơn rất nhiều so với cơ sở dữ liệu quan hệ, nguyên nhân là vì tất cả bộ dữ liệu của bảng “Order Details” (trong cơ sở dữ liệu quan hệ) đã được chuyển đổi thành dữ liệu thuộc tính của mỗi quan hệ “ORDERS” (mỗi quan hệ liên kết các nút thuộc nhãn “Order” và các nút thuộc nhãn “Product” trong cơ sở dữ liệu đồ thị).

4. KẾT LUẬN VÀ HƯỚNG PHÁT TRIỂN

Qua tất cả những nội dung cùng những kết quả thực nghiệm đã được trình bày, chúng tôi nhận thấy cơ sở dữ liệu đồ thị sẽ mang lại rất nhiều lợi ích so với cơ sở dữ liệu quan hệ trong việc xây dựng các hệ thống gợi ý: giúp đơn giản hóa dữ liệu, hạn chế số lượng bảng lưu trữ (khoảng 15%) và bộ dữ liệu (khoảng 65%); hạn chế việc lặp lại các bảng dữ liệu và số lượng điều kiện trong câu truy vấn, nâng cao hiệu suất truy

xuất dữ liệu. Kết quả thực nghiệm này cũng đồng nghĩa với việc chúng tôi đã minh chứng được: “Cơ sở dữ liệu đồ thị sẽ mang lại nhiều ưu thế hơn, đặc biệt là nâng cao được hiệu suất truy xuất dữ liệu so với cơ sở dữ liệu quan hệ và rất phù hợp để xây dựng các hệ thống gợi ý trong thời đại bùng nổ thông tin như hiện nay”.

Tuy nhiên, bên cạnh những thuận lợi, cơ sở dữ liệu đồ thị cũng mang lại một số khó khăn, trong đó đáng quan tâm nhất là việc xác định hướng đi giữa các nhãn sao cho câu truy vấn được tối ưu nhất và quan trọng hơn hết là phải đáp ứng được nhu cầu của mỗi cá nhân người dùng.

Trong tương lai, chúng tôi sẽ tiếp tục nghiên cứu và áp dụng cơ sở dữ liệu đồ thị trong việc xây dựng các ứng dụng hệ thống thông tin ở một số lĩnh vực khác để chứng minh cơ sở dữ liệu đồ thị không những có nhiều hữu ích trong việc xây dựng các hệ thống gợi ý mà còn trong rất nhiều lĩnh vực hay ngành nghề khác.

TÀI LIỆU THAM KHẢO

- [1] F. Alexander, J. Michael, N. Gerald, R. Florian, R. Stefan and S. Martin, *Basic Approaches in Recommendation Systems*, 2013.
- [2] F. Alexander and R. Burke, *Constraint-based Recommender Systems: Technologies and Research Issues*, 2008.
- [3] R. Burke, *Hybrid Recommender Systems: Survey and Experiments*, 2002.
- [4] G. David, N. David, O. M. Brian and T. Douglas, *Using collaborative filtering to weave an information tapestry*, 1992.
- [5] R. S. Deepak, S. S. Rajesh and D. H. L. Mandoria, *Analysis of Fraudulent in Graph Data base for Identification & Prevention*, *International Journal for Innovative Research in Science & Technology*, vol. 4, no. 5, pp. 42-49, 2017.
- [6] J. Dietmar, Z. Markus, F. Alexander and F. Gerhard, *Recommender Systems: An Introduction*, New York: Cambridge University Press, 2011.
- [7] N. H. Dũng và N. T. Nghe, *Hệ thống gợi ý sản phẩm trong bán hàng trực tuyến sử dụng kỹ thuật lọc cộng tác*, *Tạp chí Khoa học Trường Đại học Cần Thơ*, pp. 36-51, 2014.
- [8] The Enlightened DBA, *DBA's Guide to NoSQL and DATASTAX ENTERPRISE*, 2016.
- [9] L. Fabiana and R. Francesco, *Case-Based Recommender Systems: a Unifying View*, 2005.
- [10] H. Guy, *Next Generation Databases - NoSQL, NewSQL and Big Data*, Apress, 2016.

- [11] C. Hoang-Quy and J. Malek, Implementing a Recommender system with graph database, 2014.
- [12] L. T. Huyền, Phương pháp lọc cộng tác và ứng dụng trong hệ thống tin tư vấn, 2015.
- [13] R. Ian, W. Jim and E. Emil, Graph Databases, 2015.
- [14] Š. Jan, Social Network Recommendation using Graph Databases, 2015.
- [15] B. Joeran, L. Stefan, G. Bela and N. Andreas, The Architecture and Datasets of Docear's Research Paper Recommender System, 2014.
- [16] B. S. John, H. David and K. Carl, Empirical analysis of predictive algorithms for collaborative filtering, 1998.
- [17] P. Joseph, Flexible recommender systems based on graphs, 2017.
- [18] H. Michael, B. Ryan and L. William, The Definitive Guide to Graph Databases for the RDBMS Developer, 2016.
- [19] E. D. Michael, R. T. John en K. A. Joseph, Collaborative Filtering Recommender Systems, Foundations and TrendsR in Human–Computer Interaction, vol. IV, nr. 2, pp. 81-173, 2011.
- [20] B. Mike, The Current State of Graph Databases, 2012.
- [21] N. D. Phuong, Phát triển một số phương pháp lọc thông tin cho hệ tư vấn, 2011.
- [22] G. Radu, H. L. Matthew and R. Roy, Elasticsearch in Action, 2016.
- [23] P. Rashmi, Retail data analytics using Graph Database, 2018.
- [24] B. V. Rik, Learning Neo4j, Birmingham-Mumbai: Packt, 2014.
- [25] S. Sumedh, Collaborative Filtering using Weighted BiPartite Graph Projection, 2013.
- [26] The Neo4j Team, The Neo4j Developer Manual v3.1, 2017.
- [27] The Neo4j Team, The Definitive Guide to Graph Databases for the RDBMS Developer, 2016.
- [28] H. N. Tín, Phát triển một số phương pháp khuyến nghị hỗ trợ tìm kiếm thông tin học thuật dựa trên tiếp cận phân tích mạng xã hội, 2016.
- [29] Byoung-Ha Yoon, Seon-Kyu Kim and Seon-Young Kim, Use of Graph Database for the Integration of Heterogeneous Biological Data, 2017.
- [30] Callidus Software, OrientDB, [Online]. Available: <https://orientdb.com/>.
- [31] Datastax, Datastax Enterprise, [Online]. Available: <https://www.datastax.com/>.
- [32] Microsoft, Microsoft Azure, [Online]. Available: <https://azure.microsoft.com/>.
- [33] The Neo4j Team, Neo4j, [Online]. Available: <https://neo4j.com/>.