

ĐẠI HỌC QUỐC GIA TP. HỒ CHÍ MINH
TRƯỜNG ĐẠI HỌC CÔNG NGHỆ THÔNG TIN
KHOA CÔNG NGHỆ PHẦN MỀM

NGUYỄN HỮU ĐẠT

HƯỚNG DẪN SỬ DỤNG
PHẦN MỀM DEMO

KHUYẾN NGHỊ CỘNG TÁC
DỰA TRÊN TIẾP CẬN HỌC SÂU
DEEP LEARNING FOR
COLLABORATOR RECOMMENDATION

KỸ SƯ NGÀNH KỸ THUẬT PHẦN MỀM

TP. HỒ CHÍ MINH, 2020

ĐẠI HỌC QUỐC GIA TP. HỒ CHÍ MINH
TRƯỜNG ĐẠI HỌC CÔNG NGHỆ THÔNG TIN
KHOA CÔNG NGHỆ PHẦN MỀM

NGUYỄN HỮU ĐẠT – 15520104

HƯỚNG DẪN SỬ DỤNG
PHẦN MỀM DEMO

KHUYẾN NGHỊ CỘNG TÁC
DỰA TRÊN TIẾP CẬN HỌC SÂU
DEEP LEARNING FOR
COLLABORATOR RECOMMENDATION

KỸ SƯ NGÀNH KỸ THUẬT PHẦN MỀM

GIẢNG VIÊN HƯỚNG DẪN
TS. HUỖNH NGỌC TÍN

TP. HỒ CHÍ MINH, 2020

LỜI CẢM ƠN

Trước hết, em xin gửi lời cảm ơn đến quý thầy cô Khoa Công nghệ phần mềm trường Đại học Công nghệ Thông tin – ĐHQG TP.HCM. Sau hơn 4 năm tròn học tập tại trường, em đã được trang bị các kiến thức cơ bản, các kỹ năng để có thể hoàn thành Khóa luận tốt nghiệp của mình.

Đặc biệt với thầy Huỳnh Ngọc Tín, em xin cảm ơn thầy vì đã nhiệt tình, quan tâm, hướng dẫn và truyền đạt những kiến thức và kinh nghiệm cho em trong suốt thời gian thực hiện Khóa luận tốt nghiệp của mình. Bên cạnh đó, em cũng muốn gửi lời cảm ơn đến bạn bè, anh chị thuộc công ty VCGroup đã đồng viên, tận tình giúp đỡ trong lúc em thực hiện khóa luận tốt nghiệp.

Tuy nhiên, việc vấp phải sai sót trong hoàn tất Khóa luận tốt nghiệp là điều không tránh khỏi, em mong nhận được sự góp ý của quý thầy và các bạn để được hoàn thiện hơn.

Một lần nữa, em xin chân thành cảm ơn.

Thành phố Hồ Chí Minh, 31 tháng 07 năm 2020

Sinh viên

Nguyễn Hữu Đạt

MỤC LỤC

Chương 1.	LỜI GIỚI THIỆU	1
Chương 2.	HƯỚNG DẪN SỬ DỤNG	2
2.1.	Mở chương.....	2
2.2.	Yêu cầu hệ thống.....	2
2.3.	Tải và cài đặt Neo4j.....	2
2.4.	Tải và cài đặt Python	2
2.5.	Hướng dẫn cài đặt phần mềm	2
2.6.	Hướng dẫn sử dụng	2
2.6.1.	Khởi động lần đầu	2
2.6.2.	Khởi tạo dự án Neo4j mới	3
2.6.3.	Khởi tạo cơ sở dữ liệu mới	4
2.6.4.	Thêm thư viện APOC vào dự án.....	6
2.6.5.	Sửa đổi cài đặt để kết nối với phần mềm.....	8
2.6.6.	Khởi động phần mềm và tạo kết nối	10
2.6.7.	Tạo mới dự án quản lý phương pháp khuyến nghị	11
2.6.8.	Xem dữ liệu thực nghiệm	14
2.6.9.	Tạo mới phương pháp khuyến nghị	15
2.6.10.	Chạy và đánh giá phương pháp.....	18
2.1.	Kết chương.....	21
Chương 3.	KẾT.....	22

DANH MỤC HÌNH

Hình 2-1 Màn hình chính của phần mềm. Trong đó, các thành phần gồm: 1. Giới thiệu bài toán khuyến nghị cộng tác; 2. Mở tệp pdf hướng dẫn sử dụng; 3. Mở tệp giới thiệu về đề tài; 4. Thêm dự án quản lý, so sánh và đánh giá thuật toán.	3
Hình 2-2 Khởi tạo dự án Neo4j mới bằng cách nhấn vào nút “New” như trên màn hình	4
Hình 2-3 Chọn nút kế bên tên màn hình để thực hiện đổi tên dự án	4
Hình 2-4 Đổi tên dự án và nhấn nút đồng ý	4
Hình 2-5 Chọn dự án muốn khởi tạo cơ sở dữ liệu.....	5
Hình 2-6 Chọn “Create a Local Graph”	5
Hình 2-7 Tạo mới cơ sở dữ liệu.	6
Hình 2-8 Chọn “Add Plugin” để thêm APOC	7
Hình 2-9 Chọn “Install” để cài đặt APOC.....	7
Hình 2-10 Chọn “Manage” để vào trang cài đặt.....	8
Hình 2-11 Chọn “Setting” để vào trang cài đặt và thực hiện thêm các mục như 2,3,4 và chọn “Apply”	9
Hình 2-12 Kết nối cơ sở dữ liệu Neo4j	10
Hình 2-13 Khởi tạo cơ sở dữ liệu cho hệ thống khuyến nghị cộng tác.....	10
Hình 2-14 Chọn “Thêm dự án” để tạo mới dự án.....	11
Hình 2-15 Khởi tạo dự án so sánh thuật toán khuyến nghị cộng tác. Trong đó, từ 1 đến 5: xem tại Bảng 2-1; 6: Cảnh báo trường dữ liệu; 7. Nút dừng để khởi tạo dự án	13
Hình 2-16 Kết quả khởi tạo dự án.....	13
Hình 2-17 Chọn dự án cần xem dữ liệu thực nghiệm	14
Hình 2-18 Thông tin chi tiết về dữ liệu thực nghiệm.....	15
Hình 2-19 Chọn dự án cần thêm mới thuật toán.....	16
Hình 2-20 Chọn “Thêm phương pháp”	16
Hình 2-21 Chọn “Thêm phương pháp”	17

Hình 2-22 Chọn phương pháp cần thêm. Lưu ý: với Node2vec, bạn đọc có thể chọn cài đặt và thay đổi tham số để có được kết quả tiên đoán chính xác hơn tùy theo mạng đồng tác giả xây dựng.	17
Hình 2-23 Tùy chỉnh tham số của thuật toán.....	18
Hình 2-24 Chọn phương pháp cần đánh giá	19
Hình 2-25 Chọn phương pháp cần đánh giá	19
Hình 2-26 Chọn nút “Thực hiện” để khởi chạy phương pháp.....	20
Hình 2-27 Chọn nút “Thực hiện” để đánh giá phương pháp.....	20
Hình 2-28 Kết quả đánh giá phương pháp.....	21

DANH MỤC BẢNG

Bảng 2-1 Trường dữ liệu mô tả theo hình	12
--	----

DANH MỤC TỪ VIẾT TẮT, THUẬT NGỮ

Không có

Chương 1. LỜI GIỚI THIỆU

Nhằm để so sánh độ chính xác cũng như nắm rõ các bước trong các thuật toán được đem ra so sánh với phương pháp học sâu Node2vec trong giải quyết bài toán khuyến nghị cộng tác, phần mềm này là công cụ hỗ trợ nhằm mục đích chạy và so sánh kết quả dựa trên 3 phương pháp đánh giá trong truy vấn thông tin gồm: Precision (độ đo chính xác), Recall (độ bao phủ) và F-measure (độ đo trung bình giữa Precision và Recall). Nhờ vào việc so sánh này, phần mềm sẽ phô diễn sự uyển chuyển của Node2vec trong việc học và tiên đoán cộng tác so với những thuật toán khuyến nghị phổ biến.

Bên cạnh đó, phần mềm chỉ phục vụ mục đích nghiên cứu khoa học, mọi nghiên cứu khoa học đều có thể sử dụng, kế thừa và mở rộng vào việc xây dựng, đánh giá và so sánh các thuật toán khác.

Ngoài ra, phần mềm còn có thể giúp nghiên cứu viên hiểu rõ đầu vào (Input) và đầu ra (Output) trong các bước của từng thuật toán. Điều này sẽ giúp những bạn nghiên cứu viên mới bắt đầu nghiên cứu dễ dàng theo kịp nhằm đẩy nhanh tiến độ học tập trong những bước đầu tiên nghiên cứu các thuật toán khuyến nghị.

Tóm lại, phần mềm sẽ nhằm so sánh kết quả khuyến nghị của Node2vec với một số phương pháp khuyến nghị phổ biến và thể hiện rõ từng bước của các thuật toán khuyến nghị nhằm phục vụ cộng đồng học thuật.

Chương 2. HƯỚNG DẪN SỬ DỤNG

2.1. Mở chương

Chương này tập trung giới thiệu về những tính năng để có thể tương tác với phần mềm giúp bạn đọc thực hiện, đánh giá và so sánh các thuật toán khuyến nghị dùng trong khuyến nghị cộng tác.

2.2. Yêu cầu hệ thống

Phần mềm thực hiện trên máy có yêu cầu tối thiểu như sau:

- **RAM:** 16G
- **Hệ điều hành:** Windows 10 Education.
- **Processor:**
 - o **Số lượng:** 8
 - o **Loại:** Intel(R) Core(TM) i7-7700HQ CPU @2.8GHz
- **Database:** Neo4j (vui lòng cài đặt Neo4j Desktop)
- **Python:** 3.7.3

2.3. Tải và cài đặt Neo4j

Xin vui lòng xem tại trang <https://neo4j.com/developer/get-started/>

2.4. Tải và cài đặt Python

Xin vui lòng xem tại trang <https://www.python.org/>

2.5. Hướng dẫn cài đặt phần mềm

2.6. Hướng dẫn sử dụng

2.6.1. Khởi động lần đầu

Các bước thực hiện sau chỉ áp dụng ở lần sử dụng đầu tiên, ở các lần tiếp theo, bạn đọc chỉ cần thực hiện từ **Bước 5**. Các bước khởi tạo phần mềm lần đầu như sau:

- **Bước 1:** mở Neo4j Desktop, khởi tạo dự án mới (như mục 2.6.2).
- **Bước 2:** ở Neo4j Desktop, khởi tạo cơ sở dữ liệu mới (như mục 2.6.3).
- **Bước 3:** Thêm thư viện Apoc của Neo4j vào dự án (như mục 2.6.4).
- **Bước 4:** Sửa đổi cài đặt của dự án để giúp phần mềm có thể tương tác với Neo4j (như mục 2.6.5).
- **Bước 5:** Khởi động phần mềm và tạo kết nối Neo4j (như mục 2.6.6).

Giờ đây, phần mềm sẽ xuất hiện màn hình chính như hình Hình 2-1. Bạn đọc có thể tham khảo thêm các mục còn lại để nắm thông tin chi tiết.

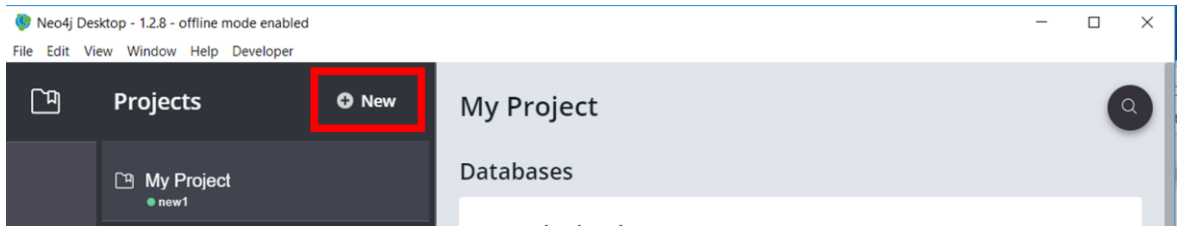


Hình 2-1 Màn hình chính của phần mềm. Trong đó, các thành phần gồm: 1. Giới thiệu bài toán khuyến nghị cộng tác; 2. Mở tệp pdf hướng dẫn sử dụng; 3. Mở tệp giới thiệu về đề tài; 4. Thêm dự án quản lý, so sánh và đánh giá thuật toán.

2.6.2. Khởi tạo dự án Neo4j mới

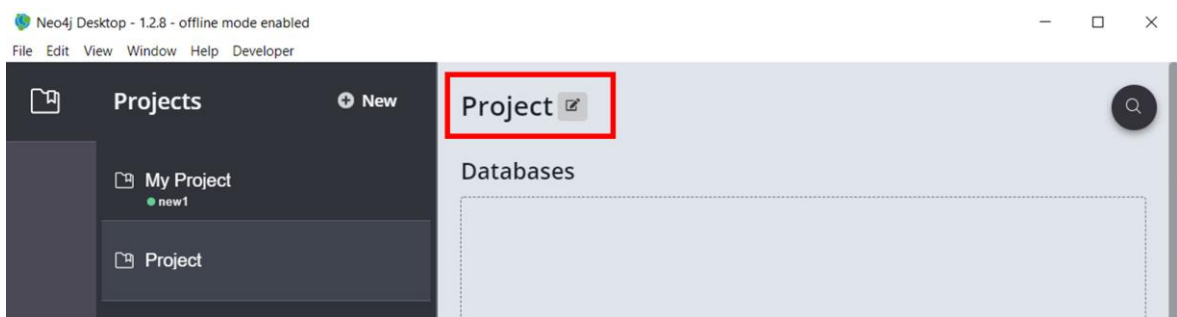
Việc khởi tạo dự án Neo4j mới chỉ cần thiết khi chưa có sẵn cơ sở dữ liệu trong lần sử dụng phần mềm đầu tiên hoặc bạn đọc có nhiều dự án với nhiều cơ sở dữ liệu khác nhau. Còn lại, bạn đọc không cần khởi tạo dự án mới trong mỗi lần sử dụng chương trình.

- **Bước 1:** khởi tạo dự án mới như tại hình

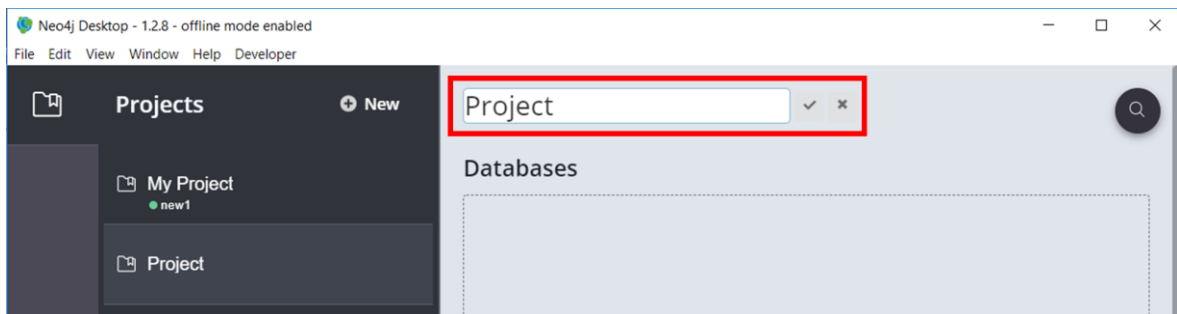


Hình 2-2 Khởi tạo dự án Neo4j mới bằng cách nhấn vào nút “New” như trên màn hình

- **Bước 2:** Đổi tên dự án muốn khởi tạo bằng cách thực hiện tuần tự như hình Hình 2-3 và hình Hình 2-4 (Không bắt buộc)



Hình 2-3 Chọn nút kế bên tên màn hình để thực hiện đổi tên dự án

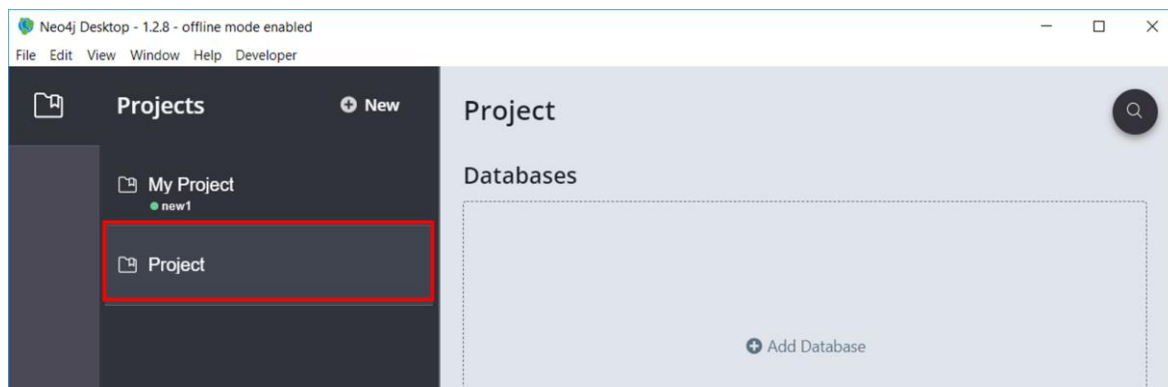


Hình 2-4 Đổi tên dự án và nhấn nút đồng ý

2.6.3. Khởi tạo cơ sở dữ liệu mới

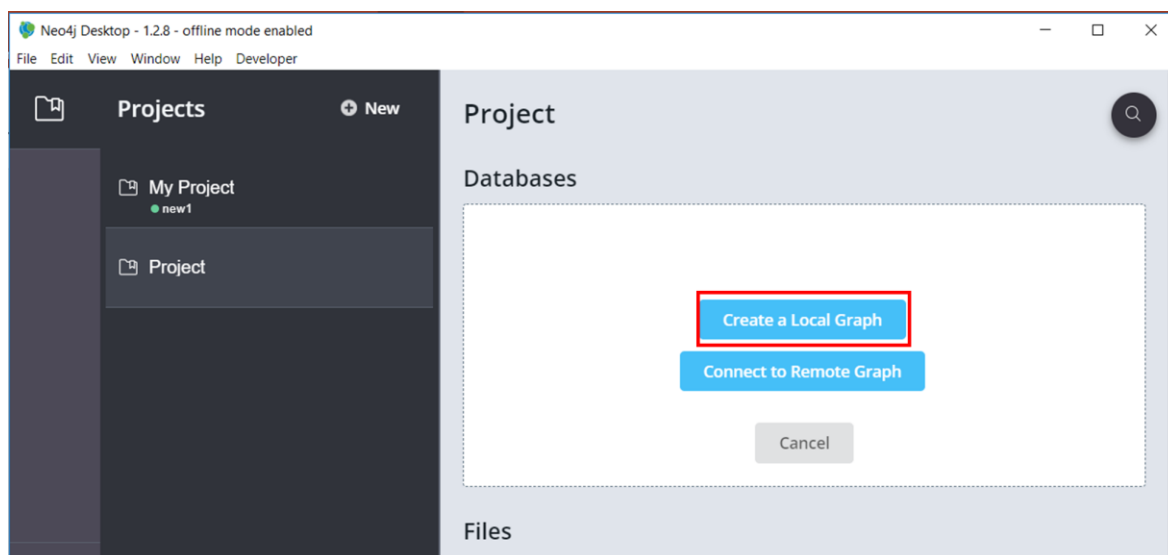
Việc khởi tạo cơ sở dữ liệu mới chỉ cần thiết khi chưa có sẵn cơ sở dữ liệu trong lần sử dụng phần mềm đầu tiên hoặc bạn đọc có nhiều dự án với nhiều cơ sở dữ liệu khác nhau. Còn lại, bạn đọc không cần khởi tạo cơ sở dữ liệu mới trong mỗi lần sử dụng chương trình.

- **Bước 1:** Chọn dự án muốn khởi tạo cơ sở dữ liệu.



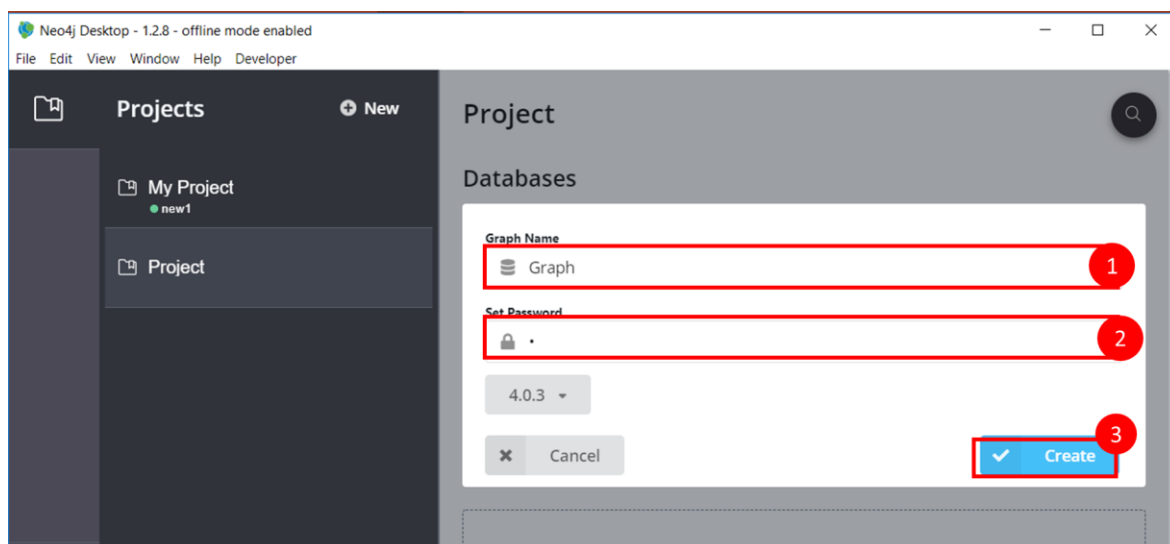
Hình 2-5 Chọn dự án muốn khởi tạo cơ sở dữ liệu

- **Bước 2:** Chọn khởi tạo cơ sở dữ liệu bằng cách nhấn “Create a Local Graph”



Hình 2-6 Chọn “Create a Local Graph”

- **Bước 3:** Lần lượt điền vào các trường và thực hiện như sau:
 - **Graph Name:** Graph (không bắt buộc)
 - **Set Password:** a (bắt buộc)
 - Chọn nút “Create”

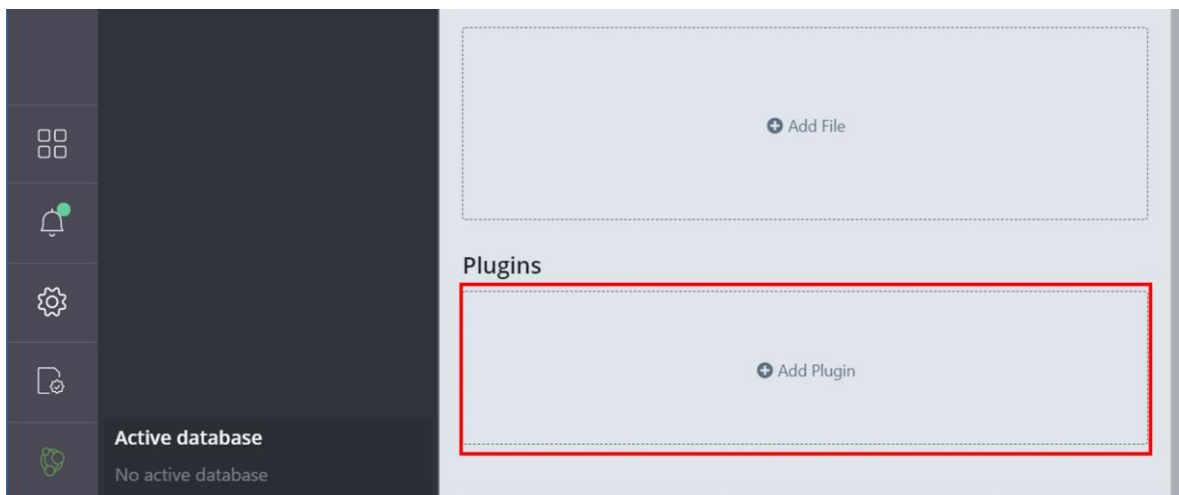


Hình 2-7 Tạo mới cơ sở dữ liệu.

2.6.4. Thêm thư viện APOC vào dự án

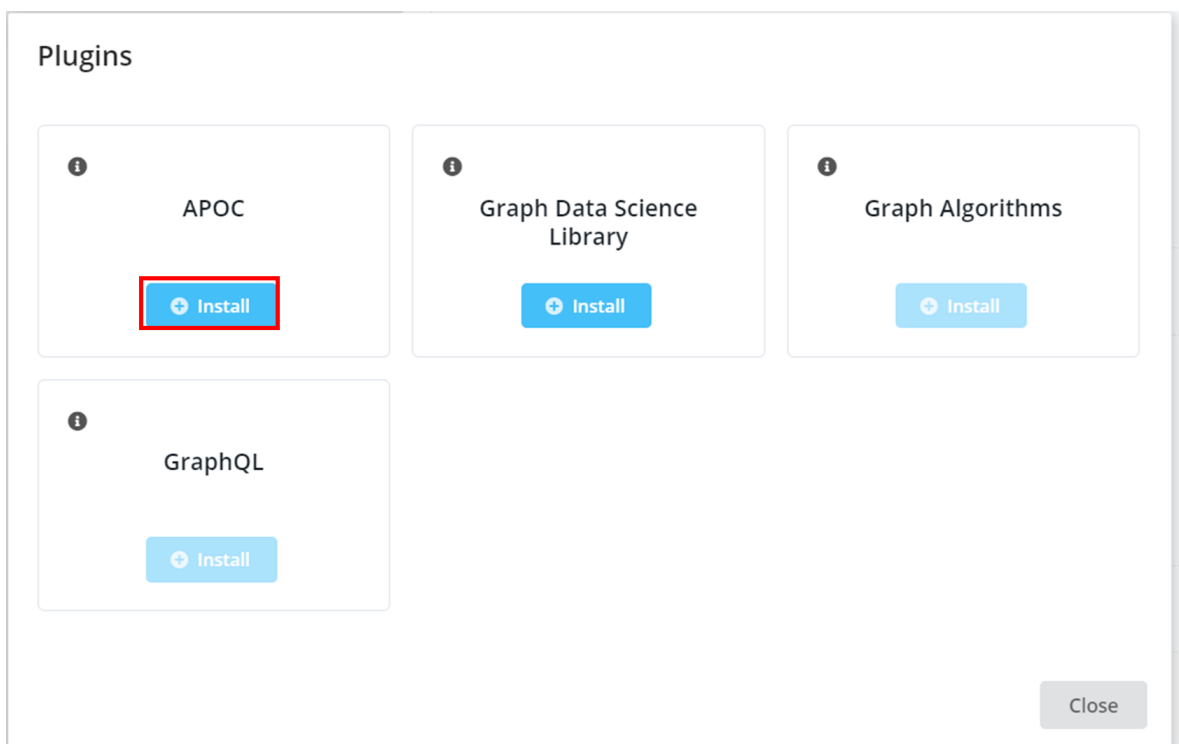
Thư viện APOC giúp phần mềm thực hiện các lệnh xóa, thêm các nút và liên kết trở nên nhanh chóng hơn nhờ vào sử dụng đa tiến trình. Để thêm thư viện APOC, bạn đọc thực hiện theo các bước như sau:

- **Bước 1:** Chọn “Add Plugin” (hình ...)



Hình 2-8 Chọn “Add Plugin” để thêm APOC

- **Bước 2:** Cài đặt APOC bằng các chọn “Install” (hình ...)

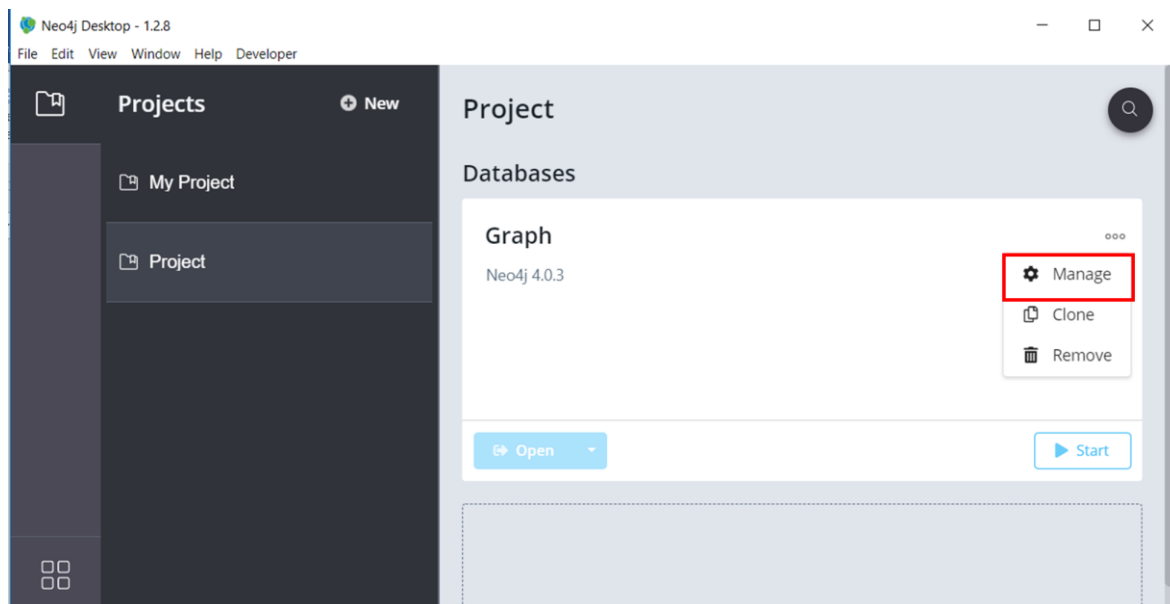


Hình 2-9 Chọn “Install” để cài đặt APOC

2.6.5. Sửa đổi cài đặt để kết nối với phần mềm

Phần mềm chỉ có thể kết nối với cơ sở dữ liệu của Neo4j khi và chỉ khi sửa đổi cài đặt của cơ sở dữ liệu như sau:

Bước 1: Vào trang quản lý dự án bằng hướng dẫn tại hình ...



Hình 2-10 Chọn “Manage” để vào trang cài đặt

Bước 2: Vào trang cài đặt bằng cách chọn “Setting” bằng hướng dẫn tại hình ... và thêm dấu “#” vào trước dòng “dbms.directories.import=import”

Bước 3: Thêm các dòng sau vào ngay sau dòng “dbms.directories.import=import” như sau:

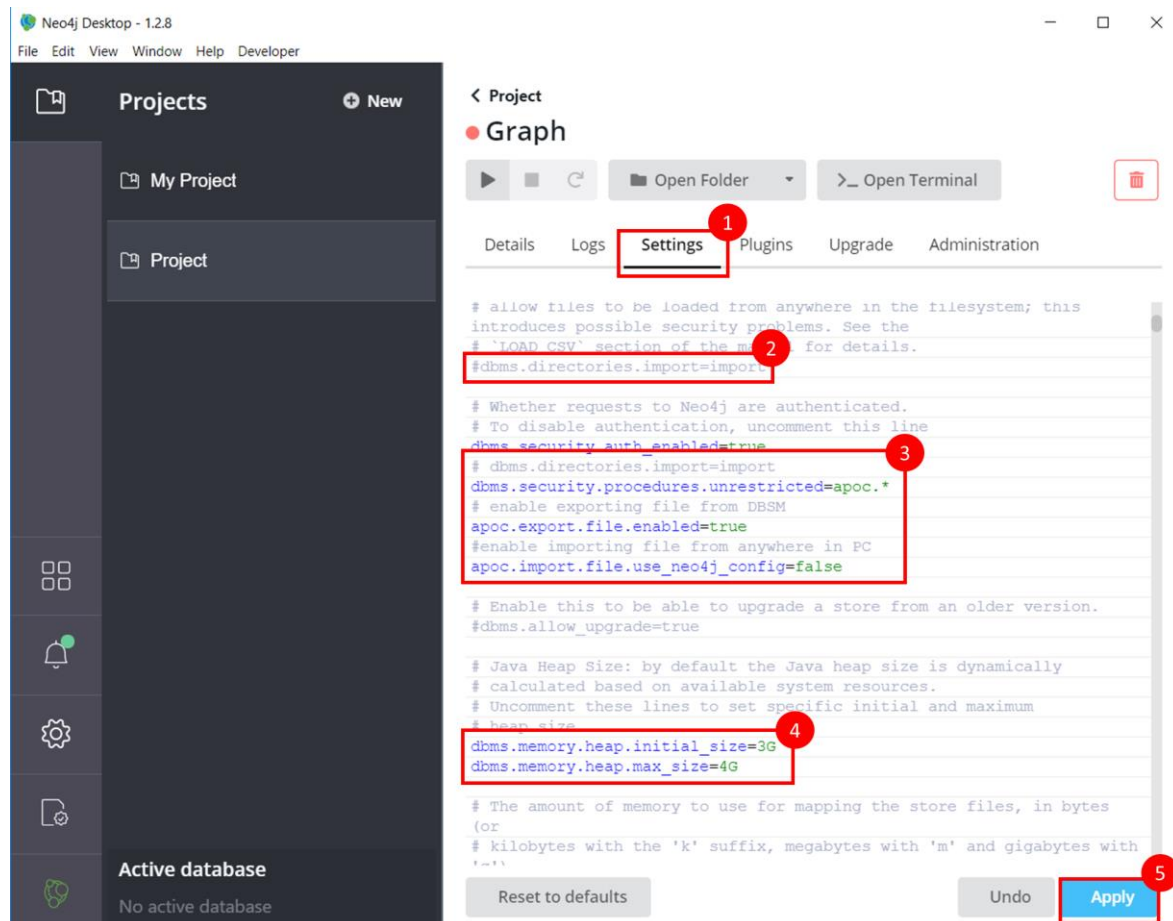
```
# dbms.directories.import=import  
  
dbms.security.procedures.unrestricted=apoc.*  
  
# enable exporting file from DBSM  
apoc.export.file.enabled=true  
  
#enable importing file from anywhere in PC  
apoc.import.file.use_neo4j_config=false
```


Bước 4: Tìm các dòng sau và chỉnh sửa như sau:

`dbms.memory.heap.initial_size=3G`

`dbms.memory.heap.max_size=4G`

Bước 5: Cập nhật các cài đặt bằng cách nhấn “Apply”

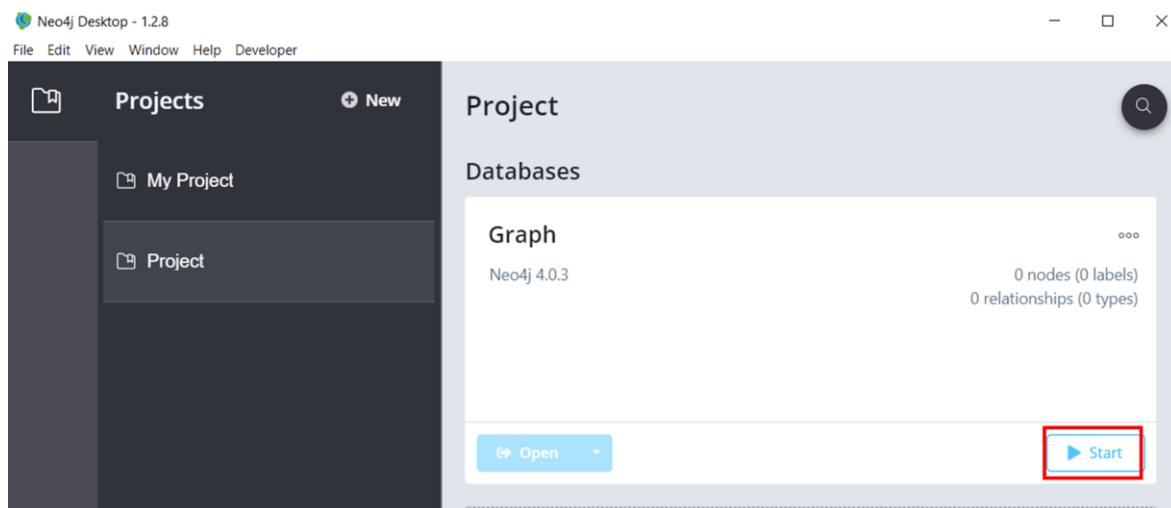


Hình 2-11 Chọn “Setting” để vào trang cài đặt và thực hiện thêm các mục như 2,3,4 và chọn “Apply”

2.6.6. Khởi động phần mềm và tạo kết nối

Phần mềm khi khởi tạo sẽ cần kết nối với cơ sở dữ liệu Neo4j nên bạn đọc cần mở Neo4j Destop và thực hiện các bước như sau:

Bước 1: Chọn nút “Start” cơ sở dữ liệu tại màn hình Neo4j Desktop:



Hình 2-12 Kết nối cơ sở dữ liệu Neo4j

Bước 2: Khởi động phần mềm và nếu xuất hiện dòng “Không tìm thấy cơ sở dữ liệu. Vui lòng khởi tạo để sang màn hình chính”. Nhấn nút “Khởi tạo” như tại hình Nếu cơ sở dữ liệu đã được khởi tạo, bạn đọc sẽ được chuyển sang màn hình chính. (**Lưu ý:** Thời gian khởi tạo phần mềm rất lâu – từ 3 đến 4 giờ đồng hồ trên máy có cấu hình như tại 2.1 nên bạn đọc không thể thực hiện thao tác trong thời gian ngắn.)

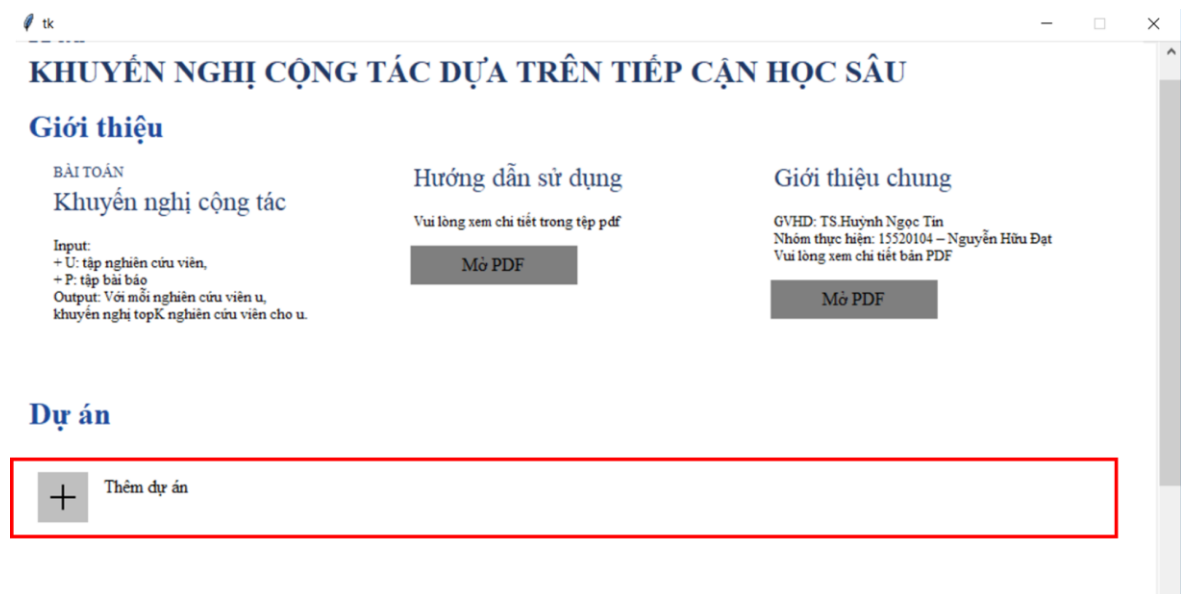


Hình 2-13 Khởi tạo cơ sở dữ liệu cho hệ thống khuyến nghị cộng tác

2.6.7. Tạo mới dự án quản lý phương pháp khuyến nghị

Để giúp bạn đọc có thể xây dựng các cơ sở dữ liệu khác nhau cho từng dự án, phần mềm có chức năng thêm mới dự án để bạn đọc có thể xây dựng nhiều dự án và ứng dụng nhiều phương pháp khác nhau cho từng dự án. Với mỗi dự án là một cơ sở dữ liệu được lưu trữ bởi Neo4j. Để tạo dự án mới, bạn đọc tiến hành theo các bước như sau:

- **Bước 1:** Chọn nút “Thêm dự án” tại màn hình chính:



Hình 2-14 Chọn “Thêm dự án” để tạo mới dự án

- **Bước 2:** Tại trang “Thêm dự án”, bạn đọc tiến hành điền các trường dữ liệu tạo bảng như sau:

STT	Trường dữ liệu	Mô tả	Ràng buộc
1	Tên dự án	Tên dự án của cơ sở dữ liệu tương ứng	
	Tập huấn luyện		
2	Năm bắt đầu	Năm bắt đầu quan sát liên kết cộng tác.	<ul style="list-style-type: none"> - Kiểu số - Sau 1995 - Nhỏ hơn năm kết thúc
3	Năm kết thúc	Năm kết thúc quan sát liên kết cộng tác.	<ul style="list-style-type: none"> - Kiểu số - Sau 1995 - Nhỏ hơn năm bắt đầu trong tập đánh giá
	Tập đánh giá		
4	Năm bắt đầu	Năm bắt đầu hình thành liên kết cộng tác thực sự xảy ra trong tương lai.	<ul style="list-style-type: none"> - Kiểu số - Sau 1995 - Nhỏ hơn năm kết thúc trong tập đánh giá
5	Năm kết thúc	Năm kết thúc hình thành liên kết cộng tác thực sự xảy ra trong tương lai.	<ul style="list-style-type: none"> - Kiểu số - Sau 1995 - Lớn nhất trong các năm đã xét

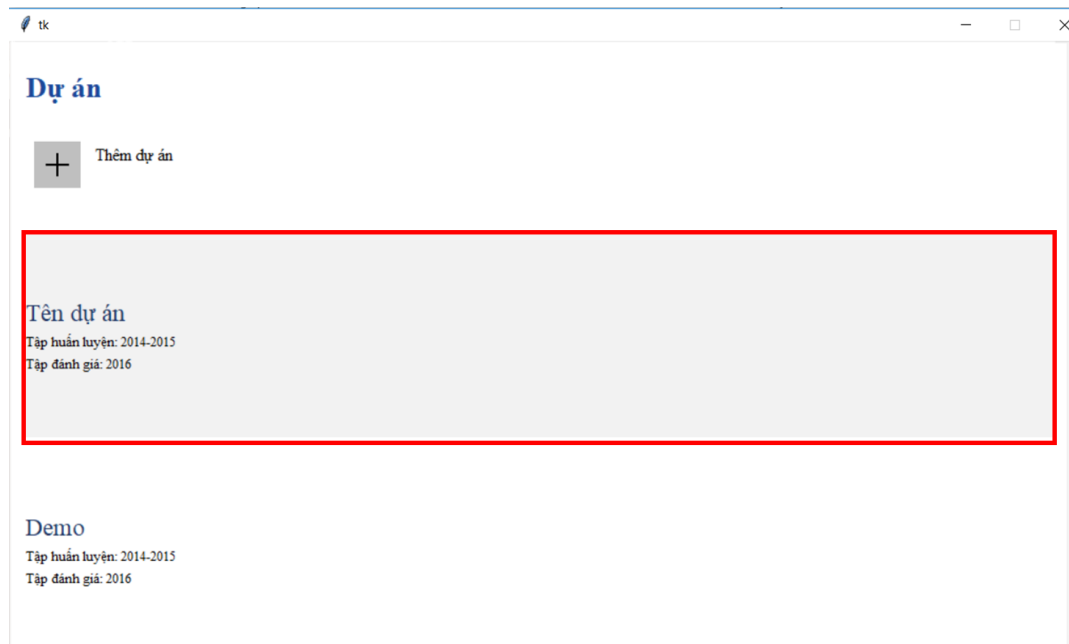
Bảng 2-1 Trường dữ liệu mô tả theo hình Hình 2-15

Hình 2-15 Khởi tạo dự án so sánh thuật toán khuyến nghị cộng tác. Trong đó, từ 1 đến 5: xem tại Bảng 2-1; 6: Cảnh báo trường dữ liệu; 7. Nút dừng để khởi tạo dự án

Hình 2-16 Kết quả khởi tạo dự án

2.6.8. Xem dữ liệu thực nghiệm

Để giúp bạn đọc có thể xem dữ liệu thực nghiệm tại từng dự án. Bạn đọc có thể chọn dự án muốn xem dữ liệu thực nghiệm



Hình 2-17 Chọn dự án cần xem dữ liệu thực nghiệm

Khi đó, dữ liệu thực nghiệm sẽ được xác định bằng số liệu xác định ở 2 cột “Tập huấn luyện” và “Tập đánh giá” (thể hiện tại hình ...).

tk

←

DỰ ÁN

Tên dự án

Tập dữ liệu

Tập huấn luyện

Gồm 319247 nghiên cứu viên với điều kiện:
 1. Là tác giả ít nhất 1 bài báo 2014-2015.
 2. Bài báo đó có ít nhất 1 đồng tác giả.

Xây dựng mạng huấn luyện:
 $CoNet_1=(U_1, Co_1)$
 Trong đó:
 + U_1 : 319247 nghiên cứu viên.
 + 202864 bài báo.
 + Co_1 : 1008797 liên kết

Tập đánh giá

Gồm 114327 nghiên cứu viên với điều kiện:
 1. Từ tập huấn luyện.
 2. Là tác giả ít nhất 1 bài báo 2016.
 3. Mỗi bài báo có ít nhất 1 đồng tác giả.

Xây dựng mạng đánh giá:
 $CoNet_2=(U_2, Co_2)$
 Trong đó:
 + U_2 : 114327 nghiên cứu viên.
 (U_2 là tập con của U_1)
 + 84340 bài báo.
 + Co_2 : 283711 liên kết (thực sự)

Thực nghiệm, đánh giá

Precision

Recall

F-measure

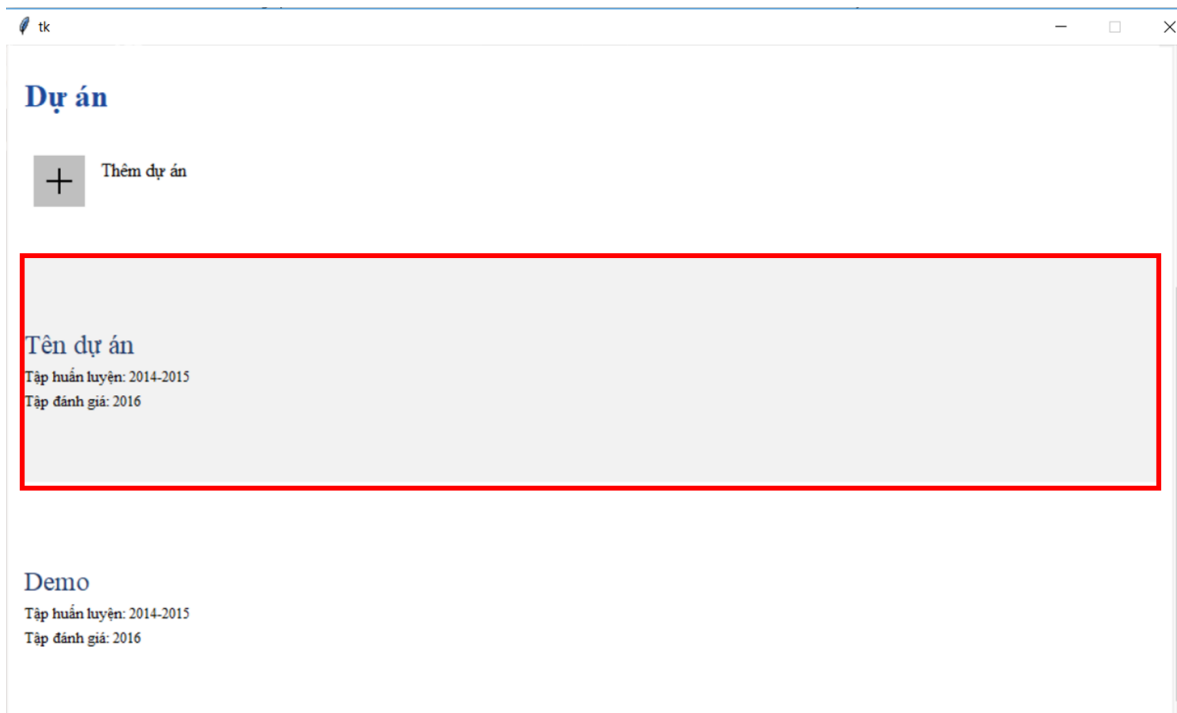
⚠ Chưa hoàn tất huấn luyện nên không thể đánh giá

Hình 2-18 Thông tin chi tiết về dữ liệu thực nghiệm

2.6.9. Tạo mới phương pháp khuyến nghị

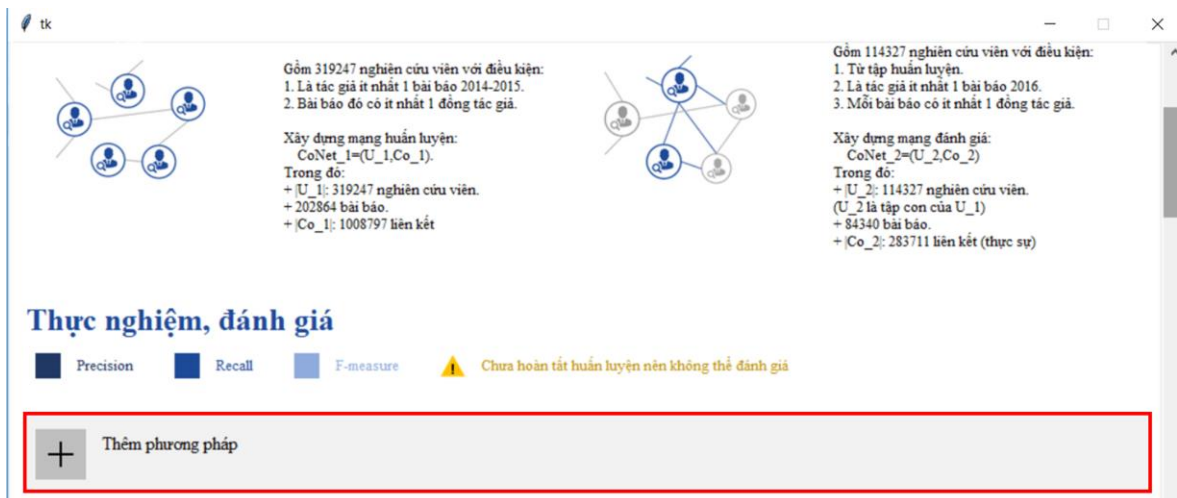
Để giúp bạn đọc có thể thực nghiệm, so sánh một số thuật toán khuyến nghị, phần mềm thiết kế chức năng thêm phương pháp khuyến nghị vào dự án. Bạn đọc có thể thực hiện theo các bước sau:

- **Bước 1:** Chọn dự án cần thêm mới thuật toán



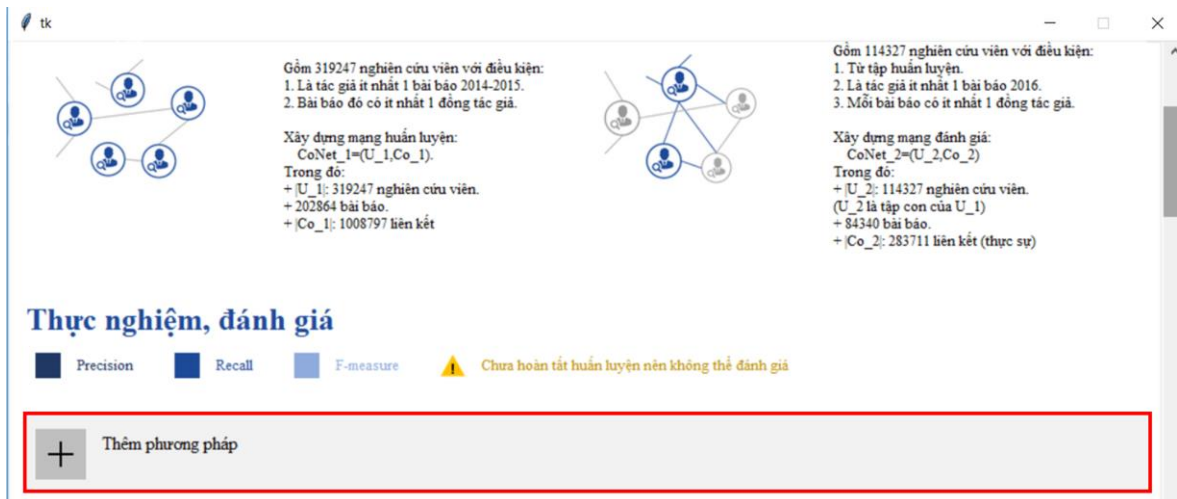
Hình 2-19 Chọn dự án cần thêm mới thuật toán

- **Bước 2:** Chọn nút “Thêm phương pháp”.



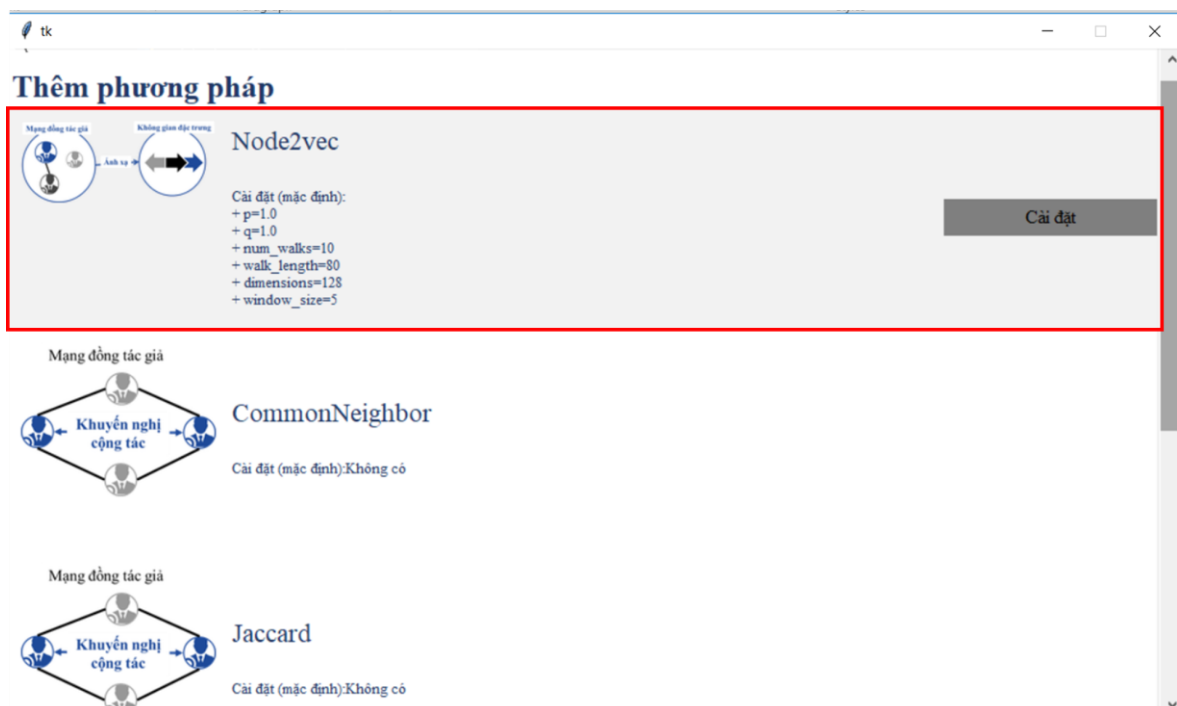
Hình 2-20 Chọn “Thêm phương pháp”

- **Bước 3:** Chọn nút “Thêm phương pháp”.



Hình 2-21 Chọn “Thêm phương pháp”

- **Bước 4:** Chọn phương pháp.



Hình 2-22 Chọn phương pháp cần thêm. **Lưu ý:** với Node2vec, bạn đọc có thể chọn cài đặt và thay đổi tham số để có được kết quả tiên đoán chính xác hơn tùy theo mạng đồng tác giả xây dựng.

- **Bước 5:** Tùy chỉnh tham số của phương pháp (Không bắt buộc).



Hình 2-23 Tùy chỉnh tham số của thuật toán.

2.6.10. Chạy và đánh giá phương pháp

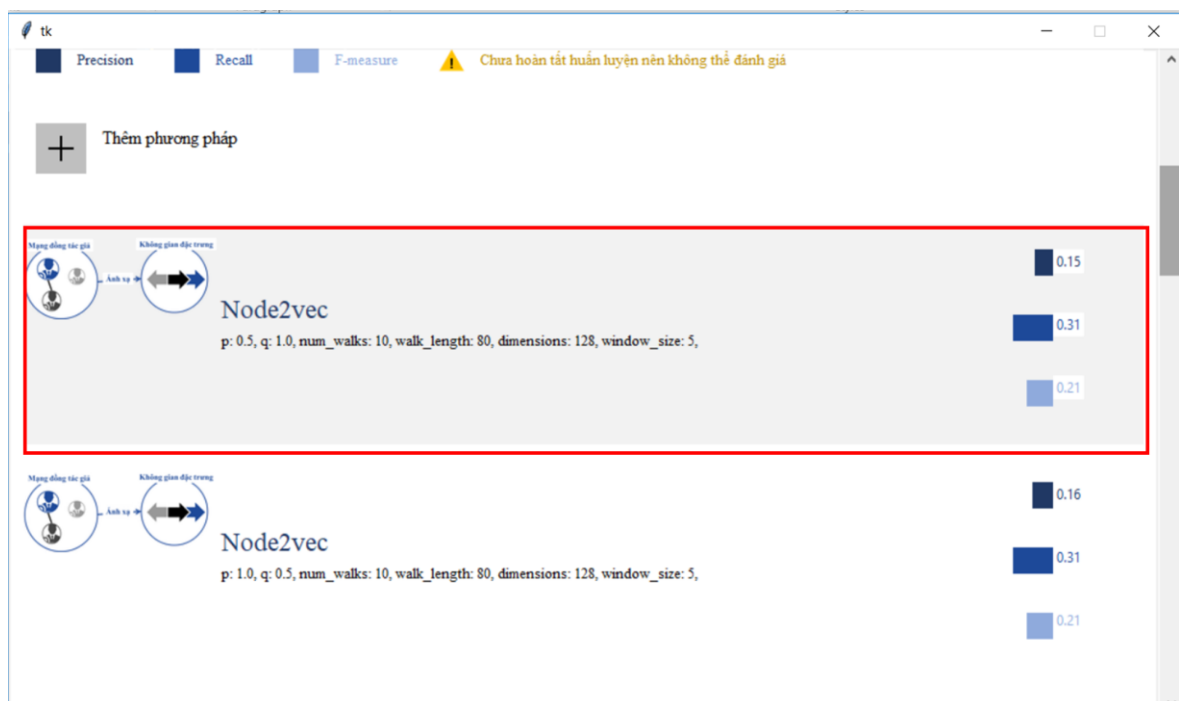
Với từng phương pháp tạo mới, bạn đọc có thể thực hiện chạy thuật toán phương pháp trên tập dữ liệu huấn luyện và đánh giá trên tập đánh giá. Các bước thực hiện như sau:

- **Bước 1:** Chọn phương pháp cần đánh giá



Hình 2-24 Chọn phương pháp cần đánh giá

- **Bước 2:** Thực hiện chạy phương pháp khuyến nghị để hệ thống thực hiện khuyến nghị.



Hình 2-25 Chọn phương pháp cần đánh giá

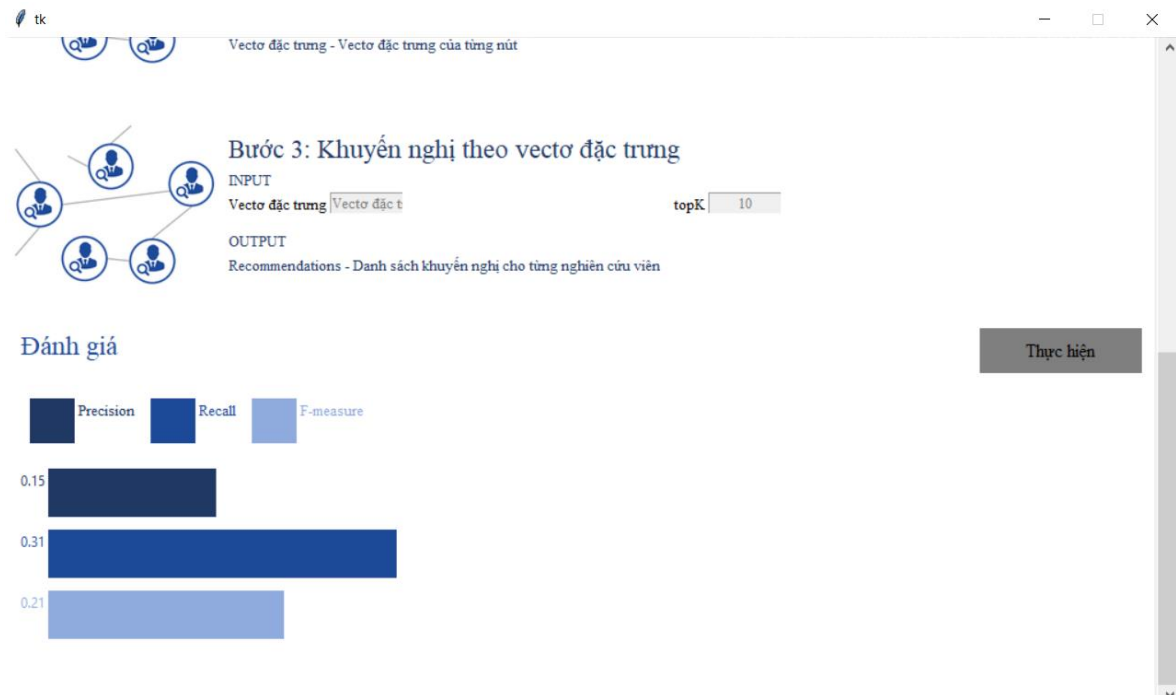


Hình 2-26 Chọn nút “Thực hiện” để khởi chạy phương pháp

- **Bước 3:** Thực hiện đánh giá phương pháp.



Hình 2-27 Chọn nút “Thực hiện” để đánh giá phương pháp



Hình 2-28 Kết quả đánh giá phương pháp

2.1. Kết chương

Với những hướng dẫn phía trên, bạn đọc có thể thêm các thuật toán, thay đổi các thông số và thực hiện, đánh giá, so sánh chúng với nhau nhằm đạt được kết quả cao nhất có thể trong khuyến nghị cộng tác.

Chương 3. KẾT

Phần mềm là kết quả đính kèm của khóa luận “Khuyến nghị cộng tác dựa trên tiếp cận học sâu” của sinh viên Khoa Công Nghệ Phần Mềm Đại học Công Nghệ Thông Tin- Đại học Quốc gia Hồ Chí Minh với mục đích demo tiếp cận học sâu cho khuyến nghị cộng tác. Hiện tại, phần mềm chưa giải quyết các bài toán khác. Nếu bạn đọc quan tâm, bạn đọc có thể tìm được mã nguồn phát triển phần mềm này tại: https://github.com/datnhemployee/khuyennghicongtac_15520104. Cảm ơn bạn đọc đã dùng phần mềm và chúc bạn đọc thành công.