TỔNG LIÊN ĐOÀN LAO ĐỘNG VIỆT NAM

**TRƯỜNG ĐẠI HỌC TÔN ĐỨC THẮNG**

**KHOA CÔNG NGHỆ THÔNG TIN**



**BÁO CÁO CUỐI KỲ**

**MÔN: XỬ LÝ DỮ LIỆU LỚN**

**XÂY DỰNG HỆ THỐNG GỢI Ý KẾT BẠN**

**(RECOMMENDER SYSTEM)**

*Người hướng dẫn*: **TS BÙI THANH HÙNG**

*Người thực hiện*: **NGUYỄN TRƯỜNG ANH THINH – 518H0571**

**LÊ HOÀNG LONG – 518H0035**

Lớp**: 18H50205**

Khoá**: 22**

THÀNH PHỐ HỒ CHÍ MINH, NĂM 2021

TỔNG LIÊN ĐOÀN LAO ĐỘNG VIỆT NAM

**TRƯỜNG ĐẠI HỌC TÔN ĐỨC THẮNG**

**KHOA CÔNG NGHỆ THÔNG TIN**



**BÁO CÁO CUỐI KỲ**

**MÔN: XỬ LÝ DỮ LIỆU LỚN**

**XÂY DỰNG HỆ THỐNG GỢI Ý KẾT BẠN**

**(RECOMMENDER SYSTEM)**

Người hướng dẫn: **TS BÙI THANH HÙNG**

Người thực hiện: **NGUYỄN TRƯỜNG ANH THỊNH – 518H0571**

**LÊ HOÀNG LONG – 518H0035**

Lớp**: 18H50205**

Khoá**: 22**

**THÀNH PHỐ HỒ CHÍ MINH, NĂM 2021**

**LỜI CẢM ƠN**

Chúng em xin gửi lời cảm ơn tới giảng viên bộ môn thầy Bùi Thanh Hùng đã dạy dỗ, giảng dạy nhiệt tình và dành thời gian quý báu giúp chúng em hiểu và làm bài dễ hơn, trong quá trình học có một vài tuần gặp khó khăn do ảnh hưởng dịch Covid-19 phải học online nhưng thầy vẫn tạo điều kiện giúp đỡ sinh viên gặp khó khăn trong quá trình làm bài. Do chưa có nhiều kinh nghiệm làm để tài, trong bài tiểu luận chắc chắn sẽ không tránh khỏi những thiếu sót. Rất mong nhận được sự nhận xét, ý kiến đóng góp, phê bình từ phía thầy và các bạn. Chúng em chúc thầy luôn nhiều sức khỏe để tiếp tục đóng góp cho sự nghiệp giáo dục của nước nhà.

**CÔNG TRÌNH ĐƯỢC HOÀN THÀNH**

**TẠI TRƯỜNG ĐẠI HỌC TÔN ĐỨC THẮNG**

Tôi xin cam đoan đây là công trình nghiên cứu của riêng tôi và được sự hướng dẫn khoa học của TS Bùi Thanh Hùng;. Các nội dung nghiên cứu, kết quả trong đề tài này là trung thực và chưa công bố dưới bất kỳ hình thức nào trước đây. Những số liệu trong các bảng biểu phục vụ cho việc phân tích, nhận xét, đánh giá được chính tác giả thu thập từ các nguồn khác nhau có ghi rõ trong phần tài liệu tham khảo.

Ngoài ra, trong luận văn còn sử dụng một số nhận xét, đánh giá cũng như số liệu của các tác giả khác, cơ quan tổ chức khác đều có trích dẫn và chú thích nguồn gốc.

**Nếu phát hiện có bất kỳ sự gian lận nào tôi xin hoàn toàn chịu trách nhiệm về nội dung luận văn của mình.** Trường Đại học Tôn Đức Thắng không liên quan đến những vi phạm tác quyền, bản quyền do tôi gây ra trong quá trình thực hiện (nếu có).

*TP. Hồ Chí Minh, ngày tháng năm*

*Tác giả*

*(ký tên và ghi rõ họ tên)*

*Nguyễn Trường Anh Thịnh*

*Lê Hoàng Long*

**TÓM TẮT**

Trong bài báo cáo này, chúng ta sẽ được cung cấp một số phương pháp khả thi để xây dựng hệ thống gợi ý kết bạn (Recommender system). Khi đi sâu vào thuật toán cũng như thực nghiệm ta cũng đã có thêm những cái nhìn khái quát về cách tiếp cận cũng như cách mà chúng hoạt động trong mô hình Recommender system. Qua đó giúp chúng ta giải quyết được bài toán gợi ý kết bạn. Đồng thời báo cáo cũng cung cấp phép đo về độ hiệu quả giữa các phương pháp được cung cấp. Với sự tiến bộ của khoa học và kỹ thuật ngày nay, hệ thống gợi ý kết bạn dần trở thành một công cụ phổ biến cho các ứng dụng mạng xã hội được hàng triệu người quan tâm như Facebook, Skype, Instagram, ...Và các phương pháp cung cấp sau đây có thể là một giải pháp khả quan để ứng dụng vào các công cụ xã hội trong đời sống thực tiễn ngày nay.

**MỤC LỤC**

[**I.** **Giới thiệu về bài toán** 10](#_Toc91259365)

[**II.** **Phân tích yêu cầu bài toán** 11](#_Toc91259366)

[**III.** **Phương pháp giải quyết** 11](#_Toc91259367)

[a. Phương pháp gợi ý dựa theo số lượng bạn chung(number of common friends): 11](#_Toc91259368)

[b. Phương pháp gợi ý dựa theo điểm ảnh hưởng(influence score): 13](#_Toc91259369)

[c. Đánh giá phương pháp: 14](#_Toc91259370)

[d. Sơ đồ tổng quát giải quyết bài toán: 14](#_Toc91259371)

[**IV.** **Thực nghiệm** 16](#_Toc91259372)

[a. Dữ liệu 16](#_Toc91259373)

[b. Xử lý dữ liệu 16](#_Toc91259374)

[c. Công nghệ sử dụng 16](#_Toc91259375)

[d. Lưu đồ giải thuật 17](#_Toc91259376)

[e. Code thực nghiệm 18](#_Toc91259377)

[**V.** **Kết luận** 21](#_Toc91259378)

[a. Kết quả đạt được 21](#_Toc91259379)

[b. Hạn chế 26](#_Toc91259380)

[c. Hướng phát triển 27](#_Toc91259381)

[**VI.** **Tài liệu tham khảo** 28](#_Toc91259382)

**DANH MỤC KÝ HIỆU VÀ CHỮ VIẾT TẮT**

**CÁC CHỨ VIẾT TẮT:**

**NOCF :** Number of common friends – Số lượng bạn chung

**IS:** Influence score – Điểm ảnh hưởng

**FRS:** Friendrecommender system – Hệ thống gợi ý kết bạn

**DANH MỤC HÌNH ẢNH**

[Hình 1 Đồ thị minh họa quan hệ bạn bè 12](#_Toc91325356)

[Hình 2 Sơ đồ tổng quát theo phương pháp IS 15](file:///C:\Users\thinh\Downloads\longlozz-1.docx#_Toc91325357)

[Hình 3 Lưu đồ giải thuật theo phương pháp IS 17](file:///C:\Users\thinh\Downloads\longlozz-1.docx#_Toc91325358)

[Hình 4 Code tạo đồ thị từ tập dữ liệu đầu vào 18](#_Toc91325359)

[Hình 5 Đồ thị được tạo từ tập dữ liệu 18](file:///C:\Users\thinh\Downloads\longlozz-1.docx#_Toc91325360)

[Hình 6 Hàm tìm các node liền kế với node đã chọn cùng kết quả 19](#_Toc91325361)

[Hình 7 Hàm tìm các node mới có node phổ biến chung vói node “0” và kết quả 19](#_Toc91325362)

[Hình 8 Hàm tìm các node phổ biến chung giữa node mới và node “0” cùng kết quả 20](#_Toc91325363)

[Hình 9 Hàm tính điểm ảnh hưởng giữa các node mới và node “0”cùng kết quả đạt dc 21](#_Toc91325364)

[Hình 10 Hàm sắp xếp và danh sách các node gợi ý cho node”0” sau khi được sắp xếp 21](file:///C:\Users\thinh\Downloads\longlozz-1.docx#_Toc91325365)

[Hình 11 Hàm xác đinh thứ tự uu tiên 22](#_Toc91325366)

[Hình 12 Hàm tính thứ tự ưu tiên trung bình của F1 và F2 23](#_Toc91325367)

[Hình 13 Code thực hiện tính mức độ ưu tiên trung bình sau 100 lần chạy 24](#_Toc91325368)

[Hình 14 Kết quả thực nghiệm 24](#_Toc91325369)

[Hình 15 Biểu đồ so sánh bậc xếp hạng trung bình giữa hai phương pháp 25](file:///C:\Users\thinh\Downloads\longlozz-1.docx#_Toc91325370)

[Hình 16 Đồ thị quan gia đình Romeo-Juliet 25](file:///C:\Users\thinh\Downloads\longlozz-1.docx#_Toc91325371)

[Hình 17 Kết quả và biểu đồ so sánh kết quả giữa 2 phương pháp 26](file:///C:\Users\thinh\Downloads\longlozz-1.docx#_Toc91325372)

**DANH MỤC CÁC BẢNG**

[Bảng 1 Mô tả công nghệ sử dụng cho bài toán gợi ý kết bạn 16](#_Toc91325551)

[Bảng 2 Bảng tự đánh giá mức độ hoàn thành 29](#_Toc91325552)

**HỆ THỐNG GỢI Ý KÉT BẠN – RECOMMENDER SYSTEM**

# **Giới thiệu về bài toán**

* Trong thời buổi công nghệ phát triển như hiện chắc hẳn ai trong số chúng ta cũng đều đã trải qua những hiện tượng này:
  + Các ứng dụng xem phim trực tuyến như netflix tự động gợi ý phim cho người dùng.
  + Youtube hay tiktok tự động chuyển các clip liên quan đến những clip bạn đã xem. Chúng cũng tự gợi ý những clip mà bạn có thể sẽ thích.
  + Khi mua một món hàng trên Shopee, Tiki, …, hệ thống sẽ tự động gợi ý những món hàng thường mua chung với nhau, hoặc gợi ý những món hàng bạn có thể sẽ thích dựa trên lịch sử mua hàng.
  + Facebook hiển thị quảng cáo những sản phẩm có liên quan đến từ khóa mà bạn vừa tìm kiếm hay thậm chí là những từ khóa trong tin nhắn của bạn.
  + Facebook gợi ý kết bạn.
* Và vẫn còn nhiều ví dụ khác về các ứng dụng sở hữu hệ thống có khả năng tự động gợi ý cho người dùng những sản phẩm họ có thể thích và quan tâm. Dựa trên các dữ liệu phù hợp từ người dùng cung cấp. Qua đó, đem lại nhiều lợi ích cho cả người dùng lẫn người cung cấp dịch vụ. Những thuật toán đằng sau các ứng dụng trên chính là những thuật toán Machine Learning có tên gọi chung là Recommender System hay Recommendation System (Hệ thống gợi ý).
* Recommender System là một lĩnh vực khá rộng của học máy, và có phần mới mẻ hơn so với Classification vì Internet chỉ mới thực sự bùng nổ vào khoảng một thập niên đổ lại đây. Có hai đối tượng chính trong Recommender System là users và items. Users là người dùng và items là các sản phẩm, ví dụ như bộ phim, bài hát, clip và ở đây đối với bài toán gợi ý kết bạn của chúng ta thì items cũng chính là các users khác. Mục đích chính của hệ thống này là dự đoán mức độ quan tâm của một user tới một item nào đó, qua đó đưa ra được các gợi ý phú hợp.

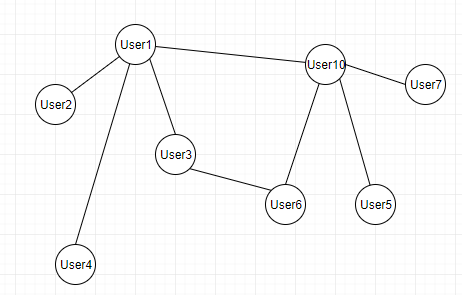
# **Phân tích yêu cầu bài toán**

* Ta có thể mô tả bài toán gợi ý kết banh như sau : Chúng ta sẽ có một đồ thị G=(V, E) trong đó V là tập hợp các đỉnh của đồ thị (bao gồm tên các user) và E là tập hợp cạnh của đồ thị (thể hiện mối quan hệ bạn bè giữa các user). Nhiệm vụ của việc gợi ý kết bạn là từ một đỉnh V bất kỳ, và một tập hợp con các liên kết với đỉnh khác, ta phải đưa ra được một tập hợp các đỉnh gợi ý theo thứ tự ưu tiên và các đỉnh đó chưa có các liên kết trực tiếp với đỉnh đã chọn.

# **Phương pháp giải quyết**

## Phương pháp gợi ý dựa theo số lượng bạn chung(number of common friends):

* + Ý tưởng cơ bản của phương pháp NOCF là xác định mức độ quan tâm của một user tới một item dựa trên các users khác gần giống với user này. Việc gần giống nhau giữa các users có thể được xác định thông qua mức độ quan tâm của các users này tới các items khác mà hệ thống đã biết. Ví dụ, A, B đều thích phim SpiderMan. Ta còn biết thêm A cũng thích phim Superman, vậy nhiều khả năng B cũng thích bộ phim này.
  + Áp dụng vào FRS, ta cần nắm được một trong những yếu tố quan trọng khi gợi ý những người bạn mới đến user theo phương pháp này chính là số lượng bạn chung và những người bạn chung giữa họ. Ví dụ như User1 có các người bạn là User2, User3, User4. User10 có các người bạn là User1, User5, User6, User7. Và User6 có người bạn là User3.



Hình 1 Đồ thị minh họa quan hệ bạn bè

* + Để tìm được những người bạn mới phù hợp cho User1 ta thực hiện các bước như sau:
    - Đầu tiên, dựa vào đồ thị trên chúng ta cần xác định được những người bạn của User1, ta có được những người bạn của User1 là User2, User3, User4, User10.
    - Tiếp theo, dựa trên nhóm những người bạn của User1, chúng ta có được danh sách gợi ý gồm những người bạn của nhóm người trên gồm: User5, User6, User7.
    - Cuối cùng ta sắp xếp mức độ ưu tiên dựa vào số bạn chung giữa những người trong danh sách trên với User1: User6(2), User5(1), User7(1).

## Phương pháp gợi ý dựa theo điểm ảnh hưởng(influence score):

* + Hãy xem xét tình huống giả định sau:
    - Bạn có 2 người bạn là Long và Thịnh.
    - Thịnh chỉ có 2 người bạn là bạn và một người khác.
    - Long thì có đến tận 9 triệu người bạn.
    - Vì Thịnh là một người rất kỹ tính và có sự chọn lọc trong việc kết bạn với người khác, và với tư cách là một người bạn của Thịnh bạn có thể có nhiều điểm chung với những người bạn khác của bạn ấy. Mặt khác, Long là một người rất phóng khoáng và thường kết bạn bừa bãi, nên có rất ít khả năng bạn có thể có được mối quan hệ thân thiết với những người bạn của bạn ấy.
  + Từ tình huống giả định trên ta đã rút ra được một ý tưởng về phương pháp giải thuật mới giúp giải quyết bài toán xây dựng hệ thống gợi ý kết bạn được gọi là kỹ thuật điểm ảnh hưởng (Influence score).
  + Ví dụ rằng User1 và User2 có 3 người bạn chung là f1, f2 và f3. Ta có thể tính được điểm bạn bè (điểm ảnh hưởng) giữa User1 đối với User2 theo công thức là:

**Score =**

Với numfriends(f) là số bạn bè mà f có. Với công thức ta thấy được điểm ảnh hưởng giữa User1 và User sẽ được tính bằng tổng nghịch đảo số lượng bạn bè của từng người bạn chung giữa User1 và User2.

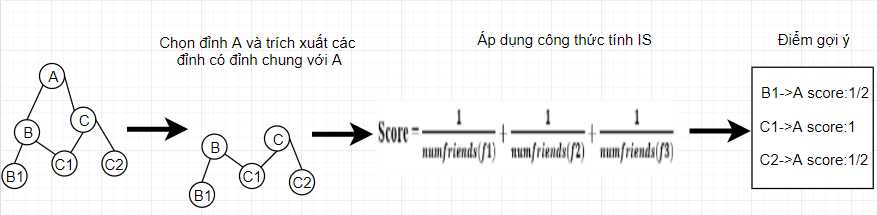
* + Quay lại tình huống giả định lúc đầu, giờ đây ta có thể có cái nhìn chính xác hơn về điểm ảnh hưởng giữa những người bạn của Thịnh và Long đối với bạn. Theo công thức đã có ta có thể tính được rằng điểm ảnh hưởng giữa bạn với những người bạn của Thịnh là 1/2 và điểm ảnh hưởng giữa bạn đối với những người bạn của Long là 1/9000000. Từ đây, hệ thống có thể đưa ra các gợi ý kết bạn với mức độ ưu tiên tùy theo số điểm ảnh hưởng.

1. **Đánh giá phương pháp:**

* Theo lý thuyết đề cập ở trên ta tạm thời thấy được phương pháp gợi ý theo IS có phần tối ưu hơn so với NOCF, vì IS không chỉ gợi ý dựa theo số lượng bạn chung mà còn tính ra được mức độ tin tưởng(điểm ảnh hưởng) dựa vào số lượng bạn của những người bạn chung. Vì vậy, có vẻ như phương pháp IS sẽ đem đến kết quả đáng tin cậy và hiệu quả cao hơn so với phương pháp NOCF.

1. **Sơ đồ tổng quát giải quyết bài toán:**

* Như đã đề cập ở trên, ta sẽ chọn phương pháp IS để giải quyết bài toán FRS.
* Sơ đồ tổng quát của phương pháp IS để giải quyết bài toán sẽ được trình bày theo sơ đồ dưới đây. Trong sơ đồ này gồm có 3 phần chính:
  + Phần 1: Biểu diễn dữ liệu dưới dạng đồ thị.
  + Phần 2: Chọn 1 đỉnh bất kỳ và trích xuất các đỉnh có liên kết ít nhất 1 đỉnh chung với đỉnh đã chọn và không liên kết trực tiếp với đỉnh đã chọn.
  + Phần 3: Áp dụng công thức IS để tính điểm ảnh hưởng giữa đỉnh đã chọn và các đỉnh đã trích xuất được.



Hình 2 Sơ đồ tổng quát theo phương pháp IS

# **Thực nghiệm**

## Dữ liệu

* + Dữ liệu được dùng để thực hiện bài toán này là bộ dữ liệu được cung cấp bởi [Online Social Networks Research @ MPI-SWS](http://socialnetworks.mpi-sws.org/data-wosn2009.html).
  + File dữ liệu “facebook.txt” bao gồm một danh sách tất cả các liên kết giữa người dùng với người dùng từ mạng Facebook New Orleans và các liên kết này là vô hướng.
  + Mỗi một hàng trong dữ liệu này chứa hai số định danh của người dùng, với số thứ nhất là số định danh của chủ thể và số thứ hai là số định danh của bạn chủ thể đó.

## Xử lý dữ liệu

* + Vì bài toán yêu cầu dùng Graph-based để tiến hành thực nghiệm. Nên tập dữ liệu facebook cần phải được xây dựng lại dưới dạng đồ thị. Ngoài việc đó ra, trong quá trình thực nghiệm chúng ta sẽ tiền xử lý dữ liệu (kiểm tra missing data) trước khi thực hiện.

## Công nghệ sử dụng

|  |  |
| --- | --- |
| Ngôn ngữ | Python 3.8 |
| Thư viện | Networkx, matplotlib, random |
| Môi trường | Google Colab |

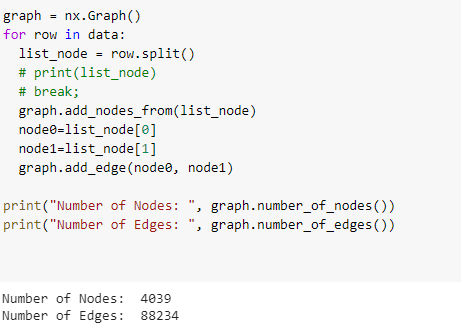
Bảng 1 Mô tả công nghệ sử dụng cho bài toán gợi ý kết bạn

## Lưu đồ giải thuật

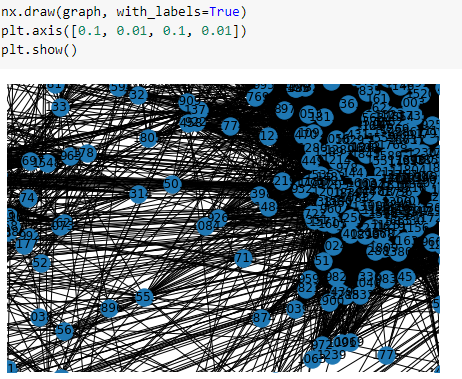
Hình 3 Lưu đồ giải thuật theo phương pháp IS

## Code thực nghiệm

* + Trước khi đi vào giải quyết bài toán ta sẽ chuẩn hóa lại dữ liệu theo dạng đồ thị (graph):

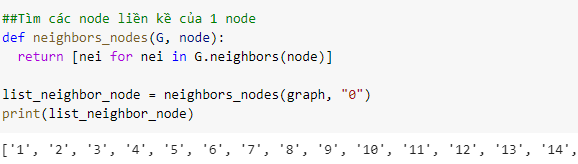
****

Hình 4 Code tạo đồ thị từ tập dữ liệu đầu vào



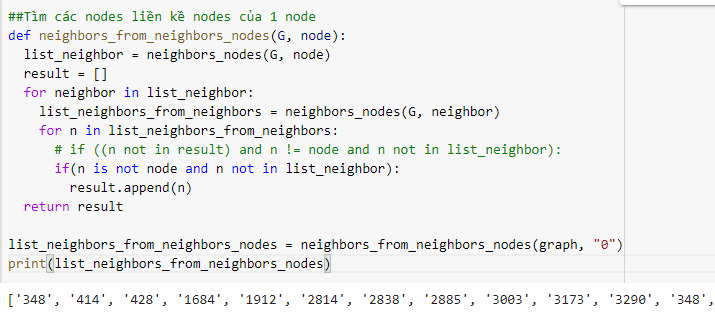
Hình 5 Đồ thị được tạo từ tập dữ liệu

* + Theo lưu đồ giải thuật đầu tiên chúng ta sẽ tìm các node liền kề với node đã chọn, ở đây ta sẽ chọn node “0” để tìm các gợi ý kết bạn.



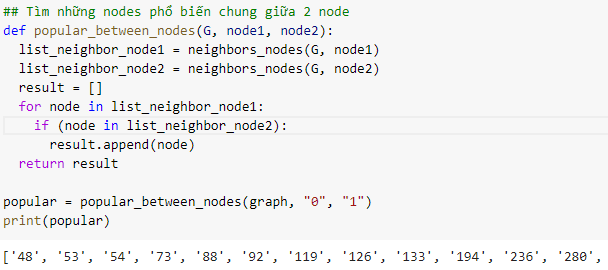
Hình 6 Hàm tìm các node liền kế với node đã chọn cùng kết quả

* + Tiếp theo ta tiếp tục tìm các node mới liền kề với từng node tìm được ở trên với điều kiện các node mới không thuộc tập hợp các node ở trên và không phải là node “0”.



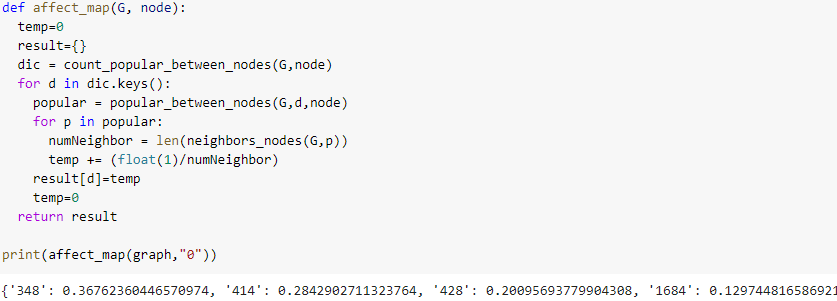
Hình 7 Hàm tìm các node mới có node phổ biến chung vói node “0” và kết quả

* + Khi đã có được các node mới, ta tiếp tục tìm các node phổ biến chung có liên kết giữa node mới và node “0”.

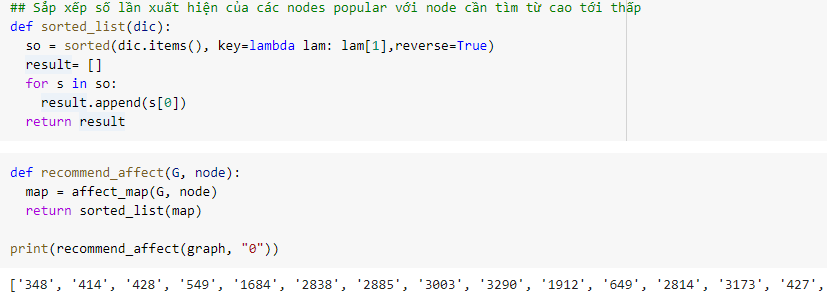


Hình 8 Hàm tìm các node phổ biến chung giữa node mới và node “0” cùng kết quả

* + Tiếp theo để tính được IS theo công thức ta cần phải tìm ra được số lượng bạn bè của những người bạn chung giữa “0” và người mới. Cụ thể ở đây, ta sẽ tìm các node liên kết với từng node phổ biến bằng cách dùng hàm neighbors\_nodes ở hình 6, sau đó dùng hàm len để lấy được độ dài (số lượng) từ node phổ biến ấy. Khi đã có được dữ liệu cần thiết ta thế dữ liệu vào công thức và tính được điểm ảnh hưởng (Influence score) giữa node mới và node “0”. Lưu ý mỗi node mới có thể có nhiều hơn 1 node chung phổ biến với node “0”, vì thế ta cần bỏ tất cả quá trình trên vào vòng lặp for để tính được từng số lượng node liên kết với node chung, qua đó tính được điểm ảnh hưởng từ tổng số đó.



Hình 9 Hàm tính điểm ảnh hưởng giữa các node mới và node “0”cùng kết quả đạt dc

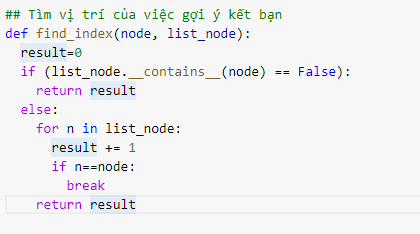
* + Cuối cùng ta sắp xếp lại các node mới được gợi ý cho node”0” dựa theo số điểm ảnh hưởng (influence score) với thứ tự lần lượt từ cao đến thấp.

Hình 10 Hàm sắp xếp và danh sách các node gợi ý cho node”0” sau khi được sắp xếp

# **Kết luận**

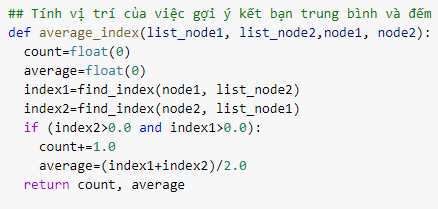
## Kết quả đạt được

* + Để kiểm tra độ chính xác của kết quả theo phương pháp IS, ta sẽ so sánh nó với thuật toán được đề xuất còn lại NOCF dựa trên đồ thị của tập dữ liệu Facebook.
  + Để tính được độ chính xác của cả 2 phương pháp FRC, ta sẽ thực hiện các bước sau đây:
    - 1. Chọn ngẫu nhiên 2 người bất kỳ đã là bạn của nhau, ta sẽ gọi họ là F1 và F2.
      2. Loại bỏ quan hệ(liên kết) giữa họ khỏi đồ thị.
      3. Tìm các gợi ý kết bạn cho F1 và F2.
      4. Xác định thứ tự ưu tiên của F1 trong danh sách gợi ý kết bạn của F2
      5. Xác định thứ tự ưu tiên của F2 trong danh sách gợi ý kết bạn của F1

****

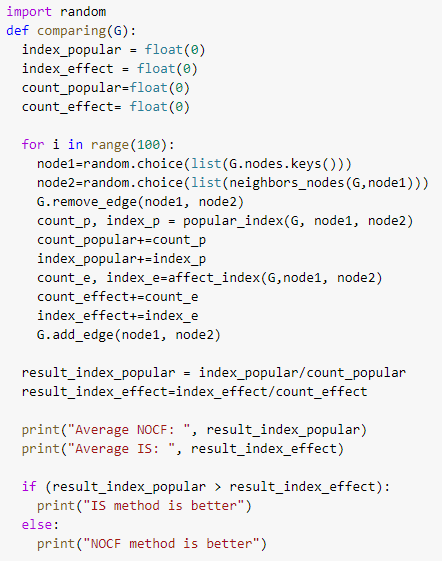
Hình 11 Hàm xác đinh thứ tự uu tiên

* + - 1. Nếu một trong hai không được đề xuất kết bạn cho nhau, bỏ cặp bạn bè này khỏi thực nghiệm.
      2. Ngược lại, ta sẽ tính tổng trung bình thứ tự ưu tiên từ thứ tự ưu tiên của F1 và F2 trong gợi ý kết bạn của nhau.



Hình 12 Hàm tính thứ tự ưu tiên trung bình của F1 và F2

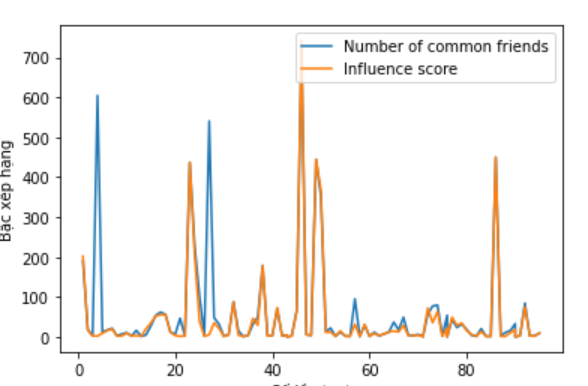
* + - 1. Sau khi có được kết quả, trả lại các liên kết trong đồ thị của F1 và F2 như ban đầu.
  + Ta sẽ thực hiện các bước trên với hai phương pháp đề ra là “Number of common friends” và “Influence score” để so sánh hiệu quả của chúng. Và vì có một số trường hợp khi hai người bạn không có trong danh sách gợi ý của người còn lại, nên bài test sẽ được lặp lại 100 lần và lấy điểm trung bình để có được kết quả chính xác và bao quát nhất.

****

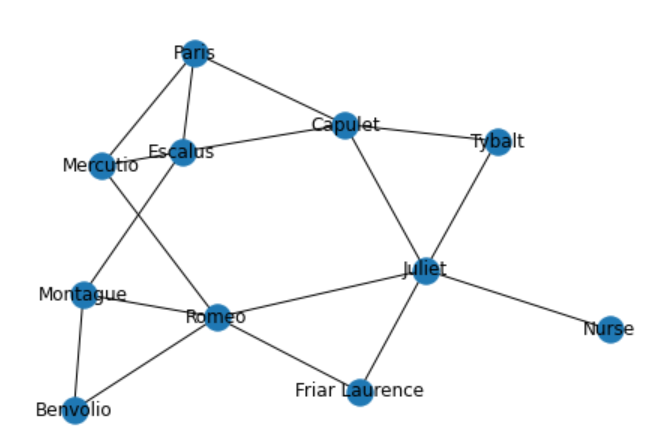
Hình 13 Code thực hiện tính mức độ ưu tiên trung bình sau 100 lần chạy



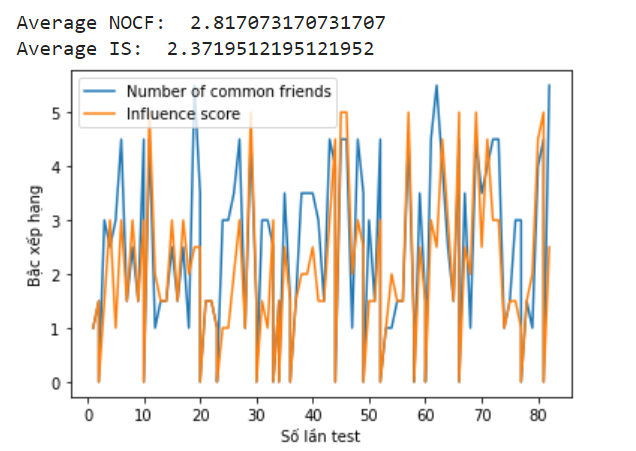
Hình 14 Kết quả thực nghiệm



Hình 15 Biểu đồ so sánh bậc xếp hạng trung bình giữa hai phương pháp

* + Để có thể đánh giá kết quả so sánh một cách khách quan hơn ta sẽ thử thực nghiệm ở đồ thị về mối quan hệ trong gia đinh giữa Romeo và Juliet.

Hình 16 Đồ thị quan gia đình Romeo-Juliet



Hình 17 Kết quả và biểu đồ so sánh kết quả giữa 2 phương pháp

* + Sau khi chạy thực nghiệm 2 thuật toán đề xuất trên toàn bộ tập dữ liệu, ta có thể thấy rằng phương pháp IS có kết quả tốt hơn so với NOCF vì thứ tự ưu tiên trung bình cho phương pháp IS có điểm số nhỏ hơn so với NOCF tương đương với thứ hạng cao hơn (ví dụ 1 nhỏ hơn 2 và thứ hạng 1 sẽ cao hơn thứ hạng 2).

## Hạn chế

* + Vì cả hai phương pháp trên đều dựa theo Collaborative Filtering nên chúng vẫn còn tồn tại những hạn chế sau đây:
    - Không thể gợi ý được cho những người dùng mới khi họ chưa có người bạn nào. Vì hệ thống gợi ý dựa trên thông tin bạn chung của những người bạn mới và đối tượng. Nên khi đối tượng không có người bạn nào, hệ thống sẽ không thể tìm được bất kỳ điểm chung nào để đưa ra gợi ý chính xác.
    - Một bộ dữ liệu thông thường sẽ không chỉ có 2 thông tin là đối tượng và bạn của họ mà còn có thêm thông tin phụ khác như độ tuổi, quốc gia, sở thích, thời gian trở thành bạn bè…. Những thông tin này có thể có ích trong việc cải thiện chất lượng của mô hình. Tuy nhiên, ở hiện tại, việc tích hợp chúng vào mô hình của chúng ta vẫn là một điều không hề dễ dàng.

## Hướng phát triển

* + Để cải tiến hơn nữa hệ thống gợi ý kết bạn này, chúng ta có thể sử dụng thêm thời gian mà mối quan hệ được hình thành(thời gian mà cả hai vừa trở thành bạn) để làm một tiêu chí đánh giá. Vì khi cả hai đã trở thành bạn với nhau trong thời gian dài thì sẽ có khả năng cao hơn hai người có nhiều điểm chung và thân thiết với nhau. Qua đó, kết quả đạt được sẽ có độ tin cậy và chính xác cao hơn khi ta thực hiện gợi ý kết bạn những người này cho người dùng khác. Đây sẽ là một hướng phát triển mới đầy hứa hẹn với dự án của chúng ta trong tương lai.

# **Tài liệu tham khảo**

* [Social networking and recommendation systems (washington.edu)](https://courses.cs.washington.edu/courses/cse140/13wi/homework/hw4/homework4.html)
* [Machine Learning cơ bản (machinelearningcoban.com)](https://machinelearningcoban.com/2017/05/24/collaborativefiltering/)
* [Online Social Networks Research @ MPI-SWS](http://socialnetworks.mpi-sws.org/data-wosn2009.html)
* [G5139059720.pdf (ijitee.org)](https://www.ijitee.org/wp-content/uploads/papers/v9i8/G5139059720.pdf)
* [Graph Plotting in Python | Set 1 - GeeksforGeeks](https://www.geeksforgeeks.org/graph-plotting-in-python-set-1/)

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
|  | **BẢNG TỰ ĐÁNH GIÁ** | | | | | | |
| Yêu cầu | |  | | Mức đánh giá | | | Điểm |
| 0% | 50% | | 75% | 100% |
| Phân tích yêu cầu bài toán | |  |  | |  | ✓ | 0.5 đ |
| Phương pháp giải quyết | |  |  | | ✓ |  | 1.75 đ |
| Hiện thực | |  |  | | ✓ |  | 4.25 đ |
| Đánh giá kết quả đạt được | |  |  | | ✓ |  | 1 đ |
| Hướng phát triển trong tương lai | |  |  | |  | ✓ | 0.25 đ |
| Trình bày báo cáo | |  |  | |  | ✓ | 1 đ |
| Làm việc nhóm | |  |  | |  | ✓ | 0.5 đ |
|  | Tổng: | | | | | | 9.25 đ |

Bảng 2 Bảng tự đánh giá mức độ hoàn thành