**Ảnh có chứa quảng trường

Mô tả được tạo tự độngĐẠI HỌC QUỐC GIA THÀNH PHỐ HỒ CHÍ MINH**

**TRƯỜNG ĐẠI HỌC CÔNG NGHỆ THÔNG TIN**

**KHOA HỆ THỐNG THÔNG TIN**

**BÁO CÁO ĐỒ ÁN CUỐI KỲ**

**MÔN MẠNG XÃ HỘI**

**Đề tài: TÌM HIỂU MỐI QUAN HỆ GIỮA NHÀ SẢN XUẤT GAME VÀ THỂ LOẠI GAME**



GVHD: ThS. Nguyễn Thị Kim Phụng

Sinh viên thực hiện:

1. Nguyễn Cao Khoa MSSV: 19521694

Tp. Hồ Chí Minh, 12/2022

**Ảnh có chứa quảng trường

Mô tả được tạo tự độngĐẠI HỌC QUỐC GIA THÀNH PHỐ HỒ CHÍ MINH**

**TRƯỜNG ĐẠI HỌC CÔNG NGHỆ THÔNG TIN**

**KHOA HỆ THỐNG THÔNG TIN**

**BÁO CÁO ĐỒ ÁN CUỐI KỲ**

**MÔN MẠNG XÃ HỘI**

**Đề tài: TÌM HIỂU MỐI QUAN HỆ GIỮA NHÀ SẢN XUẤT GAME VÀ THỂ LOẠI GAME**



GVHD: ThS. Nguyễn Thị Kim Phụng

Sinh viên thực hiện:

1. Nguyễn Cao Khoa MSSV: 19521694

Tp. Hồ Chí Minh, 12/2022

**NHẬN XÉT CỦA GIÁO VIÊN HƯỚNG DẪN**

*……., ngày……...tháng……năm 2022…*

**Người nhận xét**

*(Ký tên và ghi rõ họ tên****)***

**LỜI CẢM ƠN**

Nhóm xin chân thành cảm ơn cô Nguyễn Thị Kim Phụng - giảng viên môn Mạng Xã Hội đã nhiệt tình chỉ bảo và hướng dẫn tận tình trong suốt thời gian thực hiện đồ án môn học. Nhờ đó, chúng em đã có được nhiều kiến ​​thức bổ ích trong việc ứng dụng cũng như kỹ năng làm đồ án.

Bắt đầu từ việc tìm hiểu về mối quan hệ giữa nhà sản xuất game và thể loại game. Dựa trên những kiến ​​thức được giáo viên cung cấp trên lớp, kết hợp với việc tự nghiên cứu các công cụ và kiến ​​thức mới, nhóm cố gắng thực hiện dự án một cách tốt nhất có thể. Trong thời gian thực hiện dự án, nhóm chúng em đã sử dụng kiến ​​thức nền tảng đã tích lũy được, kết hợp học tập và nghiên cứu kiến ​​thức mới. Từ đó, nhóm của chúng em sử dụng đầy đủ các thông tin thu thập được để đưa ra báo cáo dự án tốt nhất có thể. Tuy nhiên, trong quá trình thực hiện, nhóm chúng em không tránh khỏi những thiếu sót. Vì vậy, nhóm rất mong nhận được sự góp ý của cô để nhóm em hoàn thiện hơn kiến ​​thức và chuẩn bị cho các đề tài khác trong tương lai.

Sau cùng, nhóm xin kính chúc cô thật dồi dào sức khỏe, niềm tin để tiếp tục thực hiện sứ mệnh cao đẹp là truyền đạt kiến thức cho các bạn sinh viên.

Xin chân thành cảm ơn Cô !

# DANH MỤC CÁC BẢNG, HÌNH ẢNH

**Danh mục bảng:**

[Bảng 1: Tên và ý nghĩa thuộc tính 9](#_Toc122933131)

[Bảng 2: Đổi tên các node 29](#_Toc122933132)

**Danh mục ảnh:**

[Hình 1: Số lượng nhà sản xuất, thể loại và cạnh 10](#_Toc122933261)

[Hình 2: Cho 2 cột dữ liệu vào Graph và vẽ lên đồ thị 11](#_Toc122933262)

[Hình 3: Đồ thị 2 hướng 12](#_Toc122933263)

[Hình 4: Tạo Graph dựa trên Graph đã tạo trước đó và vẽ lên đồ thị 13](#_Toc122933264)

[Hình 5: Đồ thị 1 hướng 13](#_Toc122933265)

[Hình 6: Tính degree centrality 15](#_Toc122933266)

[Hình 7: Kết quả tính được 15](#_Toc122933267)

[Hình 8: Tính betweeness centrality 16](#_Toc122933268)

[Hình 9: Kết quả tính được 16](#_Toc122933269)

[Hình 10: Tính closeness centrality 17](#_Toc122933270)

[Hình 11: Kết quả tính được 17](#_Toc122933271)

[Hình 12: Tính eigenvector centrality 18](#_Toc122933272)

[Hình 13: Kết quả tính được 18](#_Toc122933273)

[Hình 14: Tính pagerank 19](#_Toc122933274)

[Hình 15: Kết quả tính được 19](#_Toc122933275)

[Hình 16: Xuất đồ thị 1 hướng 20](#_Toc122933276)

[Hình 17: File csv đồ thị 1 hướng 20](#_Toc122933277)

[Hình 18: File csv đã được đọc vào Gephi 21](#_Toc122933278)

[Hình 19: Chạy các độ đo 21](#_Toc122933279)

[Hình 20: Kết quả tính được 22](#_Toc122933280)

[Hình 21: Kết quả tính được 23](#_Toc122933281)

[Hình 22: Kết quả tính được 24](#_Toc122933282)

[Hình 23: Kết quả tính được 25](#_Toc122933283)

[Hình 24: Kết quả tính được 26](#_Toc122933284)

[Hình 25: Xuất graph file đồ thị 1 hướng 27](#_Toc122933285)

[Hình 26: Ma trận của các node 27](#_Toc122933286)

[Hình 27: Ma trận 10 nodes 28](#_Toc122933287)

[Hình 28: Đồ thị 10 nodes đã được cho vào Gephi 28](#_Toc122933288)

[Hình 29: Đồ thị trước khi đổi trên node 29](#_Toc122933289)

[Hình 30: Đồ thị sau khi đổi tên node 29](#_Toc122933290)

[Hình 31: Tính betweeness centrality 30](#_Toc122933291)

[Hình 32: Tính closeness centrality 30](#_Toc122933292)

[Hình 33: Tính eigenvector centrality 31](#_Toc122933293)

[Hình 34: Tính pagerank 31](#_Toc122933294)

[Hình 35: Chạy thuật toán girvan newman và vẽ đồ thị 34](#_Toc122933295)

[Hình 36: Đồ thị girvan newman 34](#_Toc122933296)

[Hình 37: Chạy thuật toán louvain và vẽ đồ thị 35](#_Toc122933297)

[Hình 38: Đồ thị louvain 35](#_Toc122933298)

[Hình 39: Tính số lượng cụm 36](#_Toc122933299)

[Hình 40: Xuất cụm vào file csv 37](#_Toc122933300)

[Hình 41: Pivot cụm 0 38](#_Toc122933301)

[Hình 42: Biểu đồ dành cho nhà sản xuất 38](#_Toc122933302)

[Hình 43: Pivot cụm 1 39](#_Toc122933303)

[Hình 44: Biểu đồ dành cho nhà sản xuất 39](#_Toc122933304)

[Hình 45: Pivot cụm 2 40](#_Toc122933305)

[Hình 46: Biểu đồ dành cho nhà sản xuất 40](#_Toc122933306)

[Hình 47: Pivot cụm 3 41](#_Toc122933307)

[Hình 48: Biểu đồ dành cho nhà sản xuất 41](#_Toc122933308)

**MỤC LỤC**

[**LỜI CẢM ƠN** 4](#_Toc122933361)

[DANH MỤC CÁC BẢNG, HÌNH ẢNH 5](#_Toc122933362)

[**MỤC LỤC** 7](#_Toc122933363)

[PHẦN 1: TỔNG QUAN ĐỀ TÀI 10](#_Toc122933364)

[**1.** **Lí do chọn đề tài** 10](#_Toc122933365)

[**2.** **Mô tả dữ liệu** 10](#_Toc122933366)

[**2.1.** **Nguồn dữ liệu** 10](#_Toc122933367)

[**2.2.** **Mô tả dữ liệu** 10](#_Toc122933368)

[**2.3.** **Mô tả thuộc tính** 11](#_Toc122933369)

[PHẦN 2: ĐỒ THỊ 12](#_Toc122933370)

[**1.** **Khái niệm về Nodes và Edges** 12](#_Toc122933371)

[**1.1.** **Nodes và Edges** 12](#_Toc122933372)

[**1.2.** **Edges Direction** 12](#_Toc122933373)

[**1.3.** **Edges Weight** 12](#_Toc122933374)

[**2.** **Đồ thị 2 hướng** 12](#_Toc122933375)

[**2.1.** **Ý nghĩa đồ thị 2 hướng** 13](#_Toc122933376)

[**3.** **Đồ thị 1 hướng** 14](#_Toc122933377)

[**3.1.** **Ý nghĩa đồ thị 1 hướng** 14](#_Toc122933378)

[PHẦN 3: TÍNH TOÁN ĐỘ ĐO TRÊN ĐỒ THỊ 15](#_Toc122933379)

[**1.** **Khái niệm các độ đo** 15](#_Toc122933380)

[**2.** **Độ đo trên Google Colab** 16](#_Toc122933381)

[**2.1.** **Degree Centrality** 16](#_Toc122933382)

[**2.2.** **Betweeness Centrality** 17](#_Toc122933383)

[**2.3.** **Closeness Centrality** 18](#_Toc122933384)

[**2.4.** **EigenVector Centrality** 19](#_Toc122933385)

[**2.5.** **PageRank** 20](#_Toc122933386)

[**3.** **Độ đo trên Gephi** 21](#_Toc122933387)

[**3.1.** **Chuẩn bị dữ liệu** 21](#_Toc122933388)

[**3.2.** **Degree Centrality** 23](#_Toc122933389)

[**3.3.** **Betweeness Centrality** 24](#_Toc122933390)

[**3.4.** **Closeness Centrality** 25](#_Toc122933391)

[**3.5.** **EigenVector Centrality** 26](#_Toc122933392)

[**3.6.** **PageRank** 27](#_Toc122933393)

[PHẦN 4: TÍNH TAY ĐỘ ĐO TRÊN ĐỒ THỊ 10 NODES 28](#_Toc122933394)

[**1.** **Tạo đồ thị 10 nodes** 28](#_Toc122933395)

[**2.** **Đổi tên các node đồ thị** 30](#_Toc122933396)

[**3.** **Betweeness Centrality** 31](#_Toc122933397)

[**4.** **Closeness Centrality** 31](#_Toc122933398)

[**5.** **EigenVector Centrality** 32](#_Toc122933399)

[PHẦN 5: GRAPH MINING 33](#_Toc122933400)

[**1.** **Khái niệm các thuật toán** 33](#_Toc122933401)

[**1.1.** **Girvan Newman** 33](#_Toc122933402)

[**1.2.** **Louvain** 34](#_Toc122933403)

[**2.** **Girvan Newman** 35](#_Toc122933404)

[**3.** **Louvain** 36](#_Toc122933405)

[**3.1.** **Vẽ đồ thị** 36](#_Toc122933406)

[**3.2.** **Tổng quan các cụm** 37](#_Toc122933407)

[**3.3.** **Pivot các cụm** 39](#_Toc122933408)

[**3.3.1.** **Cụm 0** 39](#_Toc122933409)

[**3.3.2.** **Cụm 1** 40](#_Toc122933410)

[**3.3.3.** **Cụm 2** 41](#_Toc122933411)

[**3.3.4.** **Cụm 3** 42](#_Toc122933412)

[PHẦN 6: TÀI LIỆU THAM KHẢO 43](#_Toc122933413)

# PHẦN 1: TỔNG QUAN ĐỀ TÀI

1. **Lí do chọn đề tài**

Hiện nay với sự phát triển mạnh mẽ của thị trường game (cụ thể ở đây là những game trên nền tảng Xbox) trên thế giới thì game có mặt ở khắp nơi và là một phần không thể thiếu trong cuộc sống. Và với thị trường khắc nghiệt và với hàng ngàn sản phẩm game được ra mắt mỗi năm thì tính cạnh tranh là rất khốc liệt. Mỗi nhà sản xuất lại có một điểm mạnh riêng về một thể loại game nào đó hay có thể coi là họ nổi lên từ một sản phẩm game nào đó mà họ làm ra và thể loại của sản phẩm đó trở thành thế mạnh của họ từ đó. Nên việc một nhà sản xuất muốn tạo ra một sản phẩm game thì họ phải lựa chọn thể loại game trước và việc có thể biết được những thể loại là thế mạnh của các nhà sản xuất khác hay là thể loại mà họ sản xuất nhiều nhất thì có thể giúp cho công ty có thể lựa chọn chiến lược phù hợp để có thể bắt đầu phát triển một tựa game.

1. **Mô tả dữ liệu**
   1. **Nguồn dữ liệu**

Link Dataset: <https://www.kaggle.com/datasets/sidtwr/videogames-sales-dataset>

* 1. **Mô tả dữ liệu**
* Tên Dataset: **XboxOne\_GameSales.csv**
* Kích thước: **40.88 kB**
* Số cột: **10**
* Số dòng: **613** 
  1. **Mô tả thuộc tính**

Em sẽ chọn 2 cột đó là **Genre** và **Publisher** để phục vụ cho bài toán.

|  |  |
| --- | --- |
| **Tên cột** | **Ý nghĩa** |
| Genre | Thể loại game |
| Publisher | Nhà sản xuất game |

Bảng 1: Tên và ý nghĩa thuộc tính

* Số node (nhà sản xuất): **98**
* Số cạnh: **241**
* Số thể loại: **16**

**Text

Description automatically generated**

Hình 1: Số lượng nhà sản xuất, thể loại và cạnh

# PHẦN 2: ĐỒ THỊ

1. **Khái niệm về Nodes và Edges**
   1. **Nodes và Edges**

Một nút (node) được hiểu là đại điện một thực thể, một actor trong việc tạo ra liên kết mạng.

Cạnh (Edge): các cạnh thể hiện sư hiện diện của một kết nối hoặc mối quan hệ giữa hai nút.

* 1. **Edges Direction**

Có hai loại cạnh: cạnh có hướng và cạnh vô hướng. Trước khi xây dựng lên mạng cần phải giải mã dữ liệu của chúng ta chứa loại cạnh nào khi xây dựng biểu đồ mạng.

* Các cạnh được định hướng: được áp dụng từ nút này sang nút khác với một nút bắt đầu và một nút kết thúc.
* Các cạnh không định hướng: các mối quan hệ này được đáp lại bởi cả hai bên mà không có nút bắt đầu và nút kết thúc rõ ràng.
  1. **Edges Weight**

Trọng số của một cạnh là số lần cạnh đó xuất hiện giữa hai nút cụ thể.

1. **Đồ thị 2 hướng**

Cho 2 cột dữ liệu vào Graph và sau đó vẽ lên đồ thị 2 hướng.

**Graphical user interface, application

Description automatically generated**

Hình 2: Cho 2 cột dữ liệu vào Graph và vẽ lên đồ thị

Diagram

Description automatically generated

Hình 3: Đồ thị 2 hướng

* 1. **Ý nghĩa đồ thị 2 hướng**
* Các node bên trái : các nhà sản xuất game (98 nhà sản xuất).
* Các node bên phải : các thể loại game (16 thể loại game).
* Các cạnh : cho biết những thể loại mà nhà sản xuất đã làm ra cho sản phẩm của họ.

1. **Đồ thị 1 hướng**

Vẽ đồ thị 1 hướng dựa vào Graph đã tạo ra ở bước trên.

**Text

Description automatically generated**

Hình 4: Tạo Graph dựa trên Graph đã tạo trước đó và vẽ lên đồ thị

A picture containing graphical user interface

Description automatically generated

Hình 5: Đồ thị 1 hướng

* 1. **Ý nghĩa đồ thị 1 hướng**
* Các node : các nhà sản xuất game.
* Các cạnh : thể hiện sự giống nhau về thể loại game của các nhà sản xuất, tức là nếu 2 nhà sản xuất cùng sản xuất chung 1 thể loại game thì sẽ tạo thành 1 cạnh.
* Các node nằm giữa trung tâm : là các node có độ đo Betweeness, Closeness hay Pagerank cao do có các node đó có sản xuất nhiều thể loại game nhất dẫn đến có chung nhiều thể loại và liên kết với các nhà sản xuất khác nhất.

# PHẦN 3: TÍNH TOÁN ĐỘ ĐO TRÊN ĐỒ THỊ

1. **Khái niệm các độ đo**

* **Degree Centrality (Bậc):** bậc của một nút là số lượng liên kết đến các node đó.
* **Betweenness Centrality (Độ trung tâm trung gian):** tính sự ảnh hưởng của một node đối với sự lan truyền thông tin trong mạng. Một node có betweenness centrality càng lớn thì nó có vai trò quan trọng trong mạng xã hội bởi có thể nó là cầu nối liên kết giữa 2 nhóm với nhau và nếu nó bị mất thì việc trao đổi thông tin giữa hai nhóm sẽ không thể thực hiện được.
* **Closeness Centrality (Độ Trung tâm gần gũi):** độ gần gũi của mỗi node với các node khác trong mạng. Nếu một node có độ gần gũi càng thấp thì để giao tiếp với các node khác nó muốn thì phải đi qua rất nhiều node khác trong mạng.
* **Eigenvector Centrality (Vector đặc trưng trung tâm):** độ đo tương tự như bậc, tuy nhiên thay vì đếm số lượng liên kết đến node đó thì nó quan tâm đến bậc của node mà liên kết với nó. Trong một mạng, khi 2 node có cùng bậc thì đại lượng này sẽ cho biết node nào liên kết với các node quan trọng hơn ở trong mạng.
* **PageRank (Xếp hạng các node):** độ đo này là một biến thể của Eigenvector và điểm khác biệt chính đó là Pagerank hướng tới sự liên kết của node – ngược lại so với Eigenvecotor. Từ đó biết được node nào có sự liên kết nhiều nhất so với các node khác.
* **Clustering Coefficient (Hệ số phân cụm):** chính là số lượng liên kết trong một khu vực node trên tổng số liên kết có thể. Ví dụ như khu vực có 4 node mà có đến 8 liên kết thì chứng tỏ rất có thể chúng cùng trong một cụm.

1. **Độ đo trên Google Colab**
   1. **Degree Centrality**

**Text

Description automatically generated**

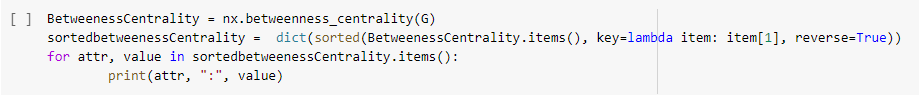
Hình 6: Tính degree centrality

A picture containing text, receipt

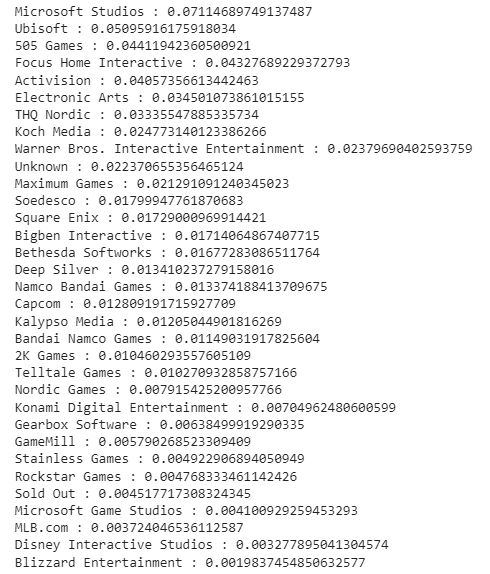
Description automatically generated

Hình 7: Kết quả tính được

* 1. **Betweeness Centrality**

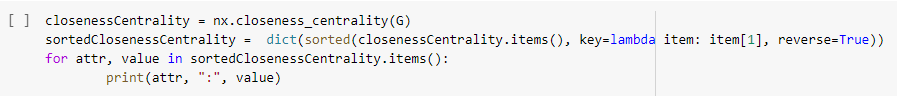
****

Hình 8: Tính betweeness centrality



Hình 9: Kết quả tính được

* 1. **Closeness Centrality**

****

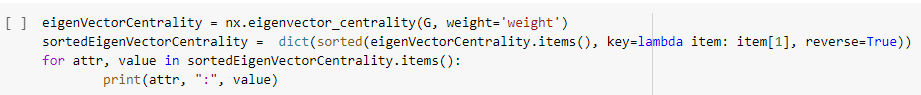
Hình 10: Tính closeness centrality

A black and white photo of a document

Description automatically generated with low confidence

Hình 11: Kết quả tính được

* 1. **EigenVector Centrality**

****

Hình 12: Tính eigenvector centrality

A page of a book

Description automatically generated with low confidence

Hình 13: Kết quả tính được

* 1. **PageRank**

**Graphical user interface, text

Description automatically generated**

Hình 14: Tính pagerank

Text

Description automatically generated

Hình 15: Kết quả tính được

1. **Độ đo trên Gephi**
   1. **Chuẩn bị dữ liệu**

**Bước 1:** Xuất đồ thị 1 hướng ra file csv.

A picture containing text

Description automatically generated

Hình 16: Xuất đồ thị 1 hướng

Table

Description automatically generated

Hình 17: File csv đồ thị 1 hướng

**Bước 2:** Cho file csv trên vào Gephi.

Graphical user interface, application

Description automatically generated

Hình 18: File csv đã được đọc vào Gephi

**Bước 3:** Run các độ đo cần thiết để Gephi tính toán.

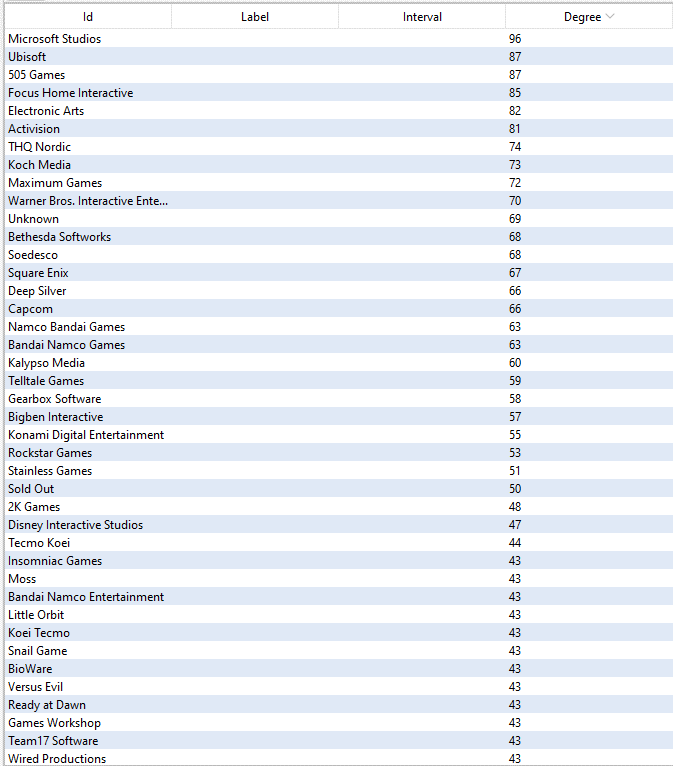
Graphical user interface, application

Description automatically generated

Hình 19: Chạy các độ đo

* 1. **Degree Centrality**

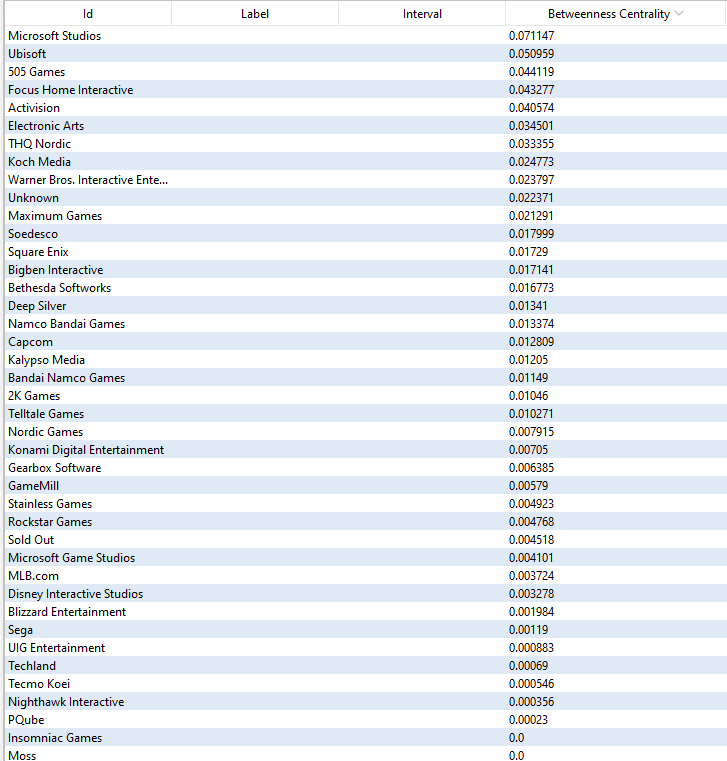
Cho kết quả giống Colab.

****

Hình 20: Kết quả tính được

* 1. **Betweeness Centrality**

Cho kết quả giống Colab.

****

Hình 21: Kết quả tính được

* 1. **Closeness Centrality**

Cho kết quả giống Colab.

**Table

Description automatically generated**

Hình 22: Kết quả tính được

* 1. **EigenVector Centrality**

Cho kết quả khác Colab (em cũng chưa biết lí do tại sao).

**Table

Description automatically generated**

Hình 23: Kết quả tính được

* 1. **PageRank**

Cho kết quả giống Colab.

**Table

Description automatically generated**

Hình 24: Kết quả tính được

# PHẦN 4: TÍNH TAY ĐỘ ĐO TRÊN ĐỒ THỊ 10 NODES

1. **Tạo đồ thị 10 nodes**

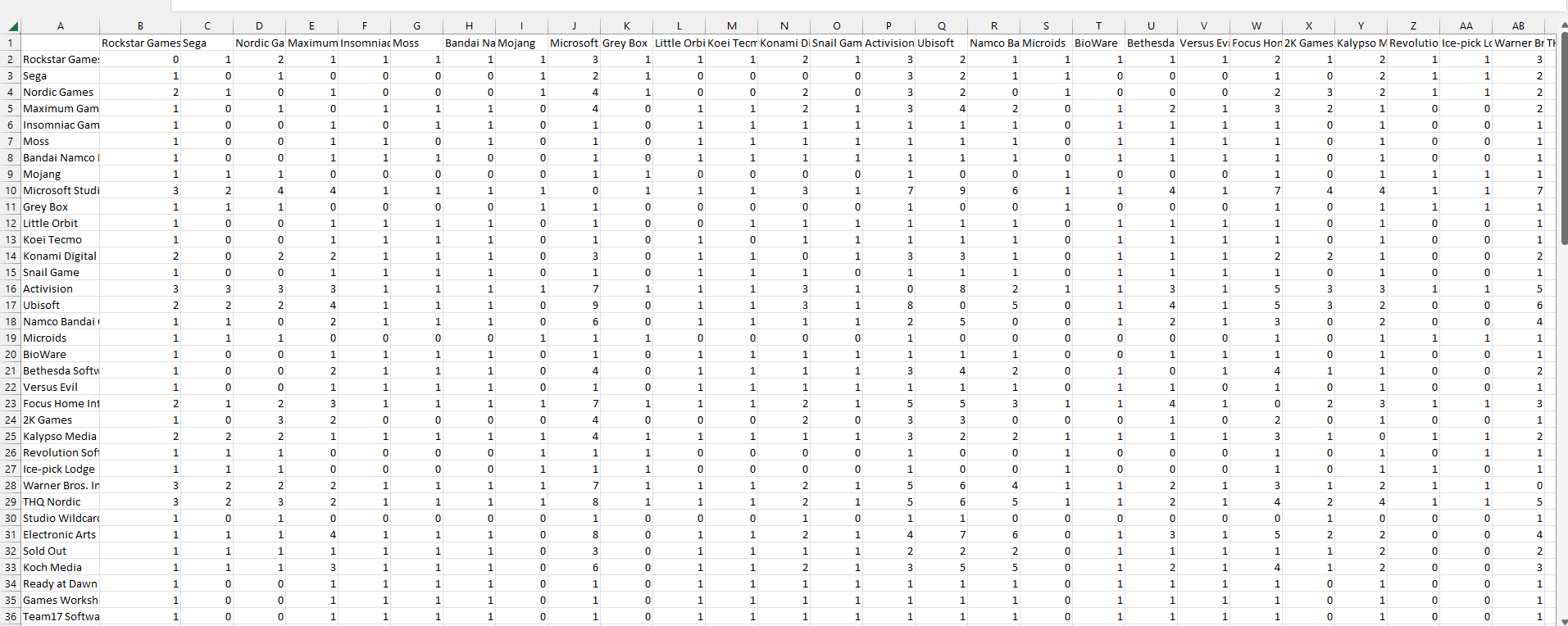
**Bước 1:** Xuất graph file từ Gephi (đồ thị tạo từ file đồ thị 1 hướng làm ở **Phần 3 mục 3.1**).

Graphical user interface, application, Word

Description automatically generated

Hình 25: Xuất graph file đồ thị 1 hướng

**Bước 2:** Sau khi xuất file sẽ được 1 file ma trận các node của đồ thị.



Hình 26: Ma trận của các node

**Bước 3:** Chọn 10 giá trị đầu của ma trận và cho vào 1 file csv riêng.

A picture containing chart

Description automatically generated

Hình 27: Ma trận 10 nodes

**Bước 4:** Đọc file csv trên vào Gephi để tạo đồ thị 10 nodes.

Radar chart

Description automatically generated with medium confidence

Hình 28: Đồ thị 10 nodes đã được cho vào Gephi

1. **Đổi tên các node đồ thị**

Chart, radar chart

Description automatically generated

Hình 29: Đồ thị trước khi đổi trên node

Chart, radar chart

Description automatically generated

Hình 30: Đồ thị sau khi đổi tên node

|  |  |
| --- | --- |
| **Định nghĩa tên node** | |
| Rockstar Games | A |
| Sega | B |
| Nordic Games | C |
| Maximum Games | D |
| Insomniac Games | E |
| Moss | F |
| Bandai Namco Entertainment | G |
| Mojang | H |
| Microsoft Studios | I |
| Grey Box | J |

Bảng 2: Đổi tên các node

1. **Betweeness Centrality**

**Table

Description automatically generated with low confidence**

Hình 31: Tính betweeness centrality

1. **Closeness Centrality**

**Chart

Description automatically generated**

Hình 32: Tính closeness centrality

1. **EigenVector Centrality**

**Chart

Description automatically generated**

**Chart

Description automatically generated with medium confidence**

Hình 33: Tính eigenvector centrality

1. **PageRank**

**Table, Excel

Description automatically generated**

Hình 34: Tính pagerank

# PHẦN 5: GRAPH MINING

1. **Khái niệm các thuật toán**
   1. **Girvan Newman**

Thuật toán Girvan Newman dựa trên quan niệm cho rằng khi các cộng đồng được gắn kết với nhau thì đường đi giữa cộng đồng này đến cộng đồng khác sẽ đi qua các cạnh nối giữa các cộng đồng với tần suất cao. Mục đích chính của thuật toán là tìm những cạnh nối đó. Thay vì việc xây dựng cộng đồng bằng cách thêm vào các cạnh mạnh mẽ nhất, chúng ta sẽ xây dựng bằng cách loại bỏ dần dần các cạnh nối từ đồ thị ban đầu. Khi đó, các cộng đồng trong mạng sẽ bị ngắt kết nối với nhau, ta có thể xác định được cách phân vùng đồ thị thành các phần nhỏ riêng rẽ. Để làm được việc này, điều quan trọng nhất của thuật toán là việc tính toán như thế nào, sử dụng tính chất nào để phát hiện ra những cạnh nối này, từ đó loại bỏ chúng ra khỏi đồ thị. Thuật toán lần đầu tiên được đề xuất bởi Freeman. Theo Freeman, các cạnh được coi là cạnh có số lượng con đường ngắn nhất giữa các cặp đỉnh khác nhau chạy qua nó. Cạnh nối có ảnh hưởng rất lớn đến dòng chảy của thông tin giữa các nút khác, đặc biệt là trong trường hợp thông tin lưu truyền trong mạng chủ yếu theo con đường ngắn nhất. Thuật toán điển hình nhất trong các thuật toán chia này là thuật toán Girvan-Newman.

Nếu một mạng lưới bao gồm các cộng đồng hoặc nhóm chúng chỉ được liên kết

nối yếu bằng một nhóm cạnh, thì tất cả các đường đi ngắn nhất giữa các cộng

đồng khác nhau sẽ phải đi dọc theo một trong số ít các cạnh thuộc nhóm cạnh

đó.Vì vậy, các cạnh kết nối các cộng đồng sẽ là cạnh có độ trung gian cao. Bằng

cách loại bỏ các cạnh, thuật toán Girvan-Newman tách được thành các nhóm

riêng biệt. Thuật toán được thực hiện theo các bước sau:

* **Bước 1:** Tính độ đo trung gian cho tất cả các cạnh trong mạng.
* **Bước 2:** Hủy bỏ các cạnh có độ trung gian cao nhất.
* **Bước 3:** Tính lại độ trung gian cho tất cả các cạnh bị ảnh hưởng theo các cạnh đã loại bỏ.
* **Bước 4:** Lặp lại từ bước 2 cho đến khi không còn các cạnh trung gian.
  1. **Louvain**

Thuật toán Louvain là một phương pháp phân chia cộng đồng và thực hiện lặp đi lặp lại việc phân chia cộng đồng nhiều lần để có được mô đun tối đa của toàn bộ mạng. Nó gán các cộng đồng khác nhau cho mỗi đỉnh và lặp lại việc hợp nhất các đỉnh dựa trên mức độ tăng của độ đo đơn thể. Thuật toán được lặp đi lặp lại cho đến khi không thể cải tiến thêm nữa. Độ phức tạp tính toán của thuật toán là O(nlog(n)). Các bước chính của thuật toán như sau:

* **Bước 1:** Đầu tiên, gán một cộng đồng khác nhau cho mỗi đỉnh của mạng, vì vậy trong phân vùng ban đầu, số lượng cộng đồng nhiều như số đỉnh.
* **Bước 2:** Đối với mỗi đỉnh, xem xét các đỉnh lân cận của nó và đánh giá giá trị tính mô đun sau khi loại bỏ một đỉnh khỏi cộng đồng của đỉnh đó và đặt đỉnh đó vào một trong những cộng đồng lân cận của nó. Nếu ∆Q dương, đỉnh vẫn ở trong cộng đồng ban đầu của nó, nếu không đỉnh được đặt trong cộng đồng cập nhật.  
  **Bước 3:** Lặp lại Bước 2 đến khi cộng đồng của tất cả các đỉnh không còn thay đổi.
* **Bước 4:** Xây dựng một đồ thị mới và mỗi đỉnh đại diện cho một cộng đồng được phân vùng bởi Bước 3. Thực hiện Bước 2 và Bước 3 liên tục cho đến khi đạt được giá trị mô đun lớn nhất.

1. **Girvan Newman**

**Text

Description automatically generated**

Hình 35: Chạy thuật toán girvan newman và vẽ đồ thị

Diagram

Description automatically generated

Hình 36: Đồ thị girvan newman

1. **Louvain**
   1. **Vẽ đồ thị**

**Text

Description automatically generated**

Hình 37: Chạy thuật toán louvain và vẽ đồ thị

**Diagram

Description automatically generated**

Hình 38: Đồ thị louvain

* 1. **Tổng quan các cụm**

**Text

Description automatically generated**

Hình 39: Tính số lượng cụm

Ta có số lượng cụm là **4.**

**Graphical user interface, text, application

Description automatically generated**

Xuất 4 cụm và mỗi cụm bao gồm tên nhà sản xuất và thể loại vào 4 file csv để tiến hành nhận xét.

Text

Description automatically generated with medium confidence

Hình 40: Xuất cụm vào file csv

* 1. **Pivot các cụm**
     1. **Cụm 0**

Hình 41: Pivot cụm 0

**Nhận xét:** Các nhà sản xuất ở cụm này chú trọng vào sản xuất thể loại game **ACTION**.

Hình 42: Biểu đồ dành cho nhà sản xuất

Đây là biểu đồ có thể giúp cho những nhà sản xuất có cái nhìn tổng quan hơn về những nhà sản xuất làm thể loại **ACTION**.

* + 1. **Cụm 1**

Hình 43: Pivot cụm 1

**Nhận xét:** Các nhà sản xuất ở cụm này chú trọng vào sản xuất thể loại game **SPORTS và ADVENTURE**.

Hình 44: Biểu đồ dành cho nhà sản xuất

Đây là biểu đồ có thể giúp cho những nhà sản xuất có cái nhìn tổng quan hơn về những nhà sản xuất làm thể loại **SPORTS và ADVENTURE**.

* + 1. **Cụm 2**

Hình 45: Pivot cụm 2

**Nhận xét:** Các nhà sản xuất ở cụm này chú trọng vào sản xuất thể loại game **RACING, ROLE-PLAYING, SIMULATION, ACTION VÀ PLATFORM**.

Hình 46: Biểu đồ dành cho nhà sản xuất

Đây là biểu đồ có thể giúp cho những nhà sản xuất có cái nhìn tổng quan hơn về những nhà sản xuất làm thể loại **RACING, ROLE-PLAYING, SIMULATION, ACTION VÀ PLATFORM**.

* + 1. **Cụm 3**

Hình 47: Pivot cụm 3

**Nhận xét:** Các nhà sản xuất ở cụm này chú trọng vào sản xuất thể loại game **SHOOTER VÀ MISC.**

Hình 48: Biểu đồ dành cho nhà sản xuất

Đây là biểu đồ có thể giúp cho những nhà sản xuất có cái nhìn tổng quan hơn về những nhà sản xuất làm thể loại **SHOOTER VÀ MISC.**

# PHẦN 6: TÀI LIỆU THAM KHẢO

[1] Portal.ptit.edu.vn, “NGHIÊN CỨU CÁC THUẬT TOÁN RÚT GỌN ĐỒ THỊ VÀ ỨNG DỤNG ĐỂ PHÁT HIỆN CỘNG ĐỒNG TRÊN MẠNG XÃ HỘI”

<https://portal.ptit.edu.vn/saudaihoc/wp-content/uploads/2021/08/LA_Nguyễn-Xuân-Dũng.pdf> (truy cập lần cuối 26/12/2022).

[2] uet.edu.vn, “HỌ THUẬT TOÁN GIRVAN-NEWMAN TRONG PHÁT HIỆN CỘNG ĐỒNG VÀ CÀI ĐẶT THỬ NGHIỆM TRÊN MẠNG XÃ HỘI TRỰC TUYẾN” <http://www.uet.vnu.edu.vn/~thuyhq/Student_Thesis/K52_Tran_Trung_Hieu_Thesis.pdf> (truy cập lần cuối 26/12/2022).

[3] uet.edu.vn, “THUẬT TOÁN PHÂN CỤM TRONG KHAI PHÁ KHÍA CẠNH TỔ CHỨC TRONG PHÁT HIỆN QUÁ TRÌNH” <https://uet.edu.vn/~thuyhq/Student_Thesis/K55_Nguyen_Van_Quyen_Thesis.pdf> (truy cập lần cuối 26/12/2022).

[4] uit.edu.vn, “ĐỒ ÁN MẪU CỦA GIÁO VIÊN CUNG CẤP” <https://courses.uit.edu.vn> (truy cập lần cuối 26/12/2022).