**ĐẠI HỌC QUỐC GIA THÀNH PHỐ HỒ CHÍ MINH**

**TRƯỜNG ĐẠI HỌC CÔNG NGHỆ THÔNG TIN**

**KHOA HỆ THỐNG THÔNG TIN**

****

****

**BÁO CÁO ĐỒ ÁN**

**Môn: Mạng xã hội**

**Đề tài: Flipkart Mobiles**

**Lớp: IS353.N11**

**GVHD: Nguyễn Thị Kim Phụng**

**Sinh viên: Nguyễn Hữu Thắng - 19522208**

**🙡 TP. Hồ Chí Minh, 12/2022 🙣**

# Lời cám ơn

Đầu tiên, em xin gửi lời cảm ơn chân thành đến tập thể quý Thầy Cô Trường Đại học Công nghệ thông tin – Đại học Quốc gia TP.HCM và quý Thầy Cô khoa Hệ thống thông tin đã giúp cho em có những kiến thức cơ bản làm nền tảng để thực hiện đề tài này.

Đặc biệt, em xin gửi lời cảm ơn và lòng biết ơn sâu sắc nhất tới Cô *Nguyễn Thị Kim Phụng*. Cô đã trực tiếp hướng dẫn tận tình, sửa chữa và đóng góp nhiều ý kiến quý báo giúp em hoàn thành tốt báo cáo môn học của mình.

Trong thời gian một học kỳ thực hiện đề tài, em đã vận dụng những kiến thức nền tảng đã tích lũy đồng thời kết hợp với việc học hỏi và nghiên cứu những kiến thức mới. Từ đó, em vận dụng tối đa những gì đã thu thập được để hoàn thành một báo cáo đồ án tốt nhất. Tuy nhiên, trong quá trình thực hiện, em không tránh khỏi những thiếu sót. Chính vì vậy, em rất mong nhận được những sự góp ý từ phía Cô nhằm hoàn thiện những kiến thức mà em đã học tập và là hành trang để em thực hiện tiếp các đề tài khác trong tương lai.

Xin chân thành cảm ơn các quý Thầy Cô!

# Nhận xét của giảng viên

………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………

**Mục lục**

Contents

[Lời cám ơn 1](#_Toc123334365)

[Nhận xét của giảng viên 2](#_Toc123334366)

[1. Tổng quan 4](#_Toc123334367)

[1.1. Giới thiệu 4](#_Toc123334368)

[1.2. Xác định bài toán 4](#_Toc123334369)

[2. Dữ liệu 5](#_Toc123334370)

[2.1. Giới thiệu nguồn dữ liệu 5](#_Toc123334371)

[2.2. Mô tả dữ liệu 5](#_Toc123334372)

[2.3. Xử lý và phân tích dữ liệu 6](#_Toc123334373)

[ Làm sạch dữ liệu 6](#_Toc123334374)

[ Chuyển đổi DataFrame thành đồ thị 7](#_Toc123334375)

[3. Thuật toán Centrality 13](#_Toc123334376)

[3.1. Page rank 13](#_Toc123334377)

[ Gephi 13](#_Toc123334378)

[ Python 13](#_Toc123334379)

[3.2. Closeness centrality 14](#_Toc123334380)

[ Gephi 14](#_Toc123334381)

[ Python 14](#_Toc123334382)

[3.3. Betweenness centrality 15](#_Toc123334383)

[ Gephi 15](#_Toc123334384)

[ Python 15](#_Toc123334385)

[4. Thuật toán phân cụm 16](#_Toc123334386)

[4.1. Louvain 16](#_Toc123334387)

[ Python 16](#_Toc123334388)

[ Gephi 20](#_Toc123334389)

[4.2. Girvan Newman 20](#_Toc123334390)

[4.3. KMean 22](#_Toc123334391)

[5. Trích 10 Nodes 26](#_Toc123334392)

[5.1. Trích xuất 26](#_Toc123334393)

[5.2. Đồ thị 1 phía 27](#_Toc123334394)

[5.3. Đồ thị 2 phía 29](#_Toc123334395)

[5.4. Tính tay 29](#_Toc123334396)

[6. Trích 10 Nodes theo yêu cầu của cô 30](#_Toc123334397)

[6.1. Trích xuất 30](#_Toc123334398)

[6.2. Tính tay 31](#_Toc123334399)

[7. Tham khảo 31](#_Toc123334400)

# Tổng quan

## Giới thiệu

Điện thoại ngày nay bất kể từ thương hiệu gì, đều được trang bị nhiều dung lượng RAM khác nhau. Nên mua điện thoại RAM bao nhiêu là đủ? Chúng ta hãy cùng phân tích độ phổ biến sản xuất của các loại dung lượng RAM theo từng hãng (nhãn hiệu).

## Xác định bài toán

* Input: Tập dữ liệu chứa thông số kỹ thuật của nhiều thương hiệu Di động khác nhau ở Ấn Độ đã được lấy từ một trang web thương mại điện tử 'Flipkart'. Bộ dữ liệu này có 2647 mẫu với 8 thuộc tính.
* Output: Đưa ra độ đo, đưa ra cộng đồng phục vụ cho việc phân tích mạng xã hội “Flipkart Mobiles”

# Dữ liệu

## Giới thiệu nguồn dữ liệu

* Link Dataset: <https://www.kaggle.com/datasets/devsubhash/flipkart-mobiles-dataset>
* Dữ liệu gồm 2647 mẫu với 8 thuộc tính.

## Mô tả dữ liệu

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **STT** | **Thuộc tính** | **Kiểu dữ liệu** | **Mô tả** | **Giá trị** |
| 1 | Brand | Char | Tên của nhà sản xuất điện thoại di động |  |
| 2 | Model | Char | Số kiểu của Điện thoại di động |  |
| 3 | Color | Char | Màu sắc của mô hình. |  |
| 4 | Memory | Char | RAM(4GB,6GB,8GB, v.v.) | 4GB,6GB,8GB, v.v. |
| 5 | Storage | Char | ROM(32GB,64GB,128GB,256GB, v.v.) | 32GB,64GB,128GB,256GB, v.v. |
| 6 | Rating | Int | Xếp hạng của mô hình dựa trên đánh giá (trong số 5). Giá trị Thiếu hoặc Không cho biết không có xếp hạng nào cho mô hình. |  |
| 7 | Selling Price | Int | Giá bán/Giá chiết khấu của mô hình |  |
| 8 | Original Price |  | Giá thực tế của kiểu máy tính bằng INR |  |

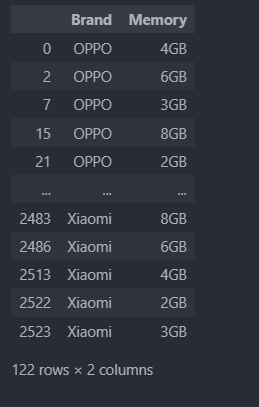
*Bảng 2.1. Bảng mô tả dữ liệu*

## Xử lý và phân tích dữ liệu

### Làm sạch dữ liệu

Bộ dữ liệu nhìn chung khá đầy đủ nhưng có thuộc tính bị thiếu dữ liệu, cần loại bỏ khỏi dataset.



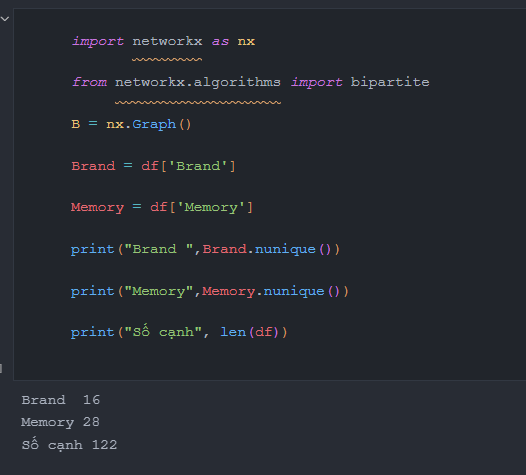
**

*Hình 2.2. Đọc dữ liệu từ file csv và loại bỏ dữ liệu trùng và rỗng*

Dữ liệu sẽ được đọc từ file csv đưa vào dataframe. Sau đó xoá bỏ dữ liệu trùng lặp và rỗng đi. Kết quả nhận được là một bộ dữ liệu 122 dòng và 2 cột

### Chuyển đổi DataFrame thành đồ thị

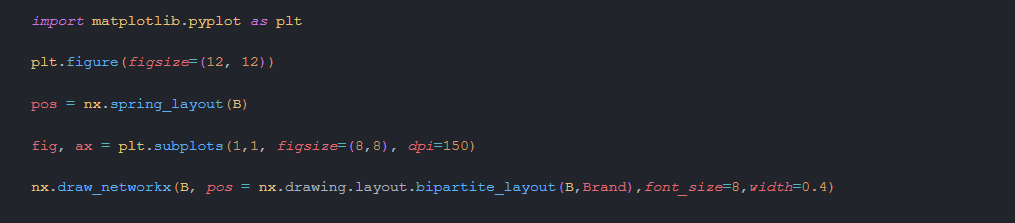
* Đồ thị 2 phía
  + Node: là loại (nhãn hiệu) điện thoại (Brand) và Loại bộ nhớ (Memory)
  + Edge: là mối quan hệ giữa việc nhãn hiệu có sản xuất loại điện thoại có bộ nhớ đó không.

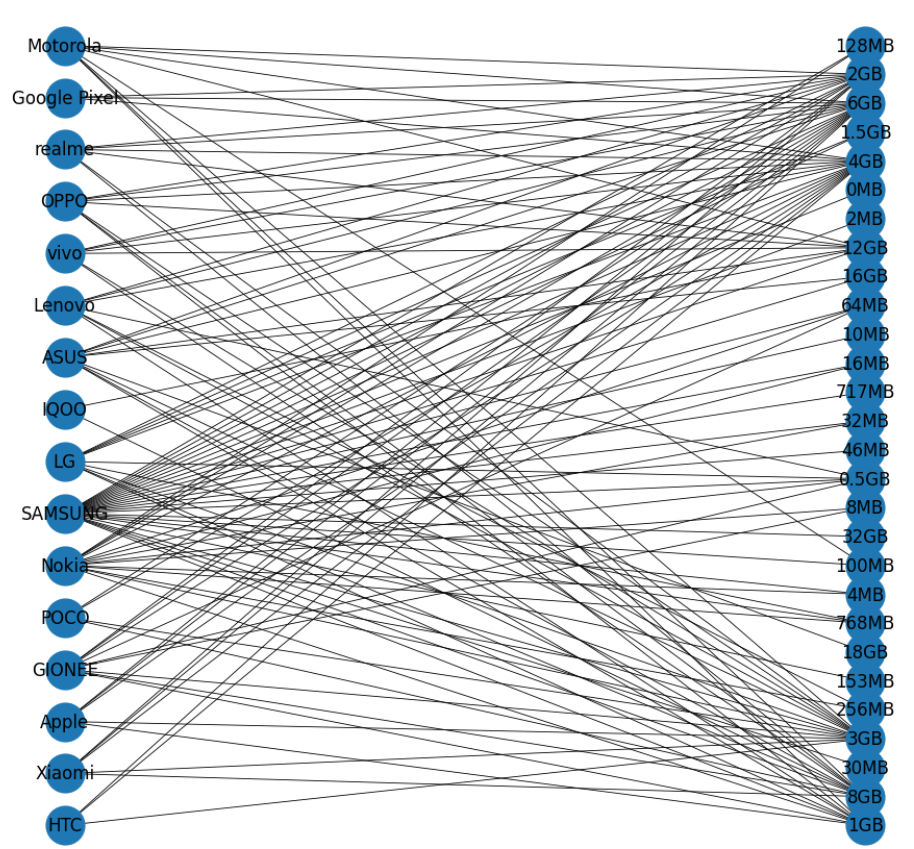
**

*Hình 2.3. Đưa dữ liệu từ DataFrame vào đồ thị vô hướng*

* Có 16 nhãn hiệu
* Số loại bộ nhớ là 28
* Số cạnh là 122

Code hiển thị đồ thị 2 phía:

*Hình 2.4. Code hiển thị đồ thị 2* phía

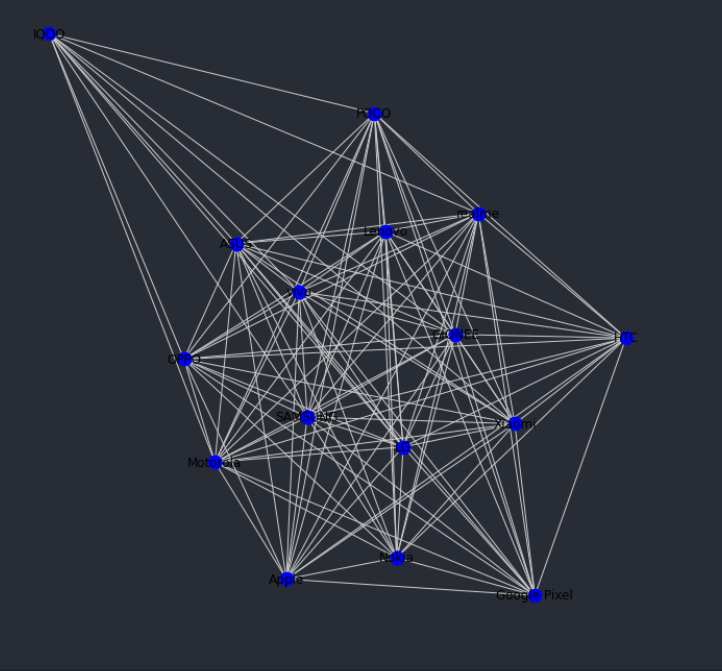


Hình *2.5. Đồ thị 2 phía*

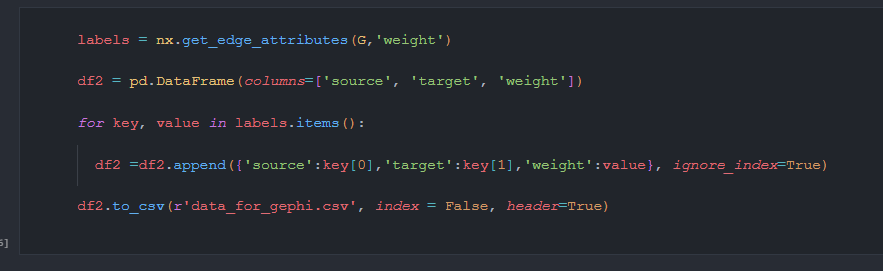
* ***Nhìn vào đồ thị ta có thể thấy một nhãn hiệu sản xuất nhiều loại điện thoại có bộ nhớ khác nhau và mỗi loại bộ nhớ cũng được nhiều nhãn hiệu sản xuất.***
* Đồ thị 1 phía
  + Node: là các nhãn hiệu (Brand)
  + Edge: Hai nhãn hiệu cùng có sản xuất chung loại bộ nhớ
  + Weight: Trọng số là số loại bộ nhớ khi hai nhãn hiệu sản xuất điện thoại cùng sản xuất.

**

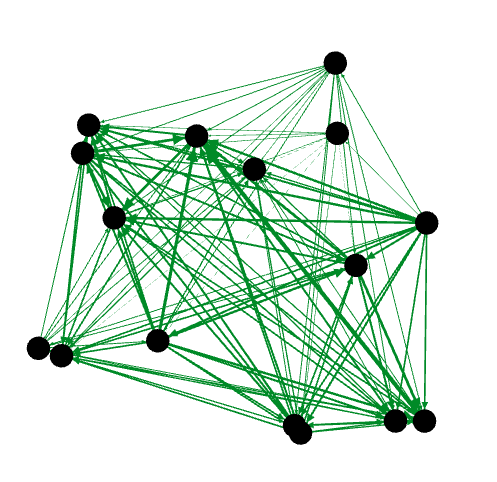
*Hình 2.6. Code hiển thị đồ thị 1 phía*

*Hình 2.7. Đồ thị 1 phía*

* + - Gephi
* Xuất dữ liệu đồ thị 1 phía ra file csv để thực hiện trên Gephi



* Đồ thị 1 phía trên Gephi

**

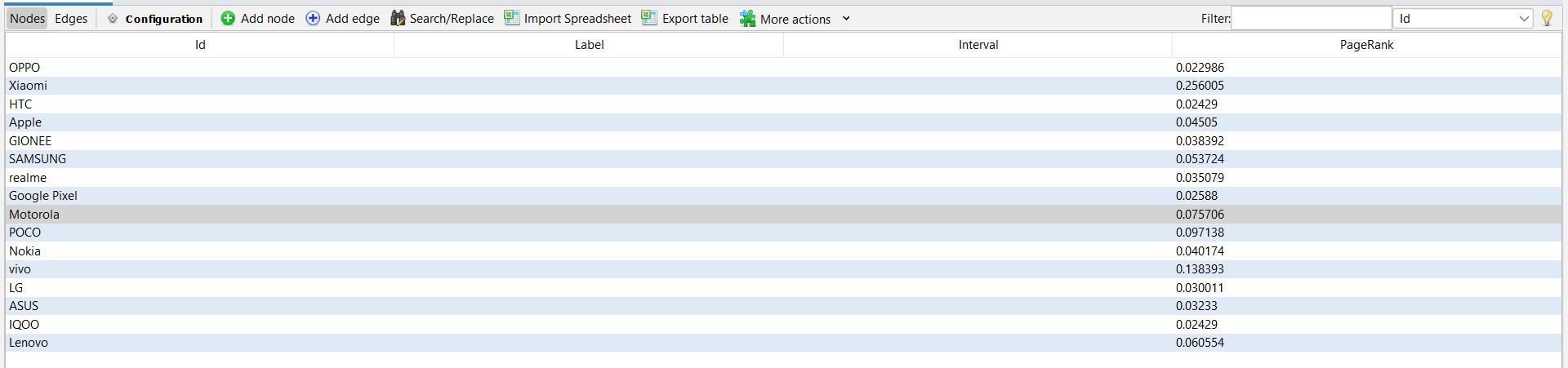
*Hình 2.8. Đồ thị 1 phía trên Gephi*

# Thuật toán Centrality

## Page rank

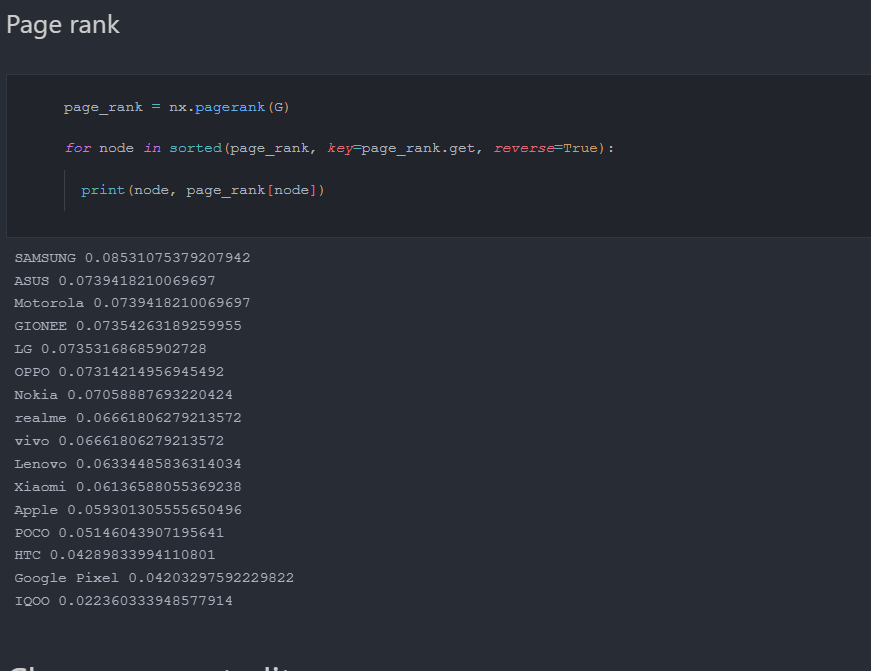
### Gephi

Kết quả tính page rank trên gephi



### Python

Kết quả tính page rank bằng python



## Closeness centrality

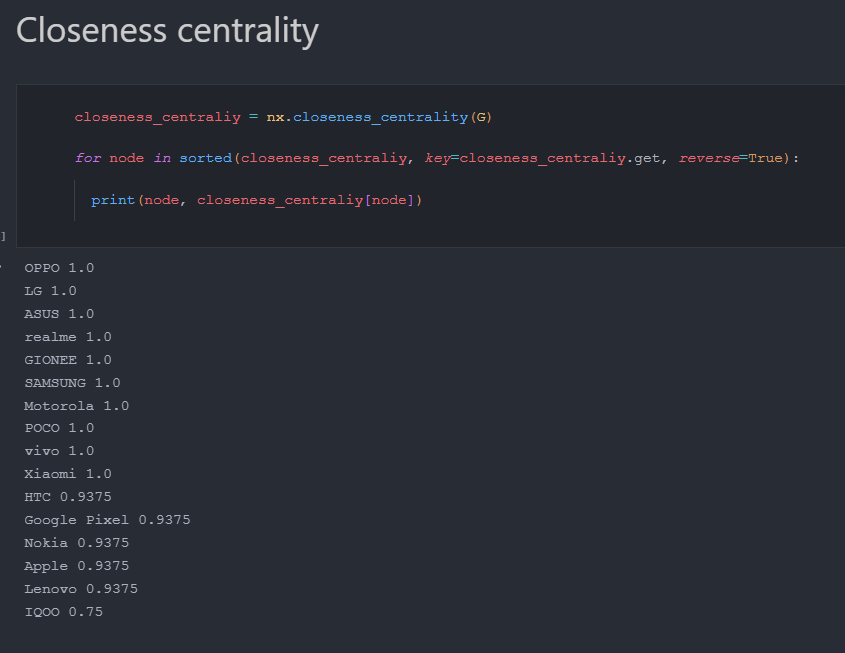
### Gephi

Graphical user interface, application

Description automatically generated

### Python

Kết quả tính Closeness centrality bằng python



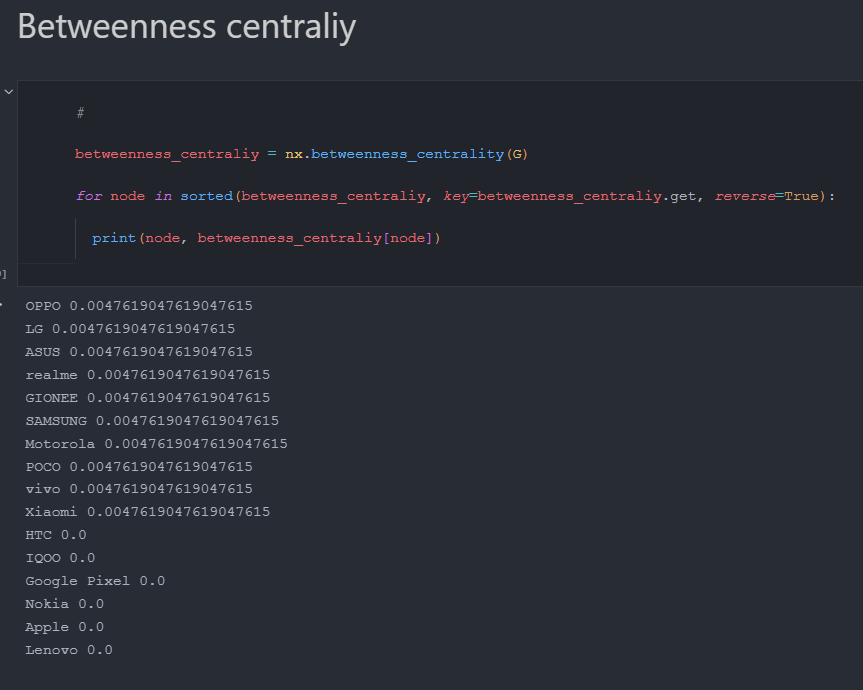
## Betweenness centrality

### Gephi

Graphical user interface

Description automatically generated with medium confidence

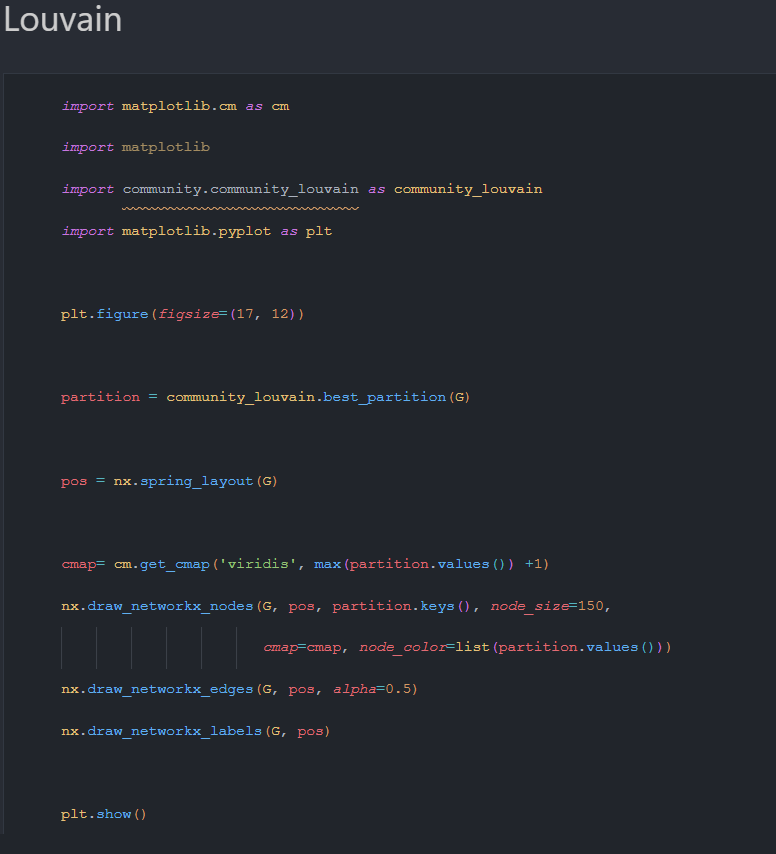
### Python



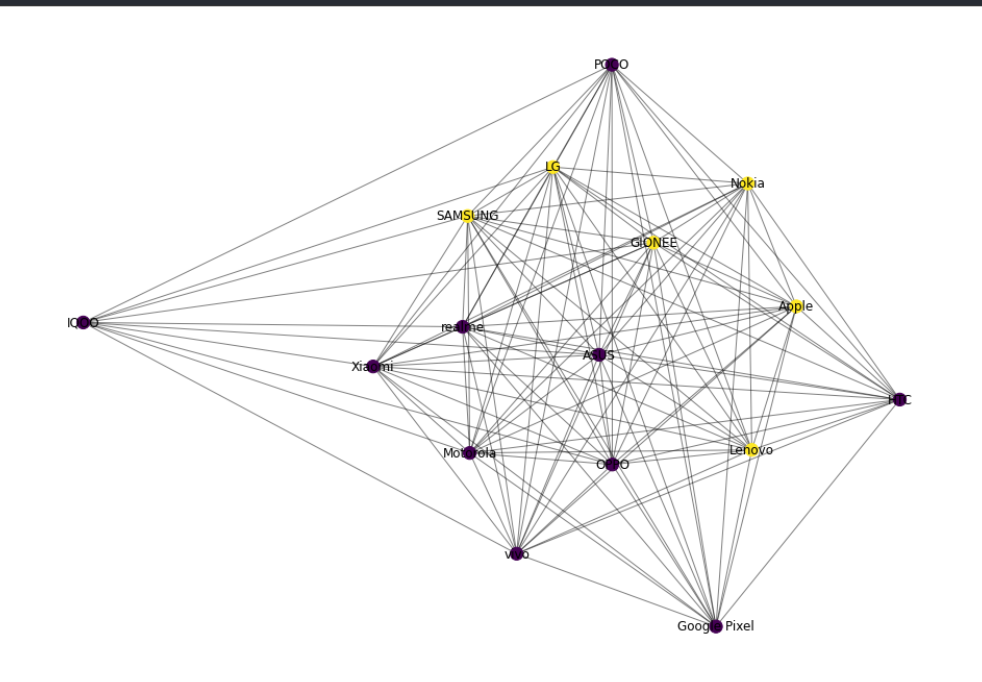
# Thuật toán phân cụm

## Louvain

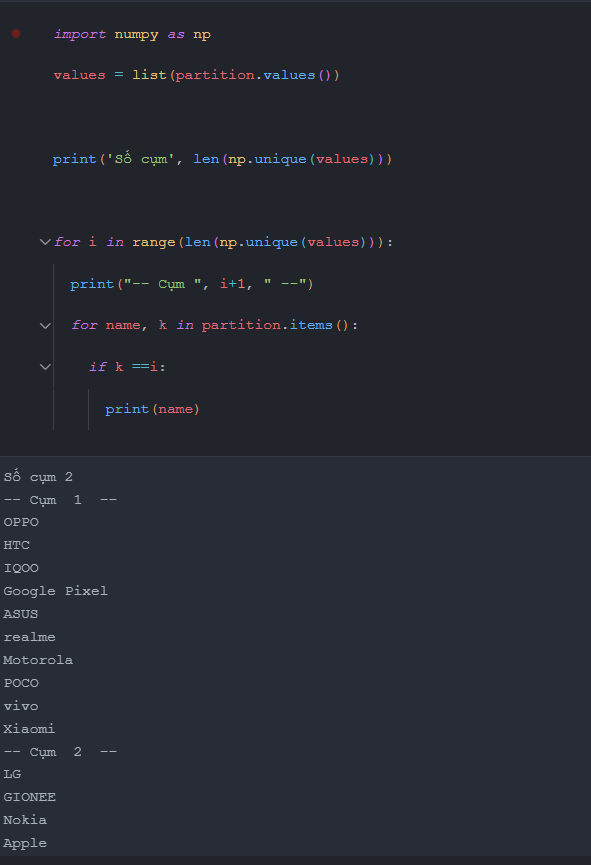
### Python



*Hình 4.1. Code thuật toán phân cụm Louvain*



*Hình 4.2. Đồ thị phân cụm Louvain*



*Hình 4.3. Code tính số cụm và nội dung cụm*

* Số cụm: 2
* -- Cụm 1 --
* OPPO
* HTC
* IQOO
* Google Pixel
* ASUS
* realme
* Motorola
* POCO
* vivo
* Xiaomi

Graphical user interface, application, Excel

Description automatically generated

*Hình 4.4. vẽ biểu đồ cho cụm 1 biểu diễn số loại bộ nhớ được sản xuất*

* -- Cụm 2 --
* LG
* GIONEE
* Nokia
* Apple
* SAMSUNG
* Lenovo

Chart

Description automatically generated with medium confidence

*Hình 4.5. vẽ biểu đồ cho cụm 2 biểu diễn loại bộ nhớ được sản xuất theo từng nhãn hiệu*

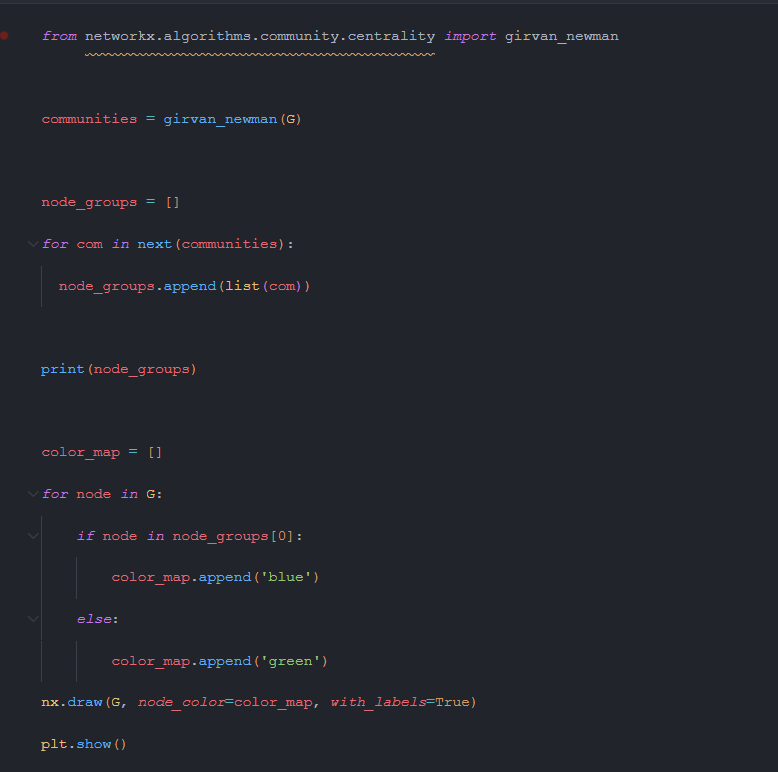
* Ý nghĩa cụm 1: các hãng sản xuất điện thoại 2GB, 3GB, 4GB, 6GB, 8GB, 12GB
* Ý nghĩa cụm 2: các hãng sản xuất điện thoại 30MB, 256MB, 0.5GB, 1GB

### Gephi

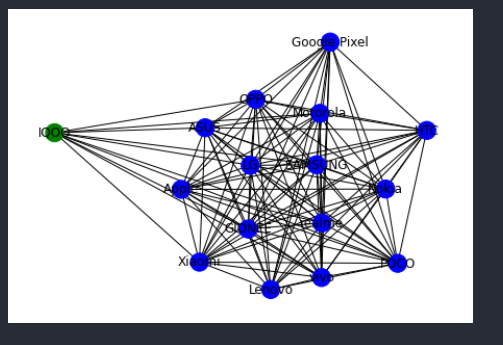
*Hình 4.4. Đồ thị phân cụm Louvain sử dụng Gephi*

## Girvan Newman

* Code thực thi



* Trực quan hóa



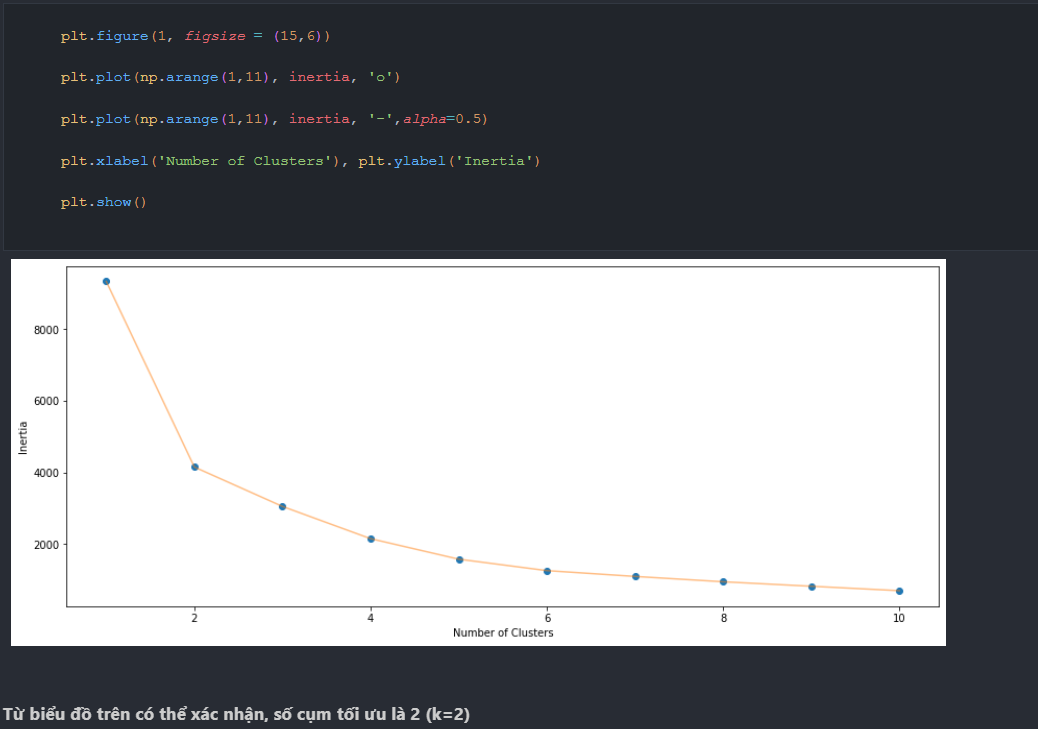
*Hình 4.5. Trực quan hóa Girvan Newman*

* Số cụm: 2
* Cụm 1 – xanh blue: ['Xiaomi', 'HTC', 'Apple', 'GIONEE', 'POCO', 'Nokia', 'SAMSUNG', 'vivo', 'LG', 'ASUS', 'Lenovo', 'OPPO', 'realme', 'Google Pixel', 'Motorola']
* Cụm 2 – xanh green: ['IQOO']
* ***Chỉ có 1 nhãn hiệu sản xuất điện thoại có bộ nhớ 30MB là*** 'IQOO'

## KMean

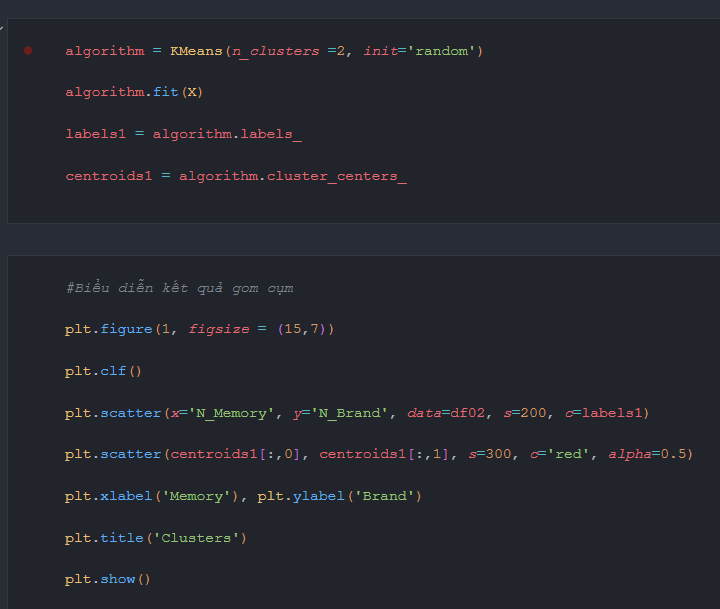
* Code thực thi

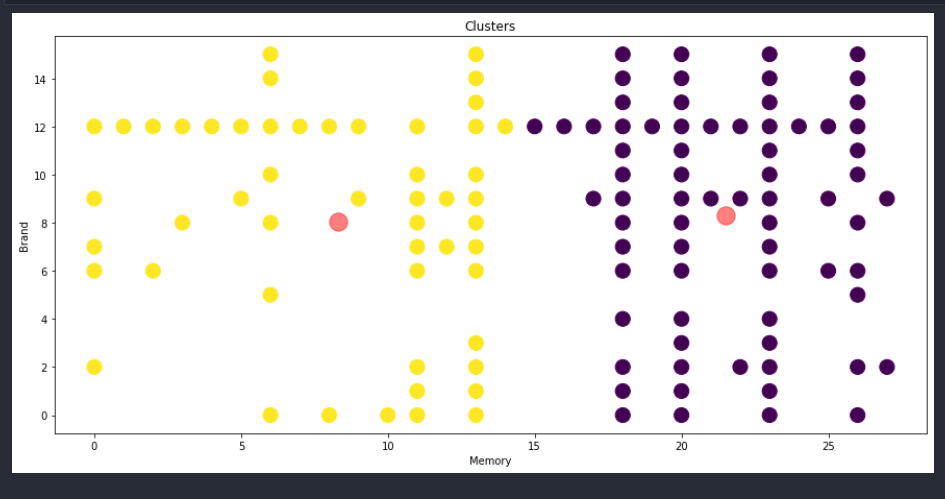




*Hình 4.6. Kết quả thực thi phương pháp ELBOW*

* Code thực thi với số cụm k = 2





*Hình 4.7. Kết quả thuật toán KMean*

# Trích 10 Nodes

* 1. Trích xuất

Từ data ban đầu

Text

Description automatically generated

*Hình 5.1*

Chọn ra 10 nhãn hiệu

Graphical user interface

Description automatically generated

*Hình 5.2*

Mapping với cột Memory để lấy đủ hết data theo 10 nhãn hiệu vừa chọn

Graphical user interface

Description automatically generated with low confidence

*Hình 5.3*

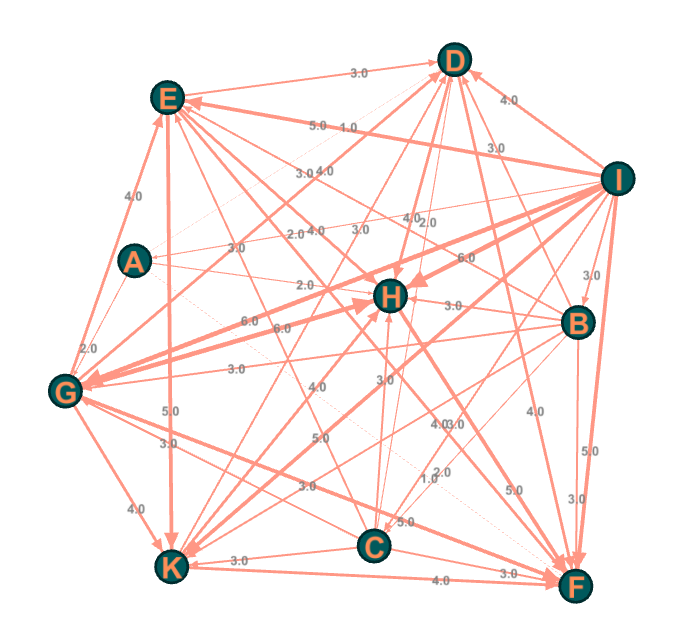
* 1. Đồ thị 1 phía

Code tương tự như ở phần data đầy đủ, ta có thể xuất được đồ thị 1 phía với 10 nodes đã chọn

Chart, radar chart

Description automatically generated

*Hình 5.3. Đồ thị 1 phía xuất bằng python*



*Hình 5.3. Đồ thị 1 phía xuất bằng gephi*

Chú thích

Table

Description automatically generated

* 1. Đồ thị 2 phía

Tương tự ta có đồ thị 2 phía

Diagram

Description automatically generated

* 1. Tính tay
     + Đồ thị và ma trận kề
     + Closeness
     + Betweeness
     + Kmean
     + Page rank
     + Eigen Vector

# Trích 10 Nodes theo yêu cầu của cô

* 1. Trích xuất

Từ danh sách tất cả các nodes ban đầu

A screenshot of a computer

Description automatically generated with medium confidence

Lọc tay ra 10 nodes bất kì và bỏ sang Workspace mới

Table

Description automatically generated with medium confidence

Đồ thị trên gephi

A picture containing chart

Description automatically generated

* 1. Tính tay
     + Đồ thị và ma trận kề
     + Closeness
     + Betweeness
     + Kmean
     + Eigen Vector

# Tham khảo

* Centrality-Measures\_Include(Eigenvectors and Eigenvalues).pdf (course)
* <https://machinelearningcoban.com/2017/01/01/kmeans/>
* <https://gtvseo.com/pagerank-la-gi/>