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

**TRƯỜNG ĐẠI HỌC CÔNG NGHỆ THÔNG TIN**

Logo, company name

Description automatically generated

**BÁO CÁO CUỐI KỲ**Môn học : Mạng Xã Hội

Học kỳ I (2022-2023)

**ĐỀ TÀI:**

**PHÂN TÍCH DỮ LIỆU NHỮNG CUỘC KHỦNG BỐ Ở THỔ NHĨ KỲ**

Sinh viên thực hiện:

Trần Nhật Tân MSSV: 19522177

GVHD: TS. Nguyễn Thị Kim Phụng

Lớp: IS353.N12.HTCL

**Thành phố Hồ Chí Minh, tháng 11 năm 2022**

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

**TRƯỜNG ĐẠI HỌC CÔNG NGHỆ THÔNG TIN**

Logo, company name

Description automatically generated

**BÁO CÁO CUỐI KỲ**Môn học : Mạng Xã Hội

Học kỳ I (2022-2023)

**ĐỀ TÀI:**

**PHÂN TÍCH DỮ LIỆU NHỮNG CUỘC KHỦNG BỐ Ở THỔ NHĨ KỲ**

Sinh viên thực hiện:

Trần Nhật Tân MSSV: 19522177

GVHD: TS. Nguyễn Thị Kim Phụng

Lớp: IS353.N12.HTCL

**Thành phố Hồ Chí Minh, tháng 11 năm 2022**

Mục lục

[**LỜI CẢM ƠN** 3](#_Toc120797475)

[**NHẬN XÉT CỦA GIẢNG VIÊN** 4](#_Toc120797476)

[**I. TỔNG QUAN** 5](#_Toc120797477)

[**1.Giới thiệu dataset** 5](#_Toc120797478)

[**2. Xác định bài toán** 5](#_Toc120797479)

[**3. Mô tả dữ liệu** 5](#_Toc120797480)

[**II. XỬ LÝ VÀ PHÂN TÍCH DỮ LIỆU** 7](#_Toc120797481)

[**1. Các thư viện được sử dụng** 7](#_Toc120797482)

[**2. Làm sạch dữ liệu** 8](#_Toc120797483)

[**3. Chuyển đổi dataframe thành đồ thị** 11](#_Toc120797484)

[**III. CÁC ĐỘ ĐO VÀ KẾT QUẢ** 16](#_Toc120797485)

[**1. Degree Centrality** 17](#_Toc120797486)

[**2. Betweenness Centrality** 18](#_Toc120797487)

[**3. Closeness Centrality** 19](#_Toc120797488)

[**4. Page Rank** 21](#_Toc120797489)

[**IV.THUẬT TOÁN PHÁT HIỆN CỘNG ĐỒNG** 23](#_Toc120797490)

[**1. Thuật toán Girvan Newman** 23](#_Toc120797491)

[**2. Thuật toán Louvain** 28](#_Toc120797492)

# **LỜI CẢM ƠN**

Trên thực tế không có sự thành công nào mà không gắn liền với những sự hỗ trợ, giúp đỡ dù ít hay nhiều, dù trực tiếp hay gián tiếp của người khác. Với lòng biết ơn sâu sắc nhất, 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 nhóm có những kiến thức cơ bản làm nền tảng để thực hiện đề tài này.

Đặc biệt nhóm chúng em xin gửi lời cảm ơn chân thành tới cô Nguyễn Thị Kim Phụng – giảng viên môn Mạng Xã Hội đã tận tình giúp đỡ, trực tiếp chỉ bảo, hướng dẫn em trong suốt quá trình làm đề tài. Nhờ đó, em đã tiếp thu được nhiều kiến thức bổ ích trong việc vận dụng cũng như kỹ năng làm đề tài. Nếu không có những lời hướng dẫn, dạy bảo của cô thì em nghĩ đề tài này của em rất khó có thể hoàn thiện được. Một lần nữa, em xin chân thành cảm ơn cô.

Cuối cùng, bản thân em đã làm việc hết công suất để hoàn thành tốt đề tài của mình. Xin chân thành cảm ơn!

# **NHẬN XÉT CỦA GIẢNG VIÊN**

# **I. TỔNG QUAN**

## **1.Giới thiệu dataset**

Khủng bố là hoạt động phá hoại, đe dọa bằng lời nói, hình ảnh hoặc video giết người do cá nhân hoặc tổ chức thực hiện làm thiệt mạng người, đặc biệt là thường dân, hoặc gây tổn thất cho xã hội và cộng đồng để tác động vào tâm lý đối phương gây hoang mang khiếp sợ, nhằm mục đích chính trị hoặc tôn giáo. Một đặc tính thống nhất của khủng bố là việc sử dụng bừa bãi bạo lực đối với những người không có khả năng chống cự với mục đích là sự nổi tiếng cho một nhóm, một phong trào, một cá nhân hoặc gây áp lực lên đối thủ chính trị buộc họ phải chấp nhận một giải pháp chính trị có lợi cho mình. Các tổ chức khủng bố có thể khai thác nỗi sợ hãi của con người để hỗ trợ đạt được những mục tiêu này. Đối tượng bị khủng bố gây thiệt hại có thể là tính mạng, sức khỏe, danh dự nhân phẩm, tài sản (của cá nhân, tổ chức hay của nhà nước) hoặc sự vững mạnh của một chính quyền nhà nước.

Khủng bố là mối quan tâm lớn đối với các quốc gia trên toàn thế giới, vì nó thường báo hiệu mở đầu cho chiến tranh và dân thường bị mắc kẹt trong cuộc chiến. Chúng ta không thể thay đổi thực tế mà chúng ta không hiểu đầy đủ, vì vậy đây là một số dữ liệu liên quan về chủ đề này. Trong đồ án này, em sẽ tập trung phân tích về độ phổ biến của những mục tiêu tấn công ở các thành phố. Môn học mạng xã hội có thể phân tích mạng phân bố các cuộc tấn công ở các thành phố trong quốc gia Thổ Nhĩ Kỳ và khai thác dữ liệu trên mạng này.

## **2. Xác định bài toán**

- Input: Tập dữ liệu ban đầu trên nguồn dữ liệu Kaggle đã qua tiền xử lý dữ liệu

- Ouput: Đưa ra độ đo, đưa ra cộng đồng phục vụ cho việc phân tích mạng xã hội “Turkey Terrorism”

## **3. Mô tả dữ liệu**

**a. Nguồn dữ liệu**

- Link dữ liệu: [terrorism-in-turkey | Kaggle](https://www.kaggle.com/datasets/senakaradeniz/terrorisminturkey)

**b. Mô tả dữ liệu**

**-** Mỗi dòng trong tập dữ liệu là một vụ khủng bố tại Thổ Nhĩ Kỳ.

- Dữ liệu gồm 547 dòng và 14 thuộc tính.

|  |  |  |
| --- | --- | --- |
| Tên cột | Kiểu dữ liệu | Ý nghĩa |
| Id | Int | Id của dữ liệu |
| Year | Int | Năm xảy ra vụ khủng bố |
| Country | String | Quốc gia (Thổ Nhĩ Kỳ) |
| Region | String | Vùng của quốc gia (Trung Đông) |
| city | String | Thành phố xảy ra vụ khủng bố |
| AttackType | String | Kiểu tấn công |
| Killed | Int | Số người tử vong |
| Wounded | Int | Số người bị thương |
| Target | String | Mục tiêu của cuộc tấn công khủng bố |
| Summary | String | Diễn biến chi tiết của cuộc khủng bố |
| Group | String | Tổ chức khủng bố |
| Target\_type | String | Loại mục tiêu của cuộc tấn công khủng bố |
| Weapon\_type | String | Loại vũ khí dùng để khủng bố |
| Motive | String | Động cơ khủng bố |

- Bảng dữ liệu:

Graphical user interface, table

Description automatically generated

**II. XỬ LÝ VÀ PHÂN TÍCH DỮ LIỆU**

## **1. Các thư viện được sử dụng**

Text

Description automatically generated

## **2. Làm sạch dữ liệu**

Dữ liệu sẽ được đọc từ file csv đưa vào dataframe. Sau đó xoá bỏ dữ liệu trùng lặp và dữ liệu trống. Kết quả ta nhận được bộ dữ liệu 511 dòng và 2 cột.

A picture containing graphical user interface

Description automatically generated

Table

Description automatically generated

Graphical user interface, text, application, email

Description automatically generated

Table

Description automatically generated

Graphical user interface, text, application

Description automatically generated

- Khi kiểm tra số lượng dữ liệu ta phân tích 162 thành phố và 18 mục tiêu tấn công với 511 cạnh.

## **3. Chuyển đổi dataframe thành đồ thị**

**a. Đồ thị 2 phía**

- Node: là tên của loại mục tiêu tấn công và tên các thành phố của nó.

- Edge: cho ta biết loại mục tiêu tấn công đó đã được thực hiện ở thành phố nào.

*Code hiển thị đồ thị 2 phía:*

A screenshot of a computer

Description automatically generated with medium confidence

*Đồ thị 2 phía:*

Diagram

Description automatically generated

- Khi nhìn vào đồ thị 2 phía thì ta thấy 1 thành phố có thể có nhiều loại mục tiêu tấn công, và 1 loại mục tiêu tấn công có thể xảy ra ở nhiều thành phố.

**b. Đồ thị 1 phía**

- Node: là các thành phố (city).

- Edge: Hai thành phố có cùng mục tiêu tấn công sẽ hợp thành 1 cạnh.

Điều này nói lên rằng các vụ khủng bố tại các thành phố khác nhau tại Thổ Nhĩ Kỳ có nhiều mục tiêu tấn công khác nhau, các thành phố khác nhau có thể có cùng mục tiêu tấn công của bọn khủng bố.

*Code hiển thị đồ thị 1 phía:*

*Graphical user interface, text, application

Description automatically generated*

*Đồ thị 1 phía:*

Chart, scatter chart

Description automatically generated

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

Graphical user interface, application

Description automatically generated

WEIGHT: số mục tiêu tấn công mà 2 thành phố giống nhau (A và B có weight là 3 thì là A và B có chung 3 loại mục tiêu tấn công).

Table

Description automatically generated

# **III. CÁC ĐỘ ĐO VÀ KẾT QUẢ**

- Sau khi import file csv dữ liệu ở trên vào Data Laboratory trên Gephi ta thu được các độ đo cần thiết để so sánh kết quả giữa Gephi và Python

**A screenshot of a computer

Description automatically generated with low confidence**

## **1. Degree Centrality**

- Hệ số này sẽ giúp chúng ta đo lường được số lượng của các mối quan hệ trực tiếp của một actor nào đó với các thành viên khác trong mạng lưới. Giá trị của hệ số này chạy từ 0.00 đến 1.00 và khi giá trị càng gần tới 1.00 thì tính trung tâm trực tiếp của actor càng lớn, tức là càng nằm ở vị trí trung tâm của mạng lưới. Công thức tính như sau:

A picture containing clock

Description automatically generated

Trong đó,

k = Tổng số các mối quan hệ trực tiếp của actor

n = Tổng số actor trong mạng lưới

**\*So sánh kết quả Python và Gephi theo giá trị Degree giảm dần:**

*Python*

Text

Description automatically generated

*Gephi*

Table

Description automatically generated

## **2. Betweenness Centrality**

- Theo quan điểm của Freeman, một actor nào đó trong mạng lưới có thể ít gắn kết với các thành viên khác trong mạng lưới (tức hệ số trung tâm trực tiếp thấp), cũng không "gần gũi" lắm với mọi thành viên trong mạng lưới (tức hệ số trung tâm lân cận thấp), nhưng lại là "cầu nối" (bridge), là "nhà trung gian" cần thiết trong mọi cuộc trao đổi trong mạng lưới.

- Hệ số này cũng đi từ 0.00 đến 1.00. Khi một actor nào đó có hệ số trung tâm trung gian càng gần đến 1.00 thì số lượng quan hệ giữa các actor khác phải "thông qua" actor này càng nhiều và do đó ảnh hưởng của actor cũng càng lớn.

Cách tính hệ số trung tâm trung gian như sau:

Text

Description automatically generated with low confidence

Trong đó

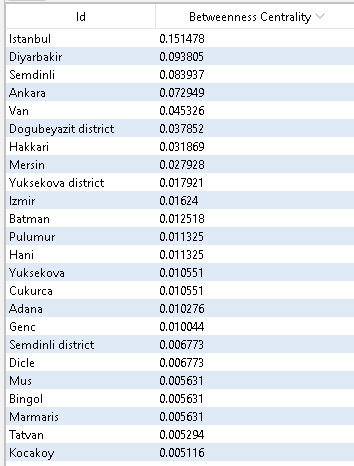
n(j,z;x) = Tổng số lần làm "trung gian" của actori

n = Tổng số actor trong mạng

*Code Python:*

Graphical user interface, text, application, email

Description automatically generated



=> Độ đo betweenness là vai trò trung gian của các mối quan hệ giữa các thương hiệu trong mạng, độ đo càng cao thì sự giao thoa của các thương hiệu càng lớn

## **3. Closeness Centrality**

- Điểm yếu của hệ số trung tâm trực tiếp là nó chỉ tính các mối quan hệ trực tiếp của actor mà thôi nên chưa chắc actor có hệ số trung tâm trực tiếp cao là người "gần gũi" với mọi thành viên khác trong mạng. Tính gần gũi hay lân cận cũng là một trong những tiêu chí quan trọng thể hiện vị thế của actor trong mạng, bởi một actor càng gần gũi với các thành viên trong mạng lưới bao nhiêu thì actor đó càng dễ có nhiều thông tin, càng có nhiều uy thế và do đó càng dễ gây ảnh hưởng lên toàn bộ mạng lưới. Để đo lường hệ số này, chúng ta sẽ tính tổng số "bước" (step) của "đoạn đường" ngắn nhất (geodesic path) mà actor phải "đi" để đến được với tất cả các thành viên khác trong mạng lưới.

- Hệ số này cũng có giá trị đi từ 0.00 đến 1.00, càng gần đến 1.00 thì actor càng gần với mọi thành viên khác trong mạng lưới, tức đoạn đường phải đi để đến với mọi actor khác càng ngắn và ngược lại. Công thức tính hệ số này như sau:

Diagram

Description automatically generated with low confidence

Trong đó:

n = Tổng số actor trong mạng lưới

∑d (x,y) = Tổng số "bước" (step) của đoạn đường ngắn nhất mà actori phải đi để đến với mọi actor trong mạng

*Code Python:*

Graphical user interface, text

Description automatically generated

**\*So sánh kết quả Python và Gephi theo giá trị Degree giảm dần:**

*Python*

*Gephi*

**Table

Description automatically generated**

=> Độ do closeness là độ gần khoảng cách của các node với các node khác trong mạng, độ đo càng cao thì thương hiệu đó xuất hiện ở cái quốc gia có chung cạnh nhiều nhất với các thương hiệu khác

## **4. Page Rank**

- PageRank là thứ hạng trang. PageRank được phát triển tại đại học Stanford bởi Lary Page (cũng bởi vậy mà có tên PageRank) và sau đó bởi Sergey Brin như một phần dự án công cụ tìm kiếm mới.Theo Google một cách tóm lược thì PageRank chỉ được đánh giá từ hệ thống liên kết đường dẫn. Trang của bạn càng nhận nhiều liên kết ( phải là dofollow ) trỏ đến thì mức độ quan trọng trang của bạn càng tăng.

Công thức PageRank có dạng như sau:

A picture containing text, watch, clock, gauge

Description automatically generated

Trong đó:

d: hằng số Google quy định. Thông thường, d = 0.85.

PR(pj): PageRank của các đỉnh đi vào đỉnh i.

L(j): số link out của các đỉnh đi vào đỉnh i.

*Code Python:*

Text

Description automatically generated

**\*So sánh kết quả Python và Gephi theo giá trị Degree giảm dần:**

*Python*

Table

Description automatically generatedGraphical user interface, text, application, chat or text message

Description automatically generated

*Gephi*

Table

Description automatically generated

# **IV.THUẬT TOÁN PHÁT HIỆN CỘNG ĐỒNG**

## **1. Thuật toán Girvan Newman**

- 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. Để tìm các cạnh trong mạng nối hai đỉnh thuộc hai cộng đồng khác nhau, khái quát đây là cạnh có độ trung gian cao, và xác định độ do trung gian này bằng cách tính số đường đi ngắn nhất giữa các cặp đỉnh mà có qua nó. Với một đồ thị m cạnh và n 13 đỉnh thì thời gian tính toán cho giai đoạn này là O(mn) .Với đồ thị có trọng số, độ đo trung gian của cạnh có trọng số đơn giản được tính bằng độ đo trung gian của cạnh không có trọng số chia cho trọng số của cạnh đó. 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 đô 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:

1. Tính độ đo trung gian cho tất cả các cạnh trong mạng.

2. Hủy bỏ các cạnh có độ trung gian cao nhất.

3. Tính lại độ trung gian cho tất cả các cạnh bị ảnh hưởng theo các cạnh đã loại bỏ.

4. Lặp lại từ bước 2 cho đến khi không còn các cạnh trung gian.

*Code cài đặt và sử dụng thuật toán:*

Graphical user interface, text, application

Description automatically generated

A screenshot of a computer

Description automatically generated with medium confidence

Số cụm 10

**Danh sách các cụm:**

**Cụm 1:** ["Campbell's", 'Tropicana Slim', 'Sarimi', 'Weh Lih', 'Itomen', 'Four Seas', 'Xiao Ban Mian', 'iNoodle', 'Sunlight', 'Paldo', 'Sao Tao', 'Nagao Noodle', 'Unif / Tung-I', 'Happy Cook', "Annie Chun's", 'Daraz', 'Unif-100', 'Ogasawara', 'Prima Taste', 'La Moderna', 'Omachi', 'Ikeda Shoku', 'Salam Mie', 'Healtimie', 'Crystal Noodle', 'Goku Uma', 'Daikoku', 'Sugakiya Foods', 'Nissin', 'Super Bihun', "World O' Noodle", 'Global Inspiration', 'Mum Ngon', 'Right Foods', 'Sakura Noodle', 'Maitri', 'Bon Go Jang', 'Eat & Go', 'Lotus Foods', 'Emart', 'Kimura', 'Cap Atoom Bulan', 'ChoripDong', 'Jackpot Teriyaki', 'Mr. Noodles', 'Chuan Wei Wang', 'Fujiwara', 'No Name', 'Takamori Kosan', 'Noodle Time', 'Lishan', 'Sau Tao', 'Authentically Asian', 'Tung-I', 'KOKA', 'Ottogi', 'JFC', 'Western Family', 'Kiki Noodle', 'Tablemark', "Mama Pat's", 'Men-Sunaoshi', 'Gomex', 'Paldo Vina', 'Fuji Mengyo', 'Good Tto Leu Foods', 'Golden Wheat', 'Oni Hot Pot', 'Lipton', 'Mexi-Ramen', 'JML', 'Long Jun Hang', 'Chewy', 'Tasty Bite', 'Unif', 'Tradition', 'Fu Chang Chinese Noodle Company', "Han's South Korea", "Farmer's Heart", 'Nakaya Shouten', 'S&S', 'Great Value', 'Love Cook', 'HoMyeonDang', 'New Touch', 'Sanrio', 'Sunny Maid', 'Thai Kitchen', 'Kailo', 'Seven & I', "Dr. McDougall's", 'Hankow', 'Guava Story', 'Baixiang Noodles', 'Kuriki', 'Ajinatori', 'Tokachimen Koubou', 'Higashimaru', 'Foodmon', 'SuperMi', 'Pop Bihun', 'Ohsung', 'Vina Acecook', 'Nan Hsing', 'Gefen', 'Hua Feng', 'Mitoku', 'Westbrae', 'Acecook', 'Wang', 'Springlife', 'Sokensha', 'Oyatsu', 'Jingqi', 'Pulmuone', 'Yamamori', 'Star Anise Foods', 'US Canning', 'Vedan', 'MyKuali', 'Haioreum', 'ABC', 'Tseng Noodles', 'GGE', 'Just Way', 'Myojo', 'GaGa', 'Migawon', 'Wei Wei', 'Higashi', 'Indomie', 'Miliket', 'Doll', 'Sura', 'President', 'Smack', 'Chikara', 'Sutah', 'Tiger', 'Thai Smile', 'La Fonte', 'IbuRamen', 'Ishimaru', 'Hosoonyi', 'Rhee Bros Assi', 'United', 'Nan Jie Cun', 'Nagatanien', 'Sakurai Foods', 'Kumamoto', 'Takamori', 'GS25', 'Long Kow', 'O Sung', 'Chering Chang', 'Vifon', "Mom's Dry Noodle", 'Fuku', 'Thai Pavilion', 'Roland', 'CJ CheilJedang', 'Pamana', 'Rocket Brand', 'Wu-Mu', 'Sahmyook', 'Tokushima Seifun', 'Hi-Myon', 'Happy Family', 'Kang Shi Fu', 'Liang Cheng Mai', 'Wei Chuan', 'Pringles', 'Wugudaochang', 'Seven-Eleven', 'Xiuhe', 'Binh Tay', 'Amianda', 'Prima', 'One Dish Asia', 'Hua Feng Noodle Expert', 'Azami', 'Deshome', 'Song Hak', 'Tayho', 'Komforte Chockolates', 'Samyang Foods', 'Master Kong', 'Six Fortune', 'Chorip Dong', 'Choumama', 'Daifuku', 'Thien Houng Foods', 'Lucky Me!', 'Osaka Ramen', 'Ten-In', 'Teriyaki Time', 'Koyo', 'Sichuan Baijia', 'Unzen', 'Koka', 'Little Cook', 'Yamachan', 'Sanyo Foods', 'Forest Noodles', 'Lishan Food Manufacturing', 'Liu Quan', 'Yamamoto', 'Chencun', 'Tokyo Noodle', 'Goku-Uma', 'Amino', 'Menraku', 'Peyang', 'Plats Du Chef', 'Sun Noodle', 'Hsin Tung Yang', 'Noah Foods', 'Qin Zong', 'Saigon Ve Wong', 'Maruchan', 'Wei Lih', 'TTL', 'Baijia', "Ripe'n'Dry", 'Souper', 'Mr. Udon', 'A-One', 'Ve Wong', 'Uni-President', 'Mie Sedaap', 'Marutai', 'Sanpo', 'Kamfen', 'Miracle Noodle', 'Samurai Ramen', 'Q', 'Sempio', 'Sichuan Guangyou', 'Snapdragon', 'Shirakiku', 'Payless', 'Yamadai', 'Dongwon', 'Nongshim', 'Sakurai', 'China Best', 'Quickchow', 'New Way', 'Kim Ve Wong', 'Unif Tung-I', 'Dream Kitchen', 'A-Sha Dry Noodle', 'Torishi', 'Samyang', 'Wu Mu', 'Sapporo Ichiban', 'Dragonfly', 'Gau Do', 'Fortune', 'Itsuki', 'Mi Sedaap']

**Cụm 2:** ['Hao Way', 'MyOri', 'Ruski', 'Uncle Sun', 'Pran', 'Singa-Me', 'Ah Lai', 'Thai Choice', 'Sunlee', 'Aroi', 'Premiere', 'FMF', 'Thai Chef', 'Tao Kae Noi', 'The Ramen Rater Select', 'Pama', 'ORee Garden', 'Cintan', 'Zow Zow', 'Saji', 'Golden Wonder', "Sainsbury's", 'Batchelors', 'Asia Gold', 'Adabi', "Vit's", 'GreeNoodle', 'President Rice', 'Yum Yum', 'Daddy', 'Knorr', 'Mama', '7 Select', 'Tiger Tiger', 'Tesco', 'J.J.', 'Ibumie', 'MAMA', 'Mi E-Zee', 'Kin-Dee', 'Urban Noodle', 'Ko-Lee', 'Wai Wai', 'A1', "Mr. Lee's Noodles", 'TRDP', 'E-Zee', 'CarJEN', 'Maggi', 'Mug Shot', 'Sawadee', 'Kabuto Noodles', 'Morre', 'Patanjali', 'The Bridge', 'Bonasia', "Chaudhary's Wai Wai", '1 To 3 Noodles', 'Super', "Kim's Bowl", 'Mee Jang', 'Papa', 'Royal Umbrella', "Ching's Secret", 'E-mi', 'Fashion Foods', 'Fantastic', '7 Select/Nissin', 'Boss', 'Trident', 'Suimin', 'Curry Prince', 'Pot Noodle', 'Fashion Food', 'Bamee', 'Nyor Nyar', 'Nanyang Chef', 'Econsave', 'Mamee', 'Shan']

**Cụm 3** ['Yum-Mie', 'Lele']

**Cụm 4** ['Asian Thai Foods']

**Cụm 5** ['Conimex', 'iMee', 'Unox', 'Koh Thai']

**Cụm 6** ['Pirkka']

**Cụm 7** ['The Kitchen Food', 'Lee Fah Mee']

**Cụm 8** ['ICA', 'Katoz'],

**Cụm 9** ['Baltix'],

**Cụm 10** ['Golden Mie']

**Ví dụ về cụm 2 và cụm 5:**

**Cụm 2:** ['Hao Way', 'MyOri', 'Ruski', 'Uncle Sun', 'Pran', 'Singa-Me', 'Ah Lai', 'Thai Choice', 'Sunlee', 'Aroi', 'Premiere', 'FMF', 'Thai Chef', 'Tao Kae Noi', 'The Ramen Rater Select', 'Pama', 'ORee Garden', 'Cintan', 'Zow Zow', 'Saji', 'Golden Wonder', "Sainsbury's", 'Batchelors', 'Asia Gold', 'Adabi', "Vit's", 'GreeNoodle', 'President Rice', 'Yum Yum', 'Daddy', 'Knorr', 'Mama', '7 Select', 'Tiger Tiger', 'Tesco', 'J.J.', 'Ibumie', 'MAMA', 'Mi E-Zee', 'Kin-Dee', 'Urban Noodle', 'Ko-Lee', 'Wai Wai', 'A1', "Mr. Lee's Noodles", 'TRDP', 'E-Zee', 'CarJEN', 'Maggi', 'Mug Shot', 'Sawadee', 'Kabuto Noodles', 'Morre', 'Patanjali', 'The Bridge', 'Bonasia', "Chaudhary's Wai Wai", '1 To 3 Noodles', 'Super', "Kim's Bowl", 'Mee Jang', 'Papa', 'Royal Umbrella', "Ching's Secret", 'E-mi', 'Fashion Foods', 'Fantastic', '7 Select/Nissin', 'Boss', 'Trident', 'Suimin', 'Curry Prince', 'Pot Noodle', 'Fashion Food', 'Bamee', 'Nyor Nyar', 'Nanyang Chef', 'Econsave', 'Mamee', 'Shan']

***>Ý nghĩa :*** *Các thương hiệu đa phần đều có mặt ở Malaysia*

**Cụm 5** ['Conimex', 'iMee', 'Unox', 'Koh Thai']:

***>Ý nghĩa :*** *Các thương hiệu đa phần đều có mặt ở Netherlands*

## **2. Thuật toán Louvain**

Phương pháp Louvain để phát hiện cộng đồng là một thuật toán để phát hiện các cộng đồng trong mạng. Nó tối đa hóa điểm mô-đun cho mỗi cộng đồng, trong đó mô-đun định lượng chất lượng của việc gán các nút cho cộng đồng. Điều này có nghĩa là đánh giá mức độ kết nối của các nút trong cộng đồng với mật độ cao hơn như thế nào so với mức độ kết nối của chúng trong một mạng ngẫu nhiên. Thuật toán Louvain là một thuật toán phân cụm phân cấp, hợp nhất một cách đệ quy các cộng đồng thành một nút duy nhất và thực hiện phân cụm mô-đun trên các đồ thị cô đọng

*Code cài đặt và sử dụng thuật toán Louvain*Text

Description automatically generated

*Đồ thị phân cụm sử dụng Louvain*

Chart, scatter chart

Description automatically generated

Graphical user interface, text

Description automatically generated

*Danh sách các cụm*

Table

Description automatically generated with medium confidence

Số cụm: 17

Ví dụ vài cụm:

**Cụm thứ 1 (47 brand )** Just Way, Wei Lih, Uni-President, Mom's Dry Noodle, Happy Cook, TTL, Kiki Noodle, GGE, Sau Tao, Ve Wong, Little Cook, Xiao Ban Mian, Unif Tung-I, Love Cook, Tseng Noodles, A-Sha Dry Noodle, Choumama, Jingqi, Oni Hot Pot, Weh Lih, Vedan , Forest Noodles, Ten-In, Chering Chang, Wu-Mu, Lishan Food Manufacturing, Deshome, Long Jun Hang, Unif, China Best, Amianda, Long Kow, Unif /Tung-I, Souper, Wu Mu, Lishan, Wei Chuan, Wei Wei, Sunlight, Nan Hsing, Hsin Tung Yang, INoodle, Q, Sunny Maid, Tiger, Kim Ve Wong, Tung-I

***>Ý nghĩa*** *: Các thương hiệu trên đa phần đều có mặt ở Taiwan*

**Cụm thứ 2 (4 brand** ) Conimex, Koh Thai, Unox, iMee

***>Ý nghĩa :*** *Các thương hiệu đa phần đều có mặt ở Netherlands*

**Cụm thứ 3 (37 brand )** Nissin, Sichuan Guangyou, Master Kong, Kang Shi Fu, Xiuhe, Sichuan Baijia, Liang Cheng Ma, Hua Feng, Liu Quan, Chuan Wei Wang, Kamfen, Qin Zong, JML, Fuku, Doll, Four Seas, Bonasia, Asia Gold, Baijia, Thai Chef, Baixiang Noodles, Wugudaochang, Chencun, Lucky Me!, Unif-100, Hankow, Chewy, Pamana, Golden Wheat, Nan Jie Cun, Springlife, Dragonfly, Kailo, Payless, Sao Tao, Quickchow, Rocket Brand, Hua Feng Noodle Expert

***>Ý nghĩa :*** *Các thương hiệu đa phần đều có mặt ở HongKong, China, Japan*

*Biểu diễn Lovain bằng Gephi*