

ĐẠI HỌC QUỐC GIA THÀNH PHỐ HỒ CHÍ MINH

TRƯỜNG ĐẠI HỌC CÔNG NGHỆ THÔNG TIN

-----🙞🙜🕮🙞🙜-----



BÁO CÁO ĐỒ ÁN MẠNG XÃ HỘI

**COMPUTER PARTS (CPUs)**

Mã lớp: IS353.M12.HTCL

**Giảng viên:** ThS. Nguyễn Thị Kim Phụng

**Sinh viên thực hiện:**

|  |  |
| --- | --- |
| **MSSV** | **Họ tên** |
| 17520528 | Dương Phúc Huân |

Thành phố Hồ Chí Minh, tháng 11 năm 2020

**LỜI CẢM ƠN**

Đầu tiên, em xin gửi lời cảm ơn chân thành đến quý Thầy cô giảng viên 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 chúng 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 ThS. Nguyễn Thị Kim Phụng, người đã hướng dẫn cho em trong suốt thời gian làm đề tài. Cô đã trực tiếp hướng dẫn tận tình, sửa chữa và đóng góp nhiều ý kiến quý báu giúp nhóm chúng em có thể hoàn thành tốt Báo cáo Đồ án môn học.

Trong thời gian một học kỳ thực hiện đề tài, nhóm chúng 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 thầy cô, bạn bè cũng như nhiều nguồn tài liệu tham khảo. Từ đó, nhóm chúng em vận dụng tối đa những gì đã thu thập được để hoàn thành một báo cáo đồ án tốt nhất.

Tuy nhiên, vì kiến thức chuyên môn còn hạn chế và bản thân còn thiếu nhiều kinh nghiệm thực tiễn nên nội dung của báo cáo không tránh khỏi những thiếu sót, nhóm em rất mong nhận được sự góp ý, chỉ bảo thêm của Cô nhằm hoàn thiện những kiến thức của mình để nhóm chúng em có thể dùng làm hành trang thực hiện tiếp các đề tài khác trong tương lai cũng như là trong việc học tập và làm việc sau này.

Một lần nữa xin gửi đến Cô và các bạn lời cảm ơn chân thành và tốt đẹp nhất!

Thành phố Hồ Chí Minh, tháng 12 năm 2021

***Sinh viên thực hiện***

**MỤC LỤC**

[PHẦN I: GIỚI THIỆU DATASET 1](#_Toc90459209)

[1.1. Mô tả dữ liệu 1](#_Toc90459210)

[1.2. Xác định bài toán 1](#_Toc90459211)

[PHẦN II: CƠ SỞ LÝ THUYẾT 1](#_Toc90459212)

[2.1. Lý thuyết về Nodes và Edges 1](#_Toc90459213)

[*2.1.1. Thế nào Nodes và Edges 1*](#_Toc90459214)

[*2.1.2. Edges Direction 2*](#_Toc90459215)

[*2.1.3. Edge Weight 2*](#_Toc90459216)

[2.2. Lý thuyết các độ đo trong mạng xã hội 3](#_Toc90459217)

[*2.2.1. Độ đo Degree Centrality* 3](#_Toc90459218)

[*2.1.2. Độ đo Betweenness Centrality* 4](#_Toc90459219)

[*2.1.3. Độ đo Closeness Centrality* 4](#_Toc90459220)

[*2.1.4. Độ đo Clustering Coefficient* 5](#_Toc90459221)

[2.3. Lý thuyết các thuật toán dùng trong mạng xã hội 6](#_Toc90459222)

[*2.3.1 Cộng đồng 6*](#_Toc90459223)

[*2.3.2. Một số thuật toán khám phá cộng đồng 7*](#_Toc90459224)

[*2.3.2.1. Thuật toán Page Rank 7*](#_Toc90459225)

[*2.3.2.2. Thuật toán Girvan Newman 7*](#_Toc90459226)

[*2.3.2.3. Thuật toán K-Mean 8*](#_Toc90459227)

[PHẦN III: XỬ LÍ VÀ PHÂN TÍCH DỮ LIỆU 8](#_Toc90459228)

[3.1. Làm sạch dữ liệu 8](#_Toc90459229)

[3.2. Chuyển đổi dataframe thành đồ thị 9](#_Toc90459230)

[PHẦN IV: XẾP HẠNG THEO ĐỘ ĐO 12](#_Toc90459231)

[4.1. Độ đo PageRank 12](#_Toc90459232)

[*4.1.1. Thực thi trên Python* 12](#_Toc90459233)

[*4.1.2. Thực thi trên Gephi* 13](#_Toc90459234)

[*4.1.3. So sánh kết quả* 13](#_Toc90459235)

[*4.1.4. Nhận xét* 13](#_Toc90459236)

[4.2. Độ đo Closeness Centrality 13](#_Toc90459237)

[*4.2.1. Thực thi trên Python* 13](#_Toc90459238)

[*4.2.2. Thực thi trên Gephi* 14](#_Toc90459239)

[*4.2.3. So sánh kết quả* 14](#_Toc90459240)

[*4.2.4. Nhận xét* 14](#_Toc90459241)

[4.3. Độ đo Clustering Coefficient 14](#_Toc90459242)

[*4.3.1. Thực thi trên Python* 14](#_Toc90459243)

[*4.3.2. Thực thi trên Gephi* 15](#_Toc90459244)

[*4.3.3. So sánh kết quả* 15](#_Toc90459245)

[*4.3.4. Nhận xét* 15](#_Toc90459246)

[PHẦN V: THUẬT TOÁN PHÁT HIỆN CỘNG ĐỒNG 16](#_Toc90459247)

[5.1. Thuật toán Louvain 16](#_Toc90459248)

[*5.1.1. Thực thi trên Gephi* 16](#_Toc90459249)

[*5.1.2. Thực thi trên Python* 16](#_Toc90459250)

[*5.1.3. Ý nghĩa các cụm* 18](#_Toc90459251)

[5.2. Thuật toán Girvan Newman 20](#_Toc90459252)

[*5.2.1. Thực thi trên Python* 20](#_Toc90459253)

[*5.2.3. Ý nghĩa các cụm* 21](#_Toc90459254)

[5.3. Thuật toán K-Mean 22](#_Toc90459255)

[*5.3.1. Thực thi trên Python* 22](#_Toc90459256)

[*5.3.2. Ý nghĩa các cụm* 25](#_Toc90459257)

[5.4. Tổng kết kết quả 28](#_Toc90459258)

[TÀI LIỆU THAM KHẢO 28](#_Toc90459259)

# 

**PHẦN I: GIỚI THIỆU DATASET**

**1.1. Mô tả dữ liệu**

**Nguồn dữ liệu:**

Link Dataset: <https://www.kaggle.com/iliassekkaf/computerparts?select=Intel_CPUs.csv>

**Mô tả dữ liệu:**

Dataset gồm 2283 dòng dữ liệu, 45 cột thuộc tính. Mỗi hàng mô tả thông tin kỹ thuật của một CPU.

Cột thuộc tính, ý nghĩa thuộc tính sử dụng:

|  |  |
| --- | --- |
| Tên thuộc tính | Ý nghĩa |
| Lithography (Tiến trình) | Tiến trình sản xuất chip bán dẫn CPU. Mỗi bóng bán dẫn có kích thước siêu nhỏ có đơn vị là nm, Khi giảm kích thước của mỗi bóng bán dẫn (Transistor) đem đến nhiều lợi ích đáng kể như tốc độ xử lý thông tin tăng lên, kích thước giảm xuống, năng lượng tiêu thụ cũng giảm đi. |
| TDP  (Mức tiêu thụ điện năng) | Mức điện năng tiêu thụ đơn vị là W, khi chỉ số này càng cao thì CPU của bạn sẽ tiêu tốn điện năng càng lớn. |

**1.2. 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 “Computer Parts (CPUs)”.

**PHẦN II: CƠ SỞ LÝ THUYẾT**

**2.1. Lý thuyết về Nodes và Edges**

***2.1.1. Thế nào 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 hiển diện của một kết nối hoặc mối quan hệ giữa hai nút.

Diagram

Description automatically generated

***2.1.2. Edges Direction***

Có hai loại cạnh: cạnh có hướng và cạnh vô hướng. Trước khi bạn xây dựng lên mạng bạn cần phải giải mã dữ liệu của bạn 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.

**Ví dụ:** Nếu một khách hàng (nút bắt đầu) thanh toán cho một cửa hàng (nút kết thúc) cho một món hàng, thì mối quan hệ đó không nhất thiết phải được đáp lại bởi vì cửa hàng không trả tiền cho khách hàng.

* 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.

**Ví dụ:** nếu hai người là bạn bè trên Facebook, mối quan hệ đó là vô hướng. Đó là bởi vì có thể nói Người A là bạn với Người B, nhưng cũng có thể nói Người B là bạn với Người A.

***2.1.3. Edge Weight***

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

**Ví dụ:** nếu Người A mua một ly cà phê từ một quán cà phê 3 lần, thì cạnh nối Người A và quán cà phê sẽ có trọng số là 3. Tuy nhiên, nếu Người B chỉ mua cà phê từ quán cà phê một lần thì cạnh nối Người B và quán cà phê sẽ có trọng số bằng 1.

**2.2. Lý thuyết các độ đo trong mạng xã hội**

**Biện pháp trung tâm (Centrality Measures):** Tính trung tâm là một tập hợp các chỉ số được sử dụng để xác định mức độ quan trọng và ảnh hưởng của một nút cụ thể đối với toàn bộ trang. Các biện pháp trung tâm được sử dụng trên các nút cụ thể trong mạng và không cung cấp thông tin ở cấp độ mạng. Ví dụ 3 độ do trung tâm dưới đây:

***2.2.1. Độ đo Degree Centrality***

Độ của một nút là số cạnh mà nút đó có. Số đo này giúp ta đo số lượng của các mối quan hệ trực tiếp của một tác nhân nào đó với các thành viên khác trong mạng xã hội.

CD(v) =

Trong đó:

* n: là số đỉnh của đồ thị
* Deg(v): tổng số các liên kết trực tiếp đến đỉnh v (bậc của đỉnh).

**Ví dụ:** nếu nút A có các cạnh nối nó với Nút B và Nút D, thì độ của nút A là 2.

Ảnh có chứa văn bản, đồng hồ, đồ họa véc-tơ

Mô tả được tạo tự động

Tuy nhiên, trong một mạng có định hướng, thực tế có ba thước đo mức độ khác nhau. Bởi vì các cạnh này có nút bắt đầu và nút kết thúc, độ trong (số cạnh mà nút là nút kết thúc), độ ngoài (số cạnh mà một nút là nút bắt đầu của) và độ (số cạnh a nút là nút bắt đầu hoặc nút kết thúc của) có thể được tính toán

***2.1.2. Độ đo*** ***Betweenness Centrality***

Betweenness Centrality: đo lường tầm quan trọng của các kết nối của một nút trong việc cho phép các nút đến các nút khác. Khoảng giữa của một nút là số đường đi ngắn nhất mà nút được đưa vào chia cho tổng số đường đi ngắn nhất. Điều này sẽ cung cấp tỷ lệ phần trăm đường dẫn ngắn nhất trong mạng mà nút đang ở trong đó.

Số đo trung tâm trung gian xác nhận một tác nhân nào đó trong mạng có thể có thể có ít gắn kết với các thành viên khác trong mạng xã hội (số đo bậc trung tâm thấp), cũng không gần gũi lắm với các thành viên khác (số đo trung tâm lân cận thấp), nhưng lại là cầu nối (bridge) hay nhà trung gian cần thiết trong mọi cuộc trao đổi trong mạng.

Độ đo Betweenness Centrality được tính bằng công thức sau:

CB (v) =

Trong đó:

* là số đường đi ngắn nhất đi từ đỉnh s đến đỉnh t của toàn mạng
* = là số đường đi ngắn nhất đi từ đỉnh s đến đỉnh t và đi qua đỉnh v
* Số đo trung gian tương đối được tính theo cách sau: Số đo trung gian cho đồ thị vô hướng:

CB (x) =

***2.1.3. Độ đo Closeness Centrality***

Mức độ gần gũi là độ đo đo lường mức độ kết nối của một nút với mọi nút khác trong mạng. Mức độ gần gũi của một nút là số bước trung bình cần thiết để tiếp cận mọi nút khác trong mạng. Bước nhảy là đường đi của một cạnh từ nút này đến nút khác. Ví dụ, như được thấy trong sơ đồ bên dưới, Nút A được kết nối với Nút B và Nút B được kết nối với Nút C. Để Nút A đến được Nút C, nó sẽ mất hai bước.

A picture containing diagram

Description automatically generated

Trong một đồ thị liên thông, độ đo gần trung tâm của một nút là thước đo mức độ trung tâm trong mạng, được tính bằng nghịch đảo của tổng độ dài của các đường đi ngắn nhất giữa nút và tất cả các nút khác trong biểu đồ. Do đó, một nút càng ở trung tâm thì nó càng gần với tất cả các nút khác. Nếu một nút có độ gần gũi càng thấp thì để giao tiếp với các nút khác nó muốn thì phải đi qua rất nhiều nút khác trong mạng.

Độ đo Closeness Centrality được tính bằng công thức sau:

Trong đó:

* j
* d(i, j) là số đường đi ngắn nhất giữa 2 nodes i và j trong mạng
* N là số lượng nodes có trong mạng

***2.1.4. Độ đo Clustering Coefficient***

Clustering Coefficient là độ đo mức độ mà các nút trong biểu đồ có xu hướng tập hợp lại với nhau. Bằng chứng cho thấy rằng trong hầu hết các mạng trong thế giới thực, và đặc biệt là các mạng xã hội, các nút có xu hướng tạo ra các nhóm liên kết chặt chẽ với đặc điểm là mật độ quan hệ tương đối cao; khả năng này có xu hướng lớn hơn xác suất trung bình của một hòa được thiết lập ngẫu nhiên giữa hai nút.

A screenshot of a video game

Description automatically generated with medium confidence

Độ đo Clustering Centrality được tính bằng công thức sau:

Trong đó:

* là số cạnh giữa các lắng giềng (neighbors)
* là bậc(degree) của node i

**2.3. Lý thuyết các thuật toán dùng trong mạng xã hội**

***2.3.1 Cộng đồng***

Cộng đồng được tạo từ các cá nhân sao cho các cá nhân trong cùng một nhóm sẽ tương tác với nhau thường xuyên hơn với các cá nhân nằm ngoài nhóm. Có ba cách tiếp cận để khám phá cộng đồng trong mạng xã hội:

* Dựa trên cấu trúc mạng xã hội để khám phá cộng đồng.
* Dựa trên thông tin trao đổi theo các tương tác trên mạng xã hội.
* Dựa trên vừa cấu trúc và nội dung trao đổi.

***2.3.2. Một số thuật toán khám phá cộng đồng***

***2.3.2.1. Thuật toán Page Rank***

Pagerank là thuật toán phân tích các liên kết được dùng trong Google Search để xếp hạng các trang web.

* Thuật toán này chỉ định giá trị nhất định cho mỗi thành phần của một tập hợp các văn bản liên kết với nhau, ví dụ như World Wide Web.
* Mục đích "đo" tầm quan trọng tương đối của các liên kết trong tập hợp đó.
* Áp dụng cho bất kỳ tập hợp văn bản nào có trích dẫn đối ứng và liên kết cụ thể.
* Giá trị (weight) mà nó gán cho bất kỳ thành phần E được gọi là PageRank của E và ký hiệu là {\displaystyle PR(E).}PR (E)

***2.3.2.2. Thuật toán 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 loạn 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

***2.3.2.3. Thuật toán K-Mean***

Thuật toán chia tập hợp D gồm có n đối tượng thành k cụm C (k ≤ n) cho trước dựa vào độ đo thích hợp. Mỗi cụm được xác đinh bởi trọng tâm của cụm. Một đối tượng được gom vào cụm Ci nếu khoảng cách của đối tượng đó tới tâm của cụm Ci là nhỏ nhất. Sau mỗi bước tâm của cụm sẽ được cập nhật lại. Quá trình lặp lại cho đến khi hàm mục tiêu bé hơn một ngưỡng cho phép hoặc tâm của cụm không thay đổi.

* Ưu điểm
* Nhanh
* Phù hợp với các cụm hình cầu
* Hạn chế
* Phải xác định trước k
* Không đảm bảo được tối ưu toàn cục
* Không thể xử lý nhiễu và ngoại lai
* Chỉ áp dụng được khi tính được trọng tâm

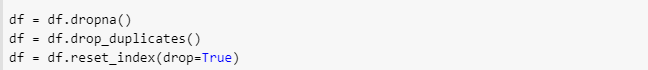
**PHẦN III: XỬ LÍ VÀ PHÂN TÍCH DỮ LIỆU**

**3.1. Làm sạch dữ liệu**

Bộ dữ liệu nhìn chung có vài thuộc tính bị thiếu dữ liệu, cần loại bỏ khỏi dataset đồng thời chỉ lấy những thuộc tính cần sử dụng là “Lithography” và “TDP”.

Text

Description automatically generated with medium confidence



Thông tin dữ liệu sau khi làm sạch

Text

Description automatically generated

Dữ liệu sẽ được đọc từ file csv đưa vào Pandas dataframe. Sau đó xoá bỏ dữ liệu trùng lặp và rỗng đi. Kết quả nhận được là một bộ dữ liệu 408 dòng và 2 cột.

**3.2. Chuyển đổi dataframe thành đồ thị**

Đồ thị 2 phía:

*Node: là tiến trình sản xuất (Lithography) và mức điện năng tiêu thụ (TDP).*

*Edge: mối quan hệ giữa việc mức điện năng tiêu thụ xuất hiện trên tiến trình sản xuất ra CPU.*

Graphical user interface, text, application

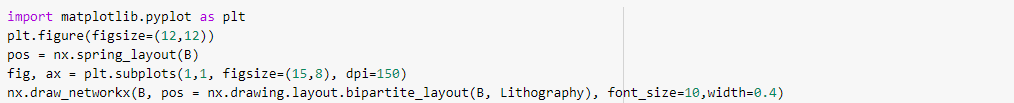
Description automatically generated

Graphical user interface, text, application

Description automatically generated

*Có thể thấy có 9 tiến trình (Lithograhy); Số mức điện tiêu thụ (TDP) là 230; Số cạnh là 408.*

Hiện thị đồ thị 2 phía:



A close-up of a graph

Description automatically generated with low confidence

*Nhìn vào đồ thị có thể thấy 1 mức điện năng tiêu thụ sẽ xuất hiện trên nhiều tiến trình, 1 tiến trình có nhiều mức điện năng tiêu thụ.*

Đồ thị 1 phía

*Node: là mức tiêu thụ điện năng (TDP)*

*Edge: Hai mức tiêu thụ điện năng cùng xuất hiện trên cùng tiến trình sẽ được nối với nhau để tạo thành cạnh, ý nghĩa thể hiện sự phổ biến của các mức điện tiêu thụ trên tiến trình.*

*Weight: Trọng số là số tiến trình trùng nhau mà hai mức điện tiêu cùng xuất hiện.*

*Text

Description automatically generated with medium confidence*

Diagram

Description automatically generated

Xuất dữ liệu đồ thị 1 phía ra file csv và thực hiện trên Gephi:

A picture containing text

Description automatically generated

Kết quả thực thi trên Gephi:

A picture containing background pattern

Description automatically generated

**PHẦN IV: XẾP HẠNG THEO ĐỘ ĐO**

**4.1. Độ đo PageRank**

***4.1.1. Thực thi trên Python***

Kết quả của Top 10 PageRank cao nhất:

Graphical user interface, text, application

Description automatically generated

***4.1.2. Thực thi trên Gephi***

Kết quả của Top 10 PageRank cao nhất:

Table

Description automatically generated

***4.1.3. So sánh kết quả***

* Kết quả gần như không có sự khác biệt giữa code python và Gephi. Hai kết quả gần như trùng khớp cả về thông số và thứ hạng. Chỉ trừ trường hợp có 2 hay nhiều node có PageRank bằng nhau, thứ hạng của chúng được sắp xếp ngẫu nhiên dẫn đến thứ hạng có chút khác biệt giữa code python và Gephi.

***4.1.4. Nhận xét***

* Mức điện năng tiêu thụ (node) có PageRank càng cao thì chúng là mức điện năng tiêu thụ CPU của càng nhiều tiến trình, các tiến trình này là tiến trình chung càng phổ biến.

**4.2. Độ đo Closeness Centrality**

***4.2.1. Thực thi trên Python***

Kết quả của Top 10 Closeness Centrality cao nhất:

Graphical user interface, text, application

Description automatically generated

***4.2.2. Thực thi trên Gephi***

**Table

Description automatically generated**

***4.2.3. So sánh kết quả***

* Kết quả gần như không có sự khác biệt giữa code python và Gephi. Hai kết quả gần như trùng khớp cả về thông số và thứ hạng. Chỉ trừ trường hợp có 2 hay nhiều node có Closeness Centrality bằng nhau, thứ hạng của chúng được sắp xếp ngẫu nhiên dẫn đến thứ hạng có chút khác biệt giữa code python và Gephi.

***4.2.4. Nhận xét***

* Mức năng lượng tiêu thụ điện (TDP) CPU có PageRank càng cao thì thì chúng là mức năng lượng CPU tìm thấy ở các tiến trình (Lithography) chung càng phổ biến.

**4.3. Độ đo Clustering Coefficient**

***4.3.1. Thực thi trên Python***

Kết quả của Top 10 Clustering Coefficient cao nhất:

Graphical user interface, text, application

Description automatically generated

***4.3.2. Thực thi trên Gephi***

Kết quả của Top 10 Clustering Coefficient cao nhất:

**Table

Description automatically generated**

***4.3.3. So sánh kết quả***

Kết quả về thông số không có sự khác biệt giữa code python và Gephi. Còn về thứ hạng do có lên đến 148 nodes có Clustering Coefficient bằng 1 được sắp xếp ngẫu nhiên giữa code Python và Gephi khác nhau nên dẫn đến thứ hạng không giống nhau.

***4.3.4. Nhận xét***

* Mức năng lượng tiêu thụ điện (TDP) CPU có Clustering Coefficient càng cao thì thì chúng là mức năng lượng CPU tìm thấy ở càng ít tiến trình (Lithography).

**PHẦN V: THUẬT TOÁN PHÁT HIỆN CỘNG ĐỒNG**

**5.1. Thuật toán Louvain**

***5.1.1. Thực thi trên Gephi***

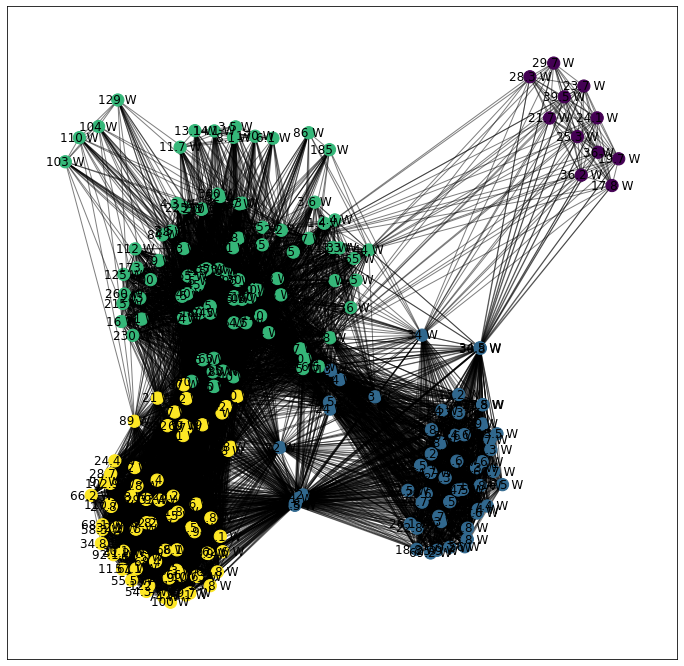
Background pattern

Description automatically generated

***5.1.2. Thực thi trên Python***

Text

Description automatically generated with medium confidence





A picture containing shape

Description automatically generated

Chart, scatter chart

Description automatically generated

**Số cụm: 4**

Cụm 1: ['29.7 W', '25.3 W', '39.5 W', '36 W', '36.2 W', '28.3 W', '24.1 W', '23.7 W', '21.7 W', '19.7 W', '17.8 W']

Cụm 2: ['7 W', '12 W', '10.63 W', '11.61 W', '28.7 W', '68.4 W', '30.4 W', '62.6 W', '23.2 W', '32.2 W', '12.1 W', '21 W', '59.8 W', '57.1 W', '89 W', '54.3 W', '25.8 W', '26.9 W', '30 W', '82 W', '20.1 W', '68.1 W', '32 W', '81.9 W', '66.1 W', '76 W', '31.2 W', '61.5 W', '38 W', '24.5 W', '29.9 W', '69.7 W', '21.8 W', '29.1 W', '69 W', '102.9 W', '81.8 W', '66.2 W', '110.7 W', '92.1 W', '58 W', '20.5 W', '91 W', '57 W', '122 W', '100 W', '97 W', '107 W', '77 W', '87 W', '74 W', '66 W', '92 W', '48 W', '19 W', '8.3 W', '66.8 W', '34.8 W', '58.3 W', '55.5 W', '52.8 W', '33.4 W', '32.1 W', '24.4 W', '28.9 W', '23.8 W', '63.5 W', '27.8 W']

Cụm 3: ['4.5 W', '15 W', '130 W', '4.3 W', '10 W', '55 W', '88 W', '35 W', '5.5 W', '6 W', '6.5 W', '7.5 W', '11.5 W', '37 W', '17 W', '5 W', '115 W', '27 W', '84 W', '65 W', '95 W', '103 W', '40 W', '86 W', '28 W', '70 W', '73 W', '31 W', '18 W', '53 W', '51 W', '25 W', '4 W', '45 W', '150 W', '140 W', '112 W', '2.2 W', '3.3 W', '3.5 W', '3 W', '90 W', '104 W', '75 W', '155 W', '185 W', '9 W', '20 W', '9.5 W', '16 W', '8 W', '8.1 W', '6.1 W', '8.5 W', '11.7 W', '13.1 W', '14.1 W', '0.65 W', '2 W', '3.6 W', '1.3 W', '1.4 W', '2.4 W', '2.5 W', '170 W', '260 W', '300 W', '80 W', '47 W', '135 W', '85 W', '50 W', '105 W', '120 W', '160 W', '145 W', '52 W', '60 W', '79 W', '110 W', '129 W', '165 W', '44 W', '225 W', '245 W', '270 W', '230 W', '215 W', '43 W', '54 W', '136 W', '125 W', '200 W', '205 W', '173 W']

Cụm 4: ['61 W', '22.5 W', '27.3 W', '20.8 W', '26.1 W', '19.5 W', '9.7 W', '18.3 W', '71.8 W', '22.8 W', '17.5 W', '34.5 W', '69.2 W', '30.7 W', '27.5 W', '15.8 W', '30.8 W', '66.7 W', '25.9 W', '22 W', '14.5 W', '24.6 W', '14 W', '64 W', '23 W', '13.2 W', '21.5 W', '57.8 W', '16.8 W', '33 W', '54.7 W', '15.5 W', '29 W', '51.6 W', '34 W', '13 W', '24 W', '72 W', '77.5 W', '29.6 W', '65.8 W', '27.4 W', '25.4 W', '68 W', '23.3 W', '59.2 W', '21.3 W', '19.2 W', '56 W', '20.6 W', '23.6 W', '18.8 W', '26.7 W', '16.5 W', '17.6 W', '25.7 W']

***5.1.3. Ý nghĩa các cụm***

Biểu đồ thể hiện đăc trưng của các cụm trên Python, kết quả hoàn toàn giống khi sử dụng excel để vẽ biểu đồ (file Louvain.xlsx đính kèm báo cáo):

Graphical user interface, text, application

Description automatically generated

Chart, bar chart, histogram

Description automatically generated

Ý nghĩa cụm thứ 1: các mức điện năng cùng xuất hiện trên tiến trình 250nm.

Ý nghĩa cụm thứ 2: các mức điện năng cùng xuất hiện trên tiến trình 180nm.

Ý nghĩa cụm thứ 3: các mức điện năng cùng xuất hiện trên tiến trình nhỏ là 14nm, 22nm, 45nm, 32nm.

Ý nghĩa cụm thứ 4: các mức điện năng cùng xuất hiện trên tiến trình 180nm.

**5.2. Thuật toán Girvan Newman**

***5.2.1. Thực thi trên Python***

Graphical user interface, text, application

Description automatically generated

Scatter chart

Description automatically generated

**Số cụm: 2**

Cụm 1:{'66.2 W', '29 W', '215 W', '30.8 W', '20.1 W', '19 W', '59.2 W', '23 W', '112 W', '3.5 W', '69.2 W', '30.4 W', '4.5 W', '270 W', '103 W', '25.8 W', '1.4 W', '18.3 W', '10 W', '80 W', '25.9 W', '129 W', '68 W', '57 W', '82 W', '28.9 W', '120 W', '64 W', '7.5 W', '28.7 W', '4 W', '90 W', '3.6 W', '8.5 W', '2.4 W', '5 W', '92 W', '58.3 W', '170 W', '30 W', '16.5 W', '136 W', '65 W', '140 W', '23.2 W', '2.2 W', '81.9 W', '27.5 W', '18 W', '21.3 W', '91 W', '97 W', '55 W', '30.7 W', '44 W', '31 W', '200 W', '92.1 W', '52 W', '8.1 W', '135 W', '68.4 W', '16.8 W', '13.1 W', '50 W', '130 W', '11.5 W', '115 W', '63.5 W', '107 W', '300 W', '27.4 W', '75 W', '150 W', '54.3 W', '45 W', '70 W', '13 W', '33.4 W', '6 W', '25 W', '3 W', '205 W', '20 W', '125 W', '102.9 W', '66.8 W', '61 W', '100 W', '56 W', '173 W', '12 W', '5.5 W', '22.5 W', '34.5 W', '34.8 W', '26.7 W', '25.7 W', '145 W', '19.5 W', '95 W', '9.7 W', '77 W', '2.5 W', '81.8 W', '6.1 W', '55.5 W', '24.5 W', '8 W', '32.1 W', '89 W', '6.5 W', '22.8 W', '32 W', '11.61 W', '14 W', '7 W', '33 W', '15 W', '58 W', '37 W', '25.4 W', '66 W', '47 W', '29.9 W', '110.7 W', '57.8 W', '24.6 W', '8.3 W', '24 W', '0.65 W', '59.8 W', '12.1 W', '71.8 W', '245 W', '14.5 W', '20.6 W', '79 W', '29.1 W', '21 W', '3.3 W', '69 W', '22 W', '18.8 W', '66.1 W', '13.2 W', '10.63 W', '60 W', '66.7 W', '4.3 W', '76 W', '86 W', '27 W', '9 W', '77.5 W', '260 W', '165 W', '57.1 W', '17.5 W', '20.5 W', '9.5 W', '88 W', '19.2 W', '84 W', '23.6 W', '26.1 W', '20.8 W', '26.9 W', '15.5 W', '23.3 W', '73 W', '61.5 W', '11.7 W', '15.8 W', '53 W', '31.2 W', '29.6 W', '68.1 W', '35 W', '43 W', '62.6 W', '230 W', '38 W', '17 W', '32.2 W', '21.5 W', '122 W', '40 W', '110 W', '87 W', '21.8 W', '27.3 W', '52.8 W', '34 W', '1.3 W', '74 W', '155 W', '54 W', '185 W', '51 W', '85 W', '105 W', '65.8 W', '27.8 W', '2 W', '104 W', '16 W', '23.8 W', '51.6 W', '28 W', '48 W', '54.7 W', '69.7 W', '17.6 W', '225 W', '24.4 W', '160 W', '72 W', '14.1 W'}

Cụm 2: {'39.5 W', '28.3 W', '19.7 W', '25.3 W', '29.7 W', '17.8 W', '36.2 W', '21.7 W', '24.1 W', '23.7 W', '36 W'}

***5.2.3. Ý nghĩa các cụm***

Biểu đồ thể hiện đăc trưng của các cụm trên Python, kết quả hoàn toàn giống khi sử dụng excel để vẽ biểu đồ (file Girvan\_Newman.xlsx đính kèm báo cáo):

Text

Description automatically generated with low confidence

Chart, bar chart, histogram

Description automatically generated

Ý nghĩa cụm thứ 1: các mức điện năng xuất hiện trên các tiến trình nhỏ hơn 250 (nm).

Ý nghĩa cụm thứ 2: các mức điện năng chỉ xuất hiện trên tiến trình 250nm.

**5.3. Thuật toán K-Mean**

***5.3.1. Thực thi trên Python***

Loại bỏ kí tự chuỗi, chuyển thuộc tính qua kiểu float





Graphical user interface, text, application

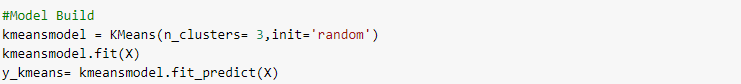
Description automatically generated

A picture containing chart

Description automatically generated

Ta thấy từ k=3 trở lên đường biểu diễn giá trị distortion/inertia giảm dần theo tuyến tính. Theo phương pháp khuỷu tay chọn k=3.

Xây dựng model, tiến hành phân cụm:



Text

Description automatically generated with medium confidence

A picture containing icon

Description automatically generated

**Số cụm: 3**

Cụm thứ 1: ['7.0 W', '12.0 W', '10.63 W', '11.61 W', '28.7 W', '68.4 W', '30.4 W', '62.6 W', '23.2 W', '32.2 W', '61.0 W', '12.1 W', '22.5 W', '59.8 W', '27.3 W', '20.8 W', '57.1 W', '20.8 W', '26.1 W', '19.5 W', '89.0 W', '54.3 W', '9.7 W', '25.8 W', '18.3 W', '71.8 W', '26.9 W', '22.8 W', '17.0 W', '30.0 W', '17.5 W', '34.5 W', '82.0 W', '69.2 W', '30.7 W', '20.1 W', '68.1 W', '27.5 W', '19.5 W', '32.0 W', '15.8 W', '30.8 W', '66.7 W', '25.9 W', '22.0 W', '14.5 W', '29.7 W', '81.9 W', '66.1 W', '24.6 W', '35.0 W', '14.0 W', '28.0 W', '64.0 W', '23.0 W', '76.0 W', '13.2 W', '31.2 W', '61.5 W', '38.0 W', '21.5 W', '70.0 W', '24.5 W', '25.3 W', '29.9 W', '69.7 W', '61.0 W', '34.5 W', '57.8 W', '16.8 W', '21.8 W', '29.1 W', '33.0 W', '69.0 W', '102.9 W', '54.7 W', '15.5 W', '21.0 W', '29.0 W', '81.8 W', '66.2 W', '110.7 W', '92.1 W', '58.0 W', '51.6 W', '34.0 W', '20.5 W', '122.0 W', '90.0 W', '100.0 W', '97.0 W', '91.0 W', '107.0 W', '40.0 W', '85.0 W', '74.0 W', '72.0 W', '80.0 W', '60.0 W', '66.0 W', '65.0 W', '57.0 W', '92.0 W', '77.5 W', '55.0 W', '87.0 W', '29.6 W', '65.8 W', '27.4 W', '34.0 W', '72.0 W', '25.4 W', '39.5 W', '68.0 W', '77.0 W', '23.3 W', '59.2 W', '21.3 W', '36.0 W', '48.0 W', '19.2 W', '36.2 W', '44.0 W', '30.8 W', '56.0 W', '8.3 W', '20.0 W', '20.6 W', '23.6 W', '66.8 W', '6.5 W', '28.0 W', '24.0 W', '34.8 W', '18.8 W', '26.7 W', '16.5 W', '22.0 W', '17.6 W', '25.7 W', '58.3 W', '15.0 W', '28.3 W', '55.5 W', '19.0 W', '27.0 W', '52.8 W', '33.4 W', '13.0 W', '24.1 W', '32.1 W', '23.7 W', '24.4 W', '28.9 W', '21.7 W', '23.8 W', '63.5 W', '19.7 W', '27.8 W', '17.8 W']

Cụm thứ 2: ['4.5 W', '15.0 W', '4.3 W', '10.0 W', '55.0 W', '35.0 W', '5.5 W', '6.0 W', '6.5 W', '10.0 W', '7.5 W', '11.5 W', '37.0 W', '15.0 W', '17.0 W', '5.0 W', '7.5 W', '27.0 W', '65.0 W', '21.0 W', '40.0 W', '40.0 W', '65.0 W', '35.0 W', '65.0 W', '35.0 W', '35.0 W', '65.0 W', '17.0 W', '73.0 W', '31.0 W', '18.0 W', '53.0 W', '51.0 W', '35.0 W', '25.0 W', '4.5 W', '4.0 W', '28.0 W', '45.0 W', '65.0 W', '57.0 W', '2.2 W', '3.3 W', '3.5 W', '3.0 W', '10.0 W', '7.0 W', '75.0 W', '6.0 W', '9.0 W', '14.0 W', '12.0 W', '20.0 W', '9.5 W', '13.0 W', '18.0 W', '31.0 W', '32.0 W', '24.0 W', '21.5 W', '21.0 W', '17.0 W', '16.0 W', '11.5 W', '9.0 W', '9.5 W', '5.0 W', '8.0 W', '6.5 W', '8.1 W', '6.1 W', '8.5 W', '11.7 W', '13.1 W', '14.1 W', '5.0 W', '12.0 W', '0.65 W', '2.0 W', '3.6 W', '1.3 W', '2.2 W', '7.0 W', '1.4 W', '4.0 W', '5.5 W', '3.0 W', '6.5 W', '10.0 W', '13.0 W', '2.4 W', '4.5 W', '8.5 W', '2.5 W', '8.0 W', '80.0 W', '80.0 W', '47.0 W', '80.0 W', '72.0 W', '50.0 W', '75.0 W', '55.0 W', '52.0 W', '75.0 W', '65.0 W', '45.0 W', '50.0 W', '70.0 W', '60.0 W', '70.0 W', '60.0 W', '50.0 W', '73.0 W', '79.0 W', '25.0 W', '82.0 W', '69.0 W', '77.0 W', '20.0 W', '45.0 W', '20.0 W', '55.0 W', '80.0 W', '15.0 W', '60.0 W', '20.0 W', '40.0 W', '80.0 W', '38.0 W', '50.0 W', '30.0 W', '45.0 W', '40.0 W', '50.0 W', '23.0 W', '48.0 W', '30.0 W', '55.0 W', '28.0 W', '47.0 W', '27.0 W', '43.0 W', '27.0 W', '6.0 W', '5.5 W', '9.0 W', '10.0 W', '34.0 W', '54.0 W', '25.0 W', '17.0 W', '28.0 W', '17.0 W', '82.0 W', '25.0 W', '75.0 W', '44.0 W', '44.0 W', '55.0 W', '19.0 W', '54.0 W', '73.0 W', '30.0 W', '19.0 W', '70.0 W', '60.0 W']

Cụm thứ 3: '130.0 W', '88.0 W', '115.0 W', '84.0 W', '95.0 W', '103.0 W', '130.0 W', '89.0 W', '86.0 W', '130.0 W', '95.0 W', '91.0 W', '150.0 W', '130.0 W', '140.0 W', '140.0 W', '112.0 W', '104.0 W', '155.0 W', '185.0 W', '170.0 W', '260.0 W', '300.0 W', '95.0 W', '135.0 W', '115.0 W', '85.0 W', '135.0 W', '90.0 W', '105.0 W', '120.0 W', '160.0 W', '145.0 W', '85.0 W', '105.0 W', '135.0 W', '90.0 W', '120.0 W', '160.0 W', '145.0 W', '150.0 W', '95.0 W', '115.0 W', '95.0 W', '84.0 W', '87.0 W', '110.0 W', '135.0 W', '95.0 W', '90.0 W', '105.0 W', '130.0 W', '120.0 W', '85.0 W', '150.0 W', '129.0 W', '165.0 W', '105.0 W', '150.0 W', '120.0 W', '225.0 W', '245.0 W', '270.0 W', '245.0 W', '230.0 W', '215.0 W', '105.0 W', '115.0 W', '88.0 W', '87.0 W', '136.0 W', '125.0 W', '130.0 W', '150.0 W', '165.0 W', '200.0 W', '205.0 W', '173.0 W', '165.0 W', '155.0 W']

***5.3.2. Ý nghĩa các cụm***

Biểu đồ thể hiện đăc trưng của các cụm trên Python, kết quả hoàn toàn giống khi sử dụng excel để vẽ biểu đồ (file Girvan\_Newman.xlsx đính kèm báo cáo):

Graphical user interface, text, application

Description automatically generated

Graphical user interface, text, application

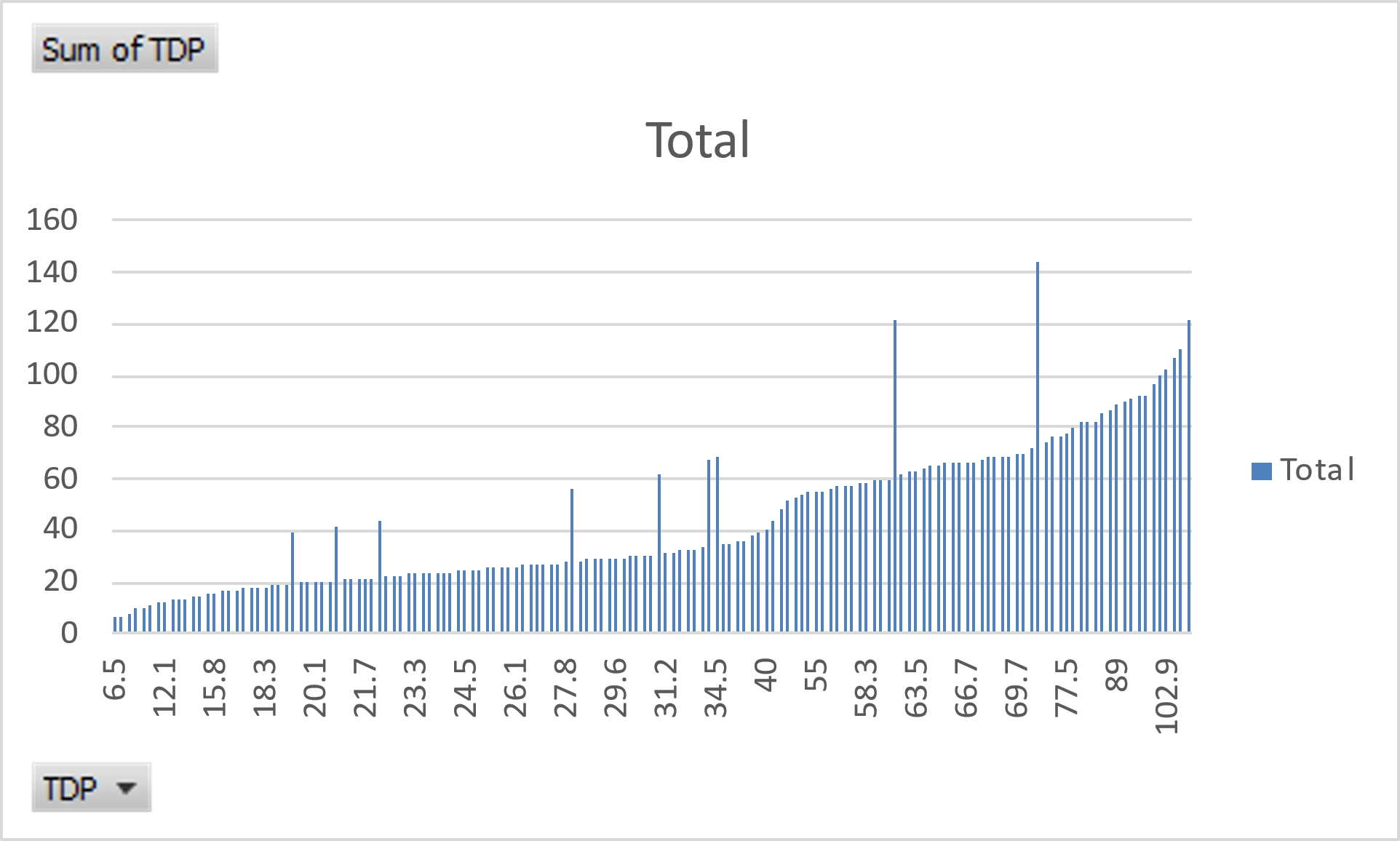
Description automatically generated

Chart, bar chart

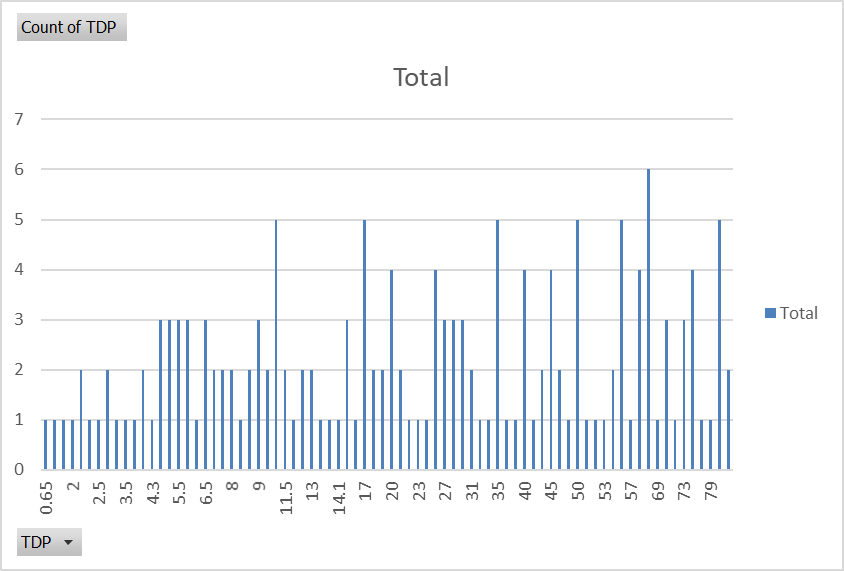
Description automatically generated

Biểu đồ đếm số lượng TDP từng cụm đế so sánh tương quan với Lithography:

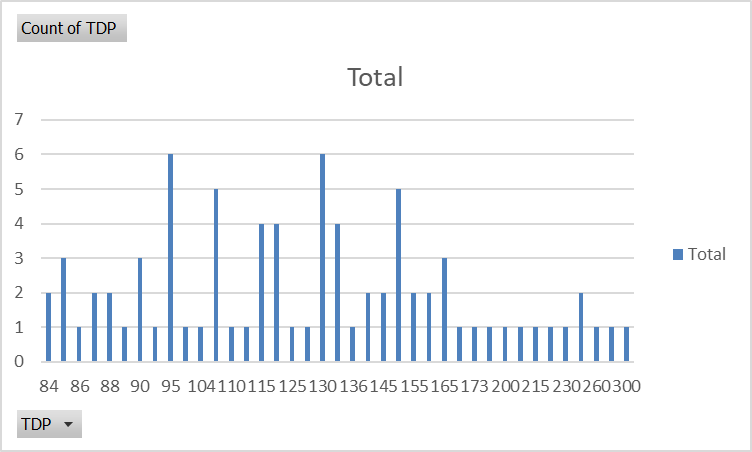
*Cụm 1:*



*Cụm 2:*



*Cụm 3:*

**

Ý nghĩa cụm thứ 1: các mức điện năng tiêu thụ nhỏ xuất hiện trên các tiến trình lớn.

Ý nghĩa cụm thứ 2: các mức điện năng tiêu thụ nhỏ xuất hiện trên các tiến trình nhỏ.

Ý nghĩa cụm thứ 3: các mức điện năng tiêu thụ lớn xuất hiện trên các tiến trình nhỏ.

**5.4. Tổng kết kết quả**

* Số lượng cộng đồng của 3 thuật toán: Louvain: 4, Givan Newman: 2, K-Mean: 3.
* Có duy nhất 1 cộng đồng giống nhau hoàn toàn giữa 2 thuật toán đó là cộng đồng thứ 1(cụm thứ 1) của Louvain và cộng đồng thứ 2(cụm thứ 2) của Givan Newman.
* Các thuật toán đa phần cho ra kết quả số lượng cộng động, phẩn tử ở mỗi cộng đồng khác nhau và mang đặc trưng khác nhau.
* Thuật toán Louvain phân ra nhiều cộng đồng nhất và đặc trưng mỗi cộng đồng rất rõ ràng.

**TÀI LIỆU THAM KHẢO**

[1] [*https://towardsdatascience.com/how-to-get-started-with-social-network-analysis-6d527685d374*](https://towardsdatascience.com/how-to-get-started-with-social-network-analysis-6d527685d374)

[2][*https://scikit-learn.org/stable/modules/generated/sklearn.cluster.KMeans.html*](https://scikit-learn.org/stable/modules/generated/sklearn.cluster.KMeans.html)

[3][*https://www.geeksforgeeks.org/elbow-method-for-optimal-value-of-k-in-kmeans/*](https://www.geeksforgeeks.org/elbow-method-for-optimal-value-of-k-in-kmeans/)

[1][*http://hnkh.tlu.edu.vn/Portals/7/2014/NewFiles/005.pdf*](http://hnkh.tlu.edu.vn/Portals/7/2014/NewFiles/005.pdf)

-----------------------------------------------------END------------------------------------------------