Project Report

On

CLUSTER HYPERGRAPH

Submitted

in partial fulfilment of the requirements for the degree of

MASTER OF SCIENCE

in

MATHEMATICS

by

ARCHANA K S

(Register No. SM22MAT002)

(2023-2024)

Under the Supervision of

JOSMY THOMAS

DEPARTMENT OF MATHEMATICS

ST. TERESA'S COLLEGE (AUTONOMOUS)

ERNAKULAM, KOCHI - 682011

APRIL 2024

ST. TERESA'S COLLEGE (AUTONOMOUS), ERNAKULAM

CERTIFICATE

This is to certify that the dissertation entitled, **CLUSTER HYPERGRAPH** is a bonafide record of the work done by Ms. **ARCHANA K S** under my guidance as partial fulfillment of the award of the degree of **Master of Science in Mathematics** at St. Teresa's College (Autonomous), Ernakulam affiliated to Mahatma Gandhi University, Kottayam. No part of this work has been submitted for any other degree elsewhere.

Date:
Place: Ernakulam

JOSMY THOMAS
Assistant Professor,
Department of Mathematics,
St. Teresa's College(Autonomous),
Ernakulam.

Nisha Oommen
Assistant Professor and HOD,
Department of Mathematics ans Statistics,
St. Teresa's College(Autonomous),
Ernakulam.

External Examiners

1:..............................

2: ....

DECLARATION

I hereby declare that the work presented in this project is based on the original work done by me under the guidance of **JOSMY THOMAS**, Assistant Professor, Department of Mathematics, St. Teresa's College(Autonomous), Ernakulam and has not been included in any other project submitted previously for the award of any degree.

Ernakulam. **ARCHANA K S**

Date: **SM22MAT002**

ACKNOWLEDGEMENTS

I must mention several individuals who encouraged me to carry out this work. Their continuous invaluable knowledgeable guidance throughout this study helped me to complete the work up to this stage
I am very grateful to my project guide Josmi Thomas for the immense help during the period of work

In addition, the very energetic and competitive atmosphere of the Department had much to do with this work. I acknowledge with thanks to faculty, teaching and non-teaching staff of the department, and Colleagues.

I am also very thankful to the HOD for their valuable suggestions and critical examination of work during the progress.

Ernakulam. **ARCHANA K S**
Date: **SM22MAT002**

Contents

Introduction

The graphs are the perfect representation of the relationship among the data. This concept was introduced by the great mathematician L. Euler, often known as the father of graph theory when he famously solved the Koingsberg bridge problem in 1736. but since the graph can only capture the interaction between pairs of vertices, they fall short in many cases and they are not able to model the complexity in some networks. To overcome this limitation, C. Berge introduced hypergraphs in 1961, allowing edges to encompass any number of vertices. cluster hypergraph is introduced to generalized hypergraph in which clustering of nodes is allowed and this development was done by Samanta El in 2020.

In this project on cluster hypergraphs, we first deal with the basic terminologies related and the association of a hypergraph with a graph, along a real-world example. Then we move on to the basic definition, and types of cluster hypergraphs and its completeness property. With a brief discussion on competition hypergraphs, we conclude the conceptual discussion to showcase two applications, one on cluster hypergraph and the other on competition cluster hypergraph.

PRELIMINARIES

0.1 Semi-directed Graph

Let V be a nonempty set of elements, called vertices or nodes. Also, let $E = E_1 \cup \overrightarrow{E}_2$ where $E_1 \subset V \times V$ is a set of unordered pairs of vertices, i.e., $E_1 = \{(u, v) \mid u, v \in V\}$, called a set of undirected edges, and $\overrightarrow{E}_2 \subset V \times V$ is a set of ordered pairs of vertices, $\overrightarrow{E}_2 = \{(a, b) \mid a, b \in V\}$, called a set of directed edges. Here, $G = (V, E_1, \overrightarrow{E}_2)$, is said to be a semidirected graph.

0.1.1 Degree

The degree of a vertex u is denoted as a triplet $D(u) = (d(u), d^+(u), d^-(u))$ where $d(u)$ is the number of all incident edges of u in E_1, $d^+(u)$ is the number of out-directed edges of $\overrightarrow{E}_2$ from the vertex u, and $d^-(u)$ is the number of in-directed edges of $\overrightarrow{E}_2$ towards the vertex u. Now, the incidence number of a vertex u is denoted as $\mathrm{in}(u)$ and defined as $\mathrm{in}(u) = d(u) + d^+(u) - d^-(u)$.

Example

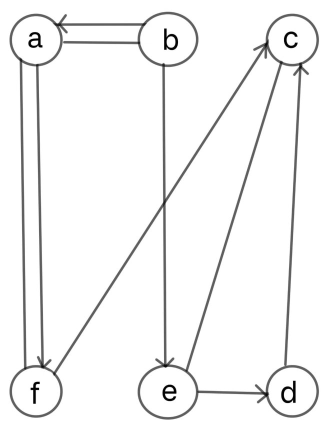

Figure 1: Semi-directed graph

The degree of a vertex of the graph, as shown in Figure 1, is given as D(a) = (2,1,1). Now, the incident number is in (a) = 2+1-1 = 2.

0.1.2 Complete Semi-directed Graph

In a semi-directed graph, if there are all three types of connections i.e., out-directed edges, in-directed edges, and undirected edges between any two vertices. then the graph is called a complete semi-directed graph.

0.1.3 Complete Incident Semi-directed Graph

A semi-directed graph is said to be a complete incidence semi-directed graph if every pair of vertices is connected by at least one edge (undirected or directed), and the incidence number of all vertices is equal.

Example

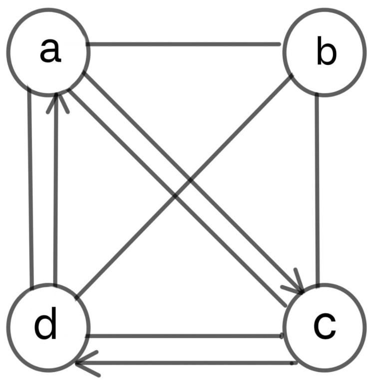

Figure 2

In Figure 2, there are connections (undirected or directed) between every pair of vertices and the incidence number of each of the vertices is 3, it is a complete-incidence semidirected graph.

0.1.4 Neighbourhood, Out-neighbourhood, In-neighbourhood

Neighborhood, out-neighborhood, and in-neighborhood of a vertex u in a semidirected graph $G = (V, E_1, \overrightarrow{E}_2)$ are denoted as $N(u)$, $N^+(u)$, and $N^-(u)$ and defined as follows:

$$N(u) = \{v \in V \mid (u, v) \in E_1\}$$

$$N^+(u) = \{v \in V \mid (u, v) \in \overrightarrow{E}_2\}$$

$$N^-(u) = \{v \in V \mid (v, u) \in \overrightarrow{E}_2\}$$

Example

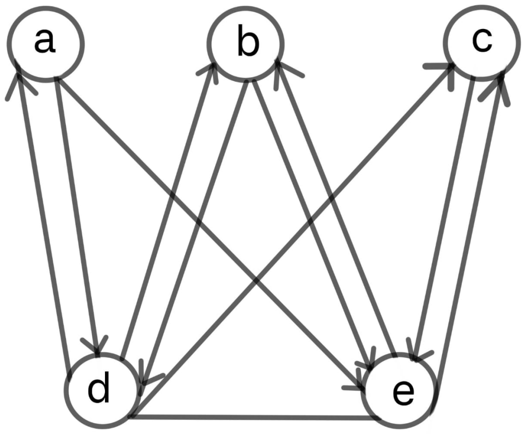

Figure 3

From Figure 3, we get:

$$N(d) = \{e\},$$
$$N^+(d) = \{a, b, c\},$$
$$N^+(e) = \{b, c\},$$
$$N^-(d) = \{a, b\},$$
$$N^-(e) = \{a, b, c\}.$$

Thus, $N^+(d)$ is a maximal out-neighborhood set and $N^-(e)$ is a maximal in-neighborhood set of the semidirected graph assumed in Figure 3.

0.1.5 m-step Neighborhood, m-step Out-neighborhood, m-step In-neighborhood

The m-step neighbourhood, m-step out-neighbourhood and m-step in-neighbourhood of a vertex u in a semidirected graph $G = (V, E_1, \overrightarrow{E}_2)$ are denoted as $N_m(u)$, $N_m^+(u)$, and $N_m^-(u)$ and defined as follows:

$$N_m(u) = \{v_m \in V : \text{for all paths such that } u - v_1 - v_2 - \cdots - v_m\},$$
$$N_m^+(u) = \{v_m \in V : \text{for all paths such that } u \to v_1 \to v_2 \to \cdots \to v_m\},$$
$$N_m^-(u) = \{v_m \in V : \text{for all paths such that } u \leftarrow v_1 \leftarrow v_2 \leftarrow \cdots \leftarrow v_m\}.$$

Example

In Figure 3, $N_2(d) = \emptyset$, $N_2^+(d) = \{e\}$, and $N_2^-(d) = \{e\}$.

0.1.6 m-step Semi-directed Graph

The m-step semi-directed graph $G_m = (V_m, E_m^1, \overrightarrow{E}_m^2)$ of a semi-directed graph $G = (V, E_1, \overrightarrow{E}_2)$ is defined as follows:

1. Vertex set of G_m is $V_m = V$

2. Edge set of G_m is $E_m^1 = \{(u, v_m) : v_m \in N_m(u)\}$,$\overrightarrow{E}_m^2 = \{(u, v_m) : v_m \in N_m^+(u) or N_m^-(u)\}$

Example

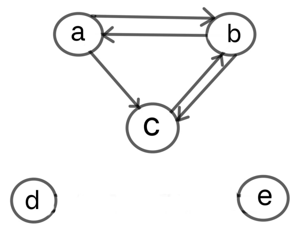

Figure 4: 2-step semi-directed graph of figure 3

Chapter 1

HYPRGRAPH

1.1 Hypergraph

Let $X = \{x_1, x_2, \ldots, x_n\}$ be a finite set, and let $E = \{e_1, e_2, \ldots, e_m\}$ be a family of subsets of X such that $e_i \neq \varnothing$ for $i = 1, 2, \ldots, m$. We have $\sum_{i=1}^{m} e_i = X$. The pair $H = (X, E)$ is called a hypergraph with vertex set X and hyperedge set E. The elements $\{x_1, x_2, \ldots, x_n\}$ of X are vertices of hypergraph H, and the sets $\{e_1, e_2, \ldots, e_m\}$ are hyperedges of hypergraph H.

1.1.1 Example

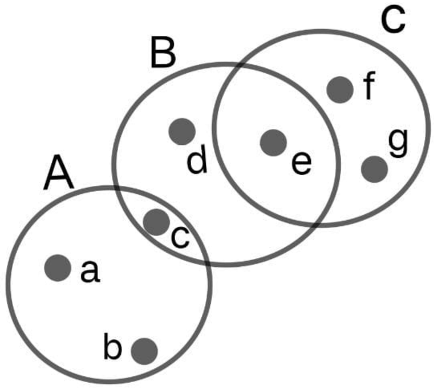

Figure 1.1: Hypergraph

Here set {A, B, C} are the hyperedges and set {a,b,c,d,e,f,g} are the vertices of the above hypergraph. where the hyperedge A contains the vertices {a,b,c}. similarily the hyperedges B and C contain the vertices {c,d,e} and {e,f,g}.

1.2 Terminology

1.Order of a hypergraph: Is the total number of vertices in a hypergraph. In Figure 1.1 order of the hypergraph is seven.

2.Size of a hypergraph: Is the total number of edges in a hypergraph. In Figure 1.1 size of the hypergraph is three.

3.Incident: A vertex v is said to be incident to an edge e if $v \in e$, similarly an edge e is said to be incident to a vertex v if $v \in e$. In Figure 1.1 vertex b is incident to hyperedge A and hyperedge A is incident to vertex b. since $b \in A$.

4.Pendant vertex: A vertex incident to exactly one edge. Then that vertex is called the incident vertex. No pendant vertex in Figure 1.1

5.Included edge: Included edge is an edge that is the subset of another edge. No included edge in Figure 1.1.

6.Multiple edges : An edge that has the same set of vertices as another edge.

7.Loop: Hyperedge with a single vertex

8.Simple hypergraph: Hypergraph without loops, included edges, multiple edges. Hypergraph in the Figure 1.1 is not a simple hypergraph

1.3 Graph And Hypergraph Association

1.3.1 Clique Graph Of A Hypergraph

Let $H = (V, E)$ be a hypergraph. we define the clique graph $G(H)$ as a simple graph on V, with an edge between $u, v \in V$ if there is a hyperedge $e \in E$ with $u, v \in e$. The name clique graph comes from the fact that if $\{v_1, v_2, .., v_n\}$ is a hyperedge in H, then $v_1, v_2, .., v_n$ form a clique in G.

Example

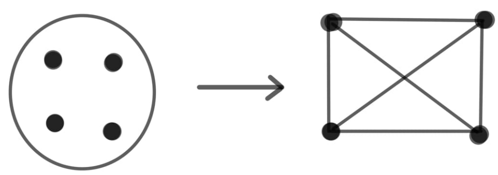

Figure 1.2: Hypergraph H and clique graph G(H)

1.3.2 Maximal Hypergraph Associated To Graph

Let $G = (V, E)$ be a simple graph. Let

$$C_1, \ldots, C_k$$

be the maximal cliques of G, and let $E_1 = \{\{u, v\} : u, v \in C_i forsomei\}$,

$$E_2 = \{C_1, \ldots, C_k\}$$

The maximal hypergraph associated $H(G)$ to G, is a hypergraph on V with hyperedges $(E \setminus E_1) \cup E_2$

Example

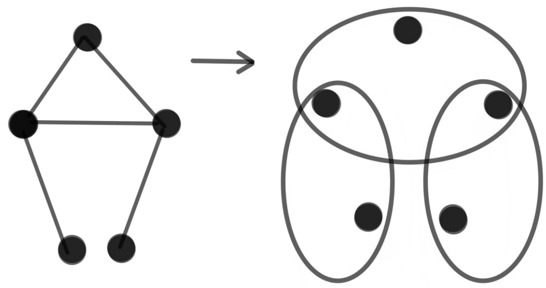

Figure 1.3: Graph G and maximal hypergraph H(G)

1.4 Complex Networks As Hypergraph

The study of complex networks represents an important area of multidisciplinary research involving physics, mathematics, chemistry, biology, social sciences, and information sciences, among others. These systems are commonly represented utilizing simple or directed graphs that consist of sets of nodes representing the objects joined together in pairs by links if the corresponding nodes are related by some kind of relationship. In some cases, the use of simple or directed graphs to represent complex networks does not provide a complete description of the real-world systems. A natural way of representing these systems is to use a generalization of graphs known as hypergraphs.

1.4.1 Real-life Example of Hypergraph

An example of complex systems for which hypergraph representation is necessary is the food web.

Trophic relations in ecological systems are normally represented through the use of food webs, which are oriented graphs (digraphs) where the species is represented by the nodes and links represent trophic relations between species. Another way of representing food webs is employing competition graphs C(G), which have the same set of nodes as the food web but two nodes are connected if, and only if, the corresponding species compete for the same prey in the food web. In the competition graph, we can only determine if two connected species share any prey in common, but it doesn't provide any information about the composition of the whole group of species that compete for common prey. To solve this problem a competition hypergraph has been proposed in which species are represented by the nodes in the food web and groups of species are represented by the hyperedges. In which the group of species competes for common prey. It has been shown that in many cases competition hypernetworks yield a more detailed description of the predation relations among the species in the food web than competition graphs. A food web and its competition network and hyper-network are illustrated in Fig below

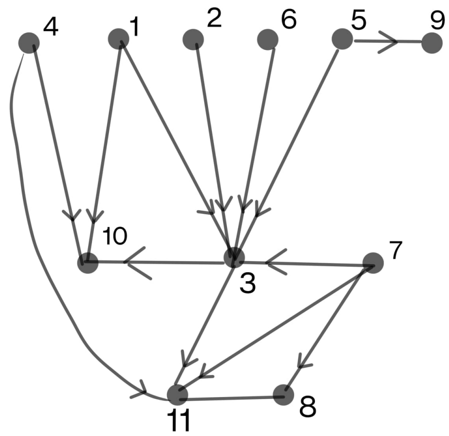

Figure 1.4: food web

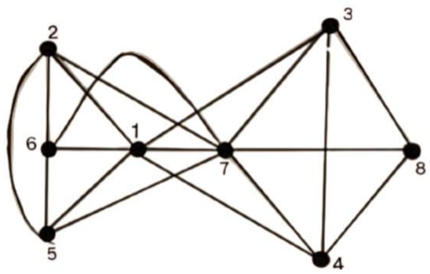

Figure 1.5: Competence network of food web

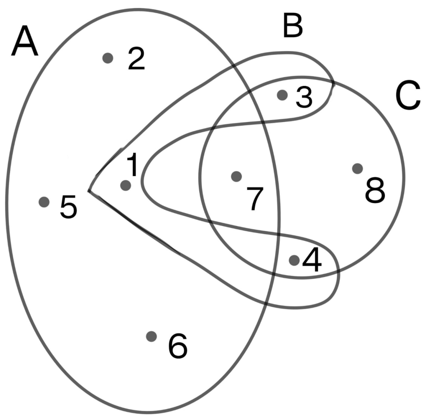

Figure 1.6: Competence hypergraph of figure 1.5

Chapter 2

CLUSTER HYPERGRAPH

2.1 Cluster Hypergraph

Let X be a nonempty set, and let V_x be a subset of $\mathcal{P}(X)$ such that

1. $\phi \notin E$.
2. For each element $e \in E$, there exists at least one element $v \in V_x$ such that $v \in e$.

Then, $G = (V_x, E)$ is said to be a cluster hypergraph where V_x is the vertex set and E is the multi-hyperedge set.

Example

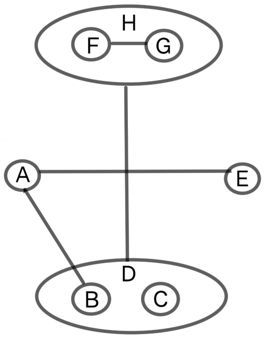

Figure 2.1

Let $X = \{A, B, C, E, F, G\}$, $V_x = \{\{A\}, \{B\}, \{C\}, D = \{B, C\}, H = \{F, G\}, \{E\}, \{F\}, \{G\}\}$, and $E = \{\{\{A\}, \{B\}\}, \{\{A\}, \{E\}\}, \{\{B, C\}, \{B\}\}, \{\{B, C\}, \{C\}\}, \{\{F\}, \{G\}\}, \{\{F, G\}, \{F\}\}, \{\{F, G\}, \{G\}\}, \{F, G\}, \{B, C\}\}$. It can be easily verified that for each element e in E, there exists an element $v \in V_x$ such that

$v \in e$. For example $\{\{A\}, \{E\}\}$ in E, there exists an element $\{A\}$ in V_x such that $\{A\} \in \{\{A\}, \{E\}\}$.

Remark

1. The vertex set of a cluster hypergraph may contain a group of people/individuals as a node (cluster node), and all the people in the network are assumed as simple nodes. This concept is helpful to assume any organization or group as a node in any network. Also, it is assumed that each node inside a cluster node is automatically connected to the cluster node, but these inside nodes may not be connected.

2. In the virtual representation of any cluster hypergraph, the cluster nodes are assumed as separate nodes, and their connections in representation are shown below, which is the virtual representation of Figure 2.1.

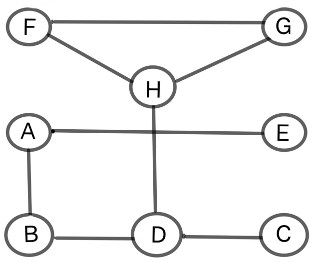

Figure 2.2

2.1.1 Degree

Let $X = \{x_1, x_2, \ldots, x_m\}$ be a non-empty set and $G = (V_X, E)$ be a cluster hypergraph where $V_X = \{v_1, v_2, \ldots, v_n\}$ is the set of nodes, i.e., $v_i \in \mathcal{P}(X)$ for $i = 1, 2, \ldots, n$, and $E = \{e_1, e_2, \ldots, e_k\}$ is the set of edges, i.e., $e_i \in \mathcal{P}(\mathcal{P}(X))$ for $i = 1, 2, \ldots, k$.

The degree of node v_i contained in the edges e_{ij}, $j = 1, 2, \ldots, p$, denoted as $d(v_i)$, is defined as:

$$d(v_i) = |\sum_{j=1}^{p} e_{ij}|$$

Example

In the cluster hypergraph (figure 2.1) degree of c is 1.

Theorem

Let $G = (V_X, E)$ be a cluster hypergraph where $|X| = k$. Then, the total degree of the cluster hypergraph is less than $(2^k - 1) \cdot 2^{2^k-2}$.

Proof:

Let $G = (V_X, E)$ be a cluster hypergraph where $|X| = k$. Thus, the maximum number of nodes of G is $2^k - 1$. Now, in cluster hypergraphs, there may be edges containing a single node, double nodes, and so on. Thus, the contribution of edges containing n nodes is $n\times$ (number of edges). Therefore, the maximum contribution of all edges is

$$1\binom{2k-1}{1} + 2\binom{2k-1}{2} + 3\binom{2k-1}{3} + \ldots + (2k-1)\binom{2k-1}{2k-1}.$$

Hence the result.

The number of connections to a node is termed as the degree of the node. If the node is in any of the cluster nodes then the node has a separate effect of connection. For this, we introduced and defined another term for the node, an effective node.

2.1.2 Effective Degree

Let $X = \{x_1, x_2, \ldots, x_m\}$ be a non-empty set and $G = (V_X, E)$ be a cluster hypergraph where $V_X = \{v_1, v_2, \ldots, v_n\}$ is the set of nodes, i.e., $v_i \in \mathcal{P}(X)$ for $i = 1, 2, \ldots, n$, and $E = \{e_1, e_2, \ldots, e_k\}$ is the set of edges, i.e., $e_i \in \mathcal{P}(\mathcal{P}(X))$ for $i = 1, 2, \ldots, k$.

Also, let CV_i be the cluster node containing the simple node V_i. Now, the effective degree of a simple node V_i is denoted as $\mathrm{ed}(V_i)$ and is defined as:

$$\mathrm{ed}(V_i) = d(V_i) + \frac{1}{l}\sum_{i=1}^{p} d(CV_i)$$

where l is the number of cluster nodes containing V_i.

Example

In the cluster hypergraph (Figure 1.1), the degree of node C is 1, but the effective degree of node C is $1 + \frac{2}{1} = 3$.

Theorem

Let $G = (V_X, E)$ be a cluster hypergraph where $|X| = k$. Then, the total effective degree of the cluster hypergraph is less than $\delta + (2^k - 1)(2^k - 2)$, where δ is the sum of the maximum degrees of nodes, and l is the number of cluster nodes in the hypergraph.

Proof:

Let $G = (V_X, E)$ be a cluster hypergraph where $X = (x_1, x_2, \ldots, x_k)$. Thus, the maximum number of nodes of G is $2^k - 1$. Now, an effective degree is

$$\mathrm{ed}(x_i) = d(x_i) + \frac{\sum_{l=1}^{l} d(c_{x_i})}{l}$$

where l is the number of cluster nodes containing x_i. Here, the sum of the effective degree is

$$\sum_{i=1}^{k} \mathrm{ed}(x_i) = \sum_{k=1}^{k} \left[d(v_i) + \left(\sum_{i=1}^{k} \frac{d(c_{x_i})}{l} \right) \right] = \sum_{i=1}^{k} d(v_i) + \sum_{i=1}^{k} \left(\frac{\sum_{k=1}^{k} d(c_{x_i})}{l} \right)$$

Let us suppose the total number of cluster nodes in the hypergraph is l. Now, the degree of a cluster node is always less than $2^k - 2$. Thus, the sum of the degrees of all cluster nodes is less than or equal to $l \times 2^k - 2$. One simple node may belong to all cluster nodes. Thus, the sum of all effective degrees is less than or equal to $\delta + (2^k - 1)(2^k - 2)$, where $\delta = (2^k - 1)2^{2^{k-2}}$ and l is the number of cluster nodes in the hypergraph. Hence, the result is true.

2.2 Types Of Cluster Hypergraph

Depending on the cluster node sizes and their edges, cluster hypergraphs are classified into different categories. To classify, maximal nodes are to be defined. The maximal nodes are those nodes that are not contained in any other cluster nodes. A simple node may be termed a maximal node if it does not belong to any other node.

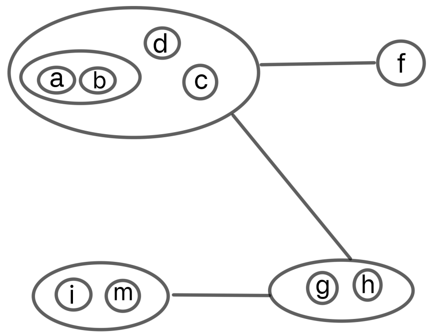

Figure 2.3

In Figure 2.3 the node f is simple as well as maximal but the node a,b,d,c is maximal.

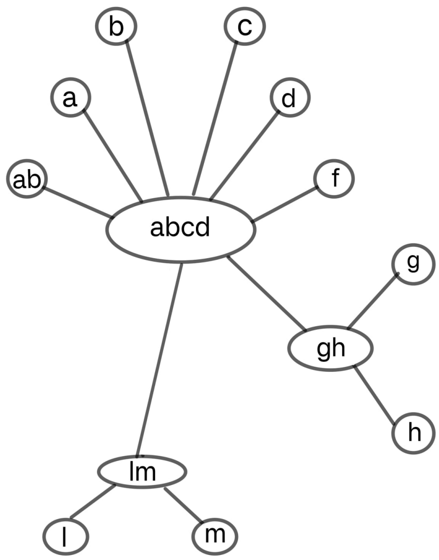

Figure 2.4: virtual representation of figure 2.3

2.2.1 Uniform Cluster Hypergraph

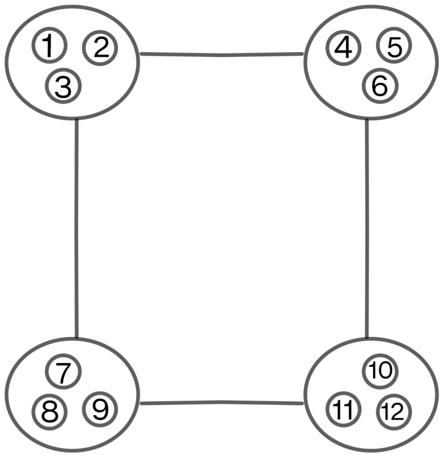

Figure 2.5: (2,3)-uniform cluster hypergraph

The cluster hypergraph is called (m,n) uniform cluster hypergraph if each edge contains m maximal nodes and there will be n simple nodes in each maximal node.

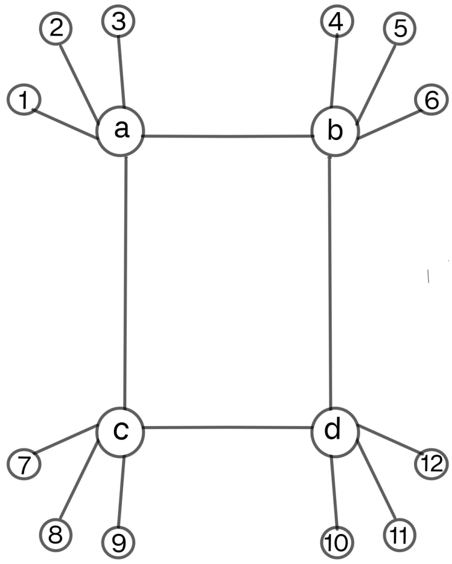

Figure 2.6: virtual representation of figure 2.5

2.2.2 Cluster Connected Cluster Hypergraph

A cluster hypergraph is said to be a cluster-connected cluster hypergraph (CCCH) if only the maximal nodes are connected by an edge. For an example consider the figure 2.3 which is a cluster connected cluster hypergraph. Where each maximal nodes are connected to its internal nodes.

2.3 Completeness Property Of Cluster Hypergraph

Let X be a non-empty set containing n elements. A cluster hypergraph $G(V, E)$ on X contains maximum $|P(X) - \emptyset|$ number of vertices, i.e., $|V| = 2^n - 1$, and for the complete cluster hypergraph, the number of edges is $|P(P(X) - \emptyset) - \emptyset| = 2^{2^n} - 1 - 1$. The completeness properties of different types of cluster graphs are discussed as follows.

2.3.1 Complete CCCH

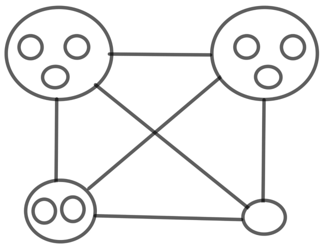

Figure 2.7: Complete CCCH

A cluster hypergraph is called complete CCCH if any two maximal nodes are connected by an edge. The above example shows a complete CCCH

2.3.2 Complete Uniform Cluster Hypergraph

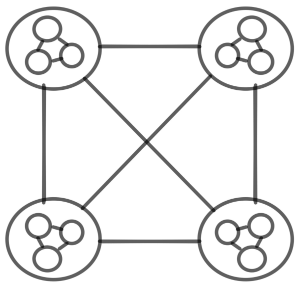

Figure 2.8: Complete uniform Cluster Hypergraph

A complete uniform cluster hypergraph is a (m,n) uniform cluster hypergraph if any two edges are connected by an edge. An edge also connect any two simple nodes within the maximal nodes.The above figure shows a complete (2,3) uniform cluter hypergraph.

Theorem

A complete (m, n) -uniform cluster hypergraph having x cluster nodes contains $x \times \binom{n}{m} + \binom{x}{m}$ edges.

Proof Let us consider a complete (m, n)-uniform cluster hypergraph having x cluster nodes. The hypergraph has x cluster nodes containing n nodes per cluster. Thus, the total number of simple nodes is $x \times n$. Also, the graph is complete. Therefore, each edge contains exactly m nodes. Hence, the total number of edges per cluster is $\binom{n}{m}$. Also, the total number of edges among x clusters is $\binom{x}{m}$. Thus, the total number of edges in a complete (m, n)-uniform cluster hypergraph having x cluster nodes contains $x \times \binom{n}{m} + \binom{x}{m}$ edges.

Chapter 3

COMPETITION CLUSTER HYPERGRAPH

3.1 Competition Cluster Hypergraph

Competition cluster hypergraphs of semidirected graphs are characterized by the adjacency of vertices connected by undirected edges forming a cluster. When these adjacent vertices share common out-directed neighbors, they are also considered adjacent in competition cluster hypergraphs. The formal definition is provided below.

3.1.1 Definition

Let $G = (X, E_1, \vec{E}_2)$ be a semi-directed graph where X is a nonempty vertex set, E_1 is the set of undirected edges, and $\vec{E}_2$ is the set of directed edges. Now, the competition cluster hypergraph of G is denoted as $C(G) = (V_x, E)$ where $V_x \subseteq \mathcal{P}(X)$ is the vertex set of $C(G)$ such that $X \subseteq V_x$, $\{x_i, x_j\} \in V_x$ if $(x_i, x_j) \in E_1$, and $\{x_1, x_2, \ldots, x_m\} \in V_x$ if $\{x_1, x_2, \ldots, x_m\}$ forms a maximal clique in G and E which is the hyperedge set if there exists an edge containing vertices $x_1, x_2, \ldots, x_m$ and $N^+(x_1) \cap N^+(x_2) \cap \ldots \cap N^+(x_m) \neq \emptyset$ and $m = 2, \ldots, |X|$.

Example

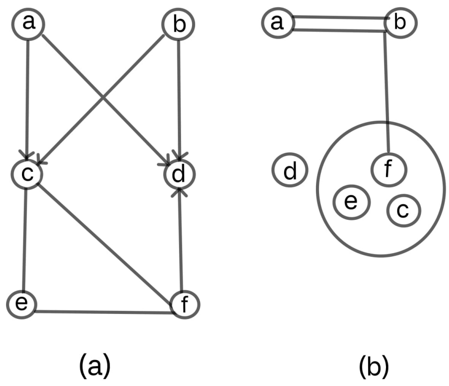

Figure 3.1: semidirected graph and the corresponding competition cluster hypergraph

Consider a semidirected graph depicted in Figure 3.1(a). The associated competition cluster hypergraph is illustrated in Figure(b). In this semidirected graph, let $X = \{a, b, c, d, e, f\}$, and note that $\{e, f, c\}$ forms a clique. Consequently, the vertex set of the corresponding competition hypergraph is denoted as $V_x = X \cup \{e, f, c\}$. Furthermore, observe that $N^+(a) \cap N^+(b) = \{a\}$ and $N^+(a) \cap N^+(b) \cap N^+(f) = \{d\}$. Thus, the set of edges in $C(G)$ is represented as $E = \{ab, abf\}$.

3.1.2 Proposition

Let $G = (X, E_1, \vec{E}_2)$ be a semi-directed graph, and let the corresponding competition cluster hypergraph of G be denoted as $C(G) = (V_X, E)$. The number of edges in $C(G)$ is equal to the number of a maximal in-degree set of vertices in G with cardinality greater than one.

Proof

Let $G = (X, E_1, \vec{E}_2)$ represent a semidirected graph with X as a non-empty set of vertices. Denote the corresponding competition cluster hypergraph of G as $C(G) = (V_X, E)$. In competition cluster hypergraphs, an edge is formed between two vertices x and y if they share at least one common vertex in the semi-directed graph. Moreover, if a third vertex, denoted as z, shares the same common vertex, then the edge will encompass all three vertices x, y, and z, and so forth. Hence, the number of edges in $C(G)$ equals the count of maximal in-degree sets of vertices in G with a cardinality exceeding one.

3.1.3 Remark

Consider a semi directed graph $G = (X, E_1, \vec{E}_2)$ and the corresponding competition cluster hypergraph of G be denoted as $C(G) = (V_x, E)$. The number of cliques formed by undirected edges in G is equal to the number of cluster nodes in $C(G)$.

3.1.4 Theorem

Let $G = (X, E_1, \vec{E}_2)$ be a semidirected graph and the corresponding competition cluster hypergraph of G be C(G) = (Vx, E). The number of maximal nodes in C(G) is equal to the sum of nodes that are not adjacent to other vertices by undirected edges in G + (the number of undirected edges that are not part of any cliques in G) + (number of maximal cliques in G)

Proof

Let $G = (X, E_1, \vec{E}_2)$ be a semi-directed graph where X is a nonempty vertex set and let the corresponding competition cluster hypergraph of G is denoted as $C(G) = (V_x, E)$.

Case 1: $E_1 = \emptyset$: If the semi-directed graph G has no undirected edges, then each node in C(G) is a simple node. Hence, the statement is obvious.

Case 2: $E_1 \neq \emptyset$: In this case, G contains undirected edges. These edges may construct maximal cliques or simple undirected edges. Each maximal clique in G will correspond to one cluster node in C(G). The undirected edges that are not part of any cluster also correspond to cluster nodes containing two simple nodes. Hence, the number of maximal nodes in C(G) is equal to the number of nodes that are not adjacent to other vertices by undirected edges in G, + the number of undirected edges that are not part of any cliques in G, + the number of maximal cliques in G

3.2 m-step Competition Cluster Hypergraph

The m-step competition cluster hypergraph of G denoted as $C_m(G) = (V_x, E)$, is defined as follows. Let $G = (X, E_1, \vec{E}_2)$ be a semi-directed graph, where X is a nonempty vertex set, E_1 is the set of undirected edges, and $\vec{E}_2$ is the set of directed edges. The vertex set of $C_m(G)$, denoted as V_X, is a subset of the power set of X, such that $X \subseteq V_X$. An element $\{x_i, x_j\} \in V_X$ if $(x_i, x_j) \in E_1$, and $\{x_1, x_2, \ldots, x_m\} \in V_X$ if $\{x_1, x_2, \ldots, x_m\}$ forms a maximal clique in G. The hyperedge set E consists of hyperedges containing vertices $x_1, x_2, \ldots, x_m$ if $N_m^+(x_1) \cap N_m^+(x_2) \cap \ldots \cap N_m^+(x_m) \neq \emptyset$, where $m = 2, \ldots, |X|$.

Example

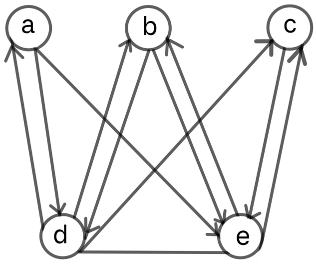

Figure 3.2

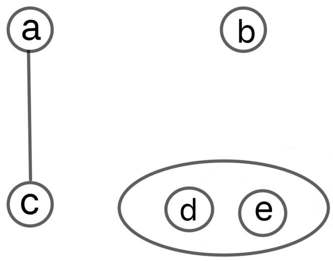

Figure 3.3: 2-step competition cluster hypergraph

Consider a semi-directed graph, as shown in Figure 3.2, and the corresponding 2-step competition cluster hypergraph $C_m(G)$ is shown in Figure 3.3. In the semi-directed graph, $X = \{a, b, c, d, e\}$, and also the undirected edge joining d and e $\in E_1$. So, the vertex set of the corresponding 2-step competition hypergraph is $V_x = X \cup \{d, e\}$. Now, $N_2^+(a) \cap N_2^+(c) = \{b\}$. Thus, the edge set of $C_m(G)$ is $E = \{ac\}$.

3.3 Isolated Node

A node is called an isolated node in a cluster hypergraph if there don't exist edges to that node from other nodes in the graph.

3.3.1 Isolated Maximal Node

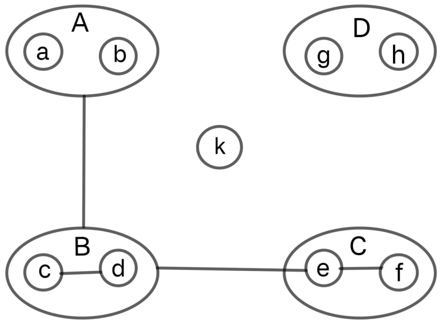

Figure 3.4: Cluster hypergraph with two isolated nodes

A maximal node (simple or cluster) becomes a maximal isolated node when it has no incident edges, i.e. it is entirely disconnected from all other maximal nodes within the cluster hypergraph. In Figure 3.4, we can identify that the singleton set {k} and the clusters D = {{g} {h}} are examples of such isolated maximal nodes.

3.3.2 Isolated Node In A Cluster

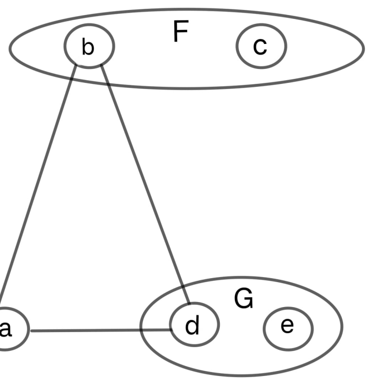

Figure 3.5: cluster hypergraph with one isolated node

Within a cluster hypergraph, a simple node might exist in isolation within a cluster node. Consider a cluster hypergraph as shown in Figure 3.5, it becomes evident that the only node $\{c\}$ stands as an isolated entity within the cluster $F = \{\{c\}, \{b\}\}$.

3.4 Competition Number

Consider a cluster hypergraph G. Then, the competition number k of G is the minimum number of k maximal isolated nodes with G which form the competition graph of a semidirected graph.

3.4.1 Algorithm

The steps to find the competition number of a cluster hypergraph are given as follows:

Step 1. Consider a cluster hypergraph G

Step 2: Construct the necessary directed or undirected edges to create a corresponding semi-directed graph G' from the original hypergraph G.

Step 3: If additional nodes are required to achieve C(G') = G, consider these nodes as isolated nodes.

Step 4. The minimum number of maximal isolated nodes will be the competition number of G

3.4.2 Example

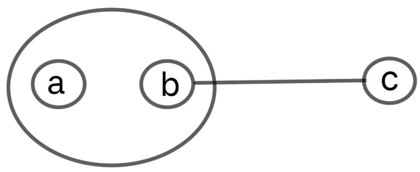

Figure 3.6: A cluster hypergraph G

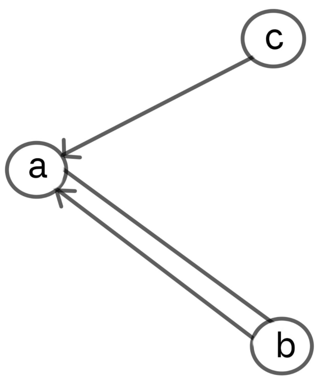

Figure 3.7: Corresponding semi-directed graph of G

let cluster hypergraph $G = (V_x, E)$ (figure:3.6), where $V_x = \{\{a\}, \{b\}, \{a, b\}, \{c\}\}$ and $E = \{\{\{c\}, \{b\}\}\}$, we can construct the corresponding semi-directed graph (figure:3.7). Notably, there are no isolated nodes in this graph. Consequently, the competition number of G is 0.

3.4.3 Proposition

The competition number of a (2,2)-uniform cluster hypergraph is zero.

Proof

Consider a (2,2)-uniform cluster hypergraph. We need to prove that its competition number is zero.
since each hyperedge contains exactly two vertices and each vertex is incident to exactly two edges. let's start drawing the corressponding semi-directed graph.where two vertices are adjacent if and only if they belong to a common hyperedge. And also each hyperedge has atmost one vertex common. Therefore, no isolated vertex in the graph. which implies competition number is equal to zero.
This complete the proof

3.5 Proposition

Let G be a semi-directed graph and G_m be the m-step semi-directed graph of G, then $C(G_m) = C_m(G)$.

Proof

Given a semi-directed graph G and its m-step semidirected graph G_m, we note that both graphs share the same vertex set. Suppose (u, v) belongs to the competition cluster of G_m. Then, there exist edges $(u, \overset{\rightarrow}{x_1}), (v, \overset{\rightarrow}{x_1}); (u, \overset{\rightarrow}{x_2}), (v, \overset{\rightarrow}{x_2}); \ldots, (u, \overset{\rightarrow}{x_n}), (v, \overset{\rightarrow}{x_n})$ for some integer n. Now consider the intersection of the out-neighborhoods of u and v, denoted as $N^+(u)N^+(v)$, which consists of nodes $x_1, x_2, \ldots, x_n$. Since an edge (u, x_1) in G_m implies the existence of a path of length m from u to $\mathrm{x}_1, and similarly, for (\mathrm{v, x1}) in \mathrm{G}_m$, we conclude that (u, v) belongs to the $C_m(G)$. Conversely, if an edge (x, y) is in $C_m(G)$, then $C(G_m) = C_m(G)$.

Chapter 4

APPLICATION

I am working on two applications: one involves the use of cluster hypergraphs, and the other involves the application of competition cluster hypergraphs.

4.1 Cluster Hypergraph

I demonstrate the application of cluster hypergraphs to a real-life problem by examining five types of cancer and their top 10 mutated genes. The collected data is presented in Table 1. In this representation, genes are depicted as nodes or vertices, cancers are depicted as clusters of nodes, and tumor suppressor genes are depicted as hyperedges.

CANCERS	GENES
OVARIAN CANCER	PIK3CA, CTNNB1, OPCML, BRCA1, BRCA2, TP53, PTEN, BRAF,ATM, EGFR
PANCREATIC CANCER	SMAD4, TP53, KRAS, BRCA1, BRCA2, ATM, CHEK2, STK11, TSC1, PTEN
PROSTATE CANCER	PTEN, CHEK2, CDH1, BRCA2, KLF6, ZFHX3, MAD1L1, HOXB13, ATM, TP53
BREAST CANCER	BRCA2, RAD51, PIK3CA, ATM, BARD1, TP53, RAD54L, CHEK2, KRAS, BRCA1
BLADDER CANCER	FGFR3, RB1, ATM, KRAS, TP53, EGFR, PTEN, BRCA1, PIK3CA, CTNNB1

Table 4.1: TABLE 1

The tumor suppressor genes among the 25 genes in our data include BRCA2, TP53, BRCA1, BRCA2, PIK3CA, PTEN, and CDH1.

The cluster hypergraph obtained from the data above is as follows:

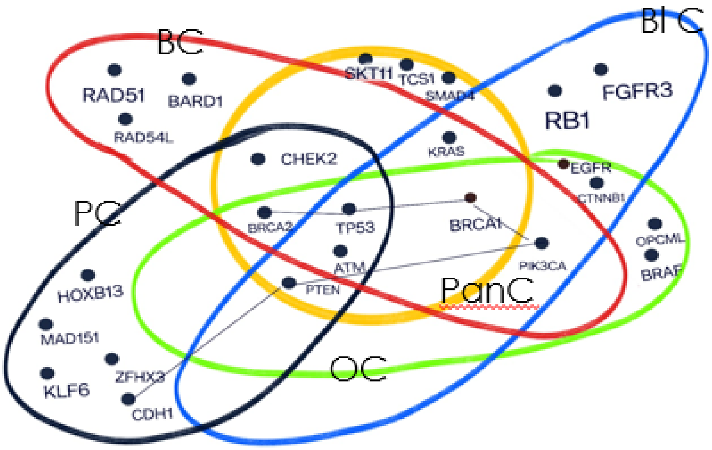

Figure 4.1: Cluster hypergraph of our data

Now degree and effective degree are calculated for the above clustering hypergraph. The degree and effective degree calculation are shown in the following table.

GENES	DEGREE	EFFECTIVE DEGREE
KRAS	3	$3+(30)/3 = 13$
PIK3CA	4	$4+(30)/3 = 14$
TP53	6	$6+(50)/5 = 16$
EGFR	2	$2+(20)/2 = 12$
ATM	5	$5+(50)/5 = 15$
BRCA1	5	$5+(50)/5 = 15$
BRCA2	5	$5+(50)/5 = 15$
CHEK2	3	$3+(30)/3 = 13$
CTNNB1	2	$2+(20)/2 = 12$
PTEN	5	$5+(40)/4 = 15$
OPCML	1	$1+(10)/1 = 10$
BRAF	1	$1+(10)/1 = 10$
RAD51	1	$1+(10)/1 = 10$
BARD1	1	$1+(10)/1 = 10$
RAD54L	1	$1+(10)/1 = 10$
CDH1	1	$1+(10)/1 = 10$
KLF6	1	$1+(10)/1 = 10$
ZFHX3	1	$1+(10)/1 = 10$
MAD1L1	1	$1+(10)/1 = 10$
HOXB13	1	$1+(10)/1 = 10$
FGFR3	1	$1+(10)/1 = 10$
RB1	1	$1+(10)/1 = 10$
SMAD4	1	$1+(10)/1 = 10$
STK11	1	$1+(10)/1 = 10$
TSC1	1	$1+(10)/1 = 10$

Table 4.2: TABLE 2

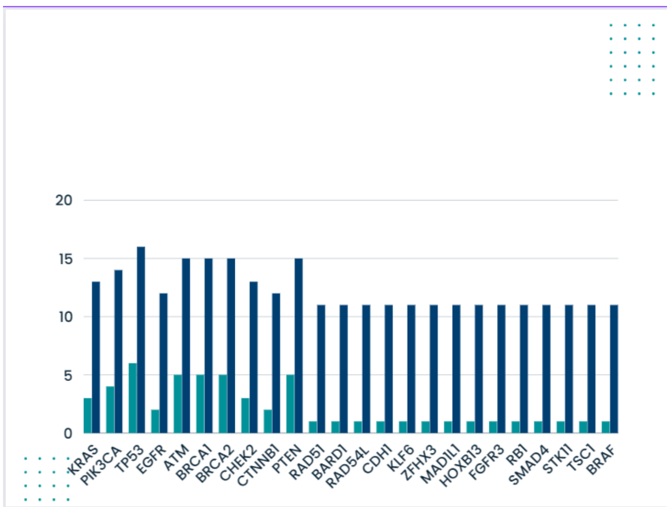

Figure 4.2: COMPARISON BETWEEN DEGREE AND EFFECTIVE DEGREE

By observing the comparison between the degree and effective degree, it is evident that the gene with the highest degree also possesses the highest effective degree. This can be attributed to our consideration of only the top ten genes for each cancer, resulting in each cluster containing ten genes. moreover, our attention was also directed toward tumor suppressor genes among the datasets of genes. Notably, TP53 emerges as the gene with the highest effective degree. Consequently, we identify TP53 as the most important gene among the five cancers when focusing on their top 10 mutated genes.

4.2 Competition Cluster Hypergraph

The network is designed to represent COVID-19-affected areas and their competition. Affected places are nodes in a semidirected graph, with "COVID-19" as the source node. Additionally, major carbon emission countries are included in this network, with a source node labeled "carbon emission". Undirected edges connect the countries that are affected by the same country, and directed edges are from both source nodes to all other nodes in the network.

COUNTRY	TOTAL CASES	AFFECTED FROM
USA	98,525,870	China
India	44,676,087	China
France	37,716,837	China
Germany	36980,883	
Brazil	35,751,411	

Table 4.3: Top 5 COVID-19 affected countries

COUNTRY	$CO_2 Emission$
China	12667.43
US	4853.78
India	2693.03
Russia	1909.04
Japan	1082.65
Indonesia	692.24
Iran	686.42
Germany	673.60
South Korea	635.50
Saudi Arabia	607.91
Canada	582.07
Mexico	487.77
Turkey	481.25
Brazil	466.77
South Africa	404.97
Australia	393.16
UK	340.61
Vietnam	327.91
Italy	322.95
France	315.30

Table 4.4: Top 10 countries of carbon emissions

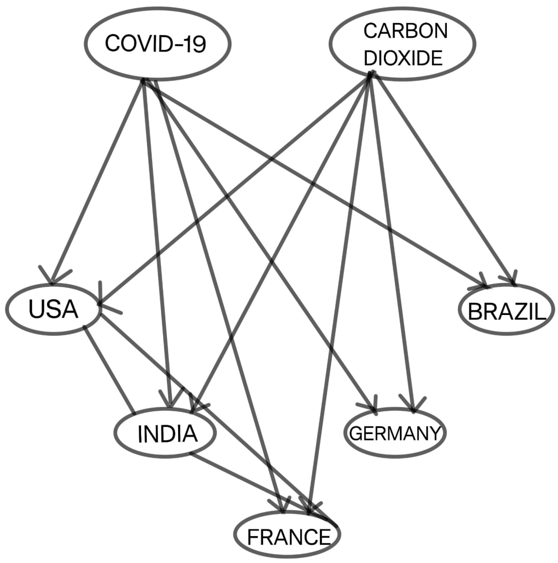

Figure 4.3: A semidirected graph

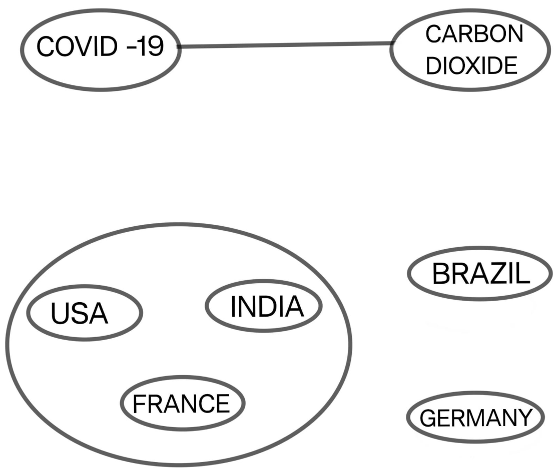

Figure 4.4: Competition cluster hypergraph of figure

From tables 4.3 and 4.4, the corresponding semi-directed is shown in the figure 4.3, and its corresponding competition cluster hypergraph is shown in the figure 4.4

The step-by-step process to find out the competition cluster hypergraph is given as follows:

Step 1. First construct a Semidirected Graph from the Six highest COVID-19-affected countries which are assumed as nodes along with two fictitious nodes COVID-19 and CO_2 emissions. All the assumed countries are affected by COVID-19 and CO_2 emissions. Thus, there will be direct edges from COVID-19 and CO_2 emissions to all the nodes, undirected edges connect the countries that are affected by the same country. (see Table 4.3 and Figure 4.4).

Step 2. In the resultant competition hypergraphs, cliques will form cluster nodes. USA-India-France is the cluster node for this case (Figure 4.4). Between two nodes, there will be edges if the nodes have common out-neighbourhoods in semidirected graphs. Hence, the nodes COVID-19 and CO_2 emissions will have one edge.

Conclusion

This project explores the topic cluster hypergraph. For the application, I focus on analyzing five cancers and their top ten mutated genes, and among these genes, I give more focus on the tumor suppressor genes. Constructed cluster hypergraphs based on these data. From that representation, TP53 emerges as the critical gene in the network. Interestingly observed that the gene that has the highest degree also has the highest effective degree. Similarity arises because each cluster node contains ten nodes. Here I didn't consider all mutated genes of the corresponding cancer if we do so then the network will be on complex. To illustrate the application of competition cluster hypergraph considered two cases COVID-19 and CO_2 emission. and obtain a competition cluster hypergraph from the semi-directed graph drawn from the data provided.

REFERENCES

[1] Goh, K.-I., Cusick, M. E., Valle, D., Childs, B., Vidal, M., Barabási, A.-L. (2007).*The Human Disease Network.* "Proceedings of the National Academy of Sciences of the United States of America".

[2] Aittokallio, T., Schwikowski, B. (2006). *Graph-based methods for analyzing networks in cell biology.*

[3] Maity, A., Das, K., Samanta, S., Mondal, S., Dubey, V. (2021). *A study of cluster hypergraphs and its properties.*

[4] Samanta, S., Muhiuddin, G., Alanazi, A. M., Das, K. (2020).*A Mathematical Approach on Representation of Competitions: Competition Cluster Hypergraphs.*

[5] Estrada, E., Rodríguez-Velázquez, J. A. (2020).*Complex Networks as Hypergraphs. Mathematical Problems in Engineering.*

[6] Chitra, U. (2017).*Random Walks on Hypergraphs with Applications to Disease-Gene Prioritization.*

[7] Busseniers, E. (2014).*General Centrality in a Hypergraph.*