Jagua Publication

Current Research in Interdisciplinary Studies, 4(3):1–11, 2025. DOI: 10.58614/cris431

Current Research in Interdisciplinary Studies

Journal homepage: https://www.jpub.org/journal-details.php?journal-id=40

ISSN (E): 2583-1887
Current Research in Interdisciplinary Studies
Jagua Publication

Research Article

Similarity HyperGraph and Similarity SuperHyperGraph in Medical Science

Takaaki Fujita[1*]

[1] *Independent Researcher, Shinjuku, Shinjuku-ku, Tokyo, Japan.*
[*] *Corresponding author: Takaaki.fujita060@gmail.com*

Article Info

Keywords: *Superhypergraph, Hypergraph, Similarity SuperHyperGraph, Similarity HyperGraph, Similarity Graph*

Received: *28.07.2025*
Accepted: *19.08.2025*
Published: *25.08.2025*

Abstract

A similarity graph represents items as vertices, with edges connecting pairs whose similarity exceeds a specified threshold or meets k-nearest-neighbor criteria. Applications of similarity graphs have been explored in domains such as medical science. A *hypergraph* extends this notion by allowing each *hyperedge* to join any nonempty subset of vertices simultaneously [1–4]. A *SuperHyperGraph* further introduces a hierarchy by iterating the powerset construction, thereby capturing nested, multi-scale relationships among vertices and edges. In this paper, we investigate *Similarity HyperGraphs* and *Similarity SuperHyperGraphs*. These extensions make it possible to study similarity graphs in a more deeply hierarchical context.

1. Introduction

We begin by fixing notation and recalling foundational definitions that will be used throughout this paper. Unless otherwise specified, all graphs are assumed to be finite. For more extensive treatments, see the referenced works.

1.1. Hypergraphs and SuperHyperGraphs

Graph theory is the mathematical study of networks of vertices and edges representing relationships or connections [5–7]. A *hypergraph* extends the notion of a graph by permitting each *hyperedge* to join any nonempty subset of vertices at once [1, 3, 4, 8, 9]. Hypergraphs have been applied in a wide range of domains, and various mathematical properties and graph algorithms have been developed to analyze them [9–12].

A *SuperHyperGraph* further builds a hierarchy by iterating the powerset construction, thus capturing nested, multi-scale relationships among vertices and edges [13–18].

Definition 1.1 (Base set)**.** *Let V_0 be a finite set, called the* base set*. All subsequent vertex and edge collections are drawn from V_0 or its iterated powersets.*

Definition 1.2 (Powerset)**.** *For any set X, its* powerset *is*

$$\mathscr{P}(X) = \{A : A \subseteq X\}.$$

How To Cite: T. Fujita. Similarity HyperGraph and Similarity SuperHyperGraph in Medical Science. *Current Research in Interdisciplinary Studies*, 4(3):1–11, 2025. https://doi.org/10.58614/cris431

Definition 1.3 (Hypergraph)**.** *[1, 3] A* hypergraph *is a pair* $H = (V, E)$ *where*

- V *is a finite set of* vertices*, and*
- $E \subseteq \mathscr{P}(V) \setminus \{\emptyset\}$ *is a finite collection of nonempty subsets of* V*, called* hyperedges*.*

Definition 1.4 (Iterated powerset)**.** *[19–21] Define recursively for* $k \geq 0$*:*

$$\mathscr{P}^0(V_0) = V_0, \quad \mathscr{P}^{k+1}(V_0) = \mathscr{P}\big(\mathscr{P}^k(V_0)\big).$$

We write $\mathscr{P}_n(V_0)$ *for* $\mathscr{P}^n(V_0)$ *and denote by* $\mathscr{P}_n^*(V_0)$ *its collection of nonempty subsets.*

Example 1.5 (Iterated Powerset in Pharmaceutical Regimen Design)**.** *We illustrate the iterated powerset construction (Definition 2.2) in the context of combination drug therapies.*

Level 0 (Base set of drugs).

$$V_0 = \{\mathrm{A}, \mathrm{B}, \mathrm{C}\},$$

where A*,* B*, and* C *are three distinct drug compounds.*

Level 1 (Single- and multi-drug formulations).

$$\mathscr{P}^1(V_0) = \mathscr{P}(V_0) = \big\{\{A\},\{B\},\{C\},\{A,B\},\{A,C\},\{B,C\},\{A,B,C\}\big\}.$$

Interpretation:

- $\{A\},\{B\},\{C\}$ *are monotherapies.*
- $\{A,B\},\{A,C\},\{B,C\}$ *are two-drug combinations.*
- $\{A,B,C\}$ *is a three-drug cocktail.*

Select three representative formulations:

$$F_1 = \{A,B\}, \quad F_2 = \{B,C\}, \quad F_3 = \{A,C\}.$$

Level 2 (Two-phase regimens).

$$\mathscr{P}^2(V_0) = \mathscr{P}\big(\mathscr{P}^1(V_0)\big).$$

Choose three clinically relevant two-phase regimens, each a set of two formulations:

$$R_1 = \{F_1, F_2\}, \quad R_2 = \{F_2, F_3\}, \quad R_3 = \{F_1, F_3\}.$$

Here R_1 *means administer formulation* F_1 *in phase 1 and* F_2 *in phase 2, and so on.*

Level 3 (Multi-cycle treatment programs).

$$\mathscr{P}^3(V_0) = \mathscr{P}\big(\mathscr{P}^2(V_0)\big).$$

Select two full treatment programs, each a set of two regimens:

$$P_1 = \{R_1, R_2\}, \qquad P_2 = \{R_2, R_3\}.$$

Here P_1 *represents a bi-cycle protocol using* R_1 *in cycle 1 and* R_2 *in cycle 2.*

Thus the iterated powersets $\mathscr{P}^0(V_0)$*,* $\mathscr{P}^1(V_0)$*,* $\mathscr{P}^2(V_0)$*,* $\mathscr{P}^3(V_0)$ *model, respectively, individual drugs, formulations, regimens, and multi-cycle programs in a hierarchical fashion.*

Definition 1.6 (n–SuperHyperGraph)**.** *[14, 22, 23] An* n–SuperHyperGraph *is a pair*

$$\mathrm{SuHyG}^{(n)} = \big(V, E\big), \quad V, E \subseteq \mathscr{P}^n(V_0),$$

where each element of V *is called an* n–supervertex *and each element of* E *an* n–superedge*.*

Example 1.7 (Polytherapy Regimens as a 2-SuperHyperGraph)**.** *We model combination drug therapies in a chronic disease setting as a 2-SuperHyperGraph.*

Base set of drugs. *Let the base set of active pharmaceutical ingredients be*

$$V_0 = \{\mathrm{A}, \mathrm{B}, \mathrm{C}\},$$

where A $=$ *Drug A,* B $=$ *Drug B, and* C $=$ *Drug C.*

Level-1 combinations (supervertices). *First-order combinations are subsets of* V_0*. We select three clinically relevant two-drug combinations:*

$$R_1 = \{\mathrm{A},\mathrm{B}\}, \quad R_2 = \{\mathrm{B},\mathrm{C}\}, \quad R_3 = \{\mathrm{A},\mathrm{C}\}.$$

These form the level-1 supervertices:

$$V = \{R_1, R_2, R_3\} \subseteq \mathscr{P}^1(V_0).$$

Level-2 regimens (superedges). *Regimens are pairs of combination therapies chosen for multi-phase treatment. We define two representative regimens:*

$$e_{12} = \{R_1, R_2\}, \quad e_{23} = \{R_2, R_3\}.$$

Thus the set of 2-superedges is

$$E \;=\; \{e_{12}, e_{23}\} \;\subseteq\; \mathscr{P}(V).$$

Interpretation. *– V_0 is the pool of available drugs.*
– Each R_i is a two-drug combination tested in clinical trials.
– Each e_{ij} is a sequential treatment regimen, administering combination R_i followed by R_j.

Summary. *Hence*

$$\mathrm{SuHyG}^{(2)} = \big(V, E\big)$$

$$= \Big(\{R_1,R_2,R_3\}, \{\{R_1,R_2\},\{R_2,R_3\}\}\Big)$$

is a 2-SuperHyperGraph capturing both combination therapies (supervertices) and phased regimens (superedges) in a pharmaceutical context.

1.2. Similarity Graph

A similarity graph represents items as vertices, with edges between pairs whose similarity exceeds a threshold or satisfies k-NN constraints (cf.[24–31]).

Definition 1.8 (Similarity Graph)**.** *Let X be a finite set of objects and let*

$$\mathrm{sim}\colon X \times X \;\longrightarrow\; \mathbb{R}$$

be a (not necessarily symmetric) similarity function. We present two standard constructions of a similarity graph *on X.*

(1) Thresholded Similarity Graph. *Fix a threshold $\tau \in \mathbb{R}$. Define the directed, weighted graph*

$$G_\tau = \big(V, E_\tau, w\big)$$

by

$$V = X, \quad E_\tau = \big\{(u,v) \in X \times X : u \neq v,\ \mathrm{sim}(u,v) \geq \tau\big\},$$

$$w(u,v) = \mathrm{sim}(u,v).$$

If sim *is symmetric, one may regard G_τ as an undirected graph by identifying each (u,v) with the edge $\{u,v\}$.*

(2) k-Nearest-Neighbor Similarity Graph. *Fix an integer $1 \leq k < |X|$. For each $u \in X$, let*

$$N_k(u) \subseteq X \setminus \{u\}$$

be the set of k elements v maximizing $\mathrm{sim}(u,v)$ *(breaking ties arbitrarily). Then define*

$$G^{(k)} = \big(V, E^{(k)}, w\big), \quad E^{(k)} = \big\{(u,v) : v \in N_k(u)\big\}, \quad w(u,v) = \mathrm{sim}(u,v).$$

Again, symmetry of sim *allows one to treat $G^{(k)}$ as undirected by converting each ordered pair into an unordered edge.*

In both cases, the resulting graph encodes pairwise affinities among the elements of X, with edge-weights given by the original similarity scores.

Example 1.9 (Patient Similarity Graph in a Breast Cancer Cohort)**.** *Let $X = \{P_1, P_2, \ldots, P_{100}\}$ be a cohort of 100 breast cancer patients. For each patient P_i, we measure a gene-expression profile*

$$x_i = \big(x_i^1, x_i^2, \ldots, x_i^{500}\big) \;\in\; \mathbb{R}^{500},$$

where we have preselected the 500 most variable genes across the cohort.
We define the similarity function by the Pearson correlation of mean-centered profiles:

$$\mathrm{sim}(P_i, P_j) = \frac{\sum_{g=1}^{500} \big(x_i^g - \bar{x}_i\big)\big(x_j^g - \bar{x}_j\big)}{\sqrt{\sum_{g=1}^{500} (x_i^g - \bar{x}_i)^2}\,\sqrt{\sum_{g=1}^{500} (x_j^g - \bar{x}_j)^2}},$$

where $\bar{x}_i = \frac{1}{500}\sum_{g=1}^{500} x_i^g$.

Thresholded Similarity Graph. *Choose a high-stringency threshold $\tau = 0.85$. The resulting graph*

$$G_{0.85} = (V, E_{0.85}, w)$$

is given by

$$V = X, \quad E_{0.85} = \big\{(P_i,P_j) \mid i \neq j,\ \mathrm{sim}(P_i,P_j) \geq 0.85\big\}, \quad w(P_i,P_j) = \mathrm{sim}(P_i,P_j).$$

In this graph, edges connect only those patient-pairs whose expression correlation exceeds 0.85. Analysis of $G_{0.85}$ often reveals three major connected components, corresponding to the known molecular subtypes (Luminal A, Luminal B, Basal-like) in breast cancer.

***k*-Nearest-Neighbor Similarity Graph.** *Alternatively, fix $k = 5$. For each P_i, let $N_5(P_i)$ be its five most highly correlated neighbors:*

$$N_5(P_i) = \underset{P_j \neq P_i}{\arg\operatorname{top}5} \big\{\mathrm{sim}(P_i,P_j)\big\}.$$

Then the 5-NN graph

$$G^{(5)} = (V, E^{(5)}, w)$$

is defined by

$$E^{(5)} = \{(P_i,P_j) \mid P_j \in N_5(P_i)\}, \quad w(P_i,P_j) = \mathrm{sim}(P_i,P_j).$$

This construction ensures each patient node has out-degree exactly 5, capturing its strongest transcriptomic affinities. Symmetrizing $G^{(5)}$ by retaining $\{P_i,P_j\}$ whenever either (P_i,P_j) or (P_j,P_i) appears yields an undirected network suitable for community detection of patient subgroups.

Example 1.10 (Drug–Drug Similarity Graph Based on Target-Profile Overlap)**.** *Consider the set of eight antihypertensive drugs*

$$D = \{\mathit{Lisinopril}, \mathit{Enalapril}, \mathit{Captopril}, \mathit{Losartan}, \mathit{Valsartan}, \mathit{Metoprolol}, \mathit{Atenolol}, \mathit{Amlodipine}\}.$$

For each $d \in D$ let

$$T(d) \subseteq \{\mathit{ACE}, \mathit{AGTR1}, \beta_1, \text{L-type } Ca^{2+}\}$$

be its known protein-target set:

$$\begin{aligned} T(\mathit{Lisinopril}) = T(\mathit{Enalapril}) = T(\mathit{Captopril}) &= \{\mathit{ACE}\}, \\ T(\mathit{Losartan}) = T(\mathit{Valsartan}) &= \{\mathit{AGTR1}\}, \\ T(\mathit{Metoprolol}) = T(\mathit{Atenolol}) &= \{\beta_1\}, \\ T(\mathit{Amlodipine}) &= \{\text{L-type } Ca^{2+}\}. \end{aligned}$$

Define the Jaccard similarity

$$\mathrm{sim}(d_i,d_j) = \frac{\big|T(d_i) \cap T(d_j)\big|}{\big|T(d_i) \cup T(d_j)\big|}.$$

Threshold-based similarity graph. *With threshold $\tau = 0.5$, we form*

$$G_\tau = \big(V, E_\tau, w\big),$$

where

$$V = D, \quad E_\tau = \big\{\,\{d_i,d_j\} \subseteq D : \mathrm{sim}(d_i,d_j) \geq 0.5\big\}, \quad w(\{d_i,d_j\}) = \mathrm{sim}(d_i,d_j).$$

In $G_{0.5}$, the three ACE inhibitors form a 3-clique (similarity = 1), the two ARBs form a 2-clique, the two β-blockers form a 2-clique, and Amlodipine is isolated. This graph recovers the known pharmacological classes purely from target overlap.

***k*–Nearest-Neighbor similarity graph.** *Alternatively, for $k = 2$ define for each $d \in D$*

$$N_2(d) = \underset{d' \neq d}{\arg\operatorname{top}2} \{\mathrm{sim}(d,d')\}.$$

Then the directed 2-NN graph

$$G^{(2)} = \big(V, E^{(2)}, w\big), \quad E^{(2)} = \big\{(d,d') : d' \in N_2(d)\big\},$$

ensures each drug has out-degree 2. Symmetrizing by replacing each arc (d,d') with the undirected edge $\{d,d'\}$ whenever either direction occurs yields an undirected network that again clusters drugs by mechanism of action.

2. Result: Similarity HyperGraph

A similarity hypergraph is a hypergraph whose hyperedges consist of all subsets of items whose pairwise similarity exceeds a given threshold, thereby revealing high-affinity clusters.

Definition 2.1 (Similarity Hypergraph)**.** *Let X be a finite set of objects and*

$$\mathrm{sim}\colon X \times X \longrightarrow R$$

a symmetric similarity measure. Fix a threshold $\tau \in R$. The thresholded similarity hypergraph *is the pair*

$$H_\tau = \big(X,\, E_\tau\big),$$

where

$$E_\tau = \big\{\, S \subseteq X : |S| \geq 2,\ \forall u, v \in S,\ u \neq v \implies \mathrm{sim}(u,v) \geq \tau \big\}.$$

Each $S \in E_\tau$ is a hyperedge *consisting of all mutually similar items above the threshold.*

Example 2.2 (Somatic Mutation Similarity Hypergraph in a Colorectal Cancer Cohort)**.** *Let $X = \{P_1, P_2, \ldots, P_{50}\}$ be a cohort of 50 colorectal cancer patients. For each patient P_i, we record the set of somatic mutations*

$$M_i \subseteq G,$$

where $G = \{g_1, g_2, \ldots, g_{200}\}$ is the list of the 200 most frequently mutated genes in colorectal cancer. We define the similarity between two patients by the Jaccard index on their mutation-sets:

$$\mathrm{sim}(P_i, P_j) = \frac{|M_i \cap M_j|}{|M_i \cup M_j|} \in [0,1].$$

Fix a threshold $\tau = 0.5$. Then by Definition 2.1 the thresholded similarity hypergraph is

$$H_{0.5} = \big(X, E_{0.5}\big),$$

$$E_{0.5} = \big\{\, S \subseteq X : |S| \geq 2,\ \forall P_u, P_v \in S,\ u \neq v \implies \mathrm{sim}(P_u, P_v) \geq 0.5 \big\}.$$

Two illustrative hyperedges are:

$$S_1 = \{P_3, P_{17}, P_{29}\}, \quad \textit{with } M_3 \cap M_{17} \cap M_{29} = \{APC, TP53, KRAS\},$$

and pairwise Jaccard indices $\mathrm{sim}(P_3, P_{17}) = 0.62$, $\mathrm{sim}(P_3, P_{29}) = 0.58$, $\mathrm{sim}(P_{17}, P_{29}) = 0.60$.

$$S_2 = \{P_8, P_{19}, P_{24}, P_{45}\}, \quad \textit{with shared mutations } \{SMAD4, PIK3CA\},$$

and all pairwise $\mathrm{sim} \geq 0.50$.
In this hypergraph, each hyperedge S identifies a subgroup of patients whose mutational landscapes are all mutually similar above 0.5. Such high-confidence mutation clusters often correspond to biologically distinct subtypes or shared pathway-level dysregulations in colorectal cancer.

Example 2.3 (Gut Microbiome Similarity Hypergraph in an IBD Cohort)**.** *Let $X = \{P_1, P_2, \ldots, P_{60}\}$ be a cohort of 60 pediatric patients with inflammatory bowel disease (IBD) sampled in the RISK study. For each patient P_i, we profile the gut microbiome by estimating the relative abundances of the top 100 bacterial taxa:*

$$\mathbf{m}_i = \big(m_i^1, \ldots, m_i^{100}\big) \in \Delta^{99},$$

where $\Delta^{99} = \{x \in R^{100}_{\geq 0} : \sum x^g = 1\}$. We define a symmetric similarity measure between two microbiome profiles by the Spearman rank correlation:

$$\mathrm{sim}(P_i, P_j) = \rho_{\mathrm{Spearman}}\big(\mathbf{m}_i, \mathbf{m}_j\big) \in [-1, 1].$$

Fix a high-stringency threshold $\tau = 0.80$. By Definition 2.1, the thresholded similarity hypergraph is

$$H_{0.80} = \big(X, E_{0.80}\big), \quad E_{0.80} = \big\{\, S \subseteq X : |S| \geq 2,\ \forall P_u, P_v \in S,\ u \neq v \implies \mathrm{sim}(P_u, P_v) \geq 0.80 \big\}.$$

Two representative hyperedges are:

- $S_1 = \{P_4, P_{19}, P_{27}, P_{33}\}$. *These patients' microbiomes share a high Firmicutes–Bacteroidetes ratio and strong co-occurrence of* Faecalibacterium *and* Roseburia*. Pairwise Spearman correlations range from 0.82 to 0.87.*
- $S_2 = \{P_8, P_{15}, P_{42}\}$. *Characterized by elevated* Escherichia *and reduced diversity, this subgroup exhibits pairwise correlations 0.81, 0.83, and 0.85.*

In $H_{0.80}$, each hyperedge S captures a cluster of IBD patients whose gut microbial community structures are mutually similar above 0.80. Analysis of these hyperedges can reveal microbiome-defined patient subtypes with distinct clinical trajectories and treatment responses.

Theorem 2.4 (Generalization of the Similarity Graph)**.** *Let $G_\tau = (X, E_\tau^{(2)})$ be the usual thresholded similarity graph with*

$$E_\tau^{(2)} = \big\{\{u,v\} \subseteq X : u \neq v,\ \mathrm{sim}(u,v) \geq \tau\big\}.$$

Then:

1. *$H_\tau = (X, E_\tau)$ is a hypergraph in the sense of Definition 1.3.*
2. *The* 2-section *(or primal graph) of H_τ,*

$$G^{(2)}(H_\tau) = \big(X,\ \{\{u,v\} \subseteq X : \exists S \in E_\tau,\ \{u,v\} \subseteq S\}\big),$$

coincides with G_τ. Hence H_τ strictly generalizes the similarity graph.

Proof. **(1) Hypergraph property.** By construction X is finite and

$$E_\tau \ \subseteq\ \mathscr{P}(X) \setminus \{\emptyset\},$$

and each $S \in E_\tau$ has $|S| \geq 2$. Thus H_τ satisfies Definition 1.3.

(2) Recovery of the similarity graph. By definition of E_τ, any pair $\{u,v\}$ with $\mathrm{sim}(u,v) \geq \tau$ belongs to at least one hyperedge $S = \{u,v\} \in E_\tau$. Conversely, if $\{u,v\} \subseteq S$ for some $S \in E_\tau$, then $\mathrm{sim}(u,v) \geq \tau$. Hence the edge-set of the 2-section of H_τ is exactly $\{\{u,v\} \mid \mathrm{sim}(u,v) \geq \tau\}$, which is $E_\tau^{(2)}$. This shows $G^{(2)}(H_\tau) = G_\tau$, proving that H_τ generalizes the thresholded similarity graph. □

3. Result: Similarity SuperHyperGraph

A similarity superhypergraph generalizes this concept to iterated powerset vertices, with superedges grouping nested subsets whose flattened elements all exceed the similarity threshold, capturing multi-level affinity structure.

Definition 3.1 (Similarity n–SuperHyperGraph)**.** *Let X be a finite set and*

$$\mathrm{sim}\colon X \times X \ \longrightarrow\ R$$

a symmetric similarity measure. Fix a threshold $\tau \in R$. Write $\mathscr{P}^0(X) = X$ and recursively $\mathscr{P}^{k+1}(X) = \mathscr{P}\big(\mathscr{P}^k(X)\big)$. Define the flattening map

$$\varphi_n : \mathscr{P}^n(X) \longrightarrow X, \quad \varphi_n(U) = \bigcup_{\substack{U_1 \in \mathscr{P}^{n-1}(X) \\ U_1 \in U}} \varphi_{n-1}(U_1),$$

with $\varphi_0 = \mathrm{id}_X$.
For each $n \geq 1$, the similarity n-SuperHyperGraph *at threshold τ is*

$$H_\tau^{(n)} = \big(V_n,\ E_n\big),$$

where

$$V_n = \mathscr{P}^n(X), \quad E_n = \Big\{ S \subseteq V_n : |S| \geq 2 \text{ and } \min_{\substack{U \neq V \\ U,V \in S}} \min_{x \in \varphi_n(U),\, y \in \varphi_n(V)} \mathrm{sim}(x,y)\ \geq\ \tau \Big\}.$$

Each $S \in E_n$ is an n–superedge whose members are mutually similar (via flattening) above the threshold.

Example 3.2 (Echocardiographic Phenotype 2–SuperHyperGraph in a Heart Failure Cohort)**.** *Let*

$$X = \{P_1, P_2, \ldots, P_{80}\}$$

be a cohort of 80 patients with chronic heart failure. Each patient P_i is characterized by an 8-dimensional echocardiographic feature vector $\mathbf{e}_i \in R^8$ (e.g. left-ventricular ejection fraction, left atrial volume, E/A ratio, etc.). We define a symmetric similarity measure by the Pearson correlation:

$$\mathrm{sim}(P_i, P_j) = \frac{\mathrm{Cov}(\mathbf{e}_i, \mathbf{e}_j)}{\sigma(\mathbf{e}_i)\,\sigma(\mathbf{e}_j)}\ \in [-1, 1].$$

Level 1 (Hyperedges in $H_\tau^{(1)}$). *Fix a threshold $\tau_1 = 0.75$. By Definition 2.1, the first-level hypergraph*

$$H_{\tau_1}^{(1)} = \big(X, E_1\big)$$

has

$$E_1 = \big\{ S \subseteq X : |S| \geq 3,\ \forall P_u, P_v \in S,\ u \neq v \implies \mathrm{sim}(P_u, P_v) \geq 0.75 \big\}.$$

Three prominent hyperedges are:

- *S_{HFpEF} (n=6): patients with preserved ejection fraction and concentric remodeling, pairwise correlations* 0.78–0.82.
- *S_{HFrEF} (n=5): patients with reduced ejection fraction and dilated ventricles, correlations* 0.76–0.80.
- *S_{Valve} (n=3): patients with significant mitral regurgitation, correlations* 0.75–0.77.

Level 2 (Superedges in $H_\tau^{(2)}$). *We form the second-level vertex set $V_2 = \mathscr{P}^2(X) = \mathscr{P}\big(\mathscr{P}(X)\big)$, so that each element of V_2 is a subset of first-level hyperedges. Define the flattening map φ_2 as in Definition 3.1. Fix a (slightly relaxed) threshold $\tau_2 = 0.70$. Then by Definition 3.1 the 2–SuperHyperGraph*

$$H_{\tau_2}^{(2)} = \big(V_2, E_2\big)$$

has

$$E_2 = \Big\{ S \subseteq V_2 : |S| \geq 2 \text{ and } \min_{\substack{U \neq V \\ U,V \in S}} \min_{x \in \varphi_2(U), y \in \varphi_2(V)} \mathrm{sim}(x,y) \;\geq\; \tau_2 \Big\}.$$

Two illustrative superedges are:

$$U_1 = \{S_{\mathrm{HFpEF}}\}, \quad U_2 = \{S_{\mathrm{HFrEF}}\}, \quad U_3 = \{S_{\mathrm{Valve}}\},$$

and one finds

$$\min_{x \in S_{\mathrm{HFpEF}}, y \in S_{\mathrm{HFrEF}}} \mathrm{sim}(x,y) = 0.73 \;\geq\; \tau_2, \qquad \min_{x \in S_{\mathrm{HFpEF}}, y \in S_{\mathrm{Valve}}} \mathrm{sim}(x,y) = 0.71 \;\geq\; \tau_2,$$

whereas $\min_{x \in S_{\mathrm{HFrEF}}, y \in S_{\mathrm{Valve}}} \mathrm{sim}(x,y) = 0.69 < \tau_2$. *Hence*

$$\{U_1, U_2\} \in E_2, \quad \{U_1, U_3\} \in E_2, \quad \{U_2, U_3\} \notin E_2.$$

Thus $H_{\tau_2}^{(2)}$ encodes second-order groupings of patient-clusters whose flattened echocardiographic phenotypes remain mutually similar above 0.70, revealing hierarchical structure in heart-failure subtypes.

Example 3.3 (Echocardiographic Phenotype 3–SuperHyperGraph in a Heart Failure Cohort)**.** *Continuing from Example 3.2, let*

$$V_3 = \mathscr{P}^3(X) = \mathscr{P}\big(V_2\big)$$

and retain the same flattening maps φ_3 of Definition 3.1. We now choose a relaxed threshold $\tau_3 = 0.65$. By that definition, the 3–SuperHyperGraph

$$H_{\tau_3}^{(3)} \;=\; \big(V_3,\, E_3\big)$$

has

$$E_3 = \Big\{ S \subseteq V_3 : |S| \geq 2 \quad \text{and}$$

$$\min_{\substack{U \neq V \\ U,V \in S}} \min_{x \in \varphi_3(U), y \in \varphi_3(V)} \mathrm{sim}(x,y) \;\geq\; \tau_3 \Big\}.$$

From Example 3.2 we had the three level-2 supervertices

$$U_1 = \{S_{\mathrm{HFpEF}}\}, \quad U_2 = \{S_{\mathrm{HFrEF}}\}, \quad U_3 = \{S_{\mathrm{Valve}}\},$$

each a subset of V_2. Their pairwise minimal flattening-similarities were

$$\min_{x \in \varphi_2(U_1), y \in \varphi_2(U_2)} \mathrm{sim}(x,y) \;=\; 0.73,$$

$$\min_{x \in \varphi_2(U_1), y \in \varphi_2(U_3)} \mathrm{sim}(x,y) \;=\; 0.71,$$

$$\min_{x \in \varphi_2(U_2), y \in \varphi_2(U_3)} \mathrm{sim}(x,y) \;=\; 0.69.$$

Since all three values exceed $\tau_3 = 0.65$, the unique level-3 superedge is

$$S_3 \;=\; \{U_1, U_2, U_3\} \;\in\; E_3.$$

Thus

$$H_{\tau_3}^{(3)} = \big(V_3,\, \{S_3\}\big)$$

captures a third-order grouping of echocardiographic phenotypes—aggregating preserved-EF, reduced-EF, and valvular-remodeling patient clusters into a single super-cluster whose flattened member profiles all remain mutually correlated above 0.65.

Theorem 3.4 (Similarity n–SuperHyperGraph is an n–SuperHyperGraph)**.** *For every $n \geq 1$, the pair $H_\tau^{(n)} = (V_n, E_n)$ of Definition 3.1 is an n–SuperHyperGraph in the sense of Definition 1.3 and the iterated-powerset construction. Moreover, successive flattenings recover the usual similarity hypergraph at level* 1.

Proof. **(i) n–SuperHyperGraph structure.** By construction $V_n = \mathscr{P}^n(X)$ is finite, and

$$E_n \subseteq \mathscr{P}(V_n) \setminus \{\emptyset\},$$

with each $S \in E_n$ satisfying $|S| \geq 2$. Hence $H_\tau^{(n)}$ meets the definition of a hypergraph on $\mathscr{P}^n(X)$. Since its vertex- and edge-sets lie in the n-th iterated powerset of X, it is precisely an n–SuperHyperGraph.

(ii) Recovery of similarity hypergraph at level 1. Define $H_\tau^{(1)} = (V_1, E_1)$. Here $V_1 = \mathscr{P}^1(X) = \mathscr{P}(X)$ and

$$E_1 = \big\{ \{x,y\} \subseteq X : x \neq y,\ \mathrm{sim}(x,y) \geq \tau \big\},$$

so $H_\tau^{(1)}$ is exactly the thresholded similarity hypergraph of Definition 1.3 restricted to edges of size two. For $n > 1$, the 1-flattening φ_1 maps each supervertex $\{x\} \in V_1$ back to x, and carries each n-superedge to a hyperedge in $H_\tau^{(1)}$. Iterating flattenings $\varphi_n \circ \cdots \circ \varphi_1$ therefore recovers the original similarity hypergraph on X.

(iii) Inductive consistency. Suppose $H_\tau^{(k)}$ is a k–SuperHyperGraph recovering similarity hypergraph under successive flattenings. Then $H_\tau^{(k+1)}$ is built on $V_{k+1} = \mathscr{P}(V_k)$ with hyperedges defined by the same pairwise-similarity criterion (via φ_{k+1}). The argument in (i) shows it is a hypergraph on $\mathscr{P}^{k+1}(X)$, and the 1-flattening φ_1 again reduces it to $H_\tau^{(k)}$. By induction, all levels are n–SuperHyperGraphs that generalize the similarity hypergraph. □

Theorem 3.5 (Monotonicity in the Similarity Threshold)**.** *Let X be a finite set and $\mathrm{sim} : X \times X \to \mathbb{R}$ a symmetric similarity measure. For each $n \geq 1$ and threshold τ, define*

$$V_n = \mathscr{P}^n(X),$$

$$E_n(\tau) = \Big\{ S \subseteq V_n : |S| \geq 2,\ \min_{\substack{U \neq V \\ U,V \in S}} \big(\min_{x \in \bigcup U,\, y \in \bigcup V} \mathrm{sim}(x,y) \big) \geq \tau \Big\}.$$

If $0 \leq \tau_1 \leq \tau_2$, then

$$E_n(\tau_2) \subseteq E_n(\tau_1).$$

Proof. Take any $S \in E_n(\tau_2)$. By definition, for every pair of distinct supervertices $U, V \in S$,

$$\min_{x \in \bigcup U,\, y \in \bigcup V} \mathrm{sim}(x,y) \geq \tau_2.$$

Since $\tau_2 \geq \tau_1$, it follows directly that

$$\min_{x \in \bigcup U,\, y \in \bigcup V} \mathrm{sim}(x,y) \geq \tau_1,$$

for the same U, V. Hence S also satisfies the condition for membership in $E_n(\tau_1)$. As S was arbitrary in $E_n(\tau_2)$, we conclude $E_n(\tau_2) \subseteq E_n(\tau_1)$. □

Theorem 3.6 (Consistency under Flattening)**.** *With the same notation as above, define for each $n \geq 1$ the "flattening" of an n-supervertex $U \in V_n$ by*

$$\varphi(U) = \bigcup_{u \in U} \varphi(u), \quad \varphi(x) = \{x\} \text{ for } x \in X.$$

*Then the n-*section*—the graph on V_{n-1} whose edges are all pairs $\{U,V\} \subseteq V_{n-1}$ contained in some $S \in E_n(\tau)$—coincides with the hypergraph $(V_{n-1}, E_{n-1}(\tau))$.*

Proof. By construction, an edge $\{U,V\} \subseteq V_{n-1}$ appears in the n-section if and only if there exists an $S \in E_n(\tau)$ with $\{U,V\} \subseteq S$. That membership in $E_n(\tau)$ means

$$\min_{x \in \varphi(U),\, y \in \varphi(V)} \mathrm{sim}(x,y) \geq \tau.$$

But this is exactly the condition for $\{U,V\}$ to belong to $E_{n-1}(\tau)$. Hence the edge-sets coincide. The vertex-sets are both V_{n-1}, so the two hypergraphs agree. □

Theorem 3.7 (Preservation of Connectivity)**.** *If the thresholded similarity graph on X,*

$$G_\tau = \big(X, \{\{x,y\} : x \neq y,\ \mathrm{sim}(x,y) \geq \tau\}\big),$$

is connected, then for every $n \geq 1$, the 2-section of $(V_n, E_n(\tau))$ is also connected.

Proof. We prove by induction on n. For $n = 1$, the 2-section of $E_1(\tau)$ is exactly G_τ, which is connected by assumption.
Now assume the 2-section of $(V_k, E_k(\tau))$ is connected. By Theorem 3.6, the 2-section of $(V_{k+1}, E_{k+1}(\tau))$ coincides with $(V_k, E_k(\tau))$. Hence it is also connected. This completes the induction and the proof. □

4. Conclusion and Future Work

In this paper, we investigated *Similarity HyperGraphs* and *Similarity SuperHyperGraphs*. While the treatment is preliminary, we have successfully examined their basic mathematical structure and properties.
Looking ahead, we plan to enrich these models by incorporating uncertainty-based frameworks such as Fuzzy Sets [32, 33], Vague Sets [34, 35], Neutrosophic Sets [36, 37], QuadriPartitioned Neutrosophic Set[38, 39], Soft sets[40, 41], Rough Sets[42, 43], HyperSoft Set[44–46] and Plithogenic Sets [47, 48]. We also hope that further progress will be made in computational experiments and algorithmic investigations.

Funding

This work was undertaken without any external or institutional financial support.

Acknowledgments

I am grateful to all colleagues and mentors whose insights, feedback, and encouragement have contributed to the development of this research. I also extend my thanks to the institutions that provided facilities and resources. Finally, I appreciate the readers whose interest inspires further exploration.

Data Availability

This manuscript presents a theoretical exposition and does not rely on empirical datasets. Future empirical investigations are encouraged to evaluate and build upon the concepts introduced here.

Ethical Approval

As this study is purely theoretical and involves no human or animal participants, formal ethical approval was not necessary.

Conflicts of Interest

The author declares there are no conflicts of interest related to this work or its publication.

Author Contributions

The entire manuscript has been conceptualized and written by the sole author.

Disclaimer (Generative AI Tools)

Generative AI technologies were employed exclusively for language refinement and proofreading. Their use was confined to ethical and appropriate contexts.

Code Availability

No software code or computational tools were developed for this study.

Disclaimer (Computational Tools)

No computer-assisted proofs, symbolic computations, or automated theorem-proving software (e.g., Mathematica, SageMath, Coq) were used. All proofs and derivations were carried out manually.

Disclaimer (Limitations and Claims)

The theoretical frameworks discussed have not yet been empirically validated. Readers are advised to independently verify any referenced assertions. The findings are valid only under the specific assumptions outlined herein; extending them may require additional research. The views expressed in this paper are solely those of the author and do not necessarily reflect the positions of affiliated organizations.

Clinical Trial

This study did not involve any clinical trials.

Consent to Participate

Not applicable.

References

[1] Claude Berge. *Hypergraphs: combinatorics of finite sets*, volume 45. Elsevier, 1984.

[2] Alain Bretto. Introduction to hypergraph theory and its use in engineering and image processing. *Advances in Imaging and Electron Physics*, 131:1–64, 2004. URL https://api.semanticscholar.org/CorpusID:115532388.

[3] Alain Bretto. Hypergraph theory. *An introduction. Mathematical Engineering. Cham: Springer*, 1, 2013.

[4] Yifan Feng, Haoxuan You, Zizhao Zhang, Rongrong Ji, and Yue Gao. Hypergraph neural networks. In *Proceedings of the AAAI conference on artificial intelligence*, volume 33, pages 3558–3565, 2019.

[5] Reinhard Diestel. Graduate texts in mathematics: Graph theory.

[6] MS Sunitha and Sunil Mathew. Fuzzy graph theory: a survey. *Annals of Pure and Applied mathematics*, 4(1):92–110, 2013.

[7] Jonathan L Gross, Jay Yellen, and Mark Anderson. *Graph theory and its applications*. Chapman and Hall/CRC, 2018.

[8] Yifan Feng, Jiashu Han, Shihui Ying, and Yue Gao. Hypergraph isomorphism computation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2024.

[9] Yuxin Wang, Quan Gan, Xipeng Qiu, Xuanjing Huang, and David Wipf. From hypergraph energy functions to hypergraph neural networks. In *International Conference on Machine Learning*, pages 35605–35623. PMLR, 2023.

[10] Luotao Liu, Feng Huang, Xuan Liu, Zhankun Xiong, Menglu Li, Congzhi Song, and Wen Zhang. Multi-view contrastive learning hypergraph neural network for drug-microbe-disease association prediction. In *International Joint Conference on Artificial Intelligence*, 2023. URL https://api.semanticscholar.org/CorpusID:260845944.

[11] Wei Li, Bin Xiang, Fan Yang, Yuchen Rong, Yanbin Yin, Jianhua Yao, and Han Zhang. scmhnn: a novel hypergraph neural network for integrative analysis of single-cell epigenomic, transcriptomic and proteomic data. *Briefings in bioinformatics*, 24 6, 2023. URL https://api.semanticscholar.org/CorpusID:265031925.

[12] Shuai Wang, Jiayi Shen, Athanasios Efthymiou, Stevan Rudinac, Monika Kackovic, Nachoem Wijnberg, and Marcel Worring. Prototype-enhanced hypergraph learning for heterogeneous information networks. In *International Conference on Multimedia Modeling*, pages 462–476. Springer, 2024.

[13] Takaaki Fujita and Florentin Smarandache. A concise study of some superhypergraph classes. *Neutrosophic Sets and Systems*, 77: 548–593, 2024. URL https://fs.unm.edu/nss8/index.php/111/article/view/5416.

[14] Mohammad Hamidi, Florentin Smarandache, and Elham Davneshvar. Spectrum of superhypergraphs via flows. *Journal of Mathematics*, 2022(1):9158912, 2022.

[15] Masoud Ghods, Zahra Rostami, and Florentin Smarandache. Introduction to neutrosophic restricted superhypergraphs and neutrosophic restricted superhypertrees and several of their properties. *Neutrosophic Sets and Systems*, 50:480–487, 2022.

[16] N. B. Nalawade, M. S. Bapat, S. G. Jakkewad, G. A. Dhanorkar, and D. J. Bhosale. Structural properties of zero-divisor hypergraph and superhypergraph over $\mathbb{Z}_n$: Girth and helly property. *Panamerican Mathematical Journal*, 35(4S):485, 2025. ISSN 1064-9735.

[17] Mohammad Hamidi, Florentin Smarandache, and Mohadeseh Taghinezhad. *Decision Making Based on Valued Fuzzy Superhypergraphs*. Infinite Study, 2023.

[18] Mohammed Alqahtani. Intuitionistic fuzzy quasi-supergraph integration for social network decision making. *International Journal of Analysis and Applications*, 23:137–137, 2025.

[19] Florentin Smarandache. *SuperHyperFunction, SuperHyperStructure, Neutrosophic SuperHyperFunction and Neutrosophic SuperHyperStructure: Current understanding and future directions*. Infinite Study, 2023.

[20] Ajoy Kanti Das, Rajat Das, Suman Das, Bijoy Krishna Debnath, Carlos Granados, Bimal Shil, and Rakhal Das. A comprehensive study of neutrosophic superhyper bci-semigroups and their algebraic significance. *Transactions on Fuzzy Sets and Systems*, 8(2):80, 2025.

[21] Florentin Smarandache. Superhyperstructure & neutrosophic superhyperstructure, 2024. URL https://fs.unm.edu/SHS/. Accessed: 2024-12-01.

[22] Takaaki Fujita. Review of probabilistic hypergraph and probabilistic superhypergraph. *Spectrum of Operational Research*, 3(1): 319–338, 2025.

[23] Florentin Smarandache. *Introduction to the n-SuperHyperGraph-the most general form of graph today*. Infinite Study, 2022.

[24] Zheng Liu, Xiaohan Li, Hao Peng, Lifang He, and Philip S Yu. Heterogeneous similarity graph neural network on electronic health records. In *2020 IEEE international conference on big data (big data)*, pages 1196–1205. IEEE, 2020.

[25] Heloisa Oss Boll, Ali Amirahmadi, Amira Soliman, Stefan Byttner, and Mariana Recamonde-Mendoza. Graph neural networks for heart failure prediction on an ehr-based patient similarity graph. *arXiv preprint arXiv:2411.19742*, 2024.

[26] WenXiang Li and KL Eddie Law. Interpretable ehr disease prediction system based on disease experts and patient similarity graph (de-psg). In *International Conference on Artificial Neural Networks*, pages 87–102. Springer, 2024.

[27] Amara Tariq, Leon Su, Bhavik Patel, and Imon Banerjee. Prediction of transfusion among in-patient population using temporal pattern based clinical similarity graphs. In *AMIA Annual Symposium Proceedings*, volume 2023, page 679, 2024.

[28] Yakub Sebastian, Jason Thomas Chew, Xun Ting Tiong, Valliappan Raman, Alan Yean Yip Fong, and Patrick Hang Hui Then. Anomaly detection from diabetes similarity graphs using community detection and bayesian techniques. In *Proceedings of the 12th International Conference on Ubiquitous Information Management and Communication*, pages 1–9, 2018.

[29] Owen Mayer and Matthew C Stamm. Exposing fake images with forensic similarity graphs. *IEEE Journal of Selected Topics in Signal Processing*, 14(5):1049–1064, 2020.

[30] Rasmus Knappe, Henrik Bulskov, and Troels Andreasen. Similarity graphs. In *Foundations of Intelligent Systems: 14th International Symposium, ISMIS 2003, Maebashi City, Japan, October 28-31, 2003. Proceedings 14*, pages 668–672. Springer, 2003.

[31] Hao Ding, Michael Sharpnack, Chao Wang, Kun Huang, and Raghu Machiraju. Integrative cancer patient stratification via subspace merging. *Bioinformatics*, 35(10):1653–1659, 2019.

[32] Lotfi A Zadeh. Fuzzy sets. *Information and control*, 8(3):338–353, 1965.

[33] TM Nishad, Talal Ali Al-Hawary, and B Mohamed Harif. General fuzzy graphs. *Ratio Mathematica*, 47, 2023.

[34] W-L Gau and Daniel J Buehrer. Vague sets. *IEEE transactions on systems, man, and cybernetics*, 23(2):610–614, 1993.

[35] Humberto Bustince and P Burillo. Vague sets are intuitionistic fuzzy sets. *Fuzzy sets and systems*, 79(3):403–405, 1996.

[36] Said Broumi, Mohamed Talea, Assia Bakali, and Florentin Smarandache. Single valued neutrosophic graphs. *Journal of New theory*, (10):86–101, 2016.

[37] Ajoy Kanti Das, Nandini Gupta, Carlos Granados, Rakhal Das, and Suman Das. Neutrosophic approach to water quality assessment: A case study of gomati river, the largest river in tripura, india. *Neutrosophic Systems with Applications*, 22:1–12, 2024.

[38] R Radha, A Stanis Arul Mary, and Florentin Smarandache. Quadripartitioned neutrosophic pythagorean soft set. *International Journal of Neutrosophic Science (IJNS) Volume 14, 2021*, page 11, 2021.

[39] Satham Hussain, Jahir Hussain, Isnaini Rosyida, and Said Broumi. Quadripartitioned neutrosophic soft graphs. In *Handbook of Research on Advances and Applications of Fuzzy Sets and Logic*, pages 771–795. IGI Global, 2022.

[40] Dmitriy Molodtsov. Soft set theory-first results. *Computers & mathematics with applications*, 37(4-5):19–31, 1999.

[41] Pradip Kumar Maji, Ranjit Biswas, and A Ranjan Roy. Soft set theory. *Computers & mathematics with applications*, 45(4-5):555–562, 2003.

[42] Zdzisław Pawlak. Rough sets. *International journal of computer & information sciences*, 11:341–356, 1982.

[43] Zdzisław Pawlak. *Rough sets: Theoretical aspects of reasoning about data*, volume 9. Springer Science & Business Media, 2012.

[44] Florentin Smarandache. Extension of soft set to hypersoft set, and then to plithogenic hypersoft set. *Neutrosophic sets and systems*, 22 (1):168–170, 2018.

[45] Muhammad Ihsan, Atiqe Ur Rahman, and Muhammad Haris Saeed. Hypersoft expert set with application in decision making for recruitment process. 2021. URL https://api.semanticscholar.org/CorpusID:244276059.

[46] Sagvan Y Musa and Baravan A Asaad. Bipolar hypersoft homeomorphism maps and bipolar hypersoft compact spaces. *International Journal of Neutrosophic Science (IJNS)*, 19(2), 2022.

[47] Nivetha Martin. Plithogenic swara-topsis decision making on food processing methods with different normalization techniques. *Advances in Decision Making*, 69, 2022.

[48] P Sathya, Nivetha Martin, and Florentine Smarandache. Plithogenic forest hypersoft sets in plithogenic contradiction based multi-criteria decision making. *Neutrosophic Sets and Systems*, 73:668–693, 2024.