SuperHyperGraph Attention Networks

Takaaki Fujita [1] *

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

Abstract

Graph Attention Networks (GAT) employ self-attention to aggregate neighboring node features in graphs, effectively capturing structural dependencies. HyperGraph Attention Networks (HGAT) extend this mechanism to hypergraphs by alternating attention-based vertex-to-hyperedge and hyperedge-to-vertex updates, modeling higher-order relationships. In this work, we introduce the n-SuperHyperGraph Attention Network, which leverages SuperHyperGraphs—a hierarchical generalization of hypergraphs—to perform multi-tier attention among supervertices and superedges. Our investigation is purely theoretical; empirical validation via computational experiments is left for future study.

Keywords: HyperGraph, SuperHyperGraph, Graph Attention Network, HyperGraph Attention Network, SuperHyperGraph Attention Networks

Structure of this paper

The format of this paper is described below.

1 Preliminaries

We collect here the basic notions and notation used throughout the paper. Unless stated otherwise, all underlying sets are finite.

1.1 SuperHyperGraphs

Graph theory studies mathematical properties of graphs—structures of vertices connected by edges—to model complex pairwise relationships in diverse contexts [1, 2]. A hypergraph generalizes a graph by allowing edges to connect any number of vertices, thereby modeling higher-order group interactions [3–7]. An n-SuperHyperGraph further generalizes hypergraphs hierarchically, using the n-th iterated powerset to define supervertices and superedges that capture multi-level connections [8–16]. An n-SuperHyperGraph is known to generalize multigraphs [17, 18], multi-hypergraphs [19, 20], subset-vertex graphs [21, 22], hypergraphs, power set graphs [23, 24], and more, offering an intuitive framework for representing real-world hierarchical network structures(cf. [25]). Therefore, n-SuperHyperGraphs have been extensively studied and applied across diverse fields and from multiple perspectives [9, 26–33]. Below we present definitions of these and related concepts. Let H be a nonempty set and $n \in \mathbb{N}$.

Definition 1.1 (Powerset)**.** (cf. [34]) Let S be any set. The *powerset* of S, denoted $\mathcal{P}(S)$, is the collection of all subsets of S:

$$\mathcal{P}(S) = \{ A \mid A \subseteq S \}.$$

In particular, $\emptyset \in \mathcal{P}(S)$ and $S \in \mathcal{P}(S)$.

Definition 1.2 (Nonempty Powerset)**.** Let S be any set. The *nonempty powerset* of S, denoted $\mathcal{P}^*(S)$, is

$$\mathcal{P}^*(S) = \{ A \mid A \subseteq S,\ A \neq \emptyset \}.$$

Definition 1.3 (Hypergraph). [5,6,35] A *hypergraph* $H = (V(H), E(H))$ consists of

- A finite vertex set $V(H)$.
- A finite collection $E(H)$ of nonempty subsets of $V(H)$, called hyperedges.

Hypergraphs are well suited to model higher-order interactions among elements of $V(H)$.

Example 1.4 (Student Enrollment Hypergraph). Let $V(H) = \{\text{Alice, Bob, Carol, Dave}\}$ be the set of students. We define three courses as hyperedges:

$$e_{\text{Math}} = \{\text{Alice, Bob, Carol}\}, \quad e_{\text{Physics}} = \{\text{Bob, Dave}\}, \quad e_{\text{History}} = \{\text{Alice, Carol}\}.$$

Then

$$E(H) = \{\, e_{\text{Math}}, e_{\text{Physics}}, e_{\text{History}}\},$$

and

$$H = \bigl(V(H),\ E(H)\bigr)$$

is a hypergraph modeling student–course enrollment, where each hyperedge groups all students enrolled in the corresponding course.

Definition 1.5 (n-th Iterated Powerset). (cf. [36–39]) For a set H and integer $n \geq 1$, the *n-th iterated powerset* of H, denoted $\mathcal{P}^n(H)$, is defined recursively by

$$\mathcal{P}^1(H) = \mathcal{P}(H), \quad \mathcal{P}^{n+1}(H) = \mathcal{P}\bigl(\mathcal{P}^n(H)\bigr).$$

Its nonempty analogue is given by

$$\mathcal{P}^*_1(H) = \mathcal{P}^*(H), \quad \mathcal{P}^*_{n+1}(H) = \mathcal{P}^*\bigl(\mathcal{P}^*_n(H)\bigr).$$

Example 1.6 (Iterated Powersets of $\{1,2\}$). Let

$$H = \{1,2\}.$$

Then the first iterated powerset is

$$\mathcal{P}^1(H) = \mathcal{P}(H) = \bigl\{\emptyset,\ \{1\},\ \{2\},\ \{1,2\}\bigr\}.$$

The second iterated powerset is

$$\mathcal{P}^2(H) = \mathcal{P}\bigl(\mathcal{P}^1(H)\bigr),$$

which consists of all $2^4 = 16$ subsets of $\mathcal{P}^1(H)$. For instance,

$$\emptyset, \quad \{\emptyset\}, \quad \{\{1\},\{2\}\}, \quad \{\emptyset,\{1\},\{2\},\{1,2\}\}$$

are elements of $\mathcal{P}^2(H)$. The third iterated powerset $\mathcal{P}^3(H) = \mathcal{P}(\mathcal{P}^2(H))$ then contains $2^{16} = 65{,}536$ elements, each a subset of $\mathcal{P}^2(H)$.

Definition 1.7 (n-SuperHyperGraph). [11,40–43] Let V_0 be a finite base set. Define iteratively

$$\mathcal{P}^0(V_0) = V_0, \quad \mathcal{P}^{k+1}(V_0) = \mathcal{P}\bigl(\mathcal{P}^k(V_0)\bigr).$$

An *n-SuperHyperGraph* is a pair

$$\mathrm{SuHG}^{(n)} = (V, E), \quad V, E \subseteq \mathcal{P}^n(V_0),$$

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

Example 1.8 (Corporate Team and Department Structure). Let the base set V_0 be the roster of employees in a company:

$$V_0 = \{\text{Ayano, Ichiro, Shinya, Dave, Eve}\}.$$

First, form the set of informal project teams (elements of $\mathcal{P}^1(V_0)$):

$$T_1 = \{\text{Ayano, Ichiro}\}, \quad T_2 = \{\text{Shinya, Dave}\}, \quad T_3 = \{\text{Ichiro, Shinya, Eve}\}.$$

Next, form the set of formal departments as collections of teams (elements of $\mathcal{P}^2(V_0)$):

$$D_1 = \{T_1, T_2\}, \quad D_2 = \{T_2, T_3\}.$$

Finally, define cross-department working groups, also as collections of teams:

$$W_1 = \{T_1, T_3\}, \quad W_2 = \{T_1, T_2, T_3\}.$$

Then

$$V = \{D_1, D_2\}, \quad E = \{W_1, W_2\} \subseteq \mathcal{P}^2(V_0).$$

Thus $\mathrm{SuHG}^{(2)} = (V, E)$ is a 2-SuperHyperGraph whose supervertices D_i are departments (sets of teams), and whose superedges W_j are working groups (sets of teams) spanning multiple departments.

1.2 Graph Attention Network

Graph Attention Networks (GATs) enable each node to aggregate feature information from its neighbors by computing attention weights across multiple heads, thereby capturing structural dependencies within the graph [44–49]. Related concepts include Graph Neural Networks [50–52], Recurrent Neural Networks [53–56], and Dynamic Neural Networks [57–59].

Definition 1.9 (Graph Attention Network). (cf. [60–62]) Let $G = (V, E)$ be a graph with $N = |V|$ vertices and edge set $E \subseteq V \times V$. Suppose each vertex $i \in V$ has an initial feature vector $\mathbf{h}_i^{(0)} \in \mathbb{R}^{F_0}$. A *Graph Attention Network* with L layers and K attention heads per layer is defined recursively for $l = 0, \ldots, L-1$ by:

$$\mathbf{h}_i^{(l+1)} = \Big\|_{k=1}^{K} \sigma\Big(\sum_{j \in \mathcal{N}(i)} \alpha_{ij}^{(l,k)} W^{(l,k)} \mathbf{h}_j^{(l)} \Big) \quad \text{for each } i \in V,$$

where:

- $W^{(l,k)} \in \mathbb{R}^{F_{l+1}/K \times F_l}$ is a learnable weight matrix for head k in layer l.
- For each head k, a learnable attention vector $\mathbf{a}^{(l,k)} \in \mathbb{R}^{2(F_{l+1}/K)}$ and LeakyReLU activation yields unnormalized scores

$$e_{ij}^{(l,k)} = \mathrm{LeakyReLU}\big((\mathbf{a}^{(l,k)})^{\top} \big[W^{(l,k)} \mathbf{h}_i^{(l)} \parallel W^{(l,k)} \mathbf{h}_j^{(l)} \big] \big).$$

- The normalized attention coefficients are

$$\alpha_{ij}^{(l,k)} = \frac{\exp\big(e_{ij}^{(l,k)}\big)}{\sum_{m \in \mathcal{N}(i)} \exp\big(e_{im}^{(l,k)}\big)}, \quad \mathcal{N}(i) = \{ j \mid (i, j) \in E \}.$$

- σ is a nonlinear activation function (e.g. ReLU).
- $\|$ denotes vector concatenation across the K heads, so $\mathbf{h}_i^{(l+1)} \in \mathbb{R}^{F_{l+1}}$.

After L layers, the final node representations $\{\mathbf{h}_i^{(L)}\}_{i \in V}$ can be used for downstream tasks such as node classification or link prediction.

Example 1.10 (Traffic Speed Forecasting with GAT)**.** (cf. [63, 64]) Consider a road network represented as a graph $G = (V, E)$, where each vertex $i \in V$ is a traffic sensor on a highway, and an undirected edge $(i, j) \in E$ exists if sensors i and j are located on adjacent road segments. At each discrete time t, sensor i records its average vehicle speed v_i^t. We construct the input feature vector

$$\mathbf{h}_i^{(0)} = \left[v_i^{t-T+1},\, v_i^{t-T+2},\, \dots,\, v_i^t\right] \in \mathbb{R}^T$$

from the past T time-steps. A Graph Attention Network with L layers then computes

$$\mathbf{h}_i^{(l+1)} = \|_{k=1}^{K} \sigma\Big(\sum_{j \in \mathcal{N}(i)} \alpha_{ij}^{(l,k)}\, W^{(l,k)}\, \mathbf{h}_j^{(l)}\Big), \quad l = 0, \dots, L-1,$$

where $\mathcal{N}(i)$ are the neighboring sensors of i, and $\alpha_{ij}^{(l,k)}$ are attention weights learned per head k. Finally, a readout MLP maps $\mathbf{h}_i^{(L)}$ to a one-step-ahead speed prediction $\hat{v}_i^{t+1}$. By assigning higher attention to sensors whose past speeds better explain local traffic dynamics, the GAT captures spatial dependencies and yields more accurate short-term forecasting compared to fixed-weight graph convolutions.

1.3 HyperGraph attention network

A HyperGraph Attention Network (HGAT) extends this mechanism to hypergraphs by alternating attention-based aggregation between vertices and hyperedges, thereby modeling higher-order relationships among groups of nodes (cf. [65–69]).

Definition 1.11 (Hypergraph Attention Network (HGAT))**.** (cf. [70–72]) A *Hypergraph Attention Network* (HGAT) operates on a hypergraph

$$G = (V, \mathcal{E}, W),$$

where V is the set of N vertices, $\mathcal{E}$ the set of M hyperedges, and W assigns weights to hyperedges. Let

$$A \in \{0, 1\}^{N \times M}$$

be the incidence matrix with $A_{ij} = 1$ iff vertex i belongs to hyperedge j, and let

$$H^{(l)} \in \mathbb{R}^{N \times d}, \quad E^{(l)} \in \mathbb{R}^{M \times d}$$

be the feature matrices of vertices and hyperedges at layer l.

HGAT alternates two attention-based aggregation modules:

(1) Attentive vertex→hyperedge aggregation Choose a trainable projection $W \in \mathbb{R}^{d \times d'}$ and attention mechanism $a : \mathbb{R}^{d'} \times \mathbb{R}^{d'} \to \mathbb{R}$. Compute raw coefficients

$$\alpha_{ij} = \exp\big(a(h_i^{(l)} W,\, e_j^{(l)} W)\big), \quad j \in \mathcal{N}(i),$$

where $\mathcal{N}(i) = \{j \mid A_{ij} = 1\}$. Normalize and mask by A:

$$\mathbf{A} = A \odot \operatorname{softmax}\big(\mathrm{ReLU}(H^{(l)} W\, (E^{(l)} W)^T)\big) \in [0, 1]^{N \times M}.$$

Then update hyperedge features:

$$E^{(l+1)} = \sigma\big(\mathbf{A}^T H^{(l)}\big),$$

where σ is an activation function (e.g. ReLU).

(2) Attentive hyperedge→vertex aggregation Similarly, with projection W_1, compute

$$\mathbf{B} = A^T \odot \operatorname{softmax}\big(\mathrm{ReLU}(E^{(l)} W_1\, (H^{(l)} W_1)^T)\big) \in [0, 1]^{M \times N},$$

and update vertex features:

$$H^{(l+1)} = \sigma\big(\mathbf{B}^T E^{(l)}\big).$$

HGAT Layer: One HGAT layer applies (1) then (2) in sequence, passing messages vertex→hyperedge→vertex to capture high-order correlations. A full HGAT stacks L such layers, yielding final vertex embeddings $H^{(L)}$.

Example 1.12 (Document Classification via Word Hypergraph)**.** (cf. [73, 74]) Consider a corpus of N documents $V = \{d_1, \ldots, d_N\}$. Let the vocabulary be $\mathcal{W} = \{w_1, \ldots, w_M\}$. We build the hypergraph

$$G = (V, \mathcal{E}), \quad \mathcal{E} = \left\{ e_j \mid e_j = \{ d_i \in V \mid w_j \text{ appears in } d_i \} \right\}.$$

Each vertex d_i is represented by a TF–IDF feature vector $\mathbf{h}_i^{(0)} \in \mathbb{R}^F$, and each word hyperedge e_j by the one-hot indicator $\mathbf{e}_j^{(0)} \in \mathbb{R}^M$ (or by a learned embedding of word w_j).

An HGAT layer then alternates:

(1) Document→Word attention: Compute unnormalized scores $\alpha_{ij} = \exp\big(a(\mathbf{h}_i^{(l)} W,\ \mathbf{e}_j^{(l)} W)\big)$ for each (d_i, e_j) with $d_i \in e_j$, normalize and mask by the incidence matrix $A \in \{0, 1\}^{N \times M}$:

$$\mathbf{A} = A \odot \operatorname{softmax}\big(\operatorname{LeakyReLU}(H^{(l)} W\, (E^{(l)} W)^T)\big), \quad E^{(l+1)} = \sigma(\mathbf{A}^T H^{(l)}).$$

(2) Word→Document attention: Similarly, with projection W_1,

$$\mathbf{B} = A^T \odot \operatorname{softmax}\big(\operatorname{LeakyReLU}(E^{(l)} W_1\, (H^{(l)} W_1)^T)\big), \quad H^{(l+1)} = \sigma(\mathbf{B}^T E^{(l)}).$$

After L layers, the final document embeddings $\mathbf{h}_i^{(L)}$ are fed into a softmax classifier to predict each document's topic label. By attending over word-based hyperedges, HGAT captures both the individual document content and higher-order relationships among documents sharing the same words.

2 Main Results: n-SuperHyperGraph Attention Network (n-SuHGAT)

The n-SuperHyperGraph Attention Network (n-SuHGAT) generalizes further to n-superhypergraphs, performing iterative attention across n-level supervertices and superedges to capture deep hierarchical interactions in complex data structures.

Definition 2.1 (n-SuperHyperGraph Attention Network (n-SuHGAT))**.** Let $\mathrm{SuHG}^{(n)} = (V^{(n)}, E^{(n)})$ be an n-SuperHyperGraph with $|V^{(n)}| = N$, $|E^{(n)}| = M$, and incidence matrix

$$A^{(n)} \in \{0, 1\}^{N \times M}, \quad A^{(n)}_{uv} = 1 \iff n\text{-supervertex } u \in V^{(n)} \text{ lies in } n\text{-superedge } v \in E^{(n)}.$$

At layer l, let

$$H^{(l)} \in \mathbb{R}^{N \times d}, \quad E^{(l)} \in \mathbb{R}^{M \times d}$$

be feature matrices of n-supervertices and n-superedges. An *n-SuHGAT layer* performs:

(1) Supervertex→Superedge attention: Choose a learnable projection $W \in \mathbb{R}^{d \times d'}$ and attention kernel $a : \mathbb{R}^{d'} \times \mathbb{R}^{d'} \to \mathbb{R}$. Compute unnormalized scores

$$e_{uv} = a\big(h_u^{(l)} W,\ e_v^{(l)} W\big), \quad (u, v) \text{ with } A^{(n)}_{uv} = 1.$$

Normalize over each supervertex's incident superedges and mask by $A^{(n)}$:

$$\mathbf{A}^{(n)} = A^{(n)} \ \odot\ \operatorname{softmax}\big(\operatorname{LeakyReLU}(H^{(l)} W\, (E^{(l)} W)^T)\big) \ \in\ [0, 1]^{N \times M}.$$

Update superedge features:

$$E^{(l+1)} = \sigma\big((\mathbf{A}^{(n)})^T\, H^{(l)}\big).$$

(2) Superedge→Supervertex attention: With projection W_1, compute

$$\mathbf{B}^{(n)} = \big(A^{(n)}\big)^T \ \odot\ \operatorname{softmax}\big(\operatorname{LeakyReLU}(E^{(l)} W_1\, (H^{(l)} W_1)^T)\big) \ \in\ [0, 1]^{M \times N},$$

and update supervertex features:

$$H^{(l+1)} = \sigma\big((\mathbf{B}^{(n)})^T\, E^{(l)}\big).$$

Example 2.2 (Cross-Department Project Collaboration)**.** Let the base set V_0 be the set of all employees in a company:

$$V_0 = \{\text{Ayano}, \text{Ichiro}, \dots\}.$$

Form the first-level powerset $\mathcal{P}^1(V_0)$ of project teams (each a subset of employees), then the second-level powerset $\mathcal{P}^2(V_0)$ of departments (each a set of teams). We build the 2-SuperHyperGraph

$$\text{SuHG}^{(2)} = \big(V^{(2)}, E^{(2)}\big), \quad V^{(2)} = \{\text{Dept}_1, \text{Dept}_2, \dots\} \subseteq \mathcal{P}^2(V_0), \quad E^{(2)} = \{\text{Proj}_A, \text{Proj}_B, \dots\} \subseteq \mathcal{P}^2(V_0),$$

where each Dept_i is a set of teams and each Proj_j is a set of departments collaborating on a project. The incidence matrix $A^{(2)} \in \{0,1\}^{|V^{(2)}|\times|E^{(2)}|}$ indicates which department lies in which project.

We assign initial features as follows:

$$h^{(0)}_{\text{Dept}} = \frac{1}{|\text{Dept}|} \sum_{T\in\text{Dept}} \Big(\frac{1}{|T|} \sum_{v\in T} x_v\Big), \quad e^{(0)}_{\text{Proj}} = \frac{1}{|\text{Proj}|} \sum_{d\in\text{Proj}} h^{(0)}_d,$$

where $x_v \in \mathbb{R}^d$ is a learned embedding of employee v.

A 2-SuHGAT layer then alternates:

(1) Department→Project attention:

$$\mathbf{A}^{(2)} = A^{(2)} \odot \text{softmax}\big(\text{LeakyReLU}(H^{(l)}W\,(E^{(l)}W)^T)\big), \quad E^{(l+1)} = \sigma\big((\mathbf{A}^{(2)})^T H^{(l)}\big).$$

(2) Project→Department attention:

$$\mathbf{B}^{(2)} = (A^{(2)})^T \odot \text{softmax}\big(\text{LeakyReLU}(E^{(l)}W_1\,(H^{(l)}W_1)^T)\big), \quad H^{(l+1)} = \sigma\big((\mathbf{B}^{(2)})^T E^{(l)}\big).$$

After L layers, the final department embeddings $H^{(L)}$ capture both intra-department and cross-project influences, and can be used to predict metrics such as departmental performance or project success likelihood.

Example 2.3 (Cross-Divisional Strategic Initiatives)**.** Let the base set V_0 be the roster of all employees:

$$V_0 = \{\text{Ayano}, \text{Ichiro}, \dots\}.$$

First form

$$\mathcal{P}^1(V_0) = \{\text{teams}\}, \quad \mathcal{P}^2(V_0) = \{\text{departments}\}, \quad \mathcal{P}^3(V_0) = \{\text{divisions}\}.$$

We construct the 3-SuperHyperGraph

$$\text{SuHG}^{(3)} = (V^{(3)}, E^{(3)}), \quad V^{(3)} = \{\text{Division}_1, \dots, \text{Division}_P\}, \quad E^{(3)} = \{\text{Initiative}_A, \dots, \text{Initiative}_Q\},$$

where each $\text{Division}_p \subseteq \mathcal{P}^2(V_0)$ is a set of departments, and each $\text{Initiative}_q \subseteq \mathcal{P}^2(V_0)$ is the set of divisions collaborating on that initiative. The incidence matrix $A^{(3)} \in \{0,1\}^{P\times Q}$ indicates which division participates in which initiative.

We define initial features by hierarchical averaging:

$$h^{(0)}_{\text{Dept}} = \frac{1}{|\text{Dept}|} \sum_{T\in\text{Dept}} \Big(\frac{1}{|T|} \sum_{v\in T} x_v\Big), \quad h^{(0)}_{\text{Div}} = \frac{1}{|\text{Div}|} \sum_{d\in\text{Div}} h^{(0)}_d, \quad e^{(0)}_{\text{Init}} = \frac{1}{|\text{Init}|} \sum_{\delta\in\text{Init}} h^{(0)}_\delta,$$

where $x_v \in \mathbb{R}^d$ is the feature of employee v.

A 3-SuHGAT layer then alternates:

(1) Division→Initiative attention:

$$\mathbf{A}^{(3)} = A^{(3)} \odot \text{softmax}\big(\text{LeakyReLU}(H^{(l)}W\,(E^{(l)}W)^T)\big), \quad E^{(l+1)} = \sigma\big((\mathbf{A}^{(3)})^T H^{(l)}\big).$$

(2) Initiative→Division attention:

$$\mathbf{B}^{(3)} = (A^{(3)})^T \odot \mathrm{softmax}\big(\mathrm{LeakyReLU}(E^{(l)}W_1\,(H^{(l)}W_1)^T)\big), \quad H^{(l+1)} = \sigma\big((\mathbf{B}^{(3)})^T E^{(l)}\big).$$

After L layers, the final division embeddings $H^{(L)}$ encode both intra-division structure and cross-initiative influences. These representations can be used to predict each division's contribution to initiative success or to identify strategic synergy among divisions.

Example 2.4 (Global Consulting Alliance Analysis)**.** Let the base set V_0 be the roster of all employees in a global consulting firm:

$$V_0 = \{\mathrm{Alice}, \mathrm{Bob}, \ldots\}.$$

Organizational levels are defined by iterated powersets:

$$\mathcal{P}^1(V_0) = \{\text{teams}\}, \quad \mathcal{P}^2(V_0) = \{\text{departments}\}, \quad \mathcal{P}^3(V_0) = \{\text{divisions}\}, \quad \mathcal{P}^4(V_0) = \{\text{portfolios}\}.$$

We construct the 4-SuperHyperGraph $\mathrm{SuHG}^{(4)} = (V^{(4)}, E^{(4)})$, where

$$V^{(4)} = \{\mathrm{Portfolio}_1, \ldots, \mathrm{Portfolio}_P\}, \quad E^{(4)} = \{\mathrm{Alliance}_A, \ldots, \mathrm{Alliance}_Q\},$$

with each $\mathrm{Portfolio}_p \subseteq \mathcal{P}^3(V_0)$ a set of divisions, and each $\mathrm{Alliance}_q \subseteq \mathcal{P}^3(V_0)$ a set of portfolios collaborating on a global initiative. The incidence matrix

$$A^{(4)} \in \{0,1\}^{P\times Q}, \quad A^{(4)}_{pq} = 1 \iff \mathrm{Portfolio}_p \in \mathrm{Alliance}_q.$$

Initial features are defined hierarchically:

$$h^{(0)}_{\mathrm{team}} = \frac{1}{|T|}\sum_{v\in T} x_v, \quad h^{(0)}_{\mathrm{dept}} = \frac{1}{|D|}\sum_{T\in D} h^{(0)}_T, \quad h^{(0)}_{\mathrm{div}} = \frac{1}{|S|}\sum_{D\in S} h^{(0)}_D,$$

$$h^{(0)}_{\mathrm{port}} = \frac{1}{|P|}\sum_{S\in P} h^{(0)}_S, \quad e^{(0)}_{\mathrm{alli}} = \frac{1}{|A|}\sum_{p\in A} h^{(0)}_p,$$

where $x_v \in \mathbb{R}^d$ is the embedding of employee v.

A 4-SuHGAT layer alternates:

(1) Portfolio→Alliance attention:

$$\mathbf{A}^{(4)} = A^{(4)} \odot \mathrm{softmax}\big(\mathrm{LeakyReLU}(H^{(l)}W\,(E^{(l)}W)^T)\big), \quad E^{(l+1)} = \sigma\big((\mathbf{A}^{(4)})^T H^{(l)}\big).$$

(2) Alliance→Portfolio attention:

$$\mathbf{B}^{(4)} = (A^{(4)})^T \odot \mathrm{softmax}\big(\mathrm{LeakyReLU}(E^{(l)}W_1\,(H^{(l)}W_1)^T)\big), \quad H^{(l+1)} = \sigma\big((\mathbf{B}^{(4)})^T E^{(l)}\big).$$

After L layers, the final portfolio embeddings $H^{(L)}$ capture both intra-portfolio structure and cross-alliance influences, enabling predictions of alliance success metrics.

Theorem 2.5 (n-SuHGAT Generalizes GAT and HGAT)**.** *Let GAT be the special case $n = 0$ (vertices$V^{(0)} = V_0$, edges of size-2 give adjacency) and HGAT the case $n = 1$. Then:*

$$n = 0 \implies \textit{n-SuHGAT} = \textit{Graph Attention Network},$$

$$n = 1 \implies \textit{n-SuHGAT} = \textit{Hypergraph Attention Network}.$$

Moreover, every n-SuHGAT layer respects the n-SuperHyperGraph structure via the incidence matrix $A^{(n)}$.

Proof. For $n = 0$, $V^{(0)} = V_0$, $E^{(0)}$ consists of all size-2 subsets of V_0, and $A^{(0)}$ becomes the usual adjacency. The two-phase attention reduces to node→node and node→node updates, recovering exactly the GAT formulas.

For $n = 1$, $V^{(1)} = V_0$, $E^{(1)}$ are hyperedges, and $A^{(1)}$ is the incidence matrix of the hypergraph. The two-phase attention matches the HGAT aggregation rules.

In general, the use of $A^{(n)}$ to mask and normalize messages ensures that attention is only passed along valid n-supervertex–n-superedge incidences, hence preserving the n-SuperHyperGraph's structure at every layer. □

Theorem 2.6 (Permutation Equivariance). *Let* $\mathrm{SuHG}^{(n)} = (V^{(n)}, E^{(n)})$ *with incidence* $A^{(n)}$*, and let* $P_V \in \{0,1\}^{N\times N}$*,* $P_E \in \{0,1\}^{M\times M}$ *be permutation matrices acting on supervertices and superedges. Define permuted features*

$$\widetilde{H}^{(l)} = P_V\, H^{(l)}, \quad \widetilde{E}^{(l)} = P_E\, E^{(l)}, \quad \widetilde{A}^{(n)} = P_V\, A^{(n)}\, P_E^T.$$

Then one n-SuHGAT layer applied to $\big(\widetilde{H}^{(l)}, \widetilde{E}^{(l)}, \widetilde{A}^{(n)}\big)$ *yields*

$$\widetilde{E}^{(l+1)} = P_E\, E^{(l+1)}, \quad \widetilde{H}^{(l+1)} = P_V\, H^{(l+1)},$$

so the network is equivariant under simultaneous relabeling of supervertices and superedges.

Proof. In the supervertex→superedge step the unnormalized scores satisfy

$$\widetilde{H}^{(l)} W\, (\widetilde{E}^{(l)} W)^T = P_V \big(H^{(l)} W\, (E^{(l)} W)^T\big) P_E^T.$$

Applying row-wise softmax and masking by $\widetilde{A}^{(n)} = P_V A^{(n)} P_E^T$ gives $\widetilde{\mathbf{A}}^{(n)} = P_V\, \mathbf{A}^{(n)}\, P_E^T$. Hence $\widetilde{E}^{(l+1)} = \sigma((\widetilde{\mathbf{A}}^{(n)})^T\, \widetilde{H}^{(l)}) = P_E\, \sigma((\mathbf{A}^{(n)})^T\, H^{(l)}) = P_E\, E^{(l+1)}$.

Similarly for the superedge→supervertex step one shows $\widetilde{\mathbf{B}}^{(n)} = P_E\, \mathbf{B}^{(n)}\, P_V^T$ and thus $\widetilde{H}^{(l+1)} = P_V\, H^{(l+1)}$. This completes the proof of equivariance. □

Theorem 2.7 (Special-Case Reduction). *For* $n = 0$*, an n-SuHGAT layer reduces to a standard Graph Attention layer on* $G = (V_0, E^{(0)})$*. For* $n = 1$*, it reduces to a Hypergraph Attention layer on* $H = (V_0, E^{(1)})$*.*

Proof. When $n = 0$, $V^{(0)} = V_0$, $E^{(0)}$ consists of edges of size two, and $A^{(0)}$ is the adjacency matrix. The two-phase attention then aggregates node → node messages exactly as in GAT.

When $n = 1$, $V^{(1)} = V_0$, $E^{(1)}$ is the hyperedge set, and $A^{(1)}$ the incidence matrix. The supervertex→superedge and superedge→supervertex updates coincide with the HGAT formulas. □

Theorem 2.8 (Layerwise Lipschitz Continuity). *Assume the attention kernel a is Lipschitz with constant* L_a*, and* σ*,* LeakyReLU *are 1-Lipschitz. Then each n-SuHGAT layer is Lipschitz continuous in the input features* $(H^{(l)}, E^{(l)})$*. Concretely, there exists* $C > 0$ *(depending on* $\|A^{(n)}\|$*,* $\|W\|$*,* $\|W_1\|$*, and* L_a*) such that*

$$\big\|H^{(l+1)} - \widetilde{H}^{(l+1)}\big\|_F + \big\|E^{(l+1)} - \widetilde{E}^{(l+1)}\big\|_F \;\le\; C\Big(\|H^{(l)} - \widetilde{H}^{(l)}\|_F + \|E^{(l)} - \widetilde{E}^{(l)}\|_F\Big).$$

Proof. Both aggregation steps are compositions of linear maps, LeakyReLU, softmax, masking by $A^{(n)}$, and σ. Each component is Lipschitz: the linear map with weight W has constant $\|W\|$, LeakyReLU and σ are 1-Lipschitz, and softmax over each row is L_{sm}-Lipschitz in the row-vector argument. Masking by $A^{(n)}$ is non-expansive. Chaining these gives an overall layer-wise Lipschitz constant

$$C = \|W\|\, L_{\mathrm{sm}}\, \|A^{(n)}\| + \|W_1\|\, L_{\mathrm{sm}}\, \|A^{(n)}\|.$$

□ □

Theorem 2.9 (Message-Passing Equivalence). *An n-SuHGAT layer on* $\mathrm{SuHG}^{(n)}$ *is an instance of a Message-Passing Neural Network (MPNN) on the associated bipartite graph* $B = (V^{(n)} \cup E^{(n)},\, \mathcal{E}')$*, where each supervertex* $u \in V^{(n)}$ *is connected to each incident superedge* $v \in E^{(n)}$ *whenever* $A^{(n)}_{uv} = 1$*. In particular, the two attention phases*

$$H^{(l)} \to E^{(l+1)} \quad \text{and} \quad E^{(l)} \to H^{(l+1)}$$

correspond exactly to the standard MPNN message-aggregate-update steps on B.

Proof. Define B with adjacency

$$\widetilde{A} = \begin{pmatrix} 0 & A^{(n)} \\ (A^{(n)})^T & 0 \end{pmatrix},$$

so that messages flow along edges (u, v). In the first phase, each u-to-v message $\alpha_{uv}^{(n)} W\, h_u^{(l)}$ is computed, aggregated at v, and passed through σ, exactly matching the MPNN form

$$m_v = \sum_{u \in \mathcal{N}(v)} M(h_u^{(l)}, e_v^{(l)}), \quad e_v^{(l+1)} = U(m_v, e_v^{(l)}).$$

The second phase is identical but with roles swapped. Hence the full n-SuHGAT layer is a two-step MPNN on B. □

Theorem 2.10 (Row-Stochastic Attention). *In each n-SuHGAT layer, the attention matrices* $\mathbf{A}^{(n)} \in [0, 1]^{N \times M}$ *and* $\mathbf{B}^{(n)} \in [0, 1]^{M \times N}$ *are row-stochastic:*

$$\sum_{v=1}^{M} \mathbf{A}_{uv}^{(n)} = 1 \quad (\forall u), \quad \sum_{u=1}^{N} \mathbf{B}_{vu}^{(n)} = 1 \quad (\forall v).$$

Thus each update is a convex combination of neighbor features.

Proof. By definition, $\mathbf{A}^{(n)}$ is obtained by applying softmax over each row u of the masked score matrix $\mathrm{LeakyReLU}(H^{(l)} W (E^{(l)} W)^T)$. Softmax ensures $\sum_v \exp(e_{uv}) / \sum_{v'} \exp(e_{uv'}) = 1$. Masking by $A^{(n)}$ zeroes out non-neighbors but does not change the normalization over the remaining entries. An identical argument applies to $\mathbf{B}^{(n)}$. □

Theorem 2.11 (Constant-Feature Invariance). *If at layer* 0 *all supervertices share the same feature* $\mathbf{h}_u^{(0)} = \mathbf{c} \in \mathbb{R}^d$ *and all superedges share* $\mathbf{e}_v^{(0)} = \mathbf{d} \in \mathbb{R}^d$*, then for every layer l all $H^{(l)}$ rows remain equal and all $E^{(l)}$ rows remain equal.*

Proof. Assume $h_u^{(l)} = \mathbf{c}_l$ and $e_v^{(l)} = \mathbf{d}_l$ for all u, v. Then every unnormalized score

$$e_{uv} = a(\mathbf{c}_l W,\ \mathbf{d}_l W)$$

is the same constant independent of (u, v). After masking, each row of the score matrix is constant, so softmax yields $\mathbf{A}_{uv}^{(n)} = 1/\deg(u)$ for each neighbor v. Hence

$$e_v^{(l+1)} = \sigma\Big(\sum_{u \in \mathcal{N}(v)} \tfrac{1}{\deg(u)}\, \mathbf{c}_l \Big) = \sigma(\mathbf{c}_l) \ = \mathbf{d}_{l+1},$$

a single vector for all v. The second phase is identical, so $h_u^{(l+1)} = \sigma(\mathbf{d}_{l+1}) = \mathbf{c}_{l+1}$ for all u. By induction, the constant-feature property holds at every layer. □

3 Conclusion

In this work, we introduced the n-SuperHyperGraph Attention Network, which leverages SuperHyperGraphs—a hierarchical generalization of hypergraphs—to perform multi-tier attention among supervertices and superedges. In future work, we plan to integrate advanced graph algorithms, conduct empirical evaluations on diverse real-world datasets, and develop extensions of the n-SuperHyperGraph Attention Network using frameworks such as Fuzzy Graphs [75, 76], Intuitionistic Fuzzy Graphs [77–80], Spherical Fuzzy Graphs [81–83], Bidirected Graphs [84–87], Neutrosophic Graphs [88–92], and Plithogenic Graphs [93–95]. For example, we will explore extensions and algorithmic refinements of the n-SuperHyperGraph Attention Network inspired by Fuzzy Graph Attention Networks [96–101], and perform systematic performance evaluations.

Funding

This study did not receive any financial or external support from organizations or individuals.

Acknowledgments

We extend our sincere gratitude to everyone who provided insights, inspiration, and assistance throughout this research. We particularly thank our readers for their interest and acknowledge the authors of the cited works for laying the foundation that made our study possible. We also appreciate the support from individuals and institutions that provided the resources and infrastructure needed to produce and share this paper. Finally, we are grateful to all those who supported us in various ways during this project.

Data Availability

This research is purely theoretical, involving no data collection or analysis. We encourage future researchers to pursue empirical investigations to further develop and validate the concepts introduced here.

Ethical Approval

As this research is entirely theoretical in nature and does not involve human participants or animal subjects, no ethical approval is required.

Conflicts of Interest

The authors confirm that there are no conflicts of interest related to the research or its publication.

Disclaimer

This work presents theoretical concepts that have not yet undergone practical testing or validation. Future researchers are encouraged to apply and assess these ideas in empirical contexts. While every effort has been made to ensure accuracy and appropriate referencing, unintentional errors or omissions may still exist. Readers are advised to verify referenced materials on their own. The views and conclusions expressed here are the authors' own and do not necessarily reflect those of their affiliated organizations.

References

[1] Reinhard Diestel. Graph theory 3rd ed. *Graduate texts in mathematics*, 173(33):12, 2005.

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

[3] Yue Gao, Zizhao Zhang, Haojie Lin, Xibin Zhao, Shaoyi Du, and Changqing Zou. Hypergraph learning: Methods and practices. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 44(5):2548–2566, 2020.

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

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

[7] Yue Gao, Yifan Feng, Shuyi Ji, and Rongrong Ji. Hgnn+: General hypergraph neural networks. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 45(3):3181–3199, 2022.

[8] Mohammad Hamidi and Mohadeseh Taghinezhad. *Application of Superhypergraphs-Based Domination Number in Real World*. Infinite Study, 2023.

[9] Min Huang, Fenghua Li, et al. Optimizing ai-driven digital resources in vocational english learning using plithogenic n-superhypergraph structures for adaptive content recommendation. *Neutrosophic Sets and Systems*, 88:283–295, 2025.

[10] Takaaki Fujita. Knowledge superhypergraphs, multimodal superhypergraphs, lattice superhypergraphs, and hyperbolic superhypergraphs: Concepts and applications. *Journal of Operational and Strategic Analytics*, 3(2):95–119, 2025.

[11] Takaaki Fujita and Florentin Smarandache. A concise study of some superhypergraph classes. *Neutrosophic Sets and Systems*, 77:548–593, 2024.

[12] Takaaki Fujita. Superhypergraph neural networks and plithogenic graph neural networks: Theoretical foundations. *arXiv preprint arXiv:2412.01176*, 2024.

[13] Takaaki Fujita, Atiqe Ur Rahman, Arkan A. Ghaib, Talal Ali Al-Hawary, and Arif Mehmood. On the properties and illustrative examples of soft superhypergraphs and rough superhypergraphs. *Prospects for Applied Mathematics and Data Analysis*, 5(1):12–31, 2025.

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

[15] Takaaki Fujita, Muhammad Gulistan, and Arkan A. Ghaib. Modeling molecular interactions with hyper-networks and super-hyper-networks. *Advances in Research*, 26(4):294–326, 2025.

[16] Takaaki Fujita. *Advancing Uncertain Combinatorics through Graphization, Hyperization, and Uncertainization: Fuzzy, Neutrosophic, Soft, Rough, and Beyond*. Biblio Publishing, 2025.

[17] Yoshio Sano. Characterizations of competition multigraphs. *Discrete Applied Mathematics*, 157(13):2978–2982, 2009.

[18] Brian Alspach and Cun-Quan Zhang. Cycle covers of cubic multigraphs. *Discrete mathematics*, 111(1-3):11–17, 1993.

[19] Ziang Li, Jie Wu, Guojing Han, Chi Ma, and Yuenai Chen. Multi-hypergraph neural network with fusion of location information for session-based recommendation. *IAENG International Journal of Applied Mathematics*, 53(4), 2023.

[20] Zhe Yang, Liangkui Xu, and Lei Zhao. Efbh: Collaborative filtering model based on multi-hypergraph encoder. *IEEE Transactions on Consumer Electronics*, 70(1):2939–2948, 2023.

[21] Florentin Smarandache, WB Kandasamy, and K Ilanthenral. Subset vertex graphs for social networks. 2018.

[22] WB Vasantha Kandasamy, K Ilanthenral, and Florentin Smarandache. *Subset Vertex Multigraphs and Neutrosophic Multigraphs for Social Multi Networks*. Infinite Study, 2019.

[23] Yasar Nacaroglu, Nihat Akgunes, Sedat Pak, and I Naci Cangul. Some graph parameters of power set graphs. *Advances & Applications in Discrete Mathematics*, 26(2), 2021.

[24] Melody Mae Cabigting Lunar and Renson Aguilar Robles. Characterization and structure of a power set graph. *International Journal of Advanced Research and Publications*, 3(6):1–4, 2019.

[25] Takaaki Fujita and Florentin Smarandache. A superhypergraph framework for subset-vertex graphs, h-models, k-chain-free sets, power set, johnson, and kneser graphs. 2025.

[26] Eduardo Luciano Hernandez Ramos, Luis Ramiro Ayala Ayala, and Kevin Alexander Samaniego Macas. Study of factors that influence a victim's refusal to testify for sexual reasons due to external influence using plithogenic n-superhypergraphs. *Operational Research Journal*, 46(2):328–337, 2025.

[27] Takaaki Fujita. Modeling hierarchical systems in graph signal processing, electric circuits, and bond graphs via hypergraphs and superhypergraphs. *Journal of Engineering Research and Reports*, 27(5):542, 2025.

[28] Takaaki Fujita. Unifying grain boundary networks and crystal graphs: A hypergraph and superhypergraph perspective in material sciences. *Asian Journal of Advanced Research and Reports*, 19(5):344–379, 2025.

[29] Eduardo Martín Campoverde Valencia, Jessica Paola Chuisaca Vásquez, and Francisco Ángel Becerra Lois. Multineutrosophic analysis of the relationship between survival and business growth in the manufacturing sector of azuay province, 2020–2023, using plithogenic n-superhypergraphs. *Neutrosophic Sets and Systems*, 84(1):28, 2025.

[30] Takaaki Fujita and Florentin Smarandache. Neutrosophic soft n-super-hypergraphs with real-world applications. *European Journal of Pure and Applied Mathematics*, 18(3):6621, 2025.

[31] Takaaki Fujita. Chemical hypergraph and chemical superhypergraph. 2025.

[32] Julio Cesar Méndez Bravo, Claudia Jeaneth Bolanos Piedrahita, Manuel Alberto Méndez Bravo, and Luis Manuel Pilacuan-Bonete. Integrating smed and industry 4.0 to optimize processes with plithogenic n-superhypergraphs. *Neutrosophic Sets and Systems*, 84:328–340, 2025.

[33] Takaaki Fujita. Elevating uncertainty in complex networks: Bipolar neutrosophic directed superhypergraphs for multi-tier bipolar hierarchical modeling. *Authorea Preprints*, 2025.

[34] Keith Devlin. *The joy of sets: fundamentals of contemporary set theory*. Springer Science & Business Media, 2012.

[35] Muhammad Akram and Gulfam Shahzadi. Hypergraphs in m-polar fuzzy environment. *Mathematics*, 6(2):28, 2018.

[36] Florentin Smarandache. Foundation of superhyperstructure & neutrosophic superhyperstructure. *Neutrosophic Sets and Systems*, 63(1):21, 2024.

[37] 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.

[38] Takaaki Fujita. An introduction and reexamination of hyperprobability and superhyperprobability: Comprehensive overview. *Asian Journal of Probability and Statistics*, 27(5):82–109, 2025.

[39] Adel Al-Odhari. Neutrosophic power-set and neutrosophic hyper-structure of neutrosophic set of three types. *Annals of Pure and Applied Mathematics*, 31(2):125–146, 2025.

[40] Takaaki Fujita. Review of probabilistic hypergraph and probabilistic superhypergraph. *Spectrum of Operational Research*, 3(1):319–338, 2026.

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

[42] Takaaki Fujita. Extensions of multidirected graphs: Fuzzy, neutrosophic, plithogenic, rough, soft, hypergraph, and superhypergraph variants. *International Journal of Topology*, 2(3):11, 2025.

[43] Florentin Smarandache. *Extension of HyperGraph to n-SuperHyperGraph and to Plithogenic n-SuperHyperGraph, and Extension of HyperAlgebra to n-ary (Classical-/Neutro-/Anti-) HyperAlgebra*. Infinite Study, 2020.

[44] Petar Velickovic, Guillem Cucurull, Arantxa Casanova, Adriana Romero, Pietro Lio, Yoshua Bengio, et al. Graph attention networks. *stat*, 1050(20):10–48550, 2017.

[45] Petar Veličković, Guillem Cucurull, Arantxa Casanova, Adriana Romero, Pietro Lio, and Yoshua Bengio. Graph attention networks. *arXiv preprint arXiv:1710.10903*, 2017.

[46] Xiao Wang, Houye Ji, Chuan Shi, Bai Wang, Yanfang Ye, Peng Cui, and Philip S Yu. Heterogeneous graph attention network. In *The world wide web conference*, pages 2022–2032, 2019.

[47] Aristidis G Vrahatis, Konstantinos Lazaros, and Sotiris Kotsiantis. Graph attention networks: a comprehensive review of methods and applications. *Future Internet*, 16(9):318, 2024.

[48] Shaked Brody, Uri Alon, and Eran Yahav. How attentive are graph attention networks? *arXiv preprint arXiv:2105.14491*, 2021.

[49] Tianchi Yang, Linmei Hu, Chuan Shi, Houye Ji, Xiaoli Li, and Liqiang Nie. Hgat: Heterogeneous graph attention networks for semi-supervised short text classification. *ACM Transactions on Information Systems (TOIS)*, 39(3):1–29, 2021.

[50] B Vembu and S Loghambal. Pseudo-graph neural networks on ordinary differential equations. *Journal of Computational Mathematica*, 6(1):117–123, 2022.

[51] Lung-Hao Lee and Yi Lu. Multiple embeddings enhanced multi-graph neural networks for chinese healthcare named entity recognition. *IEEE Journal of Biomedical and Health Informatics*, 25(7):2801–2810, 2021.

[52] Yuzhi Song, Hailiang Ye, Ming Li, and Feilong Cao. Deep multi-graph neural networks with attention fusion for recommendation. *Expert Systems with Applications*, 191:116240, 2022.

[53] Hongjie Chen, Ryan A Rossi, Sungchul Kim, Kanak Mahadik, and Hoda Eldardiry. Probabilistic hypergraph recurrent neural networks for time-series forecasting. In *Proceedings of the 31st ACM SIGKDD Conference on Knowledge Discovery and Data Mining V. 1*, pages 82–93, 2025.

[54] Larry R Medsker, Lakhmi Jain, et al. Recurrent neural networks. *Design and applications*, 5(64-67):2, 2001.

[55] Hojjat Salehinejad, Sharan Sankar, Joseph Barfett, Errol Colak, and Shahrokh Valaee. Recent advances in recurrent neural networks. *arXiv preprint arXiv:1801.01078*, 2017.

[56] Yong Yu, Xiaosheng Si, Changhua Hu, and Jianxun Zhang. A review of recurrent neural networks: Lstm cells and network architectures. *Neural computation*, 31(7):1235–1270, 2019.

[57] Yizeng Han, Gao Huang, Shiji Song, Le Yang, Honghui Wang, and Yulin Wang. Dynamic neural networks: A survey. *IEEE transactions on pattern analysis and machine intelligence*, 44(11):7436–7456, 2021.

[58] NK Sinha, MM Gupta, and DH Rao. Dynamic neural networks: an overview. In *Proceedings of IEEE International Conference on Industrial Technology 2000 (IEEE Cat. No. 00TH8482)*, volume 1, pages 491–496. IEEE, 2000.

[59] Jifeng Guo, CL Philip Chen, Zhulin Liu, and Xixin Yang. Dynamic neural network structure: A review for its theories and applications. *IEEE Transactions on Neural Networks and Learning Systems*, 36(3):4246–4266, 2024.

[60] Liancheng He, Liang Bai, Xian Yang, Hangyuan Du, and Jiye Liang. High-order graph attention network. *Information Sciences*, 630:222–234, 2023.

[61] Yuke Wang, Ling Lu, Yang Wu, and Yinong Chen. Polymorphic graph attention network for chinese ner. *Expert Systems with Applications*, 203:117467, 2022.

[62] Yong Liu, Susen Yang, Yonghui Xu, Chunyan Miao, Min Wu, and Juyong Zhang. Contextualized graph attention network for recommendation with item knowledge graph. *IEEE Transactions on knowledge and data engineering*, 35(1):181–195, 2021.

[63] Ge Guo and Wei Yuan. Short-term traffic speed forecasting based on graph attention temporal convolutional networks. *Neurocomputing*, 410:387–393, 2020.

[64] Junho Song, Jiwon Son, Dong-hyuk Seo, Kyungsik Han, Namhyuk Kim, and Sang-Wook Kim. St-gat: A spatio-temporal graph attention network for accurate traffic speed prediction. In *Proceedings of the 31st ACM international conference on information & knowledge management*, pages 4500–4504, 2022.

[65] Junzhong Ji, Yating Ren, and Minglong Lei. Fc–hat: Hypergraph attention network for functional brain network classification. *Information Sciences*, 608:1301–1316, 2022.

[66] Zhigao Zhang, Hongmei Zhang, Zhifeng Zhang, and Bin Wang. Context-embedded hypergraph attention network and self-attention for session recommendation. *Scientific Reports*, 14(1):19413, 2024.

[67] Ling Sun, Yuan Rao, Xiangbo Zhang, Yuqian Lan, and Shuanghe Yu. Ms-hgat: memory-enhanced sequential hypergraph attention network for information diffusion prediction. In *Proceedings of the AAAI conference on artificial intelligence*, volume 36, pages 4156–4164, 2022.

[68] Ramit Sawhney, Shivam Agarwal, Arnav Wadhwa, Tyler Derr, and Rajiv Ratn Shah. Stock selection via spatiotemporal hypergraph attention network: A learning to rank approach. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 35, pages 497–504, 2021.

[69] Xiaoyi Luo, Jiaheng Peng, and Jun Liang. Directed hypergraph attention network for traffic forecasting. *IET Intelligent Transport Systems*, 16(1):85–98, 2022.

[70] Wenxuan Liu, Zizhuo Zhang, and Bang Wang. Dual-view hypergraph attention network for news recommendation. *Engineering Applications of Artificial Intelligence*, 133:108256, 2024.

[71] Jinxin Wu, Deqiang He, Jiayi Li, Jian Miao, Xianwang Li, Hongwei Li, and Sheng Shan. Temporal multi-resolution hypergraph attention network for remaining useful life prediction of rolling bearings. *Reliability Engineering & System Safety*, 247:110143, 2024.

[72] Chengbao Liu, Jie Tan, Jingwei Li, Yuan Li, and Huanjie Wang. Temporal hypergraph attention network for silicon content prediction in blast furnace. *IEEE Transactions on Instrumentation and Measurement*, 71:1–13, 2022.

[73] Kaize Ding, Jianling Wang, Jundong Li, Dingcheng Li, and Huan Liu. Be more with less: Hypergraph attention networks for inductive text classification. *arXiv preprint arXiv:2011.00387*, 2020.

[74] Tengfei Liu, Yongli Hu, Boyue Wang, Yanfeng Sun, Junbin Gao, and Baocai Yin. Hierarchical graph convolutional networks for structured long document classification. *IEEE transactions on neural networks and learning systems*, 34(10):8071–8085, 2022.

[75] Azriel Rosenfeld. Fuzzy graphs. In *Fuzzy sets and their applications to cognitive and decision processes*, pages 77–95. Elsevier, 1975.

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

[77] R Parvathi, MG Karunambigai, and Krassimir T Atanassov. Operations on intuitionistic fuzzy graphs. In *2009 IEEE international conference on fuzzy systems*, pages 1396–1401. IEEE, 2009.

[78] K Sankar and Devarasan Ezhilmaran. Balanced bipolar intuitionistic fuzzy graphs. *International Research Journal of Engineering and Technology (IRJET)*, 3(11):806–812, 2016.

[79] MG Karunambigai, M Akram, S Sivasankar, and K Palanivel. Balanced intuitionistic fuzzy graphs. *Applied Mathematical Sciences*, 7(51):2501–2514, 2013.

[80] Rajab Ali Borzooei and Hossein Rashmanlou. Ring sum in product intuitionistic fuzzy graphs. *Journal of advanced research in pure mathematics*, 7(1):16–31, 2015.

[81] Preeti Devi etc. Decision support in selecting a reliable strategy for sustainable urban transport based on laplacian energy of t-spherical fuzzy graphs. *Energies*, 2022.

[82] Muhammad Shoaib, Waqas Mahmood, Weded E Albalawi, and Faria Ahmad Shami. Notion of complex spherical fuzzy graph with application. *Journal of Function Spaces*, 2022.

[83] Muhammad Akram, Danish Saleem, and Talal Al-Hawary. Spherical fuzzy graphs with application to decision-making. *Mathematical and Computational Applications*, 25(1):8, 2020.

[84] Erling Wei, Wenliang Tang, and Xiaofeng Wang. Flows in 3-edge-connected bidirected graphs. *Frontiers of Mathematics in China*, 6:339–348, 2011.

[85] Soumalya Joardar and Atibur Rahaman. Almost complex structure on finite points from bidirected graphs. *Journal of Noncommutative Geometry*, 2023.

[86] Rui Xu and Cun-Quan Zhang. On flows in bidirected graphs. *Discrete mathematics*, 299(1-3):335–343, 2005.

[87] Er Ling Wei, Wen Liang Tang, and Dong Ye. Nowhere-zero 15-flow in 3-edge-connected bidirected graphs. *Acta Mathematica Sinica, English Series*, 30(4):649–660, 2014.

[88] Said Broumi, Mohamed Talea, Assia Bakali, Florentin Smarandache, and PK Kishore Kumar. Shortest path problem on single valued neutrosophic graphs. In *2017 international symposium on networks, computers and communications (ISNCC)*, pages 1–6. IEEE, 2017.

[89] Arindam Dey, Said Broumi, Le Hoang Son, Assia Bakali, Mohamed Talea, and Florentin Smarandache. A new algorithm for finding minimum spanning trees with undirected neutrosophic graphs. *Granular Computing*, 4:63–69, 2019.

[90] S Sivasankar and Said Broumi. Secure dominance in neutrosophic graphs. *Neutrosophic Sets and Systems*, 56:70–81, 2023.

[91] Ajoy Kanti Das, Nandini Gupta, Takaaki Fujita, Suman Das, Carlos Granados, Bijoy Krishna Debnath, and Rakhal Das. Advanced co-branding strategies for smart luxury products: A multi-criteria decision-making approach using neutrosophic sets. In *Co-Branding Strategies for Smart Luxury Products*, pages 203–258. IGI Global Scientific Publishing, 2026.

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

[93] Takaaki Fujita and Florentin Smarandache. A review of the hierarchy of plithogenic, neutrosophic, and fuzzy graphs: Survey and applications. In *Advancing Uncertain Combinatorics through Graphization, Hyperization, and Uncertainization: Fuzzy, Neutrosophic, Soft, Rough, and Beyond (Second Volume)*. Biblio Publishing, 2024.

[94] Fazeelat Sultana, Muhammad Gulistan, Mumtaz Ali, Naveed Yaqoob, Muhammad Khan, Tabasam Rashid, and Tauseef Ahmed. A study of plithogenic graphs: applications in spreading coronavirus disease (covid-19) globally. *Journal of ambient intelligence and humanized computing*, 14(10):13139–13159, 2023.

[95] Takaaki Fujita and Florentin Smarandache. Survey of planar and outerplanar graphs in fuzzy and neutrosophic graphs. *Advancing Uncertain Combinatorics through Graphization, Hyperization, and Uncertainization: Fuzzy, Neutrosophic, Soft, Rough, and Beyond*, page 366, 2024.

[96] Ru xia Liang, Qian Zhang, and Jianqiang Wang. Hierarchical fuzzy graph attention network for group recommendation. *2021 IEEE International Conference on Fuzzy Systems (FUZZ-IEEE)*, pages 1–6, 2021.

[97] Safa Ben Atitallah, Maha Driss, Wadii Boulila, and Anis Koubaa. Securing industrial iot environments: A fuzzy graph attention network for robust intrusion detection. *IEEE Open Journal of the Computer Society*, 2025.

[98] Jinming Xing, Ruilin Xing, Chang Xue, and Dongwen Luo. Enhancing link prediction with fuzzy graph attention networks and dynamic negative sampling. *arXiv preprint arXiv:2411.07482*, 2024.

[99] Weikang Zhou, Guangfu Ma, Nan Zhou, Xiaojun Liang, Cunnian Gao, Wenfeng Deng, Chunhua Yang, and Weihua Gui. Fgat: Industrial protocol states labeling with fuzzy graph attention network. *Available at SSRN 5281432*.

[100] Jinming Xing, Chang Xue, Dongwen Luo, and Ruilin Xing. Fgatt: A robust framework for wireless data imputation using fuzzy graph attention networks and transformer encoders. *arXiv preprint arXiv:2412.01979*, 2024.

[101] Jinming Xing, Dongwen Luo, Qisen Cheng, Chang Xue, and Ruilin Xing. Multi-view fuzzy graph attention networks for enhanced graph learning. *arXiv preprint arXiv:2412.17271*, 2024.