Qingyun Sun[1], Yi Huang[1], Haonan Yuan[1], Xingcheng Fu[1], Yisen Gao[1], Jia Wu[1], Shujian Yu[1], Angsheng Li[1], Jianxin Li[1], and Philip S. Yu[1]

[1]Affiliation not available

January 10, 2026

Abstract

Graph Machine Learning (GML) has emerged as a powerful paradigm for modeling complex relational data across diverse domains. However, the intrinsic irregularity, high-dimensional dependencies, and heterogeneity of graph structures pose fundamental challenges to representation learning and model generalization. Information theory provides a principled framework to quantify uncertainty, capture interdependencies, and guide model design in analyzing the representational and algorithmic foundations of GML. Information-theoretic graph learning methods have shown impressive achievements over the past years. This survey provides a comprehensive and principled review of GML through the unifying and rigorous lens of information theory. We comprehensively survey over 200 methods and provide a systematic review that advances information theory for GML comprising three progressive layers: (i) entropy-based measures for quantifying uncertainty and structural complexity in graphs, (ii) mutual information and divergence for capturing interdependencies and distributional discrepancies, and (iii) information-theoretic principles for guiding model design. By systematically connecting data characterization, relational analysis, and learning objectives, this survey offers an integrative perspective that advances the understanding of graph machine learning. We further identify emerging frontiers, theoretical gaps, and open challenges, laying the foundation for future developments at the intersection of GML and information theory.

Information-Theoretic Foundations and Advances in Graph Machine Learning: A Comprehensive Survey

QINGYUN SUN, YI HUANG, and HAONAN YUAN, Beihang Univeristy, China
XINGCHENG FU, Guangxi Normal University, China
YISEN GAO, The Hong Kong University of Science and Technology, China
JIA WU, Macquarie University, Australia
SHUJIAN YU, Vrije Universiteit Amsterdam, The Netherlands
ANGSHENG LI and JIANXIN LI, Beihang Univeristy, China
PHILIP S. YU, Department of Computer Science, University of Illinois, Chicago, USA

Graph Machine Learning (GML) has emerged as a powerful paradigm for modeling complex relational data across diverse domains. However, the intrinsic irregularity, high-dimensional dependencies, and heterogeneity of graph structures pose fundamental challenges to representation learning and model generalization. Information theory provides a principled framework to quantify uncertainty, capture interdependencies, and guide model design in analyzing the representational and algorithmic foundations of GML. Information-theoretic graph learning methods have shown impressive achievements over the past years. This survey provides a comprehensive and principled review of GML through the unifying and rigorous lens of information theory. We comprehensively survey over 200 methods and provide a systematic review that advances information theory for GML comprising three progressive layers: (i) entropy-based measures for quantifying uncertainty and structural complexity in graphs, (ii) mutual information and divergence for capturing interdependencies and distributional discrepancies, and (iii) information-theoretic principles for guiding model design. By systematically connecting data characterization, relational analysis, and learning objectives, this survey offers an integrative perspective that advances the understanding of graph machine learning. We further identify emerging frontiers, theoretical gaps, and open challenges, laying the foundation for future developments at the intersection of GML and information theory.

CCS Concepts: • **Computing methodologies → Neural networks**; **Learning latent representations**; • **Mathematics of computing → Graph algorithms**.

Additional Key Words and Phrases: Information Theory, Graph Machine Learning, Graph Neural Network

1 Introduction

Graph Machine Learning (GML) [194] has rapidly become indispensable paradigm for modeling and understanding complex relational structures across diverse domains, including social networks [80], recommendation systems [41], biological [190] and chemical networks [105]. The power of GML stems from its ability to reason over the irregular topology and multi-scale statistical dependencies inherent to graph-structured data. However, these characteristics also introduce profound challenges in representation learning, encompassing issues of generalization, robustness, and interpretability of the GML model. While the GML community [152] has responded with a proliferation of innovative architectures (*e.g.*, graph convolutional networks, attention mechanisms, graph autoencoders), a fundamental gap remains between these mechanistic designs and a principled understanding of their capabilities and limitations. Existing surveys have largely provided catalogs of techniques but have lacked a unifying lens to answer a foundational question:

Authors' Contact Information: Qingyun Sun, sunqy@buaa.edu.cn; Yi Huang, yihuang@buaa.edu.cn; Haonan Yuan, yuanhn@buaa.edu.cn, Beihang Univeristy, Beijing, China; Xingcheng Fu, fuxc@gxnu.edu.cn, Guangxi Normal University, Guilin, China; Yisen Gao, ygaodi@cse.ust.hk, The Hong Kong University of Science and Technology, Hong Kong, China; Jia Wu, jia.wu@mq.edu.au, Macquarie University, Sydney, Australia; Shujian Yu, s.yu3@vu.nl, Vrije Universiteit Amsterdam, Amsterdam, The Netherlands; Angsheng Li, angsheng@buaa.edu.cn; Jianxin Li, lijx@buaa.edu.cn, Beihang Univeristy, Beijing, China; Philip S. Yu, psyu@uic.edu, Department of Computer Science, University of Illinois, Chicago, Chicago, USA.

how do GML models navigate the information bottleneck to learn compressed yet informative representations from graph data? It is precisely at this intersection of open challenges and architectural diversity that information theory offers a transformative perspective. As a mathematical framework for quantifying information, uncertainty, and complexity, information theory provides a set of rigorous, domain-agnostic tools to formalize and address these questions. Core ideas from information theory naturally capture key challenges in GML: *how information flows, how neighborhood data is compressed, and how dependencies between distant graph components are quantified.*

Information theory [114] has a profound history and wide-reaching impact, providing an effective framework for analyzing large-scale data [67, 68]. Originally developed to study optimal communication and encoding, it has been extended to a broad spectrum of scientific and engineering disciplines, including statistical inference [55], cryptography [133], neurobiology [130, 134], and quantum computing [146], etc. At its core, information theory formalizes how information can be efficiently represented, compressed, transmitted, and inferred, closely aligning with the foundational objectives of deep learning. Over the past few decades, these two fields have maintained a close relationship in their development. Consequently, many theories and tools from information theory have been utilized in the development of deep learning and emerging AI topics [104], such as feature extraction [125, 135], statistical inference [25], and Explainable AI [168].

Following the immense success of information-theoretic methods in applying fields such as Computer Vision [31, 36] and Neural Language Processing [23, 141], there has been increasing interest in applying information-theoretic methods to the structured graph data recently. Unlike images or text, graph data is characterized by its non-Euclidean topology, inherent sparsity, and complex, multi-scale statistical dependencies. Analyzing and extracting the information from the complex node features and irregular structure is fundamental in GML, and important efforts have been invested in developing models to obtain informative representations [66, 151]. The information in the graph data is distributed across micro (node/edge) to macro (subgraph/graph) scales, making its preservation and efficient processing a more intricate problem. Information theory provides natural tools to address these challenges, including entropy to measure graph complexity, mutual information to preserve key relationships, and information-based principles to guide model design. Recent studies have explored and validated these tools in GML tasks such as node classification, link prediction, and community detection, thereby validating the potency of this intersection. However, despite rapid progress, no comprehensive survey yet systematically organizes this emerging work within a coherent framework.

Major differences and contributions. Several works have reviewed graph learning, information theory, and the application of information theory in deep learning, respectively. Yu et al. [173] reviewed fundamental information-theoretic measures and learning principles along with their practical usages in deep learning, and highlighted graph learning as an emerging research direction. Goldfeld and Polyanskiy [43] systematically surveyed the information-theoretic foundations of the Information Bottleneck (IB) principle and its role in analyzing deep neural networks, while Hu et al. [52] further reviewed more recent applications of IB in deep learning. However, none of these works provide a systematic summary of the role of information theory in graph learning. The integration of information theory and graph learning encompasses various complex theories and technical approaches, necessitating a comprehensive review of the core ideas, methodologies, and achievements in existing research. In this survey, we dive deeper into the paradigms of advancing information theory to GML by systematically analyzing how information-theoretic principles guide model design, optimization objectives, and interpretability in various graph learning scenarios.

Organization. The content of this paper is organized as follows. In Section 2, we introduce the background knowledge for information theory and graph learning. In Section 3, we summarize the current information-theoretic GML methods, starting from basic uncertainty quantification, moving through relational analysis, and culminating in principled model design, as further illustrated in Sections 4, 5, and 6. Section 7 introduces other advanced research focusing on specific applications. Finally, we summarize the main challenges and future directions for advancing the information theory for GML in Section 8 while concluding the survey in Section 9.

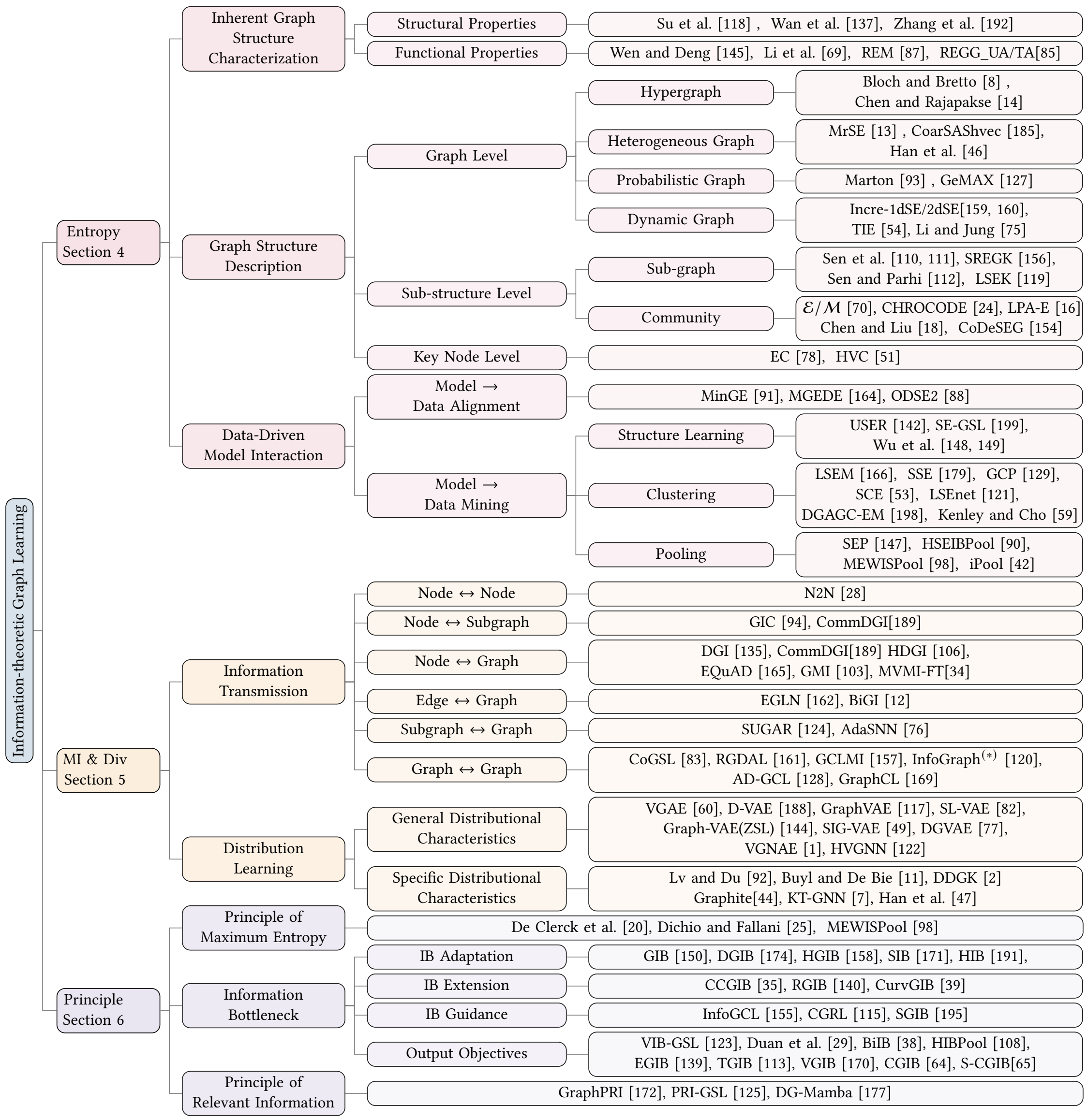

Fig. 1. A taxonomy of Information-theoretic Graph Machine Learning.

Table 1. Notations.

	Symbol	Description
Data	$G = \{\mathbf{A}, \mathbf{X}, V, E\}$	A graph with the adjacency matrix $\mathbf{A} \in \mathbb{R}^{n\times n}$, node feature matrix $\mathbf{X} \in \mathbb{R}^{n\times d}$, node set V and edge set E
	$\mathbf{Y}$	Target
	n	Number of nodes
	d_i	Degree of the i-th node
	D	Degree matrix $D = diag(d_1, d_2, \cdots, d_n)$
	L	Graph laplacian matrix
	S, T	Random variable
Distribution	$P = \{p_1, p_2, \cdots, p_n\}$	The probability set of discrete events
	$p(\cdot,\cdot)$	The joint probability
	$f(\cdot), g(\cdot)$	Probability density function
Entropy	$H_S(\cdot)$	Shannon Entropy
	$H_{R,\alpha}(\cdot)$	Rényi's α-order Entropy
	$\mathcal{H}_{vN}(\cdot)$	von Neumann Graph Entropy
	$\mathcal{T}$	Encoding tree
	$\mathcal{H}^K(\cdot)$	K-dimensional structural entropy
	$\mathcal{H}_E(\cdot)$	Edge Entropy
	$\mathcal{H}_k(\cdot,\cdot)$	Körner Graph Entropy
	$\rho_{\mathcal{P}}(\cdot)$	Normalized Residual Entropy
Mutual Information	$I(\cdot,\cdot)$	Mutual information for two random variables
	$I_\alpha(\cdot,\cdot)$	α-order Mutual information (α-relative entropy)
Divergence	$D_{KL}(\cdot,\cdot)$	Kullback-Leibler divergence between two density functions
	$D_{JS}(\cdot,\cdot)$	Jensen-Shannon divergence between two density functions
	$D_\alpha(\cdot,\cdot)$	α-divergence between two density functions
Learning	$\mathrm{GNN}(\cdot)$	Graph neural network
	Θ	Parameter space of the model
	$\mathbf{Z}$	Latent representation
	$h_i/e_i/z_i/g_i$	Node/Edge/Sub-graph/Graph representation
	l_i	The i-th hidden layer's feature representation
	$\mathcal{R}(\cdot)$	Readout function

2 Background of Graph Machine Learning and Information Theoretical Foundations

In this section, we provide a brief introduction to graph machine learning and provide an overview of commonly used quantities in information theory, including fundamental measures such as entropy, mutual information, and divergence, as well as popular information principles such as the Principle of Maximum Entropy, Information Bottleneck, and the Principle of Relevant Information. The primary notations used in this survey are summarized in Table 1.

2.1 Graph Machine Learning

A graph is denoted $G = \{V, E, \mathbf{X}, \mathbf{A}\}$, where V is the set of n nodes; E is the set of edges connecting pairs of nodes, indicating relationships between them; $\mathbf{X} \in \mathbb{R}^{n*d}$ is the node feature matrix, with each row $\mathbf{X}_i$ corresponding to the feature vector of node v_i; and $\mathbf{A} \in \mathbb{R}^{n*n}$ is the adjacency matrix, with each element $\mathbf{A}_{i,j}$ may also represent edge attributes such as weights and directions.

Graph machine learning is a subfield of machine learning that focuses on analyzing graph data with both feature and structural information. A graph neural network $\mathrm{GNN}(\cdot)$ can map a graph G to its representation $\mathbf{Z}$. With task-oriented loss function $\mathcal{L}$ and trainable parameters $\theta \in \Theta$, the objective of graph machine learning is:

$$\min_{\theta} \mathcal{L}(\mathbf{Z}), \quad \text{s.t. } \mathbf{Z} = \mathrm{GNN}_{\theta\in\Theta}(\mathbf{X}, \mathbf{A}; \theta). \tag{1}$$

Unlike traditional deep learning models based on Euclidean data, GNNs can explicitly capture the dependencies between nodes by using the **message passing** mechanism [152]. At the k-th layer of the GNN, the representation $h_v^{(k)}$ of node v is updated by aggregating messages from its neighbors $\mathcal{N}(v)$:

$$h_v^{(k)} = \text{UPDATE}^{(k)}\left(h_v^{(k-1)}, \text{AGGREGATE}^{(k)}\left(\{h_u^{(k-1)} : u \in \mathcal{N}(v)\}\right)\right). \tag{2}$$

This iterative process of message passing allows local neighborhood information to propagate through the graph, making GNNs particularly effective in tasks involving non-Euclidean data.

Table 2. Comparison and distinctions among various types of entropy.

Entropy	Structure	Key component*	Characteristic
$H_S(\cdot)$	✗	———	The most classical entropy based on probability.
$H_{R,\alpha}(\cdot)$	✗	———	Using α to control the sensitivity of probability distributions.
$\mathcal{H}_{vN}(\cdot)$	✓	$\{\lambda_i\}_{i\in\{1,\cdots,n\}}$	Extracting structural information via Laplacian eigenvalues.
$\mathcal{H}^K(\cdot)$	✓	$\mathcal{T}$	Mining the optimal hierarchical structure in the graph
$\mathcal{H}_E(\cdot)$	✓	$\{p_{ij}\}_{i,j\in\{1,\cdots,M\}}$	Calculating edge entropy using node classes of the edge.
$\mathcal{H}_k(\cdot,\cdot)$	✓	$\text{VP}(\cdot)$	Calculating graph entropy using independent sets.
$\rho_{\mathcal{P}}(\cdot)$	✓	$\mathcal{P}$	Calculating the information carried by partition $\mathcal{P}$.

*The key component represents the variables that reflect the graph information in the calculation.

2.2 Entropy

Entropy, a core concept in information theory, serves as a fundamental measure of uncertainty and disorder in data. In this section, we present various entropy definitions, including the classical Shannon's entropy measurements, graph-specific entropy measurements, and their connections. The characteristics of these entropy measures are summarized in Table 2.

2.2.1 *Classical Entropy.*

Classical entropy measures typically assume independent and identically distributed (i.i.d.) samples and do not account for the structural dependencies inherent in graph data. Well-known measures include the *Shannon's Entropy*, *Joint Entropy*, and *Rényi's α-order Entropy*.

Shannon Entropy [114] quantifies the fundamental uncertainty or average information content of a random variable. For a discrete distribution $P = \{p_i\}$, it is defined as

$$H_S(P) = -\sum_i p_i \log p_i, \tag{3}$$

while for a continuous random variable with density $p(x)$, it is given by the differential entropy

$$H_S(X) = -\int p(x) \log p(x) dx. \tag{4}$$

Joint Entropy [114], as an extension to a pair of random variables, measures the overall uncertainty of their joint system. For discrete variables (S, T), it is defined as

$$H_S(S,T) = -\sum_{i,j} p(i,j) \log p(i,j), \tag{5}$$

and for continuous variables (S, T) with joint density $p(s, t)$, it is defined as

$$H_S(S, T) = -\int\int p(s, t) \log p(s, t) ds dt. \tag{6}$$

Rényi's α-order Entropy [107] is a parametric generalization of Shannon Entropy that introduces a tunable parameter $\alpha > 0$ to control the sensitivity of the measure to different probability distributions. By varying α, one can emphasize or suppress the contribution of particular probabilities. The α-order Rényi Entropy $H_{R,\alpha}(\cdot)$ is defined as:

$$H_{R,\alpha}(X) = \frac{1}{1-\alpha} \log \sum_i p_i^\alpha; \quad \lim_{\alpha \to 1} H_{R,\alpha}(X) = H_S(X). \tag{7}$$

2.2.2 Graph-Specific Entropy. Brooks [10] described quantifying the information embedded in structures as a "grand challenge" that has persisted for fifty years of computer science. Motivated by this challenge, researchers began exploring entropy formulations that directly capture graph structural properties. To emphasize this structural perspective, we denote the graph-specific entropy by $\mathcal{H}$ rather than H.

von Neumann Graph Entropy (VNGE) [101], derived from the von Neumann entropy [136] in quantum mechanics, defines graph entropy through the spectral distribution of the normalized Laplacian. For an undirected graph G with Laplacian matrix $L = D - A$, the corresponding density matrix is defined as $\rho = L/\text{Tr}(L)$. Let $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_n$ be the eigenvalues of L, where $\text{vol}(G) = \sum_{i=1}^n \lambda_i = \text{Tr}(L)$. The VNGE is then given by:

$$\mathcal{H}_{vN}(G) = -\text{Tr}(\rho \log \rho) = -\sum_{i=1}^{n} \left(\frac{\lambda_i}{\text{vol}(G)} \log \frac{\lambda_i}{\text{vol}(G)} \right). \tag{8}$$

Structure Entropy [71] establishes a foundational framework for structural information theory by representing a graph G using a hierarchical encoding tree $\mathcal{T}$ and quantifying the uncertainty inherent in this hierarchical structure, as shown in Fig. 2. It is particularly useful for graphs with hierarchical structures, allowing for more effective encoding through hierarchical partitioning. Based on this, the structural entropy of the encoding tree $\mathcal{T}$ can be defined as follows:

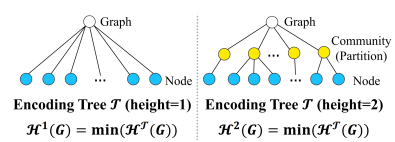

Fig. 2. Illustration of the encoding tree.

$$\mathcal{H}^{\mathcal{T}}(G) = \sum_{\alpha \in \mathcal{T}, \alpha \neq \lambda} \mathcal{H}^{\mathcal{T}}(G; \alpha) = \sum_{\alpha \in \mathcal{T}, \alpha \neq \lambda} -\frac{g_\alpha}{2|E|} \log_2 \frac{V_\alpha}{V_{\alpha^-}}, \tag{9}$$

where g_α denotes the number of edges connecting nodes in $\mathcal{T}_\alpha$ to nodes outside $\mathcal{T}_\alpha$, α^- denotes the parent node of α, and V_β is the volume of the node set $\mathcal{T}_\beta$, *i.e.*, the sum of the degrees of all nodes in $\mathcal{T}_\beta$. Then, the K-dimensional structure entropy with K specifying the height of the encoding tree can be defined as:

$$\mathcal{H}^K(G) = \min_{\{\mathcal{T} | height(\mathcal{T}) = K\}} \mathcal{H}^{\mathcal{T}}(G). \tag{10}$$

Edge Entropy [56] describes graph entropy from the perspective of edges. It can be defined with the probability p_{lm} that a node v belongs to class l:

$$p_{lm} = \frac{|\{\text{edge } e : \text{start}(e) \in V_l \wedge \text{end}(e) \in V_m\}|}{|\{\text{edge } e : \text{start}(e) \in V_l\}|}, \tag{11}$$

$$\mathcal{H}_E(G) = \sum_{l \in \{1,\ldots,M\}} \mathcal{H}_e(l) w_l = \sum_{l \in \{1,\ldots,M\}} \left(-\sum_{m \in \{1,\ldots,M\}} p_{lm} \log_M(p_{lm}) \right) w_l, \tag{12}$$

where $e \in E$ is a non-self-loop edge, V_l is the set of nodes in class l, and w_l denotes its proportion.

Körner Graph Entropy [61] quantifies the uncertainty in a probabilistic graph (G, P), where P denotes the probability distribution across its nodes. It is defined as:

$$\mathcal{H}_k(G, P) = \min_{\boldsymbol{a} \in \mathrm{VP}(G)} \sum_{i=1}^{|G|} -p_i \log(a_i), \tag{13}$$

where p_i is the probability density of node i and $\boldsymbol{a}$ lies within the node-packing polytope $\mathrm{VP}(G)$. Sun et al. [127] extended this formulation by introducing the indicator matrix $\boldsymbol{B}$ of independent sets in G. The node-packing polytope $\mathrm{VP}(G)$ is then defined as the set of vectors $\boldsymbol{a} = \boldsymbol{B}\boldsymbol{\lambda}$, with $\boldsymbol{\lambda} \geq 0$ and $\sum_i \lambda_i = 1$. This extension incorporates graph structure through independent sets, offering a refined way to characterize entropy in probabilistic graphs.

Residual Entropy [87] measures the information gain introduced by applying a partition $\mathcal{P}$ to graph G. Specifically, considering that $\mathcal{H}^1(G)$ denotes the 1-dimensional structure entropy of G under a basic "reference partition" and $\mathcal{H}^{\mathcal{P}}(G) = \mathcal{H}^{\mathcal{T}}(G)$, where $height(\mathcal{T}) = 2$, denotes the structure entropy when the partition $\mathcal{P}$ forms the hierarchical structure of the encoding tree on G. Thus, the reduction in entropy from $\mathcal{H}^1(G)$ to $\mathcal{H}^{\mathcal{P}}(G)$ captures the additional structural information revealed by partition $\mathcal{P}$. **Normalized Residual Entropy** $\rho_{\mathcal{P}}(G)$ is then defined as:

$$\rho_{\mathcal{P}}(G) = \frac{(\mathcal{H}^1(G) - \mathcal{H}^{\mathcal{P}}(G))}{\mathcal{H}^1(G)}. \tag{14}$$

2.2.3 Relationships between different entropies. Table 2 shows that traditional entropy measures exhibit only weak connections to graph structure, whereas graph-specific entropies explicitly encode structural patterns through encoding tree, Laplacian matrix, and related graph operators.

In particular, VNGE, grounded in quantum information theory, defines a graph entropy that directly operates on the Laplacian matrix. However, VNGE is computationally intensive [17], as it requires the full eigenspectrum of the Laplacian. Consequently, several works [17, 167] have proposed approximation methods to reduce computational costs while maintaining comparable effectiveness. Moreover, Liu et al. [86] investigated the relationship between 1-dimensional structural entropy and VNGE, both theoretically and empirically, by analyzing the entropy gap $\Delta\mathcal{H}(G) = \mathcal{H}^1(G) - \mathcal{H}_{vN}(G)$, and established the bound:

$$0 < \Delta\mathcal{H}(G) \leq \frac{\log_2 e}{\delta} \cdot \frac{\mathrm{Tr}(A^2)}{\mathrm{vol}(G)}, \tag{15}$$

where $\delta = \min\{d_i | d_i > 0\}$ is the minimum positive degree. It indicates that 1-dimensional structural entropy serves as an approximation of VNGE and is linearly correlated with it [86]. This highlights the distinction between spectral-based and node-based perspectives of graph information.

2.3 Divergence

Divergence measures quantify the statistical discrepancy between probability distributions. Well-known examples include *Kullback-Leibler divergence*, *Jensen-Shannon divergence*, and *α-divergence*.

Kullback-Leibler (KL) divergence [62] between two probability distributions f, g is:

$$D_{KL}(f||g) = \int_T f(t) \log \frac{f(t)}{g(t)} dt = -\int_T f(t) \log g(t) dt - H(f) = H_{CE}(f; g) - H(f), \tag{16}$$

where $H(f)$ denotes the Shannon entropy of f and $H_{CE}(\cdot;\cdot)$ denotes the cross-entropy between f and g. Notably, KL divergence is asymmetric, meaning ($D_{KL}(f||g) \neq D_{KL}(g||f)$).

Jensen-Shannon (JS) divergence [81] is a symmetric and smoothed version of the KL divergence, defined as:

$$D_{JS} = \frac{1}{2}D_{KL}\left(f||m\right) + \frac{1}{2}D_{KL}\left(g||m\right), \tag{17}$$

where $m = \frac{f+g}{2}$ is a mixture distribution of f and g. JS divergence inherits the information-theoretic properties of KL divergence while being symmetric and always finite.

α**-divergence** [107], introduced by Alfred Rényi, generalizes the KL divergence with an adjustable parameter α:

$$D_\alpha(f||g) = \frac{1}{\alpha-1}\int_T f^\alpha(t)g^{1-\alpha}(t)dt. \tag{18}$$

As $\alpha \longrightarrow 1$, D_α reduces to the standard KL divergence.

2.4 Mutual Information

Mutual Information (MI) measures the amount of information shared between two random variables. It quantifies the reduction in uncertainty of one variable given knowledge of the other.

2.4.1 Definition of Mutual Information. Mathematically, the **Mutual Information** [114] between two random variables (S,T) is defined as the KL divergence between their joint distribution $f(s,t)$ and the product of their marginals $f(s)f(t)$:

$$\begin{aligned} I(S;T) = D_{KL}(f(s,t)||f(s)f(t)) &= -\int_S\int_T f(s,t)\log\frac{f(s,t)}{f(s)f(t)}dsdt \\ &= H(S) - H(S|T) = H(T) - H(T|S) = H(S,T) - H(S|T) - H(T|S), \end{aligned} \tag{19}$$

where $H(\cdot)$ and $H(\cdot|\cdot)$ denote entropy and conditional entropy, respectively. MI is symmetric ($I(S;T) = I(T;S)$) and equals zero if and only if S and T are statistically independent.

α**-order MI** [3] generalizes $I(S;T)$ by replacing the KL divergence with the α-divergence:

$$I_\alpha(S;T) = D_\alpha(f(s,t)||f(s)f(t)). \tag{20}$$

2.4.2 Estimator of Mutual Information. Estimating MI from observational data is a fundamental challenge in both information theory and machine learning, particularly in high-dimensional settings. Besides non-parametric estimators, such as the Rényi quadratic entropy kernel estimator [104] and the matrix-based Rényi entropy estimator [79], recent studies typically adopt variational approaches, such as MINE [6], that leverage contrastive learning and noise-contrastive estimation [45] to estimate MI in a parametric manner. A discriminator $\mathcal{D}$ is trained to distinguish samples from the joint distribution $p(x,y)$ (positive pairs, $\langle\text{pos}\rangle = (x,y)$) and samples from the product of marginals $\tilde{p} = p(x)p(y)$ (negative pairs, $\langle\text{neg}\rangle = (x,y')$, with $y' \sim p(y)$). MI estimation is then reduced to optimizing objectives as follows:

- BCE (Binary Cross-Entropy):

$$\mathbb{E}_p\left[\log\mathcal{D}(\langle\text{pos}\rangle)\right] + \mathbb{E}_{\tilde{p}}\left[\log(1-\mathcal{D}(\langle\text{neg}\rangle))\right]. \tag{21}$$

- InfoNCE [100]:

$$-\mathbb{E}_p\left[\log\frac{e^{\mathcal{D}(\langle\text{pos}\rangle)/\tau}}{\sum e^{\mathcal{D}(\langle\text{pair}\rangle)/\tau}}\right]. \tag{22}$$

- JSD (Jensen-Shannon MI estimator) [99]:

$$-\mathbb{E}_p\left[\text{sp}(-\mathcal{D}(\langle\text{pos}\rangle)\right] - \mathbb{E}_{\tilde{p}}\left[\text{sp}(\mathcal{D}(\langle\text{neg}\rangle)\right], \quad \text{sp}(z) = \log(1+e^z). \tag{23}$$

- MINE (mutual information neural estimator) [6]:

$$\mathbb{E}_{p}\left[\mathcal{D}(\langle \text{pos} \rangle)\right] - \log\left(\mathbb{E}_{\tilde{p}}\left[e^{\mathcal{D}(\langle \text{neg} \rangle)}\right]\right). \tag{24}$$

Additionally, from a graph perspective, Escolano et al. [30] integrated manifold alignment with copula-based entropy estimators to efficiently estimate MI between graphs. By modeling an information channel and quantifying structural correlations through MI, their method provides a more accurate and principled measure of graph similarity.

2.5 Fundamental Information Principles

Information-theoretic principles provide a rigorous foundation for graph learning, guiding the design of models that effectively capture structural patterns, preserve relevant information, and improve robustness and generalization across tasks. In the following, we focus on three representative principles: *Maximum Entropy*, *Information Bottleneck*, and *the Principle of Relevant Information*.

2.5.1 Principle of Maximum Entropy. **Principle of Maximum Entropy (PME)** [55] states that, among all probability distributions P that satisfy a given set of constraints, the optimal distribution P^* is the one that maximizes entropy, thereby introducing the least additional assumptions beyond the known information:

$$P^* = \arg\max_{P \in \mathcal{P}} H(P). \tag{25}$$

In graph learning, PME provides a principled way to construct probabilistic graph models by maximizing uncertainty while enforcing constraints derived from observed node features, edge statistics, or topological properties, thus avoiding bias.

2.5.2 Information Bottleneck. The Information Bottleneck (IB) framework [131] provides a principled approach for balancing the trade-off between information preservation and compression. Given $x \in \mathbf{X}$, $y \in \mathbf{Y}$, and a latent representation $z \in \mathbf{Z}$ defined by $p(z|x)$ under the Markov chain $\mathbf{Y} \to \mathbf{X} \to \mathbf{Z}$, the IB objective seeks a representation $\mathbf{Z}$ that compresses $\mathbf{X}$ by $\min I(\mathbf{Z};\mathbf{X})$ ("minimal") while preserving task-relevant information by $\max I(\mathbf{Z};\mathbf{Y})$ ("sufficient").

$$\min_{p(z|x)} I(\mathbf{Z};\mathbf{X}) - \beta I(\mathbf{Z};\mathbf{Y}). \tag{26}$$

where the Lagrange multiplier β controls the degree of compression of $\mathbf{Z}$.

In deep learning, numerous studies [58, 109, 116, 132] have explored IB by examining the optimization process, learning dynamics, compression effects, and generalization abilities. In the field of GML, the traditional IB has been extended to graph-specific variants [150, 174], which account for both node attributes and structural information.

2.5.3 Principle of Relevant Information. The Principle of Relevant Information (PRI) [104] advocates learning representations that balance the complexity of the encoded information with its utility in representing the informative aspects of the data. Specifically, PRI trades off between the entropy of the learned representation $H(\mathbf{Z})$ and its descriptive power regarding $\mathbf{X}$, quantified by a divergence $D(\mathbf{Z}\|\mathbf{X})$:

$$\min_{p(z|x)} H(\mathbf{Z}) + \beta D(\mathbf{Z}||\mathbf{X}), \tag{27}$$

where β controls the amount of relevant information that $\mathbf{Z}$ extracts from $\mathbf{X}$. Notably, PRI does not require a relevant auxiliary variable $\mathbf{Y}$ and optimize directly on the original data $\mathbf{X}$.

3 Information Theory Meets Graph Machine Learning: Framework and Taxonomy

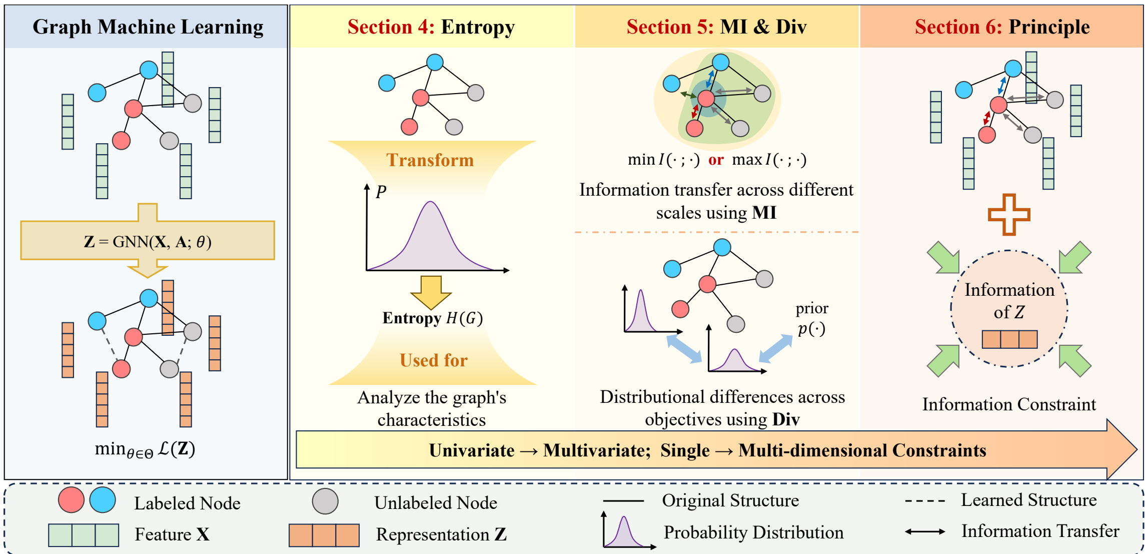

Fig. 3. We first define graph machine learning intuitively as extracting features from graph data to minimize task loss. From an information-theoretic view, we then explore three aspects: entropy for data characteristics, mutual information and divergence for information flow, and model design guided by information constraints.

This survey organizes methodologies into a hierarchical framework as shown in Fig. 3, progressing from basic uncertainty quantification to relational analysis and principled model design, supporting more effective, interpretable, and robust graph machine learning systems.

Entropy as the base metric to assess the intrinsic uncertainty and complexity of graph (Section 4). Entropy is a fundamental measure in information theory that quantifies the uncertainty of a random variable. In GML, entropy-based methods assess the intrinsic uncertainty and complexity of graph features, including node attributes, structural properties, and global characteristics. Graph data is typically more complex than regular data such as images or text, owing to intricate node interdependencies, dynamic structural variations, and multi-relational patterns. Leveraging entropy provides tools to quantify and manage such variability, enabling a deep understanding of informational content. Consequently, deeply examining the complexity, variability, and inherent randomness of graph data becomes essential for advancing effective GML models.

Capturing relationships with mutual information and divergence (Section 5). Building on entropy, mutual information quantifies the dependencies between variables in a graph, capturing relationships across structural hierarchies. Divergence measures assess dissimilarities between probability distributions, offering a complementary perspective on how different components of a graph vary. Given the hierarchical and complex nature of graph data—from local connections to global topology—analyzing these dependencies and divergences is essential. Together, these measures enable GML models to capture and preserve both the interdependencies and discrepancies inherent in graph data, leading to more informative representation learning.

Guiding model design via information principles (Section 6). At the highest level, information principles such as the Principle of Maximum Entropy, the Information Bottleneck, and the Principle of Relevant Information, integrate insights from entropy, mutual information, and divergence. These principles provide a rigorous foundation for designing GML models, guiding how representations should be compressed, regularized, and made both discriminative and robust. Specifically, they promote unbiased learning (Maximum Entropy), compress task-irrelevant features (Information Bottleneck), and balance compression with the preservation of relevant information (Principle of Relevant Information). Given the broad applicability of GML models, these principles offer a unified and general perspective for guiding the learning process while balancing model complexity, accuracy, and generalization.

4 Quantifying Intrinsic Uncertainty and Complexity: Entropy as the Base Metric

For graph data, a substantial amount of critical information is encapsulated not only in the node features but also in the structure of the graph itself, as it inherently encodes the fundamental relational connections among nodes [153]. At the foundational level, entropy, as a key measure of data uncertainty and information content [97], provides an initial understanding of the intrinsic complexity and uncertainty present in graph data by quantifying the distributions of nodes and edges. As illustrated in Fig. 4, this section demonstrates how entropy functions as a fundamental metric for quantifying uncertainty and complexity in graph machine learning from three aspects: capturing the basic characteristics, describing the underlying structural information and analyzing graphs with data-driven models.

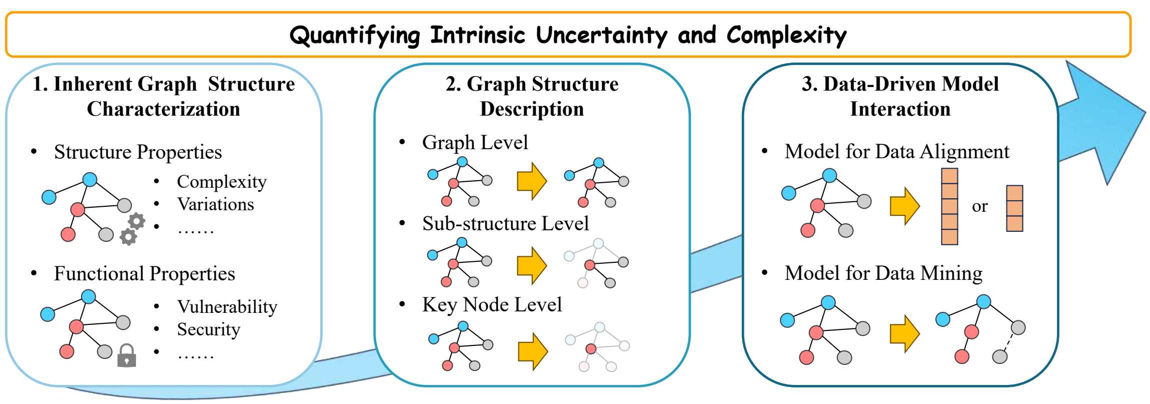

Fig. 4. Overview of Entropy as a Metric for Measuring Uncertainty and Complexity.

4.1 Inherent Graph Structure Characterization

In the context of graphs, entropy provides a principled way to quantify structural variability, including the diversity of node connections, the distribution of edge weights, and the overall complexity of the topology. By characterizing how information is distributed across the graph's structure, entropy can reveal important structural patterns, providing valuable insights for understanding, analyzing, and optimizing graph structures.

4.1.1 Structural Properties. Quantifying structural properties is central to the study of complex networks, as it provides insight into how topological patterns shape system behavior. Recent works have increasingly emphasized the use of spectral and combinatorial entropy as tools for capturing structural characteristics across multiple scales.

A major line of progress involves leveraging spectral entropy to assess topological diversity. Su et al. [118] systematically analyzed spectral entropy measures based on the eigenvalues of Laplacian and adjacency matrices,such as

BGS [9], EE [33], and DD [21], demonstrating that spectral entropy better capture global structural variations more effectively than traditional distribution-based metrics. Complementing this global perspective, Wan et al. [137] explored combinatorial entropy to characterize graph complexity through the distribution of independent sets and matchings. Their proposed NIS (Number of Independent Sets) and NM (Number of Matchings) entropy measures provide a novel perspective on structural richness based on subgraph enumeration.

Despite progress in global and local characterizations, a mesoscopic structural entropy has been absent. To fill this gap, Zhang et al. [192] introduced the clustering coefficient structural entropy, which incorporates clustering patterns into entropy calculations to more precisely quantify network complexity. Moreover, the Structural Entropy Ratio ($\Delta\mathcal{S}$) provides a robust indicator of structural stability for analyzing dynamic evolution.

4.1.2 Functional Properties . Functional properties, such as *vulnerability*, *security*, and *resilience*, reflect a graph's ability to withstand disruptions or attacks. These properties are fundamentally tied to the underlying topology, connectivity patterns, and community structure, offering critical insights into how graphs maintain stability and recover under external perturbations.

To improve the assessment of graph vulnerability, Wen and Deng [145] introduced an entropy-driven approach by embedding entropy-derived features into the vulnerability assessment process. This process uncovers subtle structural weaknesses and offers a more comprehensive characterization of community vulnerability. Extending this line of inquiry, Li et al. [69] investigated network security by introducing two novel metrics: Resistance and the Security Index, both derived from formulations grounded in graph entropy defined in Eq. (28). Experimental results demonstrate that higher values of the Security Index correspond to stronger resistance against virus propagation and structural attacks.

$$\text{Resistance}: \ \mathcal{R}(G) = \mathcal{H}^1(G) - \mathcal{H}^2(G), \quad \text{Security Index}: \ \rho(G) = \frac{\mathcal{R}(G)}{\mathcal{H}^1(G)}. \tag{28}$$

The functional role of entropy is also demonstrated in the context of attack and defense strategies. REM [87] exploits residual entropy to manipulate perceived community structures, enabling effective community deception by minimizing this entropy measure. In contrast, REGG_UA/TA [85] adopts the opposite strategy that maximizing residual entropy to enhance structural robustness, thereby enhancing a network's resilience to disruptions.

4.2 Graph Structure Description

With the development of graph neural networks, effectively extracting and representing structural information has become increasingly important. Beyond leveraging inherent node or edge attributes, modern graph models require a deeper understanding of the latent structural patterns embedded within complex networks. This necessitates analyzing structure at multiple levels of granularity—including global graph topology, localized substructures, and critical individual nodes, as illustrated in Fig. 5.

4.2.1 Graph Level. While entropy captures intrinsic structural characteristics, it also reveals how graphs are organized. Specifically, graph data appear in diverse forms such as *hypergraphs*, *heterogeneous graphs*, *probabilistic graphs*, and *dynamic graphs*, each with distinct structural assumptions and modeling challenges as illustrated in Fig. 6. Therefore, based on the characteristics and requirements of different scenarios, researchers have extended graph entropy, particularly Structure Entropy [71], to develop tailored formulations for various graph types.

Hypergraph. Hypergraphs generalize classical graphs by allowing each hyperedge to connect multiple nodes, making them effective for modeling multi-way relationships in complex systems. Bloch and Bretto [8] introduce a

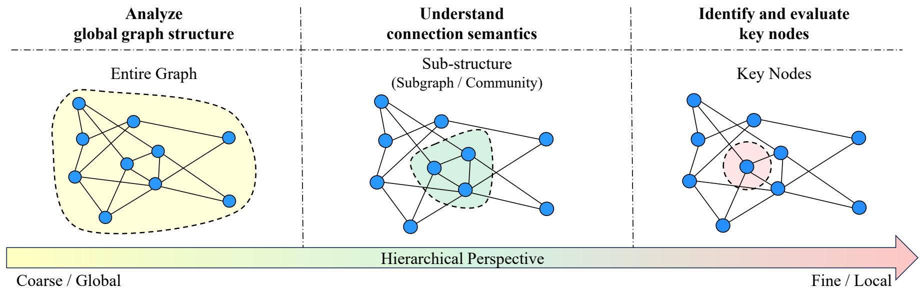

Fig. 5. Graph Structure Mining under Different Levels

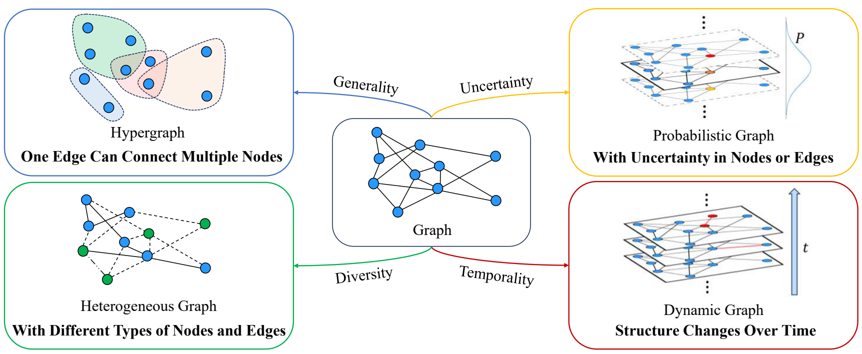

Fig. 6. The Relationship Between Different Graph Types

hypergraph entropy that decomposes a hypergraph into partial components to form an entropy vector, enabling fine-grained characterization of structural diversity. Complementarily, Chen and Rajapakse [14] develop a tensor-based entropy for uniform hypergraphs by extending von Neumann entropy using tensor formulations. It not only reflects the regularity of uniform hypergraphs but also exhibits well-defined bounds and robustness properties.

Heterogeneous Graph. Heterogeneous graphs contain multiple types of nodes and edges, enabling the modeling of diverse relationships in applications such as recommendation systems, knowledge graphs, and biological networks [15, 138]. Han et al. [46] propose a simplified version of von Neumann entropy and analyze its relationship with the Heterogeneity Index [32]. MrSE [13] extends structural entropy to the multi-relational setting through a random-surfing formulation, relaxing the classical single-relation assumption. Similarly, CoarSAShvec [185] further incorporates entropy-based principles into embedding methods for heterogeneous graphs, enhancing representation learning and broadening entropy's utility in heterogeneous graph analysis.

Probabilistic Graph. A probabilistic graph is defined as a graph (G, P), where a probability distribution P is assigned to the node set V. This formulation enriches classical graphs by incorporating uncertainty, yielding more flexible and expressive structures. Marton [93] provide a new perspective on the Shannon capacity of probabilistic graphs by integrating probability distributions with orthogonal representations. This work proposes a new upper bound and uncovers a deep connection between probabilistic graph capacity and graph entropy, extending Lovász's classical results [89]. Building on this foundation, GeMAX [127] combines orthogonal representation with the probabilistic graph entropy concept introduced by Korner et al. [61], enabling effective estimation of Körner's Graph Entropy.

Dynamic Graph. A dynamic graph is a sequence of temporally indexed graph states evolving over time [48]. Unlike other graphs, dynamic graphs incorporate a temporal dimension, substantially increasing structural complexity. Incre-1dSE/2dSE [159, 160] introduces Global Invariant and Local Difference, enabling efficient updates of 1- and 2-dimensional structural entropy via dynamic adjustments of the coding tree, rather than recomputing entropy from scratch. Focusing on the irregularity of temporal interactions, Huang et al. [54] introduce Temporal Information Entropy (TIE), which quantifies unpredictability in node interaction over time, where higher entropy values indicate more bursty or irregular behaviors. Li and Jung [75] propose a dynamic graph embedding model based on entropy to make temporally similar graphs closer in the embedding space, facilitating anomaly detection in temporal graph sequences.

4.2.2 *Sub-structure Level.*

While graph-level entropy captures global structural information, a finer-grained analysis requires examining localized substructures within the graph. This shift from global to local perspectives enables models to capture subtle connectivity patterns, highlighting two key substructures: *subgraphs* and *communities*.

Subgraph. A subgraph is a subset of nodes and edges within a large graph. Subgraph-level analysis is fundamental for capturing localized structural patterns, enabling reveal meaningful patterns for applications such as brain region detection and social network analysis. Subgraph entropy was first introduced by Sen et al. [111] to quantify the structural information encoded in the subgraphs induced by individual nodes. Extending this idea, Sen et al. [110] applied subgraph entropy to identify critical brain regions and functional components in clinical datasets. To provide a more rigorous understanding, Sen and Parhi [112] further analyzed key properties of subgraph entropy, deriving upper and lower bounds for subgraph entropy and showing through experiments that it is more stable than common centrality measures (*e.g.*, strength, eigenvector, betweenness). Considering computational efficiency, SREGK [156] uses a kernel method to efficiently compute Rényi's entropy. Based on this, LSEK [119] develops the labeled subgraph entropy graph kernel via a dynamic programming–based subgraph counting algorithm and a re-derived clustering expansion formula. This method enables fast computation of global graph entropy while integrating both topology and node label semantics.

Community. Discovering communities is important for analyzing complex graph structures, as it uncovers meaningful structural organization within graphs [22]. Traditional approaches primarily rely on topological statistics, such as node degree distributions and clustering coefficients [5, 37, 96]. These methods still have limitations under varying conditions. Recent efforts have turned to information-theoretic formulations of community detection. A representative work is $\mathcal{E}/\mathcal{M}$ [70]. Grounded in the "homophily/kinship" principle, it posits that natural communities emerge as the result of minimizing structural entropy. Another line of work shifts the focus from connectivity to node attribute similarity over connectivity. CHROCODE [24] uses chromatic entropy to detect communities based on shared attributes, addressing the issue that connectivity density may not necessarily correlate with a shared property or purpose. More recently, Propose-Select-Adjust (PSA) framework [18] extends structural entropy to decentralized environments, enabling robust community detection without centralized control. Building on these developments, CoDeSEG [154] formulates community detection as a potential game, leveraging two-dimensional structural entropy for efficient detection of both overlapping and non-overlapping communities with near-linear time complexity. LAP-E [16] improves the Label Propagation Algorithm by incorporating mutual information and a belonging coefficient, enabling the capture of indirect relationships and improving accuracy and stability in sparse or noisy networks.

4.2.3 *Key Node Level.*

Building on structural extraction at the graph and substructure levels, we further explore finer-grained key node components. Identifying and analyzing critical nodes offers essential insights into a graph's functional organization, improving both understanding and the decision-making capability of complex systems.

For instance, EC [78] combines Local Structural Entropy and Clustering Coefficient to capture uncertainty and local heterogeneity as well as to measure neighborhood cohesiveness, effectively identifying nodes that are both structurally distinct and functionally central. ALGE [197] selects key nodes for active learning based on structure relative entropy, quantifying a node's similarity to others across local and global structures. In hypergraphs, HVC [51] leverages von Neumann Entropy to identify important nodes that strongly influence information flow and structural robustness.

4.3 Data-Driven Model Interaction

Shifting from data-centric view to model-centric view, GML emphasizes how models extract, align, and refine graph information. Fundamentally, GML can be viewed as an information optimization problem in graph data space [66], where entropy naturally serves as a guiding measure. We categorize this interaction into two complementary directions: *Model for Data Alignment*, which focuses on entropy-based strategies to configure model parameters that respect underlying graph structures, and *Model for Data Mining*, where entropy supports structure-aware tasks such as encoding, clustering, and pooling.

4.3.1 Model for Data Alignment. In GML, choosing the right model configuration is essential for effectively capturing data characteristics. Recent methods leverage entropy to automatically calibrate key parameters, particularly for embedding dimension selection and model structure design. MinGE [91] applies the minimum entropy principle to node embedding dimension selection, jointly considering feature and structural entropy to compute the most informative and compact embedding dimension for each node. Extending this idea, MGEDE [164] combines attribute entropy with structural entropy to estimate both node-level and graph-level embedding dimensions, providing a more comprehensive entropy-guided dimension calibration strategy. In both cases, minimizing entropy effectively quantifies and reduces uncertainty in graphs. Entropy has also been leveraged to improve data alignment at graph level. ODSE2 [88] employs Rényi entropy–based estimators (QRE and MST) to quantify dissimilarity between graphs. By minimizing this entropy, it adaptively captures structural variations and improves alignment and generalization across graphs.

4.3.2 Model for Data Mining. While data alignment focuses on fitting models to observed structure of graph data, data mining emphasizes uncovering and preserving hidden structural patterns. Entropy-based techniques have recently been central to enhancing models' ability to extract deep, non-obvious signals from graphs.

Structure learning. Structure learning refines graph topology to better reflect intrinsic node relationships, with the aim of uncovering latent patterns beyond observable connections. By optimizing structural entropy, models can reveal hidden regularities and recover faithful latent structures. One representative approach is USER [142], which learns a cleaner graph structure by minimizing structural entropy, thereby suppressing noisy or spurious edges. Similarly, SE-GSL [199] introduces a hierarchical entropy optimization framework, maximizing 1-dimensional structural entropy for local diversity while minimizing high-dimensional structural entropy to enforce global compactness. Beyond structure refinement, entropy-based methods also distill essential graph compositions. Methods such as [148, 149] convert graphs into coding trees via entropy minimization, compressing redundant components and highlighting the most informative substructures. Moreover, Han et al. [47] leverage von Neumann entropy and the Minimum Description Length principle to build generative prototypes of graph data, capturing common substructures and global regularities.

Clustering. Graph clustering aims to reveal the latent structural organization by grouping nodes into clusters that reveal hidden structures, functional modules, and semantic groupings. LSEM [166] minimizes structural entropy to detect optimal communities in directed graphs, effectively identifying coherent sub-networks. SSE [179] extends this to semi-supervised clustering, integrating constraint-based supervision with high- and low-dimensional structural entropy for both partition-based and hierarchical clustering. Building further, GCP [129] employs a 2-dimensional structural entropy to uncover meaningful clusters under privacy constraints, while SCE [53] iteratively minimizes structural entropy with sparse embeddings and coding trees, addressing locality limitations of traditional clustering. Recent works focus on enhancing clustering flexibility and adaptivity. LSEnet [121] proposes Differentiable Structural Information, an entropy-based formulation that enables clustering without preset cluster numbers. Operating in hyperbolic space, it minimizes entropy to jointly capture node features and global structural organization. DGAGC-EM [198] integrates entropy minimization with attention mechanisms to dynamically adjust neighborhood influence. Kenley and Cho [59] leverage node-based entropy reduction to guide node clustering and structural segmentation.

Pooling. Graph pooling compresses groups of nodes into compact representations [63]. Entropy provides a natural criterion to guide information-preserving compression, ensuring that simplified graphs retain the most informative and structurally significant components. SEP [147] introduces a global pooling strategy by minimizing structural entropy to construct a hierarchical coding tree, which naturally reflects graph structure without relying on predefined compression ratios. HSEIBPool [90] improves entropy stability by scoring nodes and pooling only structurally important ones. Beyond coding-tree approaches, pooling can be formulated as information optimization. MEWISPool [98] maximizes graph output entropy to retain high informational fidelity, while iPool [42] leverages neighborhood conditional entropy to retain the most informative nodes, ensuring the coarsened graph preserves key structures and semantic patterns.

In summary, entropy plays a pivotal role in graph data: maximizing entropy promotes structural diversity and preserves information, whereas minimizing entropy uncovers regularities, reduces noise, and extracts essential patterns.

5 Uncovering Dependencies: Capturing Relationships with Mutual Information and Divergence

GML paradigms characterize how information is represented, transmitted, and optimized within graph structures. Central to these paradigms is the balance between *local* and *global* information. Mutual Information governs dependency preservation across graph components, facilitating effective information transmission, while divergence measures provide optimization principles that align *specific* data observations with *general* distributional representations, ensuring robust and generalizable models. Together, these perspectives form a unified framework that advances representation learning, generation, and inference on graphs.

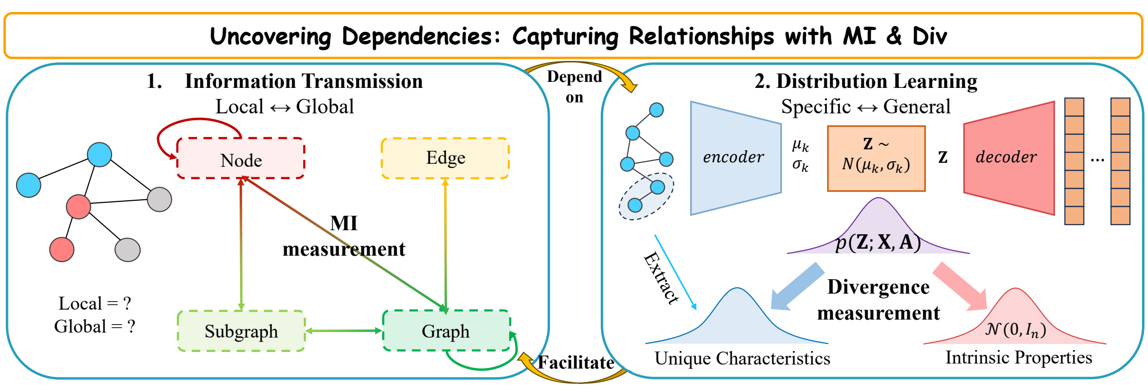

Fig. 7. Overview of Capturing Relationships with Mutual Information and Divergence

5.1 Information Transmission: Local ↔ Global

Mutual Information (MI) serves as a fundamental measure for quantifying how information flow propagates and transmits within graphs both locally and globally. DIM [50] first introduced an unsupervised learning approach based on MI maximization between global and local representations, capturing meaningful features and offering theoretical insights for information-theoretic GML. Following the perspective of multi-scale information flow, as illustrated in Fig. 8, we provide a comprehensive analysis of the mechanisms that enable effective information transmission at both local and global levels. A comparative summary of their characteristics is in Table 3.

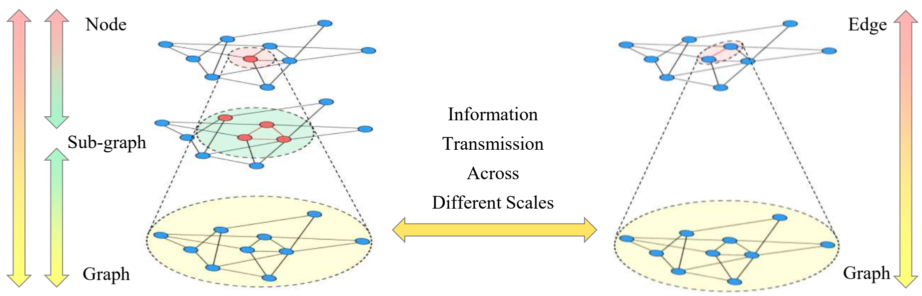

Fig. 8. The Flow of Information Across Different Scales

5.1.1 Node ↔ Node. From the perspective of information exchange between nodes, a key objective is to capture the mutual dependencies between each node and its local context. N2N [28] addresses this by maximizing the MI between a node's hidden representation and that of its surrounding neighborhood. This enables the model to learn structure-aware embeddings without explicit topology augmentation or traditional aggregation.

5.1.2 Node ↔ Subgraph. Building upon the local node-level interactions, subgraphs serve as richer representations of local graph information, bridging the gap between individual nodes and the entire graph. GIC [94] leverages coarse-grained graph information to enhance node embeddings by identifying nodes within the same cluster and maximizing the MI among them. This addresses the limitations of relying solely on either fine-grained local similarities or global similarities, effectively meaningful cluster-level representations. Extending multi-scale information flow, CommDGI [189] models the information exchange within community structures. By maximizing the MI between nodes and their corresponding communities, CommDGI captures complex relationships between nodes and community-level contexts, yielding community-aware embeddings.

5.1.3 Node ↔ Graph. Nodes encode fine-grained local information, whereas the entire graph provides holistic global context. Capturing the information flow between these two levels enhances representation consistency and yields more expressive embeddings. DGI [135] is a pioneering method that leverages MI maximization between node embeddings and a global graph summary vector to learn meaningful node representations. CommDGI [189] extends DGI to community-aware feature learning. HDGI [106] extends DGI to heterogeneous graphs by incorporating meta-paths to capture diverse semantic relationships. Furthermore, EQuAD [165] identifies DGI's ability to preserve invariant graph features and enhances it to improve Out-of-Distribution performance. Considering topology explicitly, GMI [103] jointly maximizes Feature Mutual Information and Topology-Aware Mutual Information, encouraging the preservation of both attribute-level and structural signals. MVMI-FT[34] constructs both a feature view and a topology view, maximizing MI between them to capture more comprehensive node representations that integrate node features and graph structure.

5.1.4 *Edge ↔ Graph.*

Shifting the focus from nodes to edges as fundamental relational units, the interaction between edges and the whole graph highlights how pairwise relations align with the global structural context. EGLN [162] maximizes MI between edge embeddings and the global graph structure, capturing multi-scale relational dependencies that are particularly beneficial in recommendation scenarios. Additionally, BiGI [12] extends this paradigm by modeling MI between edges and global bipartite representations, effectively capturing the global properties of bipartite structures.

5.1.5 *Subgraph ↔ Graph.*

At the graph-level, individual node embeddings insufficient for capturing higher-order structural semantics. Thus, effective information transmission between subgraphs and the entire graph is essential for learning expressive representations. SUGAR [124] introduces a reinforcement learning-based approach to dynamically extract informative subgraphs and reconstruct a "sketched graph" that reveals subgraph-level patterns. A self-supervised MI mechanism is employed to enhance subgraph representations, ensuring they accurately reflect the global graph structure. AdaSNN [76] further proposes a bi-level MI enhancement mechanism that identifies critical subgraphs within the graph and enforces that these subgraphs preserve the structural characteristics of the overall graph.

5.1.6 *Graph ↔ Graph.*

Cross-graph information flow enables models to capture dependencies beyond a single graph, enhancing representation quality and generalization ability. By examining these broader interactions, we reveal how MI transfer between graphs contributes to more effective GML. For supervised learning, CoGSL [83] introduces a compact structure learning framework that optimizes graphs via information compression. For domain adaptation scenarios, RGDAL [161] introduces a robust graph domain adaptation framework that constrains MI to filter out noisy factors, enabling cross-network node classification between source and target graphs. In self-supervised learning, InfoGraph [120] maximizes MI between node- and graph-level representations. Its enhanced variant, InfoGraph* [120], introduces a dual-encoder semi-supervised framework to refines graph-level embeddings using both supervised and unsupervised signals. While effective,InfoMax-style methods may generate redundant and non-discriminative representations [128]. To mitigate this, InfoMin methods such as GraphCL [169] encourage diverse representations by minimizing shared information between views and combining IB constraints to prevent collapse and enhance generalization. However, they may lose task-relevant information by neglecting view consistency [128]. Beyond these extremes, min-max MI frameworks jointly balance redundancy and essential information. AD-GCL [128] introduces adversarial augmentation via a trainable edge-dropping module to minimize redundant information while maintaining essential structure. Similarly, GCLMI [157] applies a min-max MI objective to filter noise, enabling adaptive, information-theoretically grounded graph learning. These approaches reflect a shift toward adaptive, information-theoretically principled self-supervised frameworks that control information flow based on task demands and data characteristics.

The above methods highlight that regulating information flow, particularly between local and global structures, is fundamental in GML. In supervised settings, MI maximization dominates, guided by labels to preserve task-relevant signals. In self-supervised scenarios, balancing information retention and suppression is key for flexible and principled control across views and hierarchical structure scales.

5.2 Distribution Learning: Specific ↔ General

Distribution learning in GML aims to balance specific instance-level characteristics with general distribution properties, enabling models to capture both local details and overarching structure.

Table 3. Comparison of different mutual information mechanisms in graph machine learning methods.

Perspective	Model	MI	Estimator	Description
Node ↔ Node	N2N [28]	$I(h_i, \{h_j\}_{j \in \mathcal{N}_i})$	InfoNCE	Learn node representation by maximizing MI between nodes and neighborhood
Node ↕ Subgraph	GIC [94]	$I(h_i, z_i)$	BCE	Preserve the coarse-grain information in the embedding space
	CommDGI[189]	$I(h_i; z_i)$	InfoNCE	Design a new community detection objective called community MI
Node ↕ Graph	DGI [135]	$I(h_i; \mathcal{R}(\{h_j\}_{j=1}^{n}))$	BCE	Maximize mutual information between node and graph representation
	HDGI [106]	$I(h_i; \mathcal{R}(\{h_j\}_{j=1}^{n}))$	BCE	Maximize MI to learn heterogeneous graph representation
	CommDGI[189]	$I(h_i; \mathcal{R}(\{h_j\}_{j=1}^{n}))$	BCE	Encode node attributes. More in node ↔ subgraph part
	MVMI-FT[34]	$I(h_i; \mathcal{R}(\{h_j\}_{j=1}^{n}))$	JSD	Maximize MI across feature and topology views
	GMI [103]	$I(h_i, g)$	JSD	Calculate MI in terms of features and edges
	EQuAD [165]	$I(h_i; \mathcal{R}(\{h_j\}_{j=1}^{n}))$	InfoNCE	Isolate spurious features via MI–based analysis
Edge ↕ Graph	EGLN [162]	$I(e_i, \mathcal{R}(\{e_j\}_{j=1}^{m}))$	BCE	Design an enhanced graph optimization function
	BiGI [12]	$I(z_i, g)$	BCE	Develop a local-global infomax objective for bipartite graphs
Subgraph ↕ Graph	SUGAR [124]	$I(z_i, \mathcal{R}(\{z_j\}_{j=1}^{n}))$	JSD	Adaptively find the most striking subgraphs using RL method
	AdaSNN [76]	$I(z_i, \mathcal{R}(\{z_j\}_{j=1}^{n}))$	JSD	Detect critical subgraph and preserve the properties of the whole graph
Graph ↕ Graph	GCLMI [157]	$I(g, g_i^{aug})$	BCE	Design a simple yet effective graph contrastive learning framework
	RGDAL[161]	$I(g^{source}, g^{target})$	BCE	A robust graph domain adaptive learning framework
	InfoGraph$^{(*)}$ [120]	$I(h_i, \mathcal{R}(\{h_j\}_{j=1}^{n}))$	JSD	Unsupervised and semi-supervised graph-level representation learning
	AD-GCL [128]	$I(f(G), f(t(G)))$	InfoNCE	Use a min-max framework to suppress redundant information.
	CoGSL [83]	$I(g_i^{aug}, g_j^{aug})$	InfoNCE	Learn the most compact structure relevant to downstream tasks

5.2.1 Preserving General Distributional Characteristics. Preserving distributional consistency between latent representations and the underlying graph structure is crucial for ensuring generalizable and robust performance in GML.

Variational Graph Auto-Encoder (VGAE) [60] introduces divergence-based regularization by minimizing the KL divergence between the approximate posterior $q(\mathbf{Z}|\mathbf{X}, \mathbf{A})$ and the prior $p(\mathbf{Z})$, ensuring that latent representations capture meaningful and generalizable graph structures. Building on this idea, extensions such as SIG-VAE [49] and DGVAE [77] enhance posterior expressiveness via hierarchical or cluster-aware designs. Similarly, VGNAE [1] and HVGNN [122] further employ divergence regularization for dynamic or hierarchical graphs. In generative modeling, D-VAE [188] and GraphVAE [117] leverage KL divergence to align latent and observed graph distributions, ensuring realistic and coherent graph generation. Beyond representation, SL-VAE [82] and related zero-shot approaches [144] use divergence minimization to facilitate knowledge transfer across unseen or uncertain graph scenarios.

5.2.2 Preserving Specific Distributional Features. Unlike methods emphasizing global consistency, approaches that preserve specific distributional features align latent representations with task-relevant or domain-specific distributions, in which divergence acts as a core objective to encode feature dependencies, variations, or fairness constraints. At the feature level, Lv and Du [92] minimize divergence between global and local feature-induced graphs to retain structurally meaningful features. At the graph level, Graphite [44] and DDGK [2] employ divergence to align latent distributions with graph topology and measure similarity across graphs. At the dataset level, Han et al. [47] minimize Jensen–Shannon divergence to learn prototype graphs that summarize global distributional properties. At the domain level, KT-GNN [7] reduces the KL divergence between heterogeneous node distributions to enable cross-domain knowledge transfer. In fairness-constrained learning, Buyl and De Bie [11] use divergence minimization to maintain distributional proximity while enforcing independence from sensitive attributes.

These approaches highlight the central role of divergence in learning distribution-aware graph representations. By **minimizing divergence** between learned and reference distributions, models preserve context-specific statistical patterns and structural dependencies, **making divergence not just a regularizer but a principled paradigm** for encoding meaningful, data-driven structure.

6 Principled Graph Learning: Harnessing Information Principles for Model Design

Information principles (*i.e.*, PME, IB and PRI) can be harnessed to improve GML models, particularly in representation learning, regularization, and robustness. Integrating these principles offers a holistic strategy for developing advanced GML models. A summary of some representative methods is shown in Table 4.

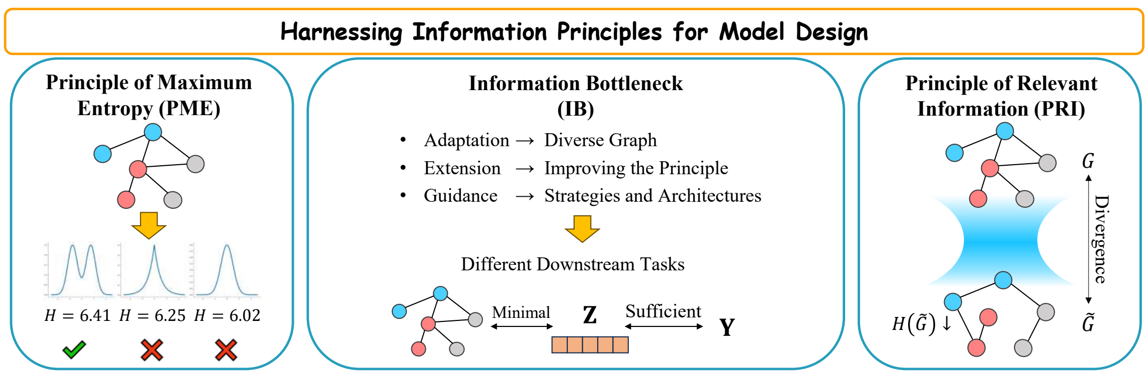

Fig. 9. Overview of Principled Graph Learning

6.1 Principle of Maximum Entropy

The Principle of Maximum Entropy (PME) provides a framework for modeling complex systems by selecting the least biased distributions consistent with known constraints, inspiring models that infer or summarize graph structures while preserving key statistics.

In social networks, Bipartite Configuration Model (BiCM) and Directed Bipartite Configuration Model (DBiCM) were employed by De Clerck et al. [20] to analyze the structural properties of large-scale bipartite graphs. These models retain local constraints while maximizing entropy over possible networks, ensuring that deviations from a random baseline reveal significant higher-order patterns. In neuroscience, Dichio and Fallani [25] used PME and Exponential Random Graph Models (ERGMs) to infer distributions over network ensembles that match observed topological features, such as degree distributions or motifs, revealing how local connectivity rules can give rise to emergent global structures in brain networks. Beyond network modeling, MEWISPool [98] interprets graph pooling as a communication problem, maximizing information transmission in a noisy channel. It selects a Maximum Weighted Independent Set based on node entropy, combining Shannon capacity with entropy maximization for a theoretically grounded pooling method.

6.2 Information Bottleneck

The Information Bottleneck (IB) principle provides a theoretical framework for learning representations that are maximally informative about a target variable while compressing input information. In GML, IB serves as a foundational tool for controlling information flow between graph structure, node features, and downstream tasks, enabling models to focus on task-relevant signals while discarding irrelevant or noisy details.

6.2.1 IB Adaptation: Handling Diverse Graph Structures. Real-world graphs exhibit substantial variability, including static and dynamic topologies, heterogeneous semantics, and high-order relationships. To address these varieties, recent studies have adapted the IB to various graph modalities, enabling robust and structure-aware representation learning.

For static graphs, GIB [150] reformulates the IB principle by analyzing both structure and feature information flows respectively, enhancing robustness against adversarial perturbations. For dynamic graphs, DGIB [174] proposes the *Minimal-Sufficient-Consensual* (MSC) condition to filter evolving graph signals over time, significantly enhancing robustness under adversarial attacks. For heterogeneous graphs, HGIB [158] addresses heterogeneous semantics by aligning multiple homogeneous subgraphs and minimizing intra-subgraph redundancy. For high-order relations, HIB [191] extends IB to hypergraphs, learning to refine noisy hypergraph structures by enhancing reliable connections and suppressing spurious ones. Beyond input adaptation, SIB [171] selects informative substructures from a complete graph, optimizing the IB objective via bi-level optimization to extract predictive subgraphs without supervision.

6.2.2 IB Extension: Improving the Principle Itself. While the IB principle provides a solid foundation for learning compact and informative representations, its direct application can face challenges such as limited representational capacity and incomplete information extraction. Recent methods have extended the core IB principle to optimize feature extraction, discrimination, and representation, as shown in Fig. 10. To improve estimation accuracy, CCGIB [35] proposed a cross-channel IB framework for multi-channel graphs, leveraging both variational upper and lower bounds to optimize mutual information. This dual-bound strategy improves representation quality by capturing both consistency and complementarity across channels. Focusing on robustness, RGIB [140] introduces the Subgraph Mutual Information (SMI) estimator, simulating attack scenarios to optimize the SMI objective ands improve unsupervised robustness on noisy graphs. CurvGIB [39] integrates domain-specific priors into the IB framework by incorporating discrete geometric properties, specifically Ricci curvature, into the IB objective. By aligning variational IB with discrete curvature flow, it facilitates structure-aware representation learning while reducing computational complexity.

6.2.3 IB Guidance: Learning Strategies and Architectures. Beyond extending the IB principle itself, another line of work leverages IB as guidance for designing learning strategies and network architectures. These approaches aim to improve the internal mechanics of graph representation learning rather than targeting specific downstream tasks. InfoGCL [155] introduces an information-aware contrastive learning framework that decouples the process into three modules: view augmentation, view encoding, and representation contrasting. Grounded in IB theory, it provides theoretical insights and practical criteria for optimizing each module, ensuring minimal information loss while retaining task-relevant content. CGRL [115] further extends contrastive learning by combining IB with an adversarial view reconstruction mechanism. Rather than relying on random node or edge perturbations, CGRL adaptively learns how to generate structurally diverse but semantically aligned views. On the architectural side, SGIB [195] proposes a differentiable

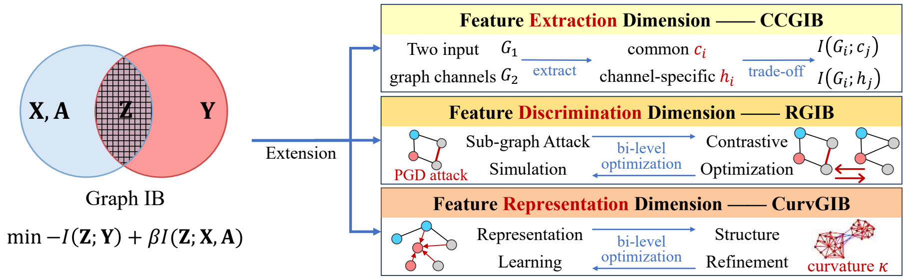

Fig. 10. Overview of Graph IB Extension from Different Dimensions

framework to mitigate noise in graph data. LGCN-SGIB [195] incorporates SGIB into a learnable GCN architecture that simultaneously refines the graph structure and optimizes feature propagation. The IB objective serves as a regularizer to reduce noisy interference, leading to improved semi-supervised node classification performance.

6.2.4 Output Objectives: Applying IB to Different Downstream Tasks. The IB principle has also been applied to align learning objectives, particularly in tasks involving graph construction or refinement. In this context, IB serves as a principled criterion for filtering redundant or noisy information while preserving task-relevant structural signals.

Structure Learning. Graph structure learning (GSL) aims to infer or refine the underlying topology to support downstream tasks when the original graph is noisy or unavailable. The IB principle offers a natural solution by balancing informativeness and compactness, guiding the discovery of minimal yet sufficient structures for robust prediction. VIB-GSL [123] applies the variational IB to generate a task-specific IB-graph by optimizing the trade-off between informativeness and compactness. The approach shows strong robustness to structural perturbations and offers a new perspective for noise-resilient structure construction. Similarly, Duan et al. [29] propose a hierarchical structure IB-based learning framework to eliminate redundant information and introduces a community influence-based fusion mechanism to enhance the base graph. The above methods significantly improve structure quality and model performance, especially in scenarios with noisy or incomplete graphs.

Dataset Condensation. Graph dataset condensation aims to compress large-scale graph data into a smaller, task-relevant one while preserving essential information. A key challenge lies in retaining multi-scale patterns across different compression levels without introducing information loss during scale transitions. To address this, BiMSGC [38] proposes a Bidirectional Multi-Scale Graph Compression framework guided by the IB principle. By estimating optimal intermediate scales and performing both large-to-small and small-to-large compression, it effectively mitigates the performance degradation commonly observed in traditional multi-scale methods.

Pooling. Graph pooling aims to generate compact graph-level representations by aggregating meaningful local structures. A major challenge lies in selecting substructures that are both discriminative and non-redundant for downstream tasks. HIBPool [108] addresses this by introducing a structure-aware discriminative pooling mechanism with the IB principle: maximizing MI with labels while minimizing redundancy with input data.

Graph Explanations. Graph explanation endeavors to uncover concise and meaningful substructures that underlie GNN predictions, enhancing model transparency and interpretability. A key objective is to strike a balance between informativeness and parsimony, capturing what matters for prediction while discarding redundancy. EGIB [139] uses a two-stage pipeline that first learns a task-agnostic explainer and then adapts it to downstream tasks, leveraging sufficient and ϵ-explanatory subgraphs to formalize explanation quality and support transferability. In dynamic scenarios, TGIB [113] integrates time-aware representations to identify influential event interactions, regulating information flow from candidate to target events via IB principle.

Subgraph Recognition. Subgraph recognition focuses on identifying informative substructures that contribute to graph-level tasks. The IB principle provides a mechanism to isolate such structures by retaining the information flow to task-relevant components. VGIB [170] proposes a Variational Graph Information Bottleneck framework that enhances subgraph recognition through noise injection and information flow regulation, improving both interpretability and performance. For pairwise relational tasks such as molecular interaction prediction, CGIB [64] extends IB via conditional MI, allowing dynamic selection of substructures based on input graph pairs. S-CGIB[65] introduces a subgraph-conditioned pre-training framework that discovers meaningful subgraphs without supervision and learns graph-level representations conditioned on these latent structures.

6.3 Principle of Relevant Information

The Principle of Relevant Information (PRI) has been applied to refine and sparsify graph structures, enabling models to learn or distill graph structures that retain essential information content with enhanced sparsity or robustness.

In graph sparsification, GraphPRI [172] formulates the task as selecting a subgraph G_s from the original graph G that preserves both structural properties while reducing edges. The optimization objective is:

$$J_{\text{Graph-PRI}} = \arg\min_{\sigma} \left(\mathcal{H}_{vN(G)}(\sigma) + \beta D_{QJS}(\sigma \parallel \rho)\right), \tag{29}$$

where $\mathcal{H}_{vN(G)}(\sigma)$ denotes the von Neumann Graph Entropy of the subgraph, and $D_{QJS}(\sigma \parallel \rho)$ is the quantum Jensen-Shannon divergence between the subgraph and the original graph. By minimizing this joint objective, GraphPRI selects the most information-preserving yet compact subgraph, leveraging PRI for the first time in graph sparsification.

Extending this principle to graph structure learning, PRI-GSL [125] aims to learn an optimized graph $\tilde{G}$ that supports downstream tasks, ensuring the learned structure retains relevant yet non-redundant information from the original graph, as shown in Eq. (30). This approach not only produces task-specific graphs but also preserves global informativeness and mitigates topological noise.

$$\mathcal{L} = \mathcal{L}_{\text{cls}} + \alpha\mathcal{L}_{\text{PRI}} = H_{\text{CE}}\left(\tilde{G}, Y\right) + \alpha\left(\mathcal{H}_{vN(G)}\left(\tilde{G}\right) + \beta D_{QJS}\left(\tilde{G} \parallel G\right)\right), \tag{30}$$

For dynamic graph modeling, DG-Mamba [177] integrates PRI into its self-supervised framework for Dynamic Graph Structure Learning. By employing PRI as an regularization term, it guides the model to learn dynamic graph structures that maximize informative temporal–spatial dependencies while minimizing redundancy:

$$\mathcal{L}_{\text{PRI}}(\hat{G}^{1:T}) = H(\hat{G}^{1:T}) + \beta \cdot D_{KL}(\hat{G}^{1:T} \parallel G^{1:T}). \tag{31}$$

7 Applications of Information-theoretic Graph Learning

In the real world, graph structures are pervasive and numerous problems across various fields can be modeled and solved using graphs, such as social networks, biological systems, and more [95]. To this end, we now shift focus to how information-theoretic GML methods are applied to real-world tasks in the following aspects:

- **Information Source Modeling**: focusing on how data is represented, traced, protected, and utilized. These challenges are fundamental for building trustworthy and effective data-driven systems.
- **Multi-Agent Learning**: focusing on designing algorithms for agents to interact and cooperate in complex environments. Key challenges include communication, policy learning, and stability.

Table 4. Comparison of methods using different principles.

Principle	Model	Maximize	Minimize	Characteristic
PME	De Clerck et al. [20] Dichio and Fallani [25]	$-\sum_{G\in\mathcal{G}} P(G)\log P(G)$	–	Obtaine the most unbiased probability density function $P(G)$
	MEWISPool [98]	$\sum_{i=1}^{\lvert V\rvert} H_S(X_i)p_i$	–	Maximize the information transmission rate in a noisy communication channel.
IB	GIB [150]	$I(\mathrm{Y};\mathrm{Z}_X)$	$I(G;\mathrm{Z}_X)$	Learn the Minimal-Sufficient representation.
	DGIB [174]	$I(\mathrm{Y}^{T+1};\mathrm{Z}^{T+1})$	$I(G^{1:T};\mathrm{Z}^{T+1})$ $I(\mathrm{Z}^{1:T},\mathrm{Z}^{T+1})$	Minimal-Sufficient-Consensual Condition.
	HGIB [158]	$I(h^i;g^j)$	$I(h^i;g^i \mid h^j)$	Learn the heterograph by splitting it into two parts.
	SIB [171]	$I(\mathrm{Y};G_{sub})$	$I(G;G_{sub})$	Extract the key subgraph from the graph using IB principle.
	HIB [191]	$I(\mathrm{Y},\mathrm{Z})$	$I(G_{structure},\mathrm{Z})$	Enhances hypergraph representations by suppressing noisy connections.
	CGIB [64]	$I(\mathrm{Y};G^1_{CIB})$	$I(G^1;G^1_{CIB}\lvert G^2)$	Extract key subgraphs in the action of a pair of graphs.
	CCGIB [35]	$I(G^i;g^i_{com})$ $I(G^i;g^{i/j}_{con})$	$I(G^i;g^j_{com})$	Extend the IB principle: for complementarity and consistency.
	RGIB [140]	$I(G;\mathrm{Z}),I(G;\mathrm{Z}')$	$I(G';\mathrm{Z}'\lvert G)$	A more robust IB method. G' is the adversarial graph.
	CurvGIB [39]	$I(\mathrm{Y};\mathrm{Z} \mid \kappa)$	$I(\mathcal{D};\mathrm{Z} \mid \kappa)$	Bi-level optimization for optimal curvature and representation.
	HIBPool [108]	$I(\mathrm{Y};S^{(L)})$	$I(\mathrm{A}^{(L)},\mathrm{X}^{(L)};S^{(L)})$	$S^{(L)}$ is the pooled representation. A novel graph pooling method.
	TGIB [113]	$I(\mathrm{Y}^k;\mathcal{R}^k)$	$I(\mathcal{R}^k;e_k,G^k)$	$\mathcal{R}^k$ is a subgraph of edge e_k's L-hop computation graph.
	CCGIB [35]	$I(G^i;g^i_{com})$ $I(G^i;g^{i/j}_{con})$	$I(G^i;g^j_{com})$	Extend the IB principle: for complementarity and consistency.
	EGIB [139]	$I(\mathrm{Z}^k;G_{sub})$ $I(\mathrm{Y}^k;G_{sub})$	$I(G_{sub},G)$ $I(G/G_{sub};\mathrm{Z})$	A two-stage explanation strategy (pretraining and finetuning)
PRI	GraphPRI [172]	–	$\mathcal{H}_{vN(G)}(\sigma)$ $D_{QJS}(\sigma \parallel \rho)$	For graph sparsification task: minimizing the number of edges.
	PRI-GSL [125]	$I(\mathrm{Y};\tilde{G})$	$H_{vN(G)}(\tilde{G})$ $D_{QJS}(\tilde{G} \parallel G)$	For structure learning task: learning an optimized graph structure.
	DG-Mamba [177]	$I(\mathrm{Y}^{T+1},\hat{\mathrm{Y}}^{T+1})$	$H(\hat{G}^{1:T})$ $D_{KL}(\hat{G}^{1:T} \parallel G^{1:T})$	For dynamic structure learning task: learning robust spatio-temporal graph.

- **Biological Applications**: focusing on adapting general learning models to biological applications requires domain knowledge since biological systems possess intrinsic complexity and specific constraints.
- **Scenario Perception & Prediction**: focusing on understanding and forecasting the evolution is vital for timely and accurate decision-making.

7.1 Information Source Modeling

Information source modeling focuses on understanding and preserving data origins for reliable and trustworthy analysis.

Source detection aims to identify the origin of a cascade from a partially observed interaction graph. Traditional methods often rely on structural heuristics, such as node degree or distance centrality. In contrast, ITE [186] proposes a

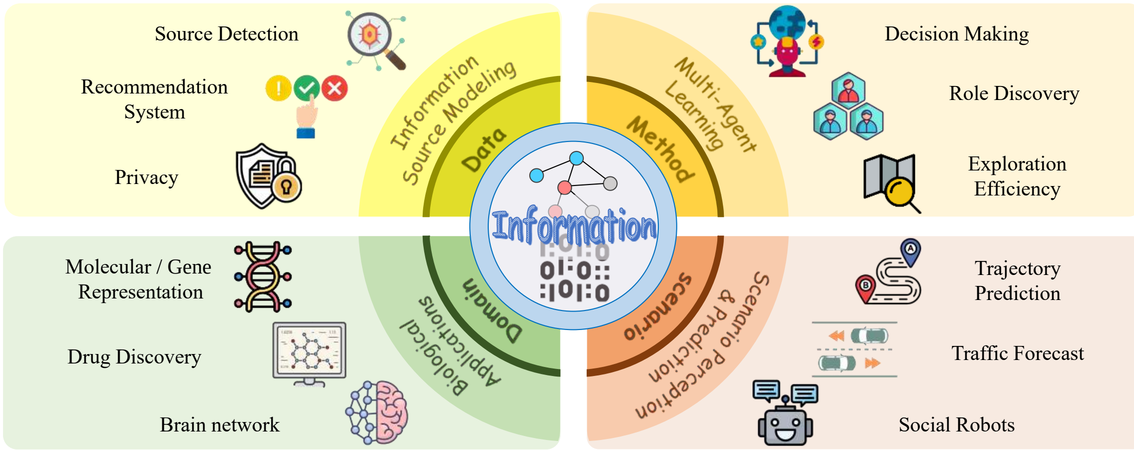

Fig. 11. Overview of Applications of Information-theoretic Graph Learning

novel metric called Infected Tree Entropy, which is inspired by structural entropy. The core idea is to identify the source node whose infection tree minimizes the encoding complexity of the structure, effectively capturing the structural divergence between potential infection patterns and the observed cascade.

Recommendation systems model users and content as information sources to personalize and optimize content delivery. SERGE [84] constructs both historical and real-time graphs and computes their graph entropies to dynamically weight their contributions. This enables the system to adaptively balance long-term preferences with recent user behavior. For enhancing diversity, DEME-GCL [196] introduces a maximum entropy-based neighbor selection strategy, promoting diverse information aggregation during message passing. To mitigate noise in user-item interactions, CGI [143] employs the IB principle to selectively retain informative nodes and edges while filtering out irrelevant ones.

Privacy preservation becomes both a technical and ethical imperative as models increasingly rely on user data. GCP [129] combines structural entropy and graph clustering with homomorphic encryption to extract the essential topology of Social Internet of Things networks while protecting user data. RM-GIB [19] applies the IB principle to design a GNN framework that jointly ensures robustness and membership privacy. In federated graph learning, PM-FedHG [187] leverages the IB principle to identify task-relevant subgraphs, reducing the exposure of structural information to Membership Inference Attacks without sacrificing predictive performance.

7.2 Multi-Agent Learning

Multi-agent reinforcement learning (MARL) involves complex interactions among agents, in which effective coordination and decision-making depend on capturing relational structures. Graph information theory offers a principled way to model these interactions by quantifying structural dependencies and uncertainty, providing a unified foundation.

Decision making lies at the core of MARL, requiring agents to learn robust policies under uncertainty and interaction complexity. SIDM [184] and SISA [183] leverage structural entropy to enhance decision-making through principled abstraction. SIDM minimizes the structural entropy of the state-action graph to reveal hierarchical relationships to reducing uncertainty. SISA uses unsupervised adaptive clustering to build an optimal encoding tree, filtering out irrelevant information. MAGI [26] further strengthens the decision-making process by addressing inter-agent communication. Built upon the IB principle, this method learns minimal yet informative message representations by maximizing MI between messages and selected actions while minimizing MI with agents' private features. This compression-enhanced communication allows agents to share only the most relevant information for coordination.

Role discovery further builds upon this structural foundation, as identifying agents' functional roles is essential for scalable and coordinated behavior. SIRD [180] leverages structural entropy to uncover roles through a hierarchical abstraction of the agent interaction graph. This enables roles to emerge from the topology of interactions, facilitating dynamic coordination and specialization.

Exploration efficiency is also important, particularly in high-dimensional environments with sparse rewards. Extending the role and decision abstractions, SI2E [182] maximizes value-conditional structural entropy, encouraging agents to explore parts of the state-action space that are both uncertain and potentially high-reward. This exploration strategy ensures exploration is guided by structural novelty and task relevance, rather than randomness.

7.3 Biological Applications

Biological systems are inherently complex and structured, characterized by rich multiple-level interactions. Modeling them requires capturing underlying graph structures while quantifying the diversity and relevance of information.

Molecular representation aims to encode molecular graphs composed of atoms and bonds into compact and informative representations. From an information-theoretic view, $\mathcal{E}^K$ [72] adopts a structural entropy minimization framework to construct cell-sample graphs from gene expression data. CGIB [64] applies the graph IB framework to capture subgraphs that are most influential for molecular properties. SMV-GNN [74] introduces a multi-view bottleneck mechanism to jointly leverage 2D graph and 3D spatial data, extracting consistent and informative structural features across modalities. MGIB [178] further extends this idea to a self-supervised pretraining setting, using the IB principle to distill core substructures that retain essential semantics while discarding task-irrelevant details.

Gene representation aims to anchor the encoding directly on the genomic structure. deDoc [73] identifies chromatin domains in high-throughput chromatin interaction data by modeling the contact map as a weighted map. By minimizing structural entropy, partitioning can capture dense intra-domain interactions, enabling fast and adaptive resolution topological associated domain identification.

Drug discovery seeks candidate molecules or targets to modulate disease processes with molecular representation. EMDIP [4] integrates uncertainty into protein–protein interaction networks by computing Shannon Entropy over annotation-context vectors, which captures the functional diversity of each protein and aids in identifying essential drugs with greater interpretability.

Brain network analysis aims to identify critical regions and high-order interactions that characterize cognition and disease. Sen et al. [110, 111] employ subgraph entropy to identify structurally significant regions in brain networks. Complementarily, Li et al. [79] leverage matrix-based Rényi entropy to capture higher-order functional interactions, revealing a synergistic organization of the resting-state brain. Dichio and Fallani [25] apply PME from a statistical perspective and model brain networks using Exponential Random Graph Models, explaining how global structures emerge from local constraints. BrainIB [193] leverages IB-guided GNNs to extract informative subgraphs from static functional magnetic resonance imaging (fMRI) data. DGAIB [27] extends the IB concept to dynamic brain graphs by incorporating attention mechanisms, effectively addressing noise and variability in brain signals.

7.4 Scenario Perception & Prediction

Understanding and predicting dynamic scenarios (*e.g.*, trajectories, traffic patterns, and social interactions) relies heavily on modeling evolving relationships. In this context, graph structural information is essential across diverse perception and prediction tasks.

Trajectory prediction aims to anticipate agents' future motion based on past behaviors and interactions. Jin et al. [57] constructs a dynamic graph and applies structural entropy minimization to detect relevant neighbors and quantify key relational patterns.

Traffic forecasting, a more macro-level variant of trajectory prediction, aims to predict future traffic conditions from historical spatiotemporal data. MultiSPANS [200] minimizes structural entropy to construct a hierarchical encoding tree of the road network, capturing complex spatial dependencies and identifying structurally significant node pairs.

Social robot detection and behavior modeling use structural information to analyze social networks and identify anomalous agents. SIASM [181] minimizes high-dimensional structural entropy to model dynamic communities and quantify user influence, and further introduces conditional structural entropy to guide adversarial follower selection. Building on similar principles, UnDBot [102] applies entropy-based analysis to multi-relational social graphs to detect latent robot communities, while SeBot [163] integrates structural entropy into contrastive learning to jointly optimize global and local subgraph representations.

8 Challenges, Open Problems, and Future Directions

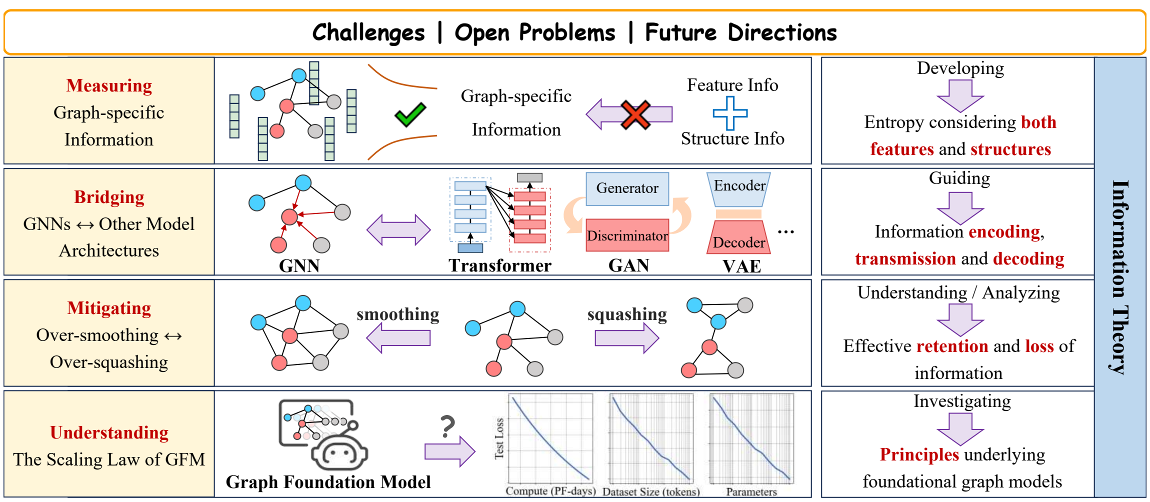

Fig. 12. Overview of Challenges, Open Problems, and Future Directions

Measuring Graph-specific information. Existing entropy calculation methods typically focus on either graph features or graph structures (Table 2), with few approaches that integrate both aspects. Future research should develop entropy measures specifically tailored for graph data that jointly consider both features and structures. Moreover, more complex types of graphs, such as graphs with multi-modal attributes, present further challenges. Designing information measures specifically tailored to these complex graphs remains an open problem.

Bridging GNNs and other model architectures. The optimal neural network architecture for graph data remains an open question. While message-passing GNNs are mainstream, graph transformers facilitate global information exchange, and generative models (*e.g.*, Graph GANs, VAEs, and diffusion models) capture underlying graph distributions. Information theory, by characterizing information encoding, transmission, and decoding, has the potential to analyze connections among these models and contribute to identifying the most suitable model architecture for graph data.

Mitigating over-smoothing and over-squashing in GNNs. Information theory can provide a new perspective for understanding over-smoothing and over-squashing phenomena in the message-passing paradigm of GNNs, helping to analyze the effective retention and loss of information during transmission and transformation [126]. Consequently, this insight can serve as a foundation for developing higher-order graph learning model architectures.

Understanding the scaling law of graph foundation models. With the rise of large language models, graph learning is increasingly focusing on graph foundation models [175, 176]. However, whether scaling laws exist for graph data remains an open question. Information theory can provide valuable insights and theoretical tools for studying the principles of these models. For instance, Gan and Liu [40] propose a reverse-bottleneck framework to understand how synthetic data influences LLM post-training from an information-theoretic perspective.

9 Conclusion

This survey principally introduces the existing methods that advance information theory for graph learning. First, we summarize commonly used information-theoretic concepts in graph learning, analyzing and highlighting their graph-specific designs, similarities, and differences. Then, we review existing approaches in GML from three perspectives: measuring intrinsic graph uncertainty and complexity, capturing multi-scale structural relationships, and guiding principled model design. We further discuss practical applications: Information Source Modeling, Multi-Agent Learning, Biological Applications, and Scenario Perception & Prediction, showing how theory translates into real-world impact. Finally, we summarize the possible future research challenges and directions for information-theoretic GML.

Acknowledgments

The corresponding author is Jianxin Li. The authors are supported in part by the National Natural Science Foundation of China through grant No.62225202, and the Fundamental Research Funds for the Central Universities.

References

[1] Seong Jin Ahn and MyoungHo Kim. 2021. Variational graph normalized autoencoders. In *CIKM*. 2827–2831.
[2] Rami Al-Rfou, Bryan Perozzi, and Dustin Zelle. 2019. Ddgk: Learning graph representations for deep divergence graph kernels. In *The Web Conference*. 37–48.
[3] Suguru Arimoto. 1977. Information measures and capacity of order α for discrete memoryless channels. *Topics in information theory* (1977).
[4] Hamid Bashiri, Hossein Rahmani, Vahid Bashiri, Dezső Módos, and Andreas Bender. 2020. EMDIP: an entropy measure to discover important proteins in PPI networks. *Computers in Biology and Medicine* 120 (2020), 103740.
[5] Alex Bavelas. 1948. A mathematical model for group structures. *Human organization* 7, 3 (1948), 16–30.
[6] Mohamed Ishmael Belghazi, Aristide Baratin, Sai Rajeswar, Sherjil Ozair, Yoshua Bengio, Aaron Courville, and R Devon Hjelm. 2018. Mine: mutual information neural estimation. *arXiv preprint arXiv:1801.04062* (2018).
[7] Wendong Bi, Bingbing Xu, Xiaoqian Sun, Li Xu, Huawei Shen, and Xueqi Cheng. 2023. Predicting the silent majority on graphs: Knowledge transferable graph neural network. In *The Web Conference*. 274–285.
[8] Isabelle Bloch and Alain Bretto. 2019. A new entropy for hypergraphs. In *DGCI*. 143–154.
[9] Samuel L Braunstein, Sibasish Ghosh, and Simone Severini. 2006. The Laplacian of a graph as a density matrix: a basic combinatorial approach to separability of mixed states. *Annals of Combinatorics* 10 (2006), 291–317.
[10] Frederick P Brooks Jr. 2003. Three great challenges for half-century-old computer science. *J. ACM* 50, 1 (2003), 25–26.

[11] Maarten Buyl and Tijl De Bie. 2021. The kl-divergence between a graph model and its fair i-projection as a fairness regularizer. In *ECML PKDD*. 351–366.

[12] Jiangxia Cao, Xixun Lin, Shu Guo, Luchen Liu, Tingwen Liu, and Bin Wang. 2021. Bipartite graph embedding via mutual information maximization. In *WSDM*. 635–643.

[13] Yuwei Cao, Hao Peng, Angsheng Li, Chenyu You, Zhifeng Hao, and Philip S Yu. 2024. Multi-Relational Structural Entropy. *arXiv preprint arXiv:2405.07096* (2024).

[14] Can Chen and Indika Rajapakse. 2020. Tensor entropy for uniform hypergraphs. *IEEE Transactions on Network Science and Engineering* 7, 4 (2020), 2889–2900.

[15] Jiaoyan Chen, Hang Dong, Janna Hastings, Ernesto Jiménez-Ruiz, Vanessa López, Pierre Monnin, Catia Pesquita, Petr Škoda, and Valentina Tamma. 2023. Knowledge graphs for the life sciences: Recent developments, challenges and opportunities. *arXiv preprint arXiv:2309.17255* (2023).

[16] Naiyue Chen, Yun Liu, Haiqiang Chen, and Junjun Cheng. 2017. Detecting communities in social networks using label propagation with information entropy. *Physica A: Statistical Mechanics and its Applications* 471 (2017), 788–798.

[17] Pin-Yu Chen, Lingfei Wu, Sijia Liu, and Indika Rajapakse. 2019. Fast incremental von neumann graph entropy computation: Theory, algorithm, and applications. In *ICML*. 1091–1101.

[18] Yang Chen and Jiamou Liu. 2019. Distributed community detection over blockchain networks based on structural entropy. In *BSCI*. 3–12.

[19] Enyan Dai, Limeng Cui, Zhengyang Wang, Xianfeng Tang, Yinghan Wang, Monica Cheng, Bing Yin, and Suhang Wang. 2023. A unified framework of graph information bottleneck for robustness and membership privacy. In *KDD*. 368–379.

[20] Bart De Clerck, Luis EC Rocha, and Filip Van Utterbeeck. 2022. Maximum entropy networks for large scale social network node analysis. *Applied Network Science* 7, 1 (2022), 68.

[21] Manlio De Domenico and Jacob Biamonte. 2016. Spectral entropies as information-theoretic tools for complex network comparison. *Physical Review X* 6, 4 (2016), 041062.

[22] Wouter De Nooy, Andrej Mrvar, and Vladimir Batagelj. 2018. *Exploratory social network analysis with Pajek*. Vol. 46. Cambridge university press.

[23] Łukasz Dębowski and Christian Bentz. 2020. Information theory and language. , 435 pages.

[24] Franck Delaplace. 2023. Entropic Detection of Chromatic Community Structures. In *International Conference on Complex Networks and Their Applications*. 223–234.

[25] Vito Dichio and Fabrizio De Vico Fallani. 2023. Statistical models of complex brain networks: a maximum entropy approach. *Reports on progress in physics* (2023).

[26] Shifei Ding, Wei Du, Ling Ding, Jian Zhang, Lili Guo, and Bo An. 2023. Robust Multi-Agent Communication With Graph Information Bottleneck Optimization. *IEEE Transactions on Pattern Analysis and Machine Intelligence* (2023).

[27] Changxu Dong and Dengdi Sun. 2024. Brain network classification based on dynamic graph attention information bottleneck. *Computer Methods and Programs in Biomedicine* 243 (2024), 107913.

[28] Wei Dong, Junsheng Wu, Yi Luo, Zongyuan Ge, and Peng Wang. 2022. Node representation learning in graph via node-to-neighbourhood mutual information maximization. In *CVPR*. 16620–16629.

[29] Liang Duan, Xiang Chen, Wenjie Liu, Daliang Liu, Kun Yue, and Angsheng Li. 2024. Structural Entropy Based Graph Structure Learning for Node Classification. In *AAAI*, Vol. 38. 8372–8379.

[30] Francisco Escolano, Edwin R Hancock, Miguel A Lozano, and Manuel Curado. 2017. The mutual information between graphs. *Pattern Recognition Letters* 87 (2017), 12–19.

[31] Francisco Escolano, Pablo Suau, and Boyán Bonev. 2009. *Information theory in computer vision and pattern recognition*. Springer.

[32] Ernesto Estrada. 2010. Quantifying network heterogeneity. *Physical Review E—Statistical, Nonlinear, and Soft Matter Physics* 82, 6 (2010), 066102.

[33] Ernesto Estrada and Naomichi Hatano. 2007. Statistical-mechanical approach to subgraph centrality in complex networks. *Chemical Physics Letters* 439, 1-3 (2007), 247–251.

[34] Xiaolong Fan, Maoguo Gong, Yue Wu, and Hao Li. 2023. Maximizing mutual information across feature and topology views for representing graphs. *IEEE Transactions on Knowledge and Data Engineering* 35, 10 (2023), 10735–10747.

[35] Xiaolong Fan, Maoguo Gong, Yue Wu, Mingyang Zhang, Hao Li, and Xiangming Jiang. 2024. CCGIB: A Cross-Channel Graph Information Bottleneck Principle. *IEEE Transactions on Neural Networks and Learning Systems* (2024).

[36] Wendy Flores-Fuentes, Oleg Sergiyenko, Julio C Rodríguez-Quiñonez, and Jesús E Miranda-Vega. 2024. Application of information theory to computer vision and image processing. , 114 pages.

[37] Linton C Freeman et al. 2002. Centrality in social networks: Conceptual clarification. *Social network: critical concepts in sociology. Londres: Routledge* 1, 3 (2002), 238–263.

[38] Xingcheng Fu, Yisen Gao, Beining Yang, Yuxuan Wu, Haodong Qian, Qingyun Sun, and Xianxian Li. 2024. Bi-Directional Multi-Scale Graph Dataset Condensation via Information Bottleneck. *arXiv preprint arXiv:2412.17355* (2024).

[39] Xingcheng Fu, Jian Wang, Yisen Gao, Qingyun Sun, Haonan Yuan, Jianxin Li, and Xianxian Li. 2024. Discrete Curvature Graph Information Bottleneck. *arXiv preprint arXiv:2412.19993* (2024).

[40] Zeyu Gan and Yong Liu. 2024. Towards a Theoretical Understanding of Synthetic Data in LLM Post-Training: A Reverse-Bottleneck Perspective. *arXiv preprint arXiv:2410.01720* (2024).

[41] Chen Gao et al. 2023. A survey of graph neural networks for recommender systems: Challenges, methods, and directions. *ACM Transactions on Recommender Systems* 1, 1 (2023), 1–51.

[42] Xing Gao, Wenrui Dai, Chenglin Li, Hongkai Xiong, and Pascal Frossard. 2021. ipool—information-based pooling in hierarchical graph neural networks. *IEEE Transactions on Neural Networks and Learning Systems* 33, 9 (2021), 5032–5044.

[43] Ziv Goldfeld and Yury Polyanskiy. 2020. The information bottleneck problem and its applications in machine learning. *IEEE Journal on Selected Areas in Information Theory* 1, 1 (2020), 19–38.

[44] Aditya Grover, Aaron Zweig, and Stefano Ermon. 2019. Graphite: Iterative generative modeling of graphs. In *ICML*. 2434–2444.

[45] Michael U Gutmann and Aapo Hyvärinen. 2012. Noise-contrastive estimation of unnormalized statistical models, with applications to natural image statistics. *Journal of machine learning research* 13, 2 (2012).

[46] Lin Han, Edwin R Hancock, and Richard C Wilson. 2011. Entropy versus heterogeneity for graphs. In *GbRPR*. 32–41.

[47] Lin Han, Richard C Wilson, and Edwin R Hancock. 2015. Generative graph prototypes from information theory. *IEEE transactions on pattern analysis and machine intelligence* 37, 10 (2015), 2013–2027.

[48] Frank Harary and Gopal Gupta. 1997. Dynamic graph models. *Mathematical and Computer Modelling* 25, 7 (1997), 79–87.

[49] Arman Hasanzadeh, Ehsan Hajiramezanali, Krishna Narayanan, Nick Duffield, Mingyuan Zhou, and Xiaoning Qian. 2019. Semi-implicit graph variational auto-encoders. *NeurIPS* 32 (2019).

[50] R Devon Hjelm, Alex Fedorov, Samuel Lavoie-Marchildon, Karan Grewal, Phil Bachman, Adam Trischler, and Yoshua Bengio. 2018. Learning deep representations by mutual information estimation and maximization. *arXiv preprint arXiv:1808.06670* (2018).

[51] Feng Hu, Kuo Tian, and Zi-Ke Zhang. 2023. Identifying vital nodes in hypergraphs based on Von Neumann entropy. *Entropy* 25, 9 (2023), 1263.

[52] Shizhe Hu, Zhengzheng Lou, Xiaoqiang Yan, and Yangdong Ye. 2024. A Survey on Information Bottleneck. *IEEE Transactions on Pattern Analysis and Machine Intelligence* (2024).

[53] Junyu Huang, Qilong Feng, Jiahui Wang, Ziyun Huang, Jinhui Xu, and Jianxin Wang. 2024. SEC: more accurate clustering algorithm via structural entropy. In *AAAI*, Vol. 38. 12583–12590.

[54] Xiaoli Huang, Jingyu Li, and Yumiao Yuan. 2024. Link Prediction in Dynamic Social Networks Combining Entropy, Causality, and a Graph Convolutional Network Model. *Entropy* 26, 6 (2024), 477.

[55] Edwin T Jaynes. 1957. Information theory and statistical mechanics. *Physical review* 106, 4 (1957), 620.

[56] Lavender Y Jiang, John Shi, Mark Cheung, Oren Wright, and José MF Moura. 2020. Edge Entropy as an Indicator of the Effectiveness of GNNs over CNNs for Node Classification. In *ASILOMAR*. 746–750.

[57] Yuhui Jin, Sixun Yang, Weifeng Lv, Haitao Yu, Sainan Zhu, and Jian Huang. 2022. Structure entropy minimization-based dynamic social interaction modeling for trajectory prediction. *Information Sciences* 614 (2022), 170–184.

[58] Kenji Kawaguchi, Zhun Deng, Xu Ji, and Jiaoyang Huang. 2023. How does information bottleneck help deep learning?. In *ICML*. 16049–16096.

[59] Edward Casey Kenley and Young-Rae Cho. 2011. Entropy-based graph clustering: Application to biological and social networks. In *ICDM*. 1116–1121.

[60] Thomas N Kipf and Max Welling. 2016. Variational graph auto-encoders. *arXiv preprint arXiv:1611.07308* (2016).

[61] Janos Korner et al. 1971. Coding of an information source having ambiguous alphabet and the entropy of graphs.. In *6th Prague conference on Information Theory, etc.* 411–425.

[62] Solomon Kullback and Richard A Leibler. 1951. On information and sufficiency. *The annals of mathematical statistics* 22, 1 (1951), 79–86.

[63] Junhyun Lee, Inyeop Lee, and Jaewoo Kang. 2019. Self-attention graph pooling. In *ICML*. 3734–3743.

[64] Namkyeong Lee, Dongmin Hyun, Gyoung S Na, Sungwon Kim, Junseok Lee, and Chanyoung Park. 2023. Conditional graph information bottleneck for molecular relational learning. In *ICML*. 18852–18871.

[65] O-Joun Lee et al. 2025. Pre-Training Graph Neural Networks on Molecules by Using Subgraph-Conditioned Graph Information Bottleneck. In *AAAI*, Vol. 39. 17204–17213.

[66] Angsheng Li. 2020. Structural information learning machinery: Learning from observing, associating, optimizing, decoding, and abstracting. *arXiv preprint arXiv:2001.09637* (2020).

[67] Angsheng Li. 2024. Principles of Artificial Intelligence: The Models, Principles, and Methods From Computing to Mousuan.

[68] Angsheng Li. 2024. Science of Artificial Intelligence: Mathematical Principles of Intelligence.

[69] Angsheng Li, Qifu Hu, Jun Liu, and Yicheng Pan. 2016. Resistance and security index of networks: structural information perspective of network security. *Scientific reports* 6, 1 (2016), 26810.

[70] Angsheng Li, Jiankou Li, and Yicheng Pan. 2015. Discovering natural communities in networks. *Physica A: Statistical Mechanics and its Applications* 436 (2015), 878–896.

[71] Angsheng Li and Yicheng Pan. 2016. Structural information and dynamical complexity of networks. *IEEE Transactions on Information Theory* 62, 6 (2016), 3290–3339.

[72] Angsheng Li, Xianchen Yin, and Yicheng Pan. 2016. Three-dimensional gene map of cancer cell types: Structural entropy minimisation principle for defining tumour subtypes. *Scientific reports* 6, 1 (2016), 20412.

[73] Angsheng Li, Xianchen Yin, Bingxiang Xu, Danyang Wang, Jimin Han, Yi Wei, Yun Deng, Ying Xiong, and Zhihua Zhang. 2018. Decoding topologically associating domains with ultra-low resolution Hi-C data by graph structural entropy. *Nature communications* 9, 1 (2018), 3265.

[74] Changsheng Li, Kaihang Mao, Shiye Wang, Ye Yuan, and Guoren Wang. 2023. Self-Supervised Graph Information Bottleneck for Multiview Molecular Embedding Learning. *IEEE Transactions on Artificial Intelligence* 5, 4 (2023), 1554–1562.

[75] Gen Li and Jason J Jung. 2021. Seizure detection from multi-channel EEG using entropy-based dynamic graph embedding. *Artificial Intelligence in Medicine* 122 (2021), 102201.

[76] Jianxin Li, Qingyun Sun, Hao Peng, Beining Yang, Jia Wu, and S Yu Philip. 2023. Adaptive subgraph neural network with reinforced critical structure mining. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 45, 7 (2023), 8063–8080.

[77] Jia Li, Jianwei Yu, Jiajin Li, Honglei Zhang, Kangfei Zhao, Yu Rong, Hong Cheng, and Junzhou Huang. 2020. Dirichlet graph variational autoencoder. *NeurIPS* 33 (2020), 5274–5283.

[78] Peng Li, Shilin Wang, Guangwu Chen, Chengqi Bao, and Guanghui Yan. 2022. Identifying key nodes in complex networks based on local structural entropy and clustering coefficient. *Mathematical Problems in Engineering* 2022, 1 (2022), 8928765.

[79] Qiang Li, Shujian Yu, Kristoffer H Madsen, Vince D Calhoun, and Armin Iraji. 2023. Higher-order organization in the human brain from matrix-based rényi's entropy. In *ICASSPW*. 1–5.

[80] Xiao Li, Li Sun, Mengjie Ling, and Yan Peng. 2023. A survey of graph neural network based recommendation in social networks. *Neurocomputing* 549 (2023), 126441.

[81] Jianhua Lin. 1991. Divergence measures based on the Shannon entropy. *IEEE Transactions on Information theory* 37, 1 (1991), 145–151.

[82] Chen Ling, Junji Jiang, Junxiang Wang, and Zhao Liang. 2022. Source localization of graph diffusion via variational autoencoders for graph inverse problems. In *KDD*. 1010–1020.

[83] Nian Liu, Xiao Wang, Lingfei Wu, Yu Chen, Xiaojie Guo, and Chuan Shi. 2022. Compact graph structure learning via mutual information compression. In *The Web Conference*. 1601–1610.

[84] Shenghao Liu, Bang Wang, and Minghua Xu. 2017. SERGE: Successive event recommendation based on graph entropy for event-based social networks. *IEEE Access* 6 (2017), 3020–3030.

[85] Wencong Liu, Jiamou Liu, Zijian Zhang, Yiwei Liu, and Liehuang Zhu. 2022. Residual Entropy-based Graph Generative Algorithms. In *AAMAS*. 816–824.

[86] Xuecheng Liu, Luoyi Fu, Xinbing Wang, and Chenghu Zhou. 2022. On the similarity between von Neumann graph entropy and structural information: Interpretation, computation, and applications. *IEEE Transactions on Information Theory* 68, 4 (2022), 2182–2202.

[87] Yiwei Liu, Jiamou Liu, Zijian Zhang, Liehuang Zhu, and Angsheng Li. 2019. REM: From structural entropy to community structure deception. *NeurIPS* 32 (2019).

[88] Lorenzo Livi. 2017. Designing labeled graph classifiers by exploiting the Rényi entropy of the dissimilarity representation. *Entropy* 19, 5 (2017), 216.

[89] László Lovász. 1979. On the Shannon capacity of a graph. *IEEE Transactions on Information theory* 25, 1 (1979), 1–7.

[90] Wei Lu, Xin Li, Pan Zhu, Sai Lv, Chenting Zhao, and Jin Wang. 2023. A Hierarchical Pooling Method Using Structural Entropy and Information Bottleneck. In *CCC*. 8611–8615.

[91] Gongxu Luo, Jianxin Li, Jianlin Su, Hao Peng, Carl Yang, Lichao Sun, Philip S Yu, and Lifang He. 2021. Graph Entropy Guided Node Embedding Dimension Selection for Graph Neural Networks. In *IJCAI*. 2767–2774.

[92] Xiaolin Lv and Liang Du. 2021. Graph-based Kullback-Leibler Divergence Minimization for Unsupervised Feature Selection. In *ICMLSC*. 109–114.

[93] Katalin Marton. 1993. On the Shannon capacity of probabilistic graphs. *Journal of Combinatorial Theory, Series B* 57, 2 (1993), 183–195.

[94] Costas Mavromatis and George Karypis. 2021. Graph infoclust: Maximizing coarse-grain mutual information in graphs. In *PAKDD*. 541–553.

[95] Basudeb Mondal and Kajal De. 2017. An overview applications of graph theory in real field. *International Journal of Scientific Research in Computer Science, Engineering and Information Technology* 2, 5 (2017), 751–759.

[96] Vincenzo Nicosia, Regino Criado, Miguel Romance, Giovanni Russo, and Vito Latora. 2012. Controlling centrality in complex networks. *Scientific reports* 2, 1 (2012), 218.

[97] Michael A Nielsen and Isaac L Chuang. 2001. *Quantum computation and quantum information*. Vol. 2. Cambridge university press Cambridge.

[98] Amirhossein Nouranizadeh, Mohammadjavad Matinkia, Mohammad Rahmati, and Reza Safabakhsh. 2021. Maximum entropy weighted independent set pooling for graph neural networks. *arXiv preprint arXiv:2107.01410* (2021).

[99] Sebastian Nowozin, Botond Cseke, and Ryota Tomioka. 2016. f-gan: Training generative neural samplers using variational divergence minimization. *NeurIPS* 29 (2016).

[100] Aaron van den Oord, Yazhe Li, and Oriol Vinyals. 2018. Representation learning with contrastive predictive coding. *arXiv preprint arXiv:1807.03748* (2018).

[101] Filippo Passerini and Simone Severini. 2008. The von Neumann entropy of networks. *arXiv preprint arXiv:0812.2597* (2008).

[102] Hao Peng, Jingyun Zhang, Xiang Huang, Zhifeng Hao, Angsheng Li, Zhengtao Yu, and Philip S Yu. 2024. Unsupervised social bot detection via structural information theory. *arXiv preprint arXiv:2404.13595* (2024).

[103] Zhen Peng, Wenbing Huang, Minnan Luo, Qinghua Zheng, Yu Rong, Tingyang Xu, and Junzhou Huang. 2020. Graph representation learning via graphical mutual information maximization. In *The Web Conference*. 259–270.

[104] José C Príncipe. 2010. *Information theoretic learning: Renyi's entropy and kernel perspectives*. Springer Science & Business Media.

[105] Patrick Reiser, Marlen Neubert, André Eberhard, Luca Torresi, Chen Zhou, Chen Shao, Houssam Metni, Clint van Hoesel, Henrik Schopmans, Timo Sommer, et al. 2022. Graph neural networks for materials science and chemistry. *Communications Materials* 3, 1 (2022), 93.

[106] Yuxiang Ren, Bo Liu, Chao Huang, Peng Dai, Liefeng Bo, and Jiawei Zhang. 2019. Heterogeneous deep graph infomax. *arXiv preprint arXiv:1911.08538* (2019).
[107] Alfréd Rényi. 1961. On measures of entropy and information. In *Proceedings of the fourth Berkeley symposium on mathematical statistics and probability, volume 1: contributions to the theory of statistics*, Vol. 4. 547–562.
[108] Kashob Kumar Roy, Amit Roy, AKM Mahbubur Rahman, M Ashraful Amin, and Amin Ahsan Ali. 2021. Structure-aware hierarchical graph pooling using information bottleneck. In *IJCNN*. 1–8.
[109] Andrew M Saxe, Yamini Bansal, Joel Dapello, Madhu Advani, Artemy Kolchinsky, Brendan D Tracey, and David D Cox. 2019. On the information bottleneck theory of deep learning. *Journal of Statistical Mechanics: Theory and Experiment* 2019, 12 (2019), 124020.
[110] Bhaskar Sen, Gail A Bernstein, Bryon A Mueller, Kathryn R Cullen, and Keshab K Parhi. 2020. Sub-graph entropy based network approaches for classifying adolescent obsessive-compulsive disorder from resting-state functional MRI. *NeuroImage: Clinical* 26 (2020), 102208.
[111] Bhaskar Sen, Shu-Hsien Chu, and Keshab K Parhi. 2019. Ranking regions, edges and classifying tasks in functional brain graphs by sub-graph entropy. *Scientific reports* 9, 1 (2019), 7628.
[112] Bhaskar Sen and Keshab K Parhi. 2020. Graph-theoretic properties of sub-graph entropy. *IEEE Signal Processing Letters* 28 (2020), 135–139.
[113] Sangwoo Seo, Sungwon Kim, Jihyeong Jung, Yoonho Lee, and Chanyoung Park. 2024. Self-Explainable Temporal Graph Networks based on Graph Information Bottleneck. In *KDD*. 2572–2583.
[114] Claude Elwood Shannon. 1948. A mathematical theory of communication. *The Bell system technical journal* 27, 3 (1948), 379–423.
[115] Yuntao Shou, Haozhi Lan, and Xiangyong Cao. 2024. Contrastive graph representation learning with adversarial cross-view reconstruction and information bottleneck. *arXiv preprint arXiv:2408.00295* (2024).
[116] Ravid Shwartz-Ziv and Naftali Tishby. 2017. Opening the black box of deep neural networks via information. *arXiv preprint arXiv:1703.00810* (2017).
[117] Martin Simonovsky and Nikos Komodakis. 2018. Graphvae: Towards generation of small graphs using variational autoencoders. In *ICANN*. 412–422.
[118] Housheng Su, Dan Chen, Gui-Jun Pan, and Zhigang Zeng. 2021. Identification of network topology variations based on spectral entropy. *IEEE Transactions on Cybernetics* 52, 10 (2021), 10468–10478.
[119] Chengyu Sun, Xing Ai, Zhihong Zhang, and Edwin R Hancock. 2023. Labeled subgraph entropy kernel. *arXiv preprint arXiv:2303.13543* (2023).
[120] Fan-Yun Sun, Jordan Hoffmann, Vikas Verma, and Jian Tang. 2019. Infograph: Unsupervised and semi-supervised graph-level representation learning via mutual information maximization. *arXiv preprint arXiv:1908.01000* (2019).
[121] Li Sun, Zhenhao Huang, Hao Peng, Yujie Wang, Chunyang Liu, and Philip S Yu. 2024. Lsenet: Lorentz structural entropy neural network for deep graph clustering. *arXiv preprint arXiv:2405.11801* (2024).
[122] Li Sun, Zhongbao Zhang, Jiawei Zhang, Feiyang Wang, Hao Peng, Sen Su, and Philip S Yu. 2021. Hyperbolic variational graph neural network for modeling dynamic graphs. In *AAAI*, Vol. 35. 4375–4383.
[123] Qingyun Sun, Jianxin Li, Hao Peng, Jia Wu, Xingcheng Fu, Cheng Ji, and S Yu Philip. 2022. Graph structure learning with variational information bottleneck. In *AAAI*, Vol. 36. 4165–4174.
[124] Qingyun Sun, Jianxin Li, Hao Peng, Jia Wu, Yuanxing Ning, Philip S Yu, and Lifang He. 2021. Sugar: Subgraph neural network with reinforcement pooling and self-supervised mutual information mechanism. In *The Web Conference*. 2081–2091.
[125] Qingyun Sun, Jianxin Li, Beining Yang, Xingcheng Fu, Hao Peng, and S Yu Philip. 2023. Self-organization preserved graph structure learning with principle of relevant information. In *AAAI*, Vol. 37. 4643–4651.
[126] Qingyun Sun, Jianxin Li, Haonan Yuan, Xingcheng Fu, Hao Peng, Cheng Ji, Qian Li, and Philip S Yu. 2022. Position-aware structure learning for graph topology-imbalance by relieving under-reaching and over-squashing. In *CIKM*. 1848–1857.
[127] Ziheng Sun, Xudong Wang, Chris Ding, and Jicong Fan. 2024. Learning Graph Representation via Graph Entropy Maximization. In *ICML*.
[128] Susheel Suresh, Pan Li, Cong Hao, and Jennifer Neville. 2021. Adversarial graph augmentation to improve graph contrastive learning. *NeurIPS* 34 (2021), 15920–15933.
[129] Youliang Tian, Zhiying Zhang, Jinbo Xiong, Lei Chen, Jianfeng Ma, and Changgen Peng. 2021. Achieving graph clustering privacy preservation based on structure entropy in social IoT. *IEEE Internet of Things Journal* 9, 4 (2021), 2761–2777.
[130] Nicholas M Timme and Christopher Lapish. 2018. A tutorial for information theory in neuroscience. *eneuro* 5, 3 (2018).
[131] Naftali Tishby, Fernando C Pereira, and William Bialek. 2000. The information bottleneck method. *arXiv preprint physics/0004057* (2000).
[132] Naftali Tishby and Noga Zaslavsky. 2015. Deep learning and the information bottleneck principle. In *2015 IEEE Information Theory Workshop*. 1–5.
[133] Himanshu Tyagi and Shun Watanabe. 2023. *Information-theoretic cryptography*. Cambridge University Press.
[134] Shinsuke Uda. 2020. Application of information theory in systems biology. *Biophysical reviews* 12, 2 (2020), 377–384.
[135] Petar Veličković, William Fedus, William L Hamilton, Pietro Liò, Yoshua Bengio, and R Devon Hjelm. 2019. Deep Graph Infomax. In *ICLR*.
[136] John Von Neumann. 2013. *Mathematische grundlagen der quantenmechanik*. Vol. 38. Springer-Verlag.
[137] Pengfei Wan, Xinzhuang Chen, Jianhua Tu, Matthias Dehmer, Shenggui Zhang, and Frank Emmert-Streib. 2020. On graph entropy measures based on the number of independent sets and matchings. *Information Sciences* 516 (2020), 491–504.
[138] Hui Wang, Zichun Le, and Xuan Gong. 2020. Recommendation system based on heterogeneous feature: A survey. *IEEE Access* 8 (2020), 170779–170793.

[139] Jihong Wang, Minnan Luo, Jundong Li, Yun Lin, Yushun Dong, Jin Song Dong, and Qinghua Zheng. 2023. Empower Post-hoc Graph Explanations with Information Bottleneck: A Pre-training and Fine-tuning Perspective. In *KDD*. 2349–2360.

[140] Jihong Wang, Minnan Luo, Jundong Li, Ziqi Liu, Jun Zhou, and Qinghua Zheng. 2023. Toward enhanced robustness in unsupervised graph representation learning: A graph information bottleneck perspective. *IEEE Transactions on Knowledge and Data Engineering* (2023).

[141] Luran Wang, Mark Gales, and Vatsal Raina. 2024. An information-theoretic approach to analyze nlp classification tasks. *arXiv preprint arXiv:2402.00978* (2024).

[142] Yifei Wang, Yupan Wang, Zeyu Zhang, Song Yang, Kaiqi Zhao, and Jiamou Liu. 2023. User: Unsupervised structural entropy-based robust graph neural network. In *AAAI*, Vol. 37. 10235–10243.

[143] Chunyu Wei, Jian Liang, Di Liu, and Fei Wang. 2022. Contrastive graph structure learning via information bottleneck for recommendation. *NeurIPS* 35 (2022), 20407–20420.

[144] Jiwei Wei, Yang Yang, Xing Xu, Yanli Ji, Xiaofeng Zhu, and Heng Tao Shen. 2021. Graph-based variational auto-encoder for generalized zero-shot learning. In *MMAsia*. 1–7.

[145] Tao Wen and Yong Deng. 2020. The vulnerability of communities in complex networks: An entropy approach. *Reliability Engineering & System Safety* 196 (2020), 106782.

[146] Mark M Wilde. 2013. *Quantum information theory*. Cambridge university press.

[147] Junran Wu, Xueyuan Chen, Ke Xu, and Shangzhe Li. 2022. Structural entropy guided graph hierarchical pooling. In *ICML*. 24017–24030.

[148] Junran Wu, Jianhao Li, Yicheng Pan, and Ke Xu. 2021. Structural Optimization Makes Graph Classification Simpler and Better. *arXiv preprint arXiv:2109.02027* (2021).

[149] Junran Wu, Shangzhe Li, Jianhao Li, Yicheng Pan, and Ke Xu. 2022. A simple yet effective method for graph classification. *arXiv preprint arXiv:2206.02404* (2022).

[150] Tailin Wu, Hongyu Ren, Pan Li, and Jure Leskovec. 2020. Graph information bottleneck. In *NeurIPS*.

[151] Zonghan Wu, Shirui Pan, Fengwen Chen, Guodong Long, Chengqi Zhang, and Philip S Yu. 2020. A comprehensive survey on graph neural networks. *IEEE Transactions on Neural Networks and Learning Systems* 32, 1 (2020), 4–24.

[152] Feng Xia, Ke Sun, Shuo Yu, Abdul Aziz, Liangtian Wan, Shirui Pan, and Huan Liu. 2021. Graph learning: A survey. *IEEE Transactions on Artificial Intelligence* 2, 2 (2021), 109–127.

[153] Feng Xia, Ke Sun, Shuo Yu, Abdul Aziz, Liangtian Wan, Shirui Pan, and Huan Liu. 2021. Graph learning: A survey. *IEEE Transactions on Artificial Intelligence* 2, 2 (2021), 109–127.

[154] Yantuan Xian, Pu Li, Hao Peng, Zhengtao Yu, Yan Xiang, and Philip S Yu. 2025. Community detection in large-scale complex networks via structural entropy game. In *The Web Conference*. 3930–3941.

[155] Dongkuan Xu, Wei Cheng, Dongsheng Luo, Haifeng Chen, and Xiang Zhang. 2021. Infogcl: Information-aware graph contrastive learning. *NeurIPS* 34 (2021), 30414–30425.

[156] Lixiang Xu, Lu Bai, Xiaoyi Jiang, Ming Tan, Daoqiang Zhang, and Bin Luo. 2021. Deep rényi entropy graph kernel. *Pattern Recognition* 111 (2021), 107668.

[157] Yuhua Xu, Junli Wang, Mingjian Guang, Chungang Yan, and Changjun Jiang. 2024. Graph contrastive learning with min-max mutual information. *Information Sciences* 665 (2024), 120378.

[158] Liang Yang, Fan Wu, Zichen Zheng, Bingxin Niu, Junhua Gu, Chuan Wang, Xiaochun Cao, and Yuanfang Guo. 2021. Heterogeneous Graph Information Bottleneck.. In *IJCAI*. 1638–1645.

[159] Runze Yang, Hao Peng, and Angsheng Li. 2022. Dynamic Measurement of Structural Entropy for Dynamic Graphs. *arXiv preprint arXiv:2207.12653* (2022).

[160] Runze Yang, Hao Peng, Chunyang Liu, and Angsheng Li. 2024. Incremental measurement of structural entropy for dynamic graphs. *Artificial Intelligence* 334 (2024), 104175.

[161] Shuiqiao Yang, Borui Cai, Taotao Cai, Xiangyu Song, Jiaojiao Jiang, Bing Li, and Jianxin Li. 2022. Robust cross-network node classification via constrained graph mutual information. *Knowledge-Based Systems* 257 (2022), 109852.

[162] Yonghui Yang, Le Wu, Richang Hong, Kun Zhang, and Meng Wang. 2021. Enhanced graph learning for collaborative filtering via mutual information maximization. In *SIGIR*. 71–80.

[163] Yingguang Yang, Qi Wu, Buyun He, Hao Peng, Renyu Yang, Zhifeng Hao, and Yong Liao. 2024. SeBot: Structural Entropy Guided Multi-View Contrastive Learning for Social Bot Detection. In *KDD*. 3841–3852.

[164] Zhenyu Yang, Ge Zhang, Jia Wu, Jian Yang, Quan Z Sheng, Hao Peng, Angsheng Li, Shan Xue, and Jianlin Su. 2023. Minimum entropy principle guided graph neural networks. In *WSDM*. 114–122.

[165] Tianjun Yao, Yongqiang Chen, Zhenhao Chen, Kai Hu, Zhiqiang Shen, and Kun Zhang. 2024. Empowering Graph Invariance Learning with Deep Spurious Infomax. *arXiv preprint arXiv:2407.11083* (2024).

[166] Yuan Yao, Yicheng Pan, Shaojiang Wang, Hongan Wang, and Waqas Nazeer. 2019. Flow-based clustering on directed graphs: A structural entropy minimization approach. *IEEE Access* 8 (2019), 152579–152591.

[167] Cheng Ye, Richard C Wilson, César H Comin, Luciano da F Costa, and Edwin R Hancock. 2014. Approximate von Neumann entropy for directed graphs. *Physical Review E* 89, 5 (2014), 052804.

[168] Zhitao Ying, Dylan Bourgeois, Jiaxuan You, Marinka Zitnik, and Jure Leskovec. 2019. Gnnexplainer: Generating explanations for graph neural networks. *NeurIPS* 32 (2019).
[169] Yuning You, Tianlong Chen, Zhangyang Wang, and Yang Shen. 2022. Bringing your own view: Graph contrastive learning without prefabricated data augmentations. In *WSDM*. 1300–1309.
[170] Junchi Yu, Jie Cao, and Ran He. 2022. Improving subgraph recognition with variational graph information bottleneck. In *CVPR*. 19396–19405.
[171] Junchi Yu, Tingyang Xu, Yu Rong, Yatao Bian, Junzhou Huang, and Ran He. 2021. Recognizing Predictive Substructures with Subgraph Information Bottleneck. *IEEE Transactions on Pattern Analysis and Machine Intelligence* (2021).
[172] Shujian Yu, Francesco Alesiani, Wenzhe Yin, Robert Jenssen, and José C Principe, Príncipe. 2022. Principle of Relevant Information for Graph Sparsification. In *UAI*.
[173] Shujian Yu, Luis G Sánchez Giraldo, and José C Príncipe. 2021. Information-Theoretic Methods in Deep Neural Networks: Recent Advances and Emerging Opportunities.. In *IJCAI*. 4669–4678.
[174] Haonan Yuan, Qingyun Sun, Xingcheng Fu, Cheng Ji, and Jianxin Li. 2024. Dynamic Graph Information Bottleneck. In *The Web Conference*.
[175] Haonan Yuan, Qingyun Sun, Junhua Shi, Xingcheng Fu, Bryan Hooi, Jianxin Li, and Philip S Yu. 2025. GRAVER: Generative Graph Vocabularies for Robust Graph Foundation Models Fine-tuning. In *NeurIPS*.
[176] Haonan Yuan, Qingyun Sun, Junhua Shi, Xingcheng Fu, Bryan Hooi, Jianxin Li, and Philip S Yu. 2025. How Much Can Transfer? BRIDGE: Bounded Multi-Domain Graph Foundation Model with Generalization Guarantees. In *ICML*.
[177] Haonan Yuan, Qingyun Sun, Zhaonan Wang, Xingcheng Fu, Cheng Ji, Yongjian Wang, Bo Jin, and Jianxin Li. 2025. Dg-mamba: Robust and efficient dynamic graph structure learning with selective state space models. In *AAAI*, Vol. 39. 22272–22280.
[178] Xuan Zang, Junjie Zhang, and Buzhou Tang. 2024. Self-supervised Pre-training via Multi-view Graph Information Bottleneck for Molecular Property Prediction. *IEEE Journal of Biomedical and Health Informatics* (2024).
[179] Guangjie Zeng, Hao Peng, Angsheng Li, Zhiwei Liu, Runze Yang, Chunyang Liu, and Lifang He. 2024. Semi-supervised clustering via structural entropy with different constraints. In *SDM*. 208–216.
[180] Xianghua Zeng, Hao Peng, and Angsheng Li. 2023. Effective and stable role-based multi-agent collaboration by structural information principles. In *AAAI*, Vol. 37. 11772–11780.
[181] Xianghua Zeng, Hao Peng, and Angsheng Li. 2024. Adversarial socialbots modeling based on structural information principles. In *AAAI*, Vol. 38. 392–400.
[182] Xianghua Zeng, Hao Peng, and Angsheng Li. 2024. Effective exploration based on the structural information principles. *NeurIPS* 37 (2024), 95291–95325.
[183] Xianghua Zeng, Hao Peng, Angsheng Li, Chunyang Liu, Lifang He, and Philip S Yu. 2023. Hierarchical state abstraction based on structural information principles. In *IJCAI*. 4549–4557.
[184] Xianghua Zeng, Hao Peng, Dingli Su, and Angsheng Li. 2024. Hierarchical Decision Making Based on Structural Information Principles. *arXiv preprint arXiv:2404.09760* (2024).
[185] Ling Zhan and Tao Jia. 2022. Coarsas2hvec: Heterogeneous information network embedding with balanced network sampling. *Entropy* 24, 2 (2022), 276.
[186] Chong Zhang, Qiang Guo, Luoyi Fu, Xiaoying Gan, and Xinbing Wang. 2021. Ite: A structural entropy based approach for source detection. In *IEEE INFOCOM*. 1–10.
[187] Chenhan Zhang, Weiqi Wang, James JQ Yu, and Shui Yu. 2023. Extracting privacy-preserving subgraphs in federated graph learning using information bottleneck. In *ASIACCS*. 109–121.
[188] Muhan Zhang, Shali Jiang, Zhicheng Cui, Roman Garnett, and Yixin Chen. 2019. D-vae: A variational autoencoder for directed acyclic graphs. *NeurIPS* 32 (2019).
[189] Tianqi Zhang, Yun Xiong, Jiawei Zhang, Yao Zhang, Yizhu Jiao, and Yangyong Zhu. 2020. CommDGI: community detection oriented deep graph infomax. In *CIKM*. 1843–1852.
[190] Xiao-Meng Zhang, Li Liang, Lin Liu, and Ming-Jing Tang. 2021. Graph neural networks and their current applications in bioinformatics. *Frontiers in genetics* 12 (2021), 690049.
[191] Zizhao Zhang, Yifan Feng, Shihui Ying, and Yue Gao. 2022. Deep hypergraph structure learning. *arXiv preprint arXiv:2208.12547* (2022).
[192] Zhaobo Zhang, Meizhu Li, and Qi Zhang. 2024. A clustering coefficient structural entropy of complex networks. *Physica A: Statistical Mechanics and its Applications* 655 (2024), 130170.
[193] Kaizhong Zheng, Shujian Yu, Baojuan Li, Robert Jenssen, and Badong Chen. 2024. Brainib: Interpretable brain network-based psychiatric diagnosis with graph information bottleneck. *IEEE Transactions on Neural Networks and Learning Systems* (2024).
[194] Xin Zheng, Yixin Liu, Zhifeng Bao, Meng Fang, Xia Hu, Alan Wee-Chung Liew, and Shirui Pan. 2023. Towards data-centric graph machine learning: Review and outlook. *arXiv preprint arXiv:2309.10979* (2023).
[195] Luying Zhong, Zhaoliang Chen, Zhihao Wu, Shide Du, Zheyi Chen, and Shiping Wang. 2023. Learnable Graph Convolutional Network With Semisupervised Graph Information Bottleneck. *IEEE Transactions on Neural Networks and Learning Systems* (2023).
[196] Jinhao Zhou, Pinghua Chen, Yunhua Chen, and Honghong Zhou. 2024. Diverse Recommendations With Maximum Entropy Neighbor Selection and Graph Contrast Learning. In *ISPA*. 981–989.

[197] Bingyu Zhu, Qingyun Sun, Jianxin Li, and Daqing Li. 2023. Precise influence evaluation in complex networks. *arXiv preprint arXiv:2310.12181* (2023).

[198] Ran Zhu, Jian Peng, Wen Huang, Yujun He, and Chengyi Tang. 2024. Dynamic graph attention-guided graph clustering with entropy minimization self-supervision. *Applied Intelligence* 54, 24 (2024), 12819–12834.

[199] Dongcheng Zou, Hao Peng, Xiang Huang, Renyu Yang, Jianxin Li, Jia Wu, Chunyang Liu, and Philip S Yu. 2023. Se-gsl: A general and effective graph structure learning framework through structural entropy optimization. In *The Web Conference*. 499–510.

[200] Dongcheng Zou, Senzhang Wang, Xuefeng Li, Hao Peng, Yuandong Wang, Chunyang Liu, Kehua Sheng, and Bo Zhang. 2024. Multispans: A multi-range spatial-temporal transformer network for traffic forecast via structural entropy optimization. In *WSDM*. 1032–1041.

Received 08 December 2025