A Survey of Structural Entropy: Theory, Methods, and Applications

Dingli Su, **Hao Peng**, **Yicheng Pan**✉ and **Angsheng Li**
Beihang University
{sudingli, penghao, yichengp, angsheng}@buaa.edu.cn

Abstract

Classical information theory, a cornerstone of artificial intelligence, is fundamentally limited by its local perspective, often analyzing pairwise interactions while ignoring the larger, hierarchical architecture of complex systems. Structural entropy (SE) presents a paradigm shift, extending Shannon entropy to quantify information on a global scale and measure the uncertainty embedded in a system's organizational hierarchy. Although its applications have broadened significantly from its origins in community detection across diverse AI domains, a systematic synthesis of its theory, computational methods, and applications is currently lacking. This survey provides a comprehensive overview of SE to fill this critical void in the literature. We offer a detailed examination of its theoretical foundations, computational frameworks, and key learning paradigms, with a focus on its integration with graph learning and reinforcement learning. Through an exploration of its diverse applications, we highlight the power of SE to advance graph-based analysis and modeling. Finally, we discuss key challenges and future research opportunities for incorporating SE principles into the development of more interpretable and theoretically grounded AI systems.

1 Introduction

Information lies at the heart of artificial intelligence. From Shannon's entropy [Shannon, 1948] guiding learning algorithms, to the Information Bottleneck [Tishby *et al.*, 2000] shaping representations–virtually every paradigm in AI can be seen as either compressing, transferring, organizing, or extracting meaning from information. Yet, while much of classical information theory in AI is concerned with quantities of uncertainty or relevance at a local or pairwise level, **Structural Entropy (SE)** offers a distinct and complementary perspective: it captures the global structure and hierarchical uncertainty embedded within complex information systems. At its core, SE provides a rigorous method to quantify and decode the uncertainty embedded in the architecture of a complex system, modeled as a graph or matrix [Li and Pan, 2016]. Its significance for the AI community is twofold, offering both a new metric and a constructive principle for optimization.

On one hand, SE serves as a fundamental metric to quantify a system's intrinsic uncertainty. A low SE signals a clear, efficient hierarchical organization, implying greater functional coherence and robustness. Conversely, a high SE indicates a system that is difficult to decompose, approaching the behavior of a random graph. This principle is directly applied in areas like privacy-preserving graph analysis, where community structures are deliberately obscured by increasing structural entropy [Liu *et al.*, 2019]. This analytical power enables us to measure properties like the amount of knowledge captured by a model [Wang *et al.*, 2023a], the emergence of collaboration between agents [Su *et al.*, 2025], or the safety and stability of an AI system [Zeng *et al.*, 2025b].

On the other hand, and perhaps more profoundly, SE provides a principled method to discover a system's optimal hierarchical structure. The theory's central insight is that minimizing structural entropy is an optimization process that decodes a system's most efficient information-theoretic abstraction, revealing a nested, hierarchical partitioning of its components. This discovered hierarchy is not arbitrary; it represents the most concise description of the system's organization, akin to finding its "true" community structure [Li and Pan, 2016] or its most effective decision hierarchy [Zeng *et al.*, 2025c]. This constructive capability has already found potent applications across the AI development pipeline—from designing optimal pooling layers in Graph Neural Networks (GNNs) [Wu *et al.*, 2022; Zou *et al.*, 2023; Duan *et al.*, 2024; Ren *et al.*, 2024] and guiding data augmentation [Wu *et al.*, 2023; Wang *et al.*, 2023b], to shaping reward functions in reinforcement learning [Huang *et al.*, 2024; Zeng *et al.*, 2023c; Zeng *et al.*, 2024c] and learning more meaningful representations [Yang *et al.*, 2023; Zeng *et al.*, 2023b; Zeng *et al.*, 2025c].

Despite the rapid proliferation and significant impact of SE-based methods across various AI domains, a comprehensive and systematic review that consolidates these advancements is currently lacking. This paper aims to fill this critical gap. We provide an extensive overview of the theoretical underpinnings of structural entropy, detail the computational methods developed for its calculation and optimization, and survey the

✉ Corresponding authors.

learning paradigms that leverage SE for enhanced graph learning and reinforcement learning. Furthermore, we explore its diverse applications in fields such as bioinformatics, transport and geoscience studies, social network analysis, and pattern recognition, as illustrated in Figure 1.

The paper is structured as follows. Section 2 outlines the basic concepts of SE. Section 3 delves into learning methods. Section 4 explores cross-domain applications. Section 5 concludes with future directions and open challenges.

2 Basic Concepts

Structural entropy offers a theoretical foundation for quantifying the complexity and hierarchical organization of structured data. At its core, this framework models a system as a finite set of interacting elements, and characterizes its structural regularities through a tree-based recursive partitioning. In this section, we introduce the basic mathematical constructs of encoding trees and the corresponding structural entropy.

2.1 Encoding Tree of a Finite Set

Let A be a finite set representing the entities in a complex system, such as nodes in a network or data points in a dataset. A natural approach to modeling the structure of such a system is through hierarchical abstraction, in which elements are recursively grouped into nested subsets. This leads to the definition of an encoding tree, which organizes the elements of A into a tree structure where each node corresponds to a subset of A, and the leaves correspond to individual elements.

Definition 2.1 (Encoding Tree of a Finite Set)**.** *Given a finite set A, an* encoding tree *of A is a rooted tree T satisfying the following conditions.*

First, the root node is denoted by the empty string λ, and corresponds to the entire set: $T_\lambda = A$.

Second, for every node $\alpha \in T$, there exists a nonempty subset $T_\alpha \subseteq A$. If α has children $\beta_0, \beta_1, \ldots, \beta_l$, then each child is denoted as $\beta_j = \alpha \cdot j$, where $\cdot$ denotes the concatenation of the index sequence α with the integer j, forming a unique code for the node in the encoding tree. The corresponding subsets $T_{\beta_0}, \ldots, T_{\beta_l}$ form a partition of T_α; that is, $T_\alpha = \bigsqcup_{j=0}^{l} T_{\beta_j}$.

Third, each leaf node $\gamma \in T$ corresponds to a singleton subset of A, that is, $T_\gamma = \{a\}$ for some $a \in A$.

Such a tree encodes a recursive decomposition of the set A, forming a layered abstraction often referred to as a spectral hierarchy or multiscale partition.

2.2 Basic Metrics on an Encoding Tree

Consider a non-negative, irreducible matrix $A \in \mathbb{R}_{\geq 0}^{n\times n}$, which represents an information system over n elements. Let $V = \{1, 2, \ldots, n\}$ denote the index set of rows and columns in A, corresponding to the objects in the system. An encoding tree for this system is defined over the set V.

To measure the structural properties of a system, we associate to A a stochastic process that models information flow. Let $\pi = (\pi_1, \ldots, \pi_n)$ be the stationary distribution of the Markov chain derived from A by row normalization. The probability of transitioning from object x to object y is defined by

$$p_{xy} = \pi_x \cdot b_{xy}, \quad \text{where } b_{xy} = \frac{a_{xy}}{\sum_{j=1}^{n} a_{xj}}.$$

Given a nonempty subset $X \subseteq V$, we define its *volume* as

$$V_X = \sum_{x \in X} \pi_x,$$

which represents the stationary probability mass of X.

The *incoming flow* into X is defined by $p_X = \sum_{y \notin X} \sum_{x \in X} p_{yx}$, representing the total probability of transitions from outside X into X.

Similarly, the *outgoing flow* from X is given by $q_X = \sum_{x \in X} \sum_{y \notin X} p_{xy}$, which quantifies the probability of escaping from X to the rest of the system.

These quantities provide a probabilistic characterization of how subsets of the system interact with each other under the dynamics encoded by A.

2.3 Structural Entropy under an Encoding Tree

To evaluate how well a given encoding tree T reflects the structure of the information system A, we define the structural entropy of A under T, which captures the information cost of describing the dynamics of A hierarchically along the tree.

Definition 2.2 (Structural Entropy under an Encoding Tree)**.** *Let A be a non-negative, irreducible matrix, and let T be an encoding tree over V. The* structural entropy *of A under T is defined by*

$$\mathcal{H}^T(A) = -\sum_{\substack{\alpha \in T \\ \alpha^- \neq \lambda}} p_\alpha \log_2 \frac{V_\alpha}{V_{\alpha^-}} = -\int_T p_\alpha \log_2 \frac{V_\alpha}{V_{\alpha^-}},$$

where $T_\alpha \subseteq V$ is the subset associated with node α, $V_\alpha = V_{T_\alpha}$ is its volume, $p_\alpha = p_{V_\alpha}$ is the incoming flow into T_α, and α^- denotes the parent node of α in the tree.

This quantity measures how well the tree structure aligns with the natural flow of information in the system. Low entropy indicates that the tree captures well-separated and coherent groupings of elements.

To assess the intrinsic complexity of an information system, we consider the optimal encoding tree that minimizes the structural entropy.

Definition 2.3 (Structural Entropy of an Information System)**.** *Given a non-negative, irreducible matrix A, the* structural entropy *of the information system is defined as*

$$\mathcal{H}(A) = \min_T \mathcal{H}^T(A),$$

where the minimum is taken over all encoding trees T over V.

This value quantifies the minimal cost of encoding the system under an optimal hierarchical abstraction. The **dimension** of structural entropy is characterized by the depth of the encoding tree T, representing the number of hierarchical partitioning levels. A key example is the **1-dimensional structural entropy (1D SE)**, corresponding to an encoding tree T_1 of depth one where the root is directly partitioned into leaf nodes representing individual entities. In this case, 1D SE simplifies to the Shannon entropy of the stationary distribution $\boldsymbol{\pi}$, i.e., $\mathcal{H}^1(A) = -\sum_{i=1}^{n} \pi_i \log_2 \pi_i$.

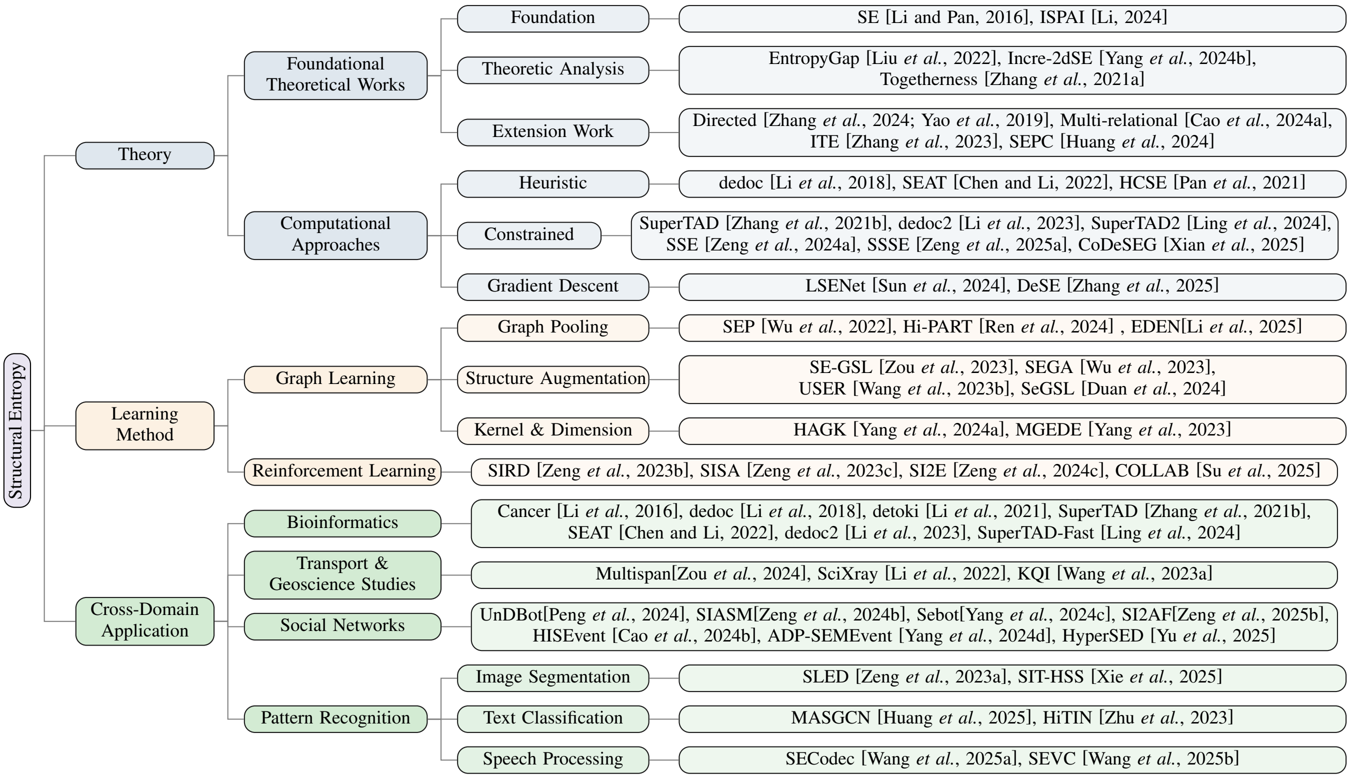

Figure 1: A taxonomy of Structural Entropy works.

3 Theoretical Foundations

3.1 Foundational Theoretical Works

The theoretical framework of SE was established through pioneering works that formalized its mathematical principles and later expanded them into a broader philosophy of information.

Foundation Theory The concept of SE was introduced by [Li and Pan, 2016] as a principled, information-theoretic framework for quantifying the complexity of a graph's structure. This seminal work defined SE as the minimum code length required to describe a network's hierarchy under a random walk process, effectively bridging graph theory with Shannon's information principles. It established that minimizing SE reveals a graph's intrinsic multi-scale organization. Building on this, [Li, 2024] significantly broadened the theoretical landscape by situating SE within the "mathematical principles of the information world." This work frames SE within the broader context of the Philosophy of AI, formulating fundamental information laws and the Information Science Principles of AI (ISPAI), establishing axiomatic foundations for a comprehensive structural information theory.

Theoretic Analysis Subsequent research has focused on elucidating SE's properties and forging connections with other established graph-theoretic concepts. In social network analysis, for example, [Zhang *et al.*, 2021a] introduced a "togetherness" metric that uses SE reduction to quantify the efficacy of network integration when adding inter-community edges. From a more fundamental perspective, [Liu *et al.*, 2022] established a significant theoretical link between SE and spectral graph theory. Their work proved that for any undirected graph, the difference between SE and the von Neumann graph entropy (VNE) is bounded within the interval $[0, \gamma]$, formally connecting SE to the spectral properties of the graph Laplacian.

Extension Work Recognizing that the original formulation targeted simple graphs, significant research has extended SE's principles to more complex network structures and dynamic scenarios. To handle directed graphs, multiple approaches have emerged; one introduced a localized 2D SE using a flow-based algorithm for financial networks [Yao *et al.*, 2019], while another proposed a random walk-based entropy (RWE) with a proven theoretical connection to VNE [Zhang *et al.*, 2024]. For heterogeneous relational data, multi-relational SE was developed using a random surfing model that jointly optimizes node and relation selection to enhance interpretability [Cao *et al.*, 2024a]. To handle evolving networks, [Yang *et al.*, 2024b] addressed the challenge of applying SE to dynamic graphs by proposing incremental computation paradigms, which include two distinct strategies for efficiently updating encoding trees and re-optimizing SE in response to network changes. The applicability of SE has also expanded into new domains, with extensions such as an "infected tree entropy" for source localization in information cascades [Zhang *et al.*, 2023] and an SE-guided probabilistic coding variant that regularizes embeddings in natural language understanding tasks [Huang *et al.*, 2024].

Discussion The evolution of SE from its core principles to a broad family of specialized variants highlights its versatility.

However, this expansion also introduces key challenges. Different SE formulations for various graph structures prioritize different network properties, and their comparative efficacy remains an open area of investigation. Incremental methods for dynamic graphs are efficient but may have limitations when faced with complex structural revolutions. A significant challenge is the development of a unified SE framework that can integrate these variants for complex, heterogeneous, and dynamic networks. While the link to spectral entropy has strengthened SE's theoretical underpinnings, a deeper exploration of its relationship with other graph-theoretic objectives—such as modularity [Newman and Girvan, 2004], the map equation [Rosvall and Bergstrom, 2008], and various centrality measures—is needed to fully situate it within network science. Future research should prioritize comprehensive comparative analyses of existing SE variants and the development of a unified framework applicable to a broader range of structures, including hypergraphs and multigraphs, to address the complexity of modern information systems.

3.2 Computational Approaches

As illustrated in Table 1 and Figure 1, the practical algorithms for SE minimization can be categorized by their optimization strategies and the types of constraints they employ.

Heuristic Methods Initial approaches to SE minimization relied on heuristic strategies to address various flat and hierarchical clustering problems. The first SE-based chromatin domain detector, deDoc, pioneered this area by employing a greedy merging strategy to identify genomic structures [Li *et al.*, 2018]. This principle was adapted for single-cell omics in the SEAT framework, which applies SE minimization to cell-cell graphs to reveal nested subpopulations and analyze functional diversity [Chen and Li, 2022]. In parallel, HCSE emerged as a widely used, general-purpose hierarchical clustering framework [Pan *et al.*, 2021]. By recursively isolating the sparsest levels of a hierarchy through SE-guided optimization, HCSE automatically determines the appropriate hierarchy depth without requiring hyperparameters. Its competitive performance against established methods like LOUVAIN [Newman and Girvan, 2004] on real-world networks validated heuristic SE minimization as a practical and effective approach to graph clustering.

Constrained Optimization More recent advances in SE minimization have incorporated domain-specific constraints to tackle complex biological and computational challenges. These methods can be systematically categorized by their constraint-handling mechanisms. One major category enforces spatial or index continuity constraints, which is crucial for applications like genomic domain detection where chromatin regions have an inherent linear organization. SuperTAD established this paradigm by combining interval dynamic programming with SE minimization, ensuring that detected topologically associating domains (TADs) preserve genomic contiguity [Zhang *et al.*, 2021b]. To improve efficiency, SuperTAD2 introduced matrix discretization and approximation strategies, achieving a significant speedup (e.g., 10x) while maintaining spatial continuity [Ling *et al.*, 2024]. The same continuity principles were extended to single-cell analysis in deDoc2, which uses dynamic programming to resolve contiguous domains and capture cell cycle-dependent chromatin dynamics [Li *et al.*, 2023]. A second category integrates semi-supervised constraints, such as labels and pairwise relationships, to guide the minimization process. The framework by [Zeng *et al.*, 2024a] unifies these constraints by reformulating SE to penalize violations of must-link/cannot-link pairs during encoding tree construction. To handle massive datasets, [Zeng *et al.*, 2025a] scaled this approach by introducing graph sampling and incremental cluster insertion, reducing time complexity from quadratic to nearly linear while preserving high accuracy through theoretically guaranteed approximations.

Gradient-Based Optimization and Differentiable SE A significant trend is the integration of SE principles with gradient-based optimization, enabled by the development of differentiable SE variants compatible with deep learning. These approaches typically transform the discrete nature of SE into a continuous objective by using soft assignment matrices, which represent cluster memberships probabilistically. This innovation allows SE to guide end-to-end learning via standard backpropagation. LSEnet exemplifies this by creating a differentiable SE formulation within hyperbolic space, facilitating cluster-free hierarchical partitioning [Sun *et al.*, 2024]. Similarly, the DeSE framework pioneers a "soft assignment structural entropy" that enables the co-design of graph topology and cluster assignments within a unified end-to-end model [Zhang *et al.*, 2025]. By making SE differentiable, these methods bridge the gap between information-theoretic clustering and deep learning, allowing for simultaneous optimization of representations and hierarchical structures.

Discussion The evolution of SE minimization algorithms reveals a fundamental trade-off between the efficiency of heuristics, the domain specificity of constrained optimization, and the integrative power of gradient-based methods. Heuristic approaches like HCSE and deDoc offer scalability and generality but provide less flexibility for incorporating complex domain knowledge. In contrast, constrained methods like SuperTAD and semi-supervised frameworks integrate specific knowledge to enhance relevance and accuracy, though sometimes at the cost of broader applicability. The emergence of gradient-based techniques such as LSEnet and DeSE signals a promising convergence, allowing SE to be integrated directly into deep learning pipelines. While these methods can jointly optimize representations and partitions, they introduce considerations of model complexity, hyperparameter sensitivity, and the interpretability of the learned structures. Distinct from these paradigms, game-theoretic approaches like CoDeSEG offer another promising direction, leveraging game theory for near-linear-time community detection [Xian *et al.*, 2025]. Ultimately, open challenges persist in achieving an optimal balance of scalability, performance, and interpretability, particularly for dynamic or heterogeneous networks. Future progress may lie in hybridizing these computational paradigms to harness their respective strengths.

Algorithm (Ref)	Levels	Constrain	Time Complexity	Github	Lang
HCSE [Pan *et al.*, 2021]	h	Unconstrained	$O(n^2)$	Link	Python
deDoc [Li *et al.*, 2018]	2/3	Unconstrained	$O(n \log^2 n)$ or $O(n^2)$	Link	Java
deDoc2 [Li *et al.*, 2023]	h	Unconstrained	$O(n^2)$	Link	Java
SSE [Zeng *et al.*, 2024a]	h	Label, Pairwise	$O(h(m \log n + n))$	Link	Python
SSSE [Zeng *et al.*, 2025a]	h	Label, Pairwise	$O(n \log^2 n + nkt)$	Link	Python
SuperTAD [Zhang *et al.*, 2021b]	h	Contiguity	$O(n^4 k^2 h)$	Link	C++
SuperTAD-Fast [Ling *et al.*, 2024]	h	Contiguity	$O(n^4 k^2 h)$	Link	C++
PYSEAT [Chen and Li, 2022]	h	Contiguity	$O(n \log n)$	Link	Python
LSENet [Sun *et al.*, 2024]	2	Unconstrained	$O(nm)$	Link	Python
CoDeSEG [Xian *et al.*, 2025]	2	Overlapping	$O(nt)$	Link	C++

Table 1: **Summary of Algorithms.** Here, n denotes the number of nodes, m is the number of edges, t is the number of iterations, k is the number of flat clusters (i.e., the number of parent nodes directly above the leaf nodes in the encoding tree). The **Levels** column indicates the number of hierarchies the algorithm can generate; h signifies that the algorithm can construct an encoding tree with an arbitrary number of levels. **Pairwise**: Specifies relationships between data points, including Must-Link (same cluster) and Cannot-Link (different clusters). **Label**: Pre-assigns specific data points to clusters as known labels. **Contiguity**: Ensures nodes in the same cluster have continuous indices. **Unconstrained**: No constraints are applied, allowing fully flexible clustering. **Overlapping**: Nodes can belong to multiple clusters.

4 Learning Methods via Structural Entropy

4.1 Graph Learning Paradigms

SE has emerged as a powerful information-theoretic tool for graph representation learning, enabling novel architectures and frameworks. It provides a principled, quantitative measure of a graph's inherent hierarchical organization, addressing a fundamental challenge in the field: how to learn from unstructured relational data in a robust and meaningful way.

Graph Pooling In graph-level tasks, pooling operations are essential for downsampling nodes to create a compact representation of the entire graph. However, traditional methods often suffer from information loss and limited expressiveness. Since the essence of a graph's function is often encoded in its multi-level community structure, methods that iteratively pool local nodes risk destroying this larger organizational map. SE Pooling [Wu *et al.*, 2022] directly addresses this by using global SE minimization to generate all cluster assignments simultaneously, preventing the cumulative damage to local structures seen in stepwise pooling. Further advancing this, the Hi-PART framework [Ren *et al.*, 2024] employs SE to construct Hierarchical Partition Trees. This approach explicitly models the nested, multi-scale nature of graph structures, enabling it to capture richer detail and surpass the expressive power of the standard Weisfeiler-Lehman test.

Structure Augmentation Graph structure augmentation, a cornerstone of modern self-supervised learning on graphs, involves creating altered "views" of a graph to teach a model what features are essential versus what is noise. Random augmentations risk discarding crucial information, so SE provides a principled way to guide this process by identifying and preserving the graph's intrinsic, low-entropy structure. A stable "anchor view" representing the graph's core semantic information is created. SEGA [Wu *et al.*, 2023] constructs such a view by finding an encoding tree with minimal SE, thus filtering out noise while maintaining the essential community hierarchy for more effective contrastive learning. Similarly, USER [Wang *et al.*, 2023b] incorporates SE directly into its objective function to learn an "innocuous graph." By minimizing SE, it seeks an intrinsic structure with well-defined communities that is robust to perturbations. Other methods use SE in a multi-stage optimization to refine graph structures. SE-GSL [Zou *et al.*, 2023] first maximizes 1D SE to enhance information content, then minimizes high-dimensional SE to abstract the graph into robust hierarchical trees, and finally uses deduction SE to guide an adaptive graph reconstruction. Extending this, [Duan *et al.*, 2024] also uses SE to build minimal-SE encoding trees, leveraging this core structure to guide the fusion of multiple graph views based on community influence.

Graph Kernels & Embedding Dimensions SE has also enhanced classical graph analysis tools. Graph kernels that measure similarity between graphs often struggle to capture complex structural patterns. The Hierarchical Abstract Graph Kernel (HAGK) [Yang *et al.*, 2024a] leverages SE to generate multi-level, abstract representations of each graph's intrinsic organization. The kernel then compares these SE-derived hierarchical structures, offering a more profound similarity measure than one based on local features alone. Furthermore, SE provides a principled solution to a persistent challenge in graph learning: determining the optimal embedding dimension. MGEDE [Yang *et al.*, 2023] calculates a novel high-level SE from multi-order adjacency matrices and minimizes it alongside attribute entropy. This process explicitly determines the most suitable number of embedding dimensions, balancing representational capacity with model complexity.

Discussion SE provides a unifying, information-theoretic foundation for graph learning, shifting the paradigm from heuristic-driven methods to a more principled approach focused on quantifying and optimizing a graph's inherent organization. By minimizing structural uncertainty, SE acts as a powerful regularizer, guiding models to learn from the essential hierarchical information within data while filtering out stochastic noise. This approach underpins diverse advancements, from constructing robust anchor views in SEGA to enabling expressive structural comparisons in kernels like

HAGK. Future work will likely focus on bridging SE with other theoretical frameworks and scaling these methods to handle the dynamics of massive real-world networks.

4.2 Reinforcement Learning Integration

Integrating SE principles into reinforcement learning RL offers novel approaches to address long-standing challenges in hierarchical learning, exploration, and multi-agent coordination. This integration equips RL agents with a mechanism to perceive and decompose complex problems, moving beyond flat state-action spaces toward a more structured, human-like understanding of their environment.

SE-Guided RL Frameworks SE provides a formal tool to automatically discover hierarchical structure in the fundamental components of an RL problem. To address the "curse of dimensionality" in State & Action Abstraction, SE helps cluster vast state or action spaces into manageable, hierarchical abstractions. The SISA framework [Zeng *et al.*, 2023c] automatically uncovers this structure by optimizing an encoding tree from a state graph, using SE to create an efficient and task-relevant state representation. In the multi-agent domain, SIRD [Zeng *et al.*, 2023b] reframes role discovery as hierarchical action space clustering by using SE minimization on an action graph to define distinct roles. For Intrinsic Motivation for Exploration, the SI2E framework [Zeng *et al.*, 2024c] formulates an intrinsic reward by maximizing value-conditional SE. This reward, derived from an SE-minimized state-action graph, encourages the agent to explore states where its actions lead to the greatest reduction in uncertainty about the environment's causal structure. More comprehensive frameworks like SIDM [Zeng *et al.*, 2025c] utilize SE at multiple stages to structure the entire Decision-Making and Skill Discovery process. COLLAB-MARL [Su *et al.*, 2025] uses SE to quantify interaction complexity, revealing emergent coalition structures by minimizing the SE of a dynamic agent interaction graph.

Discussion The integration of SE into reinforcement learning allows for a departure from "black-box" agents that learn through brute-force trial and error. It provides the mathematical tools to build a *cognitive architecture* for an agent, enabling it to parse its world into a hierarchy of concepts, skills, and goals. A structured understanding allows an agent to reason abstractly, plan over longer horizons, and transfer knowledge between related tasks—hallmarks of intelligent behavior. Frameworks like SISA and SIRD demonstrate how SE can automate the discovery of abstractions, a task that has historically required significant human engineering. Methods like SI2E show that exploration need not be random; it can be a deliberate process of information seeking aimed at resolving structural uncertainty. This results in agents that are not only more sample-efficient but also more interpretable. Looking forward, a key direction is scaling these methods to more complex, partially observable environments. Furthermore, incorporating SE into meta-RL and curriculum learning is a particularly promising frontier. This would enable agents to autonomously identify the hierarchical structure across a distribution of tasks, allowing them to learn *how to learn* by recognizing and adapting to shared structures. This could be a critical step toward developing more general and adaptive artificial intelligence.

5 Cross-Domain Applications

5.1 Bioinformatics

SE has demonstrated notable effectiveness in biological system analysis by enabling the precise detection of hierarchical structures and patterns.

The application of SE minimization in genomics began with foundational work in cancer subtyping, which correlated entropy-minimized gene expression patterns with clinical outcomes [Li *et al.*, 2016]. This principle was soon extended to the analysis of 3D genomic architecture. A major area of impact has been the identification of TADs from Hi-C data. The first SE-based detector, deDoc, pioneered this by identifying large-scale TAD-like structures [Li *et al.*, 2018]. Subsequent methods focused on refining this process. SuperTAD, for instance, employed dynamic programming to better resolve hierarchical chromatin domains [Zhang *et al.*, 2021b], and its successor, SuperTAD2, significantly accelerated this process through matrix discretization [Ling *et al.*, 2024].

The focus on hierarchical analysis has naturally translated to the higher resolution of single-cell omics. In this domain, tools like SEAT minimize global uncertainty in cell-cell graphs to improve pseudo-time inference by detecting hierarchical subpopulations across various omics datasets [Chen and Li, 2022]. SE has also been crucial for decoding chromatin dynamics at the single-cell level. deTOKI uses SE principles to analyze regulatory dynamics within individual cells [Li *et al.*, 2021], while the more recent deDoc2 uses dynamic programming to resolve how single-cell chromatin hierarchies change with the cell cycle, a significant advance in the field [Li *et al.*, 2023].

5.2 Transport and Geoscience Studies

SE minimization is also being applied to complex systems in transportation and geoscience, offering novel methods for spatio-temporal modeling and knowledge quantification.

In transportation, the transformer-based MultiSPANS framework improves traffic forecasting by using SE minimization to optimize its spatial attention mechanism. This approach generates hierarchical encoding trees of the road network, enhancing the model's interpretability and ability to capture multi-range spatial dependencies [Zou *et al.*, 2024].

In geoscience, SE provides a structural lens for analyzing the evolution of scientific knowledge. The "Scientific X-ray" method, for example, constructs "idea trees" from citation networks to visualize how scientific ideas are inherited and to identify high-potential research directions [Li *et al.*, 2022]. Building on this structural approach, another framework redefines knowledge quantification by measuring differences in hierarchical disorder within these same networks. This method produces a robust Knowledge Quantification Index that successfully identifies overlooked but influential works and Nobel Prize-winning topics. By focusing on information structure over semantic content, it provides a more objective measure of scientific impact that is less susceptible to manipulation [Wang *et al.*, 2023a].

5.3 Social Networks

SE minimization principles provide interpretable, hierarchy-aware frameworks that have significantly advanced forensic analysis, adversarial modeling, and knowledge discovery within social networks.

In bot detection, SE is applied in diverse ways. UnDBot offers an unsupervised approach, leveraging heterogeneous SE on multi-relational graphs to decode bot networks through community labeling [Peng *et al.*, 2024]. For greater robustness against adversarial attacks, SEBot integrates SE with multi-view contrastive learning to reduce uncertainty via hierarchical community encoding [Yang *et al.*, 2024c]. Moving from detection to proactive attack modeling, the SIASM framework uses conditional SE minimization to optimize follower selection, thereby maximizing a socialbot's influence while simultaneously evading detection [Zeng *et al.*, 2024b].

SE is also pivotal for analyzing information dissemination and system resilience. The SI2AF framework, for instance, evaluates the robustness of news detection systems by analyzing uncertainty and community hierarchies in user-post interactions. It introduces an SE-based influence metric that combines content relevance with community structure, enabling more realistic adversarial simulations [Zeng *et al.*, 2025b].

For social event detection, SE facilitates the identification of emergent topics without predefined event counts. HISEvent achieves this by incrementally building message graphs with 1D SE and then hierarchically detecting events using 2D SE optimization [Cao *et al.*, 2024b]. Addressing privacy, ADP-SEMEvent integrates adaptive differential privacy into the 2D SE minimization process, enabling accurate event detection while protecting user data [Yang *et al.*, 2024d]. To further improve performance, HyperSED models social messages in hyperbolic space, employing differentiable, SE-guided partitioning to significantly enhance both the accuracy and speed of event detection over previous methods [Yu *et al.*, 2025].

5.4 Pattern Recognition

SE has driven significant advances across pattern recognition by providing a principled way to model and optimize hierarchical structures in visual, textual, and speech data.

In visual computing, SE enhances feature preservation and segmentation. For object detection, SLED is an unsupervised framework that detects skin lesions by minimizing the entropy of multiscale superpixel graphs, using the resulting structural information to isolate outliers [Zeng *et al.*, 2023a]. For general-purpose segmentation, SIT-HSS first constructs a graph and then performs iterative merging, resulting in superior hierarchical superpixel segmentations that outperform other unsupervised methods [Xie *et al.*, 2025].

In natural language processing, SE enables the effective integration of syntactic structures into text models. HiTIN, a hierarchy-aware tree isomorphism network, leverages SE to inject the syntactic hierarchy of a sentence directly into its graph-based representation for more accurate hierarchical text classification [Zhu *et al.*, 2023]. Similarly, MASGCN applies an SE-based loss function to guide a graph convolutional network in capturing syntactic features for the specific task of aspect-based sentiment analysis, achieving state-of-the-art performance [Huang *et al.*, 2025].

In speech processing, SE has spurred innovations in both compression and voice conversion by identifying meaningful, cluster-based structures in audio signals. SECodec introduces a compressive speech representation by applying SE to cluster feature nodes, which allows for efficient compression with reduced distortion [Wang *et al.*, 2025a]. In a similar spirit, the SEVC framework for voice conversion uses 2D SE to group reference speech frames into semantic clusters. By mapping a source speaker's frames to these target clusters, it ensures a high-fidelity speaker transformation [Wang *et al.*, 2025b].

6 Conclusion

This survey has systematically charted the landscape of SE, consolidating its theoretical underpinnings, computational methodologies, and burgeoning applications across the field of artificial intelligence. By tracing its evolution from a novel information-theoretic measure for graph complexity to a versatile optimization principle integrated into advanced learning paradigms, we have aimed to fill a critical void in the literature. Our review reveals that SE offers a distinct and powerful lens for AI, shifting the focus from local, pairwise interactions to the global, hierarchical organization of complex systems. We have detailed how minimizing SE serves as a constructive principle to uncover a system's most efficient information-theoretic abstraction, a capability that has been effectively harnessed in graph learning for tasks like pooling and structure augmentation, and in reinforcement learning for discovering hierarchical state-action spaces and guiding exploration. The diverse cross-domain applications—from decoding genomic architectures in bioinformatics to modeling influence in social networks and enhancing pattern recognition—underscore SE's broad utility and impact.

In conclusion, Structural Entropy provides more than just a new set of tools; it offers a principled way of thinking about information and structure in a hierarchical world. By providing a rigorous method to quantify and optimize the organizational uncertainty within data, SE holds the potential to guide the development of AI systems that are not only more powerful but also more interpretable, robust, and theoretically grounded. As the complexity of our models and the systems they interact with continues to grow, the principles of Structural Entropy will be indispensable in our quest to build the next generation of artificial intelligence.

Acknowledgments

This work has been supported in part by NSFC through grants 62322202, 62441612 and 62432006, Local Science and Technology Development Fund of Hebei Province Guided by the Central Government of China through grant 246Z0102G, the "Pioneer" and "Leading Goose" R&D Program of Zhejiang through grant 2025C02044, National Key Laboratory under grant 241-HF-D07-01, Hebei Natural Science Foundation through grant F2024210008, National Key R&D Program of China through grant 2021YFB3500700, and State Key Laboratory of Complex & Critical Software Environment (CCSE-2024ZX-20).

References

[Cao *et al.*, 2024a] Yuwei Cao, Hao Peng, Angsheng Li, Chenyu You, Zhifeng Hao, and Yu Philip. Multi-relational structural entropy. In *Proceedings of the UAI*, 2024.

[Cao *et al.*, 2024b] Yuwei Cao, Hao Peng, Zhengtao Yu, and S Yu Philip. Hierarchical and incremental structural entropy minimization for unsupervised social event detection. In *Proceedings of the AAAI*, 2024.

[Chen and Li, 2022] Lingxi Chen and Shuai Cheng Li. Incorporating cell hierarchy to decipher the functional diversity of single cells. *Nucleic Acids Research*, 2022.

[Duan *et al.*, 2024] Liang Duan, Xiang Chen, and et al. Structural entropy based graph structure learning for node classification. In *Proceedings of the AAAI*, 2024.

[Huang *et al.*, 2024] Xiang Huang, Hao Peng, Li Sun, and et al. Structural entropy guided probabilistic coding. In *Proceedings of the AAAI*, 2024.

[Huang *et al.*, 2025] Xiang Huang, Hao Peng, Shuo Sun, and et al. Multi-view attention syntactic enhanced graph convolutional network for aspect-based sentiment analysis. In *Proceedings of the DASFAA*, 2025.

[Li and Pan, 2016] Angsheng Li and Yicheng Pan. Structural information and dynamical complexity of networks. *IEEE Transactions on Information Theory*, 2016.

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

[Li *et al.*, 2018] Angsheng Li, Xianchen Yin, Bingxiang Xu, and et al. Decoding topologically associating domains with ultra-low resolution hi-c data by graph structural entropy. *Nature communications*, 2018.

[Li *et al.*, 2021] Xiao Li, Guangjie Zeng, Angsheng Li, and Zhihua Zhang. Detoki identifies and characterizes the dynamics of chromatin tad-like domains in a single cell. *Genome biology*, 2021.

[Li *et al.*, 2022] Qi Li, Xinbing Wang, and Luoyi andet al. Fu. Scientific x-ray: Scanning and quantifying the idea evolution of scientific publications. *Plos one*, 2022.

[Li *et al.*, 2023] Angsheng Li, Guangjie Zeng, Haoyu Wang, Xiao Li, and Zhihua Zhang. Dedoc2 identifies and characterizes the hierarchy and dynamics of chromatin tad-like domains in the single cells. *Advanced Science*, 2023.

[Li *et al.*, 2025] Xunkai Li, Zhengyu Wu, Kaichi Yu, Hongchao Qin, Guang Zeng, Rong-Hua Li, and Guoren Wang. Toward data-centric directed graph learning: An entropy-driven approach. In *ICML*, 2025.

[Li, 2024] Angsheng Li. *Science of Artificial Intelligence: Mathematical Principles of Intelligence* (In Chinese). Science Press, Beijing, 2024.

[Ling *et al.*, 2024] Zhao Ling, Yu Wei Zhang, and Shuai Cheng Li. Supertad-fast: Accelerating topologically associating domains detection through discretization. *Journal of Computational Biology*, 31(9), 2024.

[Liu *et al.*, 2019] Yiwei Liu, Jiamou Liu, Zijian Zhang, and et al. Rem: From structural entropy to community structure deception. In *Proceedings of the NeurIPS*, 2019.

[Liu *et al.*, 2022] Xuecheng Liu, Luoyi Fu, Xinbing Wang, and Chenghu Zhou. On the similarity between von neumann graph entropy and structural information: Interpretation, computation, and applications. *IEEE Transactions on Information Theory*, 2022.

[Newman and Girvan, 2004] Mark EJ Newman and Michelle Girvan. Finding and evaluating community structure in networks. *Physical review E*, 69(2):026113, 2004.

[Pan *et al.*, 2021] Yicheng Pan, Feng Zheng, and Bingchen Fan. An information-theoretic perspective of hierarchical clustering. *arXiv preprint arXiv:2108.06036*, 2021.

[Peng *et al.*, 2024] Hao Peng, Jingyun Zhang, Xiang Huang, and et al. Unsupervised social bot detection via structural information theory. *ACM TOIS*, 2024.

[Ren *et al.*, 2024] Yuyang Ren, Haonan Zhang, Luoyi Fu, and et al. Hi-part: Going beyond graph pooling with hierarchical partition tree for graph-level representation learning. *ACM TKDD*, 2024.

[Rosvall and Bergstrom, 2008] Martin Rosvall and Carl T Bergstrom. Maps of random walks on complex networks reveal community structure. *Proceedings of the national academy of sciences*, 2008.

[Shannon, 1948] Claude E Shannon. A mathematical theory of communication. *The Bell System Technical Journal*, 27(3):379–423, 1948.

[Su *et al.*, 2025] Dingli Su, Hao Peng, Guangjie Zeng, Pu Li, Angsheng Li, and Yicheng Pan. Emergence of cooperation in multi-agent reinforcement learning via coalition labeling and structural entropy. In *Proceedings of the SDM*, 2025.

[Sun *et al.*, 2024] Li Sun, Zhenhao Huang, and et al. Lsenet: Lorentz structural entropy neural network for deep graph clustering. In *Proceedings of the ICML*, 2024.

[Tishby *et al.*, 2000] Naftali Tishby, Fernando C Pereira, and William Bialek. The information bottleneck method. In *Proceedings of the 37th annual Allerton conference on communication, control, and computing*, pages 368–377, 2000.

[Wang *et al.*, 2023a] Xinbing Wang, Huquan Kang, Luoyi Fu, and et al. Quantifying knowledge from the perspective of information structurization. *Plos one*, 2023.

[Wang *et al.*, 2023b] Yifei Wang, Yupan Wang, Zeyu Zhang, Song Yang, Kaiqi Zhao, and Jiamou Liu. User: Unsupervised structural entropy-based robust graph neural network. In *Proceedings of the AAAI*, volume 37, 2023.

[Wang *et al.*, 2025a] Linqin Wang, Yaping Liu, Zhengtao Yu, and et al. Secodec: Structural entropy-based compressive speech representation codec for speech language models. In *Proceedings of the AAAI*, 2025.

[Wang *et al.*, 2025b] Linqin Wang, Zhengtao Yu, Shengxiang Gao, and et al. Voice conversion via structural entropy. In *Proceedings of the IEEE ICASSP*, 2025.

[Wu *et al.*, 2022] Junran Wu, Xueyuan Chen, Ke Xu, and Shangzhe Li. Structural entropy guided graph hierarchical pooling. In *Proceedings of the ICML*, 2022.

[Wu *et al.*, 2023] Junran Wu, Xueyuan Chen, Bowen Shi, Shangzhe Li, and Ke Xu. Sega: Structural entropy guided anchor view for graph contrastive learning. In *ICML*, 2023.

[Xian *et al.*, 2025] Yantuan Xian, Pu Li, and et al. Community detection in large-scale complex networks via structural entropy game. In *Proceedings of the Web Conference*, 2025.

[Xie *et al.*, 2025] Minhui Xie, Hao Peng, Pu Li, and et al. Hierarchical superpixel segmentation via structural information theory. In *Proceedings of the SDM*, 2025.

[Yang *et al.*, 2023] Zhenyu Yang, Ge Zhang, Jia Wu, and et al. Minimum entropy principle guided graph neural networks. In *Proceedings of the ACM WSDM*, 2023.

[Yang *et al.*, 2024a] Runze Yang, Hao Peng, Angsheng Li, and et al. Hierarchical abstracting graph kernel. *IEEE TKDE*, 2024.

[Yang *et al.*, 2024b] Runze Yang, Hao Peng, Chunyang Liu, and Angsheng Li. Incremental measurement of structural entropy for dynamic graphs. *Artificial Intelligence*, 2024.

[Yang *et al.*, 2024c] Yingguang Yang, Qi Wu, Buyun He, and et al. Sebot: Structural entropy guided multi-view contrastive learning for social bot detection. In *Proceedings of the ACM SIGKDD*, 2024.

[Yang *et al.*, 2024d] Zhiwei Yang, Yuecen Wei, Haoran Li, and et al. Adaptive differentially private structural entropy minimization for unsupervised social event detection. In *Proceedings of the ACM CIKM*, 2024.

[Yao *et al.*, 2019] Yuan Yao, Yicheng Pan, Shaojiang Wang, Hongan Wang, and Waqas Nazeer. Flow-based clustering on directed graphs: A structural entropy minimization approach. *IEEE Access*, 2019.

[Yu *et al.*, 2025] Xiaoyan Yu, Yifan Wei, Shuaishuai Zhou, and et al. Towards effective, efficient and unsupervised social event detection in the hyperbolic space. In *Proceedings of the AAAI*, 2025.

[Zeng *et al.*, 2023a] Guangjie Zeng, Hao Peng, Angsheng Li, and et al. Unsupervised skin lesion segmentation via structural entropy minimization on multi-scale superpixel graphs. In *Proceedings of the ICDM*, 2023.

[Zeng *et al.*, 2023b] Xianghua Zeng, Hao Peng, and Angsheng Li. Effective and stable role-based multi-agent collaboration by structural information principles. In *Proceedings of the AAAI*, 2023.

[Zeng *et al.*, 2023c] Xianghua Zeng, Hao Peng, Angsheng Li, and et al. Hierarchical state abstraction based on structural information principles. In *Proceedings of the IJCAI*, 2023.

[Zeng *et al.*, 2024a] Guangjie Zeng, Hao Peng, and et al. Semi-supervised clustering via structural entropy with different constraints. In *Proceedings of the SDM*, 2024.

[Zeng *et al.*, 2024b] Xianghua Zeng, Hao Peng, and Angsheng Li. Adversarial socialbots modeling based on structural information principles. In *Proceedings of the AAAI*, 2024.

[Zeng *et al.*, 2024c] Xianghua Zeng, Hao Peng, and Angsheng Li. Effective exploration based on the structural information principles. In *Proceedings of the NeurIPS*, 2024.

[Zeng *et al.*, 2025a] Guangjie Zeng, Hao Peng, Angsheng Li, and et al. Scalable semi-supervised clustering via structural entropy with different constraints. *IEEE TKDE*, 2025.

[Zeng *et al.*, 2025b] Xianghua Zeng, Hao Peng, and Angsheng Li. Robustness evaluation of graph-based news detection using network structural information. In *Proceedings of the ACM SIGKDD*, 2025.

[Zeng *et al.*, 2025c] Xianghua Zeng, Hao Peng, DingLi Su, and Angsheng Li. Hierarchical decision making based on structural information principles. *JMLR*, 2025.

[Zhang *et al.*, 2021a] Siyu Zhang, Jiamou Liu, Yiwei Liu, Zijian Zhang, and Bakhadyr Khoussainov. Improving togetherness using structural entropy. In *Proceedings of the MobiCASE*. Springer, 2021.

[Zhang *et al.*, 2021b] Yu Wei Zhang, Meng Bo Wang, and Shuai Cheng Li. Supertad: robust detection of hierarchical topologically associated domains with optimized structural information. *Genome biology*, 2021.

[Zhang *et al.*, 2023] Chong Zhang, Qiang Guo, Luoyi Fu, and et al. Finding the source in networks: an approach based on structural entropy. *ACMT TOIT*, 2023.

[Zhang *et al.*, 2024] Chong Zhang, Cheng Deng, Luoyi Fu, and et al. Rwe: A random walk based graph entropy for the structural complexity of directed networks. *IEEE TNSE*, 2024.

[Zhang *et al.*, 2025] Jingyun Zhang, Li Sun, Wu, Zhengtao Yu, and et al. Unsupervised graph clustering with deep structural entropy. In *Proceedings of the ACM SIGKDD*, 2025.

[Zhu *et al.*, 2023] He Zhu, Chong Zhang, Junjie Huang, Junran Wu, and Ke Xu. Hitin: Hierarchy-aware tree isomorphism network for hierarchical text classification. In *Proceedings of the ACL*, 2023.

[Zou *et al.*, 2023] Dongcheng Zou, Hao Peng, Xiang Huang, and et al. Se-gsl: A general and effective graph structure learning framework through structural entropy optimization. In *Proceedings of the ACM Web Conference*, 2023.

[Zou *et al.*, 2024] Dongcheng Zou, Senzhang Wang, and et al. Multispans: A multi-range spatial-temporal transformer network for traffic forecast via structural entropy optimization. In *Proceedings of the ACM WSDM*, 2024.