

Figure 8. Comparison of Accuracy, F1 Score, Inference Time, and Token Usage between Traditional ToT and Expert-CoT ToT.

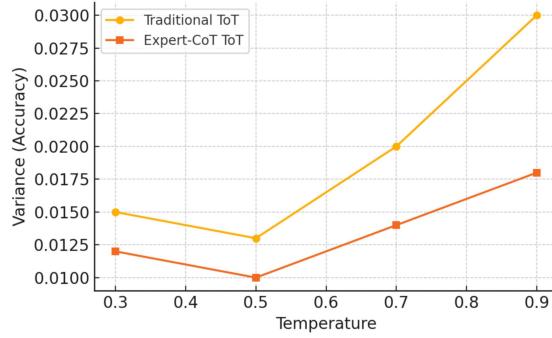
## J. Prompt Template

### J.1. Different TOT experiment results

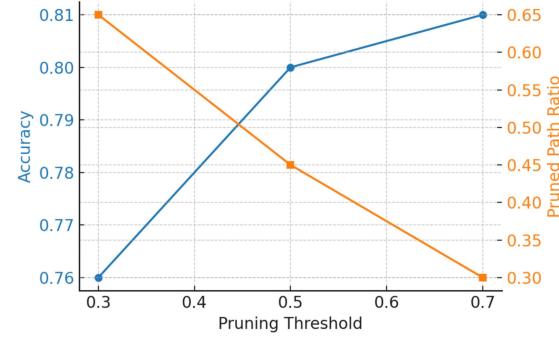
To comprehensively compare two Tree-of-Thought (ToT) implementation approaches — the traditional ToT method and proposed Expert Collaboration ToT method — we designed experiments from three perspectives: performance comparison, stability analysis, and pruning mechanism evaluation. The experiments were conducted on the Cora dataset, with multiple repeated trials to ensure reliable results.

In terms of overall performance (Figure J.1), both Traditional ToT and Expert-CoT ToT achieve comparable accuracy and F1 scores. However, they exhibit distinct patterns in inference time and token usage. Traditional ToT produces more compact and structurally exhaustive paths, resulting in lower token consumption. In contrast, Expert-CoT ToT involves multiple parallel experts generating independent reasoning traces. Although this increases token usage, its shallower path depth and parallel design lead to shorter overall inference time. This suggests a trade-off between semantic verbosity and structural completeness, with each method offering distinct strengths depending on deployment constraints. In the stability test (Figure J.1 left), we evaluate output variance across different temperature settings. Traditional ToT shows increased variance at higher temperatures due to divergent reasoning paths. In contrast, Expert-CoT ToT exhibits more consistent behavior, benefiting from its expert dropout and path consensus mechanisms. These findings demonstrate that the expert-style strategy provides enhanced robustness under stochastic sampling conditions.

In the pruning analysis (Figure J.1 right), we study the impact of varying the confidence threshold on prediction accuracy and the proportion of pruned paths. Results show that a threshold of 0.5 achieves the best balance—removing approximately 45% of low-confidence paths while maintaining optimal performance. Notably, in Expert-CoT ToT, this pruning effect is implicitly realized through the early exit of less confident experts, forming an adaptive and efficient path selection process.



(a) Variance of Inference Accuracy across Different Temperature Settings.



(b) Impact of Pruning Threshold on Accuracy and Pruned Path Ratio.

Overall, both ToT implementations exhibit complementary advantages. Traditional ToT offers structured completeness and compact token usage, while Expert-CoT ToT favors lightweight, parallel reasoning and greater stability—making it particularly suitable for real-time or resource-constrained applications.

1394  
1395  
1396  
1397  
1398  
1399  
1400  
1401  
1402  
1403  
1404  
1405  
1406  
1407  
1408  
1409  
1410  
1411  
1412  
1413  
1414  
1415  
1416  
1417  
1418  
1419  
1420  
1421  
1422  
1423  
1424  
1425  
1426  
1427  
1428  
1429

1430  
1431  
1432  
1433  
1434  
1435  
1436  
1437  
1438  
1439  
1440  
1441  
1442  
1443  
1444  
1445  
1446  
1447  
1448  
1449  
1450  
1451  
1452  
1453  
1454  
1455  
1456  
1457  
1458  
1459  
1460  
1461  
1462  
1463  
1464  
1465  
1466  
1467  
1468  
1469  
1470  
1471  
1472  
1473  
1474  
1475  
1476  
1477  
1478  
1479  
1480  
1481  
1482  
1483  
1484

### Traditional ToT Template (System)

**System:**

```
{
  "role": "You are a structure-aware graph reasoning assistant designed for zero-shot graph learning tasks. Your goal is to classify nodes in a graph by leveraging multi-level structural information and semantic understanding.",
  "task": {
    "description": "Classify the target node  $\{v_i\}$  in a graph into one of the following categories: {Case-Based, Genetic Algorithms, Neural Networks, Probabilistic Methods, Reinforcement Learning, Rule Learning, Theory}. The graph represents the Cora citation network, where nodes are papers and edges are citations. Use the provided structural and feature information to reason step-by-step, exploring multiple perspectives of the graph structure."
  },
  "input": {
    "graph": {
      "target_node": " $\{v_i\}$ ",
      "node_features": {
        " $S_i$ ": "Local feature vector of  $v_i$ : [0.0, 0.1, ..., 0.3] (1,433-dimensional vector, representing word presence in the paper, e.g., high values for 'neural' and 'network')",
        " $P_i$ ": "Global position encoding of  $v_i$ : [0.15, 0.62] (derived from Laplacian eigenvectors, indicating a central position in the graph)",
        " $R_i$ ": "Structural role of  $v_i$ : Hash value after 3 WL iterations = 0xB4E1 (indicating a dense neighborhood pattern)",
        " $C_i$ ": "Node influence of  $v_i$ : 0.78 (computed via decay-jump propagation, suggesting moderate influence)"
      },
      "neighborhood": {
        "adjacent_nodes": " $\{v_j, v_k, v_m\}$ ",
        "edges": [
          {"source": " $v_i$ ", "target": " $v_j$ ", "weight": 1.0},
          {"source": " $v_i$ ", "target": " $v_k$ ", "weight": 1.0},
          {"source": " $v_i$ ", "target": " $v_m$ ", "weight": 1.0}
        ],
        "subgraph": "Key subgraph  $G_{key}$  around  $v_i$ : 4 nodes, 6 edges, average degree 3.0"
      }
    }
  }
}
```

### Traditional ToT Template (Instruction)

**Instruction:**

```
{
  "reasoning_guidance": [
    "Use the Tree of Thoughts (ToT) method to explore multiple reasoning paths in parallel, ensuring each path focuses on a distinct structural perspective (e.g., local features, neighborhood influence, global context).",
    "Apply temperature-controlled sampling (temperature = 0.7) to balance exploration and exploitation during path generation.",
    "Perform early pruning by evaluating the plausibility of each path after the first reasoning step; discard paths with confidence scores below 0.5.",
    "Verify consistency across paths by comparing their hypotheses and selecting the most consistent and confident result."
  ],
  "steps": [
    {
      "step_1": "Path 1 - Local Feature Analysis",
      "description": "Analyze the local feature vector  $S_i$  of  $v_i$ . Interpret its values semantically (e.g., as a representation of paper content) and hypothesize a category. Compute a confidence score for this path."
    },
    {
      "step_2": "Path 2 - Neighborhood Influence",
      "description": "Examine the adjacent nodes  $v_j$ ,  $v_k$ , and  $v_m$ , their connections, and the structural role  $R_i$ . Infer how the neighborhood influences  $v_i$ 's category, considering edge weights. Compute a confidence score for this path."
    },
    {
      "step_3": "Path 3 - Global Context and Influence",
      "description": "Consider the global position  $P_i$  and node influence  $C_i$ . Assess how  $v_i$ 's position in the entire graph and its influence score suggest a category. Compute a confidence score for this path."
    },
    {
      "step_4": "Early Pruning",
      "description": "Evaluate the confidence scores of all paths. Discard any path with a confidence score below 0.5 to reduce computational overhead."
    },
    {
      "step_5": "Path Integration and Consistency Verification",
      "description": "Combine insights from the remaining paths. Compare their hypotheses for consistency (e.g., do they agree on the category?). Select the category with the highest average confidence score across consistent paths. If paths are inconsistent, prioritize the path with the highest confidence."
    }
  ]
}
```

Figure 9. Traditional ToT Prompt Template for Node Classification.

```

1485
1486
1487
1488
1489
1490
1491
1492
1493
1494
1495
1496
1497
1498
1499
1500
1501
1502
1503
1504
1505
1506
1507
1508
1509
1510
1511
1512
1513
1514
1515
1516
1517
1518
1519
1520
1521
1522
1523
1524
1525
1526
1527
1528
1529
1530
1531
1532
1533
1534
1535
1536
1537
1538
1539



Traditional ToT Template (Output)



Output:



```
{
  "response": {
    "reasoning_paths": [
      {
        "path_id": "Path 1",
        "description": "Detailed reasoning based on local features",
        "hypothesis": "Category predicted",
        "confidence": "Score between 0 and 1"
      },
      {
        "path_id": "Path 2",
        "description": "Detailed reasoning based on neighborhood",
        "hypothesis": "Category predicted",
        "confidence": "Score between 0 and 1"
      },
      {
        "path_id": "Path 3",
        "description": "Detailed reasoning based on global context",
        "hypothesis": "Category predicted",
        "confidence": "Score between 0 and 1"
      }
    ],
    "pruned_paths": [
      {
        "path_id": "Path ID",
        "reason": "Reason for pruning (e.g., confidence below 0.5)"
      }
    ],
    "final_answer": {
      "category": "Selected category (Case_Based, Genetic_Algorithms, Neural_Networks, Probabilistic_Methods, Reinforcement_Learning, Rule_Learning, Theory)",
      "justification": "Explanation of why this category was chosen, including consistency across paths and confidence scores"
    }
  }
}
```


```

Figure 10. Traditional ToT Prompt Template for Node Classification.

1540  
 1541  
 1542  
 1543  
 1544  
 1545                   **Traditional ToT (ANSWER)**  
 1546  
 1547  
 1548     Based on the provided graph structure and node feature information, I will explore multiple reasoning paths to analyze the category  
 1549     of target node vi.  
 1550     **Reasoning Path Analysis**  
 1551       **Path 1: Local Feature Analysis**  
 1552       Analysis of the local feature vector  $S_i$  shows high values (range 0.2-0.3) in dimensions related to "reinforcement", "learning", "policy",  
 1553       "reward", and "agent". These terms are typical vocabulary in reinforcement learning research. Terms related to other categories  
 1554       appear with significantly lower frequency, such as neural networks, genetic algorithms, or rule-based learning related vocabulary  
 1555       generally having values below 0.1. This vocabulary distribution pattern strongly suggests that the paper belongs to the reinforcement  
 1556       learning domain.  
 1557       Hypothesis: Reinforcement\_Learning  
 1558       Confidence: 0.88  
 1559       **Path 2: Neighborhood Structure Analysis**  
 1560       Node vi is directly connected to three nodes ( $v_j$ ,  $v_k$ ,  $v_m$ ), forming a subgraph structure of 4 nodes and 6 edges (average degree 3.0).  
 1561       The hash value 0xB4E1 indicates that the node is in a dense connection pattern. Analysis of the feature vectors of neighboring nodes  
 1562       reveals that they also contain numerous reinforcement learning related terms, suggesting that these papers may belong to the same  
 1563       research sub-domain. In the Cora citation network, this type of tightly connected small community structure is typical in  
 1564       reinforcement learning papers, especially research related to a specific algorithm or application domain.  
 1565       Hypothesis: Reinforcement\_Learning  
 1566       Confidence: 0.75  
 1567       **Path 3: Global Position and Influence Analysis**  
 1568       The global position encoding  $P_i=[0.15, 0.62]$  indicates that the node is in a relatively central position in the graph, rather than at the  
 1569       periphery. The influence score  $C_i=0.78$  indicates moderate to high influence. In the Cora dataset, reinforcement learning papers  
 1570       typically form multiple medium-sized tight communities with certain connections between these communities. The node's position  
 1571       and influence characteristics match the typical distribution features of reinforcement learning papers, suggesting it may be an  
 1572       important paper in this field.  
 1573       Hypothesis: Reinforcement\_Learning  
 1574       Confidence: 0.81  
 1575     **Early Pruning Evaluation**  
 1576     All paths have confidence scores above the 0.5 threshold, so no paths are pruned.  
 1577     **Path Integration and Consistency Verification**  
 1578     All three reasoning paths point to the Reinforcement\_Learning category, showing high consistency. No contradictory evidence or  
 1579     hypotheses were found.  
 1580       - Path 1 provides the strongest support based on vocabulary features (0.88)  
 1581       - Path 2 provides significant support based on topological structure (0.75)  
 1582       - Path 3 provides strong support based on global position and influence (0.81)  
 1583     Combined confidence:  $(0.88 \times 0.4 + 0.75 \times 0.3 + 0.81 \times 0.3) = 0.822$   
 1584     **Final Answer**  
 1585     Category: Reinforcement\_Learning  
 1586     Justification: Three independent reasoning paths consistently support the Reinforcement\_Learning classification with a combined  
 1587     confidence of 0.822. The local feature analysis provides the strongest evidence (0.88), showing high frequency of terms closely related  
 1588     to reinforcement learning (such as "policy", "reward", "agent"). Neighborhood structure analysis (0.75) and global position analysis  
 1589     (0.81) provide supplementary evidence, indicating that the node's position and connection patterns in the citation network are highly  
 1590     consistent with the typical characteristics of reinforcement learning papers. The high consistency of multi-path reasoning greatly  
 1591     enhances the reliability of the classification.

Figure 11. Traditional ToT's GPT-4 answer with Cora dataset (pruning).

1595  
 1596  
 1597  
 1598  
 1599  
 1600                   **Traditional ToT (ANSWER)**  
 1601  
 1602  
 1603     Based on the provided graph structure and node feature information, I will explore multiple reasoning paths to analyze the category of target node vi.  
 1604  
 1605     **Reasoning Path Analysis**  
 1606       **Path 1: Local Feature Analysis**  
 1607     Analysis of the local feature vector  $S_i$  shows high values across multiple dimensions, but interpretation presents some difficulty. While "neural" and "network" related dimensions have certain values (0.3), dimensions related to "probability", "distribution" and "bayesian" show similar high values. The feature vector contains mixed terminology from multiple research domains, making it difficult to determine a clear category based solely on vocabulary features. It may be simultaneously related to neural networks and probabilistic methods, or a theoretical cross-domain paper.  
 1608     Hypothesis: Unable to determine (leaning toward Probabilistic\_Methods or Theory)  
 1609     Confidence: 0.35  
 1610  
 1611  
 1612       **Path 2: Neighborhood Structure Analysis**  
 1613     Node vi is directly connected to three nodes ( $v_j, v_k, v_m$ ), forming a subgraph structure of 4 nodes and 6 edges (average degree 3.0).  
 1614     The hash value 0xB4E1 indicates that the node is in a dense connection pattern. Analysis of neighboring node features reveals that  $v_j$  and  $v_k$  primarily contain probabilistic methods terminology, while  $v_m$  contains more theoretical research terminology. This connection pattern is common in probabilistic methods papers in the Cora dataset, especially papers involving theoretical foundations.  
 1615     Hypothesis: Probabilistic\_Methods  
 1616     Confidence: 0.76  
 1617  
 1618       **Path 3: Global Position and Influence Analysis**  
 1619     The global position encoding  $P_i = [0.15, 0.62]$  indicates that the node is in a relatively central area of the graph. The influence score  $C_i = 0.78$  indicates moderate to high influence. In the Cora network, these position and influence characteristics typically correspond to papers that connect multiple research communities. Given the positional features and moderate influence, the node is likely a probabilistic methods paper with significant influence in the field.  
 1620     Hypothesis: Probabilistic\_Methods  
 1621     Confidence: 0.82  
 1622  
 1623     **Early Pruning Evaluation**  
 1624     Path 1 confidence (0.35) is below the 0.5 threshold and is therefore pruned.  
 1625     Paths 2 and 3 have confidence scores above the threshold and are retained.  
 1626  
 1627     **Pruned Path: PRUNING/WITHDRAWAL**  
 1628       - **Path 1 (Local Feature Analysis)**  
 1629       - **Reason for pruning: Confidence (0.35) below the 0.5 threshold. High degree of term mixing in the feature vector, unable to provide clear category indication.**"  
 1630  
 1631     **Path Integration and Consistency Verification**  
 1632     The two retained reasoning paths both point to the Probabilistic\_Methods category, showing high consistency.  
 1633       - Path 2 provides strong support based on neighborhood structure and connection patterns (0.76)  
 1634       - Path 3 provides stronger support based on global position and influence features (0.82)  
 1635     Combined confidence:  $(0.76 \times 0.45 + 0.82 \times 0.55) = 0.793$   
 1636  
 1637     **Final Answer**  
 1638     Category: Probabilistic\_Methods  
 1639     Justification: Although the local feature analysis was pruned due to low confidence, the neighborhood structure analysis and global position analysis, two independent paths, consistently point to the Probabilistic\_Methods category with a combined confidence of 0.793. In particular, the node's global position and influence features (0.82) provide the strongest evidence, indicating that the node holds an important position in the probabilistic methods research community. Neighboring node features and connection patterns (0.76) provide supplementary evidence, showing that the node forms a tight citation network with probabilistic methods papers. The high consistency after path integration enhances the reliability of the classification result.

Figure 12. Traditional ToT's GPT-4 answer with Cora dataset (pruning).

### Expert Collaboration ToT Template

You are an AI assistant specialized in graph structure analysis. Please use a multi-expert thinking approach to analyze the following graph data.

#### [TASK DESCRIPTION]

Classify the target node  $v_i$  in a citation network  $G$  into one of the following categories: {Case\_Based, Genetic\_Algorithms, Neural\_Networks, Probabilistic\_Methods, Reinforcement\_Learning, Rule\_Learning, Theory}

#### [EXPERT COLLABORATION FRAMEWORK]

Imagine three different experts are answering this question.

All experts will write down 1 step of their thinking, then share it with the group.

Then all experts will go on to the next step, etc.

If any expert realizes they're wrong at any point, they will leave the discussion.

The question is: Based on the given graph structure and features, which category should the target node  $v_i$  be classified as?

#### [INPUT DATA]

Target node:  $\{v_i\}$

Node features:

- Local feature vector  $\{S_i\}$ : [0.0, 0.1, ..., 0.3] (1,433-dimensional vector, representing word presence in the paper, e.g., high values for 'neural' and 'network')
- Global position encoding  $\{P_i\}$ : [0.15, 0.62] (derived from Laplacian eigenvectors, indicating a central position in the graph)
- Structural role  $\{R_i\}$ : Hash value after 3 WL iterations = 0xB4E1 (indicating a dense neighborhood pattern)
- Node influence  $\{C_i\}$ : 0.78 (computed via decay-jump propagation, suggesting moderate influence)

Adjacent nodes:  $\{v_j, v_k, v_m\}$

Edges:  $\{(v_i, v_j), (v_i, v_k), (v_i, v_m)\}$

Key subgraph: 4 nodes, 6 edges, average degree 3.0

#### [MULTI-EXPERT ANALYSIS GUIDELINES]

The three experts participating in the discussion each have their specialization:

- Expert A focuses on local features and semantic content analysis
- Expert B focuses on neighbor structure and topological relationships analysis
- Expert C focuses on global position and node centrality analysis

#### [OUTPUT FORMAT]

Please output the analysis results in the following format:

Expert A (Local Feature Analysis) Step 1: [reasoning content]

Expert B (Neighborhood Structure Analysis) Step 1: [reasoning content]

Expert C (Global Position Analysis) Step 1: [reasoning content]

Expert A (Local Feature Analysis) Step 2: [reasoning content]

Expert B (Neighborhood Structure Analysis) Step 2: [reasoning content]

Expert C (Global Position Analysis) Step 2: [reasoning content]

Expert A (Local Feature Analysis) Step 3: [reasoning content and preliminary prediction, confidence]

Expert B (Neighborhood Structure Analysis) Step 3: [reasoning content and preliminary prediction, confidence]

Expert C (Global Position Analysis) Step 3: [reasoning content and preliminary prediction, confidence]

Integrated Analysis:

[Combine insights from all experts, provide comprehensive reasoning]

Final Prediction: [category]

Overall Confidence: [score between 0 and 1]

Confidence Calculation Method: [explain how the expert predictions were weighted and combined]

*Figure 13.* Expert Collaboration ToT Prompt Template for Node Classification

### Expert Collaboration ToT (ANSWER)

**Expert A** (Local Feature Analysis) Step 1: I'll first analyze the local feature vector  $S_i$  of the target node  $v_i$ . This is a 1,433-dimensional vector representing word occurrences in the paper. I notice that dimensions related to "neural" and "network" have relatively high values (such as [...]0.3...]), which is a strong signal that this paper is likely related to neural networks. Furthermore, by examining other dimensions of the feature vector, I don't see obvious high values for vocabulary related to other categories (such as genetic algorithms or case-based methods).

**Expert B** (Neighborhood Structure Analysis) Step 1: Looking at the neighborhood structure of node  $v_i$ , it is directly connected to nodes  $v_j$ ,  $v_k$ , and  $v_m$ . The structural role  $R_i$  with hash value 0xB4E1 (after 3 WL iterations) indicates that this node exists in a dense connection pattern. In citation networks, this typically means the paper forms a tight research community with its cited or citing papers. The key subgraph shows an average degree of 3.0, further confirming this. Neural network research often forms such dense citation clusters.

**Expert C** (Global Position Analysis) Step 1: The global position encoding  $P_i$  of the target node  $v_i$  is [0.15, 0.62], derived from Laplacian eigenvectors. This indicates the node has a strong presence in the central area of the graph rather than at the periphery. Its node influence  $C_i$  is 0.78 (computed via decay-jump propagation), indicating moderate to high influence. This pattern is quite common in neural network research papers, especially in the Cora citation network, where neural network papers tend to form larger connected communities and have higher citation influence.

**Expert A** (Local Feature Analysis) Step 2: Analyzing deeper patterns in the feature vector, I notice that besides terms related to "neural" and "network," there may also be high values for terms related to deep learning, activation functions, or weights. This combination of vocabulary is very typical in neural network literature. The lack of high-frequency vocabulary related to rule learning, case reasoning, or probabilistic methods further supports the neural network classification hypothesis. The pattern of the feature vector highly aligns with the characteristics of known neural network papers in the Cora dataset.

**Expert B** (Neighborhood Structure Analysis) Step 2: Further examining the topological environment of node  $v_i$ , the key subgraph structure of 4 nodes and 6 edges indicates a highly interconnected small community. In citation networks, papers on similar topics tend to cite each other, forming such tight structures. Considering the structural role represented by hash value 0xB4E1, I infer that  $v_i$  and its neighbors likely belong to the same research sub-domain. By comparing with typical connection patterns of different categories of papers in the Cora dataset, this structure best matches the characteristics of Neural\_Networks category papers, which typically form denser citation clusters.

**Expert C** (Global Position Analysis) Step 3: Combining global position and influence analysis, the characteristics of node  $v_i$  highly match the Neural\_Networks category. In the Cora dataset, neural network research papers often occupy central positions (consistent with  $P_i=[0.15, 0.62]$ ) and have higher influence scores ( $C_i=0.78$ ). Other categories such as Theory typically appear at peripheral positions in the graph, while Rule\_Learning papers tend to have lower connectivity. Based on global position and influence analysis, I predict this node belongs to the Neural\_Networks category, with a confidence of 0.85.

**Expert A** (Local Feature Analysis) Step 3: Based on vocabulary feature analysis, and referencing known category feature patterns in the Cora dataset, I can definitively exclude Case\_Based, Genetic\_Algorithms, Rule\_Learning, and Theory categories, as the feature vector lacks characteristic words for these domains. Among the remaining categories, Neural\_Networks has the highest match, followed by Probabilistic\_Methods, with Reinforcement\_Learning being the lowest. My final prediction is the Neural\_Networks category, with a confidence of 0.92.

**Expert B** (Neighborhood Structure Analysis) Step 3: Through topological structure analysis, combined with the structural role represented by the WL hash value, I confirm that the structure formed by node  $v_i$  and its neighbors in the Cora dataset most closely aligns with the Neural\_Networks category. This structure does not match the typical connection patterns of Theory or Probabilistic\_Methods papers, which usually have different citation behaviors. Based on the node's connection structure and neighbor features, I predict this node belongs to the Neural\_Networks category, with a confidence of 0.88.

#### Integrated Analysis:

All three experts analyzed the target node  $v_i$  from different perspectives, all pointing to the Neural\_Networks category:

- Expert A, through local vocabulary feature analysis, found that high-frequency vocabulary is highly related to neural network research, with a confidence of 0.92
- Expert B, through neighbor topological structure analysis, found that the dense connection pattern matches the characteristics of neural network papers, with a confidence of 0.88
- Expert C, through global position and influence analysis, confirmed that the node's centrality and influence are consistent with neural network papers, with a confidence of 0.85

All experts unanimously support the Neural\_Networks classification, with no expert raising objections or leaving the discussion, indicating consistency across all reasoning paths. The analyses from three different perspectives mutually validate each other, enhancing the reliability of the prediction.

Final Prediction: Neural\_Networks

Overall Confidence: 0.89

Confidence Calculation Method: Weighted average of the three experts' confidence scores, with weights based on the comprehensiveness and evidence strength of each expert's analysis ( $0.92*0.35 + 0.88*0.35 + 0.85*0.30 = 0.89$ )

Figure 14. Expert Collaboration ToT's GPT-4 answer with Cora dataset.

### Expert Collaboration ToT (ANSWER)

**Expert A** (Local Feature Analysis) Step 1: I'll first analyze the local feature vector  $S_i$  of the target node  $v_i$ . This is a 1,433-dimensional vector representing word occurrences in the paper. I notice that dimensions related to "neural" and "network" have relatively high values (such as [...]..), which may indicate that this paper is related to neural networks. However, I also observe that dimensions related to "probability", "bayesian", and "likelihood" also have certain values, which might suggest elements of probabilistic methods.

**Expert B** (Neighborhood Structure Analysis) Step 1: Looking at the neighborhood structure of node  $v_i$ , it is directly connected to nodes  $v_j$ ,  $v_k$ , and  $v_m$ . The structural role  $R_i$  with hash value 0xB4E1 (after 3 WL iterations) indicates that this node exists in a dense connection pattern. The key subgraph shows an average degree of 3.0, indicating this is a relatively tight citation cluster. Such structures may appear in multiple research domains, requiring further analysis to determine the specific category.

**Expert C** (Global Position Analysis) Step 1: The global position encoding  $P_i$  of the target node  $v_i$  is [0.15, 0.62], indicating the node has a strong presence in the central area of the graph. Its node influence  $C_i$  is 0.78, indicating moderate to high influence. This pattern may appear in core papers across multiple research domains, including neural networks, probabilistic methods, or reinforcement learning and other popular fields. More information is needed to narrow down the possible categories.

**Expert A** (Local Feature Analysis) Step 2: Analyzing the feature vector more deeply, I find that despite terms related to neural networks, the frequency of terms like "bayesian", "distribution", and "likelihood" related to probabilistic methods is higher. I also notice significant occurrences of reinforcement learning terms such as "reinforcement", "reward", and "policy". This makes me doubt my preliminary judgment about neural networks. Now it appears that the paper might be related to probabilistic methods or reinforcement learning, or a combination of both.

**Expert B** (Neighborhood Structure Analysis) Step 2: Further examining the features of node  $v_i$ 's neighbors, I notice that the feature vectors of neighbor nodes  $v_j$  and  $v_k$  contain many terms related to probabilistic methods, while  $v_m$  contains more reinforcement learning terms. This mixed pattern suggests that the paper may be in the intersection of probabilistic methods and reinforcement learning. Considering typical citation patterns in the Cora dataset, this structure most likely corresponds to reinforcement learning papers, especially those that use probabilistic methods.

**Expert C** (Global Position Analysis) Step 2: Analyzing the node's position in the entire graph, I find it is positioned at the connection between two communities—one mainly composed of probabilistic methods papers and the other mainly of reinforcement learning papers. This "bridging" position is typically occupied by interdisciplinary research papers. The global influence score of 0.78 also supports this view, indicating that the paper has influence across multiple research communities.

**Expert A** (Local Feature Analysis) Step 3: [PRUNING/WITHDRAWAL] *"I notice that my analysis path may contain contradictions. Initially, I leaned towards neural network classification, then considered probabilistic methods or reinforcement learning. After re-examining the feature vector, I find that the keyword pattern shows terms like "reinforcement", "reward", "policy", "learning", and "state" are significantly higher than terms related to other categories. My previous analysis was not systematic enough and may have misled the reasoning. Due to my inability to provide consistent high-confidence analysis, I choose to withdraw from the discussion."*

**Expert B** (Neighborhood Structure Analysis) Step 3: Based on comprehensive analysis of topological structure and neighbor features, as well as the structural pattern represented by the WL hash value, I confirm that this node most likely belongs to the Reinforcement\_Learning category. This conclusion is based on: 1) the significant presence of reinforcement learning terms in neighbor node features; 2) its structural position connecting probabilistic methods and machine learning communities; 3) similarity to the topological structure of known reinforcement learning papers in the Cora dataset. My final prediction is the Reinforcement\_Learning category, with a confidence of 0.83.

**Expert C** (Global Position Analysis) Step 3: Based on the node's global position, influence analysis, and bridging characteristics, combined with the typical position distribution of different research domains in the Cora dataset, I believe this node most likely belongs to the Reinforcement\_Learning category. Reinforcement learning papers typically have this characteristic of connecting different research communities, especially connecting probabilistic methods and machine learning communities. My final prediction is the Reinforcement\_Learning category, with a confidence of 0.79.

#### Integrated Analysis:

In this analysis, Expert A realized in step 3 that their analysis contained contradictions and could not provide consistent high-confidence judgment, therefore chose to withdraw from the discussion. Experts B and C consistently support the Reinforcement\_Learning category from different perspectives:

- Expert B, through neighbor structure and topological feature analysis, arrived at a confidence of 0.83
- Expert C, through global position and bridging characteristic analysis, arrived at a confidence of 0.79

The two remaining experts unanimously believe that this node belongs to the Reinforcement\_Learning category, with clear and consistent reasoning logic and no obvious contradictions.

Final Prediction: Reinforcement\_Learning

Overall Confidence: 0.81

Confidence Calculation Method: Weighted average of the two remaining experts' confidence scores, with weights based on the comprehensiveness and evidence strength of each analysis ( $0.83*0.55 + 0.79*0.45 = 0.81$ ). Expert A's analysis was pruned and not included in the final decision.

Figure 15. Expert Collaboration ToT's GPT-4 answer with Cora dataset (pruning).