**Reviewer #1**

**Questions**

* **1. Overall Evaluation**
  + Reject
* **3. Novelty**
  + Novel solution
* **4. Importance: select all that apply:**
  + The paper contains controversial ideas and/or will generate interesting discussion
* **5. Summary of contribution (in a few sentences)**
  + The paper proposes a novel framework for graph database verification, combining semantic features and structural elements. It does so by using activated prompts for adaptive amplification perturbations.
* **6. Describe in detail all strong points, labeled S1, S2, S3, etc.**
  + S1: The illustrations throughout the paper assist in understanding the components of the framework.  
    S2: The logic behind the loss functions is well-explained.
* **7. Describe in detail all opportunities for improvement, labeled O1, O2, O3, etc.**
  + O1: A real-world running example or use-case would be incredibly helpful in demonstrating the motivation for the problem, deciding whether it is relevant, and understanding the different stages of the framework. Currently, the discussions are rather abstract, making it difficult to follow, especially for non-experts in this specific subfield.  
    O2: The paper lacks formality and theory. The threat model and problem are not formally defined, and there are no guarantees for the algorithms. All these significantly weaken the paper.  
    O3: There is no related work survey, making the paper difficult to position with respect to work already done in this field and determine its novelty.  
    O4: Section 4 presents the threat model for the problem. Is this a novel model that has not been studied (I assume so since there are no citations)? The authors should clarify this and, again, attach a real-world concrete example that demonstrate this is a realistic model.  
    O5: The experimental study should include a dataset with over 1MM edges.  
    O6: It seems a bit suspicious that the new framework performs better than existing SOTA systems across the board in all examined aspects. The authors should include a convincing explanation for why that is and see if there are any aspects that are not improved by their approach.
* **8. Please rank the three most critical strong points and/or opportunities for improvement that led to your overall evaluation rating, e.g., "{S1, O7, O4}"**
  + O3, O2, O1
* **9. Do you have specific questions to authors to answer in author feedback? Use Q1, Q2, Q3, etc. for questions.**
  + No.
* **10. Minor remarks: Use this field to describe minor remarks that are not listed as opportunities for improvement, e.g., to highlight typos, formatting problems, or minor technical issues, labeled M1, M2, M3, etc.**
  + Section 7: RQ4: Whether the our scheme -> Whether our scheme
* **11. Required changes for revision: List required changes for a revision, if appropriate, by identifying subset of previously specified opportunities for improvement (e.g., O1, O3, O6). Authors can respond to this in their proposed initial revision plan along with the author feedback.**
  + O1-O6
* **12. Inclusive writing: does the paper follow the SIGMOD inclusive writing guidelines? See https://dbdni.github.io/pages/inclusivewriting.html for more information. If not, please provide suggestions on how to improve the writing of the paper.**
  + I agree
* **14. Availability of artifacts: All papers were asked to make their code, data, scripts, and notebooks available anonymously if possible. You can find this information in the paper summary on CMT when clicking on the Paper ID next to the title. The call for papers emphasizes the need for artifacts especially for E&A Papers and DI&DS Papers. Are the artifacts available? Do you have any comments or feedback on the artifacts?**
  + The source code is available in the provided link.

**Reviewer #2**

**Questions**

* **1. Overall Evaluation**
  + Reject
* **3. Novelty**
  + Novel solution
* **4. Importance: select all that apply:**
  + SIGMOD attendees will learn something interesting from the paper
* **5. Summary of contribution (in a few sentences)**
  + This paper introduces PA^2Perturb, a novel method for verifying ownership of graph datasets used in training Graph Neural Networks (GNNs). The core idea is to inject a "damage-free" and imperceptible backdoor into the graph data before it is released. If a third party trains a GNN on this protected data, the model inherits this backdoor. The original data owner can then verify their ownership by querying the suspect GNN with a specific activated prompt. Experiments on standard benchmarks and GNNs show high verification success, robustness to pruning/anomaly defenses, and negligible impact on clean accuracy, demonstrating transferability and practicality.
* **6. Describe in detail all strong points, labeled S1, S2, S3, etc.**
  + S1. The proposed methodology is novel. In particular, the proposed interactive verification mechanism, where the data owner can trigger a specific response from a suspect model, can be both more secure and less susceptible to accidental discovery.  
      
    S2. The presented solutions are intuitive and convincing. Specifically, the proposed bi-level optimization framework, which simultaneously works to maximize the verifiability of the watermark while minimizing the impact on the GNN's predictive performance, is a sophisticated design under a clear threat model.  
      
    S3. The experiments involve multiple real datasets, prior solutions such as watermarking and backdoor attack, as well as an ablation study.
* **7. Describe in detail all opportunities for improvement, labeled O1, O2, O3, etc.**
  + O1. The paper primarily considers a passive adversary who simply trains a model on the watermarked data. A more sophisticated adversary might employ techniques to proactively attack the proposed scheme, such as model pruning or fine-tuning, which could potentially destroy the watermark. Moreover, the adversary could train a model with a mixture of clean and watermarked data. The robustness of PA^2Perturb against these more active defense mechanisms remains to be studied.  
      
    O2. Some design choices made in the paper are not formally justified, e.g., the selection of "hard nodes" based on high prediction entropy, whose effectiveness might be data-dependent. Further, once the adversary is aware of this design, they may actively prune such nodes in an effort to defeat the proposed scheme. Similarly, the selection of hyperparameter \tau (in CASR construction) is not well justified, either.  
      
    O3. The experimental results mainly report VSR. Could the prompt accidentally trigger on a non-watermarked model? Adding other metrics would make the results more convincing (notably, false positive rate, f1, etc.)
* **8. Please rank the three most critical strong points and/or opportunities for improvement that led to your overall evaluation rating, e.g., "{S1, O7, O4}"**
  + {O1, O2, S1}
* **9. Do you have specific questions to authors to answer in author feedback? Use Q1, Q2, Q3, etc. for questions.**
  + Q1. Figure 2 is rather difficult to read. Could you explain more on the distinction between the different types of nodes (e.g., "Hard Node," "P-nodes," "T-nodes") and the flow of the process?  
      
    Q2. The proposed solution needs to solve a bi-level optimization problem, which is probably rather difficult. Is the solution guaranteed to converge? Are there any bounds of the result quality?
* **10. Minor remarks: Use this field to describe minor remarks that are not listed as opportunities for improvement, e.g., to highlight typos, formatting problems, or minor technical issues, labeled M1, M2, M3, etc.**
  + M1. The "damage-free" is probably an overstatement, since the proposed scheme perturbed the data, which introduces damage. "Utility-preserving" might be a more accurate description.  
      
    M2. In Eq. 21, it's better to add a reference for the MMD concept.
* **11. Required changes for revision: List required changes for a revision, if appropriate, by identifying subset of previously specified opportunities for improvement (e.g., O1, O3, O6). Authors can respond to this in their proposed initial revision plan along with the author feedback.**
  + As discussed in O1 and O2, the main issue with the paper is that the adversary model is rather weak. To strengthen the paper, I suggest the authors analyze the proposed solution under an informed attacker who is aware of the design of the proposed scheme. Specifically, the authors could test white-box attackers who apply targeted defenses, e.g., adversarial training with graph data augmentation, differentially private training such as DP-SGD, knowledge distillation, calibration/temperature scaling, post-training quantization, and training on a mixture of clean and watermarked data.  
      
    However, this is a huge task, and I'm not optimistic that it can be done within the revision timeframe.
* **12. Inclusive writing: does the paper follow the SIGMOD inclusive writing guidelines? See https://dbdni.github.io/pages/inclusivewriting.html for more information. If not, please provide suggestions on how to improve the writing of the paper.**
  + I agree
* **14. Availability of artifacts: All papers were asked to make their code, data, scripts, and notebooks available anonymously if possible. You can find this information in the paper summary on CMT when clicking on the Paper ID next to the title. The call for papers emphasizes the need for artifacts especially for E&A Papers and DI&DS Papers. Are the artifacts available? Do you have any comments or feedback on the artifacts?**
  + https://anonymous.4open.science/r/PA2Perturb-4F14

**Reviewer #3**

**Questions**

* **1. Overall Evaluation**
  + Reject
* **3. Novelty**
  + No novelty
* **4. Importance: select all that apply:**
  + SIGMOD attendees will learn something interesting from the paper
* **5. Summary of contribution (in a few sentences)**
  + The paper aims to solve the challenge of reliably verifying ownership of graph datasets used for training GNNs, achieving stealthiness and transferability, because non-intrusive DOV is unstable while intrusive triggers degrade model performance. PA2Perturb constructs a prompt-sensitive imperceptible subgraph, injects it into context-aware stable regions, and employs prompts to activate and adaptively amplify the signal. Experiments on Cora, Citeseer, Pubmed, and Flickr report verification success rates >95%, preserved task accuracy, and robustness to pruning/OD/RIGBD.
* **6. Describe in detail all strong points, labeled S1, S2, S3, etc.**
  + S1. The paper is oriented toward a practical problem of detecting unauthorized graph datasets used in GNN model training.  
      
    S2. The paper tackles the problem with PA2Perturb, a method that combines semantic- and distribution-constrained subgraph injection with prompt-based activation to achieve stealthy injection and effective detection while preserving GNN performance.  
      
    S3. Experiments on multiple benchmark datasets (Cora, Citeseer, Pubmed, Flickr) under varied attacks demonstrate the method’s verification success and robustness over the baselines.
* **7. Describe in detail all opportunities for improvement, labeled O1, O2, O3, etc.**
  + O1. For GNN models trained on unauthorized graph datasets, the dataset owner may not have access to the models, which raises a feasibility concern for the problem addressed in the paper. It would be beneficial to discuss this as a limitation or to propose special methods for this setting.  
      
    O2. The current watermarking method is highly task-specific, meaning detection is only effective when the released watermarked graph is used for training models on the same type of task as the one used during watermark generation (e.g., node classification). If the watermarked graph is repurposed for a fundamentally different task (e.g., graph clustering), the watermark trigger effectiveness can degrade or even fail. It is beneficial to explore watermark designs that are more task-agnostic or adaptable across different graph learning objectives, thereby improving robustness in varied application scenarios.  
      
    O3. While the paper claims strong stealthiness for CIS, the evaluation does not quantitatively assess resilience against structural anomaly detection methods (e.g., subgraph rarity or motif-based analysis). The current experiments focus mainly on feature-space similarity and perturbation removal, which may overestimate stealthiness in scenarios with rigorous data auditing. It would be beneficial to include structural anomaly detection methods to verify the stealthiness of CIS.  
      
    O4. If the surrogate model is leaked, an attacker could identify vulnerable nodes and remove the injected subgraph. This scenario is not discussed, and addressing it would strengthen the analysis of CIS’s robustness.  
      
    O5. The method depends on a surrogate model to identify “hard nodes.” If the surrogate model is flawed or poorly trained, the whole perturbation process could be misled. The paper evaluates architecture transfer (Table 6) but doesn’t analyze sensitivity to a weakly trained or mis-specified surrogate. It could be beneficial to include discussion on this dependency of the surrogate model.  
      
    O6. Hyperparameter sensitivity and specification:  
      
    O6.1. The method involves multiple hyperparameters (e.g., top-k hard nodes, K-hop subgraph size), but no sensitivity analysis is provided to clarify their influence on performance or guide parameter selection.  
      
    O6.2. As an important component of the approach, the initialization strategy and dimension of the prompt matrix are unspecified and untested. Given the dual-prompt matrix’s role in bi-level optimization, this could cause large performance differences under different settings.  
      
    O6.3. The context-aware stable region selection relies on manually set thresholds (e.g., semantic similarity τ\_s, degree difference τ\_d, top-k node selection), which could strongly affect both stealthiness and verification success. The paper does not justify these values or test robustness against variations.  
      
    O7. The proposed method relies on bi-level alternating optimization between the CIS generator and the dual prompt generator, which introduces concerns regarding convergence, stability, and efficiency. Bi-level optimization is generally NP-hard and often requires heuristic approximations [1], raising the risk of suboptimal local minima. However, the paper does not provide analysis of convergence behavior, stability across different settings, or computational overhead compared to single-stage baselines. Given that bi-level optimization is typically time-consuming and sensitive to hyperparameter choices, a discussion of these theoretical limitations and empirical evidence on stability would strengthen the work’s technical soundness and clarify its practicality for large-scale or time-constrained scenarios.  
      
    O8. The framework's transferability in realistic scenarios with significant model heterogeneity. Nodes identified as "hard" by a simpler surrogate model might be "easy" for a much more powerful suspect model, potentially reducing verification effectiveness. The paper's claims would be strengthened by evaluating the method's robustness when a significant capability gap exists between the surrogate and suspect models.  
      
    O9. There is no direct validation that CIS-modified nodes remain indistinguishable in downstream tasks without the trigger prompt. It could be helpful to include an experiment evaluating their performance in the absence of the watermark prompt to confirm that the watermark is only inducible via prompting, and does not affect normal task behavior.  
      
    O10. Despite the abstract’s critique, experiments compare almost exclusively with backdoor-type and one prompt baseline; readers can’t judge whether intrusive prompting actually outperforms modern non-intrusive verification. It would strengthen the work to include direct comparisons against state-of-the-art non-intrusive verification methods.  
      
    O11. Typos:  
      
    O11.1. In Table 3, the “prue” should be “prune”.  
      
    O11.2. “the adaptive attacker” in section 4 should be “The adaptive attacker”.  
      
    O11.3. In section 7, “the our scheme” in RQ4.  
      
      
    [1]Yihua Zhang, Prashant Khanduri, Ioannis Tsaknakis, Yuguang Yao, Mingyi Hong, and Sijia Liu. 2024. An Introduction to Bilevel Optimization: Foundations and applications in signal processing and machine learning. IEEE Signal Processing Magazine 41, 1 (January 2024), 38–59.
* **8. Please rank the three most critical strong points and/or opportunities for improvement that led to your overall evaluation rating, e.g., "{S1, O7, O4}"**
  + O1-3
* **9. Do you have specific questions to authors to answer in author feedback? Use Q1, Q2, Q3, etc. for questions.**
  + No.
* **11. Required changes for revision: List required changes for a revision, if appropriate, by identifying subset of previously specified opportunities for improvement (e.g., O1, O3, O6). Authors can respond to this in their proposed initial revision plan along with the author feedback.**
  + All
* **12. Inclusive writing: does the paper follow the SIGMOD inclusive writing guidelines? See https://dbdni.github.io/pages/inclusivewriting.html for more information. If not, please provide suggestions on how to improve the writing of the paper.**
  + I agree
* **14. Availability of artifacts: All papers were asked to make their code, data, scripts, and notebooks available anonymously if possible. You can find this information in the paper summary on CMT when clicking on the Paper ID next to the title. The call for papers emphasizes the need for artifacts especially for E&A Papers and DI&DS Papers. Are the artifacts available? Do you have any comments or feedback on the artifacts?**
  + Didn't check.