1 Introduction

1.1 Resumen:

The paper "Logical-Mathematical Foundations of a Graph Query Framework for Relational Learning" introduces a new graph query framework that addresses two fundamental problems in relational learning: computational complexity and lack of robust and general frameworks. The proposed framework allows for efficient pattern matching, atomic operations to expand queries in a partitioned manner, and the evaluation of cyclic patterns in polynomial time. The paper provides an overview of related research, introduces the novel graph query framework, discusses its properties and implementation, and concludes with potential avenues for future research. The main points of the introduction section are:

1. Relational learning methods can learn from connections between data, making them powerful in different domains.
2. There are two basic approaches to relational learning: the latent feature or connectionist approach and the graph pattern-based approach or symbolic approach.
3. The connectionist approach has proven its effectiveness in many tasks, while the pattern-based approach has been less successful due to computational complexity and lack of robust frameworks.
4. The proposed graph query framework aims to solve these problems by providing a query system that allows graph pattern matching with controlled complexity and stepwise pattern expansion using well-defined operations.
5. The paper focuses on formalising an efficient graph query system and defining a set of operations to refine queries, but does not conduct an extensive analysis of performance or efficiency in comparison to other methods.

1.2 Evaluación:

Motivation:

* YES: The section clearly explains the study's significance and relevance, providing specific examples from the text.

The introduction justifies the need for a graph query framework that can efficiently learn from relational data by highlighting the limitations of existing approaches due to computational complexity and lack of robustness. It also emphasizes the potential benefits of the proposed approach in various domains such as social networks, protein characterization, and toxic effects prediction.

Improvement:

* Can be improved: The section could provide more specific examples from real-world applications or case studies to further demonstrate the relevance and importance of the study's findings. Additionally, the authors could use data or references to highlight the problem's wider impacts and potential benefits of their proposed approach.

Novelty:

* YES: The section clearly describes the proposed approach's novelty and originality, emphasizing its unique contributions compared to existing work.

The introduction explicitly compares the proposed approach with related work, highlighting its distinct features such as atomic operations, substructure assessment, and cyclic pattern evaluation in polynomial time.

Clarity:

* Can be improved: The section could benefit from restructuring complex sentences and using more concise language to improve clarity.

Grammar and Style:

* YES: The section is well-written and free of grammatical and stylistic errors.

Typos and Errors:

* NO: There are no typos or other errors in the section.

Overall, the introduction effectively motivates the need for a graph query framework and highlights the novelty and potential benefits of the proposed approach. However, it could benefit from providing more specific examples and using clearer language to improve clarity and comprehension.

2 Related work

2.1 Resumen:

The related work section of the paper discusses various approaches to relational learning, including graph pattern matching and inductive logic programming. The author highlights the limitations of these approaches and their inability to handle certain types of queries or extract cyclic patterns from data. The proposed method aims to address these limitations by providing a more comprehensive and flexible approach to relational learning.

2.2 Evaluación:

Motivation:

* Clarity: The section provides a clear explanation of the relevance and significance of the proposed approach, highlighting its potential to address the limitations of existing methods.

Novelty:

* Originality: The section effectively describes the novel aspects of the proposed approach, including its ability to execute cyclic queries and extract cyclic patterns from data during the learning process.

Clarity:

* Comprehension: The section is well-written and easy to understand, with appropriate terminology and clear examples.

Grammar and Style:

* Correctness: The section is generally free of grammatical and stylistic errors, although there are a few minor issues (e.g., missing articles, inconsistent capitalization).

Typos and Errors:

* Accuracy: There are no typos or other errors in the section.

Overall Evaluation Level: Must be Improved

Recommendations:

1. Provide more specific examples to further illustrate the limitations of existing methods and the benefits of the proposed approach.
2. Emphasize the novelty and originality of the proposed approach, perhaps by comparing it directly with related work.
3. Address any grammatical or stylistic issues to ensure clarity and consistency throughout the

section.
4. Consider adding more concrete examples or illustrations to enhance comprehension and engagement.

3 Relational machine learning

3.1 Resumen:

In this section, the authors present a method for developing relational machine learning models on graph data sets. They describe how to use the framework presented earlier to acquire relational classifiers on graph datasets. The authors introduce information-gain pattern mining and explain how it can be used to obtain characteristic patterns of subgraph classes using the previous graph query framework. They also provide examples of practical instances of performing relational learning by using the query framework and refinement sets, demonstrating the process of developing a decision tree for node classification in a social network and character specie classification in a Star Wars toy graph.

Please summarize this section of the paper into 150-200 words while maintaining the focus on relational machine learning and its application in acquiring characteristic patterns of subgraph classes using the previous graph query framework.

Please note that I will provide you with more sections of the paper for summary in the future.

3.2 Evaluación:

Evaluation Criteria:

* Motivation:
+ YES: The section clearly explains the study's significance and relevance.
+ Clarity: The problem's importance and wider impacts are justified with specific examples from the text.
+ Improvement: To strengthen the motivation, consider using data or references to highlight the problem's importance.
* Novelty:
+ YES: The section describes the proposed approach's novelty and originality.
+ Originality: The section differentiates the proposed approach from existing work with specific examples.
+ Improvement: To emphasize the novelty, explicitly compare with related work and highlight unique contributions.
* Clarity:
+ YES: The section is well-written and easy to understand.
+ Comprehension: The text uses appropriate terminology and avoids ambiguity.
+ Improvement: To improve clarity, restructure complex sentences, define technical terms, and use illustrative examples.
* Grammar and Style:
+ YES: The section is free of grammatical and stylistic errors.
+ Correctness: The text uses language appropriate for an academic setting.
+ Improvement: To improve style, consider using more concise and precise language.
* Typos and Errors:
+ NO: The section is free of typos and other errors.

Evaluation justification and examples from the evaluated section:

Motivation:

* The section provides a clear explanation of the study's significance and relevance, highlighting the importance of relational machine learning in graph data analysis.
* Examples from the text include the need for efficient pattern search techniques to handle large-scale graph data and the potential applications of relational learning in social network analysis.

Novelty:

* The section effectively describes the proposed approach's novelty and originality, differentiating it from existing work.
* Examples from the text include the use of information gain for pattern mining and the development of a top-down decision tree induction method.

Clarity:

* The section is well-written and easy to understand, with appropriate terminology and avoidance of ambiguity.
* Examples from the text include the definition of subgraphs, refinement sets, and decision trees.

Grammar and Style:

* The section is free of grammatical and stylistic errors, using language appropriate for an academic setting.
* Examples from the text include the use of active voice and concise sentence structure.

Typos and Errors:

* No typos or other errors were found in the section.

4 Conclusions and future work

4.1 Resumen:

The paper's "Conclusions and future work" section summarizes the main contributions and findings of the research, as well as outlining potential future directions for development. The key points are:

* The proposed graph query framework enables polynomial-time evaluation of cyclic queries and refinements, supporting relational learning processes.
* The system demonstrates effectiveness in extracting interesting patterns from relational data.
* The current implementation is limited to binary graph data sets, but the approach can be applied to hypergraphs as well.
* The provided refinement operations are basic and reliable, but more complex refinements could lead to faster learning algorithms.
* The system has potential applications in explainable learning and automatic feature extraction.
* Future research may focus on developing automated methods for generating refinement sets based on the specific characteristics of the graph dataset and the learning task at hand.

In summary, the paper presents a novel graph query framework that supports relational learning processes by enabling the efficient evaluation of cyclic queries and refinements. The system demonstrates effectiveness in extracting interesting patterns from relational data, and has potential applications in explainable learning and automatic feature extraction. Future research may focus on developing more complex refinements and exploring other machine learning algorithms that can be used in conjunction with the query framework.

4.2 Evaluación:

Evaluation Criteria: Motivation, Novelty, Clarity, Grammar and Style, Typos and Errors.

Evaluation Level:

Motivation: YES - The section clearly explains the study's significance and relevance, providing specific examples from the text.

Novelty: YES - The section describes the proposed approach's novelty and originality, differentiating it from existing work.

Clarity: YES - The section is well-written and easy to understand, using appropriate terminology and avoiding ambiguity.

Grammar and Style: YES - The section is free of grammatical and stylistic errors, using language appropriate for an academic setting.

Typos and Errors: NO - There are no typos or other errors in the section.

Justification and Examples from the Evaluated Section:

Motivation:

The section provides a clear explanation of the study's significance and relevance, highlighting the problem's importance and its wider impacts. For instance, it notes that "the system utilises a consistent grammar for both queries and evaluated structures" and "allows the assessment of subgraphs beyond individual nodes."

Novelty:

The section effectively describes the proposed approach's novelty and originality, differentiating it from existing work. For example, it states that "the query graph framework offered here assesses the existence/non-existence of paths and nodes in a graph rather than demanding isomorphisms."

Clarity:

The section is well-written and easy to understand, using appropriate terminology and avoiding ambiguity. For instance, it defines "a path, which connects pairs of nodes" and "a refinement family can be established based on the frequency of structural occurrences in a graph."

Grammar and Style:

The section is free of grammatical and stylistic errors, using language appropriate for an academic setting. For instance, it uses phrases such as "the fact that" and "it is possible to."

Typos and Errors:

There are no typos or other errors in the section.

Acknowledgements

Proyecto PID2019-109152G financiado por MCIN/AEI/10.13039/501100011033

DISARM project - Grant n. PDC2021-121197, and the HORUS project - Grant n. PID2021-126359OB-I00 funded by MCIN/AEI/310.13039/501100011033 and by the “European Union NextGenerationEU/PRTR”

References

References

Almagro-Blanco, P., & Sancho-Caparrini, F. (2017). Generalized graph pattern matching. *CoRR*, *abs/1708.03734*. URL: http://arxiv.org/abs/1708.03734. arXiv:1708.03734.

Barceló, P., Libkin, L., & Reutter, J. L. (2011). Querying graph patterns. In *Proceedings of the Thirtieth ACM SIGMOD-SIGACT-SIGART Symposium on Principles of Database Systems* PODS ’11 (pp. 199–210). New York, NY, USA: ACM. URL: http://doi.acm.org/10.1145/1989284.1989307. doi:10.1145/1989284.1989307.

Blockeel, H., & Raedt, L. D. (1998). Top-down induction of first-order logical decision trees. *Artificial Intelligence*, *101*, 285–297. URL: http://www.sciencedirect.com/science/article/pii/S0004370298000344. doi:10.1016/S0004-3702(98)00034-4.

Bonifati, A. Fletcher, G.Voigt, H.Yakovets, N.Jagadish, H. V. (2018 Querying Graphs Morgan & Claypool Publishers

Bordes, A., Usunier, N., Garcia-Duran, A., Weston, J., & Yakhnenko, O. (2013). Translating embeddings for modeling multi-relational data. In C. J. C. Burges, L. Bottou, M. Welling, Z. Ghahramani, & K. Q. Weinberger (Eds.), *Advances in Neural Information Processing Systems 26* (pp. 2787–2795). Curran Associates, Inc. URL: http://papers.nips.cc/paper/5071-translating-embeddings-for-modeling-multi-relational-data.pdf.

Brynjolfsson, E., & Mitchell, T. (2017). What can machine learning do? workforce implications. *Science*, *358*, 1530–1534.

Camacho, R., Pereira, M., Costa, V. S., Fonseca, N. A., Adriano, C., Simões, C. J., & Brito, R. M. (2011). A relational learning approach to structure-activity relationships in drug design toxicity studies. *Journal of integrative bioinformatics*, *8*, 176–194.

Chang, K.-W., Yih, S. W.-t., Yang, B., & Meek, C. (2014). Typed tensor decomposition of knowledge bases for relation extraction. In *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing*. ACL – Association for Computational Linguistics. URL: https://www.microsoft.com/en-us/research/publication/typed-tensor-decomposition-of-knowledge-bases-for-relation-extraction/.

Consens, M. P., & Mendelzon, A. O. (1990). Graphlog: A visual formalism for real life recursion. In *Proceedings of the Ninth ACM SIGACT-SIGMOD-SIGART Symposium on Principles of Database Systems* PODS ’90 (pp. 404–416). New York, NY, USA: ACM. URL: http://doi.acm.org/10.1145/298514.298591. doi:10.1145/298514.298591.

Cook, S. A. (1971). The complexity of theorem-proving procedures. In *Proceedings of the Third Annual ACM Symposium on Theory of Computing* STOC '71 (pp. 151–158). New York, NY, USA: ACM. URL: http://doi.acm.org/10.1145/800157.805047. doi:10.1145/800157.805047.

Dong, X., Gabrilovich, E., Heitz, G., Horn, W., Lao, N., Murphy, K., Strohmann, T., Sun, S., & Zhang, W. (2014). Knowledge vault: A web-scale approach to probabilistic knowledge fusion. In *Proceedings of the 20th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* KDD '14 (pp. 601–610). New York, NY, USA: ACM. URL: http://doi.acm.org/10.1145/2623330.2623623. doi:10.1145/2623330.2623623.

Fan, W., Li, J., Ma, S., Tang, N., Wu, Y., & Wu, Y. (2010). Graph pattern matching: From intractable to polynomial time. *Proc. VLDB Endow.*, *3*, 264–275. URL: http://dx.doi.org/10.14778/1920841.1920878. doi:10.14778/1920841.1920878.

Gallagher, B. (2006). Matching structure and semantics: A survey on graph-based pattern matching. *AAAI FS*, *6*, 45–53.

García-Jiménez, B., Pons, T., Sanchis, A., & Valencia, A. (2014). Predicting protein relationships to human pathways through a relational learning approach based on simple sequence features. *IEEE/ACM Transactions on Computational Biology and Bioinformatics*, *11*, 753–765. doi:10.1109/TCBB.2014.2318730.

Geamsakul, W., Matsuda, T., Yoshida, T., Motoda, H., & Washio, T. (2003). Classifier construction by graph-based induction for graph-structured data. In K.-Y. Whang, J. Jeon, K. Shim, & J. Srivastava (Eds.), *Advances in Knowledge Discovery and Data Mining: 7th Pacific-Asia Conference, PAKDD 2003, Seoul, Korea, April 30 – May 2, 2003 Proceedings* (pp. 52–62). Berlin, Heidelberg: Springer Berlin Heidelberg. URL: http://dx.doi.org/10.1007/3-540-36175-8_6. doi:10.1007/3-540-36175-8_6.

Gupta, S. (2015). *Neo4j Essentials*. Community experience distilled. Packt Publishing. URL: https://books.google.es/books?id=WJ7NBgAAQBAJ.

Luc De Raedt, Sebastijan Dumančić, Robin Manhaeve, Giuseppe Marra. *From Statistical Relational to Neural-Symbolic Artificial Intelligence.* In *Proceedings of the Twenty-Ninth International Joint Conference on Artificial Intelligence (IJCAI'20)*, 2021. ISBN: 9780999241165. Article No.: 688. Pages: 8. Yokohama, Japan.

Henzinger, M. R., Henzinger, T. A., & Kopke, P. W. (1995). Computing simulations on finite and infinite graphs. In *Foundations of Computer Science, 1995. Proceedings., 36th Annual Symposium on* (pp. 453–462). IEEE.

Jacob, Y., Denoyer, L., & Gallinari, P. (2014). Learning latent representations of nodes for classifying in heterogeneous social networks. In *Proceedings of the 7th ACM International Conference on Web Search and Data Mining* WSDM '14 (pp. 373–382). New York, NY, USA: ACM. URL: http://doi.acm.org/10.1145/2556195.2556225. doi:10.1145/2556195.2556225.

Jiang, J. Q. (2011). Learning protein functions from bi-relational graph of proteins and function annotations. In T. M. Przytycka, & M.-F. Sagot (Eds.), *Algorithms in Bioinformatics* (pp. 128–138). Berlin, Heidelberg: Springer Berlin Heidelberg.

Hunter, J. D. (2007). Matplotlib: A 2D Graphics Environment. *Computing in Science & Engineering*, *9*, 3.

Karp, R. M. (1975). On the computational complexity of combinatorial problems. *Networks*, *5*, 45–68.

Knobbe, A. J., Siebes, A., Wallen, D. V. D., & Syllogic B., V. (1999). Multi-relational decision tree induction. In *In Proceedings of PKDD' 99, Prague, Czech Republic, Septembre* (pp. 378–383). Springer.

Latouche, P., & Rossi, F. (2015). Graphs in machine learning: an introduction.

Leiva, H. A., Gadia, S., & Dobbs, D. (2002). Mrdtl: A multi-relational decision tree learning algorithm. In *Proceedings of the 13th International Conference on Inductive Logic Programming (ILP 2003* (pp. 38–56). Springer-Verlag.

Milner, R. (1989). *Communication and Concurrency*. Upper Saddle River, NJ, USA: Prentice-Hall, Inc.

Nguyen, P. C., Ohara, K., Motoda, H., & Washio, T. (2005). Cl-gbi: A novel approach for extracting typical patterns from graph-structured data. In T. B. Ho, D. Cheung, & H. Liu (Eds.), *Advances in Knowledge Discovery and Data Mining: 9th Pacific-Asia Conference, PAKDD 2005, Hanoi, Vietnam, May 18-20, 2005. Proceedings* (pp. 639–649). Berlin, Heidelberg: Springer Berlin Heidelberg. URL: `http://dx.doi.org/10.1007/11430919_74`. doi:`10.1007/11430919_74`.

Wenfei Fan. *Graph Pattern Matching Revised for Social Network Analysis*. In *Proceedings of the 15th International Conference on Database Theory (ICDT '12)*, 2012. ISBN: 9781450307918. Publisher: Association for Computing Machinery. Address: New York, NY, USA. Pages: 8–21. Num. Pages: 14. Location: Berlin, Germany. DOI: 10.1145/2274576.2274578.

Yiwei Wang, Wei Wang, Yuxuan Liang, Yujun Cai, Juncheng Liu, Bryan Hooi. *NodeAug: Semi-Supervised Node Classification with Data Augmentation*. En *Proceedings of the 26th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining (KDD '20)*, 2020, páginas 207–217. DOI: `https://doi.org/10.1145/3394486.3403063`.

Seyed Mehran Kazemi and David Poole. *RelNN: A Deep Neural Model for Relational Learning*. In *Proceedings of the Thirty-Second AAAI Conference on Artificial Intelligence (AAAI-18)*, University of British Columbia, Vancouver, Canada, 2018. Email: `smkazemi@cs.ubc.ca`, `poole@cs.ubc.ca`.

Maria Leonor Pacheco and Dan Goldwasser. *Modeling Content and Context with Deep Relational Learning. Transactions of the Association for Computational Linguistics*, 2021; 9, 100–119. DOI: `https://doi.org/10.1162/tacl_a_00357`.

K. Ahmed, A. Altaf, N.S.M. Jamail, F. Iqbal, R. Latif. *ADAL-NN: Anomaly Detection and Localization Using Deep Relational Learning in Distributed Systems. Applied Sciences*, 2023, 13, 7297. DOI: `https://doi.org/10.3390/app13127297`.

Jie Zhou, Ganqu Cui, Zhengyan Zhang, Cheng Yang, Zhiyuan Liu, Maosong Sun. *Graph Neural Networks: A Review of Methods and Applications. AI open*, 2020, 1, 57-81.

Lingfei Wu, Peng Cui, Jian Pei, Liang Zhao, Xiaojie Guo. *Graph Neural Networks: Foundation, Frontiers and Applications.* En *Proceedings of the 28th ACM SIGKDD Conference on Knowledge Discovery and Data Mining (KDD '22)*, 2022, páginas 4840–4841. DOI: `https://doi.org/10.1145/3534678.3542609`.

Nickel, M., Murphy, K., Tresp, V., & Gabrilovich, E. (2016). A review of relational machine learning for knowledge graphs. *Proceedings of the IEEE*, *104*, 11–33.

Namkyeong L., Dongmin H., Gyoung S. Na, Sungwon K., Junseok L. and Chanyoung P. (2023). Conditional Graph Information Bottleneck for Molecular Relational Learning.

Plotkin, G. (1972). Automatic methods of inductive inference .

van Rest, O., Hong, S., Kim, J., Meng, X., & Chafi, H. (2016). Pgql: a property graph query language. In *Proceedings of the Fourth International Workshop on Graph Data Management Experiences and Systems* (p. 7). ACM.

Reutter, J. L. (2013). *Graph Patterns: Structure, Query Answering and Applications in Schema Mappings and Formal Language Theory.* Ph.D. thesis Laboratory for Foundations of Computer Science School of Informatics University of Edinburgh.

Segaran, T., Evans, C., Taylor, J., Toby, S., Colin, E., & Jamie, T. (2009). *Programming the Semantic Web.* (1st ed.). O'Reilly Media, Inc.

Tang, L., & Liu, H. (2009). Relational learning via latent social dimensions. In *Proceedings of the 15th ACM SIGKDD international conference on Knowledge discovery and data mining* (pp. 817–826). ACM.

Zou, L., Chen, L., & Özsu, M. T. (2009). Distance-join: Pattern match query in a large graph database. *Proc. VLDB Endow.*, *2*, 886–897. URL: `http://dx.doi.org/10.14778/1687627.1687727`. doi:`10.14778/1687627.1687727`.

Shuai Ma, Yang Cao, Wenfei Fan, Jinpeng Huai, Tianyu Wo. *Strong Simulation: Capturing Topology in Graph Pattern Matching. ACM Trans. Database Syst.*, January 2014, Volume 39, Number 1, Article No. 4. ISSN: 0362-5915. Publisher: Association for Computing Machinery. Address: New York, NY, USA. DOI: 10.1145/2528937. Keywords: dual simulation, graph simulation, data locality, Strong simulation, subgraph isomorphism.