Towards efficient multilayer network data management

Georgios Panayiotou, Matteo Magnani, Bruno Pinaud

Noname manuscript No.
(will be inserted by the editor)

Towards efficient multilayer network data management

Georgios Panayiotou · Matteo Magnani · Bruno Pinaud

Abstract Real-world multilayer networks can be very large and there can be multiple choices regarding what should be modeled as a layer. Therefore, there is a need for their effective storage and manipulation. Currently, multilayer network analysis software use different data structures and manipulation operators. We aim to categorize operators in order to assess which structures work best for certain operator classes and data features. In this work, we propose a preliminary taxonomy of layer and data manipulation operators. We also design and execute a benchmark of select software and operators to identify potential for optimization.

Keywords Multilayer networks · Multiplex networks · Data management

1 Introduction

Multilayer networks have become increasingly popular across various disciplines for representing, manipulating and analyzing complex systems, such as brain [1] and ecological [2] networks. Typical applications also include social networks, which often consist of millions of actors and associated relationships when collected from online sources [3]. With large multilayer network data availability increasing, an open challenge is to provide a framework for effective storage and access, which will help reducing data preprocessing costs [4] and support development of interactive analysis systems [5].

Using established database management systems for multilayer network data storage is not ideal, as the state-of-the-art standards, i.e. relational and graph database systems, lack methods to represent, manipulate and analyze multilayer network data. On the other hand, there is a lack of consensus on the appropriate data structure for multilayer network data storage, as quite a few different alternatives exist within analysis and visualization libraries. More importantly, there

G. Panayiotou · M. Magnani
InfoLab, Dept. of Information Technology, Uppsala University, Sweden

B. Pinaud
Univ. Bordeaux, CNRS, Bordeaux INP, LaBRI, UMR 5800, France

is an imbalance between scalability of a software's underlying data structures and process efficiency; depending on the chosen data model, the performance of even simple layer manipulation operators vastly differs (cf. Sect. 3).

Despite increasing interest in multilayer networks, work on multilayer network data management remains scarce. Previous work by the data engineering community on heterogeneous information networks [6] neither includes the concept of layer nor has looked into specialized storage solutions to optimize layer operations. In the recently proposed multilayer graphs, an extension for property graphs [7], the term layer is used, but instead refers to levels of nesting in the network. Also recently, a data mining approach was proposed, based on converting EER-diagrams into multilayer networks [8]; however, the previous does not directly address the impact of different data structures for multilayer networks.

With the previous issues in mind, this work provides basis for a data management framework natively supporting multilayer networks, by considering storage, access and manipulation of both vertices and edges in the network, and the layers themselves. We propose a taxonomy covering and extending layer and data manipulation operators found in popular multilayer network software [9–12]. Finally, we provide a benchmark of select operators' performance on currently available libraries and database models able to represent a multilayer network. We aim to both compare alternative data management approaches from a scalability and efficiency perspective, and spotlight processes in need of further research.

2 Data management taxonomy

We provide a taxonomy of tasks for multilayer network representation and manipulation, covering and extending operators found in major multilayer network software previously cited, which introduces a common basis for comparing various layer-supporting data structures. Our taxonomy differs from the recently introduced visualization-centered one [5]; our proposal focuses on data management operators, as typically provided by database management systems.

A preliminary taxonomy can be seen in Table 1. Our operator examples consider a generic multilayer network model similar to the one by Kivelä et al. [13], for a multilayer network $M = (V_M, E_M, V, \mathbf{L})$ with d dimensions. We use δ to denote a dimension of the network, v to denote a vertex, l to denote a single layer and σ to denote a predicate.

Tasks in the layer definition category focus on redefining the layer structure by creating and deleting dimensions. Layer manipulation operators also create new layers (or views thereof), but they instead derive new layers based on existing layers' topological features, e.g. layer flattening, projection and difference.

We consider layer query as a special class, as it arguably falls under both layer and data manipulation categories. Operators here essentially obtain a subset of the nodes and edges that satisfy a condition related to either attributes or topological features.

Finally, operators in the data manipulation category, similarly to manipulating a single-layer graph database, include adding and removing nodes or edges, while also considering the layers they associate to.

Table 1 Preliminary taxonomy of layer-supporting data model operators

Operator class	**Examples**
Layer definition	create/delete-dimension(δ) create/delete-layer(δ)
Layer manipulation	flatten-layer(δ, l_s, l_t) project-layer(δ, l_s, l_t) diff-layer(δ, l_s, l_t)
Layer query	filter-layer$_\sigma(l)$
Data manipulation	add/update/remove-node(v, l) add/update/remove-edge(v_s, l_s, v_t, l_t)

3 Benchmark

Multilayer networks can be implemented using data structures as different as tensors, dictionary-based adjacency lists, tables in relational databases, or combinations of the previous, as with graph database systems. We perform a comparison of these structures on the operators in our taxonomy, aiming to discover processes in need of optimization. Our primary focus is on layer manipulation and query operators, as we expect their performance within different systems to vary with respect to the underlying data structure and whether that is layer-native.

As an example, we consider the performance of the layer aggregation operator in `MuxViz` [9], `multinet` [10] and `Pymnet` [11] libraries. This experiment is done for randomly generated two-layer Erdős-Rényi multilayer networks, with network size (nodes associated to each layer) ranging between 100–100,000 and average node degree ≈ 4. Note that `multinet` and `MuxViz` are libraries for R, while `Pymnet` is a library for Python, which can slightly affect performance.

Fig. 1 confirms the aforementioned imbalance between data structure efficiency and scalability, as systems exhibit various behaviours related to their underlying data structures; notice that some systems can handle larger networks, but at the expense of efficiency.

To expand our benchmark, we will also compare our findings with equivalent database models able to represent multilayer networks. Namely, we consider a relational database model representing each element of the multilayer network quadruple as its own relation and a graph database model operating on the network defined by (V_M, E_M).

4 Conclusion and future work

Despite only dealing with preliminary results, we can already stress the importance of the chosen data structure, as this choice can determine which networks we can practically handle. These results should be complemented with a comparison of additional established multilayer network operators on both the aforementioned libraries and database models, in order to explore the strengths and weaknesses of various multilayer network management approaches. Finally, future work includes defining a minimal set of necessary operators for multilayer network data and a rule-based approach to transform multilayer networks into database models.

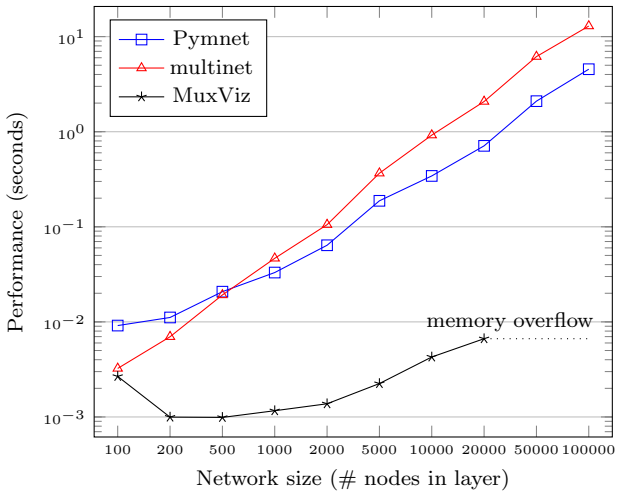

Fig. 1 Comparison of three multilayer network libraries for performance of two layer network aggregation over network size.

Acknowledgements This work has been partly funded by eSSENCE, an e-Science collaboration funded as a strategic research area of Sweden, and the FRÖ program by the French Institute in Sweden.

References

1. M. De Domenico, GigaScience **6**(5), 1 (2017)
2. S. Timóteo, M. Correia, S. Rodríguez-Echeverría, H. Freitas, R. Heleno, Nature Communications **9**, 140 (2018)
3. M.E. Dickison, M. Magnani, L. Rossi, *Multilayer Social Networks* (Cambridge University Press, 2016)
4. R. Interdonato, M. Magnani, D. Perna, A. Tagarelli, D. Vega, Computer Science Review **36**, 100246 (2020)
5. F. McGee, M. Ghoniem, B. Otjacques, B. Renoust, D. Archambault, A. Kerren, B. Pinaud, G. Melançon, M. Pohl, T. von Landesberger, *Visual Analysis of Multilayer Networks*. Synthesis Lectures on Visualization (Springer International Publishing, Cham, 2021)
6. C. Shi, Y. Li, J. Zhang, Y. Sun, P.S. Yu, IEEE Transactions on Knowledge and Data Engineering **29**(1), 17 (2017)
7. R. Angles, A. Hogan, O. Lassila, C. Rojas, D. Schwabe, P. Szekely, D. Vrgoč, in *Proceedings of the 5th ACM SIGMOD Joint International Workshop on Graph Data Management Experiences & Systems (GRADES) and Network Data Analytics (NDA)* (Association for Computing Machinery, New York, NY, USA, 2022), GRADES-NDA '22, pp. 1–6
8. A. Santra, K. Komar, S. Bhowmick, S. Chakravarthy, Data & Knowledge Engineering **141**, 102058 (2022)
9. M. De Domenico, M.A. Porter, A. Arenas, Journal of Complex Networks **3**(2), 159 (2015)
10. M. Magnani, L. Rossi, D. Vega, Journal of Statistical Software **98**, 1 (2021)
11. Multilayer Networks Library for Python (Pymnet) — Multilayer Networks Library 0.1 documentation. URL http://www.mkivela.com/pymnet/
12. D. Auber, D. Archambault, R. Bourqui, M. Delest, J. Dubois, A. Lambert, P. Mary, M. Mathiaut, G. Melançon, B. Pinaud, B. Renoust, J. Vallet, in *Encyclopedia of Social Network Analysis and Mining*, ed. by R. Alhajj, J. Rokne (Springer, New York, NY, 2017), pp. 1–28
13. M. Kivela, A. Arenas, M. Barthelemy, J.P. Gleeson, Y. Moreno, M.A. Porter, Journal of Complex Networks **2**(3), 203 (2014)