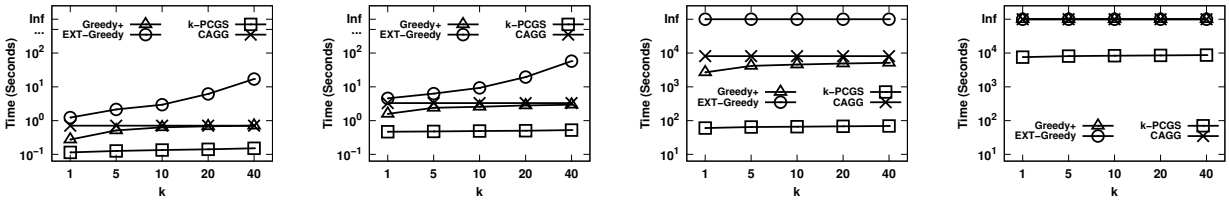
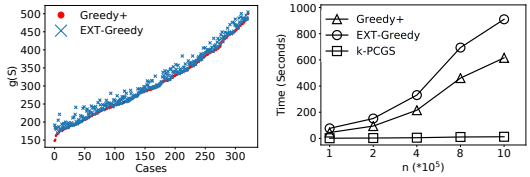


(a) ANIM (b) IMAGE (c) YAGO (d) UCI

Figure 6: Efficiency evaluation of Greedy+, EXT-Greedy, k-PCGS **and** CAGG **methods on DAGs**



(a) Approximation evaluation (b) Scalability test

Figure 7: Evaluation on synthetic DAGs.

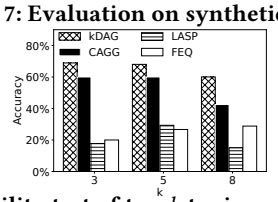


Figure 8: Usability test of top-k topic recommendation.

Top-k diversification. There exist several studies on the top-k graph diversification [5, 10, 12, 13, 21, 22, 29–31]. Qin et al.[21] investigate the diversified top-k search results in a graph. Ranu et al. [22] propose NB-Index to solve the top-k representative queries on graph databases. Based on a level-wise subgraph search, DSQL is proposed for top-k diversified subgraph querying in [29]. Long et al. [30] find top-k maximal cliques that can cover most number of nodes in a graph. Different from the above works on the top-k diversification on graph databases, subgraph queries, and cliques, this paper studies the graph summarization problem in hierarchical DAGs.

9 CONCLUSION

In this paper, we formulate and study a new kDAG-problem, which finds k representative vertices to summarize a hierarchical DAG associated with vertex weights. Due to the problem NP-hardness, we propose efficient greedy algorithms to tackle it. In addition, we develop two improved algorithms to find better answers with a theoretical guarantee in quality and be faster with theoretical complexity analysis, respectively. The k-PCGS method is scalable based on the candidate pruning and DAG compression. Extensive experiments validate the effectiveness and efficiency of our proposed algorithms.

ACKNOWLEDGEMENT

This paper is supported by NSFC 61702435, HK RGC GRF 12200917, 12201518, 12232716, 12201119, HK RGC CRF C6030-18G, and Guangdong Basic and Applied Basic Research Foundation (Project No. 2019B1515130001). Xin Huang is the corresponding author.

REFERENCES

[1] http://disease-ontology.org.
[2] https://www.snomed.org/snomed-ct/five-step-briefing.
[3] https://dl.acm.org/sig/sigmod.
[4] Leman Akoglu, Duen Horng Chau, U Kang, Danai Koutra, and Christos Faloutsos. 2012. Opavion: Mining and visualization in large graphs. In *SIGMOD*. 717–720.
[5] Ilio Catallo, Eleonora Ciceri, Piero Fraternali, Davide Martinenghi, and Marco Tagliasacchi. 2013. Top-k diversity queries over bounded regions. *TODS* 38, 2 (2013), 10.
[6] Šejla Čebirić, François Goasdoué, and Ioana Manolescu. 2015. Query-oriented summarization of RDF graphs. *PVLDB* 8, 12 (2015), 2012–2015.
[7] Jia Deng, Wei Dong, Richard Socher, Li-Jia Li, Kai Li, and Li Fei-Fei. 2009. Imagenet: A large-scale hierarchical image database. In *2009 IEEE conference on computer vision and pattern recognition*. 248–255.
[8] Georgios Fakas, Zhi Cai, and Nikos Mamoulis. 2015. Diverse and proportional size-l object summaries for keyword search. In *SIGMOD*. 363–375.
[9] Georgios J Fakas, Zhi Cai, and Nikos Mamoulis. 2011. Size-l object summaries for relational keyword search. *PVLDB* 5, 3 (2011), 229–240.
[10] Wenfei Fan, Xin Wang, and Yinghui Wu. 2013. Diversified top-k graph pattern matching. *PVLDB* 6, 13 (2013), 1510–1521.
[11] Xiangyang Gou, Lei Zou, Chenxingyu Zhao, and Tong Yang. 2019. Fast and Accurate Graph Stream Summarization. In *ICDE*. 1118–1129.
[12] Jinbin Huang, Xin Huang, Yuanyuan Zhu, and Jianliang Xu. 2019. Parameter-free Structural Diversity Search. In *WISE*. 677–693.
[13] Xin Huang, Hong Cheng, Rong-Hua Li, Lu Qin, and Jeffrey Xu Yu. 2013. Top-k structural diversity search in large networks. *PVLDB* 6, 13 (2013), 1618–1629.
[14] Xin Huang, Byron Choi, Jianliang Xu, William K Cheung, Yanchun Zhang, and Jiming Liu. 2017. Ontology-based Graph Visualization for Summarized View. In *CIKM*. 2115–2118.
[15] X Jing, JJ Cimino, et al. 2014. A Complementary Graphical Method for Reducing and Analyzing Large Data Sets. *Methods of information in medicine* 53, 3 (2014), 173–185.
[16] Xia Jing and James J Cimino. 2011. Graphical methods for reducing, visualizing and analyzing large data sets using hierarchical terminologies. In *AMIA Annual Symposium Proceedings*, Vol. 2011. 635.
[17] Richard M Karp. 1972. Reducibility among combinatorial problems. In *Complexity of computer computations*. 85–103.
[18] K Ashwin Kumar and Petros Efstathopoulos. 2018. Utility-driven graph summarization. *PVLDB* 12, 4 (2018), 335–347.
[19] Matthieu Latapy. 2008. Main-memory triangle computations for very large (sparse (power-law)) graphs. *Theor. Comput. Sci.* 407, 1-3 (2008), 458–473.
[20] Farzaneh Mahdisoltani, Joanna Biega, and Fabian M Suchanek. 2013. Yago3: A knowledge base from multilingual wikipedias.
[21] Lu Qin, Jeffrey Xu Yu, and Lijun Chang. 2012. Diversifying top-k results. *PVLDB* 5, 11 (2012), 1124–1135.
[22] Sayan Ranu, Minh Hoang, and Ambuj Singh. 2014. Answering top-k representative queries on graph databases. In *SIGMOD*. 1163–1174.
[23] Ryan A. Rossi and Nesreen K. Ahmed. 2015. The Network Data Repository with Interactive Graph Analytics and Visualization. In *AAAI*. 4292–4293.
[24] Yossi Rubner, Carlo Tomasi, and Leonidas J Guibas. 1998. A metric for distributions with applications to image databases. In *ICCV*. 59–66.
[25] Yufei Tao, Yuanbing Li, and Guoliang Li. 2019. Interactive graph search. In *SIGMOD*. 1393–1410.
[26] Yuanyuan Tian, Richard A Hankins, and Jignesh M Patel. 2008. Efficient aggregation for graph summarization. In *SIGMOD*. 567–580.
[27] You Wu, Junyang Gao, Pankaj K Agarwal, and Jun Yang. 2017. Finding diverse, high-value representatives on a surface of answers. *PVLDB* 10, 7 (2017), 793–804.
[28] X Yang, CM Procopiuc, and D Srivastava. 2011. Summary graphs for relational database schemas. *PVLDB* 4, 11 (2011), 899–910.
[29] Zhengwei Yang, Ada Wai-Chee Fu, and Ruifeng Liu. 2016. Diversified top-k subgraph querying in a large graph. In *SIGMOD*. 1167–1182.
[30] Long Yuan, Lu Qin, Xuemin Lin, Lijun Chang, and Wenjie Zhang. 2016. Diversified top-k clique search. *The VLDB Journal* 25, 2 (2016), 171–196.
[31] Tao Zhou, Zoltán Kuscsik, Jian-Guo Liu, Matúš Medo, Joseph Rushton Wakeling, and Yi-Cheng Zhang. 2010. Solving the apparent diversity-accuracy dilemma of recommender systems. *PNAS* 107, 10 (2010), 4511–4515.
[32] Xuliang Zhu, Xin Huang, Byron Choi, Jianliang Xu, William K. Cheung, Yanchun Zhang, and Jiming Liu. 2020. Ontology-based Graph Visualization for Summarized View. arXiv:arXiv:2008.03053