

# The Power of Summarization in Graph Mining and Learning: Smaller Data, Faster Methods, More Interpretability

Danai Koutra  
University of Michigan  
dkoutra@umich.edu

## ABSTRACT

Our ability to generate, collect, and archive data related to everyday activities, such as interacting on social media, browsing the web, and monitoring well-being, is rapidly increasing. Getting the most benefit from this large-scale data requires analysis of patterns it contains, which is computationally intensive or even intractable. Summarization techniques produce compact data representations (summaries) that enable faster processing by complex algorithms and queries.

This talk will cover summarization of interconnected data (graphs) [3], which can represent a variety of natural processes (e.g., friendships, communication). I will present an overview of my group's work on bridging the gap between research on summarized network representations and real-world problems. Examples include summarization of massive knowledge graphs for refinement [2] and on-device querying [4], summarization of graph streams for persistent activity detection [1], and summarization within graph neural networks for fast, interpretable classification [5]. I will conclude with open challenges and opportunities for future research.

## PVLDB Reference Format:

Danai Koutra. The Power of Summarization in Graph Mining and Learning: Smaller Data, Faster Methods, More Interpretability. PVLDB, 14(13): 3416-3416, 2021.  
doi:10.14778/3484224.3484238

## BIOGRAPHY

Danai Koutra is an Associate Director of the Michigan Institute for Data Science (MIDAS) and a Morris Wellman Associate Professor in Computer Science and Engineering at the University of Michigan, where she leads the Graph Exploration and Mining at Scale (GEMS) Lab. Her research focuses on practical and scalable methods for large-scale real networks, and her interests include graph summarization, knowledge graph mining, graph learning, similarity and alignment, and anomaly detection. She has won an NSF CAREER award, an ARO Young Investigator award, the 2020 SIGKDD Rising Star Award, research faculty awards from Google, Amazon, Facebook, and Adobe, a Precision Health Investigator award, the 2016 ACM SIGKDD Dissertation award, and an honorable mention for the SCS Doctoral Dissertation Award (CMU). She holds one "rate-1" patent on bipartite graph alignment, and has multiple papers in top data mining conferences, including 8 award-winning papers.

This work is licensed under the Creative Commons BY-NC-ND 4.0 International License. Visit <https://creativecommons.org/licenses/by-nc-nd/4.0/> to view a copy of this license. For any use beyond those covered by this license, obtain permission by emailing [info@vldb.org](mailto:info@vldb.org). Copyright is held by the owner/author(s). Publication rights licensed to the VLDB Endowment.

Proceedings of the VLDB Endowment, Vol. 14, No. 13 ISSN 2150-8097.  
doi:10.14778/3484224.3484238

She is the Secretary of the new SIAG on Data Science, an Associate Editor of ACM TKDD, and has served multiple times in the organizing committees of all the major data mining conferences. She has worked at IBM, Microsoft Research, and Technicolor. She earned her Ph.D. and M.S. in Computer Science from CMU in 2015 and her diploma in Electrical and Computer Engineering at the National Technical University of Athens in 2010.

## REFERENCES

- [1] Caleb Belth, Xinyi Zheng, and Danai Koutra. 2020. *Mining Persistent Activity in Continually Evolving Networks*. 934–944.
- [2] Caleb Belth, Xinyi Zheng, Jilles Vreeken, and Danai Koutra. 2020. What is Normal, What is Strange, and What is Missing in a Knowledge Graph: Unified Characterization via Inductive Summarization. In *The ACM Web Conference (WWW)*. 1115–1126.
- [3] Yike Liu, Tara Safavi, Abhilash Dighe, and Danai Koutra. 2018. Graph Summarization Methods and Applications: A Survey. *ACM Computing Surveys* 51, 3 (2018), 62:1–62:34.
- [4] Tara Safavi, Davide Mottin, Caleb Belth, Emmanuel Müller, Lukas Faber, and Danai Koutra. 2019. Personalized Knowledge Graph Summarization: From the Cloud to Your Pocket. In *IEEE International Conference on Data Mining (ICDM)*. 528–537.
- [5] Yujun Yan, Jiong Zhu, Marlena Duda, Eric Solarz, Chandra Sripada, and Danai Koutra. 2019. GroupINN: Grouping-Based Interpretable Neural Network for Classification of Limited, Noisy Brain Data. In *ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD)*. 772–782.