Summarizing Graphs at Multiple Scales: New Trends

Danai Koutra
Computer Science and Engineering
University of Michigan
dkoutra@umich.edu

Jilles Vreeken
CISPA Helmholtz Center
for Information Security
jv@cispa.saarland

Francesco Bonchi
ISI Foundation
Turin, Italy
francesco.bonchi@isi.it

***Abstract*—Recent advances in computing resources have made it possible to collect enormous amounts of interconnected data, such as social media interactions, web activity, knowledge bases, product and service purchases, autonomous vehicle routing, smart home sensor data, and more. The massive scale and complexity of this data, however, not only vastly surpasses human processing power, but also goes beyond limitations with regard to computation and storage. That is, there is an urgent need for methods and tools that *summarize* large interconnected data to enable faster computations, storage reduction, interactive large-scale visualization and understanding, and pattern discovery.**

Network summarization—which aims to find a small representation of an original, larger graph—features a variety of methods with different goals and for different input data representations (e.g., attributed graphs, time-evolving or streaming graphs, heterogeneous graphs). The objective of this tutorial is to give a systematic overview of methods for summarizing and explaining graphs at different scales: the node-group level, the network level, and the multi-network level. We emphasize the current challenges, present real-world applications, and highlight the open research problems in this vibrant research area.

I. Tutorial Overview

Our tutorial is structured into three parts. In the first part we focus on an important, though under-studied topic, namely that of summarizing and explaining a *subset of nodes* in a larger graph. These nodes can either be given to us beforehand (e.g. hand-picked) or discovered by an independent algorithm, and we are asked to summarize them using the graph [3], [2]. Alternatively, the task of discovering those sets of nodes can be integrated (i.e., subgroup or bump discovery [6]); the goal becomes to discover descriptions (e.g. graph queries) that identify subsets of nodes that we can summarize well given the graph structure [4]. Both approaches, and especially their combination, are of particular interest in interactive systems where users want the system to explain the specific part of the graph that is of interest [11].

The second part focuses on methods for summarizing a single graph (e.g., a snapshot or aggregate network) [9], as a whole. We discuss both methods that use solely the structure of a graph, and methods that also leverage side information, such as node and edge attributes [13], [8], [10]. In both cases, we provide a taxonomy of the approaches based on their key methodological ideas (e.g., group-based vs. influence-based vs. pattern-based), output type (e.g., supergraph vs. sparsified graph vs. compressed graph), and main objective (e.g., storage, efficiency, visualization).

In the third part we turn to multi-network summarization. In addition to covering scalable techniques tailored to large time-evolving or streaming networks [9], [1], [5], [12], [14], we present recent advances in summarizing multiple *disparate* networks simultaneously in order to construct domain-specific summaries [7] or model the networks' co-evolution [15].

The slides and more information for this tutorial are available at http://web.eecs.umich.edu/~dkoutra/tut/icdm18.html.

II. The Tutors

Danai Koutra is an Assistant Professor in CSE at University of Michigan, where she leads the Graph Exploration and Mining at Scale (GEMS) lab. Jilles Vreeken is tenured Faculty at the CISPA Helmholtz Center for Information Security, as well as Senior Researcher at the Max Planck Institute for Informatics. Francesco Bonchi is Research Leader at the ISI Foundation in Turin, Italy, and Research Director for Big Data and Data Science at Eurecat in Barcelona, Spain.

Selected References

[1] Bijaya Adhikari, Yao Zhang, Aditya Bharadwaj, and B Aditya Prakash. Condensing temporal networks using propagation. In *SDM*, pages 417–425. SIAM, 2017.

[2] Florian Adriaens, Jefrey Lijffijt, and Tijl De Bie. Subjectively interesting connecting trees. In *ECML/PKDD*, pages 53–69, 2017.

[3] Leman Akoglu, Jilles Vreeken, Hanghang Tong, Nikolaj Tatti, and Christos Faloutsos. Mining Connection Pathways for Marked Nodes in Large Graphs. In *SDM*. SIAM, 2013.

[4] Martin Atzmueller. Local exceptionality detection on social interaction networks. In *ECML/PKDD*, pages 298–302, 2016.

[5] Sorour E. Amiri, Liangzhe Chen, and B. Aditya Prakash. Efficiently summarizing attributed diffusion networks. *DAMI*, 05 2018.

[6] Aristides Gionis, Michael Mathioudakis, and Antti Ukkonen. Bump hunting in the dark: Local discrepancy maximization on graphs. *IEEE Trans. Knowl. Data Eng.*, 29(3):529–542, 2017.

[7] Di Jin and Danai Koutra. Exploratory analysis of graph data by leveraging domain knowledge. In *ICDM*. IEEE, 2017.

[8] Danai Koutra, U Kang, Jilles Vreeken, and Christos Faloutsos. Summarizing and Understanding Large Graphs. In *Statistical Analysis and Data Mining*. John Wiley & Sons, Inc., 2015.

[9] Yike Liu, Tara Safavi, Abhilash Dighe, and Danai Koutra. Graph summarization methods and applications: A survey. *ACM Comput. Surv.*, 51(3):62:1–62:34, 2018.

[10] Yasir Mehmood, Nicola Barbieri, Francesco Bonchi, and Antti Ukkonen. CSI: Community-level Social Influence Analysis. In *ECML/PKDD*, pages 48–63. Springer, 2013.

[11] Robert Pienta, Minsuk Kahng, Zhiyuan Lin, Jilles Vreeken, Partha P. Talukdar, James Abello, Ganesh Parameswaran, and Duen Horng Chau. FACETS: adaptive local exploration of large graphs. In *SDM*, 2017.

[12] Qiang Qu, Siyuan Liu, Christian S. Jensen, Feida Zhu, and Christos Faloutsos. Interestingness-driven diffusion process summarization in dynamic networks. In *ECML/PKDD*, pages 597–613, 2014.

[13] Matteo Riondato, David García-Soriano, and Francesco Bonchi. Graph summarization with quality guarantees. In *ICDM*, 2014.

[14] Neil Shah, Danai Koutra, Lisa Jin, Tianmin Zou, Brian Gallagher, and Christos Faloutsos. On summarizing large-scale dynamic graphs. *IEEE Data Eng. Bull.*, 40(3):75–88, 2017.

[15] Wenchao Yu, Charu C. Aggarwal, and Wei Wang. Modeling co-evolution across multiple networks. In *SDM*, pages 675–683, 2018.

978-1-5386-9159-5/18/$31.00 ©2018 IEEE
DOI 10.1109/ICDM.2018.00141