Network Alignment: Recent Advances and Future Directions

Si Zhang
sizhang2@illinois.edu
University of Illinois at Urbana-Champaign

Hanghang Tong
htong@illinois.edu
University of Illinois at Urbana-Champaign

ABSTRACT

In the era of big data, networks are often from multiple sources such as the social networks of diverse platforms (e.g., Facebook, Twitter), protein-protein interaction (PPI) networks of different tissues, transaction networks at multiple financial institutes and knowledge graphs derived from a variety of knowledge bases (e.g., DBpedia, Freebase, etc.). The very first step before exploring insights from these multi-sourced networks is to integrate and unify different networks. In general, network alignment is such a task that aims to uncover the correspondences among nodes across different graphs. The challenges of network alignment include: (1) the heterogeneity of the multi-sourced networks, e.g., different structural patterns, (2) the variety of the real-world networks, e.g., how to leverage the rich contextual information, and (3) the computational complexity. The goal of this tutorial is to (1) provide a comprehensive overview of the recent advances in network alignment, and (2) identify the open challenges and future trends. We believe this can be beneficial to numerous application problems, and attract both researchers and practitioners from both data mining area and other interdisciplinary areas. In particular, we start with introducing the backgrounds, problem definition and key challenges of network alignment. Next, our emphases will be on (1) the recent techniques on addressing network alignment problem and other related problems with a careful balance between the algorithms and applications, and (2) the open challenges and future trends.

1 INTENDED AUDIENCE

All researchers and practitioners engaged in big data researches (e.g., graph mining and related domains such as social network analysis, bioinformatics, cybersecurity, knowledge graphs) are welcome. No prior knowledge on specific algorithms is required. The audiences are assumed to have the basic knowledge on linear algebra and machine learning. The tutorial aims to achieve a good balance between the introductory and advanced materials (40% novice, 30% intermediate, 30% advanced).

2 PRESENTER BIOGRAPHY

The presenters and contributors include Si Zhang and Hanghang Tong. Their biographies and expertises are elaborated as follows.

Si Zhang. He is currently a Ph.D student in the Department of Computer Science at University of Illinois at Urbana-Champaign. He received his MS degree in Computer Engineering from Arizona State University in 2015 and B.Eng degree in ECE from Xi'an Jiaotong University in 2014. His current research interests lie in the large-scale data mining and machine learning, especially graph mining such as network alignment, dense subgraph detection and graph neural networks. He has been working on network alignment for 4 years, which results in several publications at major conferences and journals. He has served as a program committee member in top data mining and artificial intelligence venues (e.g., ICML, AAAI, IJCAI, WWW, PAKDD, SDM, etc). For more information, please refer to his personal website at https://sizhang2.web.illinois.edu/.

Hanghang Tong. He is an associate professor with the Department of Computer Science at University of Illinois at Urbana-Champaign. Before that he was an associate professor at School of Computing, Informatics, and Decision Systems Engineering (CIDSE), Arizona State University. He received his M.Sc. and Ph.D. degrees from Carnegie Mellon University in 2008 and 2009, both in Machine Learning. His research interest is in large scale data mining for graphs and multimedia. He has received several awards, including SDM/IBM Early Career Data Mining Research award (2018), NSF CAREER award (2017), ICDM 10-Year Highest Impact Paper award (2015) and four best paper awards (TUP'14, CIKM'12, SDM'08, ICDM'06). He is the Editor-in-Chief of SIGKDD Explorations (ACM), an action editor of Data Mining and Knowledge Discovery (Springer), and an associate editor of Knowledge and Information Systems (Springer) and Neurocomputing Journal (Elsevier); and has served as a program committee member in multiple data mining, database and artificial intelligence venues (e.g., SIGKDD, SIGMOD, AAAI, WWW, CIKM, etc.). For more information, please refer to his personal website at http://tonghanghang.org/. He has given several tutorials at top-tier conferences, such as IEEE Big Data 2015, SDM 2016, WSDM 2018, KDD 2018 (http://www.public.asu.edu/~liangyue/team-tutorial.html), etc.

3 OUTLINE OF THE COVERED TOPICS

- Introduction (20 minutes)
  - Motivations
  - Problem definitions and related settings
  - Key challenges
  - Traditional solutions and limitations
- Part I: Recent Network Alignment Algorithms (90 minutes)
  - Pairwise network alignment
  - Collective network alignment
  - High-order network alignment
  - Hierarchical network alignment
  - Knowledge graph alignment
  - Cross-layer dependency inference
- Part II: Network Alignment Applications (40 minutes)
  - Applications in social analysis
  - Applications in bioinformatics
  - Applications in knowledge completion

CIKM '20, October 19–23, 2020, Virtual Event, Ireland

ACM ISBN 978-1-4503-6859-9/20/10.
https://doi.org/10.1145/3340531.3412168

- Applications in security
- Part III: Future Research Directions (30 minutes)
  - Big network alignment
  - Adversarial network alignment
  - Active network alignment
  - Integrated network alignment

4 DESCRIPTIONS OF THE COVERED TOPICS

4.1 Part I: Network Alignment Algorithms

In this part, we review the state-of-the-art techniques on network alignment. We first categorize network alignment algorithms into the following scenarios. (1) *Pairwise network alignment*. It aims to find the node correspondence across two networks. Various methods have been proposed including: alignment consistency based methods [1, 8, 19], embedding based methods [6, 9, 22], optimal transport based methods [10, 14]. (2) *Collective network alignment*, to collectively align multiple networks. Representative works include the alignment consistency based methods [18, 21], embedding based methods [3, 4]. (3) *High-order network alignment*. Different from pairwise network alignment that maximizes the number of conserved edges among the aligned nodes, the number of higher-order substructures (e.g., triangles) is maximized [11]. (4) *Hierarchical network alignment*, to simultaneously unveil the correspondence among nodes and clusters at different resolutions. We present a recent method [20] which uses the multilevel optimization and multi-resolution matrix factorization [7]. In addition, we present another two closely related topics. These include: (1) knowledge graph alignment to align entities across different knowledge graphs, which can be further categorized into knowledge graph embedding based methods [12, 23], and graph neural network based methods [13, 16], and (2) cross-layer dependency inference which aims to infer the node dependencies in the multi-layered networks [2].

4.2 Part II: Network Alignment Applications

In this part, we will present how network alignment can be used in numerous application domains including: (1) *social analysis* where network alignment is used to unveil unique users on different social platforms [17], (2) *bioinformatics* where network alignment is often used to align different PPI networks for identifying functionally similar regions across different species and transfer the knowledge which further benefits studying more sophisticated problems (e.g., gene-disease associations) [5], (3) *knowledge completion* where entities in knowledge graphs are aligned to construct a unified knowledge base [23], and (4) *security* where multi-sourced information can be integrated to recognize adversarial activities [15].

4.3 Part III: Future Research Directions

In this part, we will summarize some open challenges and future trends in this field, including (1) *big network alignment* that aims to address the 4Vs characteristics (volume, variety, velocity and veracity) of big networks, (2) *adversarial network alignment* that leverages adversarial learning techniques to improve the alignment effectiveness and robustness, (3) *active network alignment* that can interact with other labeling oracles (e.g., humans) to deal with difficult alignment cases, and (4) *integrated network alignment* that instead of using network alignment as a separate pre-processing, integrates with other network mining tasks, e.g., explainable network alignment, fair network alignment, etc.

5 ACKNOWLEDGEMENTS

This work is supported by National Science Foundation under grant No. 1947135, and 2003924 by the NSF Program on Fairness in AI in collaboration with Amazon under award No. 1939725, and Department of Homeland Security under Grant Award Number 17STQAC00001-03-03. The content of the information in this document does not necessarily reflect the position or the policy of the Government or Amazon, and no official endorsement should be inferred. The U.S. Government is authorized to reproduce and distribute reprints for Government purposes notwithstanding any copyright notation here on.

REFERENCES

[1] Mohsen Bayati, David F Gleich, Amin Saberi, and Ying Wang. 2013. Message-passing algorithms for sparse network alignment. *ACM TKDD* (2013).
[2] Chen Chen, Hanghang Tong, Lei Xie, Lei Ying, and Qing He. 2016. FASCINATE: fast cross-layer dependency inference on multi-layered networks. In *KDD*.
[3] Xiaokai Chu, Xinxin Fan, Di Yao, Zhihua Zhu, Jianhui Huang, and Jingping Bi. 2019. Cross-Network Embedding for Multi-Network Alignment. In *WWW*. ACM.
[4] Boxin Du and Hanghang Tong. 2019. MrMine: Multi-resolution Multi-network Embedding. In *CIKM*.
[5] Fazle E Faisal, Lei Meng, Joseph Crawford, and Tijana Milenković. 2015. The post-genomic era of biological network alignment. *EURASIP Journal on Bioinformatics and Systems Biology* (2015).
[6] Mark Heimann, Haoming Shen, Tara Safavi, and Danai Koutra. 2018. Regal: Representation learning-based graph alignment. In *CIKM*. ACM.
[7] Risi Kondor, Nedelina Teneva, and Vikas Garg. 2014. Multiresolution matrix factorization. In *ICML*.
[8] Danai Koutra, Hanghang Tong, and David Lubensky. 2013. Big-align: Fast bipartite graph alignment. In *ICDM*. IEEE.
[9] Li Liu, William K Cheung, Xin Li, and Lejian Liao. 2016. Aligning Users across Social Networks Using Network Embedding.. In *IJCAI*.
[10] Hermina Petric Maretic, Mireille EL Gheche, Giovanni Chierchia, and Pascal Frossard. 2019. GOT: An Optimal Transport framework for Graph comparison. *arXiv* (2019).
[11] Shahin Mohammadi, David F Gleich, Tamara G Kolda, and Ananth Grama. 2017. Triangular Alignment TAME: A Tensor-Based Approach for Higher-Order Network Alignment. *IEEE/ACM TCBB* (2017).
[12] Zequn Sun, Wei Hu, Qingheng Zhang, and Yuzhong Qu. 2018. Bootstrapping Entity Alignment with Knowledge Graph Embedding.. In *IJCAI*.
[13] Zhichun Wang, Qingsong Lv, Xiaohan Lan, and Yu Zhang. 2018. Cross-lingual knowledge graph alignment via graph convolutional networks. In *EMNLP*.
[14] Hongteng Xu, Dixin Luo, Hongyuan Zha, and Lawrence Carin. 2019. Gromov-wasserstein learning for graph matching and node embedding. *arXiv* (2019).
[15] Jiejun Xu, Hanghang Tong, Tsai-Ching Lu, Jingrui He, and Nadya Bliss. 2018. GTA 3 2018: Workshop on Graph Techniques for Adversarial Activity Analytics. In *WSDM*. ACM.
[16] Kun Xu, Liwei Wang, Mo Yu, Yansong Feng, Yan Song, Zhiguo Wang, and Dong Yu. 2019. Cross-lingual knowledge graph alignment via graph matching neural network. *arXiv* (2019).
[17] Reza Zafarani and Huan Liu. 2013. Connecting users across social media sites: a behavioral-modeling approach. In *KDD*. ACM.
[18] Jiawei Zhang and S Yu Philip. 2015. Multiple anonymized social networks alignment. In *ICDM*. IEEE.
[19] Si Zhang and Hanghang Tong. 2016. Final: Fast attributed network alignment. In *KDD*. ACM.
[20] Si Zhang, Hanghang Tong, Ross Maciejewski, and Tina Eliassi-Rad. 2019. Multilevel Network Alignment. In *WWW*. ACM.
[21] Yutao Zhang, Jie Tang, Zhilin Yang, Jian Pei, and Philip S Yu. 2015. Cosnet: Connecting heterogeneous social networks with local and global consistency. In *KDD*. ACM.
[22] Fan Zhou, Lei Liu, Kunpeng Zhang, Goce Trajcevski, Jin Wu, and Ting Zhong. 2018. Deeplink: A deep learning approach for user identity linkage. In *INFOCOM*. IEEE.
[23] Hao Zhu, Ruobing Xie, Zhiyuan Liu, and Maosong Sun. 2017. Iterative Entity Alignment via Joint Knowledge Embeddings.. In *IJCAI*.