

The 1st International Workshop on Federated Learning with Graph Data (FedGraph)

Carl Yang*
Emory University
Atlanta, GA, USA
j.carlyang@emory.edu

Xiaoxiao Li
University of British Columbia
Vancouver, BC, Canada
xiaoxiao.li@ece.ubc.ca

Nathalie Baracaldo
IBM Research
San Jose, CA, USA
baracald@us.ibm.com

Neil Shah
Snap Research
Seattle, WA, USA
nshah@snap.com

Chaoyang He
FedML Inc.
Los Angeles, CA, USA
ch@fedml.ai

Lingjuan Lyu
Sony AI
Tokyo, Japan
lingjuanlvsmile@gmail.com

Lichao Sun
Lehigh University
Bethlehem, PA, USA
lis221@lehigh.edu

Salman Avestimehr
University of Southern California
Los Angeles, CA, USA
avestime@usc.edu

ABSTRACT

The field of graph data mining, one of the most important AI research areas, has been revolutionized by graph neural networks (GNNs), which benefit from training on real-world graph data with millions to billions of nodes and links. Unfortunately, the training data and process of GNNs involving graphs beyond millions of nodes are extremely costly on a centralized server, if not impossible. Moreover, due to the increasing concerns about data privacy, emerging data from realistic applications are naturally fragmented, forming distributed private graphs of multiple “data silos”, among which direct transferring of data is forbidden. The nascent field of federated learning (FL), which aims to enable individual clients to jointly train their models while keeping their local data decentralized and completely private, is a promising paradigm for large-scale distributed and private training of GNNs. FEDGRAPH2022 aims to bring together researchers from different backgrounds with a common interest in how to extend current FL algorithms to operate with graph data models such as GNNs. FL is an extremely hot topic of large commercial interest and has been intensively explored for machine learning with visual and textual data. The exploration from graph mining researchers and industrial practitioners is timely catching up just recently. There are many unexplored challenges and opportunities, which urges the establishment of an organized and open community to collaboratively advance the science behind it. The prospective participants of this workshop will include researchers and practitioners from both graph mining and federated learning communities, whose interests include, but are not limited

to: graph analysis and mining, heterogeneous network modeling, complex data mining, large-scale machine learning, distributed systems, optimization, meta-learning, reinforcement learning, privacy, robustness, explainability, fairness, ethics, and trustworthiness.

ACM Reference Format:

Carl Yang, Xiaoxiao Li, Nathalie Baracaldo, Neil Shah, Chaoyang He, Lingjuan Lyu, Lichao Sun, and Salman Avestimehr. 2022. The 1st International Workshop on Federated Learning with Graph Data (FedGraph). In *Proceedings of the 31st ACM International Conference on Information and Knowledge Management (CIKM '22)*, October 17–21, 2022, Atlanta, GA, USA. ACM, New York, NY, USA, 2 pages. <https://doi.org/10.1145/3511808.3557495>

1 INTRODUCTION

This workshop is a forum for exchanging ideas and methods for federated learning with graph data, developing new common understandings of the problems at hand, sharing of data sets where applicable, and leveraging existing knowledge from different disciplines. The goal is to bring together researchers from academia, industry, and government, to create a forum for discussing recent advances and future directions. In doing so, we aim to better understand the overarching principles and the limitations of our current knowledge and methods, and to inspire research on new algorithms and techniques for federated learning with graph data.

We encourage submissions that span the spectrum from theoretical analysis to algorithms and implementation, to applications, empirical studies, and reflection papers. Topics of interest include, but are not limited to:

- How to achieve near optimal utility that is close to the centralized training setting with distributed graph datasets?
- How to handle more complex data correlation and heterogeneity under the context of graph mining?
- How to scale up the number of clients when the graph structures require heavy communications?
- How to reduce and compress communications among clients by exploring the architectures of graph models to achieve training efficiency?

*Corresponding author/Leading organizer.

Permission to make digital or hard copies of part or all of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for third-party components of this work must be honored. For all other uses, contact the owner/author(s).

CIKM '22, October 17–21, 2022, Atlanta, GA, USA

© 2022 Copyright held by the owner/author(s).

ACM ISBN 978-1-4503-9236-5/22/10.

<https://doi.org/10.1145/3511808.3557495>

- How to rigorously protect clients' graph data privacy during collaboration?
- How to properly evaluate innovative GNN models and FL algorithms towards realistic applications such as knowledge graph completion, recommendation in e-commerce networks and diagnosis prediction in healthcare networks?
- How to extend the concepts and algorithms of FedGraph to a broader range of complex data beyond classic graphs, such as heterogeneous networks, spatiotemporal networks, text-rich networks, multi-view networks, point clouds, trees, manifolds, and fractals?
- How to conduct federated learning with graph algorithms beyond GNNs such as graph kernels, belief propagation and spectral analysis?
- How to enhance the optimization of existing FL systems based on graph mining principles and techniques?
- How to guarantee fairness, ethics, and trustworthiness in the FedGraph context?

We welcome many kinds of papers, such as, but not limited to: novel research papers, demo papers, work-in-progress papers, visionary papers (white papers), appraisal papers of existing methods and tools (e.g., lessons learned), evaluatory papers which revisit validity of domain assumptions, relevant work that has been previously published and work that will be presented at the main conference. The accepted papers are published on the workshop's website and not considered archival for resubmission purposes. While all accepted papers will be presented with posters, high-quality accepted papers will also have the opportunity to participate in the oral/spotlight presentation, and win our Best Paper Award(s).

2 INVITED SPEAKERS

Dr. Bo Li is an Assistant Professor in the Department of Computer Science at the University of Illinois at Urbana-Champaign, and the recipient of the Symantec Research Labs Fellowship, MIT Technology Review TR-35 award, Intel Rising Star award, several Amazon Research Awards, NSF CAREER Award, and best paper awards in several machine learning and security conferences. Her research focuses on both theoretical and practical aspects of security, machine learning, privacy, game theory, and trustworthy machine learning. Her work has been featured by major publications and media outlets such as Nature, Wired, Fortune, and New York Times. **Dr. Nicolas Papernot** is an Assistant Professor in the Department of Electrical and Computer Engineering and the Department of Computer Science at the University of Toronto. He is also a faculty member at the Vector Institute where he holds a Canada CIFAR AI Chair, and a faculty affiliate at the Schwartz Reisman Institute. His research interests span the security and privacy of machine learning. His work on differentially private machine learning received a best paper award at ICLR 2017.

Dr. Peter Kairouz is a Research Scientist at Google, where he leads research efforts focused on federated learning and privacy-preserving technologies. He is the recipient of the 2012 Roberto Padovani Scholarship from Qualcomm's Research Center, the 2015 ACM SIGMETRICS Best Paper Award, the 2015 Qualcomm Innovation Fellowship Finalist Award, and the 2016 Harold L. Olesen Award for Excellence in Undergraduate Teaching from UIUC.

Dr. Xing Xie is currently a Senior Principal Research Manager at Microsoft Research Asia, and a guest Ph.D. advisor at the University of Science and Technology of China. During the past years, he has published over 300 papers, won the ACM SIGKDD China 2021 test of time award, the 10-year impact award honorable mention in ACM SIGSPATIAL 2020, the 10-year impact award in ACM SIGSPATIAL 2019, the best student paper award in KDD 2016, and the best paper awards in ICDM 2013 and UIC 2010. In Oct. 2009, he founded the SIGSPATIAL China chapter which was the first regional chapter of ACM SIGSPATIAL. He is a Fellow of China Computer Federation (CCF), and a Distinguished Member of ACM.

Dr. Carl Yang is an Assistant Professor in Emory University. He received his Ph.D. in Computer Science at University of Illinois, Urbana-Champaign in 2020, and B.Eng. in Computer Science and Engineering at Zhejiang University in 2014. His research interests span graph data mining, applied machine learning, knowledge graphs and federated learning, with applications in recommender systems, biomedical informatics, neuroscience and healthcare. He has received the Dissertation Completion Fellowship of UIUC in 2020, the Best Paper Award of ICDM in 2020, the Dissertation Award Finalist of KDD in 2021, and the Best Paper Award of KDD Health Day in 2022.

Dr. Salman Avestimehr is a Dean's Professor, the inaugural director of the USC-Amazon Center on Secure and Trusted Machine Learning (Trusted AI), and the director of the Information Theory and Machine Learning (vITAL) research lab at the ECE Department of University of Southern California. Dr. Avestimehr has received a number of awards for his research, including the James L. Massey Research & Teaching Award from IEEE Information Theory Society, an Information Theory Society and Communication Society Joint Paper Award, a Presidential Early Career Award for Scientists and Engineers (PECASE) from the White House (President Obama), a Young Investigator Program (YIP) award from the U.S. Air Force Office of Scientific Research, a National Science Foundation CAREER award, the David J. Sakrison Memorial Prize, and several Best Paper Awards at Conferences. He is a fellow of IEEE.

3 PROGRAM SUMMARY

The workshop has summoned a program committee of 21 confirmed members and received a total of 16 submissions. We plan to accept about 10 papers, among which about 6 papers will be presented as oral/spotlight and 1-2 will receive our best paper awards. We have also invited two authors, Yuexiang Xie and Liuyi Yao, of FederatedScale-GNN, the best paper of KDD 2022 ADS track, to give an in-person demo/tutorial of their new FedGraph platform.

ACKNOWLEDGEMENT

With pride and gratitude, we thank the sponsors of JZTData Technology, Tencent, FedML, and USC-Amazon Center, who financially supported our cash prizes for best papers, honorariums to speakers and travel awards to student volunteers. We also thank the student volunteers/organizers— Han Xie (Emory University), Hejie Cui (Emory University), Zishan Gu (Columbia University) and Emir Ceyani (University of Southern California)— for their efforts in helping the organization of FEDGRAPH2022.