

FedKDD 2025: The 2025 International Joint Workshop on Federated Learning for Data Mining and Graph Analytics

Carl Yang
Emory University
Atlanta, United States
j.carlyang@emory.edu

Guancheng Wan
Emory University
Atlanta, United States
gwan4@emory.edu

Zexi Li
Zhejiang University
Hangzhou, China
zexi.li@zju.edu.cn

Tianchi Liao
Sun Yat-Sen University
Guangzhou, China
liaotch@mail2.sysu.edu.cn

Zhuangdi Zhu
George Mason University
Fairfax, United States
zzhu24@gmu.edu

Zheng Xu
Google Research
Mountain View, United States
xuzheng@google.com

Junyuan Hong
University of Texas at Austin
Austin, United States
jyhong@utexas.edu

Chenyu You
Stony Brook University
Stony Brook, United States
chenyu.you@stonybrook.edu

Nathalie Baracaldo
IBM
San Jose, United States
baracald@us.ibm.com

Neil Shah
Snap
Seattle, United States
nshah@snap.com

Salman Avestimehr
University of Southern California
Los Angeles, United States
avestimehr@gmail.com

Jiayu Zhou
Michigan State University
East Lansing, United States
jiayuz@msu.edu

Abstract

Deep Learning has facilitated various high-stakes applications such as crime detection, urban planning, drug discovery, and healthcare. Its continuous success hinges on learning from massive data in miscellaneous sources, ranging from data with independent distributions to graph-structured data capturing intricate inter-sample relationships. Scaling up the data access requires global collaboration from distributed data owners. Yet, centralizing all data sources to an untrustworthy centralized server will put users' data at risk of privacy leakage or regulation violation. Federated Learning (FL) is a *de facto* decentralized learning framework that enables knowledge aggregation from distributed users without exposing private data. Though promising advances are witnessed for FL, new challenges are emerging when integrating FL with the rising needs and opportunities in data mining, graph analytics, foundation models, generative AI, and new interdisciplinary applications in science. By hosting this workshop, we aim to attract a broad range of audiences, including researchers and practitioners from academia and industry interested in the emergent challenges in FL. As an effort to advance the fundamental development of FL, this workshop will encourage ideas exchange on the trustworthiness, scalability, and robustness of distributed data mining and graph analytics and their emergent challenges.

Permission to make digital or hard copies of all or part 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).
KDD '25, Toronto, ON, Canada

© 2025 Copyright held by the owner/author(s).
ACM ISBN 979-8-4007-1454-2/2025/08
<https://doi.org/10.1145/3711896.3737861>

CCS Concepts

• Information systems → Data mining.

Keywords

Federated Learning, Distributed Data Mining, Trustworthiness, Applications, Graph Analytics

ACM Reference Format:

Carl Yang, Guancheng Wan, Zexi Li, Tianchi Liao, Zhuangdi Zhu, Zheng Xu, Junyuan Hong, Chenyu You, Nathalie Baracaldo, Neil Shah, Salman Avestimehr, and Jiayu Zhou. 2025. FedKDD 2025: The 2025 International Joint Workshop on Federated Learning for Data Mining and Graph Analytics. In *Proceedings of the 31st ACM SIGKDD Conference on Knowledge Discovery and Data Mining V.2 (KDD '25)*, August 3–7, 2025, Toronto, ON, Canada. ACM, New York, NY, USA, 2 pages. <https://doi.org/10.1145/3711896.3737861>

1 Introduction

Extracting and analyzing knowledge from a vast array of distributed data sources is crucial for the development of robust AI models to satisfy the increasing demands of advanced analytics. Traditional methods of centralized learning encounter numerous obstacles in distributed learning environments, primarily due to concerns over privacy and stringent regulatory standards such as the GDPR and HIPAA that govern the exchange of sensitive data. The core issue lies in the distributed nature of data and the regulations restricting the sharing of confidential information. Federated learning (FL), a novel approach, has risen to prominence by enabling the extraction of insights from sensitive, distributed data sources through the consolidation of knowledge from distributed models instead of direct data sharing, thereby addressing privacy concerns effectively.

Despite its undeniable success in addressing real-world challenges, applying FL to practical data mining tasks is fraught with

difficulties due to varying data characteristics. Among various data forms that fuel up AI developments, *graph* serves as a cornerstone of data mining due to its fundamental and flexible structures to model various types of real-world data into entities and relations. Graph analytics, fueled by recent advances in machine learning (e.g., graph neural networks and graph transformers), have been widely practiced in crucial application domains such as physics, material sciences, social sciences, finance, transportation, and public health. While pioneering FL developments were built upon images, texts, and tabular data sources from independent distributions, it is imperative to bridge FL with graph analytics to further expand the spectrum of practical FL that advances data mining with comprehensive data characteristics. Furthermore, FL faces hurdles stemming from adversarial computational nodes, obscured yet tainted training data, erratic network conditions, inconstant system dynamics, and diverse computational capabilities. These factors compromise the scalability and reliability of FL systems.

The workshop will delve into both foundational and advanced issues in FL, incorporating a focus on graph analytics within the following four categories. Different from the previous workshops, FedGraph at ICDM 2023 or Federated Learning for Distributed Data Mining at KDD 2023, the workshop will bring in new challenges accompanied by the emergence of large models, generative AI, and new interdisciplinary applications in science.

1) Scaling laws of FL facing increasingly larger and more heterogeneous data, models, and computation-or-communication resources. Essentially, we look for studies on how the effectiveness of existing or new FL algorithms changes by the scaling. **2) Safety.** Problems and solutions for the security, privacy, and social alignment of FL systems and the resultant models. Especially, when training large generative AI models, the potential risks and countermeasures in FL systems are welcome to discuss. **3) Graph Analytics.** Intuitions and solutions to close the gap between centralized and decentralized graph analysis. **4) Other high-stakes applications.** Explorations on novel research problems of FL and FL algorithms for real-world applications. Moreover, we aim to attract high-quality original research of federated learning with applications, evaluation, and algorithms. We also plan to invite open discussions on controversial yet crucial topics regarding FL systems and discuss their barriers in data mining.

2 Audience and Schedule

Diversity Commitment. During the selection of organizers and speakers, we commit to encouraging all forms of diversity. Both groups achieved gender parity and are also diverse with respect to affiliations, races, and nationalities. The full scale of scientific seniority is covered, including PhD candidates, senior research scientists, as well as assistant and full professors. We also cover diverse research backgrounds (graph mining, federated learning, machine learning, systems, privacy, and applications), and our invited talks cover a diverse set of perspectives related to federated learning for data mining and graph analytics.

Schedule. The format of the workshop is a half-day event (4 hours). We will invite 4 keynote speakers who are leading experts from academia and industry. The potential speakers include Prof. Qiang Yang from HKUST and WeBank, Prof. Fei Wang from

Cornell, Prof. Gauri Joshi from CMU, Prof. Giulia Fanti from CMU, and Prof. Ermin Wei from Northwestern. We will have contributed papers as both oral and poster presentations for accepted ones. We will highlight outstanding papers in the award ceremony. This is intended to encourage presenting high-quality work that interests the community.

Table 1: Workshop schedule.

| Time | Event |
|---------------------|--------------------------------------|
| 8:00 am - 8:10 am | Opening Ceremony |
| 8:10 am - 8:50 am | Keynote Talk 1 |
| 8:50 am - 9:30 am | Keynote Talk 2 |
| 9:30 am - 10:10 am | Keynote Talk 3 |
| 10:10 am - 10:50 am | Coffee Break and Poster Presentation |
| 10:50 am - 11:30 am | Keynote Talk 4 |
| 11:30 am - 11:50 am | Best Paper Presentation |
| 11:50 am - 12:00 pm | Award Ceremony and Closing Session |

3 Submission Guidelines

Important Dates. Paper submissions: May 26th. Paper notifications: June 22th. Workshop date: August 4.

We invite short technical papers - up to 5 pages including references and unlimited pages of appendix. All manuscripts should be submitted in a single PDF file including all content, figures, tables, and references, following the new Standard ACM Conference Proceedings Template. Additionally, papers must be in the two-column format, with the recommended setting for Latex file: `\documentclass[sigconf, review]{acmart}`. Papers should be submitted to openreview website at <https://openreview.net/group?id=KDD.org/2025/Workshop/FedKDD>. 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). We will also present accepted papers on our website. More details in our website: <https://fedkdd.github.io/fedkdd2025/>.

4 Review Process

The review process for contributed material will be double-blind to reduce institutional and author bias, and the program will be selected with an eye to establishing broad coverage of research areas while preserving merit awarded by the double-blind reviewing process. We will also attract a wide and diverse program committee. Each submission will be reviewed by at least 3 reviewers. Decisions will be made in a transparent way by the organizers. To discourage the presentation of already finalized machine learning work, novelty and scope for discussion will be explicit aspects of the reviews. Conflicts of interest will be avoided by asking reviewers to state their conflicts.