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

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

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

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

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

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

Neil Shah
Snap
Seatle, 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.

## CCS CONCEPTS

• **Information systems → Data mining**.

*Primary contact.
†Also with TensorOpera, Inc.


*KDD '24, August 25–29, 2024, Barcelona, Spain*

ACM ISBN 979-8-4007-0490-1/24/08.
https://doi.org/10.1145/3637528.3671490

## KEYWORDS

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

**ACM Reference Format:**
Junyuan Hong, Carl Yang*, Zhuangdi Zhu, Zheng Xu, Nathalie Baracaldo, Neil Shah, Salman Avestimehr, and Jiayu Zhou. 2024. FedKDD: International Joint Workshop on Federated Learning for Data Mining and Graph Analytics. In *Proceedings of the 30th ACM SIGKDD Conference on Knowledge Discovery and Data Mining (KDD '24), August 25–29, 2024, Barcelona, Spain.* ACM, New York, NY, USA, 2 pages. https://doi.org/10.1145/3637528.3671490

## 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 heterogenous 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.

We envision a good amount of attendees (60+) and submissions (30+) to our workshop, coming from but not limited to several national and international research teams from both academia and industry. Based on our knowledge, well-known groups interested in the Federated Learning for data mining and graph (besides those of the organizers and invited speakers of this workshop) include and are not limited to

**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. 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 |
|---|---|
| 2:00 pm - 2:10 pm | Opening Ceremony |
| 2:10 pm - 2:50 am | Keynote Talk 1 |
| 2:50 pm - 3:30 pm | Keynote Talk 2 |
| 3:30 pm - 3:45 pm | Oral Presentation 1 |
| 3:45 pm - 4:00 pm | Oral Presentation 2 |
| 4:00 pm - 4:10 pm | Coffee Break |
| 4:10 pm - 4:40 pm | Poster Presentation |
| 4:40 am - 5:20 am | Keynote Talk 3 |
| 5:20 am - 5:50 pm | Panel Discussion |
| 5:50 pm - 6:00 pm | Award Ceremony and Closing Session |

## 3 SUBMISSION GUIDELINES

**Important Dates.** Paper submissions: June 11th. Paper notifications: June 26th. Workshop date: August 26th.

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/2024/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/.

## 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 2 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.

## ACKNOWLEDGEMENT

J. Zhou is supported by the National Science Foundation (NSF) under Grant IIS-2212174 and IIS-1749940.