FedEgo: Privacy-preserving Personalized Federated Graph Learning with Ego-graphs | ACM Transactions on Knowledge Discovery from Data

Introduction and Background

- **Graph Neural Networks (GNNs):** GNNs are effective in distilling information from graph data, crucial for tasks like node classification and link prediction. Traditional GNN works focused on centralized data, not considering data silo issues in real-world scenarios.
- **Federated Learning (FL) and Challenges:** FL is a technique that separates machine learning implementation from the need for direct data sharing, ideal for training models collaboratively while preserving data privacy. However, FL faces challenges due to the statistical heterogeneity among clients, leading to poor performance with naive federated algorithms like FedAvg.

FedEgo Framework

- **Overview:** FedEgo is designed to address these challenges using ego-graphs. It enables clients to train local models and contribute to a global model. The framework uses GraphSAGE over ego-graphs and **Mixup** for privacy concerns, integrating personalization into learning with an adaptive mixing coefficient strategy.
- **Ego-graphs:** These are subgraphs centered on a node with up to k-hop neighbors, providing a way to utilize structural information without compromising the original graph's privacy.
- **Personalization Approaches:** Various approaches for personalization in FL are highlighted. These include methods that use neural networks with base layers for federated averaging and personalization layers for individual customization. Other

strategies mentioned involve multi-task learning, mixing global and local models, and adding proximal terms for local fine-tuning.

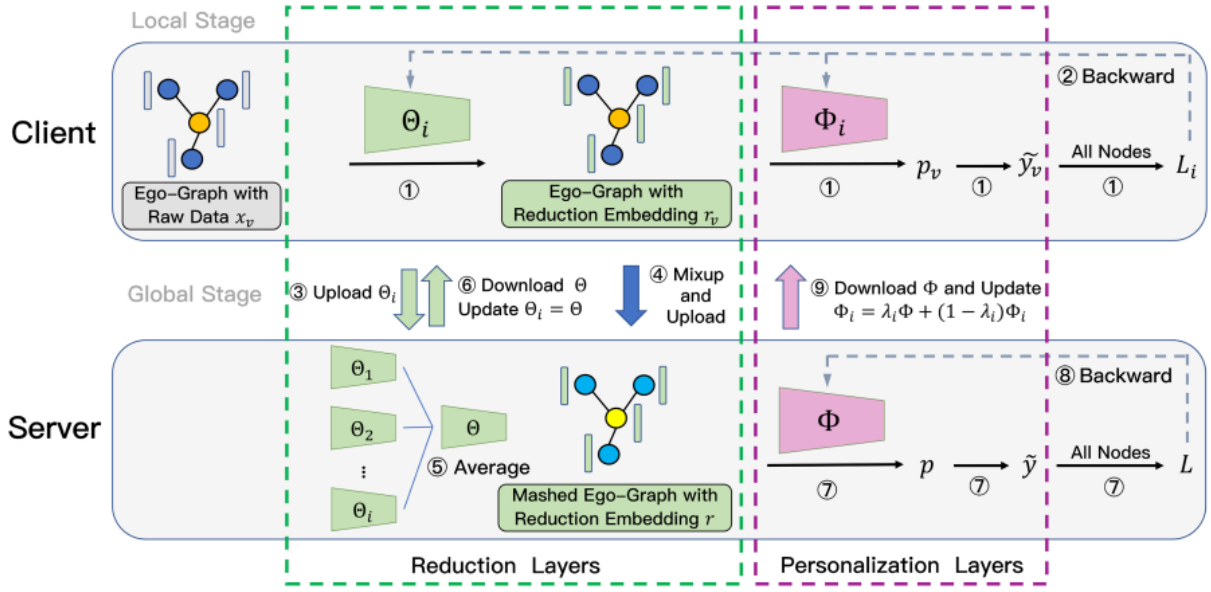
- **Adaptation to Local Data:** The paper emphasizes the need for models in federated learning to adapt to local data distributions. This is crucial due to the statistical heterogeneity often present in FL scenarios, where each client's data may follow a different distribution.

Key Contributions

1. **Privacy-Preserving Learning:** Uses ego-graphs to preserve data anonymity in terms of structure and features.
2. **Handling Non-IID Data:** Develops a global model that captures structural information within ego-graphs and addresses the non-IID issue.
3. **Personalization Strategy:** An adaptive approach to balance collaboration benefits and statistical heterogeneity disadvantages.
4. **Extensive Testing:** Demonstrates superior performance of FedEgo under non-IID scenarios through extensive experiments.

Related Work

- **Federated Learning on Graphs:** Various frameworks like GraphFL, D-FedGNN, and FedSage+ have been proposed, each with its approach to federated graph learning. However, FedEgo differentiates itself by focusing on privacy and personalization aspects.
- **Personalized Federated Learning:** The concept has been widely explored, with methods involving a mix of global and local models, multi-task learning, and local fine-tuning.



Framework Design

- **Dual Network Architecture:** FedEgo features different network architectures for clients and the central server. Clients have both reduction layers and personalization layers, while the server only has personalization layers.
- **Reduction Layers:** These layers at the client-side are designed to extract shared low-dimensional embeddings from local data, ensuring scalability and efficiency.
- **Personalization Layers:** These layers, present in both the clients and the server, are used for personalized graph mining, allowing the model to adapt to the specific data characteristics of each client.

Local and Global Stages

The training process in FedEgo is divided into local and global stages:

1. **Local Stage:**
    - Clients feed local ego-graphs into their reduction layers to obtain low-dimensional embeddings.
    - Mixup is applied over these ego-graphs to generate 'mashed ego-graphs' within each batch, enhancing data privacy.
    - The embeddings from the reduction layers are then fed into personalization layers for further processing.

- Each client calculates loss and updates parameters in both layers.

2. **Global Stage:**
   - Clients upload parameters from reduction layers and mashed ego-graphs to the server for collaborative learning.
   - The server aggregates these parameters using the FedAvg algorithm and updates the global personalization layers by training on the mashed ego-graphs.
   - All parameters are then sent back to clients, who update their personalization layers by mixing local and global weights.

Privacy Considerations

- **Anonymity in Data:** FedEgo maintains the anonymity of the data in terms of both structure and features. Ego-graphs ensure that only local topological information is extracted, preventing recovery of the original graph structure by the server or other clients.
- **Feature Anonymity:** The mashed ego-graphs contain only the mashed embeddings and local structure, which protects the transmission of raw data and ensures privacy.

Adaptive Mixing Coefficient

- Clients adaptively mix local and global model weights to achieve better personalization. The mixing coefficient is determined based on the difference between local and global data distributions, enabling each client to find an optimal balance for its unique data scenario.

Experimentation

- **Datasets and Settings:** Tested on datasets like Cora, Citeseer, CoraFull, Wiki, and FedDBLP. The experiments aimed to assess both personalization and generalization abilities of the model.

- **Comparison Methods:** Compared with methods like Local Only, FedAvg, FedProx, GraphFL, D-FedGNN, FedGCN, FedSage, and FedSage+.
- **Results:** FedEgo consistently outperformed other methods in terms of personalization and generalization abilities, particularly under non-IID scenarios.

Conclusion

FedEgo presents a significant advancement in federated graph learning, focusing on privacy preservation and addressing the non-IID nature of graph data. Its innovative use of ego-graphs and adaptive personalization strategy sets it apart from existing methods, making it effective for real-world applications where data privacy and diversity are critical concerns.