NANYANG TECHNOLOGICAL UNIVERSITY SINGAPORE

School of Computer Science and Engineering
College of Engineering

Efficient Multi-Objective P2P Federated Learning

Student: Pok Jin Hwee Supervisor: Prof. Anupam Chattopadhyay

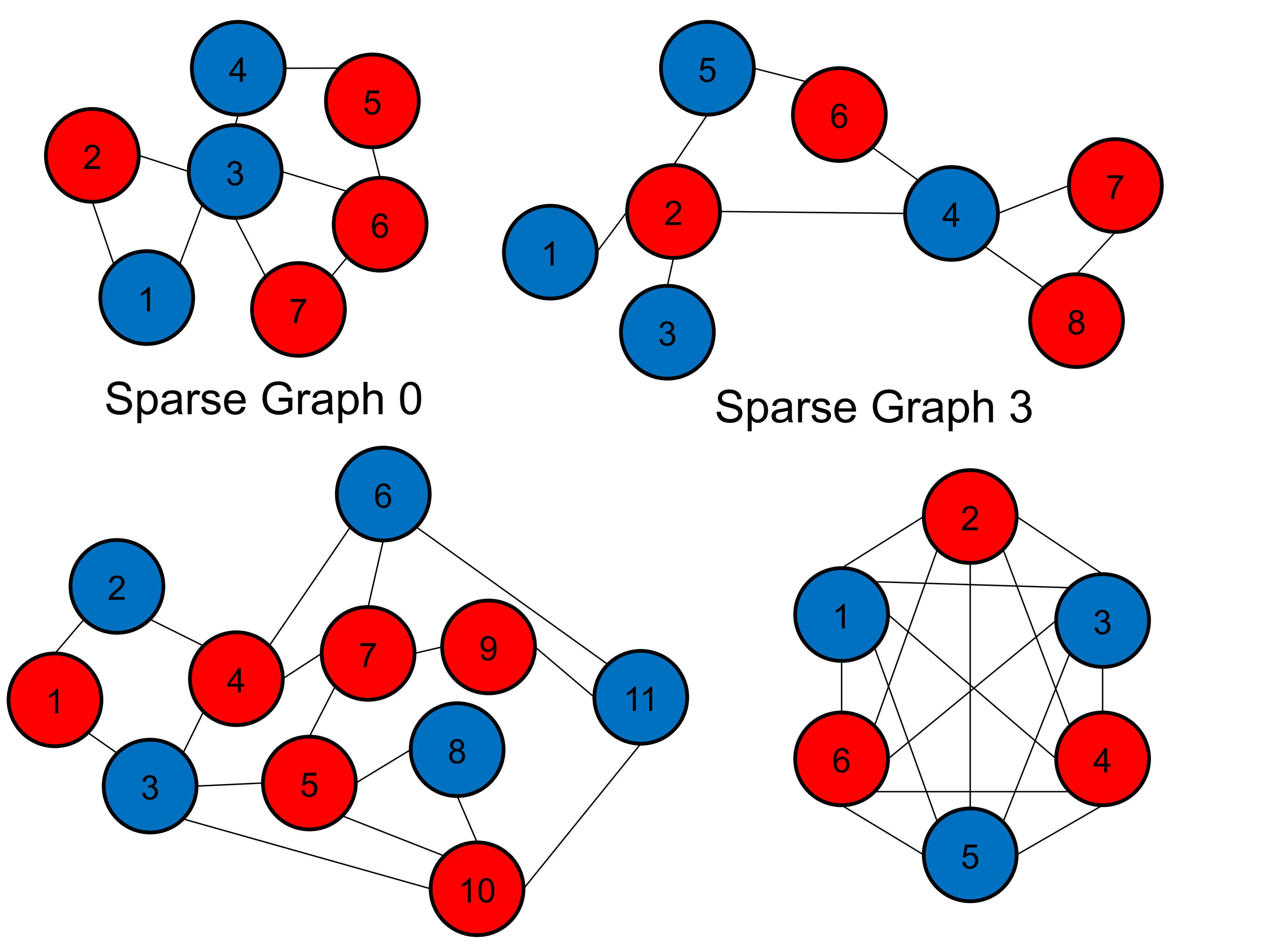
Project Objectives:

Training a Machine Learning (ML) model requires large amount of real-world data in order to obtain optimal performance and accuracy. The data that is collected for similar ML objectives may be used interchangeably. For example, in Natural Language Processing, the same dataset can be used for sentiment analysis and next word prediction. By sharing the data each device holds, each ML model will be exposed to a larger dataset, which in turns improve the model performance. However, sharing of data violates data privacy and protection laws. Therefore, in this project, a modified version of Floyd-Warshall algorithm is proposed and incorporated into Federated Learning.

A global adjacency matrix is computed at the start of the training process. This matrix contains the shortest path from source vertex i to destination vertex j and a Boolean variable that represents the similarity of objectives between each vertices. This is represented in the following graphical networks using colours. Vertices with the same colour denote that they have similar objectives.

Experiments and Results:

Graphs for Experiment 1:



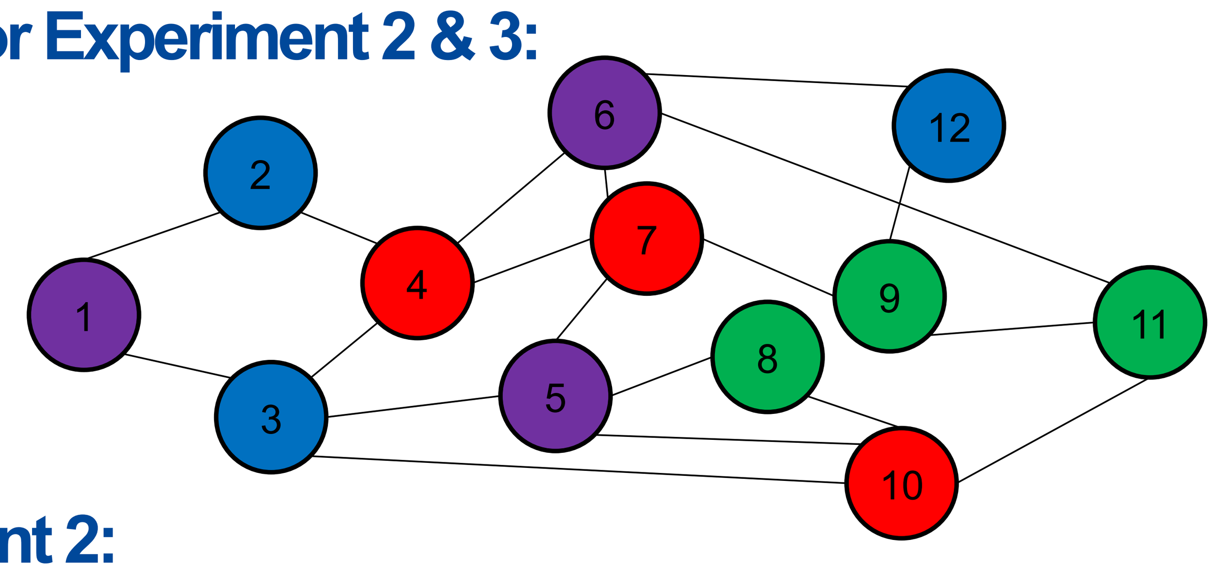
Sparse Graph 0

Sparse Graph 3

Densely Connected Graph 1

Fully Connected Graph 2

Graphs for Experiment 2 & 3:



Experiment 2:

Time Taken (s)	Objective 1	Objective 2	Objective 3	Objective 4	Sequential
Training	1842.591	2012.495	2854.703	1133.522	11553.00
Communication	2.084	2.100	2.085	2.091	23.743

For each of the objectives in the table above, training is done in parallel. As a result, the following is achieved:

1. Time taken for training to be completed is the longest duration required for Objective 3.
2. Parallel training achieves a speedup of 4.0473x over sequential training.

Experiment 3:

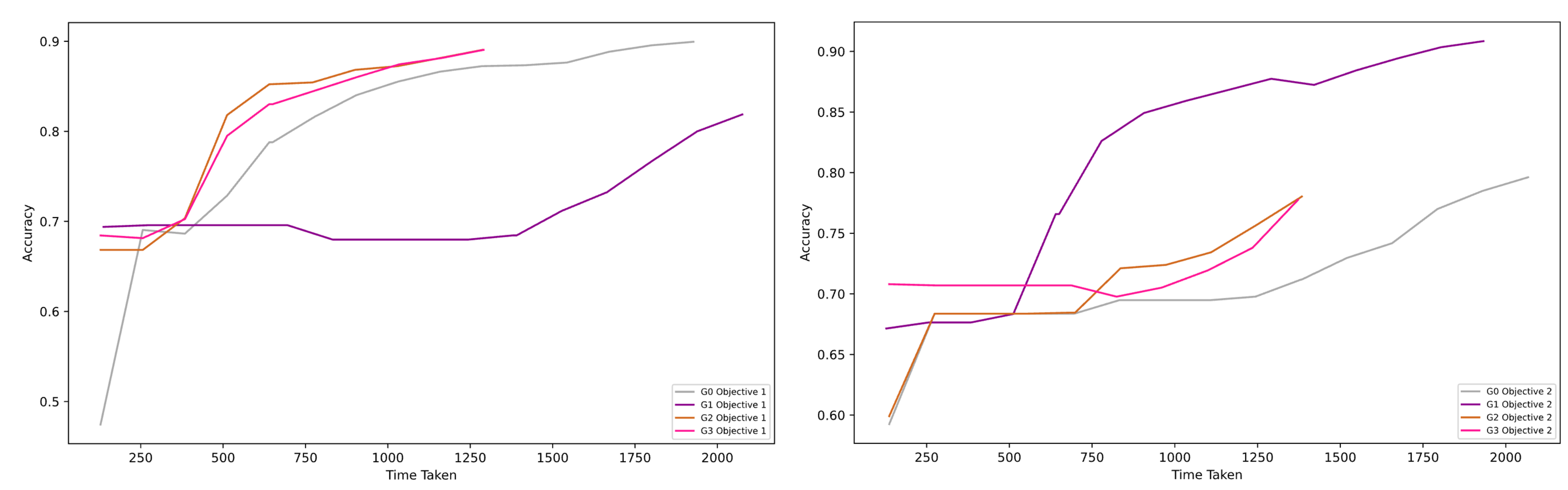
SRC → DEST	3 → 2	12 → 2	2 → 3	12 → 3	2 → 12	3 → 12
Training (s)	1838.001	1825.691	1836.556	1831.689	1836.243	1834.537
Communication (s)	4.142	6.193	4.130	8.192	6.146	8.162
Accuracy (%)	**96.386**	92.435	**86.964**	87.977	90.510	91.456

SRC → DEST	3 → 12 → 2	12 → 3 → 2	2 → 12 → 3	12 → 2 → 3	2 → 3 → 12	3 → 2 → 12
Training (s)	2755.668	2751.988	2760.300	2753.644	2755.807	2758.187
Communication (s)	14.140	12.156	14.107	11.139	12.212	9.554
Accuracy (%)	**95.947**	96.116	**84.431**	95.711	94.428	**96.150**

The blue vertices in the network is used to determine the necessity to involve all relevant vertices in the training process in Experiment 3. From the results of the experiment, the following statements can be concluded:

1. The ML model is not required to train on all relevant peers to achieve the best model performance.
   - The accuracy of the model after training on 2 out of 3 relevant vertices (**96.386%**) is the highest among all ML models.
2. The order of vertices on which the ML model trains on is important.
   - This is highlighted by the difference in accuracy obtained when training began on vertex 2 and on vertex 3.

Experiment 1:



For all vertices stored in the shortest path:

1. Training is not enforced on every vertices along the path
2. When vertices have similar objectives, the dataset it contains will be relevant to other vertices with similar objectives
3. By selectively identifying the vertices where training occurs, total amount of training time required is significantly reduced.

Conclusion:

From the results, we made the following conclusions:

1. The optimal model converges faster and less number of training epochs are needed to achieve high accuracies in fully connect graphs.
2. There is a significant speedup when training multiple objectives in parallel.
3. Not all relevant vertices are required to be involved in the training process to achieve a high performance ML model.
4. The set of vertices and permutation have a significant impact on the overall performance of the models as updates to model weights are not associative in nature.

References: Bonawitz, Keith, et al. "Towards federated learning at scale: System design." Proceedings of machine learning and systems 1 (2019): 374-388; Demetrescu, Camil, and Giuseppe F. Italiano. "A new approach to dynamic all pairs shortest paths." Journal of the ACM (JACM) 51.6 (2004): 968-992.