



## CS561 Spring 2023 - Research Project

**Title:** *Benchmarking Large Graph Processing Systems*

### **Background:**

A graph is a fundamental data structure that is widely used for modeling complicated data in many different application domains such as social networks, transportation networks, bibliographical networks, knowledge bases, and so on. With the advent of big data, large graphs with millions or billions of nodes and edges have become very common. Graph analytics is an important data discovery technique. Hence, many systems have been proposed to analyze large-scale graphs which is not a trivial task. In this work, we vision to (i) propose a classification among the popular graph processing systems based on their processing model, memory/storage requirement, and application requirements; (ii) conduct an in-depth study of the existing literature on benchmarks for graph data management; and (iii) experimentally evaluate the performance of popular large graph processing systems.

### **Objective:**

The goal of this project is to conduct an extensive survey on state-of-the-art graph processing systems and benchmarks to gain insight into different systems' applicability on different graph datasets. Some key objectives are:

- (a) Go over the current literature [1] in-depth about large graph processing benchmarks and systems.
- (b) Propose a taxonomy of the existing systems for future practitioners and researchers.
- (c) Some popular single-machine out-of-core graph processing systems are *GraphChi*, *Apache Giraph*, *GraphLab*, *X-Stream*, *TurboGraph*, and *GridGraph*. On the other hand, popular distributed shared-memory systems are Pregel, Trinity, GraphLab, GraphX and PowerGraph. Study the key design principles of these systems.
- (d) Select 7-8 popular systems (both single-machine and distributed) and experimentally evaluate their performance on several large datasets.
- (e) Gain insights from the experimental analysis to come up with new research challenges and direction.

**Responsible Mentor:** *Tarikul Islam Papon (papon@bu.edu)*

### **References:**

- [1] Liu, N., Li, Ds., Zhang, Ym. *et al.* Large-scale graph processing systems: a survey. *Front Inform Technol Electron Eng* **21**, 384–404 (2020). <https://doi.org/10.1631/FITEE.1900127>