Final Project Report

Guillaume Macquart de Terline

Network Science and Graph Learning

01 Janvier 2026

Contents

Abstract

This report includes the answers to my homework for the final project of the Network Science and Graph Learning course. The code used to perform the analysis is available in the GitHub repository: https://github.com/Gdeterline/Network-Science-and-Graph-Learning.

Introduction

Context

Network Science has emerged as a fundamental discipline for understanding the structure and dynamics of complex systems. From biological interactions and technological infrastructures to social relationships, graph representations allow us to model entities as nodes and their interactions as edges, revealing patterns that are not observable when analyzing individual components in isolation.

In particular, the study of online social networks such as Facebook has gained significant traction. These networks provide a rich testing ground for algorithms in Graph Learning, such as link prediction, label propagation, and community detection. Understanding the topology of social graphs helps in analyzing information diffusion, influence maximization, and the formation of social bubbles.

This project focuses on the analysis of real-world social network data using utilizing various metrics and algorithms from the field of Graph Theory and Machine Learning. By analyzing the Facebook 100 dataset, and applying various algorithms, we aim to uncover the underlying organizational principles of these university-based social communities.

Considered Dataset

The dataset analyzed in this project is the well-known **Facebook 100** dataset. This collection consists of complete friendship networks from 100 American colleges and universities as they existed in September 2005. The anonymized snapshot contains the connections among $n = 1,208,316$ users, and a total of $m = 93,969,074$ edges (unweighted and undirected) between them.

The data consists of 100 graph files in GML (Graph Modeling Language) format, located in the `data/` directory. This includes networks ranging from colleges (e.g., *Caltech* with 762 nodes) to larger universities (e.g., *UCLA* with 20466 nodes, etc.).

Each graph represents a single university's social network:

- **Nodes**: Represent individual persons (students, faculty, or staff).
- **Edges**: Represent "friendship" links between two individuals.
- **Attributes**: Nodes are typically annotated with metadata, including:
  - Person's Status (Undergraduate, Graduate, Summer Student, Faculty, Staff or Alumni)
  - Dormitory/House (if any)
  - Major (if any)
  - Gender (M or F)
  - Graduation Year

Question 1: Reading

The purpose of this question is to read and understand the following three papers:

- *Social structure of Facebook networks* by Traud, A. L., Mucha, P. J. & Porter, M. A. (Physica A: Statistical Mechanics and its Applications 391, 4165 – 4180, 2012). The main goal of this paper is to study the social structure of Facebook networks at 100 American colleges and universities at a single point in time. The authors investigate how user attributes such as gender, major, dormitory, and graduation year influence the formation of friendship links, specifically focusing on assortativity and community structure.

- *Comparing community structure to characteristics in online collegiate social networks* by Traud, A. L., Kelsic, E. D., Mucha, P. J. & Porter, M. A. (SIAM Review 53, 526–543, 2011). This paper explores the community structure within the Facebook networks of five US universities. The study quantifies the correlation between topological communities and metadata, highlighting how different demographic factors drive clustering at different institutions.

- *Assembling thefacebook: Using heterogeneity to understand online social network assembly* by Jacobs, A. Z., Way, S. F., Ugander, J. & Clauset, A. (Proceedings of the ACM Web Science Conference, WebSci '15, 18:1–18:10, 2015). This paper examines the evolution and assembly of the early Facebook network by analyzing the temporal sequence of user sign-ups and link formations. The authors propose a model that leverages node heterogeneity to explain the network's growth patterns and structural properties.

Question 2: Social Network Analysis with the Facebook100 Dataset

Let us consider three networks from the FB100 dataset: `Caltech` (with 762 nodes in the LCC), `MIT` (which has 6402 nodes in the LCC), and `Johns Hopkins` (which has 5157 nodes in the LCC).

The figures 1, 2, and 3 show the degree distributions for each of these three networks on a log-log scale.
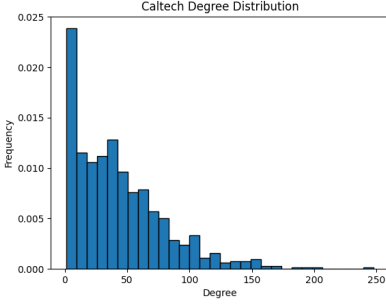

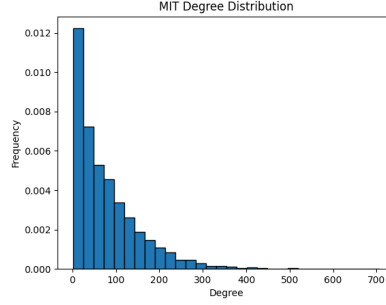Figure 1: Degree distribution for Caltech network

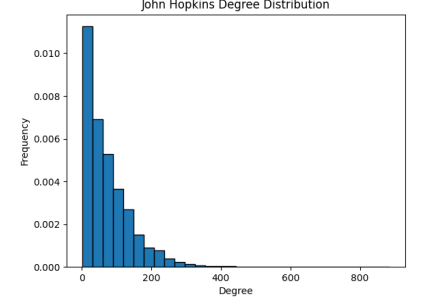Figure 2: Degree distribution for MIT network

Figure 3: Degree distribution for Johns Hopkins network

[ANALYSIS TO COMPLETE]

Given these three networks, we now seek to analyze their graph topologies. Thus, we seek to compute the global clustering coefficient and mean local clustering coefficient for each of the 3 networks. In addition, we compute the edge density of each network. The results are summarized in Table 1.

College/University	Global Clustering Coefficient	Mean Local Clustering Coefficient	Edge Density
Caltech	0.291283	0.409294	0.056404
MIT	0.180288	0.271219	0.012118
John Hopkins	0.193161	0.268393	0.013910

Table 1: Clustering Coefficients (Global and Mean Local) and Edge Density for Caltech, MIT, and Johns Hopkins networks

[ANALYSIS TO COMPLETE]

To assess the similarities and differences in local structures across these networks, we scatter plot the degree versus local clustering coefficient for each network node. The resulting plots are shown in Figures 4, 5, and 6.
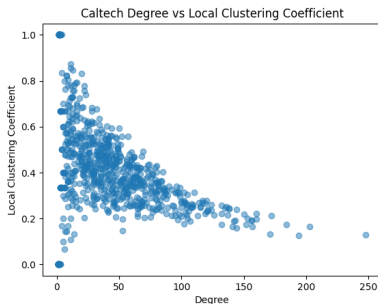

Figure 4: Degree vs Local Clustering Coefficient for Caltech network
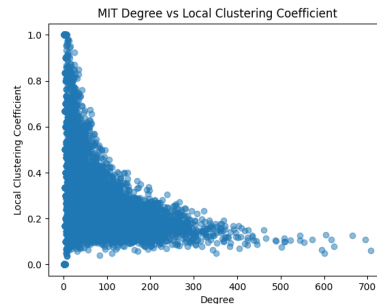

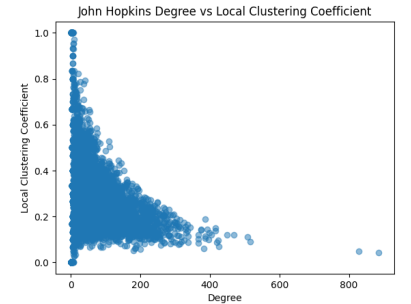Figure 5: Degree vs Local Clustering Coefficient for MIT network

Figure 6: Degree vs Local Clustering Coefficient for Johns Hopkins network

[ANALYSIS TO COMPLETE]

Question 3: Assortativity Analysis with the Facebook100 Dataset

In this section, we compute, then analyze the assortativity of all the 100 networks in the Facebook100 dataset.