College network

Course Assignment N.13

Paolo Deidda (paolo.deidda@usi.ch)
Andrea Luca Perugini (andrea.perugini@usi.ch)
https://github.com/USI-Projects-Collection/DA-College_network.git

April 17, 2025

Contents

Introduction

This document outlines the structure of the analysis on a dataset of private messages exchanged on a UC Irvine social network.

To run the code and reproduce the figures or outputs, please refer to the `README.md` file for setup and execution instructions.

1 Data Analysis

1.1 Setup and Data Loading

We utilized a dataset provided by the University of California, Irvine, comprising private messages exchanged among users within an online social network. The dataset includes 1899 unique nodes (users) and 59835 edges (messages). The data was loaded into a pandas DataFrame and preprocessed by converting timestamps from Unix time to readable datetime formats, facilitating temporal analysis.

1.2 Data Exploration

1.2.1 Basic Statistics

The dataset consists of 1899 unique users exchanging messages, resulting in 59,835 interactions. We computed essential statistics:

Total number of users (nodes): 1899

Total number of messages (edges): 59,835

Time span of data collection was determined by inspecting the minimum and maximum timestamps, indicating active periods of communication.

1.2.2 Message Frequency Over Time

To explore message distribution over time, messages were grouped and counted by day. Figure 1 clearly demonstrates peaks and troughs, possibly corresponding to academic activities or social events influencing user interactions. This analysis provides insights into temporal dynamics affecting messaging behavior within the network.

Figure 1: Daily Message Frequency

1.3 Network Construction

1.3.1 Degree Analysis

We analyzed node degrees to determine connectivity patterns within the network. Nodes with high degrees represent highly interactive users. The degree distribution shown in Figure 2 exhibits a typical scale-free network characteristic, with many users having few interactions and a small subset having numerous interactions, reflecting real-world social networks.

Figure 2: Degree Distribution

1.3.2 Identifying Isolated and One-Way Nodes

Further examination revealed isolated nodes (users who neither sent nor received messages) and one-way communication patterns (users who either only sent or only received messages). Identifying these nodes helps to understand user engagement and potential network fragmentation.

1.4 User Analysis

1.4.1 Top Senders and Receivers

We determined users with the highest message activity, distinguishing top senders and receivers. Understanding these roles highlights influential users within the network, possibly indicating user centrality or authority.

1.4.2 PageRank Centrality

We calculated PageRank scores to identify influential nodes based on network interactions. High PageRank scores correspond to users frequently contacted by others, emphasizing their significance within the social structure.

1.4.3 Edge Reciprocity

Edge reciprocity was analyzed to assess mutual interactions. High reciprocity indicates robust two-way communication, suggesting deeper social connections among users.

1.5 Temporal Analysis

1.5.1 Messages by Weekday

Analyzing message frequency by weekdays (Figure 3), we observed a clear pattern showing increased messaging during weekdays, which could be aligned with student schedules and reduced activity during weekends.

Figure 3: Messages by Weekday

1.5.2 Top User Activity Analysis

The hourly activity of top users was analyzed (Figure 4), revealing peak interaction times, which could indicate user availability or preferred communication periods. Such insights can guide future studies on user behavior or network usage patterns.

Figure 4: Hourly Activity of Top Users

2   Netwrok Analysis

3   Final Considerations