

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

1	Data Analysis	2
1.1	Setup and Data Loading	2
1.2	Data Exploration	2
1.3	Message Volume Over Time	2
1.4	Temporal Analysis	2
1.5	Top Users vs Least Users Analysis	3

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

To initiate the analysis, it was essential to first examine the structure of the dataset. The dataset contains private messages exchanged on an online social network at UC Irvine. Each entry represents a message characterized by the following attributes:

- SRC: Identifier of the sender
- TGT: Identifier of the receiver
- UNIXTS: Timestamp of the message in Unix time format

Given the non-human-readable format of UNIX timestamps, these were converted into datetime objects. To ensure temporal accuracy in the context of UC Irvine, all timestamps were further adjusted from UTC to the Pacific Time Zone (America/Los_Angeles), accounting for daylight saving time.

1.2 Data Exploration

As an initial step, the following questions were posed: *"How large is the dataset? How many users are involved and over what time span does the activity occur?"*

- Total number of messages: 59,835
- Total number of unique users: 1,899
- Temporal coverage: From April 15, 2004 to October 26, 2004

These findings confirmed that the dataset encompasses over six months of communication among nearly two thousand participants.

1.3 Message Volume Over Time

To identify communication trends, the number of messages exchanged per day was aggregated. As shown in Figure 1, notable fluctuations were observed, with certain peaks that may coincide with academic milestones or periods of heightened social activity.

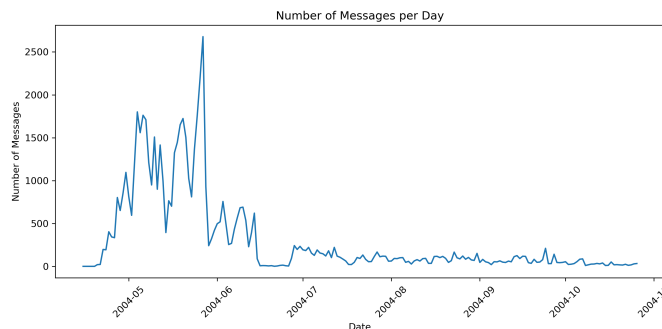


Figure 1: Daily message volume over time

1.4 Temporal Analysis

During the preliminary analysis of hourly messaging patterns, an anomaly was noted: a substantial number of messages appeared to be sent between 5:00 and 8:00 a.m. UTC—an unusual timeframe for student activity.

This observation suggested a potential misalignment between recorded and actual local times. Considering the dataset originates from UC Irvine, the hypothesis emerged that timestamps required conversion to Pacific Time. Upon applying this correction, the messaging peak appropriately shifted to between 10:00 a.m. and 1:00 p.m.

To develop a comprehensive understanding of temporal behavior, message frequencies were further analyzed across hours of the day, weekdays, and months. The resulting trends are presented collectively in Figure 2.

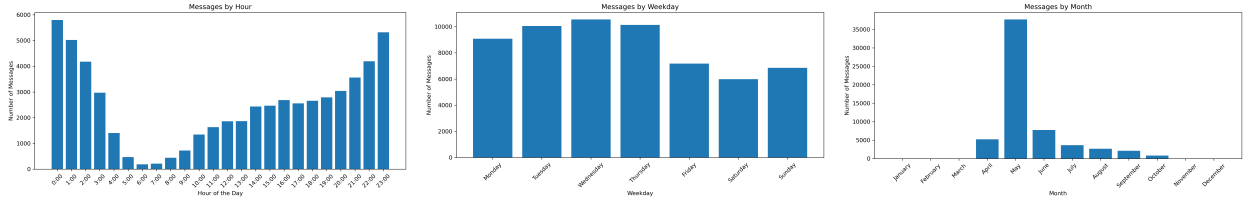


Figure 2: Temporal breakdown of message activity by hour, weekday, and month (adjusted to local time)

The analysis reveals heightened activity during mid-morning to early afternoon, reduced interaction over weekends, and seasonal variations, with communication peaking in spring and early autumn—likely aligned with the academic calendar.

1.5 Top Users vs Least Users Analysis

Following the identification of User 234 as one of the most active participants, a question arose:

”Do highly central users demonstrate distinct behavioral patterns compared to less active ones?”

To investigate this, PageRank scores were computed to identify the most and least central users. Subsequently, their hourly messaging distributions were analyzed.

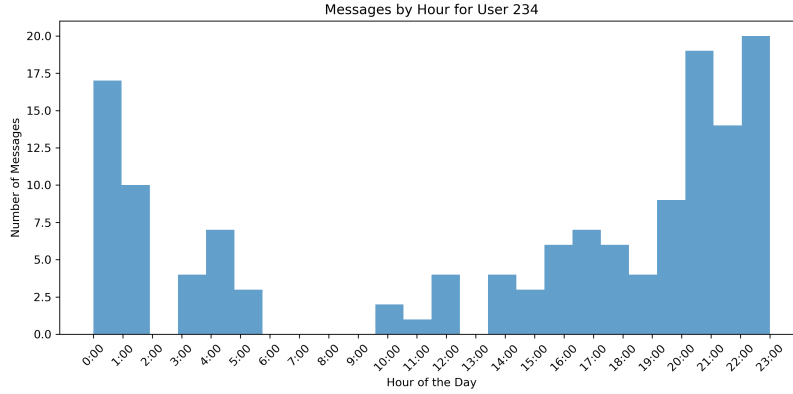


Figure 3: Hourly message distribution for User 234

As illustrated in Figure 3, User 234 maintained a consistent level of activity throughout the day, with noticeable intensity in the evening—potentially indicating a central role within the network.

To broaden the perspective, the average hourly activity of the top 10 users (by PageRank) was compared against that of the bottom 1500 users.

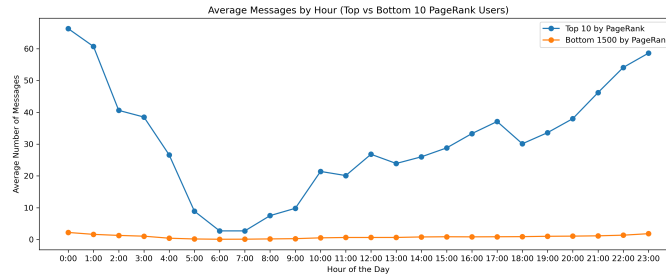


Figure 4: Comparison of average hourly activity between top and bottom PageRank users

As shown in Figure 4, top-ranked users exhibited a more consistent and structured communication pattern, particularly active during typical daytime hours. In contrast, lower-ranked users displayed sporadic and irregular messaging behavior, further reinforcing the relevance of centrality in understanding user roles within the network.