ECS 273 Final Project

Professor:	**Kwan-Liu Ma**
TA:	**Yun-Hsin Kuo**
Working Group:	**Musheng He**
	Yifeng Shi

ECS 273 Visual Analytics
Winter Quarter 2023
Course Final Project

Contents

0 Source Code

TBD

1 Motivation

Social network analysis is recently commonly seen in social research and is important for particular purposes and even plays a crucial role in these areas[4]. The egocentric network represents the network relationship between a particular individual, who we called ego, and the people associated with it[2]. Visualizations of the ego network can be of interest not only to analysts and data experts but also to the people whose data is being analyzed[7].

a) providing users with people they may know. To expand the abilities of the social media application and provide users with more intelligent functionalities helping them expand their social circles. Recommending people they might know or met in real life is a fundamental but critical one.

b) Criminal investigation experts analyze the social network branding of potential criminals or known criminals to analyze possible criminal or illegal activities. By analyzing the relationships between a suspicious criminal or known criminal, experts are enabled to find out the potential partners and target these potential criminals or are likely to generate constructive conclusions by analyzing the pattern of relationships among them.

In this Project, we construct a social network visual analytics based on the dataset $ego - Facebook$ provided by Stanford University Large Dataset Collection [6].

2 Objectives

a) **General overview of one's friends.** One can view all his/her direct friends and group friends using personal information like hobbies, language, etc. Also, provides statistical information corresponding to the values of chosen tag.

b) **Discovery of people one may know.** The ego user could find more people they may know who are more socially distant. Experts, like data analysts, and criminal investigators, could reason potential relevant users to the target user. Ideally speaking, we could provide the automated using advanced machine learning models to predict and recommend people the ego may know, but it means that more work and effort is needed to invest into the project. Considering the time and resources limitation, we leave it as our future work.

3 Challenges

3.1 Complexity of data

The relationships between users are represented by a pair of node indices $\{index_{node1}, index_{node2}\}$, and our procedure needs to Restore these complex relationships by parsing all the relationship pairs. Furthermore, in the dataset we are going to use, the values of a certain attribute or tag are listed one by one, like:

Age	1
Age	2
Age	3
Last name	A
Last name	B
Last name	C

And the features of a particular user are represented as:

155	0	0	1	0	1	0
369	1	0	0	0	0	1

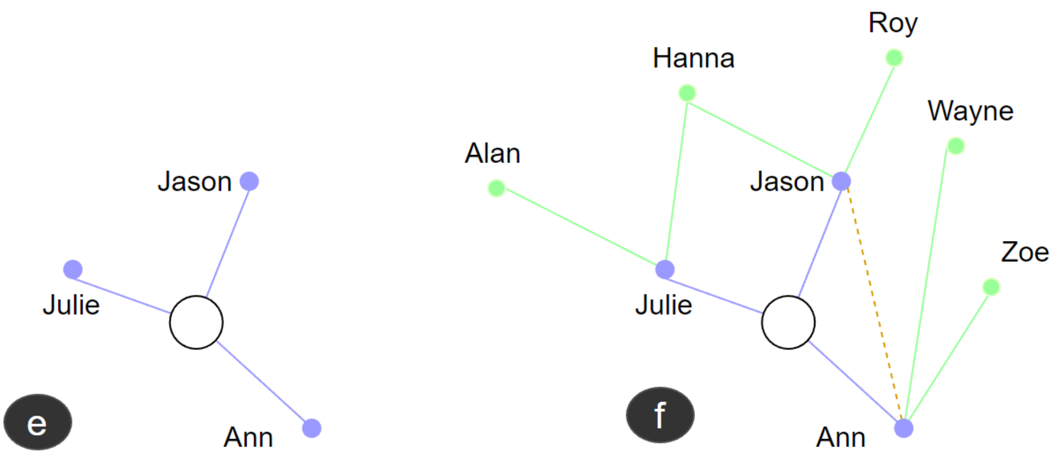

Figure 1: Framework

3.2 The complexity of Visualization

Especially for the k-hop graph we provide for users to find people the ego may know. When $k = 1$, all nodes in the graph are directly connected with the ego node, as shown in Figure 1 e. However, when we increase the value of k to 2, more people will be shown in the graph, as shown in Figure 1 f. Here are two things we need to notice, 1) the node "Hanna" is known by both "Julie" and "Jason" in the inner layer, so "Hanna" is more likely to be the one the ego node may know offline and we need to express this significant relationship; 2) Look at the dotted edge in color brown, although it presents the second-layer relationship, but since the ego node has known node "Jason" and "Ann", so it is a useless edge and will lead to vision clutter, and thus we need to filter out such edges from the data filled into the graph. Furthermore, it is obvious to see the complexity of visualization increases dramatically with the increment of variable k.

Based on the above example and explanation, it is not hard to reason what will happen if there are many features and some features may have tons of values.

4 Background

Analyzing social networks enable analysts to learn the relationships between different nodes within the network and inspire insightful thoughts[3]. To study the whole network, the analyst needs to gather the data of all the nodes present in that network while to learn personal networks or egocentric networks, the analysts need to consider the tied nodes only. In this project, we focus on the visual representation and interaction of $personalnetworks$, a type of egocentric network consisting of family, friends, and acquaintances surrounding a focal person.

In this project, we introduced two conventional definitions of the subjects we describe:

a) **Ego** represents the particular user we are going to analyze the relationship with.

b) **Alters** all connected friends.

5 Related Work

5.1 Ego Network *vs* Whole Network

Common tasks of Social Network Analysis involve the identification of the most influential, prestigious, or central actors, using statistical measures; the identification of hubs and authorities, using link analysis algorithms; discovering communities, using community detection techniques, and how information propagates in the network, using diffusion algorithms[5].

Whilst social network analysis puts more emphasis on the key features of the whole network (e.g., topology, centrality, ...), personal network analysis focuses on the relevant features of the ego's social relationships. Although the analysis of ego networks usually doesn't pay too much attention to relationships between alters and is thus unable to provide a complete analysis of the social network of users, it typically studies in greater detail the properties of the individual links between ego and alters[1].

5.2 Analysis of Egocentric Network

Many researchers have attempted to analyze the structural properties of ego networks, Their results suggest that there is a constrained relationship between the absolute number of individuals that the ego can preserve in the network and the emotional intensity of the relationship that the ego can preserve.[2].

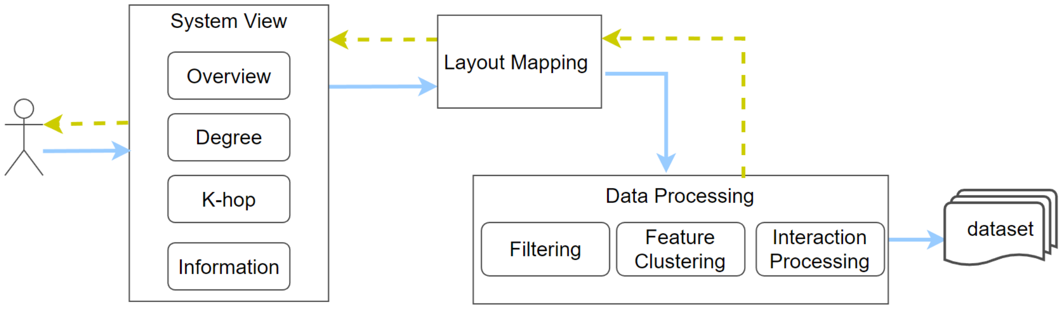

Figure 2: Framework

6 Overview

From the system to users, the back-end server will load documents in the dataset and suing the Data Processing module to filter, trim, and cluster data. The Layout Mapping module is responsible for mapping data received from the Data Processing module and project data to fit the requirement of the front-end side to display.

From users to the system, users' operations will generate requests to the back-end server and eventually be processed by the Data Processing module to satisfy users' requests, data will be returned to the Layout Mapping module to process and thus facilitate the display of the front-end side.

6.1 Interface Design

As we can see in Figure 3, there are four parts provided in this project. *a* provides an overview of ego's social network; *b* shows the average degrees of each tag value corresponding to the tag selected in Figure 3 *a*; *c* presents a customized function enabling users to find people that they may know; The tag information of the user selected in Figure 3 *c* will be shown in *d*.

7 Visualization

Generally speaking, we can separate the interface into two parts: *Overview* and *Discovery*. The *Overview* part consists of Figure 3 *a* and 3 *b*, and the latter is composed of Figure 3 *c* and 3 *d*.

7.1 Overview Part

In Figure 3 *a*, we provide a force-directed graph to display the relationships among the ego and all directly connected friends and also relationships among these directly connected friends. A select box with values of personal information items is provided for users to choose from. After choosing a tag, all directly connected friends will be grouped into several clusters corresponding to the values of the chosen tag.

Furthermore, when choosing a particular tag, the average degree of each cluster, which represents each value of the selected tag, will be shown in Figure 3 *b*, and the number of

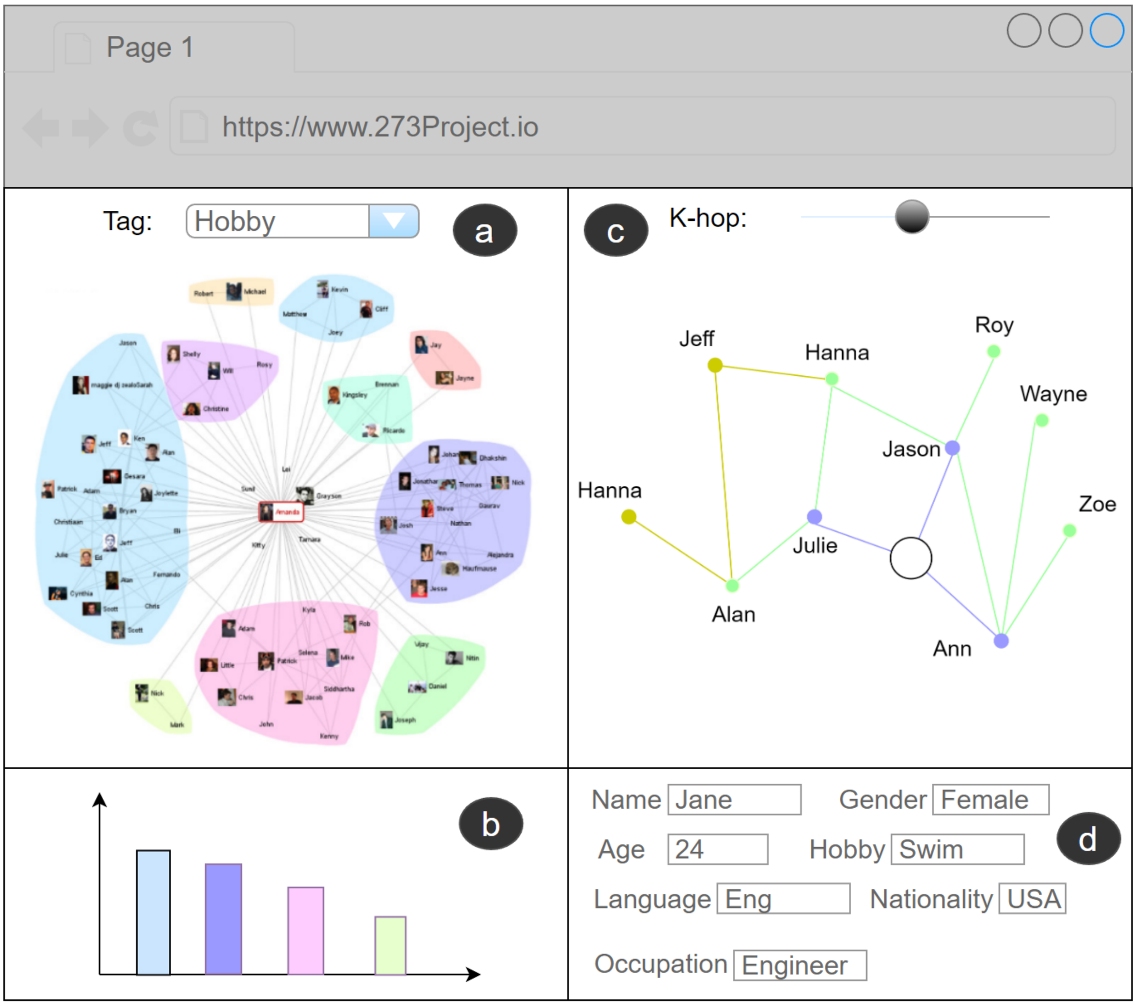

Figure 3: Interface Design

friends in each cluster will also be displayed in this figure.

7.2 Discovery Part

For the discovery part, users can discover k-hop(In this project,k=3), distance users who might be potential friends of the ego. After showing k-hop users, the current operator can view the personal information by clicking a node representing a person, and get the information of the person shown in Figure 3 *d*.

8 Analysis Method

For the ego user or an expert, firstly, one can look at the overview of known friends, and choose an interested tag, let's say the hobby, to cluster all friends into several groups based on their hobbies. By checking friends in the same group, one can notice that several friends share the same hobby, and realized that if one is also interested in a particular hobby, one can interact with friends in that group and hence increase their intimacy. Below the overview graph, a histogram graph showing the average degree and the number of people in each group can help the operator get the whole picture of friend circles.

Increasing the k-hop value by dragging the scroll bar, more socially distant users will be shown in the right-top figure to help operators find out potential friends of the ego user. Clicking the node shown in this figure, the personal information will be displayed helping the operator to check the specific information and determine whether the user is one known in real life.

9 Expected Results

a) Helping the operator analyze the composition of the ego's friends and facilitate finding the patterns among friends, and between the ego and friends.

b) Facilitating operators to view and analyze users further away socially and find people that ego may know.

10 Future Work

a) **Heterogeneous social network.** In this project, all nodes we analyzed represent people. In fact, there are many types of nodes, including organizations, events, locations, etc. And it would be a powerful approach if we could connect people by these heterogeneous nodes, like two people who attended the same academic lecture might know each other; A person who has been to the same place as the suspect at the same time may be the potential suspect as well.

b) **Temporal data.** If we have the temporal dimension, we could provide the evolution of users' relationships over time, which is believed that more useful and capable of inspiring insightful discoveries and ideas.

c) **Automatically predict and recommend people users may know.** Due to the time and resource constraints of this project, we only provide a limited ability to recommend people that users may know, but in fact, we can use cutting-edge machine learning and deep learning models to recommend people that users may know, help users expand their circle of friends or provide new discoveries to analysts.

References

[1] Valerio Arnaboldi, Andrea Guazzini, and Andrea Passarella. Egocentric online social networks: Analysis of key features and prediction of tie strength in facebook. *Computer Communications*, 36(10-11):1130–1144, 2013.

[2] Qiang Lu, Jing Huang, Yifan Ge, Dajiu Wen, Bin Chen, and Ye Yu. Egovis: A visual analysis system for social networks based on egocentric research. *International Journal of Cooperative Information Systems*, 29(01n02):1930003, 2020.

[3] Sadia Majeed, Muhammad Uzair, Usman Qamar, and Aftab Farooq. Social network analysis visualization tools: A comparative review. In *2020 IEEE 23rd International Multitopic Conference (INMIC)*, pages 1–6. IEEE, 2020.

[4] Christopher McCarty, José Luis Molina, Claudia Aguilar, and Laura Rota. A comparison of social network mapping and personal network visualization. *Field methods*, 19(2):145–162, 2007.

[5] Shazia Tabassum, Fabiola SF Pereira, Sofia Fernandes, and João Gama. Social network analysis: An overview. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, 8(5):e1256, 2018.

[6] Stanford University. Stanford large network dataset collection.

[7] Fernanda B Viégas and Judith Donath. Social network visualization: Can we go beyond the graph. In *Workshop on social networks, CSCW*, volume 4, pages 6–10, 2004.