Tensor Factorisation for Group Detection

Ange Kokotakis, Romain Ramez
directed by Rodrigo Cabral Farias

November 2023

Abstract

The aim of this project is to detect groups of people with a given dataset of interactions in a community. For this purpose we will represent the dataset as a tensor and use non-negative tensor factorization to approximate it as a product of different matrix from which we can get information about groups in this community.

1 Introduction

In this project, we use datasets similar to Figure 1 in which each line represents an interaction between two individuals: $id1$ and $id2$ at a certain $time$.

time	id1	id2
31220	58	63
31220	59	64
31220	63	66
31220	85	190
31220	85	214
31220	102	115
31220	191	199
31220	191	214
31240	58	63
31240	63	66
31240	85	190
31240	85	214
31240	102	115
31240	143	192
31240	188	194
31240	191	199

Figure 1: dataset

From this dataset we create matrices $X \in \{0,1\}^{I \times I}$ which represent interactions between people during a certain interval of time where I is equal to the number of people in the community and each coefficient is equal to :

$$X_{i,j} = \begin{cases} 1 & \text{if there is an interaction between person } i \text{ and } j \text{ during a given interval of time} \\ 0 & \text{otherwise} \end{cases}$$

And then we stack these matrices with different interval t_k of time to create our tensor $Y \in \{0,1\}^{I \times I \times K}$ where K is the number of intervals of time. Finally the tensor containing the dataset has its coefficients equal to :

$$Y_{i,j,k} = \begin{cases} 1 & \text{if there is an interaction between person } i \text{ and } j \text{ at interval of time } t_k \\ 0 & \text{otherwise} \end{cases}$$

2 Model

In order to approximate this tensor we create 3 matrices :

$$U \in \mathbb{R}_+^{I \times R},\ V \in \mathbb{R}_+^{I \times R},\ W \in \mathbb{R}_+^{K \times R}$$

where :

- I is the number of individuals in the community,
- K is the number of interval of time,
- R is the number of groups in the community (we choose one randomly at first we will see later how to choose it correctly).

The U and V matrices represent the membership level of a person to a certain group and W represents in which interval of time a group has been active.

We define $S \in \mathbb{R}_+^{I \times I \times K}$