Week 5

☑ Examine in which Repo emoji appears first

☑ Sort out the vector for the repo

☑ Examine if we can set up a graph for github users, and check the corresponding relations with that Graph

Determine to also use count each as the vector distribution

Error!!

Can not read value at 0 in block -1 in file

Sort out the vector for the Repo

Solution:

1. First determine the maximum length, them perform the vector based on [0, max[]
2. Then perform the vector based on classification

A very interesting result? First appear true position is

But the average length here, is:

5.448427973110346
5.0

(Need to figureout why!!)

median: 0.15899653356065582
mean: 0.0

Data Exploration: with Issueid:

Possible Problem (also exists in previous, as caution):

In previous calculation, for every issueid, there is possible action of edit/created and ... so there may involve repetitive calculation. In every post you should check the latest event/ instead of the

根据conversation的数量进行过滤，然后random sample

长度80%在五条以上的repo，分析的Repo所有的conversation的

算conversation长度，rankking CDF 80% conversation 100个conversation, 80%的 conversation>=5%, 80%