Todo list

1. Count

☑ Count emoji variance per repo (groupby rid, count variance, return float data)

☑ Count post/issue/comment emoji 占比

☑ Count Repo Type Distribution (Like PR event distribution)

2. Plot

☑ Plot total events 占比

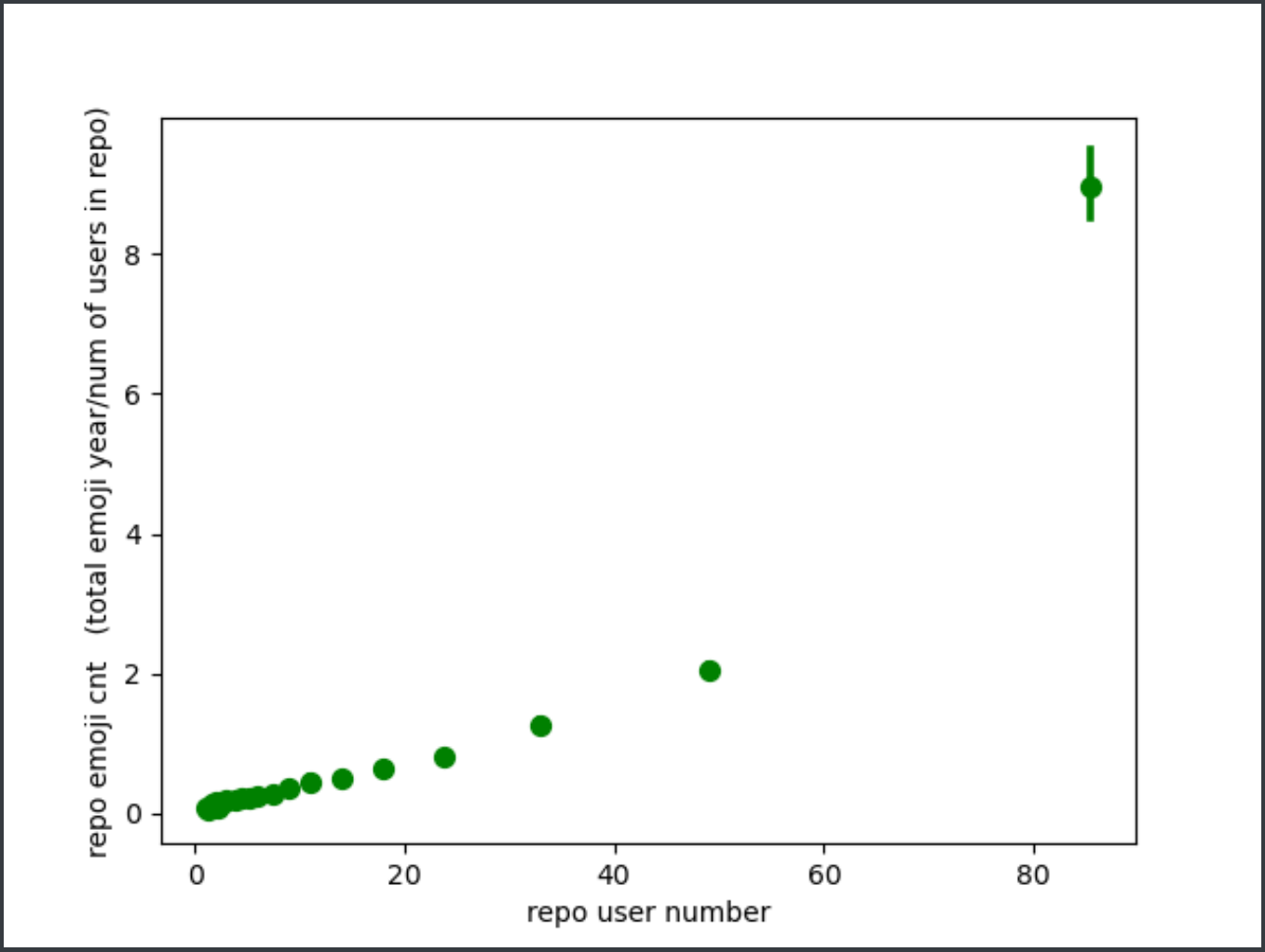
**

avg calculated by repo total/number of users

X: average events of developers in one repo : Y average number emoji usage in Posts of developers in a repo

the median user events of repo: 5.827586206896552

the median user emojis of one repo 0.03225806451612903



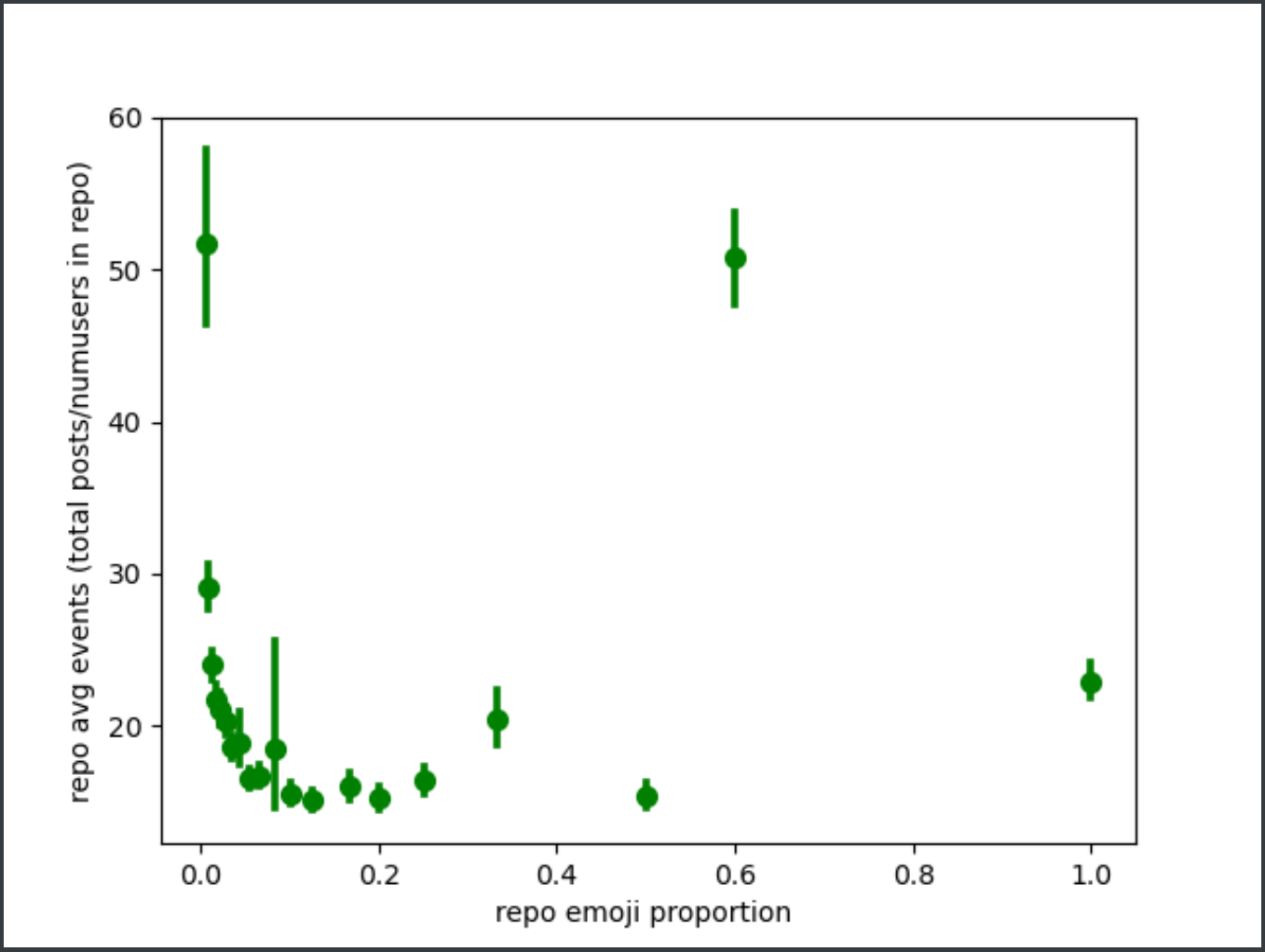
(Correction: typo here, X should be repo total events)

This shows a rather linear relationship between the repo total events and repo

X: emoji posts proportion in one repo: Y: average events of developers in one repo

the median user events of repo: 6.0

the median X of one repo 0.07407407407407407



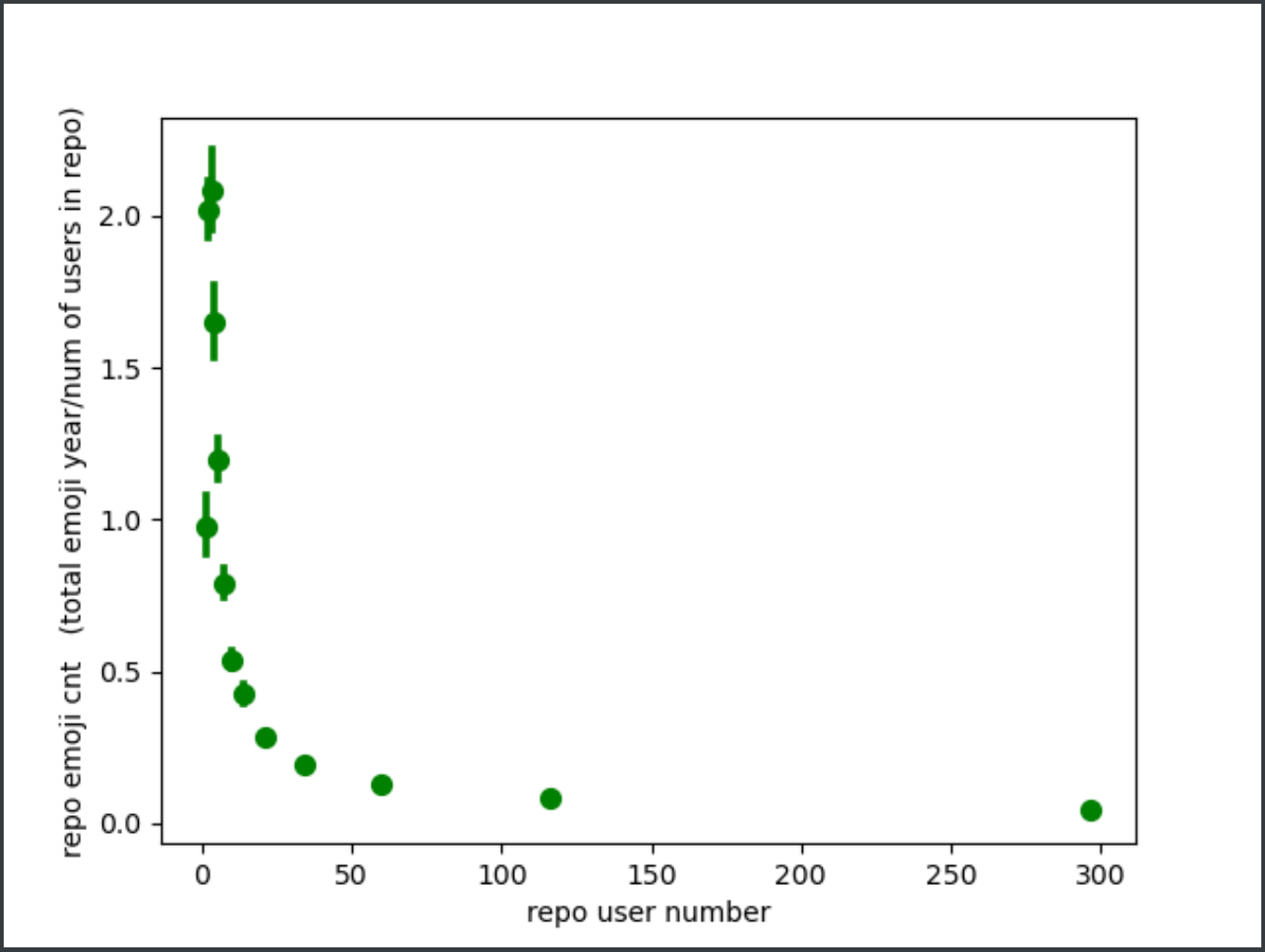
This shows when emoji used ratio is low (normally it should be low as the median value is 0.07, and we can look at the 80%),

when the repo tends to use the emoji more, the average user work events is lower.

X: Number of Users, Y: repo average emoji usage

此处number of user计算为countDistinct

he median number of users repo: 3.0
the median user emojis of one repo 0.032258064516129032



This shows the repo user number increases, the repo average emoji cnt will decrease (it is very likely,)

X轴缩小，把别的metric和size加起来分析

calculate user with events includes conmmunication posts

issues communication

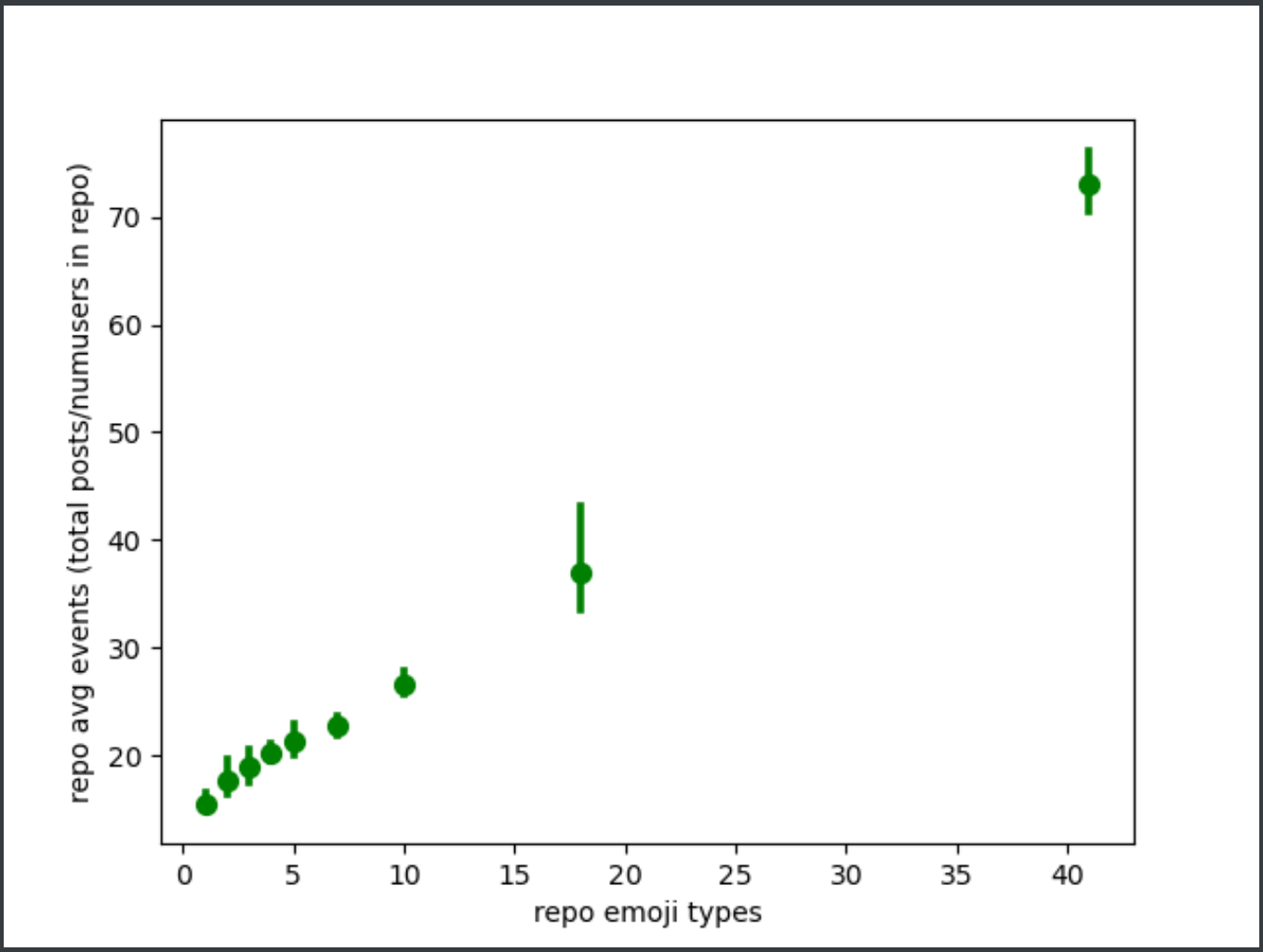
进一步缩小到pr, review comment,

team emoji

X: Repo used emoji types, Y: repo average events

调过来

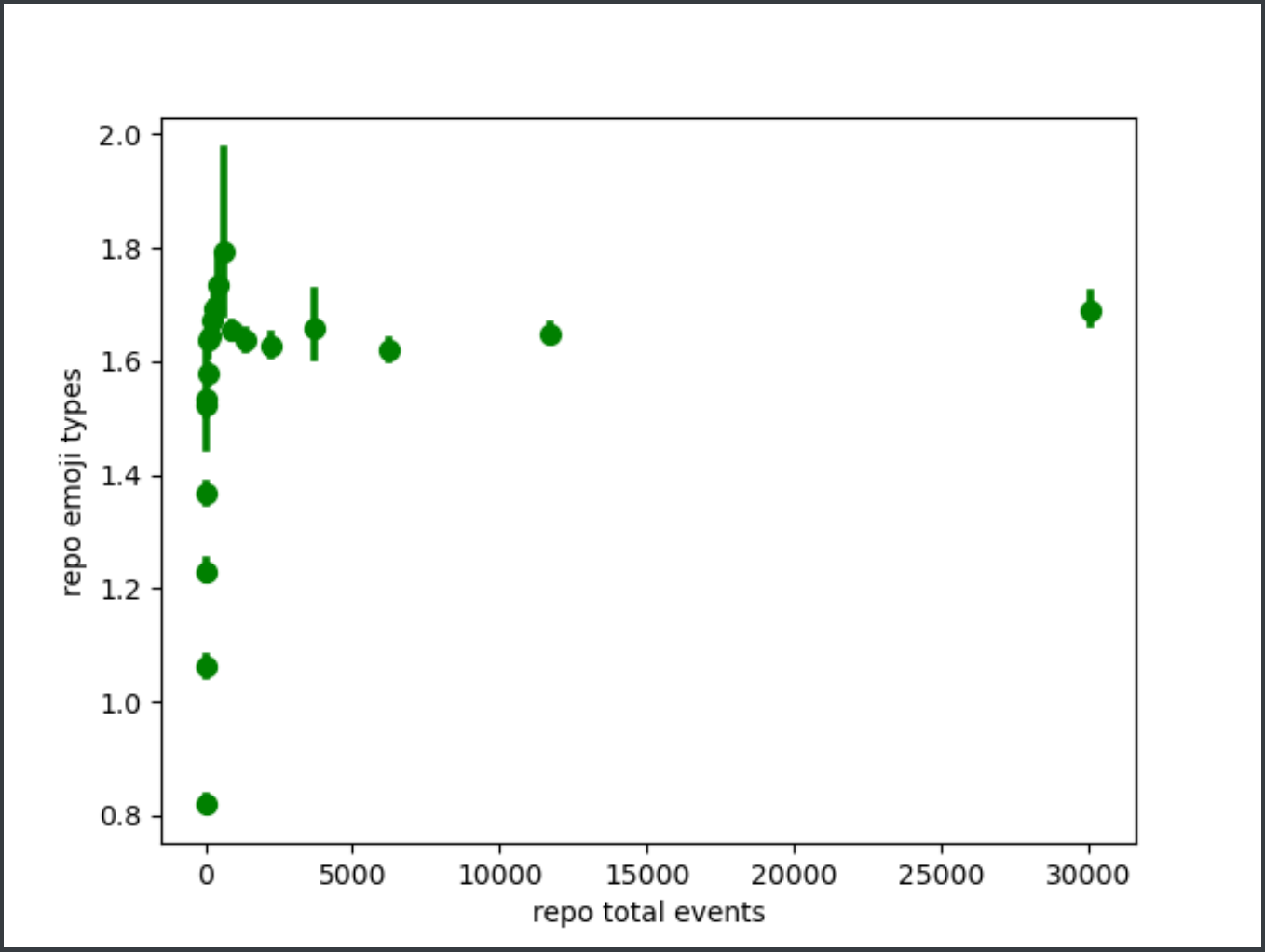
the median user events of repo: 6.0
the median user emojis of one repo 2.0



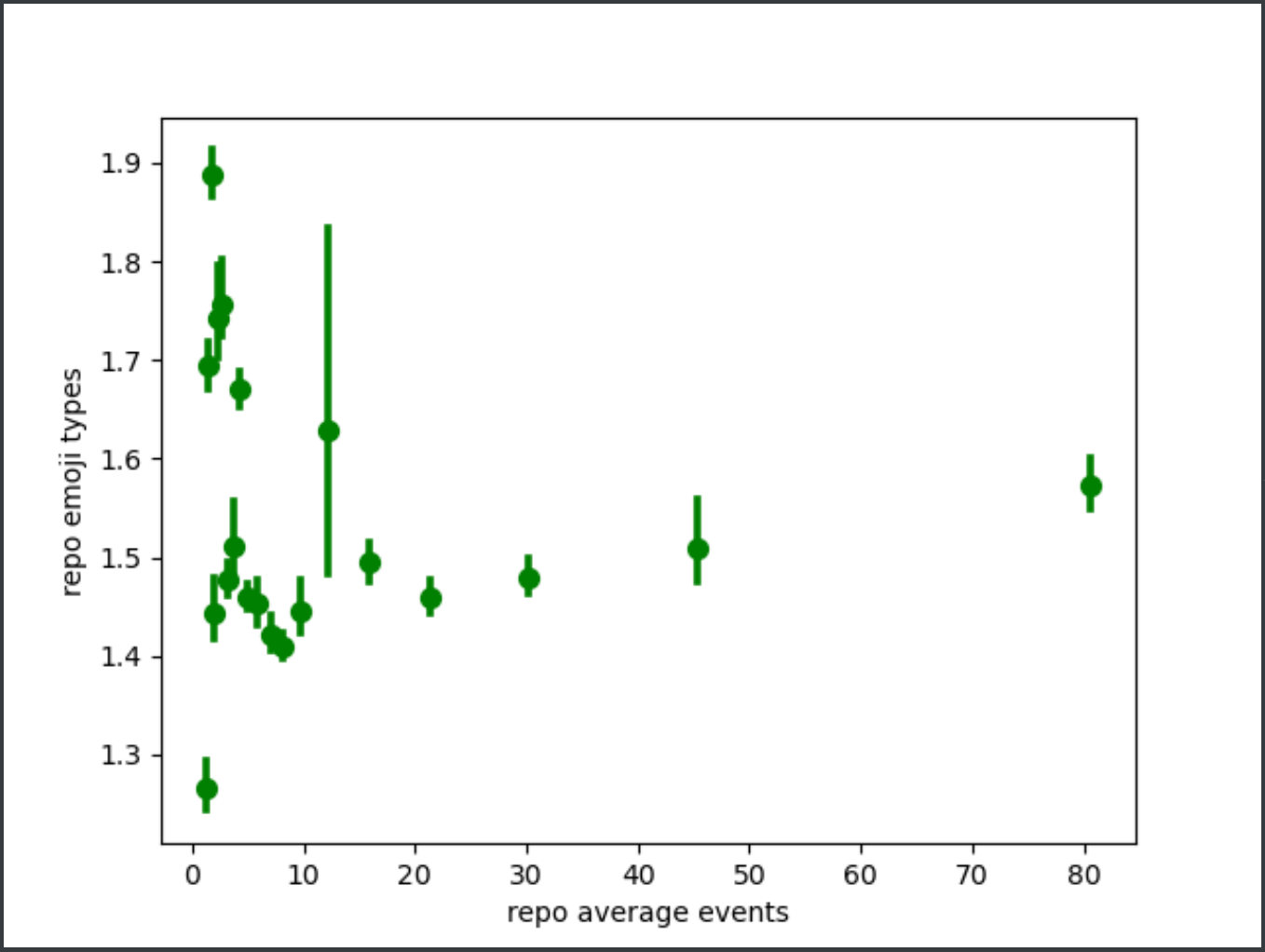
We can see there is a rather postive correlation between the repo emoji types & repo average events.

New sec

Y: Repo User emoji type (averaged by first calculating the_), X: repo total events



X: repo average events Y: repo emoji



based on median value of average events

assumption: repo with more issues

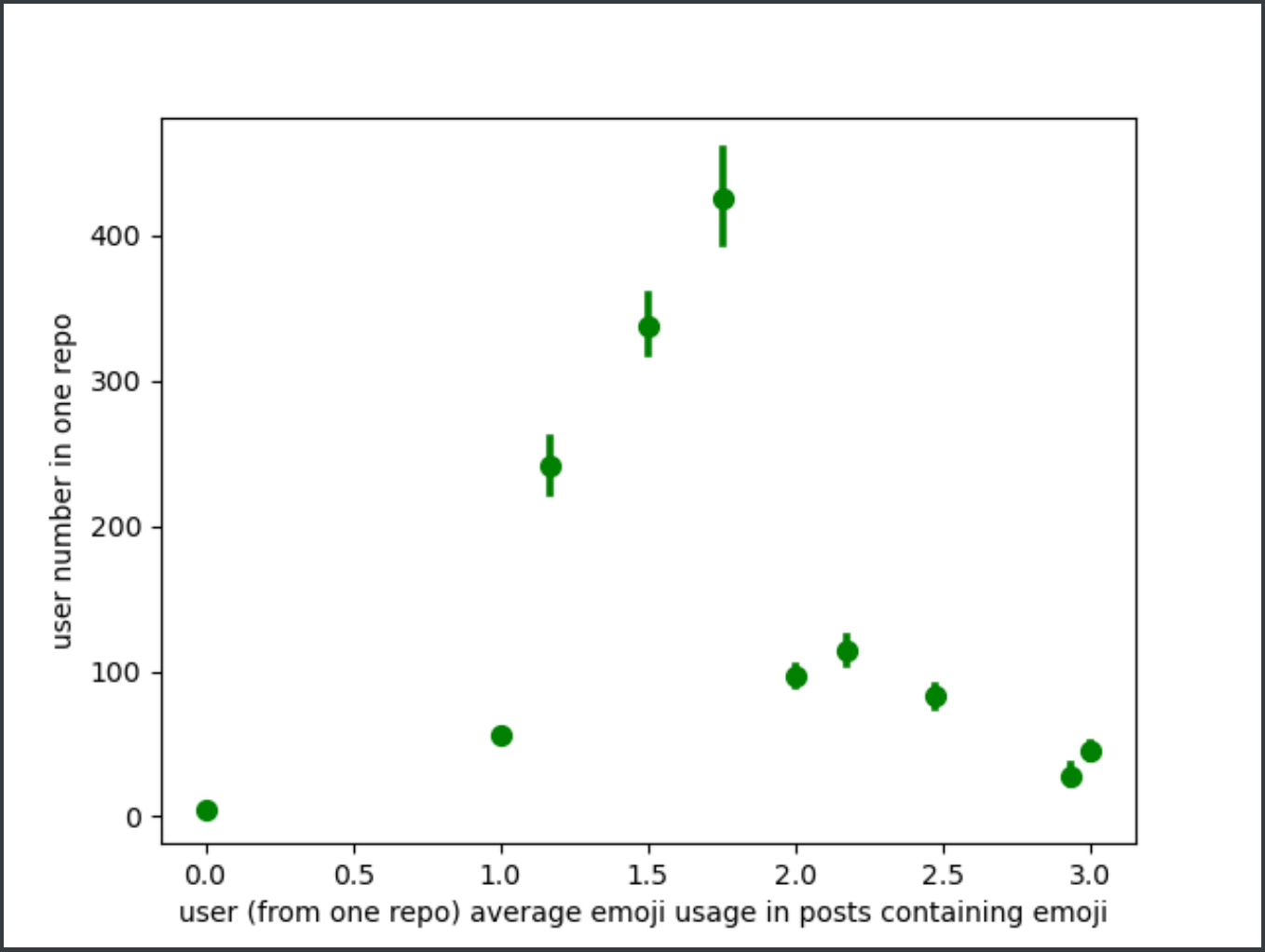
Classfiy based on repo user number: for example more users in one repo yield more emoji types average

含有emoji的conversation的比例 并到一个conversation里

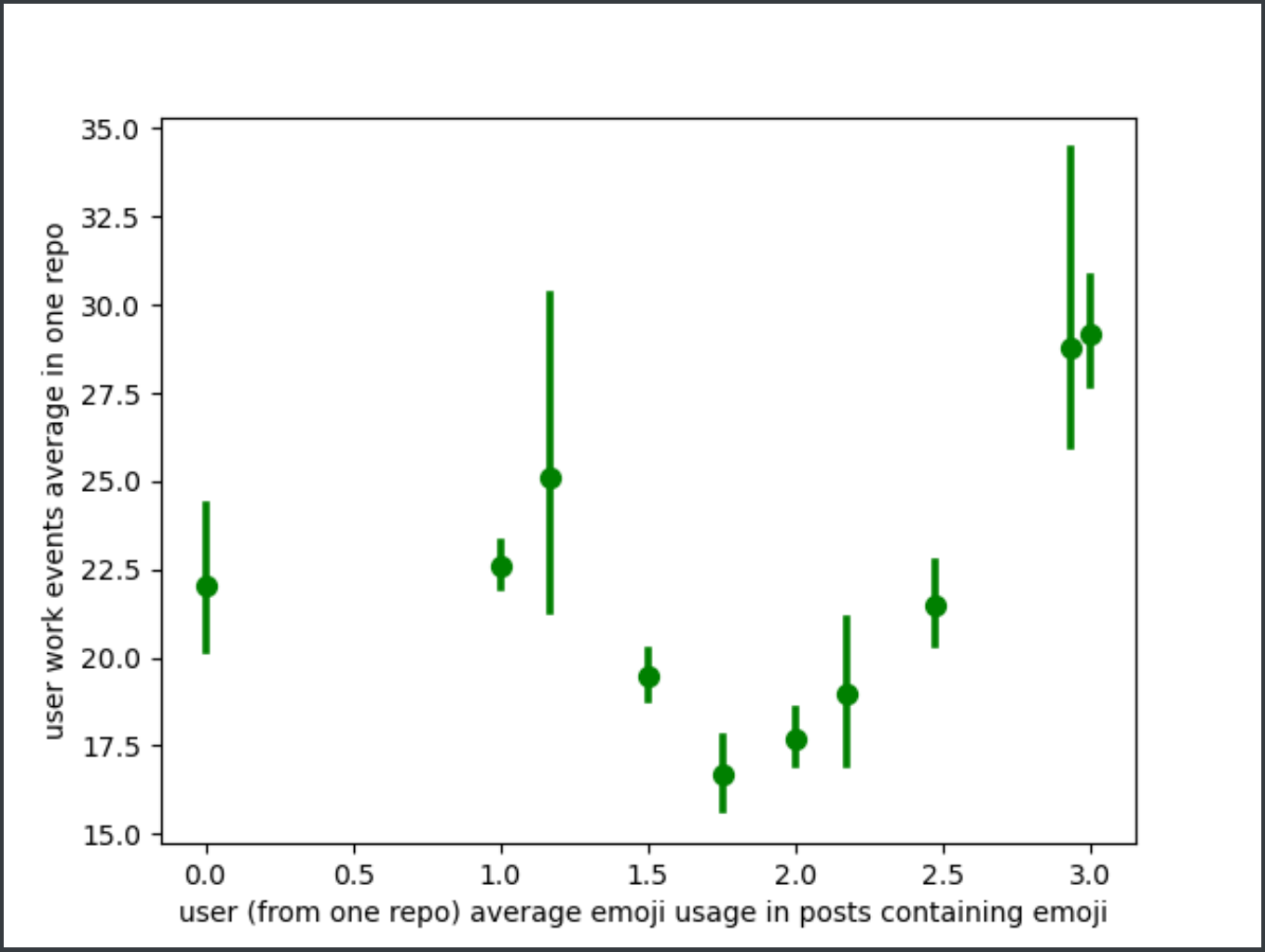
一个conversation里有多个posts, normalize

conversation里有多快出现emoji,

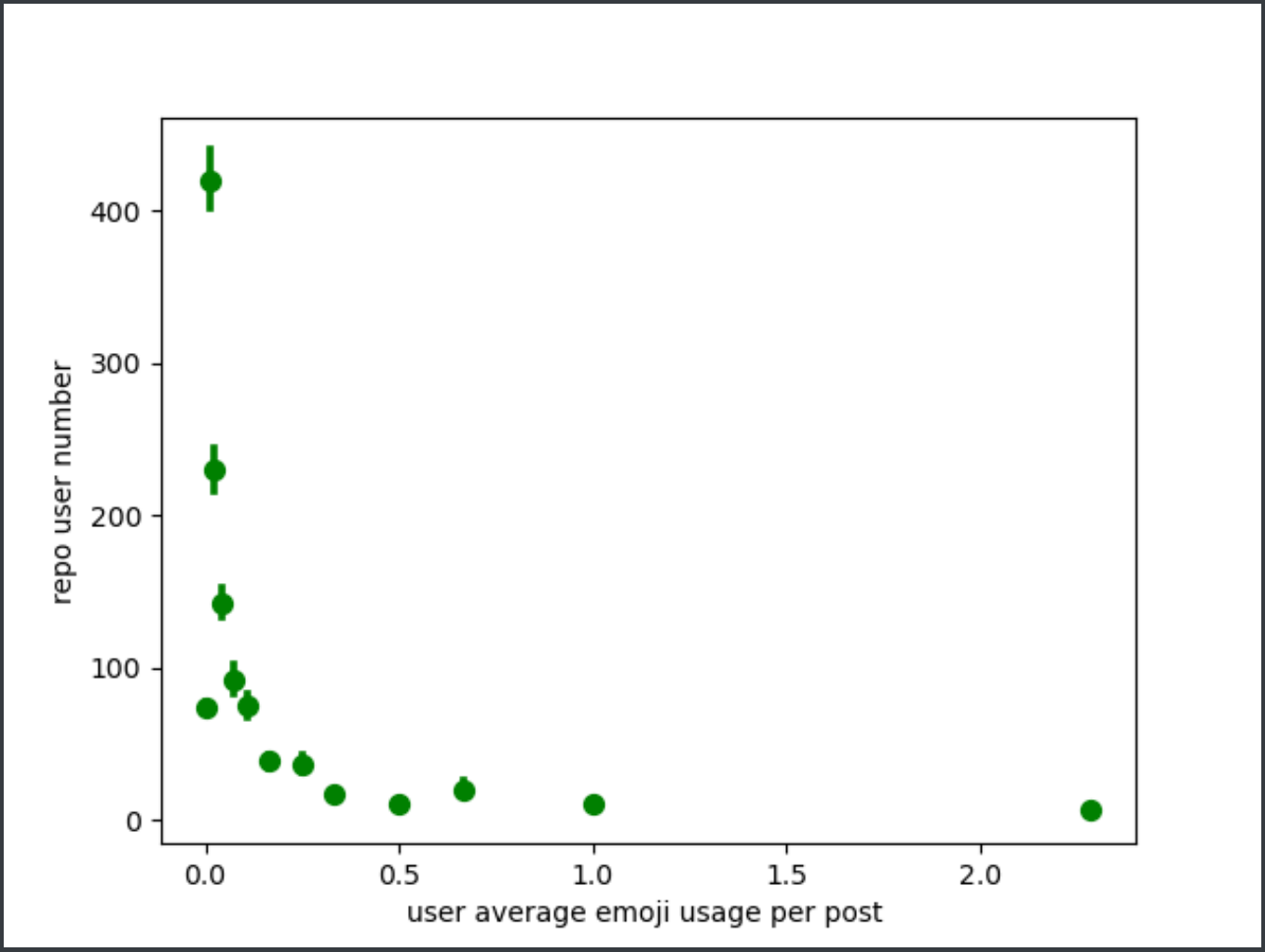
X: repo user average used emoji count



X: user average emoji usage in posts containing emoji Y: user work events average (one year) in one repo



X: Emoji per post Y user number


repo user number
400
300
200
100
0
0.0
0.5
1.0
1.5
2.0
user average emoji usage per post