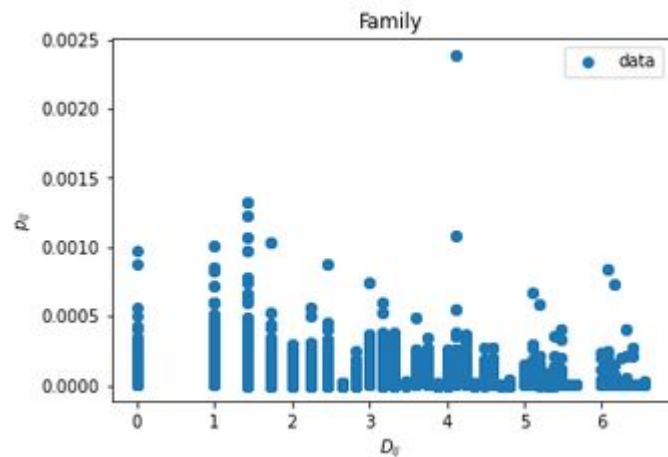
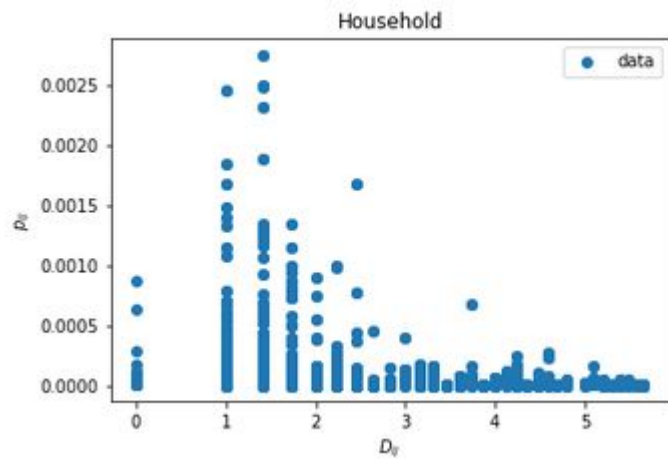
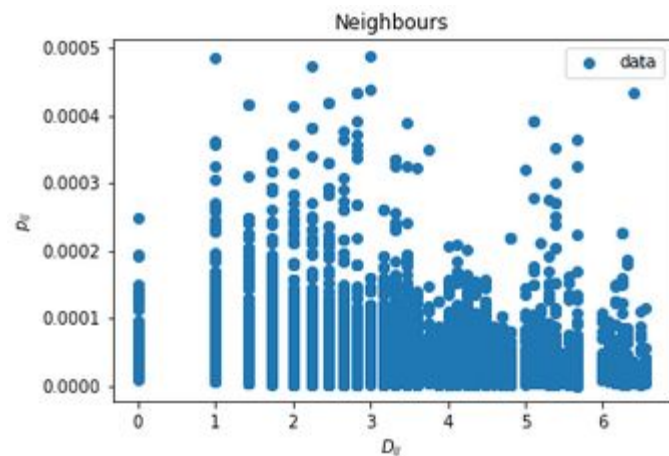
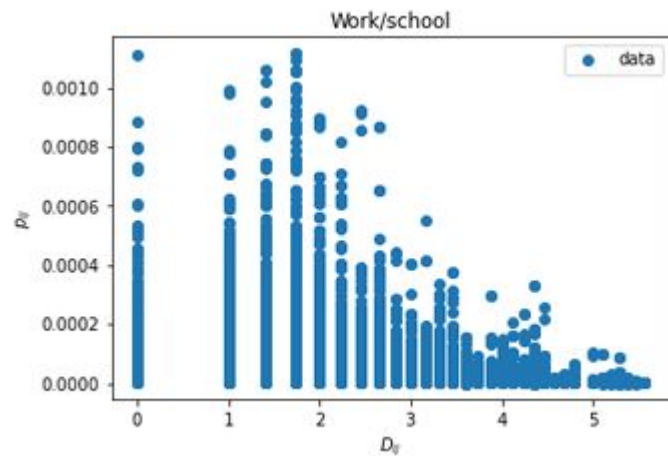
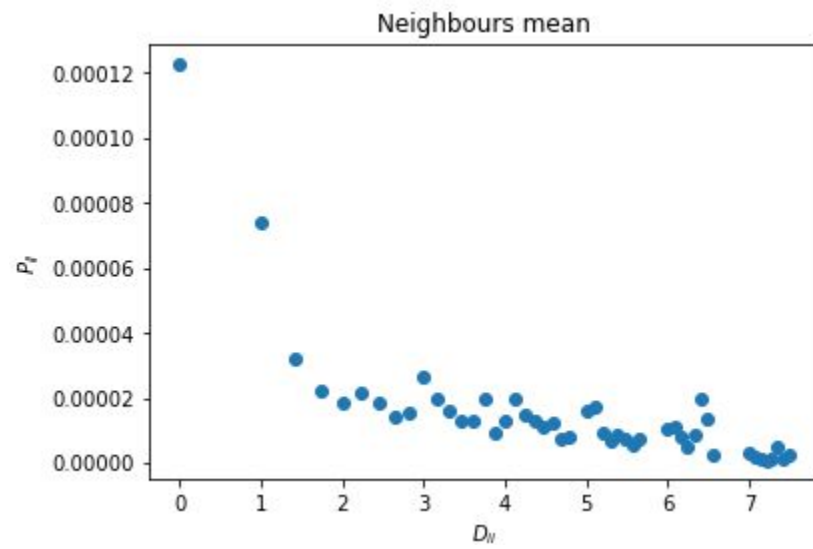
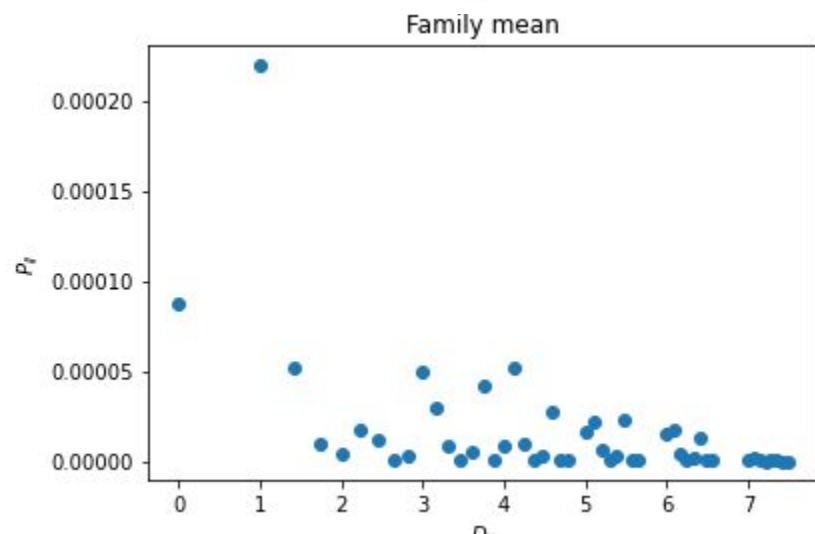
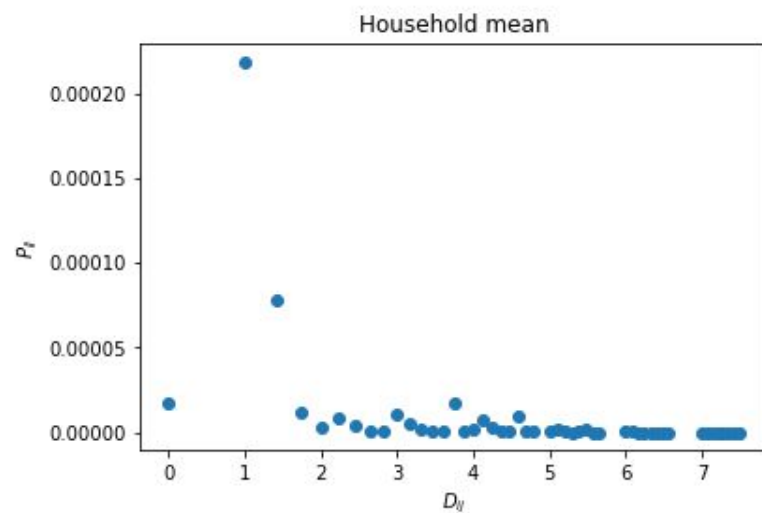
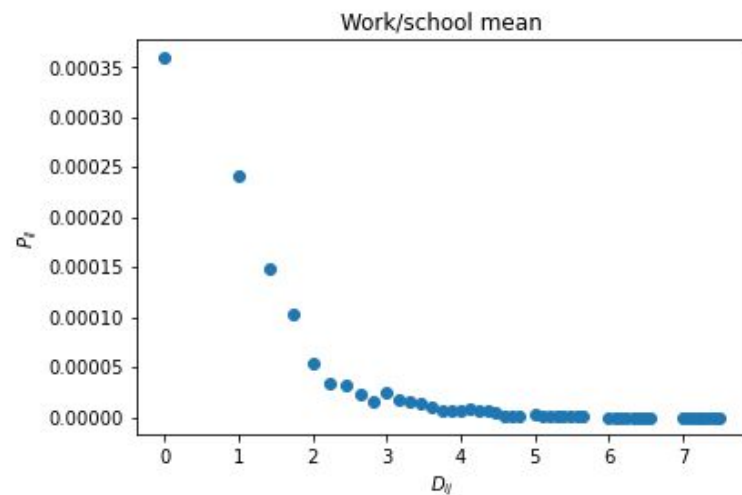


Probability - Distance

- Looking at the relationship between probability and distance
- Taking average of every distance to get better insight
- Distribution

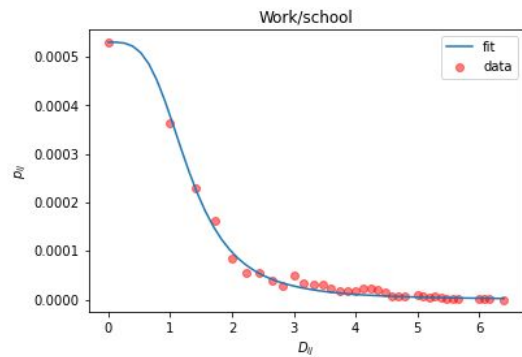




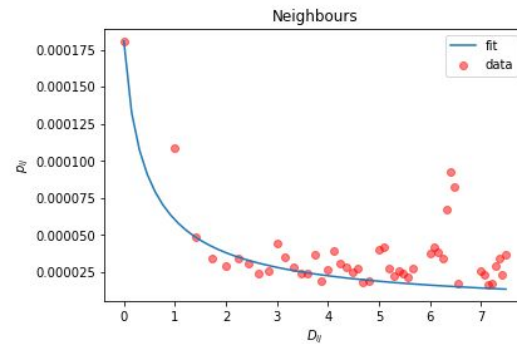
Fitting function

- First method was using $\text{mean}(p_{ij})$ where $d_{ij} = 0$ instead of 1
- Try to fit it with scipy to get the parameters α and b

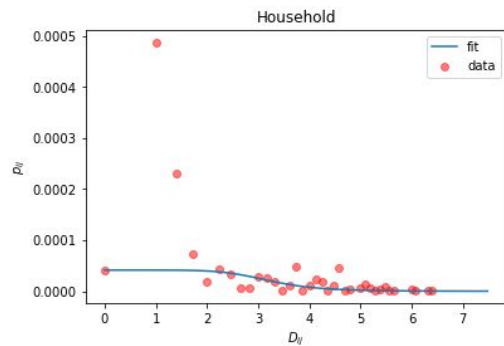
$$p_{ij} = \frac{1}{1 + [b^{-1} d(x_i, x_j)]^\alpha} \longrightarrow p_{ij} = \frac{a}{1 + (\frac{1}{b} * D_{ij})^\alpha}$$



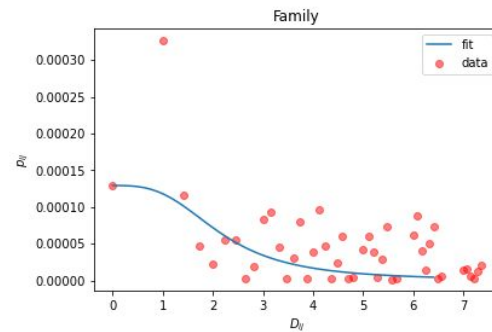
$b = 3.46352199$ $\alpha = 1.30000613$



$b = 0.90522746$ $\alpha = 0.46531558$

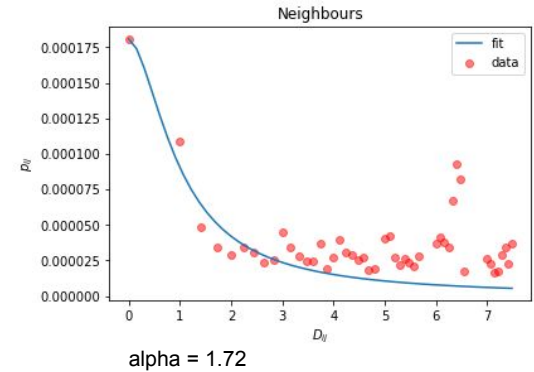
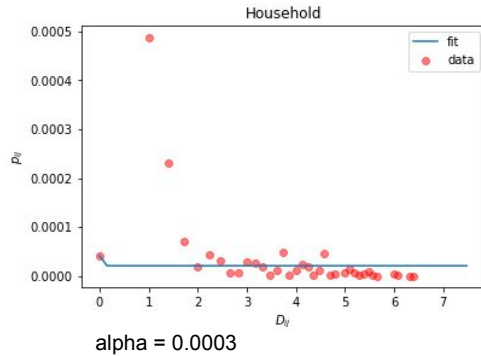
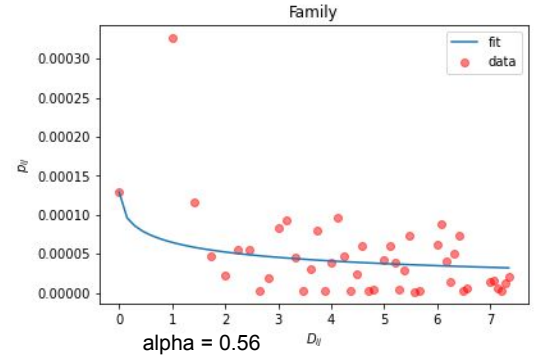
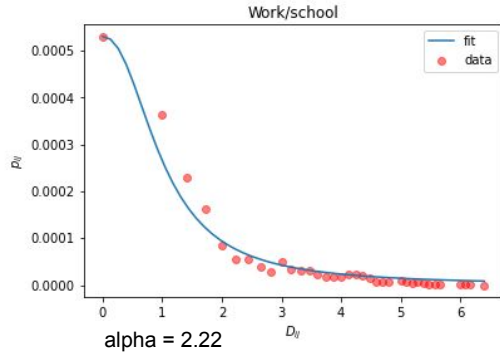


$b = 6.63827622$ $\alpha = 3.21663278$

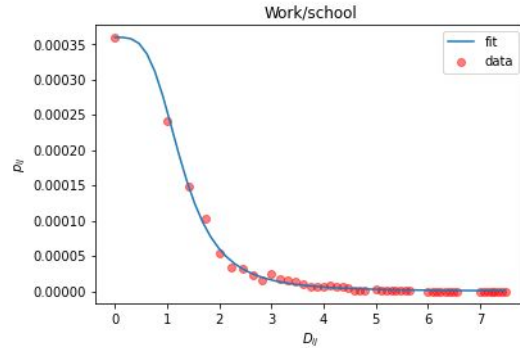


$b = 3.07415509$ $\alpha = 2.15199069$

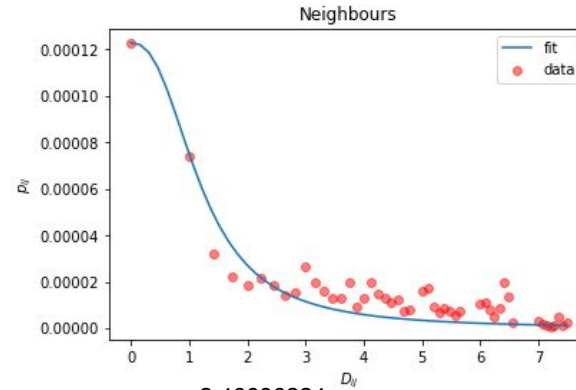
Keeping b at 1 and only varying alpha



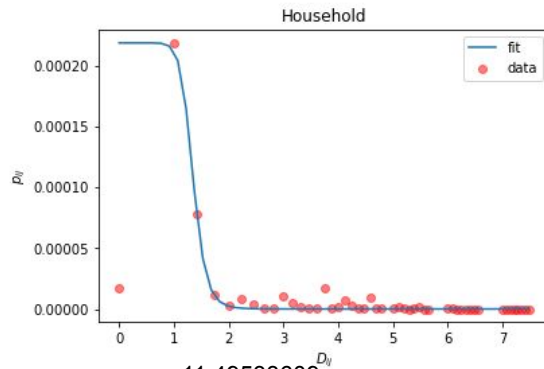
$a = \max(p_{ij})$ instead of $a = p_{ij} \Rightarrow d_{ij} = 0$



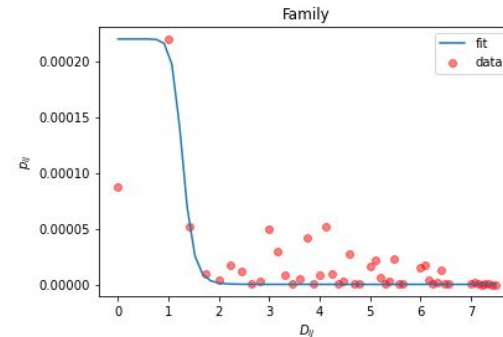
$a = 3.61188973$
 $b = 1.28011541$



$a = 2.46636224$
 $b = 1.23221235$

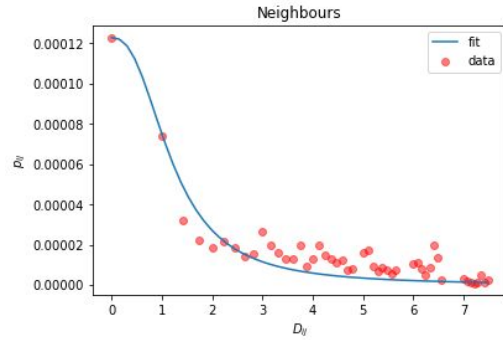


$a = 11.49533609$
 $b = 1.34599449$

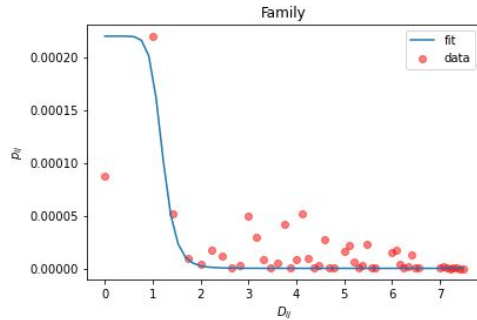


$a = 11.77616785$
 $b = 1.28549145$

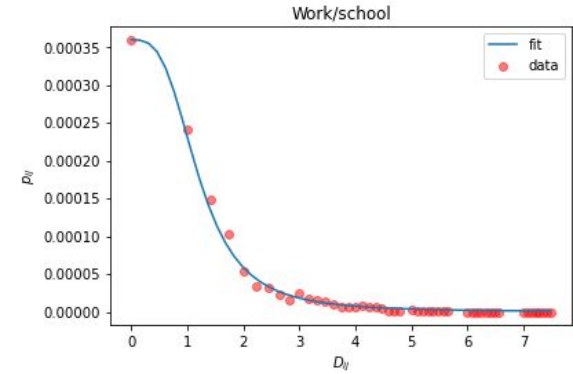
Keep b fixed for comparison



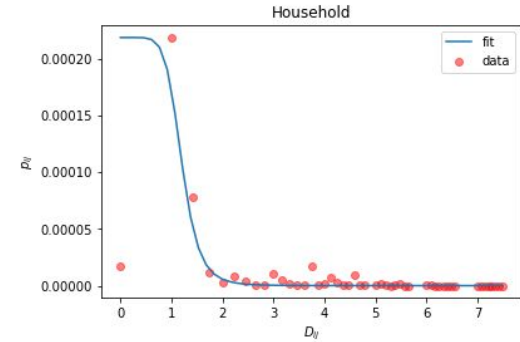
2.46636224



8.8234499



3.20626269



7.10136322

Important factors

- Based on paper: 'Birds of a feather'
- Looking at most important features of prediction probability
- Re-writing p_{ij} function for logistic and linear regression:

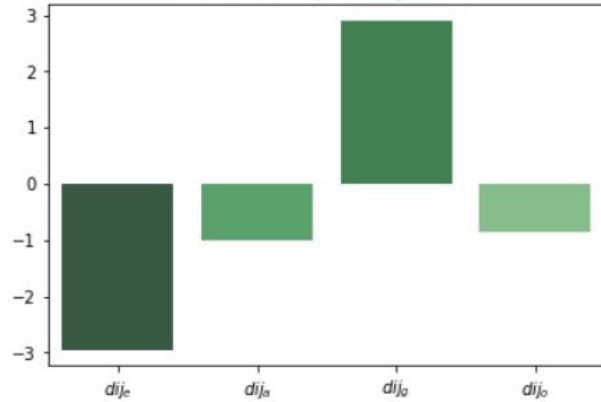
$$p_{ij} = \frac{a}{D_{ij-A} * X_1 + D_{ij-G} * X_2 + D_{ij-E} * X_3 + D_{ij-El} * X_4}$$

$$p_{ij} = D_{ij-A} * X_1 + D_{ij-G} * X_2 + D_{ij-E} * X_3 + D_{ij-El} * X_4$$

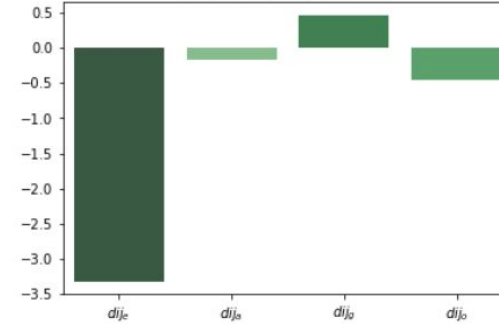
- Using linear regression and logistic regression to get a idea of the feature importance

Most important predictors logistic

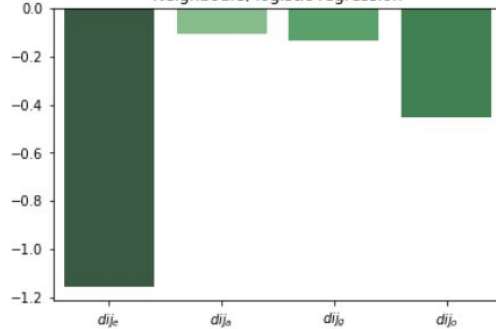
Household, logistic regression



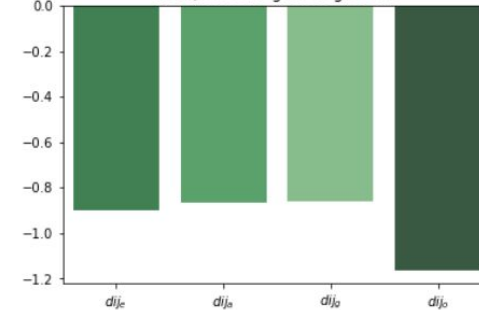
Family, logistic regression



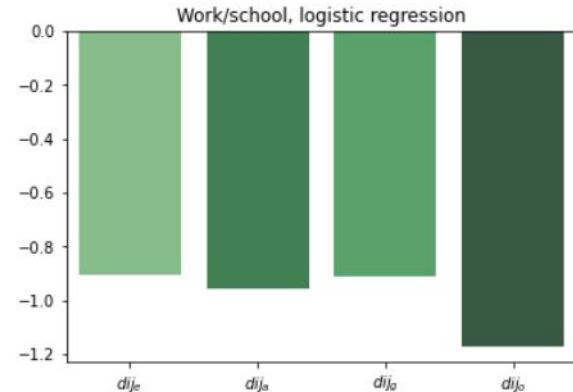
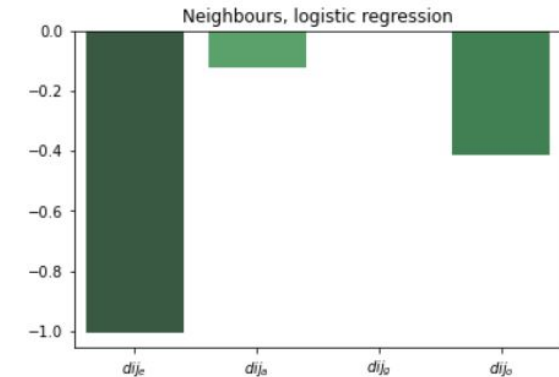
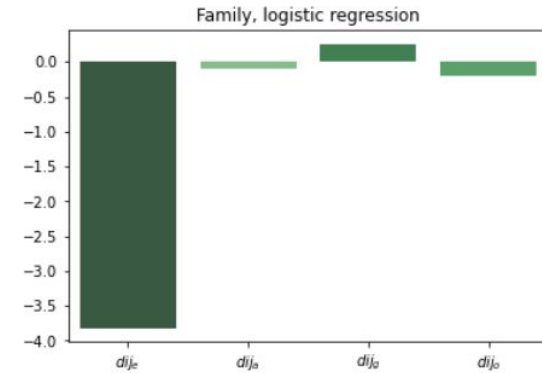
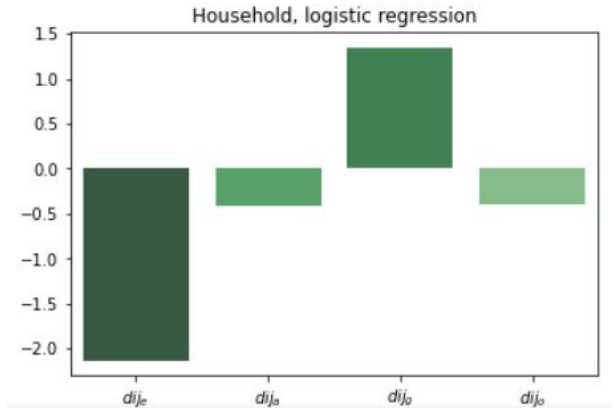
Neighbours, logistic regression



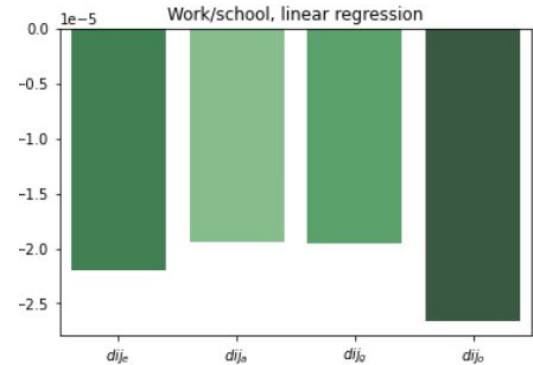
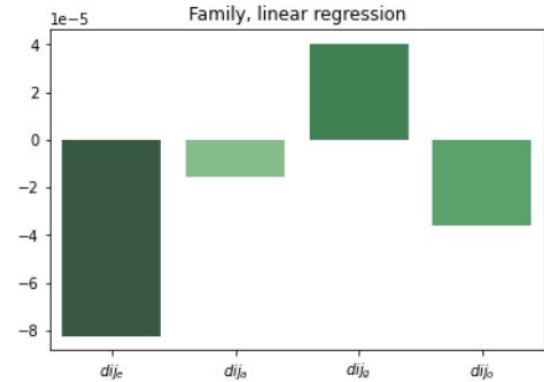
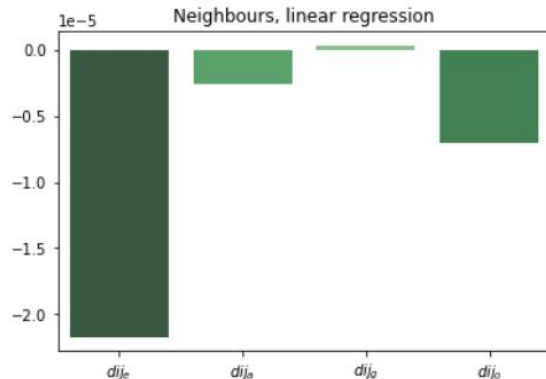
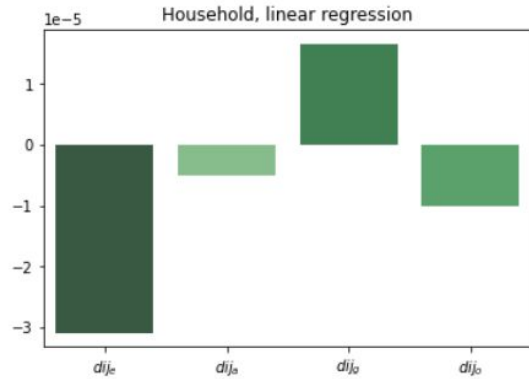
Work/school, logistic regression



Most important predictors logistic - balanced



Most important predictors linear



Back to literature

- Birds of a feather papers states:
 - Neighbourhood: ethnicity, race, religion, and family background
 - Work/school: In general, ties formed among co-workers tend to be more heterogeneous in race and religion than ties formed elsewhere
 - Family: Ties are much more likely to be same race, same ethnicity, and same religion

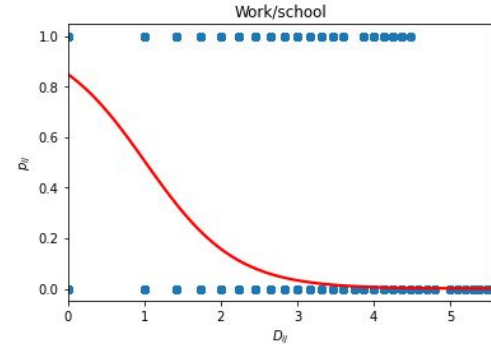
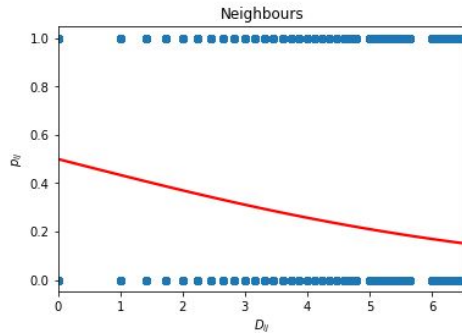
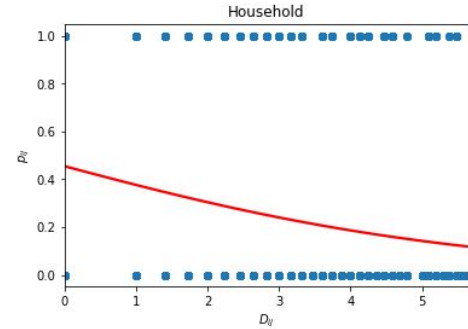
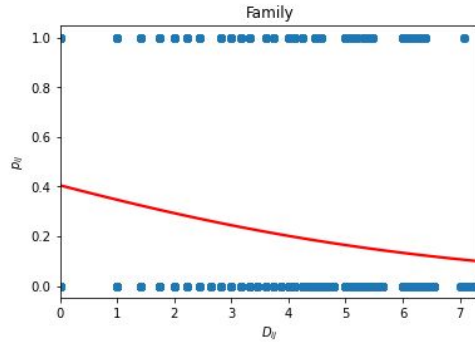
Notes

To-do

- Interlayer correlation
- Finishing spatial data, by looking at other attributes aswell:
 - Age
 - Education
 - Gender

Make a logistic regression (maximum entropy)

- Use cut off point of $a/2$, after the removal of outliers



<https://www.cbs.nl/nl-nl/longread/statistische-trends/2021/huishoudensprognose-2021-2070-groei-aantal-huishoudens-houdt-aan?onepage=true>