6.3732 PSet 4 Part 1

Stephen Andrews

April 11, 2025

4.1:

a) There are 1,280,459 rows in the dataset, and the number of unique cases is 539,593.

b) Similarly, the number of offenders is also 539,593.

c) The total number of crime events is 1164836. For each year the totals are as follows.

Year	Crime Event
2003	110556
2004	121374
2005	172664
2006	185840
2007	196151
2008	199288
2009	178959
2010	4

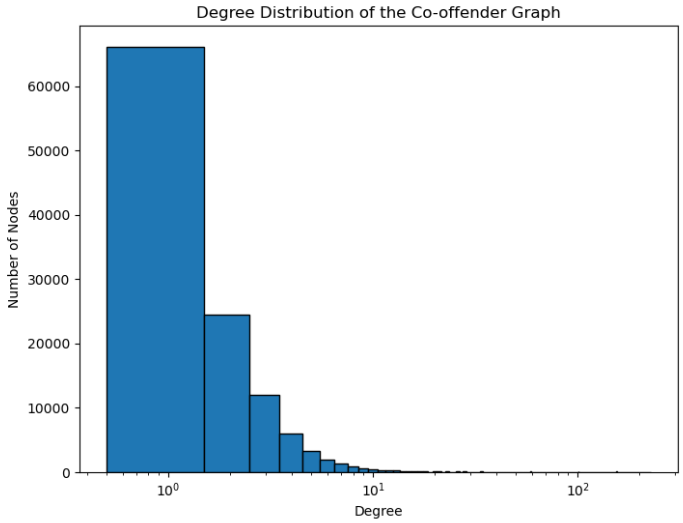
d) The following table shows the top 10 crimes with the most number of offenders and the municipalities in which they were in.

Degree	Frequency	Municipality
27849	156	66023
876159	102	12072
445040	77	66023
23526	60	75017
60815	60	54048
212285	57	37067
754187	56	95032
539385	52	66023
23610	51	78005
453024	46	66023

e) The number of nodes is 121159. Therefore, the number of solo offenders is 539593-121159 = 418434. The number of edges is 178413.

(f)

The degree distribution is as follows. The log-scaled plot shows a heavy-tailed distribution with many nodes having low degree and a few nodes with high degree, which suggests the plot may follow the power law degree distribution.



(g)

There are 36098 connected components.

(h)

The largest connected component has 19924 nodes.

(i)

The new degree distribution is as follows. From the entire entire co-offender graph, many nodes are isolated or in small components, so there's a large spike at very low degree. This skews the degree distribution significantly on the left side. In contrast, the largest connected component excludes those isolated nodes, so its distribution shifts toward higher degrees overall and reveals a slightly different tail shape.