Analysis of Orlando Theft Crimes

Network analysis of the crimes in and around the city of Orlando, Florida by means of clustering.

Brenden Morton

University of Central Florida

Network Optimization

Dr. Wei Zhang

December 8, 2020

Throughout the history of human gathering, individuals or groups of individuals have committed theft crimes. In general, theft can be described as the reallocation of one’s goods in a non-consensual manner with the intent to deprive the rightful owner of personal property. Although the motivations for this genre of crime can develop from both economic and non-economic circumstances, there is always a victim. According to American sociologist Richard Quinney, there is a distinct relationship between a society and the crimes that occurs within it. Reflecting on this claim, I thought it would be interesting to see the relationships between a subset of a largely evident type of crime on a geographical scale centered in and around our community. In this paper, the crimes in the city of Orlando, Florida between the years of 2014 and 2020 are thoroughly examined through the use of network and data clustering algorithms to provide an insight into where exactly theft is most prominent and whether or not there are relationships between theft crimes. In the case of this report, all crimes that involve the loss of property that are not explicitly defined as theft such as motor vehicle theft and burglary will be generalized as theft. The reason behind is that it is necessary to include these crimes since they involve similar, if not the same, motivations as regular theft.

To collect the geographical relationship between theft crimes around the city of Orlando, Florida, network clustering algorithms, as well as some data clustering algorithms, are employed. Markov Chain Clustering and Spectral clustering will be used for aggregating the dataset of crimes based on the locational positioning of the crime in terms of a street intersection. Additionally, the Mean-Shift data clustering algorithm is used to gain a perspective on the distribution of theft crimes across Orlando. Additional software such as Cytoscape will be used to visualize the graph constructed by the unique streets of the dataset as well as to illustrate the clusters formed by the clustering algorithms. Crude representations of the data are developed by the Networkx Python 3 package are also used to exemplify some dense regions in the network in a different perspective.

The city of Orlando provides a multitude of varying datasets on their website data.cityoforlando.net. Here, a few different datasets for criminal activity were available for public use, however, the most extensive dataset is the one used in this paper. In this dataset resides roughly 230,000 crimes that occurred in the city bounds of Orlando starting in the year 2014 and ending in September of 2020. According to the description of the dataset, all crime data conforms to FBI standards in terms of the definition of the crime. In the case of this report, the crimes of theft, motor vehicle theft, and burglary are verified to be accurate descriptions of the crime. In general, the data from this dataset can be exported to a csv file which delimits 10 different records for each of the 230,000 crimes. The following is the list of attributes that are shown for each crime: Case Number, Case Date Time, Case Location, Case Offense Location Type, Case Offense Category, Case Offense Type, Case Offense Charge Type, Case Disposition, Status and Location. The attributes of interest in the analysis of the data included only the following: Case Location, Case Offense Type and Location. Case Location refers to the street intersection of the crime. The dataset consisted of two forms for Case Location where only one of which was adequate for analysis. The Case Location could be given in the form of [street 1 / street 2] which can easily be tokenized and analyzed with an adjacency matrix or of the form [Block of X on Street] which can be tokenized but not in an advantageous way. The latter form does not provide enough information to be used for network analysis. In the case of the former form of [street 1 / street 2], these streets can be tokenized and provides some intersection in Orlando that works well with an adjacency matrix. Whereas the form [Block of X on Street] represents a residential crime and this data is not able to be easily translated to a position in an adjacency matrix. To remedy the situation, I filtered out this form such that only the crimes with valid [street 1, street 2] form existed in the dataset. The next attribute of interest is the Case Offense Type. This attribute can be used to filter out any crimes that do not belong to the subset of theft which includes theft, motor vehicle theft and burglary. The last attribute of interest is the Location of the crime. This differs from the Case Location in that it represents the longitude and latitude coordinates of the crime in the form [long. / lat.].

<Brief introduction of the implementation and details of your results>

Find some images maybe 3 or more?

Mat plot lib of streets

MCL clusters maybe some tables

- meaning of those clusters, the more streets in the cluster the higher the crime activity of that intersection

Cytoscape mcl and spectral

Write a bit about each image

Citations

1. MetroWest Public Safety. (created 2016, last updated 2020) City of Orlando Crimes [Crimes in Orlando, Florida]. Retrieved 23:44, December 1, 2020, from <https://data.cityoforlando.net/Orlando-Police/City-Of-Orlando-Crimes/hm2t-fd4m>
2. Wikipedia contributors. (2020, November 28). Crime. In *Wikipedia, The Free Encyclopedia*. Retrieved 04:43, December 1, 2020, from <https://en.wikipedia.org/w/index.php?title=Crime&oldid=991063902>