**Overview**

The solution is designed around an algorithm to assist in reduction of gang-related gun violence. The algorithm, implemented using open-source software packages, performs social network analysis of arrest records. The open source softwares include: R Tool, R Statistical, Node XL, MYSQL, ItextSharp, Log4Net, and .Net Framework. The contributions of Nielsen and TCS include the development of a scalable and user-friendly software solution which can be used by the law enforcement agencies with minimal intervention. Specifically the team delivered a solution addressing loading, processing, and visualization of the data. The solution was developed with the understanding no long-term support from Nielsen or TCS would be provided.

|  |  |  |
| --- | --- | --- |
| [**Category**](#Reference!A1) | **Requirement Description** | **Supporting Information** |
|  |  |  |
| **F - Data** | The user is responsible for uploading 4 files: Arrest.csv, Gang.csv, POI.csv, and Victim.csv. "Input Files Requirements Document.xlsx" contains the specific templates and details. Note the Gang, POI, and Victim files are all *OPTIONAL*. | [Input Files Requirements Document](../../../private/var/folders/9r/5drz8z_96hb0pdynl3yxqj8w0000gn/T/TemporaryItems/Input%20Files%20Requirements%20Document.xlsx) |
| **F - Data** | Master.csv is the result of combining all 4 of the user inputed files using the PersonID as the primary key. | Master.csv, Arrest.csv, and Ties.csv will always have the same # of rows. |
| **F - Data** | Ties.csv is a subset of the Arrest.csv file. It only contains the first 2 column; PersonID and CaseID. This file is used by the R code to generate the social network. |  |
| **F - Data** | Append the most recent attribute entry (Victim, Gang, POI) to each Person. The solution must be able to support and filter on a person with multiple gang affiliations and victim descriptions. |  |
| **F - Data** | Append characters to the beginning of PersonID and CaseID | Append the letter 'a' to the beginning of Person; Append the letter 'b' to the beginning of Case |
| **F - Data** | R code must output a NetworkGraph.graphml file to generate the visualization of the social network. | The R code must be tweaked to generate a NetworkGraph.graphml which is required by NodeXL |
| **F - Data** | Default field length should be used for each field | This is at the discretion of the development team, no requirement from the professor |
| **F - Data** | R code must output a NetworkGraph.gml file to provide to Andy for analysis | This is in addition to the .graphml needed for NodeXL |
| **F - Data** | Names will be entered as a whole name | The whole name is entered as a single entry; filtering will be on the whole name |
| **F - Data** | The application Ties.csv file needs to support up to 5 million rows. | The projected NYC data will contain 4.5 million rows. |
|  |  |  |
| **F - DataCleansing** | All columns marked as required in the "Input Files Requirements Document.xlsx" must contain data of the specified type. If there are any errors, the program will kick back to the user identifying the errors and request they fix them and re-upload the file. | [Input Files Requirements Document](../../../private/var/folders/9r/5drz8z_96hb0pdynl3yxqj8w0000gn/T/TemporaryItems/Input%20Files%20Requirements%20Document.xlsx) |
|  |  |  |
| **F - Assumptions** | There are no specific requirements to modify the Professor's R code. The system curently takes the ties.csv as input and produces the .graphml file required to produce the visualization of the social network. | The team can make any changes necessary to the R code as long as the results are the same |
|  |  |  |
|  |  |  |
| **F - Reporting** | The interface must include a visualation of the social network. The network will show nodes which represent people that have been arrested and ties between the nodes representing a relationship between people meaning they were arrested together. |  |
| **F - Reporting** | The interface must include a spreadsheet-like output of the social network. This will include the data from the Master.csv. | PersonID, CaseID, Arrest Date, Event Type, Address, District, DOB, Name, Offdes, Gender, Race, Gang Name, Victim Desc |
| **F - System Functionality** | The visualization and the spreadsheet-like interface must link. For example, when the user clicks on a node in the visualization, the row in the spreadsheet is highlighted or selected as well (and visa-versa). |  |
| **F - System Functionality** | Orphan Nodes: Any nodes that are not connected to any other nodes in the network will not be shown in the visualization |  |
| **F - Reporting** | The resulting social netowrk that is visualized must include 3 key metrics: • Degree Centrality (count of # of ties) • Betweeness Centrality (how much of a broker/in between people you are) • Modularity (groups people in clumps/clusters of people) | *\*need to follow up with Andy on 2 points… which of these does he want applied to the whole netowrk vs. a single node, and clearly define Modularity and Betweeness metrics. \* where on the app should they be shown?* |
| **F - Reporting** | Each filter on the visualization screen must have the ability to change the size and/or color of the nodes that meet the filter's criteria. |  |
| **F - Reporting** | Users need the capability to download reports. | The professor has requested PDF formatting. Export .gml to upload in gephi required |
| **F - System Functionality** | Users need the capability to save their workspace/ search criteria or upload that of another user. Workspace includes all filters on input screen, filters on graph, NetworkGraph.graphml, and graph object. | Save at least the previous 3 sessions |
|  |  |  |
| **F - System Functionality** | There are 2 required layers of filtering. The first layer filters the full arrest data set to determine the seed list. The second layer filters the resulting social network (seed list + desired # of handshakes) |  |
| **F - System Functionality** | The users should have the capability to select multiple values for the gang and district filters. |  |
| **F - System Functionality** | All filters should provide cascading functionality |  |
| **F - System Functionality** | 1st (seed list) filter will contain the following: Person IDs (manual entry and POI.csv upload), gang name, # of handshakes, district, arrest date range. | EITHER some POI upload/manual entry OR gang OR district, must be selected from the first filter set |
| **F - System Functionality** | Filters will have an intersecting relationship. For example, if the user selected district 1 and the nortenos gang, the result would only be the nortenos that were arrested in district 1. | We can explore providing the option for intersection or union as a part of the interface. |
| **F - System Functionality** | 2nd (results) filter will contain all of the filters as outlined in the "Input Files Requirements Document.xlsx". | [Input Files Requirements Document](../../../private/var/folders/9r/5drz8z_96hb0pdynl3yxqj8w0000gn/T/TemporaryItems/Input%20Files%20Requirements%20Document.xlsx) |
|  |  |  |
| **NF - Security** | No application level security needed. The application will be installed on local machines within a police department. Users with access to these machines will also have access to this application. |  |
| **C - Anticipated Workplace Environment** | Application must be able to run on a standard Windows Machine (XP and Windows 7) |  |
| **NF - Security** | All PII data needs to be encrypted | SHA-2 |