ISSN 2409-2665
Journal of Logistics, Informatics and Service Science
Vol. 9 (2022) No.1, pp.367-404
DOI:10.33168/LISS.2022.0122

A Dynamic Visual Analytical Approach for Crime - Transformation to Better Decision Making

Neelu Chaudhary [1], Hardeo Kumar Thakur[1], and Rinky Dwivedi [2]

[1] CST, Manav Rachna University, Faridabad, India
[2] CSE, Maharaja Surajmal Institute of Technology, Delhi, India

neelu@mru.edu.in

Abstract. Crime often disrupts normal life; people can experience lots of complex emotions. Everyone reacts differently in the adverse situations that can impact a person's mental well-being. The effect of crime last a long time and it doesn't depend on how 'serious' the crime was. Presently, digitization is created new incentives for the violence or new avenues for its arbitration, social media has made victimization visible. Crime analysis is a law enforcement function that involves systematic analysis for identifying the patterns and prints in crime and disorder. Information on patterns can help Criminologist and law enforcement agency to address violence or deploy resources in a law effective manner. Crime at a topographical location is an integral function of cultural issues, religious issues and international policies. Sharp rise or alarming number of cases consequences the degradation of government trust. The present research focusses on studying the Crime dataset of San Francisco city over last ten years. The research intends to analyse how crime is distributed across the city, the density of crime incidences, the weakness, and strengths of administration in particular - amount of cases occurrence and number of cases resolved, variation of crime activities in an area, triumph proportion of forces in solving a crime and dominance of crime on weekends. Mapping crime using Geographical Information System allows crime analyst to identify crime hotspots along with other trends and patterns. Crime analysts wield crime mapping for terrestrial profiling and formulating policies of allocating constabularies' perambulations and dispatching them in case of emergencies.

Keywords: Dynamic graph, crime, weighted bipartite network, attributes, social network

1. Introduction

Crime impacts a plethora on the life of people, some individuals developed emotionally unbalanced afterwards. Every now and then, they develop extensive complications such as unhappiness or nervousness related sickness and a few people suffers an unadorned, long-lasting rejoinder afterwards identified as post-traumatic stress disorder. These robust sentiments can make them feel more anxious and confused sometime people projects impression quite normal but may suddenly start to fall apart. Other people might have physical symptoms such as struggling to sleep or feeling ill. Many victims' loose self-confidence and start blaming themselves or feel too embarrassed. As a result, safety analysts must conduct a thorough investigation of crime events to identify elements that contribute to the occurrence of a crime. Officers can use crime data analysis to maintain track of criminal activities, uncover similarities between happenings, deploy resources, and make faster decisions. An enormous amount of crime data has been gathered and made public due to the rapid development of computer networks and storage technologies; such big data can give fresh insights for better recommendations (Yoon et al., 2020) and especially analysing crime trends (Feng et al., 2018).

Dynamic Graph mining (Gupta et al., 2015) is an innovation, interdisciplinary, and rapidly expanding discipline of study and analysis that aims to develop paradigms and techniques for extracting meaningful information and hidden patterns from large amounts of data from various sectors. It can assist us, not just in knowledge discovery or the discovery of new phenomena, but also in improving our comprehension of existing phenomena. We can quickly identify the pattern of crimes that occurs in a specific region and how it is connected to time using dynamic graph mining tools. The implications of visualization and statistical analysis on crime data, particularly time series data, will not only allow us to understand the pattern and trends of crimes in a nation but aid society in planning for crime prevention and reduction.

In this article, the crime statistics from San Francisco city are observed. For better results, the proposed process is initiated with feature normalization and feature selection to get the most significant characteristics, then traversed the data across time and visualized it in the form of a dynamic graph for a better understanding of varying crime statistics concerning time in different districts of the city. Finally, given a time and place, the crime categories are analysed for different districts to extract meaningful information that helps to make future strategies for improving crime demography in the city. For the analysis, the dynamic attributed graph (Cheng et al., 2017) approach is experimented with crime attributes like total incident cases, resolved cases, weekend cases, and neighbourhood cases and applied commonly adopted data analysis approaches on these attributes to retrieve hidden information from the network.

The proposed approach outperforms other forecasting techniques for daily and monthly forecast. The genesis and character of crime, its impact is used in dynamics of deep scientific research. The modus operand of the crime is if predicted and calculated then the effect of its happening can be restricted. This can be achieved by the deep studies and research. Crime visualization and analysis using dynamic attributed graph helps in discovering the explications and formulating crime prevention strategies, it incurred at various levels including tactical, operational and strategic level. Analysing such graphs is an intricate process, comparison to single label graphs because each vertex and edge is linked to a set of attributes. Many applications need the analysis of a single attributed graph that evolves. These quantitative and qualitative data analysis methods of crime analysis for instance examining laws intelligences helps in devising solution to crime problems. Crime analysts study police reports to identify patterns, series and trends as quickly as possible this analysing formula sometime predicts or forecasts future occurrences and generates reports and alerts to the agencies. Visual analytics approach helps in expanding patrol operations, supporting investigative and specialized unit operations and maximizes resource allocation. It encourages imagination, gives an insight view and ensures transparency in the system.

The other portions of the paper are as follows: Section 2 covers the literature survey that helps to understand the related work done and their limitations that we try to overcome in our study. Section 3 describes the definitions and preliminaries used in this paper. Section 4 covers the method used for visualization and analysis. The framework used is described in detail in this section. It covers all the steps from pre-processing of data to generation of dynamic network then visualization and analysis of dynamic network to achieve the desired objectives. Section 5 covers the result part of analysis. Section 6 covers the discussion part based on the intense analysis of network and in last, section 7 covers conclusion and future work.

2. Literature review

The literature purposes a combination of vertical associated with the visual analytics for the data related to crime and approaches related to dynamic graph mining and analysis. Several authors have given significant contribution in this field (Feng et al., 2018) (Chen et al., 2018) (Launtz et al., 2020) (Bailey et al., 2020) (Lina Baranauskaite 2020) they projected risk factors and crime patterns using the tabular data in a specific geographical place. The studies are substantial milestones but lacks to discuss the dynamic behaviour on time series data although investigators predicted crime categories for parameters such as time and location. Some authors have also observed at unweighted and undirected networks, with treating co-offending as a dynamic network process (Wu et al., 2020) (Charette et al., 2017) (Park et al., 2012) (Moor et al., 2018). The research (Charette et al., 2017) looks at the individual and network mechanisms that lead to long-term criminal partnerships. In their analysis, Factors such as centrality and transitivity influence the likelihood of co-offending. The Purpose of the paper (Park et al., 2012) is to show the usefulness of analytic algorithms in predicting crimes and depiction of the dynamic evolution of networks over time. With the expansion of visual analytics, new analytical reasoning tools for exploring and analysing vast amounts of data using interactive interfaces are now available (Aggarwal et al., 2016) (Bokhare et al., 2021). The author considered traditional time delays of a year or six months between assessments, may be insufficient to detect short-term relationships between peers and behaviour (Gupta et al., 2015). The author provides a framework (Thakur et al., 2017) that would make the mining of regular patterns in dynamic networks more efficient and thorough. This framework's correctness has been theoretically confirming, and its efficiency has been experimentally validating on a real-world network. The attribute value of nodes and edges, on the other hand, is disregarded. As mining in weighted dynamic networks has never been before, the research (Lahiri et al., 2008) offer a unique approach for mining maximum quasi-regular patterns on structure in this study. The motif is then analysed using four factors: modularity, cliques, centrality, and intersection, to acquire a better understanding of their nature. Bipartite networks are a highly significant type of complicated network and are used in intensive research of networks (Rim et al., 2020) (Wang et al., 2016). Within them, maximal bicliques are the sturdy structural communities imaginable (Zhang et al., 2008) (Duan et al., 2001). There are two types of nodes in bipartite networks: top nodes and bottom nodes. Only a pair of vertices belonging to distinct sets connect by edges. The authors suggest a change to the clustering coefficient in the bipartite network, by the proportion of cycles having size four. The two definitions got compared for a specific graph, with the findings indicating that modification one is superior for characterizing the network. On unweighted bipartite networks, biclique communities get discovered by considering overlapping communities and propose a technique for detecting an order-limited number of overlapping maximum bicliques covering the graph (Alzahrani et al., 2019). Further examination of these bicliques reveals that they are consistent with the ground truth and provide meaningful additional information. The bipartite network of an Internet store web platform gets examined in this study, where customers and items represent nodes and purchases made represent links. Using the open-source network analysis and visualization tool Gephi (Cherven, 2015), the study is on the modularity function (Chessa et al., 2014). The authors look at clusters of "positional equivalent" actors who fulfil comparable roles in a system (Sinclair et al., 2004). Undirected but weighted bipartite networks to explain the interactions between actors and their features. The intensity level with which actors display their characteristics is also measured. They develop a methodological approach that takes into consideration the multi-dimensional dependency that exists between groups of players. The authors presented a new model for determining user preferences and dislikes (Wang et al., 2016). Under the diversity index, experimental findings on the

Movie-Lens dataset showed that the suggested technique was far superior to the baseline method. The Bi-Louvain method, which repeatedly groups the nodes in each section by turns, fills the gap in this study (Zhou et al., 2018). This technique generates a balanced network topology with equal numbers of two types of nodes in bipartite networks. The network considers unweighted and undirected. They show that the gain of modularity of each aggregation can be compactly computed for all pairs of communities at the same time using matrix operations. Finally, the entire hierarchical community structure gets revealed. It demonstrates that the technique efficiently identifies community structure in bipartite networks using two benchmark data sets and a large-scale data set from an e-commerce firm. On dynamic weighted directed unipartite graphs, this work focuses on community mining, including community identification and change-point detection for real-world networks like e-mail, co-authorship, and networks. The authors signify the strength of weights on edges to analyse social networks (Duan et al., 2009). Communities in weighted networks are detected to identify modules strongly connected (Lu et al., 2013) (Zhou et al., 2018). Table 1 shows the comparative analysis of most related and latest research in field of crime and dynamic graphs. It covers the purpose of study, techniques used, findings and limitations of studies concerning the present research.

It can be analysed that most of the work in crime analysis and decision making consist of either tabular data or static networks. Significant proposals on dynamic graphs also involves only undirected and unweighted networks. Most remarkable research done in this area emphasize on the degree of nodes and the number of edges between nodes. Few approaches are there for analysing a dynamic attributed graph, that is, a single graph with edges, vertices, and multiple attributes that can vary over time. The unique proposition of the proposed work is that it offers a more general pattern domain in a dynamic attributed graph, that further represents frequent evolutions. It allows for the recording of changes in both values and vertices across time. These patterns show a sequence of the connected subgraph that meets the input data's frequency and non-redundancy requirements. The multi-attribute values of nodes and edges that give more accuracy to the results and enhance the visualization method. To investigate the improvements indicated in this study, an analysis referring to the work (Feng et al., 2018) is carried out on crime data. The use of a dynamic attributed graph for crime visualization and analysis helped to discover hidden patterns and trends of crime, which is missing in previous studies. The information gathered could support the formulation of crime prevention plans at both operational and strategic levels. From the discussion on previous researches in the crime analysis field it has now been established that the introduction of data visualization and analysis on dynamic attributed graph supports Criminologist and law enforcers to address violence in various paradigms.

The research further moves onwards to design the objectives of temporal analysis for crime networks at San Francisco. Right now, the designed objectives are focused on one city only however, these can be further extended to investigate other cities as well with high concentration of crime rate at different timestamps: daily, monthly, fortnight etc. The designed objectives are stated as:

(i) Ranking of police districts and crime categories based on the number of cases incidence and cases resolved.

(ii) Detection of the significance of crime categories for different districts.

Few o*bjectives are especially designed based on the crime patterns and trends to identify*:

(iii) Police districts that resolved maximum cases and the crime categories that get mostly resolved.

(iv) Police districts where most of the crimes occur in the same zone.

(v) Police districts where weekend crime cases are most.

(vi)Strongest or weakest category of crime for any police district.

Table 1. Comparative analysis of latest related work.

Research	Purpose	Techniques	Findings	Research limitations
Bailey al., 2020	To identify the central tendency and dispersion of knife offenses reported to the police using social network analysis of population-level data	Social network analysis in R modelling	Rather than gang-related criminals, prevention measures should target either frequent violent offenders or recurrent victims known to the police—and therefore more vulnerable to knife crime exposure.	The dynamics of data are neglected while analyzing static data.
Launtz et al., 2020	Effects of Gender, Age, and Group Size on Co-offending violence	Statistical analysis on tabular data	shows that group composition is important to consider in the situational generation of group violence.	Limited analysis on data, group characteristics are neglected.
Elezaj et al., 2019	Analysis of online social networks to assist criminal investigators in the investigation and prevention of crime	Graph Analysis with database connected to a hybrid ontology	A knowledge graph-based approach for collecting digital evidence from online social networks.	Not based on real-world facts and use scenarios.
Wu et al.,2020	Using dynamic graphs, provide a representative summary for numerous videos with good variety	Graph models, video summarization tools.	Designing and analysis of a multi-video summarization convolutional network in the form of a dynamic graph	Video node attributes and comments were not considered.

Yamini et al., 2019	To forecast the high likelihood of crime area by displaying crime analysis in several states around the United States	Data Mining	Multiple clustering techniques based on fuzzy clustering theory are presented to assess the crime-prone states in the United States so that it may be halted by increasing security levels in such areas.	Limited analysis to number of crime cases, with no study of crime patterns or trends.
Kurshan et al.,2020	Graph Computing for Financial Crime and Fraud Detection	Machine learning, Graph neural networks	In real-life transaction processing and criminal detection, graph-based detection algorithms are used	The ever-changing nature of fraud and the complexity of digital transaction processing systems (such as large-scale implementation requirements, real-time processing, multi-channel updates, complex data/graphs) are neglected.
Wajahat et al., 2020	Analyze participation of different communities in an online social network.	Statistical Analysis on Facebook data	To retrieve information for community involvement in social networks, both directed and undirected graphs were examined.	The category of the relations and the characteristics of the users are not taken into account.
Park et al.,2020	Social Network analysis for Global Transshipment	Network Analysis	The inclusion of information on known offenders enables the discovery of criminal behavior within the global transshipment network for improved law enforcement, and network analysis provides insight into how transshipment networks behave	Analysis based on static data and does not take into account how crime encounters evolve.

3. Notations and Definitions

The focus of the present research is on the visual analysis patterns on the occurrence of crime using dynamic graph premeditated on multiple attributes that specifies a constant stream of change. At any timestamp ‘t’ the designed graph has different attribute values of nodes and different weights on edges depending on the features that are selected for analysing the crime data.

Definition 3.1.1: Time Series of Graph (Lahiri et al., 2008): For a given sequence of T graphs G = {G1, G2,....,GT} with Gt = (V_t, E_t, W_t), where 1<=t<=T, V_t is the vertex-set, E_t is the edge-set and W_t(weight on edges) is the weight-set of the graph at a timestamp t". We define G as a time series of graphs which can be transformed into weighted dynamic network.

Definition 3.1.2: Dynamic Attributed Graph (Cheng et al., 2017): A dynamic attributed graph G = (G_{t1} , G_{t2} , ..., G_{tn}) depicts a graph's development over a period of time T = {t1, . . . , tn}. G's vertices are referred to as V. A collection of attributes A is assigned to each vertex (numerical or categorical). A domain value Da is assigned to each attribute a ∈A. $G_t = (V_t, E_t, \lambda_t)$ is an attributed undirected graph for each time t ∈T where $V_t \subseteq V$ is the collection of vertices at time t, $E_t \subseteq V_t \times V_t$ is the collection of edges at time t and λ_t: $V_t \rightarrow 2^{AD}$ is a function that assimiliates each vertex of V_t with values $AD = \cup_{a \in A} (a \times Da)$.

Definition 3.1.3: Attributed Sub-Graph (Cheng et al., 2017): -An attributed subgraph G' = (V', E', λ') of a graph G = (V , E , λ), represented G' ⊆ G , iff V' ⊆ V , E' ⊆ E , and for all v ∈ V' : λ'(v) ⊆ λ (v). G' is a connected attributed subgraph of G , as $G' \subseteq_{conn} G$, iff G'⊆ G and for every pair of u, v ∈V , there exists a relation between u and v in G.

Definition 3.1.4: Weighted Directed Dynamic graph (Duan et al.,2009): A dynamic network that is weighted and directed G = (G_1, G_2,..., G_T) is a graph sequence in which each graph Gi appears at timestep t_i for I = 1, 2,..., T. $G_i = (V_i, E_i, W_i)$ is a simple weighted directed graph with vertex set V_i and edge set E_i with weight w. E is the universal edge set of G, where $E = E_1 \cup E_2 \cup \ldots\ldots E_T$. In the same way, a universal vertex set is $V = V_1 \cup V_2 \cup \ldots\ldots V_T$.

Definition 3.1.5: Weighted Directed Bipartite Network (Sinclair et al., 2004): Weighted directed bipartite graph is defined as G (L, R, W, E) where L and R are two distinct sets of vertex such as L ≠ R and E is the edge list containing tuples (l_i , r_j , w_{ij}), where w_{ij} ∈W is the weight between nodes l_i ∈L. and r_j ∈R.
*Definition 3.1.6: Bipartite clique (Bipartite complete graph) (Zhang et al., 2008): A bipartite graph G (L, R, W, E) is called complete bipartite graph or biclique if each node l_i ∈L are connected to all nodes r_j ∈R such as |G|= n*m where |L| = n and |R| = m.*

4. Methods

4.1 Workflow illustration for the proposed visual analytics approach

The workflow of the method is illustrated with the assistance of four subsections that clarifies the complete process (see Figure 1). Firstly, the related dataset is identified and scrutinized for desired features and removal of insignificant data as related to the analysis. The data comes from publicly available criminal activity datasets from San Francisco. The San-Francisco crime data includes 2,142,685 crime events from the period of 01/01/2003 to 11/08/2017 and further approximately 4,56,071 crime incidents from 01/01/2018 to 08/05/2021. Both the datasets are merged to create a combined crime dataset from 2013 to 2021.The combined dataset is then pre-processed for better analysis that covers discretization of time series and normalization of data. Further the dynamic network is generated according to the timeline in proper layout for better insight of network. After the generation of dynamic network – various visualization and analysis techniques are applied to repossess valuable statistics from the designed network to attain the anticipated objectives.

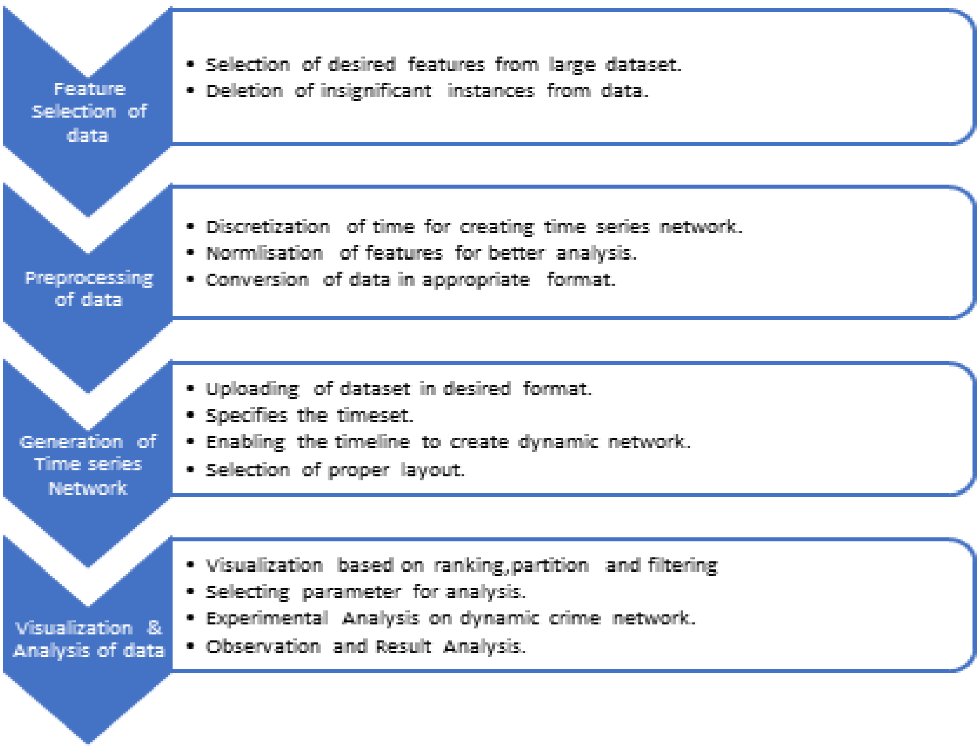

Fig 1: Workflow of dynamic crime network

4.2 Feature Selection

Every data record in the observed data set foreshows a crime and each data record has the following features:

- Dates - Date and time of the crime incident.
- Category - Type of the crime.
- Description - A brief description of any relevant details of the crime.
- DayOfWeek - Day of the week when the crime occurred.
- PdDistrict - Police Department District where the crime is incident.
- Resolution - Whether the crime incident gets resolved or not.
- Neighbourhood -Number of neighbourhood cases that occur.

Dataset from 2018 onwards contains additional features like Analysis Neighbourhood, Block Number, CAD Number, Filed Online, Incident Datetime, Report Datetime, Report Type Description report, Incident Subcategory, Supervisor District etc. These features are kept reserved and is intended to be used for future analysis.

For the present analysis a subset of feature is identified that are found most relevant to achieve the designed objectives of dynamic crime analysis, the subset mainly consists of features like category of crime category, police district, day of week, number of cases resolved and average neighbourhood cases.

4.3 Pre-processing of data

Before generating a dynamic network for visualization and analysis purposes, a series of pre-processing steps were performed to get better results.

- Discretization of data in timestamps for analysing and forecasting the overall trend of data.
- Data is normalized were needed to get better results.
- The timestamp indicates the date and time of each incident. The data is aggregate yearly as data is vast.
- Obliteration of undesirable features like incident code, coordinates.
- Collapsing of similar types of criminal activities in one category.
- Deletion of insignificant crime activities concerning their impact on society.
- Conversion of desired and filtered data in Gephi format with the dynamic timestamp.

4.4 Generation of Time series Network

The dataset gets converted in graph format with the identification of nodes and edges. A node specifies the entity of the dataset where the edge shows the relationship between different entities. All the nodes and edges contain a timestamp as the desired network evolves with time. To simplify the process, the data is aggregated in years and the aggregated values are allocated as attributes to nodes and edges. A timeline is then enabled to get a dynamic crime network. The data table in the next section describes features of a dynamic crime network.

4.4.1 Generation of Time series Network

To convert tabular data into a graph format, a set of nodes and edges is identified. Nodes are the entities that specify the domain of a network and edges show the relationship between the nodes. Two tables are designed for the data laboratory, namely- a node table and an edge table that define the features that are associated with nodes and edges respectively.

A) Node Table It contains information of nodes like node id, label, timestamp, and attribute list. Node ids are required to discriminate between each node, where the labelling is for naming police districts and crime categories. The timestamp is the additional feature that provides the dynamics to the graph. The timestamp contains information of both the date and time at which the respective node is present in the network. The attribute list contains the list of attributes of each node and their values at each timestamp. An attribute value of the node contains the information that can play a significant role in analysing the network and retrieve some meaningful information to meet objectives.

Table 2 shows the node table of the present dynamic visual analysis process where entities get divided into two sets - set one represents police districts having ten nodes: Bayview, Central, Ingleside, Mission, Northern, Park, Richmond, Southern, Taraval, and Tenderloin. Another set contains five nodes for crime categories: Larceny/theft, Robbery, kidnapping, Disorderly conduct, and other offenses.

Table 2 Node table of Dynamic crime network

Dataset	San Francisco Crime dataset 2003-March 2021
Node Type 1	Police District
Node Type2	Crime Category
Number of nodes	10(Police-District),5(Crime Category)
Timestamp	Year
Attribute-1	Number of cases-Integer
Attribute-2	Number of resolved cases-Integer
Attribute-3	Number of weekend crime cases-Integer
Attribute-4	Average neighbourhood crime cases-Float

Data aggregated year-wise starting from 2003 to March 2021 and a yearly timestamp is to discretize the data. Police District nodes also contain attributes like the total number of cases incidence, totally resolved cases, cases that occur on weekends, and an average of neighbourhood cases in a particular year.

B) Edge Table An edge in the graph specifies the connection between two distinct nodes and stipulates the relationship between these nodes. Edges can be unidirectional or bidirectional. A bidirectional edge contains the information of the mutual relationship between two nodes. To study a bidirectional relationship between two entities will be significant to analyse and retrieve more information from the known data. Table 3 shows the edge table that contains information on all the edges. Edges also get divided into two sets - Set 1 demonstrates the relationship from the crime category to police districts, where set 2 illustrates the reverse relationship. All edges are directional and weighted where weight specifies the strength of the connection between police districts and crimes. Edge from crime category to police districts resembles the total number of cases of that crime category incidence in police district in a particular year, opposite edge (from police district to crime category) reflects how many of these cases get resolved in that year.

Table 3 Edge Table of Dynamic crime network

Network	Directed, Weighted
Number of edges	100(50 of each Type)
Edge Type 1	Crime cases incident in Police districts
Edge Type 2	Crime cases resolved by police districts
Source-Type 1	Crime categories
Target type 1	Police districts
Source Type 2	Police Districts
Target Type 2	Crime categories
Timestamp	Year
Weight type 1	Number of Incidence cases
Weight Type 2	Number of Resolved cases

4.4.2 Layout Selection for better visualization

After uploading the data and generated a time series network, a layout algorithm is selected to provide the best outlook to the premeditated graph that can be better visualized and is further used for analysis. Frutcherman-Reingold the algorithm is a force-based method that works on the principle of gravitation and repulsion by extending the network farther or closer. Gravity provides better visibility by putting a graph in the middle of the network. The setting chosen will again depend heavily on the density and structure of the data set. Another feature is speed to accelerate network convergence at the cost of higher accuracy.

Figure 2 shows the resultant graph for the crime network, nodes are labelled as for selected police districts and crime categories.

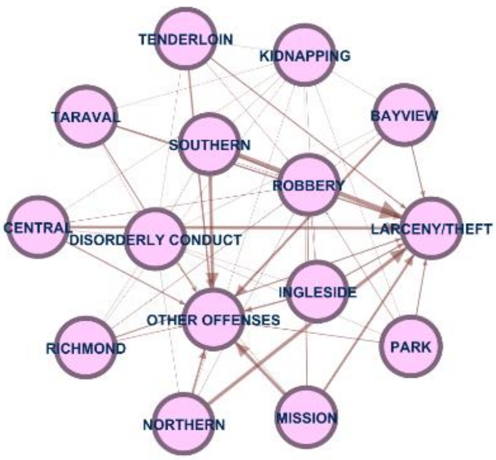

Fig. 2: Overview of dynamic crime network

4.5 Visualization of dynamic crime network

To get a better insight into network visualization and analysis plays a vital role. Where better visualization makes the complex networks look presentable and understandable, analysis retrieves the hidden information from the graph can be used for better decision making. This section covers both these aspects and the parameters on which dynamic crime network visualize and analysed.

Visualizing the variations in dynamic networks based on ranking, partition, filtering, and spatialization helps understand the network more. As both the nodes and edges in a dynamic crime network contain crucial information, it will be beneficial to visualize the variation in nodes and edges based on their attribute values. The following subsections describe how visualization based on different features helps to have more insight into the crime network.

A) Ranking of nodes colour and size based on attribute value: Nodes are the entities of the graph connected with other nodes through some relationship. Most of the studies in dynamic graph theory are focused on the visualization and analysis of their relationship. But the node also contains meaningful information in the form of attributes, which becomes more significant to visualize and analyse the variation in nodes both in the aspect of colour and size based on attributes value. The ranking process discriminates the nodes in size and colour based on the intensity of the information stored. For example, Figure 3 illustrates the scenario of 2020 where police districts are ranked based on the total number of cases incident in that year. The numeric value associated with each district node is the attribute value: the total number of crime cases in that district in 2020. The intense colour and bigger size resemble the higher number. The Central district ranked highest with the highest number of cases of 11220, where the Northern area is second highest with the approx. value of 10500. Park district was the least ranked with 2476 crime cases. This process can implement into any attribute: number of cases resolved, weekend cases, and average neighbourhood cases of the year. This feature helps us achieving objectives (i) and (iii) for ranking police districts and crime categories based on parameters.

B) Ranking of edges colour and width based on attribute value: The edge in a graph resembles the relationship between two distinct nodes. An edge present at any timestamp represents a relationship between its corresponding nodes at that instance of time. The direction of the edge specifies the flow of information between nodes. Most studies are on the presence and absence of connection between nodes and not on the attribute value of edge. As edge also contain crucial information as an attribute or weight that signifies the strength or nature of the relationship between nodes. Bidirectional edges specify the mutual relationship between nodes: to and fro relationship. The study of these edges can undoubtedly lead to better retrieval of information. In a dynamic crime network, edges are bidirectional. The weight on edges directed from the crime category to police districts represents the total cases of that crime incident in a particular year, whereas the value of the opposite edges shows the number resolved of that crime in that year. Figure 4 shows the scenario of 2021 till 31st march. The ranking is done based on the above parameters where intense colour and wider edges show that cases of Larceny/theft are much higher than other categories of crimes with the highest number in the northern district of 5819 and central districts with 4685 larceny/theft cases, whereas light colour and weaker edges show that kidnapping and robbery cases are low in all areas. This feature helps us in achieving objectives (i), (iii), and (vi) to identify areas and crime categories based on desired parameters.

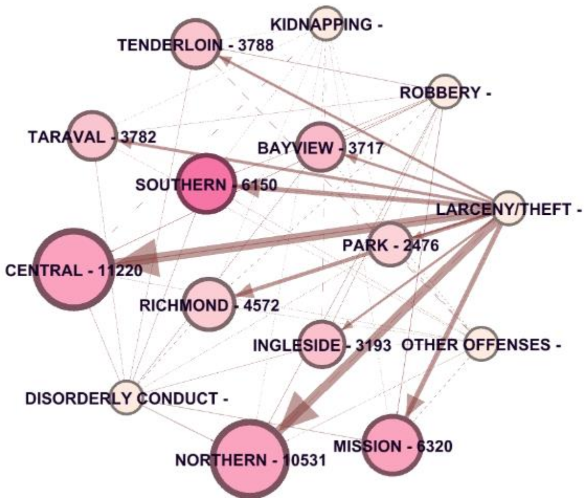

Fig. 3: Ranking of nodes (police districts) based on total number of crime cases in 2020.

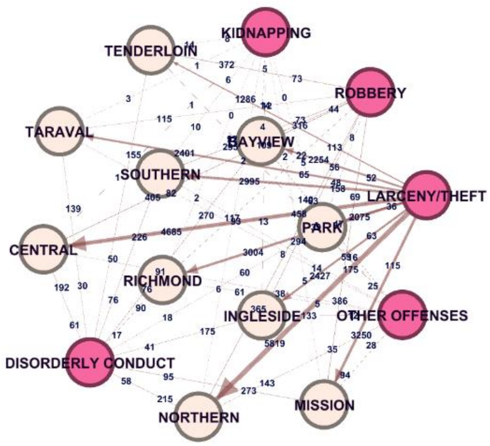

Fig 4 Ranking of edges based on crime cases of a particular category in 2021 up to march.

C) **Dynamic filtering of graph**: This process gets utilized eventually to visualize a selected portion of the graph and get a better insight into the dynamic network using a range of criteria. Most of the time, the network is vast, dense, and hard to navigate entirety. In such cases, filtering provides the necessary options to search the graph effectively for specific attributes or features. Multiple filters can be applied where more than one condition is nested. Filtering based on the attribute value of nodes and edges or for edge type, weight, mutual-edges, or self-loops. Queries designed to filter dynamic graph-based functions like equality, inter-edges, intra-edges, partition, partition-count, and range of attributes. It is a process to filter out desired information from the graph-based node information or edge information. The dynamic interval can also be applied to filter out dynamic data from the network. This feature will help meet objectives (iv) and (v) to filter out areas based on neighbourhood and weekend crime cases.

4.6 Analysis of dynamic crime network

To analysis the dynamic graphs various methods and metrics are essential to address the temporal feature of these networks. The most common technique to cater with the temporal aspect is to simply divide the entire network over time into a series of static subgraphs by discretization of dynamic graph (Zaidi et al., 2014). For example for a given graph G in a time period from t(1) to t(m-1), m different subgraphs such as G1,G2.---Gm represents the state of graph at timestamp t1,t2,…..up to tm-1.G1 represents the state of the graph in time interval t1 to t2.Similarly,G2 represents the state of network in time interval t2 to t3 and so on. The term timestamp is used to refer to individual state of graph (subgraph) according to time periods sampled from G. Such discretization of the dynamic graph into a series of static subgraphs can enable the implementation of many statistical analysis metrics to be applicable on dynamic networks (Bailey et al., 2020) (Wang et al., 2016). In the present research, statistical analysis is performed (Chessa et al., 2014) to find pattern, modularity, page rank and weighted degree of the network and explains the importance of this analysis for the better understanding of the process and retrieval of innovative information from the network.

4.6.1 Pattern analysis

Patterns are the distinctive formations created by the connection of nodes in a graph that lay the foundation of technical analysis. There can be multiple patterns in a dynamic network: complete graph, clique, periodic, regular, bipartite graph, and many more. Each type of pattern is significant and helpful in analysing and retrieving useful information from the network. The dynamic crime network bisects into two sets of nodes then there are chances that the pattern can be bipartite in nature. For further analysis, the degree of each node gets calculated proceeds by calculation of network diameter. A multimode projection (Jaroslav, 2014) is then implemented on the graph to check the exact pattern of the network.

A) Degree: The degree of a vertex vi is a total number of edges incidence on the vertex and denoted as deg(v_i). In undirected graphs, each edge contributes two degrees hence the total degree of the network is always twice the number of connections present in that graph. In the case of a directed graph, the degree of a vertex vi gets divided as in-degree(in-deg(v_i)): the number of edges incident on that vertex and as out-degree(out-deg(v_i)): the number of edges coming out from that vertex. The degree of a vertex in a directed graph is the sum of the in-degree and out-degree. In the n-regular network, the degree of each vertex vi should be equal to the degree of each vertex in a clique with n number of vertices should be n-1. For an undirected bipartite graph G (U, V, E), the degree sum of nodes for both sets should be the same and equals to the total number of edges present in the graph. Equation 1 (Wang et al., 2016) gives the degree sum formula for a bipartite network:

$$\sum_{i=1}^{U} \deg(u_i) = \sum_{i=1}^{V} \deg(v_i) = |E| \qquad (1)$$

Thus, degree is a useful feature to analyse pattern of a graph. A complete bipartite-graph or bi-clique, with |U|=m and |V|=n can be denoted as $K_{m,n}$ and degree of each vertex u_i of U equal to number of vertices in set V and vice-versa .Further for a bidirectional complete bipartite network this number becomes double as one edge contributes two type of degree i.e. in-degree to one node and out-degree to another node to which it is connected. Equation 2 and 3 (Alzahrani et al., 2019) gives the degree value:

For bipartite bidirectional clique network, $K_{m,\,n}$;

$$\deg(u_i) = 2n \text{, and} \qquad (2)$$
$$\deg(v_i) = 2m \qquad (3)$$

The dynamic crime network is consisting of the relationship between two types of nodes where one type represents the police districts another one is for crime categories, so to check whether the network follows the bipartite pattern, the dynamic degree gets calculated. Figure 5 shows the degree distribution of dynamic crime network that shows nodes belonging to a set of police districts have degree twice the number of crime category nodes where the degree of crime nodes is twice the number of police districts with an average degree of nodes is 6.667. Based on equations 2 and 3, it concluded that the crime network is following a bipartite clique pattern with weights on edges. This conclusion can help further analyse the bipartite modularity to detect the significant communities and helps to meet the objective (iii).

B) Network Diameter

Network diameter is the longest graph distance between any pair of nodes. The graph distance is the average graph distance between any pair of nodes. The calculated diameter of the dynamic crime network is 2, having an average path length is approx. 1.52. This value also indicates that the longest distance from any node is equal to 2 and for a bidirectional graph these values support the pattern of bipartite.

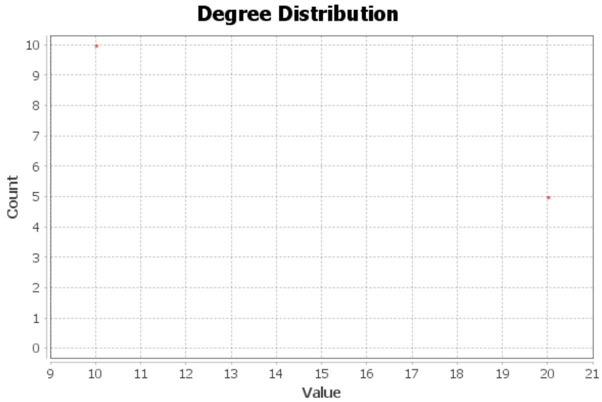

Fig. 5: Degree distribution of nodes in dynamic crime network.

Further, graph distance for any node is calculated based on betweenness centrality (Sinclair et al., 2004), closeness centrality, and eccentricity.

Betweenness Centrality In graph theory, betweenness centrality a measure of centrality in a graph based on the shortest paths. For each vertex, it is the number of shortest paths that pass through the vertex. In a directed bipartite graph, betweenness centrality gets divided into two values and where the same value indicates that these nodes belong to the same set. Figure 6 shows the normalized betweenness centrality value of all nodes. The nodes representing police districts have a value of 0.01, where crime nodes have a value of 0.09. It indicates that crime nodes are more connected than police district nodes which hold for our network.

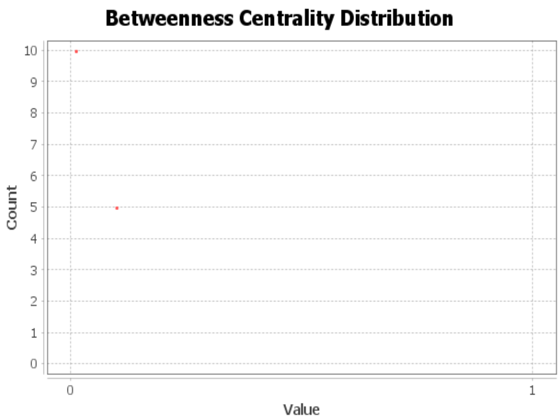

Fig. 6: Betweenness centrality distribution of network

Closeness Centrality It is also a measure to calculate the graph distance of a vertex. It is the average distance from the starting node to all other nodes in the network. For a bipartite network, this measurement gets divided into two values. Figure 7 shows the closeness centrality distribution of the network. It follows the behaviour of betweenness centrality: police district nodes have a closeness centrality of 0.60, where crime nodes have a value of 0.77. This measure also supports the bipartite clique nature of the network.

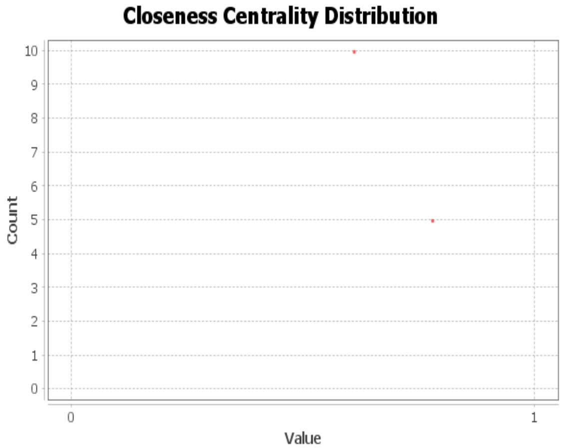

Fig. 7: Closeness centrality distribution of network

Eccentricity It is the distance from the starting node to its farthest node in the network. For a bidirectional bipartite clique, its value should be 2 for each node. Figure 8 shows the eccentricity of the dynamic crime network. All the nodes exhibit value 2. This calculation supports the theory that this pattern is of a complete bipartite graph or bipartite clique. This analysis helps us to get a better insight into detected communities required for objective (ii).

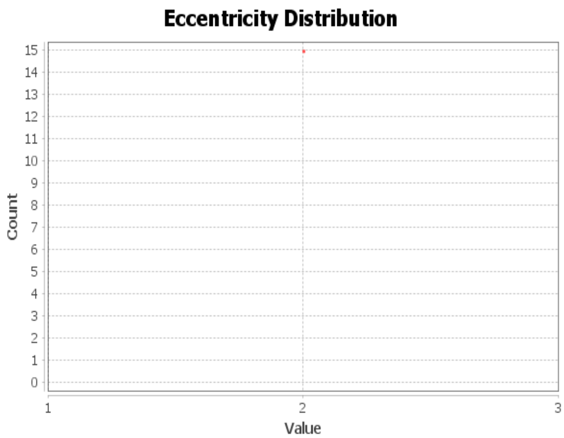

Fig. 8: Eccentricity distribution of network

C) Multi-mode Network Transformation of network After analysing the pattern, the next step is to visualize the graph in the form of the bipartite to understand the results of further analysis. Most of the time, such patterns in a network are detected but not visualized. A multi-mode network transformation method (Jaroslav, 2014) got applied for the projection/transformation of the network as bipartite. Based on the matrix multiplication approach, it allows different types of projections. This process helps project a network in bipartite or multipartite form based on the number of node types or any other attribute.

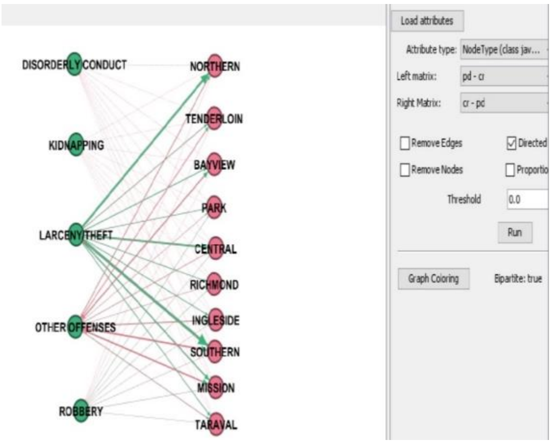

Fig. 9: Projection of dynamic crime network as bipartite network.

Figure 9 shows the result of multi-mode transformation for bipartite is correct, and the resultant graph is bipartite pattern after graph colouring of nodes based on the node type. Graph colouring mode for bipartite pattern comes true and nodes of the same set(type) have shown in the same colour.

4.6.2 Modularity

Modularity is a measure to find the strength of dividing the network into modules. In a network with a high degree of modularity, the connections between nodes in the modules are dense, and between nodes of different modules are sparse. For a weighted network, the value of modularity depends on the edge weight connecting the nodes and is in the range (-0.5 to 1). That measures the relative density of edges inside communities for that outside communities.

By finding the modularity score using the Louvain method (Blondel et al., 2008) of evolution graphs at different timestamps and plotting them, we find that at almost all discrete timestamps, the value of modularity of the network is on the lower side (<0.5). That leads to the conclusion that the graph is robustly connected and the nodes inside the community have a dense connection, but it is not that sparse outside communities. It indicates that crime categories are majorly evenly distributed in all police districts. It becomes more interesting to analyze such communities to identify the area-wise significance of a particular crime that will not be possible by tabular data. And as the network is following a bipartite pattern throughout the time series, it also indicates that two nodes of the same type have no connection with each other even if they belong to the same community. This analysis of variation in modularity to time will help to meet the objective (ii) to find the time-based significance of crimes in a particular area/district.

4.6.3 Page Rank

The rank of the page is a salient element of the dynamic network node and online pages. The score of every node is always a non-negative float number where all the scores will sum to 1(Engstrom et al., 2018). It may be expressed as a percentage sometimes. The score of a node depends on its links with other nodes. Inlinks and outlinks of the node are the linkages coming towards that node and going outside from that node. It is a variant of eigenvector value and can be utilized successfully for directed networks. The damping factor to which a node can link to other nodes gives probability Epsilon is the stopping criterion value that determines the stopping point. Equation 4(Brin et al., 1998) specifies the page rank calculation for finding all the other nodes linked to a given node. Assume node A has T1…..Tn nodes linked. These nodes will provide the targeted node a proportion of their outgoing links such as:

$$PR(A) = (1-d) + d\left(\frac{PR(T1)}{C(T1)} + \cdots \frac{PR(Tn)}{C(Tn)}\right) \qquad (4)$$

Where, PR is the page rank of nodes connected to targeted node, d is damping factor set between 0 and 1, and C(T) are the number of outgoing connections the ‘other nodes’ have.

Page rank also considers the weight of the edges. The node ranked higher has a high proportion of in-links weights compare to the total weight of the links on that node. This analysis helps us meet the objective (iv) as the in-links weights determine the resolved number of cases of a particular crime category

4.6.4 Weighted degree

The weighted degree of a given node is the sum of the weights of the edges connected to that node. The weighted degree can further get divided as a weighted-indegree and weighted-out degree. The Weighted-in degree of a given node is the sum of weights of edges incident to that node, where the weighted-out degree is the sum of weights of edges coming out from that node. The weighted degree helps to identify the strength of the connection between nodes in a network. For a bipartite network, the weighted degree of a given node specifies the solidity of relationship it has towards the nodes of another set. The analysis of varied values of weighted-in degree and the weighted-out degree to time helps to meet the objective (iii) to identify the time-based ranking of police districts and crime categories based on the number of cases incidence and number of cases resolved.

4.6.5 Spatial Analysis

Spatial transformation converts the network into a map, where all of the nodes are place according to some metric. For the weighted network, the weight of the edges used as a metric. This metric specifies the location of a node in a map based on the strength of its connecting edges. Force-based algorithms like Force Atlas, Force Atlas 2(Jacomy et al., 2014), and OpenOrd(Martin et al., 2011) consider the edge weight in placing the nodes nearer or farther in the network.

A) Force Atlas 2 It is a force-directed approach that follows the principles of repulsion, attraction, and gravity to deliver a high gradation of accuracy for small to large datasets. It works on the principle of attraction and repulsion. A higher level of attractiveness will compose the nodes, where lower attracting nodes place at a distance. The force-based process has multiple features like gravity, scaling, and overlap prevention. Apart from these, it gives a layout that is more clustered and visually more vibrant. Its advanced value of gravity assistances in towing the nodes towards the center of the graph, and subordinate merit supports the dispersion of the nodes. The scaling feature sets a revulsion level for the diffusion of the network to make it better decipherable. The prevent overlap helps to avoid the overlying among the nodes. This algorithm is the finest suitable to deliver an improved outline of the graph.

B) OpenOrd This algorithm helps to generate network graphs with high speed and is considered best for vast networks. This approach is based on the Fruchterman-Reingold algorithm (Fruchterman et al., 1991) but is much faster than later. In cases of thousands of rows of data, this algorithm enables the speedy transformation of the network structure. It works by allocating the time to five stages: Liquid, Expansion, Cooldown, Crunch, and Simmer. This process also works on the levels of attraction and repulsion to locate nodes based on varying parameters to optimize results. Spatial analysis of nodes at different time stamps helps to achieve the objective (vi) to identify the strongest and weakest category of crime in any area at points in time.

5. Results

Visualization and analysis of dynamic networks on different parameters help get a better insight into the network. New information can infer which is difficult and mostly out of scope for the tabular data. Based on the parameters we chose to visualize and analyze the dynamic crime network following results are achieved that helps to meet the objectives of this study.

5.1 Ranking of police districts and crime categories based on the number of cases incidence and cases resolved

Analysis of weighted degree helps ranking police districts and crime categories according to the number of cases that occurred and resolved. Ranking of police districts gives the idea of incident cases in a year, where the order of crime categories shows the resolved cases of that crime. Two timestamps are selected: 2014 and 2020 to find how the statistics change over time. In weighted degree distribution of nodes in 2014, an average degree calculated as 5320.40. The highest weighted degree (including both weighted-in degree and weighted-out degree) is approx. 35000. Figure 10 shows the variation in node size based on the weighted-in degree.

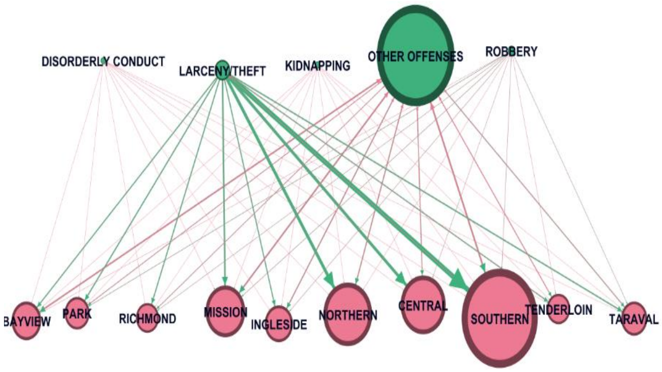

Fig. 10: Variation in node size based on weighted-in degree in 2014.

The cases incident in the southern district were highest, then in northern followed by central district. Park and tenderloin are the areas with the least number of cases. Also, maximum cases of other offenses are resolve in 2014.

Figure 11 shows the weighted-out degree distribution of all nodes. It represents Larceny/theft as the crime category that occurs most in 2014, followed by other offenses. The Southern District is the one that solves maximum cases.

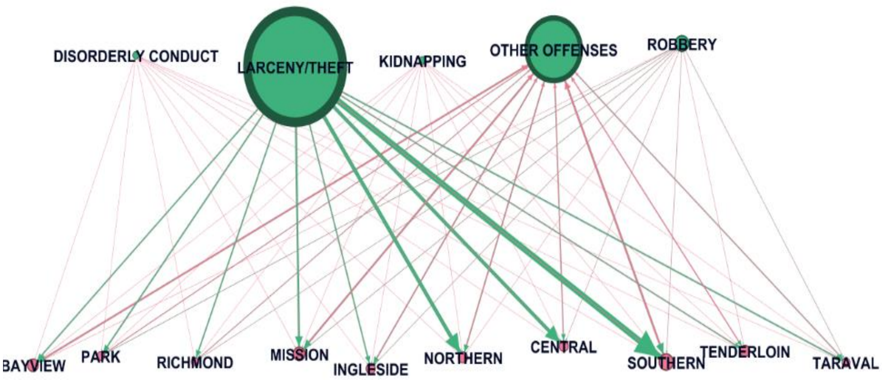

Fig. 11: Variation in node size based on weighted-out degree in 2014

Figure 12 shows the weighted degree distribution of 2020 with an average weighted degree is 2556, and the highest weighted degree is below 10000, which shows crime cases are marginally less in 2020 compare to 2014. Figure 13 shows the weighted-in degree distribution in 2020. The Northern District has the highest crime cases incident, followed by Central and then Mission.

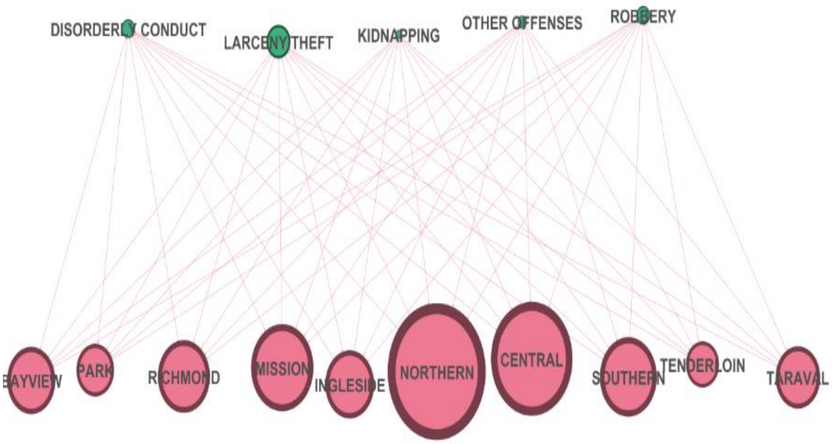

Fig. 12: Variation in node size based on weighted-in degree in 2020

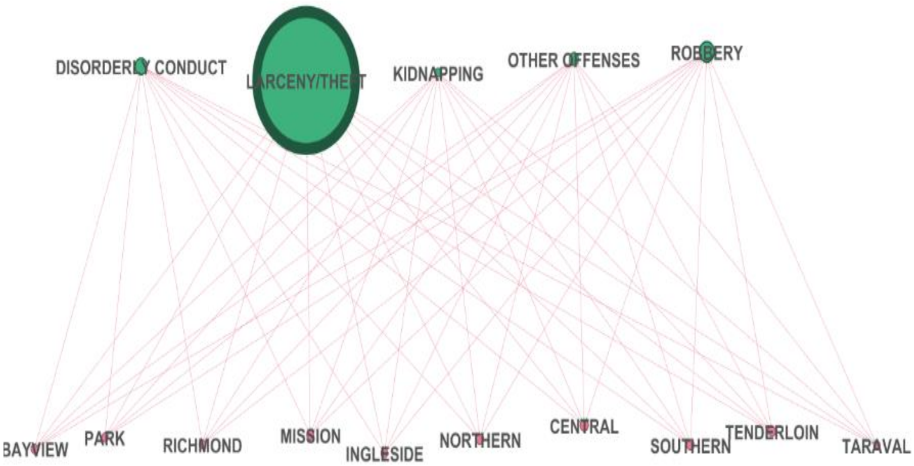

Fig. 13: Variation in node size based on weighted-out degree in 2020

Figure 14 shows the average weighted degree chart of all police districts nodes in dynamic crime network for both incident cases and resolved cases over a period of last 15 years.

It shows that the ratio of resolved cases are much lower than incident cases and southern district had much higher number of cases among all districts of San

Francisco city.

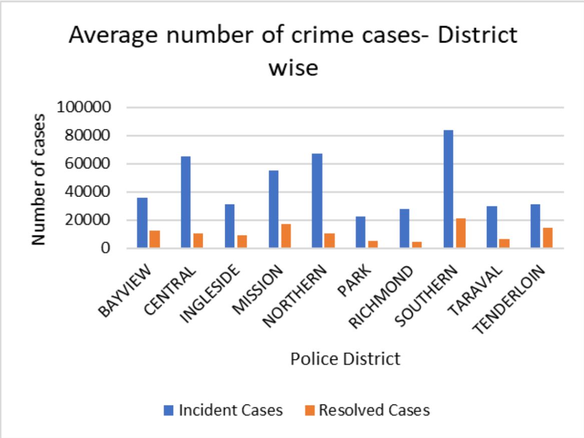

Fig. 14: Average weighted degree chart of dynamic crime network for last 15 years

5.2 Detection of the significance of crime categories for different districts

Analysis of weighted degree helps ranking police districts and crime categories according to the number of cases that occurred and resolved. The dynamic graph can facilitate us to detect communities that signify which crime category is significant in which area. For this analysis, we compare the modularity of the network at three timestamps: 2010, 2016 and 2020. The modularity score for every year is below 0.5. It shows that edges inside the modules are not that densely connected compare to outside edges. The reason is the bipartite nature of the crime network. Still, the detected communities give a better insight into crime significance in a particular area. Figure 15 shows the two communities detected in 2010 illustrates that Larceny/theft are more significant in Richmond, Northern, and Central district, whereas the rest of the crime categories are consequential in other areas. T.

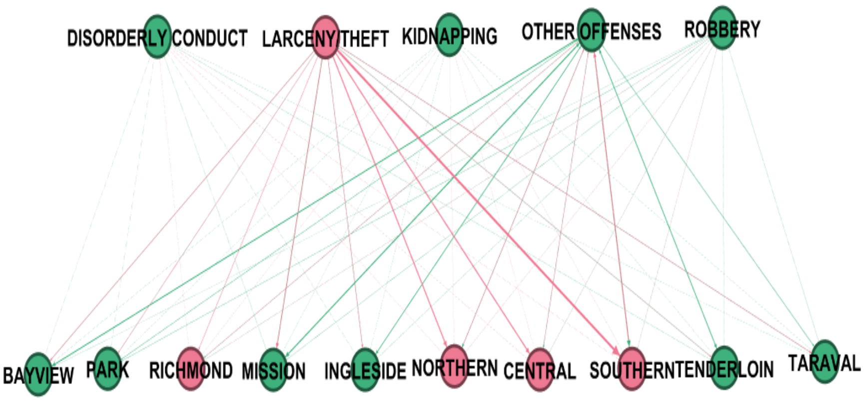

Fig. 15: Community distribution in 2010

Figure 16 shows three communities detected in 2016. It illustrates that Disorderly conduct and Larceny/ theft are significant in the Central, Northern, and Southern areas, where robbery is more prevalent in Park and Richmond. The remaining crime categories are notable in other areas.

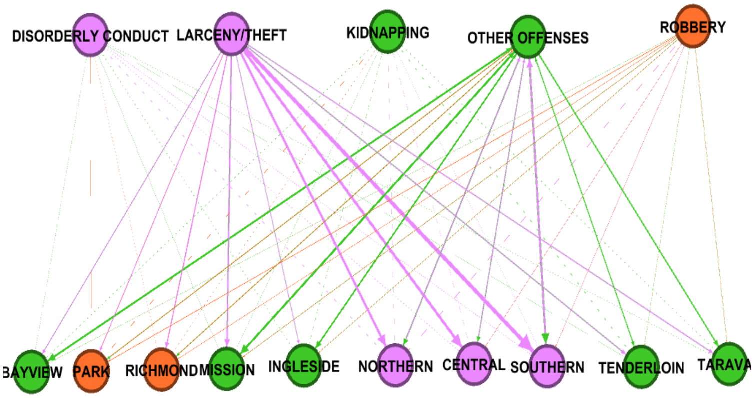

Fig. 16: Community distribution in 2016.

Figure 17 shows three communities detected in 2020 according to which Larceny/Theft is the most occurring crime in most areas except Ingleside and Tenderloin, where robbery, kidnapping, and other offenses are notable, and Bayview where disorderly conduct is prevalent. The modularity score for 2020 is better than 2010 and 2016, which signifies that the impact of a particular crime in areas of the same community is more.

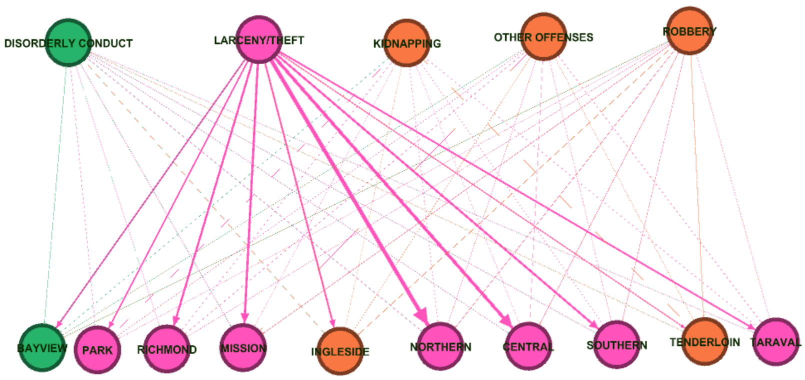

Fig. 17: Community distribution in 2020.

Figure 18 shows the modularity chart of dynamic crime network for last 15 years that helps to identify the strength of bonds between nodes over the period of time. It represents that the modularity of the dynamic crime network gets increased over time.

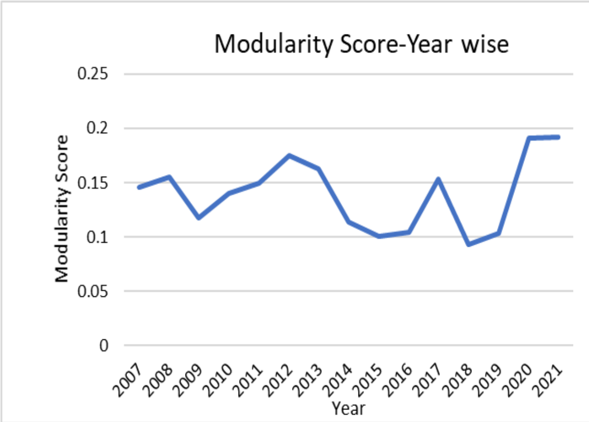

Fig. 18: Modularity chart of dynamic crime network for last 15 years

5.3 Identification of police districts that resolved maximum cases and the crime categories that get mostly resolved

By analysing the nodes for in-links and out-links by considering weights on the edges then calculating page rank that gives the proportion of in-links to the out-links of a given node, we can identify the police districts that resolved the greatest number of cases and the crime categories that are mostly resolved compared to the occurrence of that category. Again, we considered two-time stamps: 2014 and 2020 to show the variation in crime behaviour and its resolution. Figure 19 shows variation in size of nodes based on page rank value in 2014. Other offenses with the highest page rank are mostly resolved crime categories among all crime categories, and the southern district has settled the maximum number of cases with the highest page rank among all police districts.

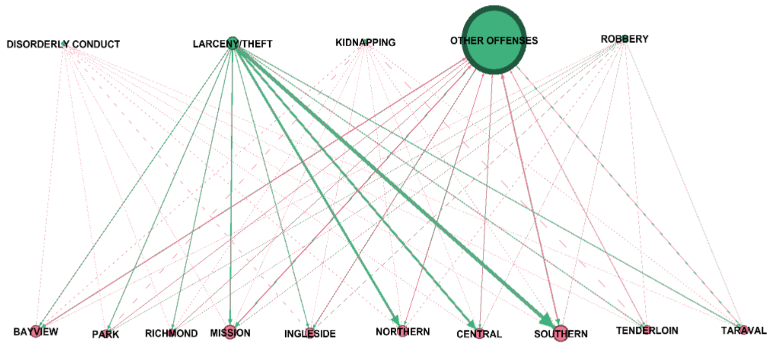

Fig. 19: Variation in node size according to page rank in 2014

Figure 20 shows the variation accordingly in 2020, where Larceny/theft is the most resolved crime category with page rank followed by disorderly conduct with page rank.

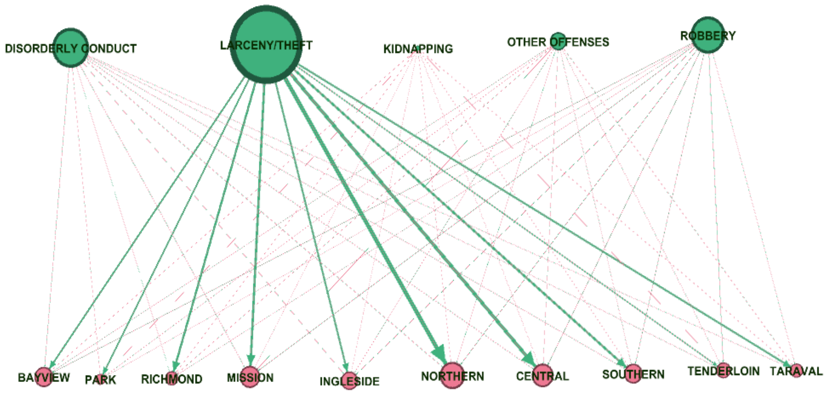

Fig. 20: Variation in node size according to page rank in 2020

Northern is the district that solved the maximum number of cases with highest page rank between police districts. Fig 21 shows the average page rank chart of police districts and crime category nodes over the last five years.

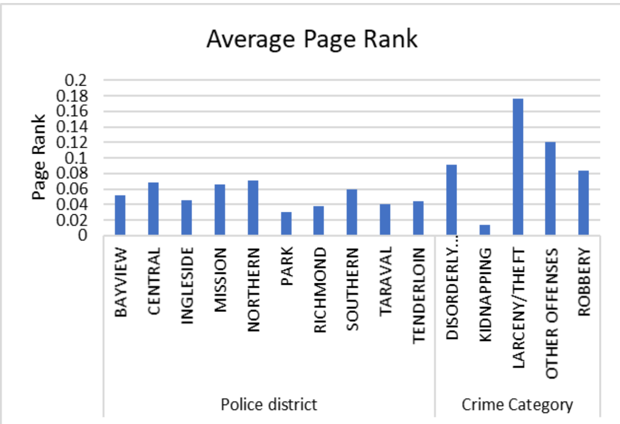

Fig. 21: Average Page rank chart of nodes for last five years.

5.4 Identification of areas where most of the crimes occur in the same zone

Identifying areas where most cases occur in the same neighbourhood is crucial information to retrieve and helps in future planning against crimes. Queries applied to strain areas based on the neighbourhood attribute of police districts that stores the average percentage of neighbourhood cases yearly. Query 1 identifies areas with the least number of the neighbourhood (<50%) cases in 2020, where query 2 filtrates the areas with maximum neighbourhood cases (>70%) of similar year.

a) Query 1

Dynamic Range

Parameters

- *range: 1.072859832021E12 - 1.6171488E12*
- *keepNull: true*

Range (neighborhood)

Parameters

- *column: neighborhood*
- *range: 21-50*

Figure 22 shows the result of the above query and filter out Richmond, Southern, and Tenderloin districts as the areas where neighbourhood cases are least or the crime occurrence are scattered evenly in all locations in 2020.

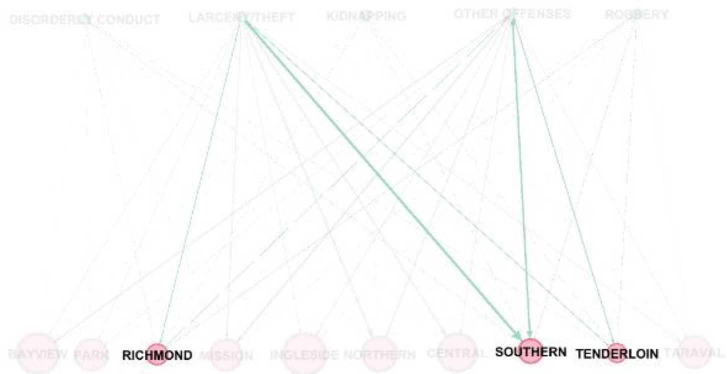

Fig. 22: Selection of nodes with average neighbourhood cases less than 50% in 2020.

b) Query 2

Dynamic Range

Parameters

- *range: 1.072859832021E12 - 1.6171488E12*
- *keepNull: true*

Range (neighborhood)

Parameters

- *column: neighborhood*
- *range: 71-81*

Figure 23 shows in Bayview and Ingleside, most of the crimes occurs in same area or neighbourhood. Fig 24 shows the average neighbourhood crime cases chart of all police districts for last 15 years. It shows that the Ingleside region of San Francisco city has maximum cases occurred in same neighbourhood as compared to other regions. Richmond has the least number of these cases.

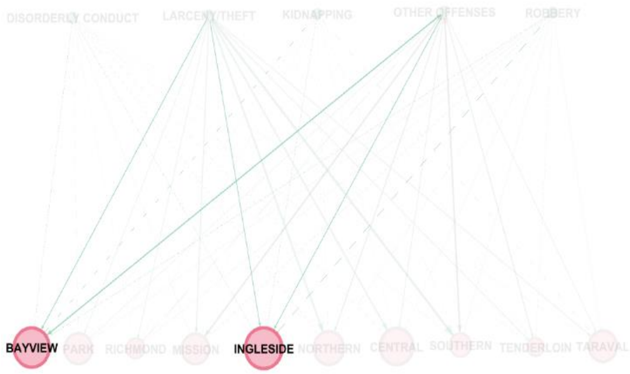

Fig. 23: Selection of nodes with average Neighbourhood cases more than 70% in 2020.

5.5 Identification of police districts where weekend crime cases are most

To identify districts where the weekend crime rate is more again vital information to retrieve. Queries applied to filtrate these areas based on the weekend attribute of police districts. Query 3 and 4 relates to strain areas for least and most weekend cases consequently in 2020.

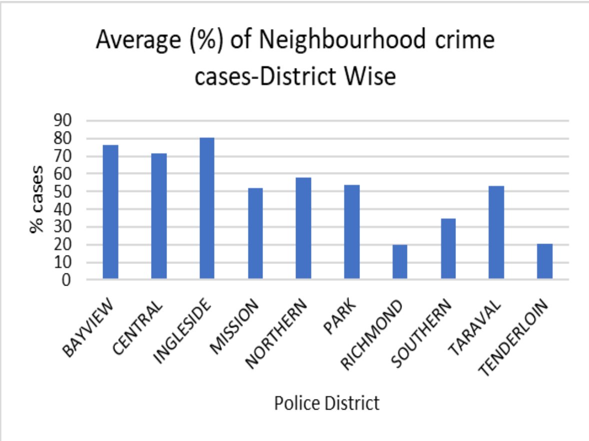

Fig. 24: Average (%) neighbourhood cases chart for police districts for last 15 years.

a) Queries 3: -

Dynamic Range

Parameters

- *range: 1.072859832021E12 - 1.6171488E12*
- *keepNull: true*

Range (weekend)

Parameters

- *column: weekend*
- range: 24.634-24.673

b) Queries 4:-

Dynamic Range

Parameters

- *range: 1.072859832021E12 - 1.6171488E12*
- *keepNull: true*

Range (weekend)

Parameter

- *column: weekend*
- range: 28-29.112007

Figure 25 shows the outcome of query 3. Ingleside and Taraval have the lowest weekend crime cases in 2020.

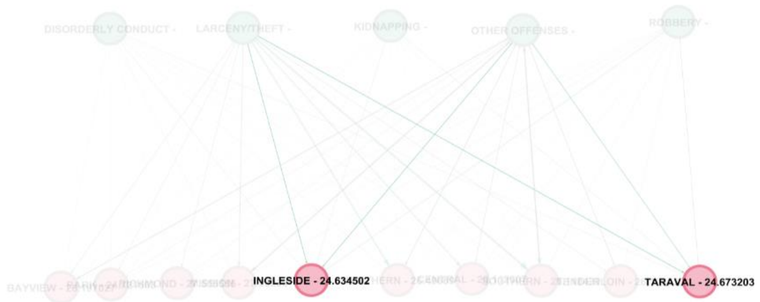

Fig. 25: Selection of nodes with least weekend crime cases in 2020.

The values associated with each police district is the percentage of crime cases occurred on weekends in 2020. Figure 26 shows the result of query 4 that filtrates the central district with the maximum number of weekend cases in 2020.

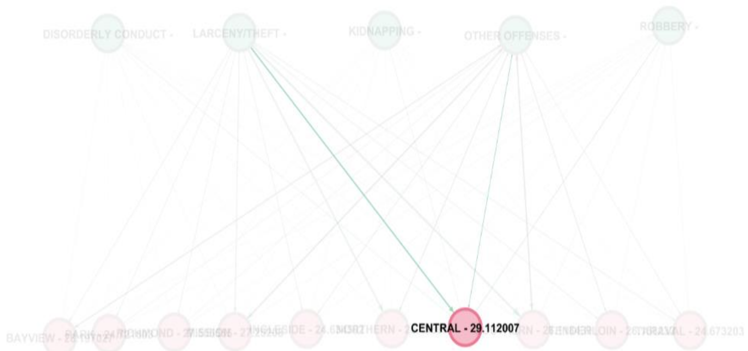

Fig. 26: Selection of nodes with most weekend crime cases in 2020.

Figure 27 shows the average weekend cases chart of all police districts for last 15 years. It represents that the central district has the maximum number of weekend cases as compared to other regions.

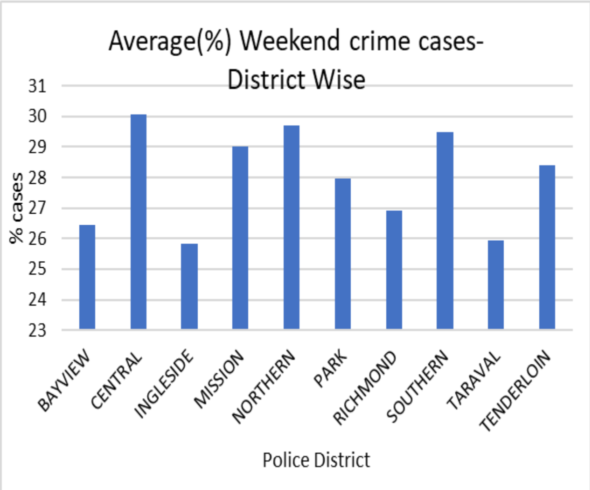

Fig. 27: Average weekend cases chart of police districts for last 15 years.

5.6 Identification of strongest/weakest category of crime for any police district

The spatialization algorithms work on the principle of attraction and repulsion. Based on the weights of the edges, nodes are placed nearer or farther to each other in the map. The node place in the center of the map is the most significant or connected to other nodes. The nodes with weak connections are placed farthest on the map. Distance between nodes also signifies the strength of the relationship between nodes. Force Atlas 2(Jacomy et al., 2014), and OpenOrd(Martin et al., 2011) are the two force-based algorithms that give approximate results when applied for spatial mapping of dynamic crime networks in 2020. Figure 28 and 29 shows the outcome of Force Atlas 2 and OpenOrd respectively with larceny/theft comes out as a most significant category of 2020 by both algorithms with most cases in central and northern districts followed by southern and mission.

Kidnapping and other offenses come out as the weakest crime category by Force atlas 2, whereas Robbery is weakest according to Open Ord.

6. Discussion on observed objectives

This section presents an explicit discussion of congregated results on all the observed objectives for the present research.

6.1 The crime graph in the southern district was higher than in other regions over time

Results of section 5.1 represented the variation in the total number of cases incidents and resolved by police districts over the years. The southern district had the highest number of cases incidents in most of the years. The average weighted degree chart of the last 15 years also supports this result and shows crime graph of the southern district is marginally higher than other districts.

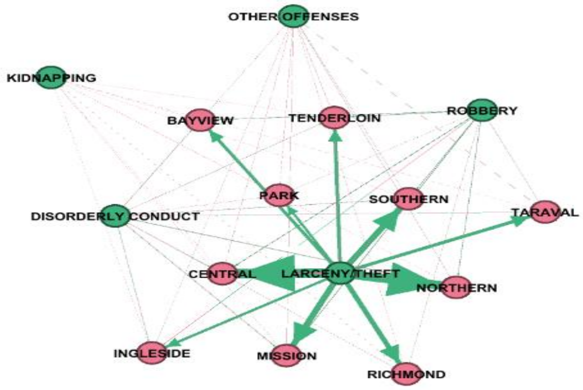

Fig. 28: Spatialization of nodes in 2020 based on Force Atlas 2.

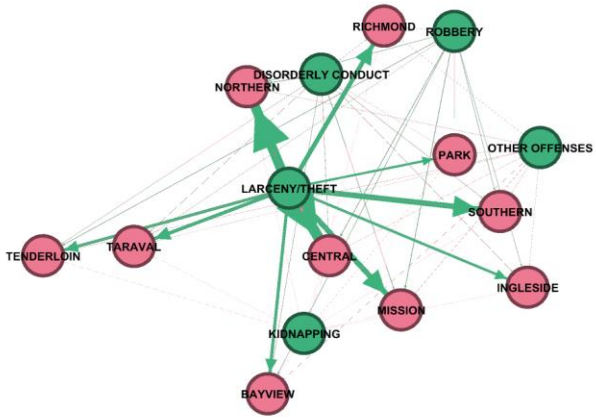

Fig. 29: Spatialization of nodes in 2020 based on OpenOrd.

6.2 The significance of a particular crime category in a region gets increased over time

Results induced in section 5.2 not only detect the significance of a particular crime category in a district at any given timestamp (year) but also show that the modularity of the dynamic crime network gets increased over time. These results conclude that the impact of a particular crime category in a district expands over the years, and areas are more prone to one or two crime categories. These results will help in making a prevention plan by considering significant crime categories that dominate that district.

6.3 Northern district remained highest in resolving cases over the years, and larceny/theft was a majorly resolved crime category

Results induced in section 5.3 signifies that the northern police district had resolved a maximum number of crime cases most of the time. Larceny/theft was a majorly work-out crime category in most years. The average page rank chart of police districts and crime categories for the last five years also supports this inference.

6.4 Neighbourhood crime cases were highest in Ingleside over the years

The results of section 5.4 revealed an intriguing fact: Ingleside district had the highest number of criminal cases in the neighbourhood or neighbouring locations maximum times. The number of neighbourhood instances was the lowest in the Richmond district. This finding gets supported by the average neighbourhood cases chart of all police districts. It's critical information because putting in place rigorous crime prevention programs in high-crime areas can help keep crime rates low in neighbouring locations as well.

6.5 Crime cases on weekends occurred maximum in the Central district and least in Ingleside and Traval

The findings of section 5.5 deduce that the percentage of criminal cases that occur on weekends is higher in the central police district for most of the time compared to other regions. Ingleside and Traval remained at the bottom for the most part. This data is essential for a better understanding of crime patterns in different districts, as it aids in decision-making and strategic planning.

6.6 Manuscript requirements (Use “Header 2” style)

According to the findings of section 5.6, larceny/theft is the most common crime category in all regions of San Francisco, while kidnapping is the least prevalent mostly. In comparison to other crime categories, this information aids in developing a thorough preventive plan for theft.

7. Conclusion and Future scope

Earlier criminal analysts did not use data driven policing and attempt to locate the felonious happenings with the help of yellow notes extracted from stack of files. Fundamentally, the system was failing in terms of – ignorance of criminal justice system, meta data about the things that matter, no data sharing, analytics tools help to make better critical decision process. Usually in the entire crime scheme - in police department, in prosecution offices, in courts and in jails decisions were made based on the experience and instincts.

Introduction of data analysis and statistical analysis helps critical decision-making process to reduce crime and improve criminal justice system. Adjudicators were making decision subjectively because public safety is the most important function of a government. The dynamic visual approach followed in this research not only helps to visualise complex networks that are easier to comprehend. Moreover, the addition of analysis helps the search for additional knowledge contain by the patterns and improves our comprehension of the network. The present research focus on the city of San Francisco and draws very productive results in terms of identifying the district having highest number of cases, modularity of dynamic crime and time period when the maximum number of cases occurred etc. In the future, the work can be extended in many dimensions- to use dynamic visual data analytics approach or making on of the most critical decision in public safety and that is the determination of Low-risk offenders. Juries all over the world shares one common perspective specifically to put threatening offenders in prison and let non-dangerous, non-violent people out. This requires analysing the crime patterns for the severity of the crimes committed by the offender. Using visual analysis approach, risk imposed by an offender can be understand in a scientific and objective way. Although risk analysis tools are available in the market but are not widely used by the jurisdiction because they are expensive to administer, time consuming, limited to local jurisdictions for which they are originally build or created and are difficult to scale or transform for other places. Further the work can also be extended to predict the crime and criminal behaviour using various machine learning algorithms.

References

Aggrawal, N., & Arora, A. (2016, October). Visualization, analysis and structural pattern infusion of DBLP co-authorship network using Gephi. *In 2016 2nd International Conference on Next Generation Computing Technologies (NGCT)* (pp. 494-500).

Alzahrani, T., & Horadam, K. (2019). Finding maximal bicliques in bipartite networks using node similarity. *Applied Network Science*, 4(1), 1-25.

Ayyappan, G., Nalini, C., & Kumaravel, A. (2016). A study on SNA: measure average degree and average weighted degree of knowledge diffusion in Gephi. *Indian Journal of Computer Science and Engineering,* 7, 230-237.

Bailey, L., Harinam, V., & Ariel, B. (2020). Victims, offenders and victim-offender overlaps of knife crime: A social network analysis approach using police records. *PLoS one*, 15(12), e0242621.

Baranauskaitė, L. (2020). Import Risks for the Country: A Case Study in Lithuania. Journal of Service, *Innovation and Sustainable Development,* 1(1&2), 51-68.

Blondel, V. D., Guillaume, J. L., Lambiotte, R., & Lefebvre, E. (2008). Fast unfolding of communities in large networks. *Journal of statistical mechanics: theory and experiment*, 2008(10), P10008.

Bokhare, A., & Metkewar, P. S. (2021). Visualization and Interpretation of Gephi and Tableau: A Comparative Study. *In Advances in Electrical and Computer Technologies: Select Proceedings of ICAECT 2020* (pp. 11-23).

Borowik, G., Wawrzyniak, Z. M., & Cichosz, P. (2018, December). Time series analysis for crime forecasting. *In 2018 26th International Conference on Systems Engineering (ICSEng)* (pp. 1-10). IEEE.

Brin, S., & Page, L. (1998). The anatomy of a large-scale hypertextual web search engine. *Computer networks and ISDN systems*, 30(1-7), 107-117.

Charette, Y., & Papachristos, A. V. (2017). The network dynamics of co-offending careers. *Social Networks*, 51, 3-13.

Chen, P., & Lu, Y. (2018). Exploring co-offending networks by considering geographic background: an investigation of electric bicycle thefts in Beijing. *The Professional Geographer*, 70(1), 73-83.

Cheng, Z., Flouvat, F., & Selmaoui-Folcher, N. (2017, May). Mining recurrent patterns in a dynamic attributed graph. *In Pacific-Asia Conference on Knowledge Discovery and Data Mining* (pp. 631-643).

Cherven, K. (2015). Mastering Gephi network visualization. Packt Publishing Ltd.

Chessa, A., Crimaldi, I., Riccaboni, M., & Trapin, L. (2014). Cluster analysis of weighted bipartite networks: a new copula-based approach. *PloS one*, 9(10), e109507.

Dawande, M., Keskinocak, P., Swaminathan, J. M., & Tayur, S. (2001). On bipartite and multipartite clique problems. *Journal of Algorithms*, 41(2), 388-403.

De Moor, S., Vandeviver, C., & Vander Beken, T. (2018). Integrating police-recorded crime data and DNA data to study serial co-offending behaviour. *European Journal of Criminology*, 15(5), 632-651.

Duan, D., Li, Y., Jin, Y., & Lu, Z. (2009, November). Community mining on dynamic weighted directed graphs. *In Proceedings of the 1st ACM international workshop on Complex networks meet information & knowledge management* (pp. 11-18).

Elezaj, O., Yayilgan, S. Y., Kalemi, E., Wendelberg, L., Abomhara, M., & Ahmed, J. (2019, December). Towards designing a knowledge graph-based framework for investigating and preventing crime on online social networks. *In International Conference on e-Democrac*y (pp. 181-195). Springer, Cham.

Engström, C., & Silvestrov, S. (2018). PageRank for networks, graphs, and Markov chains. *Theory of Probability and Mathematical Statistics*, 96, 59-82.

Feng, M., Zheng, J., Han, Y., Ren, J., & Liu, Q. (2018, July). Big data analytics and mining for crime data analysis, visualization and prediction. *In International Conference on Brain Inspired Cognitive Systems* (pp. 605-614).

Fruchterman, T. M., & Reingold, E. M. (1991). Graph drawing by force ‐ directed placement. *Software: Practice and experience*, 21(11), 1129-1164.

Gupta, A., Thakur, H. K., Jain, B., & Garg, S. (2015). Framework for mining regular patterns in dynamic networks. *International Journal of Knowledge Engineering and Data Mining*, 3(3-4), 299-314.

Gunduz, M., & Khader, B. K. (2019). Management performance assessment: A method of analytic network process. *Journal of System and Management Sciences*, 9(4), 67-90.

Jacomy, M., Venturini, T., Heymann, S., & Bastian, M. (2014). ForceAtlas2, a continuous graph layout algorithm for handy network visualization designed for the Gephi software. *PloS one*, 9(6), e98679.

Kuchar, J. Multimode Networks plugin for Gephi. [s.l.]: Open Source Community, 2014. https://github.com/jaroslav-kuchar/Multimode-Networks

Kurshan, E., Shen, H., & Yu, H. (2020, September). Financial Crime & Fraud Detection Using Graph Computing: Application Considerations & Outlook. *In 2020 Second International Conference on Transdisciplinary AI (TransAI*) (pp. 125-130).

Lahiri, M., & Berger-Wolf, T. Y. (2008, December). Mining periodic behavior in dynamic social networks. *In 2008 Eighth IEEE International Conference on Data Mining* (pp. 373-382).

Lantz, B. (2020). Co-offending group composition and violence: The impact of sex, age, and group size on co-offending violence. *Crime & Delinquency*, 66(1), 93-122.

Li, Q., Han, J., Ko, Y. O., Li, Q., & Lo, C. M. (2020). A Content Mining and Network Analysis Method for the Thematic Evolution of Educational Leadership. *Journal of System and Management Sciences*, 10(3), 86-98.

Lu, Z., Wen, Y., & Cao, G. (2013, March). Community detection in weighted networks: Algorithms and applications. *In 2013 IEEE International Conference on Pervasive Computing and Communications* (PerCom) (pp. 179-184).

Martin, S., Brown, W. M., Klavans, R., & Boyack, K. W. (2011, January). OpenOrd: an open-source toolbox for large graph layout. *In Visualization and Data Analysis 2011* (Vol. 7868, p. 786806).

Park, A. J., & Stamato, S. Z. (2020, December). Social Network Analysis of Global Transshipment: A Framework for Discovering Illegal Fishing Networks. *In 2020*

IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM) (pp. 776-783).

Park, A. J., Tsang, H. H., & Brantingham, P. L. (2012, August). Dynalink: A framework for dynamic criminal network visualization. *In 2012 European intelligence and security informatics conference* (pp. 217-224).

Rim, H., Lee, Y., & Yoo, S. (2020). Polarized public opinion responding to corporate social advocacy: Social network analysis of boycotters and advocators. *Public Relations Review*, 46(2), 101869.

Sinclair, P. (2004). Betweenness centralization for bipartite graphs. *Journal of Mathematical Sociology*, 29(1), 25-31.

Thakur, H. K., Gupta, A., Jain, B., & Ambika. (2017). Mining Maximal Quasi Regular Patterns in Weighted Dynamic Networks. *I.J. Information Technology and Computer Science*, Vol.9, No.4, pp.48-62.

Wajahat, A., Nazir, A., Akhtar, F., Qureshi, S., Razaque, F., & Shakeel, A. (2020, January). Interactively visualize and analyze social network gephi. *In 2020 3rd International Conference on Computing, Mathematics and Engineering Technologies (iCoMET)* (pp. 1-9).

Wang, W. (2013). Research on Logistics Distribution Networks of Online Retailers. *Journal of System and Management Sciences*, Vol. 3 (2013) No. 2, pp. 51-59

Wang, J., Shao, F., Wu, S., Sun, R. C., & Li, R. (2016). Weighted bipartite network projection for personalized recommendations. *Journal of Advances in Computer Networks*, 4(1), 64-69.

Weerman, F. M., Wilcox, P., & Sullivan, C. J. (2018). The short-term dynamics of peers and delinquent behavior: An analysis of bi-weekly changes within a high school student network. *Journal of quantitative criminology*, 34(2), 431-463.

Wu, J., Zhong, S. H., & Liu, Y. (2020). Dynamic graph convolutional network for multi-video summarization. *Pattern Recognition*, 107, 107382.

Yamini, M. P. C. (2019). A violent crime analysis using fuzzy c-means clustering approach. *ICTACT Journal on Soft Computing*, 9(3), 1939-1944.

Yoon, J., & Joung, S. (2020). A big data based cosmetic recommendation algorithm. *Journal of System and Management Sciences*, 10(2), 40-52.

Zaidi, F., Muelder, C., & Sallaberry, A. (2014). Analysis and visualization of dynamic networks. arXiv preprint arXiv:1409.5034.

Zhang, P., Wang, J., Li, X., Li, M., Di, Z., & Fan, Y. (2008). Clustering coefficient and community structure of bipartite networks. *Physica A: Statistical Mechanics and its Applications*, 387(27), 6869-6875.

Zhou, C., Feng, L., & Zhao, Q. (2018). A novel community detection method in bipartite networks. *Physica A: Statistical Mechanics and its Applications*, 492(15), 1679-1693.