

The Case Study of an Australian Crime Dataset

Jessica Liebig and Asha Rao

School of Mathematical and Geospatial Sciences, RMIT University, Melbourne, VIC, Australia; Emails: jessica.liebig@rmit.edu.au (J.L.); asha@rmit.edu.au (A.R.)

SUMMARY

Analysis of crime data is crucial for prevention and assessment of illegal activity. This paper is one of the first case studies of a crime dataset collected in New South Wales, Australia. We apply methods from complex network analysis to identify key aspects of criminal activity in the state of New South Wales. We further detect groups of local government areas and examine their dynamics over time. We represent our results by using various different visualisation techniques.

Keywords: Crime data, Visualisation, Networks

INTRODUCTION

The analysis of crime datasets is necessary in order to prevent and assess criminal activity [17]. Information about different types of crimes can often be found in the form of annual reports published by government bodies, but rarely in the form of publicly available datasets that may be used for research. In contrast, the New South Wales Bureau of Crime Statistics and Research in Australia has published data on criminal activity in the state of New South Wales (NSW) [4]. It contains information collected between 1995 and 2012, recording several types of offences and the local government area where they occurred.

In this paper we use this NSW crime dataset to shed light on the dynamics of criminal activity. This is one of the first case studies of this dataset along with the one presented in [6]. We make use of tools from complex network analysis and apply various visualisation techniques to present our results. A network is a mathematical representation of a system, consisting of nodes that represent different entities of the system, and edges that connect them. In a social network, for instance, nodes may represent different people whereas an edge connecting two people could represent friendship between the individuals.

Much research has been conducted in order to identify influential people and other entities in networks [7, 9, 10]. For example, in the development of marketing strategies, knowledge of important people can help to spread information about new products quickly. Similarly, such knowledge has been shown to aid the termination of the spread of disease in human and animal contact networks [8]. In this paper we apply a variation of a method finding influential people that was introduced by us in a previous publication [11] and which we used therein to successfully find the most important people in a social and a terrorist network.

In addition to identifying important locations and offence categories within the crime network, we discover that local government areas form different groups. It is often the case that certain locations associate with different groupings over time and we present several visualisations to clarify the dynamics of the discovered groups.

The rest of the paper is organised as follows: First, we give a detailed description of the dataset and its network representation. We then outline the approach used to find important areas and offence categories and visualise the results. Next we describe the process of detecting groups and show the dynamics of their structure over time. Finally we give the conclusions.

THE DATA

The dataset analysed in this paper is publicly available¹ and contains information about the different types of crime that took place in New South Wales between 1995 and 2012. It records the local government area where the crime occurred along with its offence category and the month and year of the crime. The New South Wales Bureau of Crime Statistics also provides a helpful visualisation tool for the dataset on their website². It allows the user to research various basic statistics of the local government areas and offence categories.

As outlined in the introduction we use tools from complex network analysis as a means to analyse this data and represent the given information as a network. In the case of the NSW crime data, there are 155 local government areas and 49 offence categories, that can be represented by two different types of nodes. A government area can never be linked directly to another government area. Similarly, a connection cannot be established between two offence categories and hence, links are solely found between areas and offences. For example, the scenario of a person stealing from a retail store in Bourke and two people escaping custody in the local government areas, Wagga Wagga and Upper Hunter Shire, may be represented as the network depicted in Figure 1.

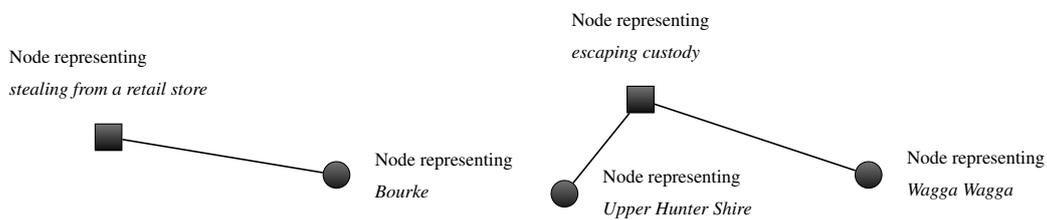


Figure 1. The network representation of the following scenario: A person steals from a retail store in Bourke and two people escape custody in the local government areas, Wagga Wagga and Upper Hunter Shire.

We are particularly interested in changes in the data between 2000 and 2012 and hence have divided the dataset into 156 networks, each covering a period of one month. Analysing each network separately and comparing the results gives valuable insights into the dynamics of criminal activity with respect to the local government areas.

IDENTIFYING CENTRAL ASPECTS TO CRIME ACTIVITY

The identification of central crime locations and offences is highly beneficial in preventing future criminal activity [17]. Knowledge of critical areas allows government agencies to target illegal activities more efficiently. By applying a combination of the two methods introduced in [11] and [12], we find the most important areas and offences in the NSW crime network.

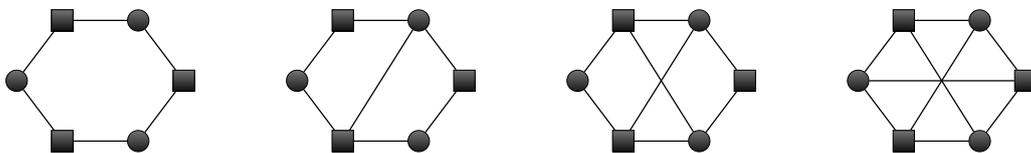


Figure 2. Finding the concentration of the four different patterns reveals (circles and squares represent locations and crimes respectively) the offences and areas that are most important. The four structures represent how well any three offence categories or locations are connected.

The process works as follows: We calculate the concentration of the four different patterns shown in Figure 2 with respect to every location and offence category (called the local clustering coefficient [11, 12]) in each network. We then compare the calculated concentrations of all nodes of the same type to the

¹<http://data.gov.au/dataset/nsw-crime-data/>

²<http://crimetool.bocsar.nsw.gov.au/bocsar/>

average concentration in the network. A location or offence type that shows a concentration that is very different to the average, plays an important role for the dynamics of the network. To be able to make the comparison to the average concentration, we calculate a score based on the mean and standard deviation of the various concentrations. For more detail on this method see [11] and [12]. The local government areas and offence categories can then be ranked accordingly. We have ranked the local government areas and offence categories for every month between January 2000 and December 2012. Note that the rankings of the local government areas are based on all 49 offence categories and vice versa.

Ranking of local government areas

A total of 155 local government areas form the Australian state of New South Wales. Their ranks range between 0 and 1 and are inversely proportional to the concentration. Thus, a rank close to zero shows that the concentration of a particular area was higher than the mean concentration. A rank close to 0.5 shows a similar concentration to the average, and a high rank (close to one) represents a concentration much lower than the average. Examination of our results shows that the rank of any individual area never fell below 0.3, meaning that the concentrations are skewed with many areas exhibiting concentrations below the average. We found that isolated and sparsely populated areas received extremely high ranks and did not show much variation over time. We have plotted the ranks of four government areas over time in Figure 3. Being able to make a clear connection between the rate of certain crimes in particular areas and their rank requires further work.

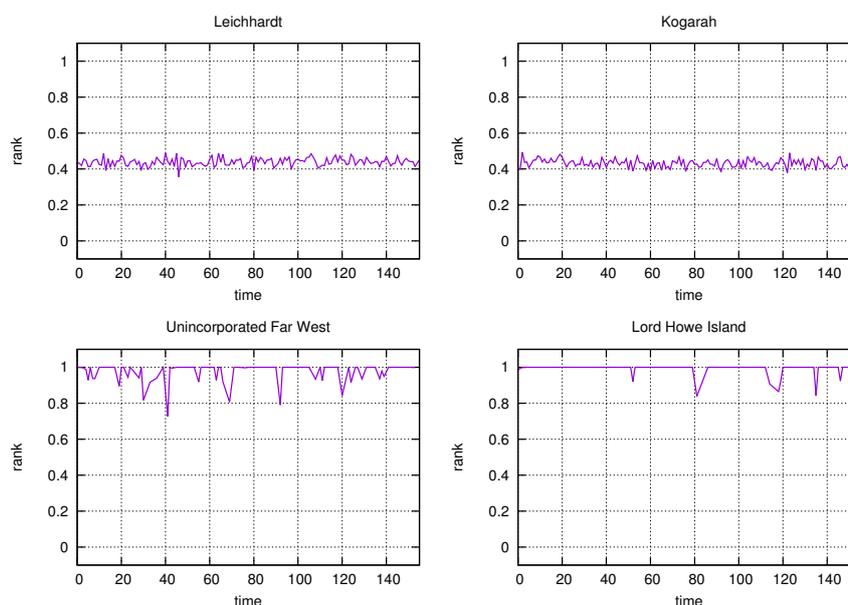


Figure 3. The graphs illustrate the rankings of four local government areas (Leichhardt, Kogarah, Unincorporated Far West, Lord Howe Island) over time.

Ranking of offence categories

Similar to the government areas, the ranks of the 49 offence categories range between 0 and 1. We observe that more common, often less serious, crimes are ranked low (close to 0) while, offences that are less common, but more serious, are given a high rank (close to 1).

The ranking of offence categories changes from month to month, however, the observed difference in the ranking of each category is generally small. Two of the lowest ranked categories in the NSW crime dataset in the years between 2000 and 2012 are *possession and use of cannabis* and *sexual offences*. Some offences that fall under *disorderly conduct* and certain *offences against justice procedures* were also ranked low throughout the 13 year period. Figure 4 shows the change in ranking of these offences together with other similar offences that fall within the same super-category.

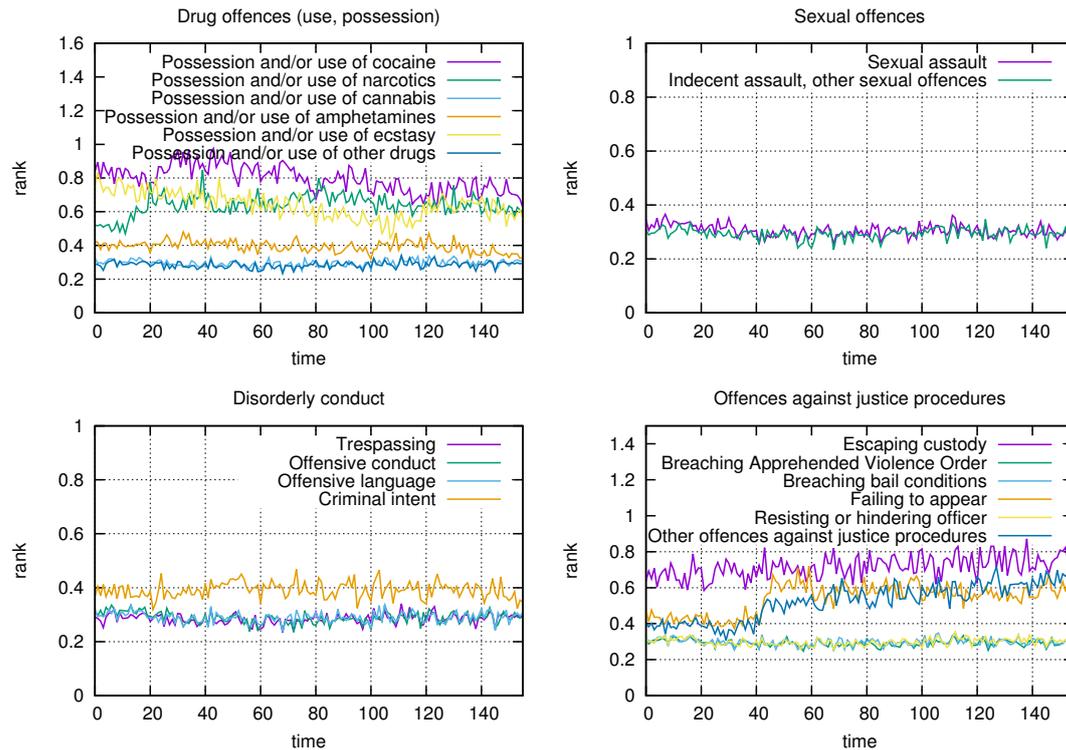


Figure 4. The rankings of offence categories that were particularly low over time together with similar categories that fall into the same super-category.

Looking at the first plot in Figure 4, we can see that the rank of the offence category *use or possession of cannabis* is much lower than that of other listed drugs. Although Australia has seen a significant decline in the use of drugs after the tightening of drug strategies in 1998, cannabis is still one of the most common and frequently used drugs [5].

Both sexual offence categories recorded in the dataset received very low rankings throughout the 13 year period. Sexual offences are a huge issue everywhere in Australia with New South Wales having the highest total number of sexual assaults reported to police [16]. According to the Australian Bureau of Statistics, 20% of women and 5% of men over the age of 15, experience sexual violence [1].

Disorderly conduct is another common offence in NSW, specifically on weekends and in connection with alcohol consumption [15]. Interestingly, the category *criminal intent*, is ranked higher than other acts of disorderly conduct. This is an indicator that in many cases the police do not pick up the planning of criminal activity.

On the other hand, *homicide* and the *dealing of cocaine* are two of the highest ranked categories (see Figure 5). According to the Australian Institute of Criminology [3], homicide incidents are currently one of the lowest crime rates in Australia and it is unlikely that a homicide remains unreported, as is often the case with domestic violence. With regards to cocaine dealing, between 2003 and 2012 cocaine arrests have accounted for less than 1.5% of national illicit drug arrests [2].

Clearly, the rank of offences reflects the severity of the crime and not the rate at which it occurs. All data indicates that more petty crimes such as *trespassing* occur more often than serious crimes such as murder.

DETECTION OF GROUPS

The detection of groups, of entities within a system, has been another field of great interest in the area of complex networks in recent years [14]. Being able to divide the local government areas of NSW into groups may further aid in the prevention of crime. Certain strategies of crime prevention that are already in place in some areas may be applied to other areas. However, a prevention scheme that works in one

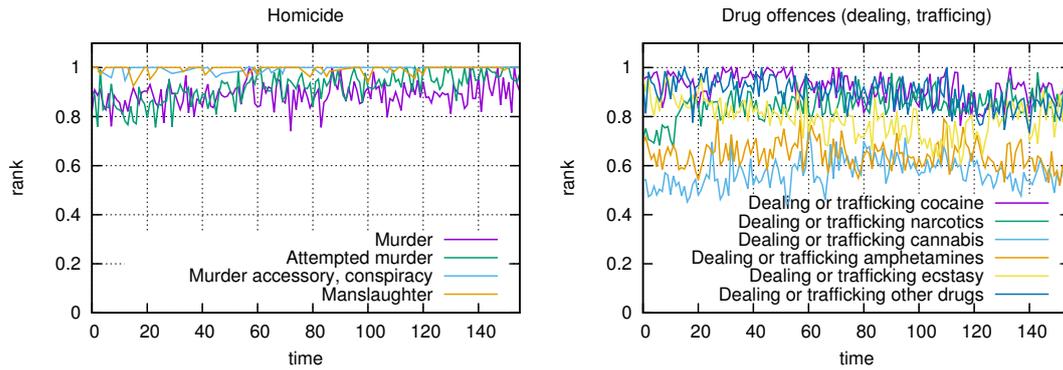


Figure 5. The highest ranked offence categories were homicide and drug dealing offences.

location is not guaranteed to be successful in another. If two local government areas are classified to be in the same group they certainly have many things in common. Therefore, a prevention strategy that works in one area is more likely to work in another area that is part of the same group.

Determining groups of local government areas and offence categories respectively, requires a simplification of the network. Without loss of generality we describe the process in terms of the local government areas. The network is simplified in the following manner: The nodes representing the 49 different offence categories are dropped and two areas are linked if one or more crimes of the same category occurred in both areas. For instance, if an attempted murder occurred in the two areas Bourke and Wagga Wagga, a connection is established between these regions. Connections are associated with an attribute that records the number of crimes in common. Once the network is simplified, it is possible to determine the most significant connections. Dropping all insignificant connections reveals the different groups as such connections often occur between groups, whereas significant links usually occur within groups. Details on how to determine the significance of a connection can be found in [13]. Note that the identified groups depend on significant connections to all 49 offence categories.

We are interested not solely in finding different groups but also in their development over time and hence have explored the dynamics in group structure of the local government areas.

We identified groups of local government areas in all 156 networks and always found two main groups as well as some smaller groups. The two largest groups usually contained government areas in the north east and south west respectively. We often found groups that only consisted of a single government area. Interestingly, such areas always received one of the highest ranks during the corresponding month.

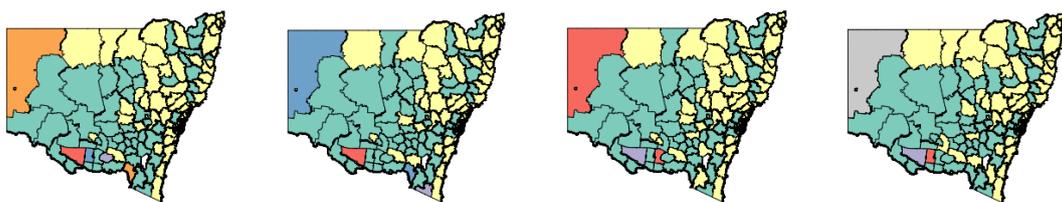


Figure 6. Maps of New South Wales and its government areas in the months August to November 2000. The different local government areas are coloured according to their group membership. The colour grey (see map on extreme right) indicates no data being available for the corresponding area in that month.

Figure 6 shows a map of NSW and its local government areas. Areas are coloured according to group membership. The largest group is coloured in blue, the second largest in green, with the size of a group determined by the number of its members and not the total area covered. The colour grey represents missing data for that month. Examination of data from October 2000 (third map in Figure 6) reveals that the areas in the largest group, coloured in green, experienced higher crime rates of trespassing than the average for NSW during that month. Trespassing happened to be the lowest ranked crime during that month. To answer the question whether this pattern continues throughout the dataset requires more

research and is left for future work.

CONCLUSION

This paper has shown how tools from complex network analysis can be applied to crime data in order to describe its dynamics. We have ranked the different offence categories and local government areas in the state of New South Wales in order to gain an understanding of the underlying mechanics of criminal activity. Different visualisation techniques were used to present the results. Being able to draw clear conclusions and find causations of the results presented in this paper requires further research.

REFERENCES

- [1] Australian Bureau of Statistics, viewed 10 December 2015, <<http://www.abs.gov.au/>>.
- [2] Australian Crime Commission 2014, *Illicit drug data report*, viewed 10 December 2015, <<https://www.crimecommission.gov.au/sites/default/files/290414-IDDR-2012-13.pdf>>.
- [3] Australian Institute of Criminology, viewed 10 December 2012, <<http://www.aic.gov.au/>>.
- [4] NSW Bureau of Crime Statistics and Research 2013, *NSW crime data*, viewed 6 December 2015, <<http://data.gov.au/dataset/nsw-crime-data>>.
- [5] United Nations Office on Drugs and Crime, viewed 10 December 2015, <www.unodc.org>.
- [6] ALZHRANI, T., AND HORADAM, K. J. Analysis of two crime-related networks derived from bipartite social networks. In *Advances in Social Networks Analysis and Mining (ASONAM), 2014 IEEE/ACM International Conference on* (2014), pp. 890–897.
- [7] ARAL, S., AND WALKER, D. Identifying influential and susceptible members of social networks. *Science* 337, 6092 (2012), 337–341.
- [8] CHEN, D., LÜ, L., SHANG, M.-S., ZHANG, Y.-C., AND ZHOU, T. Identifying influential nodes in complex networks. *Physica A* 391, 4 (2012), 1777–1787.
- [9] CHEN, D.-B., GAO, H., LÜ, L., AND ZHOU, T. Identifying influential nodes in large-scale directed networks: The role of clustering. *PLoS One* 8, 10 (2013), e77455.
- [10] KITSAK, M., GALLOS, L. K., HAVLIN, S., LILJEROS, F., MUCHNIK, L., STANLEY, H. E., AND MAKSE, H. A. Identifying influential spreaders in complex networks. *Nature Physics* 6, 11 (2010), 36.
- [11] LIEBIG, J., AND RAO, A. Identifying influential nodes in bipartite networks using the clustering coefficient. In *2014 Tenth International Conference on Signal-Image Technology and Internet-Based Systems* (2014), pp. 323–330.
- [12] LIEBIG, J., AND RAO, A. Predicting item popularity: Analysing local clustering behaviour of users. *Physica A* 442 (2016), 523–531.
- [13] LIEBIG, J., AND RAO, A. Fast extraction of the backbone of projected bipartite networks to aid community detection. *Europhysics Letters* (To appear, accepted: 25 January 2016).
- [14] NEWMAN, M. E. J. Finding community structure in networks using the eigenvectors of matrices. *Physical Review E* 74, 3 (2006), 036104.
- [15] SWEENEY, J., AND PAYNE, J. 2012, *Alcohol and disorderly conduct on Friday and Saturday nights*, viewed 10 December 2015, <<http://www.aic.gov.au/publications/current%20series/rip/1-10/15.html>>.
- [16] TARZON, C., AND QUADARA, A. 2012, *The nature and extent of sexual assault and abuse in Australia*, viewed 10 December 2015, <<http://www3.aifs.gov.au/acssa/pubs/sheets/rs5/>>.
- [17] WHITE, S., YEHLE, T., SERRANO, H., OLIVEIRA, M., AND MENEZES, R. The spatial structure of crime in urban environments. In *Social Informatics*. Springer, New York, 2014, pp. 102–111.