Bull. Aust. Math. Soc. **94** (2016), 350–352
doi:10.1017/S000497271600040X

COMPLEX NETWORK TOOLS TO ENABLE IDENTIFICATION OF A CRIMINAL COMMUNITY

PRITHEEGA MAGALINGAM

(Received 29 April 2016; first published online 16 August 2016)

2010 *Mathematics subject classification*: primary 94C15; secondary 05C82.

Keywords and phrases: criminal network, trust, shortest path, leave-one-out, suspect, investigation, betweenness centrality, intermediate node, reliance.

Retrieving criminal ties and mining evidence from an organised crime incident, for example money laundering, has been a difficult task for crime investigators due to the involvement of different groups of people and their complex relationships. Extracting the criminal associations from enormous amounts of raw data and representing them explicitly is tedious and time consuming [1, 6, 13]. A study of the complex network literature reveals that graph-based detection methods have not, as yet, been used for money laundering detection. In this research, I explore the use of complex network analysis to identify the communication associations of money laundering criminals, that is, the important people who communicate between known criminals and the reliance of the known criminals on the other individuals in a communication path.

For this purpose I use the publicly available Enron email database [5] that happens to contain the communications of 10 criminals who were convicted of money laundering crime [3]. I show that my new shortest paths network search algorithm (SPNSA) combining shortest paths and network centrality measures is better able to isolate and identify criminals' connections when compared with existing community detection algorithms and k-neighbourhood detection [9]. The SPNSA is validated using three different scenarios: (i) when the investigator knows all the criminals, (ii) when the investigator fails to detect one of the criminals and (iii) when the investigator is at the starting stage and does not have any information about the criminals, but suspects a crime is occurring. In each of these scenarios, the criminal network graphs formed using SPNSA are small and sparse and hence suitable for further investigation. The SPNSA algorithm manages to extract a criminal network with a minimum of four criminals when none of the criminals is known.

Thesis submitted to RMIT University in July 2015; degree approved on 25 September 2015; supervisors Asha Rao and Stephen Davis.

Different algorithm feeds were used in the different investigation scenarios discussed above. I show that SPNSA using a relevant algorithm feed that is related to a crime incident yields the best result. This is validated by applying a random feed to the SPNSA, and the result shows that the probability of retrieving criminals using a random feed is very low—only 1% in 1000 graphs. In another scenario, in five out of nine cases, SPNSA was successful in retrieving the left-out criminal in the resulting sub-network.

My research starts with isolating emails with ‘BCC’ recipients, with a minimum of two recipients BCC-ed, from the Enron dataset. ‘BCC’ recipients are inherently secretive and the email connections imply a trust relationship between sender and ‘BCC’ recipients. There are no studies on the usage of only those emails that have ‘BCC’ recipients to form a trust network, which led me to analyse the ‘BCC’ email group separately. SPNSA is able to identify the group of criminals and their active intermediaries in this ‘BCC’ trust network [10]. Corroborating this information with published information about the crimes that led to the collapse of Enron yields the discovery of persons of interest that were hidden between criminals and could have contributed to the money laundering activity. For validation, larger email datasets that comprise all ‘BCC’ and ‘TO/CC’ email transactions are used. On comparison with existing community detection algorithms, SPNSA is found to perform much better with regard to isolating the sub-networks that contain criminals.

Besides extracting possible people to investigate from a large network, I further identify nodes (people) important to a criminal or suspect in the shortest paths network formed. A common method of identifying important nodes is by ranking on the basis of node centrality measures [12]. Suspiciousness in a communication network can be highly dependent on the number of times a person is contacted, the number of meetings involving this person and also the location of this person. When communication links are seen as a network graph, these attributes of real life can be measured as the number of times a node occurs between criminals and other nodes or the distance of a node from these criminals [12]. In this research, I proposed a method of identifying the importance of a node, given a set of nodes of interest. In order to track important people in a criminal path that connects a criminal to their acquaintances, I explored the mathematical concept of pair dependency on the intermediate nodes, adapting the concept to criminal relationships and introducing a new source-intermediate reliance measure [11]. For the purpose of illustration, besides the Enron ‘BCC’ and ‘TO/CC’ email transactions, the Noordin Top Terrorist network was also used.

I compared the performance of the reliance measure with other importance ranking methods such as Google PageRank [4, 8] and Markov centrality [14] as well as betweenness centrality [2, 7]. The results show that the reliance measure led to a different prioritisation in terms of possible people to investigate. I also used this reliance measure to form a new network; a criminal reliance sub-network for further investigation. For this, the nodes with highest reliance that occur between the criminals or suspects and other nodes are gathered. These nodes become a set of new suspicious nodes and they are repeatedly used as algorithm feeds in the SPNSA to find new

https://doi.org/10.1017/S000497271600040X Published online by Cambridge University Press

communities. The network search will stop when there is no link found from the suspicious node to other nodes.

In conclusion, in this thesis, I demonstrated the performance of SPNSA in multiple scenarios leading to an investigator identifying criminals' connections within a large network. I showed that the criminals' importance ranking that was previously dependent on centrality measures can now be improved by using source-intermediate reliance ranking. This new shortest paths network search algorithm (SPNSA) in combination with the reliance measure can be applied to one-to-one or one-to-many relationships; for example, hyperlinks that connect different web pages and client–server links, and could be used as primary investigation tools to investigate connections between criminals in a complex network.

References

[1] A. Basu, 'Social network analysis: a methodology for studying terrorism', in: *Social Networking* (Springer, Switzerland, 2014), 215–242.

[2] U. Brandes, 'A faster algorithm for betweenness centrality', *J. Math. Soc.* **25**(2) (2001), 163–177.

[3] K. F. Brickey, 'From Enron to WorldCom and beyond: life and crime after Sarbanes–Oxley', *Wash. U. L. Q.* **81** (2003), 357–402.

[4] S. Brin and L. Page, 'The anatomy of a large-scale hypertextual web search engine', *Comput. Networks ISDN Syst.* **30**(1) (1998), 107–117.

[5] W. W. Cohen, Enron email dataset (2009) [online]. Available at: http://www.cs.cmu.edu/~enron/ (accessed 10 March 2012).

[6] W. Didimo, G. Liotta, F. Montecchiani and P. Palladino, 'An advanced network visualization system for financial crime detection', in: *Pacific Visualization Symp. (PacificVis)* (IEEE, New Jersey, 2011), 203–210.

[7] R. Geisberger, P. Sanders and D. Schultes, 'Better approximation of betweenness centrality', in: *ALENEX* (SIAM, Philadelphia, PA, 2008), 90–100.

[8] S. Kamvar, T. Haveliwala and G. Golub, 'Adaptive methods for the computation of PageRank', *Linear Algebra Appl.* **386** (2004), 51–65.

[9] P. Magalingam, S. Davis and A. Rao, 'Using shortest path to discover criminal community', *Digital Investigation* **15** (2015), 1–17.

[10] P. Magalingam, A. Rao and S. Davis, 'Identifying a criminal's network of trust', in: *2014 Tenth Int. Conf. Signal-Image Technology and Internet-Based Systems (SITIS)* (IEEE, Washington, DC, 2014), 309–316.

[11] P. Magalingam, A. Rao and S. Davis, 'Ranking the importance level of intermediaries to a criminal using a reliance measure'. arXiv:1506.06221v3.

[12] M. Newman, *Networks: An Introduction* (Oxford University Press, New York, 2009).

[13] G. Oatley and T. Crick, 'Measuring UK crime gangs', in: *2014 IEEE/ACM Int. Conf. Advances in Social Networks Analysis and Mining (ASONAM)* (IEEE, New Jersey, 2014), 253–256.

[14] S. White and P. Smyth, 'Algorithms for estimating relative importance in networks', in: *Proc. Ninth ACM SIGKDD Int. Conf. Knowledge Discovery and Data Mining* (ACM, New York, 2003), 266–275.

PRITHEEGA MAGALINGAM, Advanced Informatics School,
Level 5, Menara Razak, Universiti Teknologi Malaysia (UTM),
Jalan Sultan Yahya Petra, 54100 Kuala Lumpur, Malaysia
e-mail: mpritheega.kl@utm.my

https://doi.org/10.1017/S000497271600040X Published online by Cambridge University Press