2014 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM 2014)

Analysis of two crime-related networks derived from bipartite social networks

Taher Alzahrani and K. J. Horadam
School of Mathematical and Geospatial Sciences
RMIT University
Melbourne 3000, Australia
Email: {taher.alzahrani, kathy.horadam}@rmit.edu.au

***Abstract*—In this paper we investigate two real crime-related networks, which are both bipartite. The bipartite networks are: a spatial network where crimes of various types are committed in different local government areas; and a dark terrorist network where individuals attend events or have common affiliations. In each case we analyse the communities found by a random-walk based algorithm in the primary weighted projection network. We demonstrate that the identified communities represent meaningful information, and in particular, that the small communities found in the terrorist network represent meaningful cliques.**

***Keywords*—bipartite network, criminal (illegal) network, community detection, random walks**

I. INTRODUCTION

SOCIAL network analysis (SNA) has attracted much attention from academic researchers in the last decade, with the evolution of social media on the Internet. These social networks form complex networks whose structure exhibits substantial heterogeneities and no clear pattern [1], and include the increasingly important class of bipartite networks. A bipartite network, also called an affiliation network [2], is a network in which there are two different types of nodes, and the edges between nodes may occur only if the nodes belong to different types.

Many social networks model the interaction between two types of nodes where nodes of the same type do not interact with each other. For instance, the movie actors network, where actors and the movies they act in are the two node types [3], is bipartite. This form of complex network naturally arises in covert or illegal networks, where individuals represent one node set, while common actions, events or affiliations represent the other node set, through which links between individuals are inferred.

In recent years, there has been increasing motivation to analyse bipartite networks in general as a separate network category, and in particular to investigate their community structure. We are interested in community detection in covert or illegal networks that have such a bipartite representation.

The high performance random walk based algorithm Infomap [4] is the best algorithm for community detection in large one mode networks, as tested on LFR benchmark networks [5]. Unfortunately it is impossible to apply Infomap on a bipartite network, since it relies on the existence of a stationary distribution for the probability of the walker to be at a given node. However, projection for one node set of the bipartite network may be combined with Infomap to acquire a weighted one mode network that can be clustered by this random walk technique. This is a natural approach, as usually one node set in a bipartite network, denoted the *primary set* P, is of more interest for a particular purpose than the other node set, the *secondary set* S.

We apply these techniques to two bipartite networks as case studies. The first case study is a locality-crime network which we analyse to demonstrate very simply that the Infomap algorithm applied to the weighted projected locality network extracts more ground-truth information from the bipartite network than a competing modularity based clustering algorithm.

The second case study is the South East Asian terrorist network known as the *Noordin Top* network. The information from which this network was initially identified is in the 2006 report [6], and describes a military group with branches across South East Asia that has been involved in mounting terror attacks on the United State and its allies. This included bombing the Marriott Hotel in Jakarta in 2003; the Australian embassy in Jakarta in 2004; and restaurants in Bali in 2002 and 2005. Application of Infomap to the weighted projection network of individuals is shown to identify communities with a range of sizes, of which the smallest are cliques which do reflect ground truth in the organisation.

The rest of the paper is organized as follows: Section 2 gives some background on community detection algorithms for bipartite networks and known work on the Noordin Top network. Section 3 describes the projection method briefly, while Section 4 will present the datasets and empirical results from our analysis of the two real world bipartite networks. Finally, we conclude in Section 5 with an outlook for future work.

II. RELATED WORK

According to [7], the Infomap algorithm [4] is the best performing community detection algorithm for large one mode networks, as tested against benchmark networks generated at random with pre-specified degree distributions, number of communities and distribution of community sizes [5]. The Louvain algorithm [8] is the best-performing modularity-based algorithm against the same benchmarks. The Infomap algorithm is based on minimising the average length of a random walk on the network and cannot be directly applied to a bipartite network (see Section III). The Louvain algorithm is a two-stage algorithm based on maximising the modularity, a quality measure depending on the ratio of intra-class to inter-class edge numbers. Other modularity based algorithms have been devised to apply specifically to bipartite networks. However, modularity-based algorithms are known to suffer from a resolution limit which can prevent them from detecting smaller communities (those with internal edge numbers less than $O(\sqrt{m})$, where m is the number of edges in the network) [9]. This may mean they do not detect significant or influential small communities simply because they are small.

Unlike the situation for one mode communities, where randomised benchmark networks with pre-scribed community structure may be generated for performance testing of community detection algorithms, there are very few bipartite social networks with known community structure available for benchmarking. The most widely studied is the small *Southern women* network of 18 actors and 14 events [10]. A meta-analysis of 21 studies of this network up to 2003 appears in [11]. Most of these studies find two (sometimes overlapping) communities of women while 6 studies find a third set of unclassified women and one finds 4 groupings in total.

A statistical modelling approach to community detection in bipartite graphs has been proposed in [12]. The paper first surveys the statistical models used for modelling networks where actors attend events (some of these models are not intended for community detection), and of which only one (the exponential random graph (p^*) model) had previously been applied to the Southern women network. It discusses a latent class model, which is a "mixed Rasch model" where the number of communities of women, K, is an initial (unknown) variable, and particular choices of K are fitted by assigning different event attendance probabilities among groups, but identical attendance probabilities within groups. The choice of K is discussed at some length.

Using a Bayesian version of the latent class model, [12] tests the Southern women network for $K = 1, \ldots, 7$ latent classes, with the aim of clarifying inconsistencies in the 21 earlier studies. After considerable analysis, they decide that there are two overlapping communities, with three women in the overlap. These three women are all classified as peripheral when assigned to one or other of the two groups in the earlier studies. However, when modern (post-2003) community detection algorithms are applied to the Southern women bipartite network, the majority find 4 non-overlapping communities [13, Table 2]. These randomised algorithms search for communities which optimise a quality function and, in the process, automatically determine their optimal value of K.

In [13], both the Infomap algorithm and the Louvain algorithm are incorporated into an algorithm which projects a bipartite network onto a weighted one mode network. The paper demonstrates that the Infomap algorithm, applied to the weighted projected primary network of 18 women, identifies 4 communities, whereas the Louvain algorithm applied to the same network finds only 2 non-overlapping communities. Two small communities detected using Infomap, each containing two women, lie beneath the resolution limit mentioned above. Each consists of women identified as core members of each of the two larger communities found by the Louvain algorithm.

This work on the benchmark Southern women bipartite social network supports the ideas that peripheral actors in overlapping communities may be assigned to additional communities by modularity-based algorithms, while actors with strong ties (but beneath the resolution limit of modularity-based algorithms) may be assigned to additional small communities by the random-walk based Infomap algorithm.

Recent SNA research on the Noordin Top network [14] studies how its structure varies from 2001 to 2010 to test if social network change detection can identify significant change in a dark network's topology. Another study [15] utilizes two generic approaches, kinetic and non kinetic, to explore disruption strategies for dark networks in order to combat terrorism. In [16] the Bayesian latent class model of [12] is applied directly to the Noordin Top terrorist network for $K = 1, \ldots, 4$. The researchers find no appreciable difference in the deviance distributions for $K = 3$ and $K = 4$ and opt for the $K = 3$ model. With 3 groups, their Group 1 contains two important leaders and planners (Noordin Top and Azahari Husin), and they conclude that their Group 2 are the intermediaries who meet the planners and who train Group 3, the "footsoldiers".

We are not aware of any research that considers clustering algorithms in terror networks, or in the Noordin Top network in particular, from a bipartite network viewpoint.

III. Method

The high performance random walk based algorithm, Infomap, utilizes the information flow on a network in order to achieve its clustering. Unfortunately, it is impossible to formulate Infomap for a bipartite network, since it depends on calculation of the stationary distribution to which the random walk transition matrix converges. In a bipartite network, while a stationary distribution can be calculated for nodes in either set, the walk is periodic, and the distribution does not converge.

The stationary distribution for random walks on any network is given by the equation [17]:

$$\pi_v = \frac{d(v)}{2m} \tag{1}$$

where $d(v)$ is degree of vertex v for all $v \in V$ and m is total number of edges in the network. A random walk on a bipartite network will not converge to the stationary distribution in Equation 1 as time tends to ∞, independent of the start vertex, due to the periodicity. For instance, if you start in one node set of a bipartite network, then you will always be in that set after an even number of steps, so the probability of being at a particular node in that set after an odd number of steps is zero.

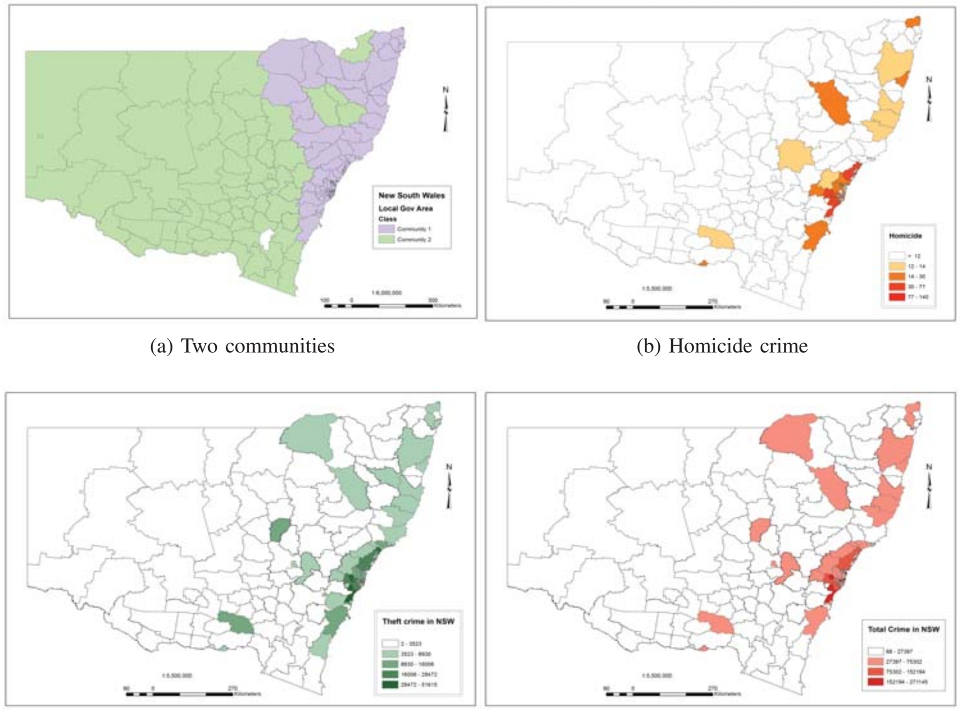
To overcome this, we use the weighted projection onto P for community detection in bipartite networks using Infomap. Either set P (usually actors) or S (usually events) can be used in the projection, but the actor set is commonly the main focus of interest.

This is motivated by the lack of existing benchmark bipartite networks. Moreover, we know of no community detection method in the literature that examines bipartite networks from a random walks perspective. We apply common neighbor similarity to the projection method for bipartite networks to acquire a weighted one mode network that can be clustered by the Infomap random walks technique. We use the common neighbor similarity scale mainly due to its simplicity and efficiency [18].

Projection is first done by finding the adjacency matrix A for P by checking for common neighbours between node i and j as follows:

$$A_{ij} = \begin{cases} 1, \text{if nodes } i \text{ and } j \text{ have a common} \\ \quad \text{neighbor} \\ 1, \text{if node } i \text{ has a neighbor which} \\ \quad \text{has no other neighbors in } P \\ \quad \text{(resulting in a self loop, } i = j) \\ 0, \ \text{otherwise} \end{cases}$$

This means that we do allow circuits, links from a node to itself, and multiple links encoded as link weighs between two given nodes. There are some measures based on the local structure of a bipartite network, which allow us to extract information that assists in finding relations between the primary set and itself in conjunction with secondary set. One of these measures is common neighbours similarity. For node i, let $\Gamma(i)$ represent its set of neighbours [18], the common neighbour similarity is defined as:



(a) Two communities

(b) Homicide crime

(c) Theft crime

(d) Total crimes

Fig. 1: Geographical map of NSW local goverment areas and related crimes. (a) The 2 communities of LGAs found by Infomap: Community 1 contains 82 LGAs and Community 2 contains 73 LGAs. The unclassified area is the Australian Capital Territory, which is not part of NSW. Underlying crime statistics are also mapped by LGA: (b) Homicide rate; (c) Theft rate and (d) Total crime rate. The correlation coefficient of the crime rates between the two communities is 0.992.

$$W_{ij} = |\Gamma(i) \cap \Gamma(j)| \quad (2)$$

so W_{ij} is the weight of the edge between node i and node j in the projected weighted network, and W_{ii} is 1 in case of a self loop.

IV. Datasets and Empirical Analysis

Two real world bipartite networks are analyzed from the community detection perspective. We show that our approach, using Infomap, uncovers more than ties between actors in the primary network P. The resulting community structures also reflect valuable information contained in S.

A. *The New South Wales crime network*

This publically available Australian crime data from the state of New South Wales (NSW) was published in 2012 [20]. It was collected by the NSW Bureau of Crime Statistics and Research from January 1995 to 2009, and it provides rich information about every crime that occurred in each month, categorised by offence type. There are 21 offence categories; some of these categories have subcategories that are related to the main category of the offences. For instance, "Homicide", as a category of offence, has four subcategories (Murder, Attempted murder, Accessory to Murder and Manslaughter) that all relate one way or another to the main category. The underlying social network of offenders is reflected in the reported crimes.

Moreover, the data reports the crime according to the local government area (LGA) it was committed in. There are 155 LGAs in NSW. The criminal bipartite network we extract has as node sets the offence categories and the LGAs that they were committed in. This bipartite network has 8761 edges, which reflects the frequency of commission of the crimes during the period 1995 – 2009. We are interested in identifying and illustrating where similar patterns of crime have occurred, and which are the more dangerous areas. As we are interested mostly in the areas where crimes have occurred, P is the geographic location and S is the main category of offence.

For comparison we applied both the Infomap and Louvain algorithms to the weighted projection on P, which has $m = 3,478,084$ edges. The Louvain algorithm did not determine any community structure at all. Consequently it is of no use for analytic purposes. However the Infomap algorithm found 2 communities of LGAs, one containing 82 LGAs and the other containing 73 LGAs. The modularity of this structure is higher than that for a single community, see Table 4.1, indicating it is a better structure, so the modularity-maximising Louvain algorithm should have found more than one community. This failure may be because an intra-community edge number is less than the modularity resolution limit of $\lfloor\sqrt{3478084}\rfloor = 1,864$. This is still to be tested.

We expect there is more frequent connection between some subset of crimes for Community 1 of LGAs versus the more frequent connection between some other subset of crimes for Community 2.

Most dramatically, when the LGAs in NSW are mapped and coloured according to community, a very strong geographical divide is visible. The 38 LGAs in the main metropolitan area, Sydney, are all in Community 1. Generally speaking, Community 1 includes the more populated LGAs and Community 2 includes the majority of rural and "Outback" LGAs. Analysis of the underlying crime statistics by LGA shows that for homicide (Fig1(b)), 90% of the shaded LGAs occur in Community 1; for theft (Fig1(c)) 85% of the shaded LGAs occur in Community 1 and for total crime rate (Fig1(d)), 86% of the shaded LGAs occur in Community1. The correlation coefficient of the crime rates between the two communities is 0.992. Deeper analysis of this network will be undertaken elsewhere.

TABLE 4.1: Comparison of algorithm performance on NSW crime network.

Algorithm	Communities	Sizes	Modularity
Louvain	1	155	0
Infomap	2	82, 73	0.026

B. *The Noordin Top terrorist network*

The Noordin Top terrorist group data linking individuals with relationships or affiliations first appeared in [6]. The ties or links between actors represent one or more common affiliations or relationships. Common attendance of actors at events was inferred from their mention together in public reports in newspapers and elsewhere. The data were coded as network data by Naval postgraduate students and the information was published in 2012 in [19]. We work with the affiliation subnetwork. It forms a bipartite network with 79 actors and 45 events (affiliations), classified into six categories (Operations, Logistics, Organizations, Training, Finance, Meeting). We excluded the actors who did not present at any of the 45 events. Weighted projection of the Noordin Top bipartite network onto the actor set P determines a network with $m = 759$ edges in total.

Our investigation using the Infomap algorithm found that there are 5 communities. For comparison we also applied the Louvain algorithm, which found 4 communities, see Table 4.2. The modularity resolution limit [9] for this network is $\lfloor\sqrt{759}\rfloor = 27$. Therefore, a community with strong ties and < 27 edges may not be detected by modularity based methods.

In fact, the largest Louvain community (of 29 actors) contains 23 of the 25 actors belonging to the first community found by the Infomap algorithm. It also contains the smallest Infomap community (a clique of 3 actors with weighted edge sum 6). The second small Infomap community (a clique of 4 actors with weighted edge sum 6) has three actors in the largest Louvain community and one in the second largest Louvain community of 16 actors. *Essentially, Infomap detects three communities inside the largest Louvain community*. The two small clique communities are approximately an order of magnitude smaller than the largest Louvain community. The smallest Louvain community (14 actors) wholly contains the third Infomap community (12 actors), and we regard them as essentially equivalent. Consequently, and guided by the results of the previous subsection,

TABLE 4.2: Communities in the Noordin Top terrorist network.

Algorithm	Communities	Sizes
Louvain	4	29, 15, 16, 14
Infomap	5	25, 30, 12, 4, 3

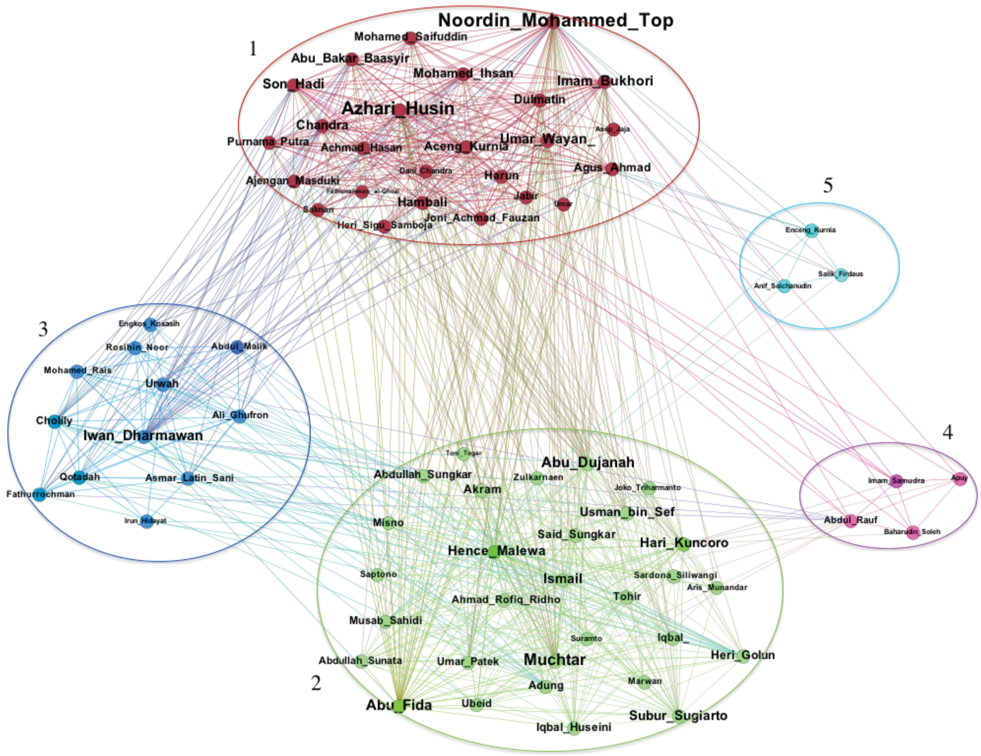


Fig. 2: Noordin Top terrorist network actor communities found using Infomap: Community 1 (red, top, 25 actors), Community 2 (green, bottom, 30 actors), Community 3 (purple, left, 12 actors), Community 4 (pink, below right, 4 actors), Community 5 (blue, above right, 3 actors). Edges are unweighted for clarity of representation.

we concentrate on the structure found by the Infomap algorithm.

The 5 communities found by the Infomap algorithm are displayed in detail in Fig 2. Community 4 contains actors Abdul Rauf, Imam Samudra, Apuy and Baharudin Soleh. Community 5 contains actors Enceng Kurnia, Anif Solchanudin and Salik Fridaus. These two small cliques have no recorded direct links between them, nor does Community 5 have any recorded direct links with Community 3. Identifying these small clique communities in the original bipartite network described in [15] recovers very sensible information. For instance, Anif Solchanudin and Salik Fridaus were trained together to be suicide bombers for Bali Bomb II in 2005. Community 4 also reflects useful information. Abdul Rauf, Imam Samudra and Apuy came from the same organization, known as Ring Baten, while Apuy and Baharudin Soleh were involved directly in the Australian Embassy bombing in 2004. The two smallest communities are new structure, not found by the Louvain algorithm or in [16].

We relate Fig 2 back to the 6 categories of events, as was done in [16] for only 3 communities. Community 1 contains the two principal leaders and planners (Noordin Top and Azhari Husin) and actors involved heavily in Operations and Training with 2-23 events. The most significant common property of this group is that 17 out of 25 of its actors were affiliated to the same Organization (Jemaah Islamiyah, a transnational Southeast Asian militant Islamist terrorist organisation linked to Al-Qaeda). Community 2s actors are involved more than other communities' actors in Finance and participated in 3-8 events. The remaining three Communities are never involved in Operations or Finance. Community 3 is characterised by having actors involved in Training with 3 or fewer events. The actors in Community 4 attended 2-6 events. Lastly, all actors in Community 5 are linked by Training events. We hope that appli-

TABLE 4.3: Top 10 actors in Noordin Top projected network by different centrality measures

Degree	Weight	Eigenvector Centrality	Betweenness
Noordin Top Mohamed (53)	Azhari Husin (115)	Noordin Top Mohamed (1)	Azhari Husin (303.968)
Azhari Husin (51)	Noordin Top Mohamed (111)	Azhari Husin (0.967)	Noordin Top Mohamed (303.27)
Muchtar (43)	Hari Kuncoro (56)	Muchtar (0.935)	Muchtar (99.034)
Abu Fida (37)	Son Hadi (53)	Abu Fida (0.861)	Son Hadi (87.691)
Abu Dujanah (36)	Iwan Dharmawan (53)	Abu Dujanah (0.855)	Iwan Dharmawan (82.837)
Iwan Dharmawan (35)	Abu Fida (52)	Imam Bukhori (0.765)	Ismail (76.103)
Imam Bukhori (33)	Muchtar (51)	Umar Wayan (0.756)	Hari kuncoro (72.404)
Umar Wayan (33)	Abu Dujanah (49)	Iwan Dharmawan (0.717)	Umar Wayan (56.725)
Son Hadi (32)	Imam Bukhori (45)	Hambali (0.698)	Imam Bukhori (54.236)
Ismail (32)	Aceng Kurnia (45)	Son Hadi (0.677)	Chandra (52.642)

cation of community merging algorithms may bring these relationships out automatically, see Section V.

Another interesting question about the communities which have been detected is, which of the most important individual actors belong to which community? A list of the Top 10 actors in the Operational subnetwork of the Noordin Top network, according to various centrality measures, appears in Table 3 of [15]. For the whole projected actor network not just the Operational network, we present similar information in Table 4.3. As with the network in [15], there is remarkable agreement amongst these measures: there are 14 actors in total in the four Top 10 lists of Table 4.3. Six of these 14 actors appear in both Table 4.3 and Table 3 in [15] and 4 of these six actors are located in our Community 1. For the 14 actors appearing in Table 4.3, 8 are in Community 1, 5 in Community 2 and 1 in Community 3. Taken together with the common membership of Jemaah Islamiyah of so many of the actors in Community 1, it is reasonable to conclude that Community 1 is the most significant.

V. Conclusions and Future Work

One of the primary contributions of this work is to apply the SNA tool of community detection to two criminal bipartite networks, through the approach of weighted projection into a random-walk based algorithm. We found that communities extracted by Infomap on the weighted primary network projected from the bipartite network reflect valuable and meaningful information in the bipartite network framework. For the NSW crime network, the LGAs in the main metropolitan areas are all classified into one community and the majority of rural and Outback LGAs are classified into the second community. In our second case study, the projected network of actors in the Noordin Top terrorist contains 5 communities, two of which are newly found small cliques, meaningful in terms of the ground truth, whose sizes lie beneath the modularity resolution limit.

The Infomap algorithm appears more sensitive than the Louvain algorithm and is able to find small communities, which is a valuable feature in criminal network analysis, as such small communities can show important dependencies.

The two networks presented in this paper are small, so we aim to test and analyze our approach in large scale networks, for example recommendation systems, specifically a users and books rating system.

One important observation made during the detailed study of these two cases is that each actor might in fact belong to more than one community when the information from the secondary network is taken into account. Neither the Infomap nor the Louvain algorithm can find overlapping communities, so investigation of overlapping communities is possible future work. Such a perspective may unfold new ways to measure the quality of the communities found. We also intend to project the two sets P and S of the bipartite network in parallel, cluster them separately using the random-walks based algorithm and recover their results within the bipartite network. This may naturally result in overlapping communities in the bipartite network. Finally we plan to compare these communities with those clusters found by modularity-based bipartite clustering and those using multi assignment clustering.

ACKNOWLEDGMENTS

We thank the reviewers for their helpful comments which have encouraged us to add more analysis of the NSW crimes network and have improved the presentation. We are very grateful to Assoc. Prof. Murray Aitken for supplying us with the cleaned affiliation network data for the Noordin Top terrorist network andl to Assoc. Prof. Chris Bellman and Sarah Taylor for running mapping software on our clustered NSW crime network. This work forms part of the PhD thesis of the first author, taken under the supervision of the second author. The first author would like to thank the Ministry of Finance of Saudi Arabia for supporting his research. The work of the second author was partly supported by Department of Defence of Australia Agreement 4500743680.

REFERENCES

[1] A. Galeotti and F. Vega-Redondo, "Complex networks and local externalities: A strategic approach," International Journal of Economic Theory, vol. 7, pp. 77-92, 2011.

[2] M. Latapy, C. Magnien, and N. D. Vecchio, "Basic notions for the analysis of large two-mode networks," Social Networks, vol. 30, pp. 31-48, 2008.

[3] L. A. N. Amaral, A. Scala, M. Barthelemy, and H. E. Stanley, "Classes of small-world networks," Proceedings of the National Academy of Sciences, vol. 97, pp. 11149-11152, 2000.

[4] M. Rosvall and C.T. Bergstrom, " Maps of random walks on complex networks reveal community structure," Proceedings of the National Academy of Sciences, 105, 1118–1123 (2008).

[5] A. Lancichinetti, S. Fortunato, and F. Radicchi, "Benchmark graphs for testing community detection algorithms," Physical Review E, vol. 78, p. 046110, 2008.

[6] International Crisis Group, "Terrorism in Indonesia: Noordin's Networks," Asia Report no. 114, Brussels, Belgium, 2006.

[7] Lancichinetti, A., Fortunato, S.: Community detection algorithms: A comparative analysis. Physical Review E 80, 056117 (2009)

[8] V. D. Blondel, J.-L. Guillaume, R. Lambiotte, and E. Lefebvre, "Fast unfolding of communities in large networks," Journal of Statistical Mechanics: Theory and Experiment, vol. 2008, p. P10008, 2008.

[9] S. Fortunato and M. Barthelemy, "Resolution limit in community detection," Proceedings of the National Academy of Sciences, vol. 104, pp. 36-41, 2007.

[10] Davis, A., Gardner, B.B., Gardner, M.R., Deep south: A Social Anthropological Study of Caste and Class University of Chicago Press Chicago, 1941.

[11] Freeman, L.C.: Finding social groups: A meta-analysis of the southern women data. Dynamic social network modeling and analysis, National Academies Press 39–97, 2003.

[12] Aitkin, M., Vu, D. and Francis, B., "Statistical modelling of the group structure of social networks", Soc. Net., vol. 38, pp.74-87, 2014.

[13] T. Alzahrani, K. J. Horadam and S. Boztas, "Community detection in bipartite networks using random walks," Proc. CompleNet 2014, Springer Studies in Computational Intelligence vol. 549, pp.157–165, 2014.

[14] S. F. Everton and D. Cunningham, "Detecting significant changes in dark networks," Behavioral Sciences of Terrorism and Political Aggression, vol. 5, pp. 94-114, 2013.

[15] N. Roberts and S. F. Everton, "Strategies for combating dark networks," Journal of Social Structure, vol. 12, 2011.

[16] Aitkin, M., Vu, D. and Francis, B., "Statistical modelling of a terrorist network ",[Online], Available: http://mathsofplanetearth.org.au/wp-content/uploads/2013/07/Aitkin-Murray.pdf, 2013.

[17] D. A. Levin, Y. Peres, and E. L. Wilmer, Markov chains and mixing times: American Mathematical Soc., 2009.

[18] T. Zhou, L. Lu, and Y. C. Zhang, "Predicting missing links via local information," The European Physical Journal B, vol. 71, pp. 623-630, 2009.

[19] S. F. Everton, Disrupting dark networks: Cambridge University Press, 2012.

[20] NSW Bureau of Crime Statistics and Research, Dataset [Online], NSW Crime data, Available: http://data.gov.au/dataset/nsw-crime-data/, 2012.