The (not so) Critical Nodes of Criminal Networks

Donatella Firmani[1](✉), Giuseppe F. Italiano[1], and Luigi Laura[2]

[1] University of Rome "Tor Vergata", Via del Politecnico 1, 00133 Rome, Italy
firmani@ing.uniroma2.it, giuseppe.italiano@uniroma2.it
[2] University of Rome "Sapienza", Via Ariosto 25, 00185 Rome, Italy
laura@dis.uniroma1.it

Abstract. One of the most basic question in the analysis of social networks is to find nodes that are of particular relevance in the network. The answer that emerged in the recent literature is that the *importance*, or *centrality*, of a node x is proportional to the number of nodes that get disconnected from the network when node x is removed. We show that while in social networks such important nodes lie in their *cores* (i.e., maximal subgraphs in which all nodes have degree higher than a certain value), this is not necessarily the case in criminal networks. This shows that nodes whose removal affects large portions of the criminal network prefer to operate from network peripheries, thus confirming the intuition of Baker and Faulkner [4]. Our results also highlight structural differences between criminal networks and other social networks, suggesting that classical definitions of importance (or centrality) in a network fail to capture the concept of key players in criminal networks.

Keywords: Articulation points · Social networks · Criminal networks

1 Introduction

In recent years several tools from Social Network Analysis (SNA) have been applied to the study of criminal networks; however, citing the words of Morselli [18]: "*Criminal networks are not simply social networks operating in criminal contexts. The covert settings that surround them call for specific interactions and relational features within and beyond the network*". Indeed, it is known that criminal networks differ substantially from social networks (in short, SNs); this is mainly due to the trade-off between security and efficiency which directly affects their underlying network structure [10]. Sparrow, in a seminal paper [23], listed four peculiar features of criminal networks: i) limited size, ii) information incompleteness (i.e., criminal network data is inevitably incomplete) iii) undefined borders, i.e. it is not easy to discover all the connections of a node;

This paper has been partially supported by MIUR, the Italian Ministry of Education, University and Research, under Project AMANDA (Algorithmics for MAssive and Networked DAta).

L.M. Aiello and D. McFarland (Eds.): SocInfo 2014 Workshops, LNCS 8852, pp. 87–96, 2015.
DOI: 10.1007/978-3-319-15168-7_12

and, iv) dynamics, that is, many of the useful networks questions depends heavily on the temporal dimension. In this scenario, SNA tools are limited, due to these intrinsic differences.

One of the most basic questions associated to the analysis of a network is *who are the key players*, or *who are the central nodes in the network*? Node-centrality measures introduced in the recent literature, such as degree, betweenness and closeness centrality, just to name a few (see [8] for a good survey) can successfully identify most key players in social networks. Quite on the contrary, in criminal networks, the most important actors do not necessarily display high centrality scores [4]. More recently, node-centrality measures that look at how some graph invariant changes when some nodes or edges are deleted have been studied for example in [9], but no results are known for criminal networks.

In this paper we show that criminal networks can suffer high disconnection when few nodes are deleted. However, while in social networks nodes whose removal disconnects most nodes from the network belong to their cores, in criminal networks this is not the case. This suggests that, on the one hand, and differently from most social networks, nodes that pulls together the network deliberately operate from network peripheries, thus being protected from detection; on the other hand, the key players in the network may not be the ones whose removal affects large portions of the network.

Related Work. The seminal work of Sparrow [23] is considered as the starting point of the academic research on the use of SNA tools to the study of criminal networks. The interested reader is referred to the books of Morselli [18] and to the very recent one of Masys [16] for a broad coverage of research on criminal networks. There are nowadays several tools devoted to the analysis of criminal networks using tools from SNA; we refer the interested reader to the survey of Xu and Chen [25] and we also cite the recent *LogAnalysis* of Ferrara et al. [11,12].

Schwartz and Rouselle [20] addressed the problem of identifying central actors in the network, building on the previous work of Borgatti [7,9] for the *key players* problem. Duijn et al. [10] focused on the dynamics of the interaction between disruption and resilience within criminal networks, concluding that the disruption of the criminal cannabis network they studied is relatively ineffective. Mainas [15] presented an exhaustive analysis of two criminal networks, a drug-trafficking and a terrorists group, which are the ones studied in this paper.

The concept of *critical nodes* has been introduced in [2], building upon the concepts of *articulation points* and *core* of a connected network, that are respectively the nodes whose removal disconnects the network, and the subset of the nodes obtained by repeatedly pruning the nodes of low (fixed) degree. Precisely, the critical nodes are defined as the articulation points belonging to the network core. Using different samples of the Autonomous Systems network[1], Ausiello

[1] The network of routers comprising the Internet can be organized into sub-graphs called Autonomous Systems (AS) and we can construct the ASes communication network from the BGP (Border Gateway Protocol) logs. More information about this dataset can be found at http://snap.stanford.edu/data/as.html.

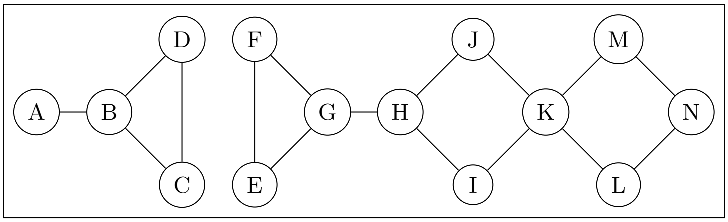

Fig. 1. An example graph with two connected components (nodes A, B, C, D and E, F, G, H, I, L, M, N) and four articulation points (B, G, H, K)

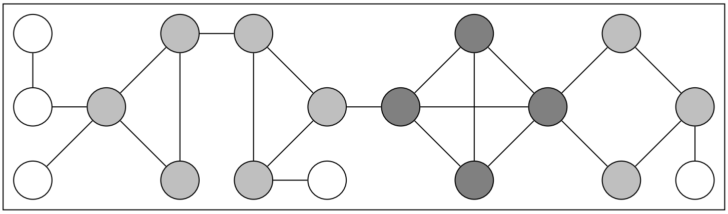

Fig. 2. An example graph, its 2-core (light and dark grey nodes), and its 3-core (dark grey nodes). Nodes in the 3-core are a subset of the nodes in the 2-core of the graph.

et al. show that: (1) the removal of few critical nodes can affect large portions of the network, thus they are central in a very strong sense [2], and (2) the critical nodes have orders of magnitude higher centrality scores than other nodes [3].

Structure of the Paper. In the next section we recall few preliminary notions. In Section 3 we discuss some properties of the main notions used in this papers: articulation points, critical nodes and network cores. Our experimental results are discussed in Section 4, whilst Section 5 addresses concluding remarks.

2 Preliminaries

Given an unirected graph $G = (V, E)$, a **connected component** is defined as a maximal set of nodes $V' \subseteq V$ such that, given $u, v \in V'$, there is at least one path between u and v in G. Furthermore an **articulation point** is defined as a node $v \in V$ such that its removal from the graph G increases the number of connected components in G (see Figure 1). A connected graph G is *biconnected* if the removal of any of its nodes leaves the graph connected. Therefore, a biconnected graph contains no articulation point.

Coreness. The concept of coreness was introduced in [21] and [6]. The k-*core* of a graph is defined as the unique subgraph obtained by recursively removing all nodes of a degree less than k. A node has coreness value l, if it belongs to the l-core but not to the $(l+1)$-core. We denote with G^k the k-core of graph G. Figure 2 shows an example graph G, its 2-core G^2, and its 3-core G^3.

Critical Nodes. As we mentioned in the previous section, we follow the definition of Ausiello et al. [2], that defines the critical nodes as articulation points of the network cores. In the following sections, unless specified otherwise, we refer as "critical nodes" to the articulation points of the 2-core.

3 Articulation Points, Critical Nodes and Cores

Before discussing the experimental results in the next section, we discuss briefly few properties of the main concepts used in of our analysis: articulation points, critical nodes and network cores. We begin by proving the following lemma, which shows the connection between articulation points and critical nodes.

Lemma 1. *Let G be an undirected connected graph. The critical nodes of G (i.e. the articulation points of G^2) are articulation points of G.*

Proof. Assume by contradiction that node x is a critical node of G, but not an articulation point of G. If x is a critical node of G this means, by definition, that the removal of x from G^2 disconnects at least one node from G^2. Consider a generic node y that gets disconnected from G^2 when node x is removed. Since the removal of x disconnects y in G^2, this means that all the paths from y to nodes in G^2/x pass through node x. Let w be a generic node in G^2/x. By hypothesis, y does not get disconnected from G if we remove x, since x is a critical node but not an articulation point; this means that y in G is connected to one or more nodes that get pruned in G^2. Let z be one of these nodes. Since x and y are connected in G^2, and G is connected, we can consider the following four paths: 1) the path from x to y in G^2, which does not use z, 2) the path from y to z in G, which does not include x, 3) the path from x to w in G^2, which does not use y nor z, 4) the path from z to w in x. If we consider these paths together, it is easy to see that we obtain a graph that includes a cycle, in which one of the following holds: either z belongs to the cycle, or another node $z' \notin G^2$, connected to z through a simple path, belongs to the cycle. In both cases, we have a node of degree two (since it belongs to a cycle), either z or z', that does not belong to G^2. This is a contradiction, since all the nodes in the cycle should belong to G^2. □

It is easy to see that the converse of Lemma 1 does not hold. Furthermore, this relation does not hold for higher level cores, that is, an articulation point of the 3-core is not necessarily an articulation point of the 2-core of the graph.

Let us define the *impact* of a node v as the number of nodes that get disconnected from the largest connected component when v is removed. Lemma 1 provides support to the intuitive fact that the critical nodes have a bigger impact than the generic articulation points: indeed, if an articulation point x gets pruned in the computation of the 2-core, this means that it is connected to other nodes that get pruned and to exactly one node of G^2. Thus, this node of G^2 is, by definition, an articulation point, since its removal disconnects x and all the nodes connected to it, and its impact is at least equal to the impact of x plus one[2]. In particular, for each articulation point not in G^2 there should be a critical node in G^2 with higher impact. In the next section we will see that, whilst all the considered real social networks confirm this intuition, criminal networks exhibit consistently a different behavior.

4 Experimental Results

In this section we describe our experimental results. In particular, before detailing our findings, we describe the datasets and the metrics we considered.

4.1 Datasets

In Table 1 we show the type, the number of nodes (n), the number of edges (m) and the repository of the networks in our dataset. Our criminal networks include `Drug-Traffic`, a network collected in November 2007 during a police investigation against a drug-trafficking group, and `Terrorists`, collected over a period of several years of an ongoing intelligence operation against terrorism. The social networks we considered include `Karate`, which is amongst the most studied in SNA after its first appearance in the work of Zachary [26], and describe the friendship relations in a karate club in a US university in the 1970s. `Science` is a co-authorship network of scientists compiled by Newman [19], whilst `Facebook` collects friends' lists of survey participants; it has been collected by McAuley and Leskovec [17]. Among social networks, we also consider two *fictitious* social networks; the first one, `Lindenstrasse`, has been used in the Graph Drawing Conference contest in '99 [14] and describes the relationships between the characters of the german soap opera Lindenstrasse. The other dataset, `Marvel`, describes the relationships in the Marvel Comics Universe (thus including, amongst many many others, superheroes like Spiderman, Capitan America, and Iron Man); it has been collected by Alberich et al. [1]. For the sake of comparison, we also consider three networks from different domain application areas: `Airlines`, that describes airlines connections in US cities, `PowerGrid`, collected by Watts and Strogatz [24], which represents the topology of the Western States Power Grid of the United States, and `ASes`, which is snapshot of the structure of the Internet at the level of autonomous systems, collected by Newman from 2006 data. We refer the interested reader to the cited repositories and references for more details about each dataset and their format.

[2] Note that, since the impact is defined as the number of nodes that gets disconnected by the main connected component, it might be the case that the pruned nodes are more than the ones in G^2, thus the impact of an articulation point might be bigger than the one of the node that connects it to G^2.

Table 1. The network datasets analyzed in this paper

Name	**Type**	**Nodes** (n)	**Edges** (m)	**Source**
`Drug-Traffic`	Criminal Network	2749	13578	Mainas [15]
`Terrorists`	Criminal Network	4275	7874	Mainas [15]
`Karate`	Social Network	34	78	GEPHI [13]
`Science`	Social Network	1589	2742	Pajek [5]
`Facebook`	Social Network	4039	88234	SNAP [22]
`Lindenstrasse`	Fictitious SN	233	325	Pajek [5]
`Marvel`	Fictitious SN	10822	314054	GEPHI [13]
`Airlines`	Other	235	1297	GEPHI [13]
`PowerGrid`	Other	4941	6594	GEPHI [13]
`ASes`	Other	22963	48436	GEPHI [13]

4.2 Metrics

We recall that the *impact* of a node v is given by the number of nodes that get disconnected from the largest connected component when v is removed. In our analysis, we compute for each network in our dataset the following information:

1. the number of articulation points, in short APs;
2. the number of critical nodes, in short CNs;
3. the number of critical nodes in the 3-core, in short CN^{3}s;
4. the number of CNs and CN^{3}s that belong to the top K articulation points, sorted by the impact, where K is respectively the number of CNs or CN^{3}s.

Let us clarify (4) with an example: the `Marvel` network has 107 APs, 3 CN, and 2 CN^{3}s. The top five APs have impact values equals to $\{39, 27, 15, 13, 12\}$. The three CNs have impact values equals to $\{39, 4, 3\}$, whilst the two CN^3 have both impact equals to 1. Of the three CNs only the first one belongs to the top three APs. Of the two CN^3, none belongs to the top two APs. We call this ratio the "membership ratio", as detailed in the followng.

Membership and Weighted Impact Ratio. We formally define the *membership ratio* as $\frac{|\text{CNs} \cap (\text{Top APs})|}{|\text{CNs}|}$. For example, this ratio equals $\frac{1}{3}$ in the case of the CNs of the `Marvel` network. We can consider this ratio as a measure of the relative importance of the CNs (and CN^3) when compared to the superset of the APs. Another measure of the relative importance of CNs (and CN^3) that we will use in the next section is the *weighted impact ratio*: the sum of the impacts of the CNs (or of the CN^3) divided by the sum of the top K APs. It is easy to verify that this value ranges from 0 to 1. For example, in the case of the `Marvel` network, the weighted impact ratio is equal to $\frac{39+4+3}{39+27+15} = \frac{46}{81} \approx 0.56$ for the CNs. In Figures 3 (CNs) and 4 (CN^{3}s) we report, for each network, the membership ratio and the weighted impact ratio as percentage values.

4.3 Experimental Findings

In Table 2 we show, for each considered network, the size and the number of articulation points of the largest connected components and of G^2 and G^3.

Table 2. In this table we report, for each network, the size of its largest connected components and its 2-core and 3-core, together with the number of articulation points in each of these components

	G		largest CC			G^2			G^3		
	n	m	n	m	APs	n	m	APs	n	m	APs
`Drug-Traffic`	2749	13578	1554	2216	280	454	1116	15	163	588	0
`Terrorists`	4275	7874	4085	6358	521	1303	3576	77	681	2440	7
`Karate`	34	78	34	78	1	33	78	1	22	55	1
`Science`	1589	2742	379	914	57	352	887	44	265	736	1
`Facebook`	4039	88234	4039	88234	11	3964	88159	7	3856	87952	1
`Lindenstrasse`	233	325	233	325	74	123	215	1	12	19	0
`Marvel`	10822	314054	10822	314054	107	10543	313775	3	9935	312565	2
`Airlines`	235	1297	235	1297	9	201	1263	2	162	1190	1
`PowerGrid`	4941	6594	4941	6594	1229	3353	5006	94	231	479	1
`ASes`	22963	48436	22963	48436	1870	14966	40439	10	4383	19678	1

We note that, in almost all the considered networks, the number of critical points is much smaller than the number of articulation points. The most notable exception is `Karate`, which has only one articulation point that is also a critical node.

In Table 3 we show the impact of nodes removal: we report for each network, the maximum impact of an articulation point (AP), of a critical node (CN), and of a critical node of the 3-core (CN^3). We also report the number of CNs and CN^{3}s that are in the top K articulation points, as in Section 4.2. It is impressive to see that in all the considered social networks the top APs are exactly the CNs, whilst this does not happen in the criminal networks. In the case of fictitious SNs, in `Lindenstrasse` there is only one critical node that is the one with maximum impact. The same does not hold in the `Marvel` network, where there are few APs and CNs, if compared to the size of the whole network. For the other network types, exactly half of the CNs belong to the top APs in both `Airlines` and `PowerGrid`; it is interesting to notice that in the Autonomous System network dataset, we have only 3 CNs in the top APs, but this three are the first, the second and the fourth, thus confirming the findings of Ausiello et al. [2].

Our findings can be summarized by the plots in Figure 3 and 4. In Figure 3 we show, for each network, the membership ratio and the weighted impact of the critical nodes ratio as percentage values. Here it is possible to distinguish, at a glance, criminal networks from social networks: all the considered social networks have the maximum values in these ratios. The same happens also for the fictitious `Linderstrasse`; in the `ASes` networks, with only three CNs in the top ten APs, the total weighted impact ratio is slightly more than 50%. Finally, in Figure 4 we consider the CN^{3}s for all the networks except `Drug-Traffic` and `Lindenstrasse`, that have no CN^{3}s, i.e. their 3-core G^3 is biconnected. We see a clear difference between the real (not fictitious) social networks and `Terrorist`.

From the results shown it seems that there is a strong difference between criminal networks and real social networks, if we focus on the relative importance of CNs and APs. The fictitious networks considered have two different behaviors:

Table 3. In this table we report, for each network, the number of APs, of CNs and CN^3 and their maximum impact. We also report, for CNs and CN^3, the number of them included in the top APs, sorted by their impact.

	1CC	G^2		G^3		Max Impact		
	APs	CNs	CNs in TOP APs	CN^3s	CN^3s in TOP APs	AP	CN	CN^3
`Drug-Traffic`	280	15	8	0	–	172	57	–
`Terrorists`	521	77	40	7	2	83	83	83
`Karate`	1	1	1	1	1	7	7	7
`Science`	57	44	44	1	1	60	60	60
`Facebook`	11	7	7	1	1	197	197	197
`Lindenstrasse`	74	1	1	0	–	13	13	–
`Marvel`	107	3	1	2	0	39	39	1
`Airlines`	9	2	1	1	1	11	11	11
`PowerGrid`	1229	94	47	1	0	106	106	8
`ASes`	1870	10	10	1	1	333	333	333

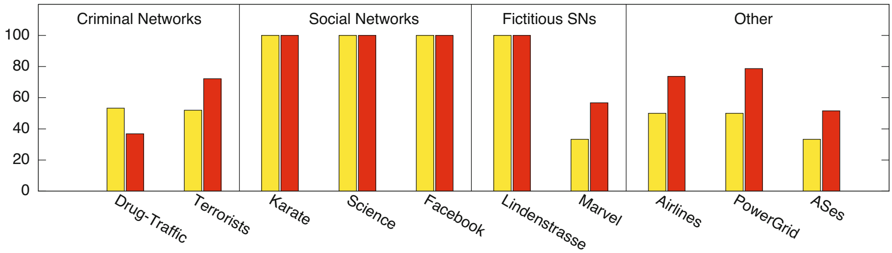

Fig. 3. Overall impact of Critical Nodes: the *membership ratio* (yellow) and the *weighted impact ratio* (red) as percentage values ranging from 0 to 100%. (Best viewed in colors).

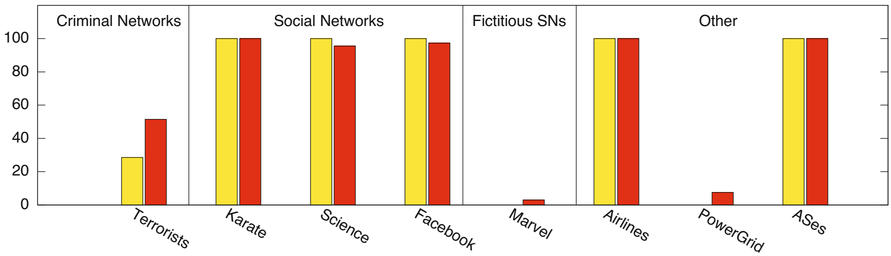

Fig. 4. Overall impact of CN^3, the Critical Nodes of the 3-core:the *membership ratio* (yellow) and the *weighted impact ratio* (red) as percentage values ranging from 0 to 100%. (Best viewed in colors).

Linderstrasse seems a real social network, whilst Marvel does not, and this appears to be consistent with the findings of Alberich et al. [1]. The networks classified as "other" exhibit a different behavior, but this seems due to their physical nature: all of them are, in some sense, infrastructure networks, and thus we do not expect them to behave like social networks for centrality aspects.

5 Conclusions

In this paper we showed that, if we focus on critical nodes, criminal networks and social networks exhibit a consistently different behavior, and critical nodes "deserve their name" only in the case of social networks. We analyzed two criminal networks, from [15], and compared them with real and fictitious social networks, and also networks from different application domains. Our findings confirm, from a different perspective, what observed by Baker and Faulkner [4]: in criminal networks important actors are not the ones whose removal affects a large portion of the network.

Acknowledgements. We are indebted to Efstathios D. Mainas for making their criminal network datasets in [15] available to us.

References

1. Alberich, R., Miro-Julia, J., Rossello, F.: Marvel Universe looks almost like a real social network. eprint arXiv:cond-mat/0202174, February 2002
2. Ausiello, G., Firmani, D., Laura, L.: Real-time analysis of critical nodes in network cores. In: IWCMC, pp. 42–46. IEEE (2012)
3. Ausiello, G., Firmani, D., Laura, L.: The (betweenness) centrality of critical nodes and network cores. In: Saracco, R., Ben Letaief, K., Gerla, M., Palazzo, S., Atzori, L. (eds.) 2013 9th International Wireless Communications and Mobile Computing Conference, IWCMC 2013, Sardinia, Italy, July 1–5, 2013, pp. 90–95. IEEE (2013)
4. Baker, W.E., Faulkner, R.R.: The social organization of conspiracy: Illegal networks in the heavy electrical equipment industry. American Sociological Review, 837–860 (1993)
5. Batagelj, V., Mrvar, A.: Pajek datasets (2006)
6. Batagelj , V., Zaveršnik, M.: Generalized cores. Preprint 799, University of Ljibljana (2002)
7. Borgatti, S.P.: The key player problem. In: Dynamic Social Network Modeling and Analysis: Workshop Summary and Papers, p. 241. National Academies Press (2003)
8. Borgatti, S.P.: Centrality and network flow. Social Networks **27**(1), 55–71 (2005)
9. Borgatti, S.P.: Identifying sets of key players in a social network. Computational & Mathematical Organization Theory **12**(1), 21–34 (2006)
10. Duijn, P.A.C., Kashirin, V., Sloot, P.M.A.: The relative ineffectiveness of criminal network disruption. Scientific Reports **4** (2014)
11. Ferrara, E., De Meo, P., Catanese, S., Fiumara, G.: Detecting criminal organizations in mobile phone networks. Expert Systems with Applications **41**(13), 5733–5750 (2014)

12. Ferrara, E., De Meo, P., Catanese, S., Fiumara, G.: Visualizing criminal networks reconstructed from mobile phone records. In: Hypertext 2014 Extended Proceedings: Late-breaking Results, Doctoral Consortium and Workshop Proceedings of the 25th ACM Hypertext and Social Media Conference (Hypertext 2014), Santiago, Chile, September 1–4, 2014 (2014)
13. GEPHI sample datasets. https://wiki.gephi.org/index.php/Datasets
14. GD99 Graph Drawing Conference Contest. http://kam.mff.cuni.cz/conferences/GD99/contest/rules.html
15. Mainas, E.D.: The analysis of criminal and terrorist organisations as social network structures: a quasi-experimental study. International Journal of Police Science & Management **14**(3), 264–282 (2012)
16. Masys, A.: Networks and network analysis for defence and security. In: Proceedings of the 2013 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining, pp. 1479–1480. ACM (2013)
17. Mcauley, J., Leskovec, J.: Discovering social circles in ego networks. ACM Trans. Knowl. Discov. Data **8**(1), 4:1–4:28 (2014)
18. Morselli, C.: Inside criminal networks. Springer (2009)
19. Newman, M.E.J.: Finding community structure in networks using the eigenvectors of matrices. Phys. Rev. E **74**, 036104 (2006)
20. Schwartz, D.M., Rouselle, T.D.A.: Using social network analysis to target criminal networks. Trends in Organized Crime **12**(2), 188–207 (2009)
21. Seidman, S.B.: Network structure and minimum degree. Social Networks **5**(5), 269–287 (1983)
22. SNAP: Stanford Network Analysis Project. http://snap.stanford.edu/
23. Sparrow, M.K.: The application of network analysis to criminal intelligence: An assessment of the prospects. Social Networks **13**(3), 251–274 (1991)
24. Watts, D.J., Strogatz, S.H.: Collective dynamics of 'small-world' networks. Nature **393**, 440–442 (1998)
25. Xu, J., Chen, H.: Criminal network analysis and visualization. Commun. ACM **48**(6), 100–107 (2005)
26. Zachary, W.W.: An information flow model for conflict and fission in small groups. Journal of Anthropological Research **33**, 452–473 (1977)