DOI:10.1145/1400181.1400198

Knowing the structure of criminal and terrorist networks could provide the technical insight needed to disrupt their activities.

BY JENNIFER XU AND HSINCHUN CHEN

The Topology of Dark Networks

SCIENTISTS FROM A variety of disciplines, including physics, sociology, biology, and computing, all explore the topological properties of complex systems that can be characterized as large-scale networks, including scientific collaborations, the Web, the Internet, electric power grids, and biological and social networks. Despite the differences in their components, functions, and size, they are surprisingly similar in topology, leading to the conjecture that many complex systems are governed by the ubiquitous "self-organizing" principle, or that the internal complexity of a system increases without being guided or managed by external sources.

Still missing from this line of research, however, is an analysis of the topology of "dark" networks hidden from view yet that could have devastating effects on our social order and economy. Terrorist organizations, drug-trafficking rings, arms-smuggling operations, gang enterprises, and many other covert networks are dark networks. Their structures are largely unknown to outsiders due to the difficulty of accessing and collecting reliable data. Do they share the same topological properties as other types of empirical networks? Do they follow the self-organizing principle? How do they achieve efficiency under constant surveillance and threat from the authorities? How robust are they against attack? Here, we explore the topological properties of several covert criminal- and terrorist-related networks, hoping to contribute to the general understanding of the structural properties of complex systems in hostile environments while providing authorities insight regarding disruptive strategies.

Topological analysis focusing on the statistical characteristics of network structure is a relatively new methodology for studying large-scale networks.[1,11] Large complex networks can be categorized into three types: random, small-world, and scale-free.[1] A number of statistics (see Table 1) have been developed to study their topology; three of which—average path length, average clustering coefficient, and degree distribution—are widely used to categorize networks.

In random networks, two arbitrary nodes are connected with a probability p; as a result each node has roughly the same number of links. Random networks are characterized by small l, small C, and bell-shaped Poisson distributions.[1] A small l means an arbitrary node can reach any other node in a few steps. A small C implies that random networks are not likely to contain clusters and groups. Studies by physicists and computer and social scientists have found that most complex systems are not random but present small-world and scale-free properties (see Albert[1] for a comprehensive review of these studies).

A small-world network has a significantly larger C than its random-network counterpart while maintaining a relatively small l.[11] Scale-free networks, on the other hand, are char-

ILLUSTRATION BY LEANDER HERZOG

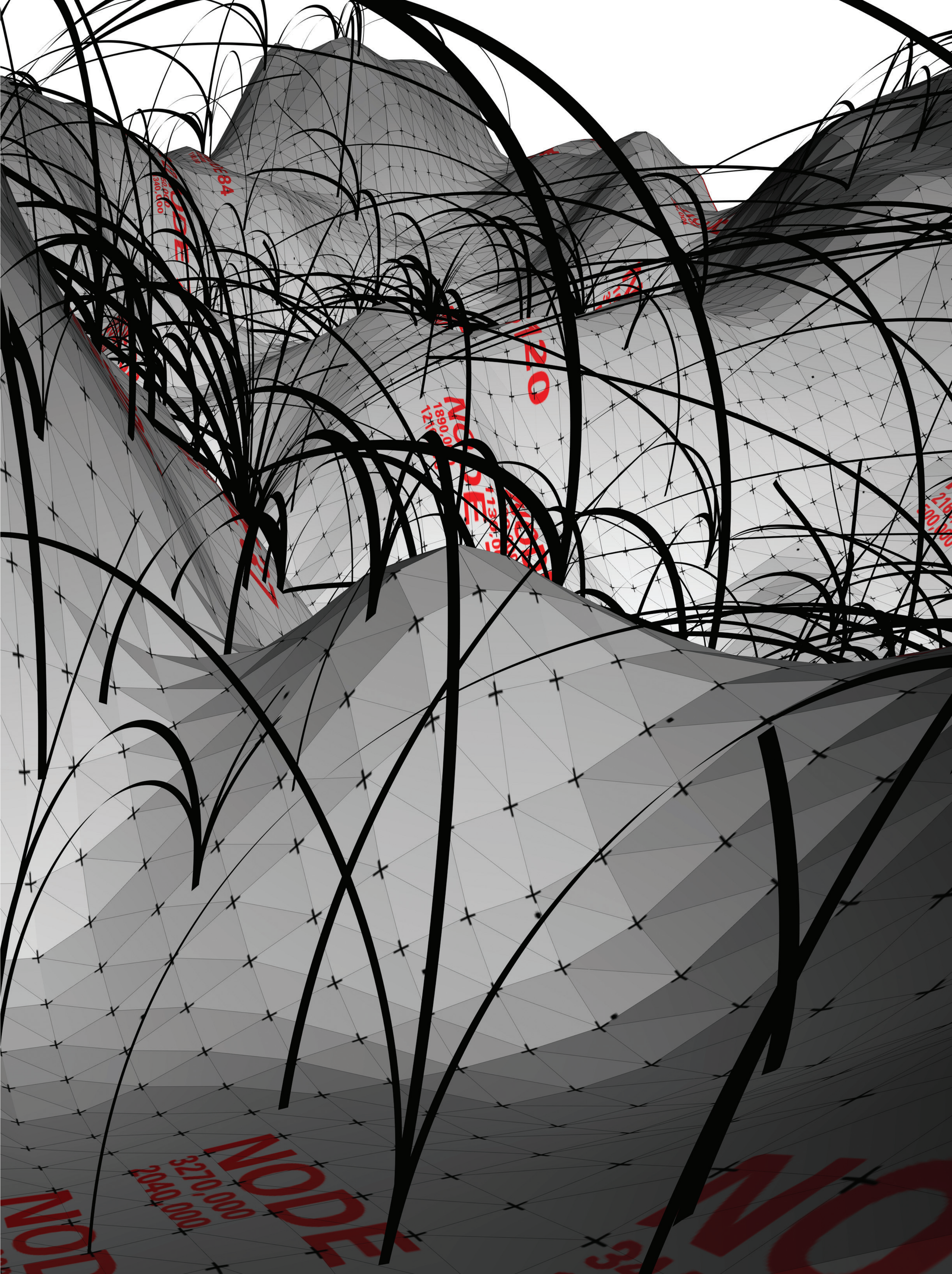

NODE
3270,000
2040,000
NODE
1890,000
120
84
340,000

Figure 1: The giant component in the GSJ Network (data courtesy of Marc Sageman[9]). The terrorists belong to one of four groups: Al Qaeda or Central Staff (pink), Core Arabs (yellow), Maghreb Arabs (blue), and Southeast Asians (green). Each circle represents one or more terrorist activities (such as the September 11 attacks and the Bali bombing) as noted.

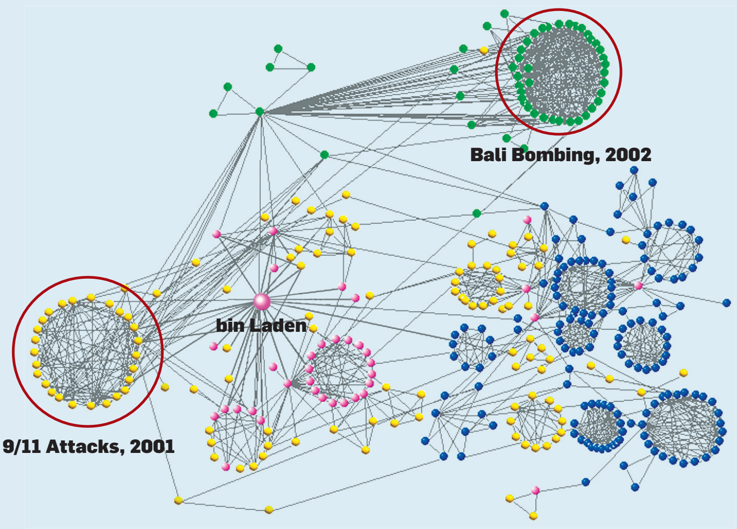


Table 1: Statistics we used for studying network topology.

Statistics	Description
Average Path Length, l[1]	The average of the lengths of the shortest paths between all pairs of nodes in a network.
Average Clustering Coefficient, C[11]	The average of all individual clustering coefficients, C_i, which is the number of links that actually exist among node i's neighbors over the possible number of links among these neighbors.
Average Degree, $<k>$[10]	The average of all individual degrees, k_i, which is the number of links that node i has.
Degree Distribution, $p(k)$[1]	The probability that an arbitrary node has exactly k links.
Link Density, d[10]	The number of links that actually exist over the possible number of links in a network.
Assortativity, r[8]	The Pearson correlation between the degrees of two adjacent nodes.
Global Efficiency, e[4]	The average of the inverses of the lengths of the shortest paths over all pairs of nodes in a network.

acterized by the power-law degree distribution, meaning that while a large percentage of nodes in the network has just a few links, a small percentage of the nodes have a large number of links.[1] Scientists conjecture that scale-free networks evolve following the self-organizing principle, where growth and preferential attachment play a key role in the emergence of the power-law distribution. Preferential attachment implies that the more links a node has, the more new links it is able to attract, manifesting the "rich-get-richer" phenomenon.

Analyzing the topology of complex systems has important implications for our understanding of nature and society. Research has shown that the function of a complex system may be affected to a great extent by its network topology.[1] For instance, the Web's short average path length makes cyberspace a convenient, navigable system in which any two Web pages are (on average) only 19 clicks away from each other. It also has been shown that the greater tendency for clustering in metabolic networks corresponds to the organization of functional modules in cells, contributing to the behavior and survival of organisms. In addition, networks with scale-free properties are highly robust against random failure and errors but notably vulnerable to targeted attacks.[5]

Methods and Data

To understand the topology and function of dark networks we studied four terrorist- and criminal-related networks:

Global Salafi Jihad (GSJ).[9] This terrorist network's 366 members (see Figure 1) include some from Osama bin Laden's Al Qaeda, connected by, perhaps, kinship, friendship, religious ties, and relationships formed after they joined. The GSJ data was provided to us by Marc Sageman, a forensic psychiatrist in private practice in Philadelphia and author of *Understanding Terror Networks*.[9] The network was constructed entirely from open-source data, including publicly available documents and transcripts of court proceedings and press, scholarly, and Web articles. Sageman scrutinized and cross-validated the information about all nodes (terrorists) and links (relationships). However, as he pointed out in his book, the data is also subject to several limitations. First, the members in the network may not be a representative sample of the global Salafi jihad. The data may be biased toward leaders and members captured or identified in attacks. Second, because most of the sources were based on retrospective accounts, the data may be subject to self-reported bias. Despite these limitations, the data provides stunning insight into clandestine terrorist organizations.

Meth World. In trafficking illegal methamphetamines,[12] this network consisted of 1,349 criminals traced and investigated by the Tucson Police Department from 1985 to 2002. Because no information about the social relationships among them is directly available, we were granted access to the police databases and retrieved all the crime incidents in which these people were involved from 1985 to 2002. We created a link between any two of them if they committed at least one crime together for which they were convicted.

Although the network was carefully validated by the crime analysts in the Tucson Police Department,[12] the co-occurrence links we generated from crime-incident records may not reflect the real relationships among the criminals. Two related criminals would appear to be unconnected if, for example, they never committed a crime together. On the other hand, a coincidental link may connect two criminals if they happened to have participated in the same crime. These two problems—missing link and coincidental link—are also common in other types of networks (such as those involving movie actors[12]) based on the co-occurrence of two nodes in the same events or activities.

Another group of 3,917 criminals involved in gang-related crimes in Tucson from 1985 to 2002.[12] As in Meth World, the links in this network were generated through co-occurrence analysis of the crime-incident records.

A terrorist Web site network ("the Dark Web"). In 2005, based on reliable government sources, we identified 104 Web sites created by four major international terrorist groups—Al-Gama'a al-Islamiyya, Hizballa, Al-Jihad, and Palestinian Islamic Jihad—fetching all of their pages and extracting all of their hyperlinks. We recognized a link between any two Web sites if at least one hyperlink existed between any two Web pages in them.

Results

Table 2 lists the basic statistics of the four elicited networks. Like many other empirical networks, each of them contains many isolated components and a single giant component. The giant component in a graph is defined as the largest connected subgraph.[1] The separation between the 356 terrorists in the GSJ network and the remaining 10 terrorists is because we found no valid evidence to connect the 10 terrorists to the giant component in the network. The giant components in Meth World and the gang network contain only 68.5% and 57.0% of the nodes, respectively. This may be because we collected the data from a single law-enforcement jurisdiction that might lack complete information about all relationships among criminals, causing missing links between the giant component and the smaller components. The isolated components in the Dark Web are possibly the result of the differences in the four terrorist groups' distinctive ideologies.

As in many other network-topology studies (such as Barabási[2]), we performed a topological analysis on only the giant component in the four elicited networks. Table 2 lists the average degrees and maximum degrees of the four networks, showing that some terrorists in the GSJ network and some terrorist Web sites in the Dark Web are extremely popular, connecting to more than 10% of their nodes.

This "assortativity" reflects the tendency for nodes to connect with others that are similarly popular in terms of link degree. The assortativity coefficients of the four networks are all significantly different from 0. The GSJ and the gang networks present positive assortativity, meaning that popular members tend to connect with other popular members. In positively assortative networks, high-degree nodes tend to cluster together as core groups,[8] a phenomenon evident in the GSJ network in which bin Laden and his closest cohorts form the core of the network and issue commands to other parts of the network.[9] In contrast, Meth World and the Dark Web have negative assortativity coefficients, or "disassortativity."

Meth World consists of drug dealers selling illegal methamphetamine to many individual buyers who do not connect with many other buyers or dealers. Moreover, studies have found that street drug-dealing organizations are led by a few high-level individuals who connect with a large number of low-level retail drug dealers.[6] Because high-degree nodes connect to low-degree nodes, Meth World is characterized by disassortative mixing patterns. On the other hand, the disassortativity in the Dark Web is the result of the fact that the popular Dark Web sites routinely receive many inbound hyperlinks from less popular Web sites.

To ascertain if the dark networks are small worlds, we calculated average path lengths, clustering coefficients, and global efficiency (see Table 3). For each network, we generated 30 random counterparts with the same number of nodes and the same number of links as in the corresponding elicited networks. We found that all of them have significantly high clustering coefficients compared to their random counterparts. Moreover, although the

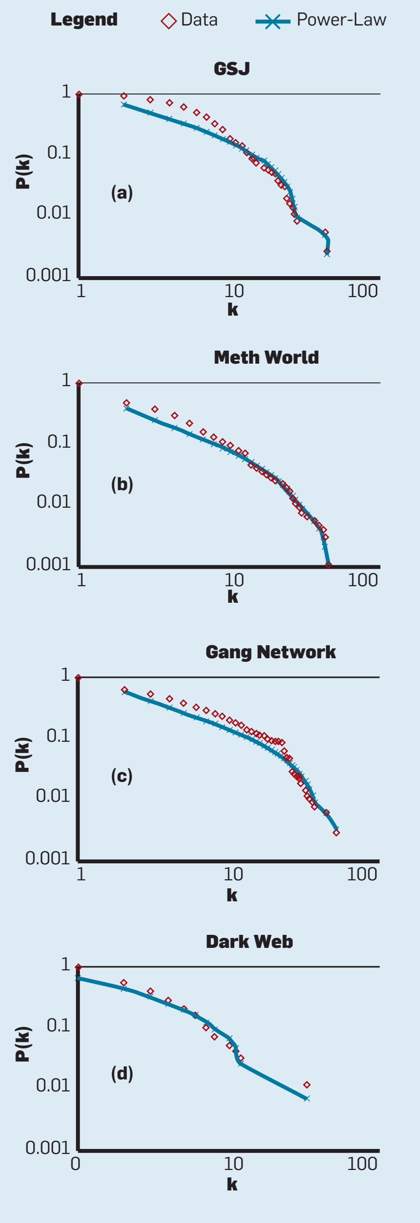
Table 2. Basic statistics and scale-free properties concerning dark networks. The numbers in parentheses in the third row are the percentage of total nodes included in the giant components. The numbers in parentheses in the fifth row are the percentage of total nodes connected to the highest-degree nodes.
**** p-value < 0.05 * p-value < 0.01**

	GSJ	Meth World	Gang Network	Dark Web
Number of Nodes, n	366	1349	3917	104
Number of Links, m	1247	4784	9051	156
Size of Giant Component	356 (97.3%)	924 (68.5%)	2231 (57.0%)	80 (77.9%)
Average Degree, $<k>$	6.97	4.62	5.74	3.88
Maximum Degree	44 (12.4%)	37 (4.0%)	51 (2.3%)	33 (41.3%)
Link Density, d	0.02	0.01	0.003	0.05
Assortativity, r	0.41**	-0.14**	0.17**	-0.24*
Power-Law Distribution Exponent, γ	1.38	1.86	1.95	1.10
Goodness of Fit, R^2	0.74	0.89	0.81	0.82

differences are statistically significant (greater than three standard deviations), the average path length of the four networks (except for the gang network) is just slightly greater than their random counterparts.

These small-world properties imply that terrorists or criminals are able to connect with any other member in a network through only a few mediators. In addition, the networks are sparse, with very low link density. These properties have important implications for the communication efficiency of the networks. Due to the increased risk of being detected by authorities as more people are involved in a network, short path length and link sparseness help lower the risk of detection and enhance efficiency of communication. As a result, the global efficiency of each network is compatible to their random-network counterparts.

Figure 2: Cumulative degree distributions: (a) GSJ network, (b) Meth World, (c) gang network, and (d) Dark Web.



On the other hand, a high clustering coefficient contributes to the local efficiency of all four dark networks. Previous studies have shown evidence of groups and teams in these networks in which members tend to have denser and stronger relationships with one another.[9,12] Communication among group members becomes more efficient, making a crime or an attack easier to plan, organize, and execute.

We also calculated the path length of other nodes to central nodes, finding that members in the three studied criminal and terrorist networks are extremely close to their leaders. For example, the terrorists in the GSJ network are on average only 2.5 links away from bin Laden himself, meaning his command is able to reach an arbitrary member through only two mediators. Similarly, the average path length to the leader of Meth World is only three links.[12] Such a short chain of command also means communication efficiency.

Special attention should be paid to the Dark Web. Despite the small size of its giant component (80 nodes), the average path length is 4.70 links, only slightly larger than the 4.20 links in the GSJ network, which has almost nine times more nodes. Since hyperlinks help visitors navigate Web pages and because terrorist Web sites are often used for soliciting new members and donations, the relatively long path length may be due to the reluctance of terrorist groups to share resources with other terrorist groups.

Moreover, the dark networks present scale-free properties with power-law degree distributions in the form of $p(k) \sim k^{-\gamma}$. Because degree-distribution curves fluctuate, we display the cumulative degree distributions, *P(k)*, in a log-log plot (see Figure 2). *P(k)* is defined as the probability that an arbitrary node has at least *k* links. Figure 2 also outlines the fitted power-law distributions. The last two rows of Table 1 report the exponent value, γ, and the goodness-of-fit, R^2, for each network. Figure 2 shows that all these networks are scale-free. The power-law distributions fit especially well at the tails. Note that the three human networks display two-regime scaling behavior, which has also been observed in other empirical networks (such as those involving scientific collaboration).[2]

Two mechanisms have been proposed to account for the emergence of two-regime power-law degree distributions during the evolution of a network.[2] First, new links may emerge between existing network members. This emergence implies that criminals or terrorists who were not related previously could become related over time. This assumption is logical since two unacquainted members could become acquainted through a third member who knows each of them. In the GSJ network, 22.6% of the links were post-joining ties formed among existing members. Second, an existing link may be

Table 3: Small-world properties of dark networks. Each network includes the metrics in the elicited network (data) and the metrics in the random graph counterpart (random). Numbers in parentheses are standard deviations.

	GSJ		Meth World		Gang Network		Dark Web	
	Data	Random	Data	Random	Data	Random	Data	Random
Average Path Length, *l*	4.20	3.23 (0.040)	6.49	4.52 (0.056)	9.56	4.59 (0.034)	4.70	3.15 (0.108)
Average Clustering Coefficient, *C*	0.55	0.020 (0.0029)	0.60	0.005 (0.0014)	0.68	0.002 (0.0005)	0.47	0.049 (0.0155)
Global Efficiency, *e*	0.28	0.33 (0.004)	0.18	0.23 (0.003)	0.12	0.23 (0.001)	0.30	0.34 (0.019)

rewired—a strong possibility in GSJ and the Dark Web. However, such rewiring would not affect Meth World or the gang network because a co-occurrence link could not be rewired once it was created.

An interesting topology-related question is what mechanisms play a role in producing the properties we observed in dark networks? Short average path length, high clustering coefficient, power-law degree distributions with two-regime scaling behavior in the human networks? That is, can we regenerate the four dark networks based on known mechanisms (such as growth and preferential attachment)? To answer, we conducted a series of simulations in which we generated 30 networks for each elicited human network based on three evolutionary mechanisms:

Growth. Starting with a small number of nodes, at each time step we add a new node to connect with existing nodes in the network;

Preferential attachment. The probability that an existing node will receive a link from the new node depends on the number of links the node already maintains. The more links it has the more likely it will receive a new link; and

New links among existing nodes. At each time step, a random pair of existing nodes may connect, depending on the number of common neighbors they have. The more common neighbors they share the more likely they will also be connected with each other.

We expected that the first two mechanisms would generate a power-law degree distribution[1] and that the third would generate a high clustering coefficient and two-regime scaling behavior.[2] Our simulations showed that the power -law degree distributions are easily regenerated, with R^2 ranging from 0.83 (the gang network) to 0.88 (GSJ). The two-regime scaling behavior was also present in the simulated networks for the human networks. However, the highest clustering coefficient in a simulation was only 0.24 (GSJ), far less than what we obtained from the elicited networks (0.55–0.68). This finding implies that some other mechanisms must have contributed to the substantially high clustering coefficients we observed in the dark networks. We suspect that member recruitment is one such mechanism. Employing active recruitment methods, subgroups of terrorists or criminals are able to attract new members into their groups. The new members quickly become acquainted with many existing members, substantially increasing the clustering coefficients.

The terrorists in the GSJ network are on average only 2.5 links away from bin Laden himself, meaning his command is able to reach an arbitrary member through only two mediators.

Caveats

A notable point is that two problems may have affected the structures of the three elicited human networks—GSJ, Meth World, and the gang network. First, they may have missing links that can cause the networks to appear to be less efficient; there may actually be hidden "shortcuts" connecting distant parts of the networks. Second, the presence of coincidental "fake" links might cause the elicited networks to be more efficient than they would otherwise be since these links are not communication channels.

To test how the results would be affected by missing links, we added various percentages of the existing links to the elicited networks based on three effects used in missing-link-prediction research:[7]

Random effect. A link is added between a randomly selected pair of nodes not originally connected;

Common neighbor effect. A link is added between a pair of unconnected nodes if they share common neighbors; the more common neighbors they share the more likely they will be connected; and

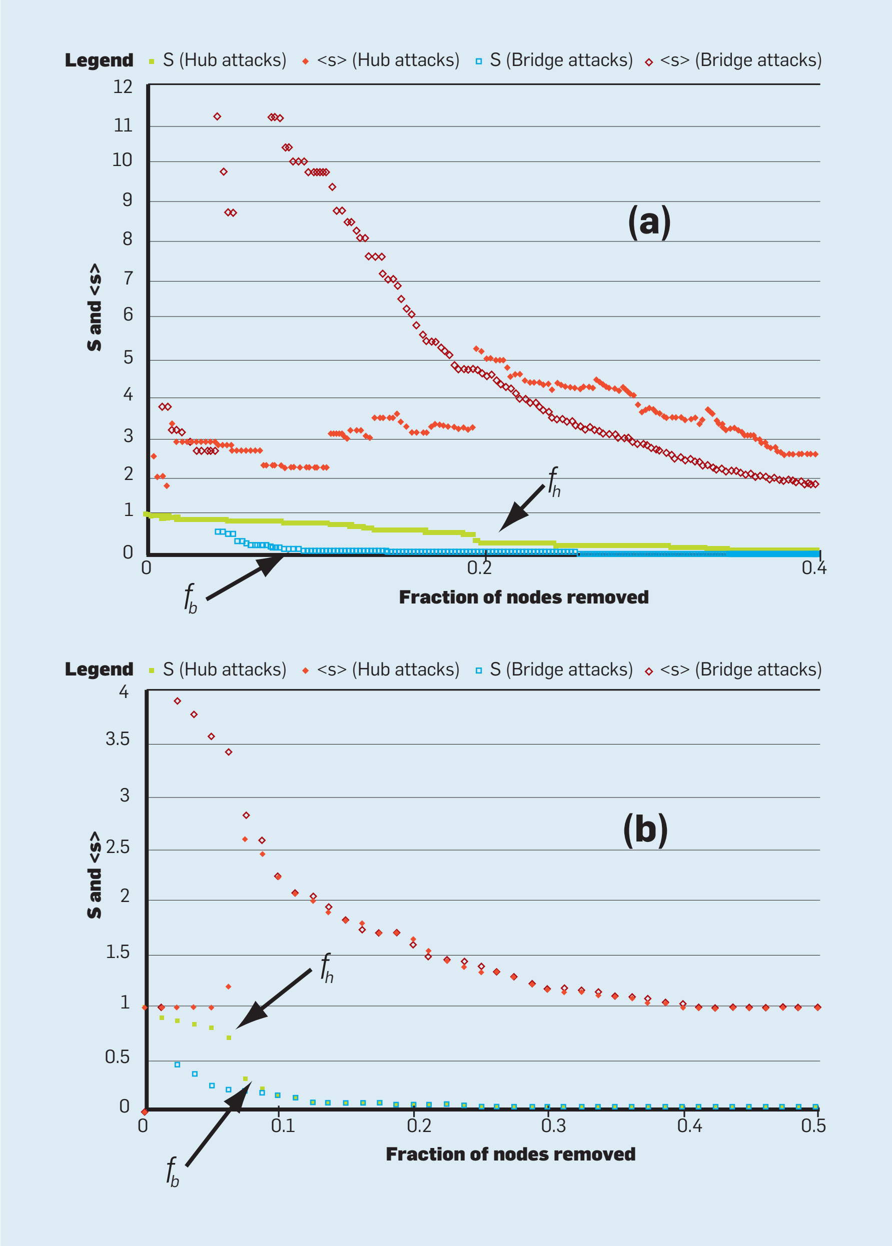
Preferential attachment effect. The probability that a pair of unconnected nodes will be linked together depends on the product of their link degrees.

We found that the small-world and scale-free properties of the four networks do not change when missing links are added. For example, when we added up to 10% of the links, the average path lengths ranged from 3.55 links (GSJ, preferential-attachment links added) to 9.45 links (the gang network, common-neighbor links added); the clustering coefficients ranged from 0.45 (GSJ, random links added) to 0.67 (the gang network, common-neighbor links added); and the R2 of power-law degree distributions ranged from 0.61 (GSJ, random links added) to 0.93 (the gang network, preferential-attachment links added).

We also randomly removed percentages of links to test the effect of "fake" links on the results, finding they were still valid even when we removed 10% of the links.

Prior research found that network topology has a significant effect on a network's robustness against failure and attacks and that scale-free networks are robust against failure (random removal of nodes).[5] Because we found that the four dark networks have scale-free properties, we tested their robustness against only targeted attacks. We simulated two types of attacks in the form of node removal: those targeting hubs and those targeting bridges. While hubs are nodes that have many links (high degree), bridges are nodes through which pass many shortest paths (high "betweenness").[10] When simulating the attacks we distinguished between two attack strategies: simultaneous removal of a fraction of the nodes based on a measure (degree or betweenness) without updating the measure after each removal and progressive removal of nodes with the measure being updated after each removal.

Figure 3: Dark-network robustness against attacks: (a) progressive attacks against the GSJ network and (b) progressive attacks against the Dark Web. Two types of attack are hub (filled markers) and bridge (empty markers).



We plotted the changes in S (the fraction of the nodes in the giant component), $<s>$ (the average size of remaining components), and average path length after some nodes are removed. We found that progressive attacks are more devastating than simultaneous attacks. Progressive attacks are similar to "cascading failures" in the Internet where an initial failure might cause a series of failures because high-traffic volume is redirected to the next bridge node.

Figure 3 (a) and (b) shows the difference between the network reactions to bridge attacks and to hub attacks. The critical points, f, at which the network falls into many small components, are marked in the figure. The behavior of Meth World and the gang network is similar to the behavior of the GSJ network, showing that these terrorist and criminal networks are more sensitive to attacks targeting bridges than to those targeting hubs ($f_b < f_h$). However, in Figure 3(b), f_b and f_h are very close, indicating that hub attacks and bridge attacks are equally effective at disrupting a one-regime scale-free network.

These results are consistent with findings from a prior study[5] that pure scale-free networks are vulnerable to both hub and bridge attacks, while small-world networks are more vulnerable to bridge attacks. In small-world networks consisting of communities and groups, many bridges may link different communities together. Intuitively, when they are removed, the network should quickly fall apart. Note that a bridge may not necessarily be a hub since a node connecting two communities can have as few as two links. Small-world networks (such as dark networks) are thus more vulnerable to bridge attacks than to hub attacks.

In the four dark networks we studied, bridges and hubs are usually not the same nodes. The rank order correlations between degree and betweenness in GSJ, Meth World, and the gang network are 0.63, 0.47, and 0.30, respectively. Note that although bridge attacks are more devastating, strategies targeting the hubs are also fairly effective since the networks have scale-free properties. Hub attacks and bridge attacks can be equally effective in tearing apart a pure scale-free network (such as the Dark Web, with a high degree-betweenness-rank-order correlation, 0.70) in which hubs function simultaneously as bridges connecting different parts of the network.

Conclusion

Dark networks (such as those involving terrorists and criminal narcotics traffickers) are hidden from nonparticipants yet could have a devastating effect on our social order and economy. Understanding their topology yields greater insight into the nature of clandestine organizations and could help develop effective disruptive strategies. However, obtaining reliable data about dark networks is extremely difficult, so our understanding of them remains largely hypothetical. To the best of our knowledge, the data sets we explore here, though subject to limitations, are the first to allow for statistical analysis of the topologies of dark networks.

To the best of our knowledge, the data sets we explore here, though subject to limitations, are the first to allow for statistical analysis of the topologies of dark networks.

We found that the covert networks we studied share many common topological properties with other types of networks. Their efficiency in terms of communication and information flow and commands can be tied to their small-world structures, which are characterized by short average path length and a high clustering coefficient. In addition, we found that due to their small-world properties, dark networks are more vulnerable to attack on their bridges that connect different communities within them than to attacks on their hubs. This finding may give authorities insight for intelligence and security purposes.

Another interesting finding about the three elicited human networks we studied is that their substantially high clustering coefficients (not always present in other empirical networks) are difficult to regenerate based on only known network effects (such as preferential attachment and small-world effects). Other mechanisms (such as recruitment) may also play an important role in network evolution. Other research has found that alternative mechanisms (such as highly optimized tolerance) may govern the evolution of many complex systems in environments characterized by high risk and uncertainty.[3] Our future research will focus on the effects of such alternative mechanisms on network topology. In addition, our findings are all based on a static view of the networks we studied; that is, we did not consider a large variety of dynamics that might have taken place in the evolution of the networks, so evolution study is definitely in our plans for future research.

Please also note that care is needed when interpreting these findings. Because dark networks are covert and largely unknown, hidden links may be missing in the elicited networks. These links may play a critical role in maintaining the function of the covert organizations. As a result, one must be extremely cautious when a decision is to be made to disrupt them.

References

1. Albert, R. and Barabási, A.-L. Statistical mechanics of complex networks. *Reviews of Modern Physics 74*, 1 (Jan. 2002), 47–97.
2. Barabási, A.-L., Jeong, H., Zéda, Z., Ravasz, E., Schubert, A., and Vicsek, T. Evolution of the social network of scientific collaborations. *Physica A, 311*, 3–4 (Aug. 2002), 590–614.
3. Carlson, J.M. and Doyle, J. Highly optimized tolerance: A mechanism for power laws in designed systems. *Physical Review E60*, 2 (Aug. 1999), 1412–1427.
4. Crucitti, P., Latora, V., Marchiori, M., and Rapisarda, A. Efficiency of scale-free networks: Error and attack tolerance. *Physica A 320* (Mar. 2003), 622–642.
5. Holme, P., Kim, B.J., Yoon, C.N., and Han, S.K. Attack vulnerability of complex networks. *Physical Review E 65*, Article No. 056109 (May 2002), 1–14.
6. Levitt, S.D. and Dubner, S.J. *Freakonomics: A Rogue Economist Explores the Hidden Side of Everything.* William Morrow, New York, 2005.
7. Liben-Nowell, D. and Kleinberg, J. The link prediction problem for social networks. *Journal of the American Society for Information Science and Technology 58*, 7 (May 2007), 1019–1031.
8. Newman, M.E.J. Mixing patterns in networks. *Physical Review E 67*, 2, Article No. 026126 (Feb. 2003), 1–13.
9. Sageman, M. *Understanding Terror Networks.* University of Pennsylvania Press, Philadelphia, PA, 2004.
10. Wasserman, S. and Faust, K. *Social Network Analysis: Methods and Applications.* Cambridge University Press, Cambridge, U.K., 1994.
11. Watts, D.J. and Strogatz, S.H. Collective dynamics of 'small-world' networks. *Nature 393*, 6684 (June 1998), 440–442.
12. Xu, J. and Chen, H. Untangling criminal networks: A case study. In *Proceedings of the First NSF/NIJ Symposium on Intelligence and Security Informatics* (Tucson, AZ, June 2–3, 2003). Springer, Berlin, Germany, 2003, 232–248.

Jennifer Xu (jxu@bentley.edu) is an assistant professor of computer information systems in Bentley College, Waltham, MA.

Hsinchun Chen (hchen@eller.arizona.edu) is McClelland Endowed Professor in the Department of Management Information Systems and head of the Artificial Intelligence Lab at the University of Arizona, Tucson, AZ.