Social Networks 13 (1991) 251–274
North-Holland

The application of network analysis to criminal intelligence: An assessment of the prospects

Malcolm K. Sparrow

Kennedy School of Government, Harvard University, 79 John F. Kennedy Street, Cambridge, MA 02138, USA

This paper explores the opportunities for the application of network analytic techniques to the problems of criminal intelligence analysis, paying particular attention to the identification of vulnerabilities in different types of criminal organization – from terrorist groups to narcotics supply networks.

A variety of concepts from the network analysis literature are considered in terms of the promise they hold for helping law enforcement agencies extract useful information from existing collections of link data. For example, six different notions of "centrality" and the three major notions of "equivalence" are examined for their relevance in revealing the mechanics and vulnerabilities of criminal enterprises.

1. Introduction

The Intelligence Analysts Training Manual of the Metropolitan Police (Scotland Yard, London) bears as its frontispiece the statement:

> Analysis is the key to the successful use of information; it transforms raw data into intelligence. It is the fourth of five stages in the intelligence process: collection, evaluation, collation, analysis, dissemination. Without the ability to perform effective and useful analysis, the intelligence process is reduced to a simple storage and retrieval system for effectively unrelated data.

Intelligence agencies, despite this obvious awareness of the importance of intelligence analysis, have remained for the most part relatively unsophisticated in their use of analytic tools and concepts. They

typically have plenty of data, much of it computerized, but comparatively little capability for extracting useful intelligence from it.

A great deal of that data can be presented in link form – that is, as a collection of nodes, with a pattern of connections. Graph theorists call them vertices and edges. Law enforcement agencies often call them entities and relationships. Whatever they are called they clearly lend themselves to analysis as networks. Some of the more obvious examples of such data types include contact reports (two or more people seen together at some specific time and place), telephone toll data, and financial transaction data (deposits, withdrawals, or transfer of funds between accounts).

The types of network questions to which intelligence analysts need answers are quite familiar to network analysts and graph theorists: "who is central in this organization?", "which names in this database appear to be aliases?", "which three individuals' removal or incapacitation would sever this drug-supply network?", "what role or roles does a specific individual appear to be playing within a criminal organization?", or "which communications links within an international terrorist fraternity are likely to be most worth monitoring?". All of these questions have fairly direct analogues in other fields, and most network analysts would be familiar with all the relevant network concepts.

Some other network questions, of significant interest to law enforcement agencies, are considerably beyond the state-of-the-art in network analysis techniques. They nevertheless remain network analysis questions. For example: "(from bank cash deposit records) what are the anomalies in the pattern of cash flow surrounding the Californian jewellery trade?", "(from contact reports) what significant changes have taken place in the supply operation for Columbian cocaine to New York City since this time last year?", or "(from a database of electronic fund transfers) where is there evidence of smurfing [1]?".

All these are network questions. So the concepts and tools of network analysis, and network analysts, probably have a lot to offer law enforcement. It is somewhat surprising and a little disappointing, therefore, to find almost no overlap between the literatures of network

[1] "Smurfing" is the breaking up of large sums of money into smaller units, and subsequent passing of each segment through multiple accounts. Used by money launderers, the practice is designed to make the money trail extremely difficult to follow.

analysis and law enforcement. The two fields have historically been quite ignorant of one another. There are just a few papers that have begun to explore the application of network analysis to intelligence analysis (for instance, Coady 1985; Howlett 1980; Davis 1981), but they have focused on relatively simple network concepts. Davis, for example, shows the importance of "liaisons", (which he calls "brokers"), in fencing operations and relates the concepts of cliques, centrality, and network density to conspiracy theory.

But in general most intelligence analysts do not have a clear idea of what the academic discipline of network analysis is. They are certainly not aware of it as an emerging discipline. Nor do they understand why it ought to be of paramount importance to their profession.

Conversely, those few network theorists who have been inclined to step outside the realm of retrospective examination of human (and occasionally animal) behavior do not appear to have stumbled into intelligence analysis. If they have, they have kept very quiet about it.

The contention of this paper is, first, that the concepts of network analysis are highly pertinent to many forms of intelligence analysis and are currently being used seldom, if at all. Second, that there is much creative and imaginative work to be done in adapting the existing concepts and tools of network analysis for direct application to intelligence databases. Third, that collaboration between the fields of network theory and intelligence analysis would benefit both a great deal. For intelligence analysis it could ultimately produce a whole new set of sophisticated tools. For network theory it will provide a new set of concrete challenges and real-world applications, not to mention some fascinating new databases (appropriately sanitized).

The purpose of this paper is to begin to deepen the dialogue between the two fields by charting some of the areas of common interest.

The sections that follow describe the state-of-the-art in law enforcement application of network analysis (Section 2); then describe the current needs of criminal intelligence analysis and the relevant properties of intelligence databases (Section 3). Section 4 examines several network analytic concepts – for instance, centrality, equivalence, strong and weak ties – and explores their application to law enforcement. Section 5, by way of conclusion, suggests what seem to be the most logical next steps in building an effective bridge between these two fields.

2. The state-of-the-art in law enforcement uses of network analysis

2.1. Anacapa charts

Use of the Anacapa charting system, as developed by Anacapa Sciences Inc., Santa Barbara, California [2], is currently the predominant form of network analysis within law enforcement. It is used particularly frequently within major fraud investigations and by Organized Crime Squads, where understanding of large and sometimes sophisticated criminal enterprises is required.

Anacapa charts constitute a two-dimensional visual representation of link data (see Harper and Harris, 1975; Howlett, 1980; Klovdahl, 1981; Coady 1985). They provide a method of making visual sense of a mass of data. They are also an extremely useful tool for communicating the results of analysis (and thus are used as briefing aids as well as aids for analysis). Anacapa charts generally depict individuals by small circles, and relationships by lines (solid or dotted according to whether the relationship is confirmed or unconfirmed). The charts may also show rectangles enclosing one or more individuals as a method of representing membership of corporations or institutions. They clearly show who is central, who is peripheral, and visually reveal chains of links connecting one individual to another. To a network analyst they look like typical network diagrams.

However, such charting systems do not actually do any analysis; they simply communicate the results. It is up to the officer preparing the chart to perform the analysis first, based upon what he knows and understands at the time. The graphical display can only communicate what the officer performing the analysis can grasp. Coady (1985) makes this quite explicit:

> Link Analysis is the graphic portrayal of investigative data, done in a manner to facilitate the understanding of large amounts of data, and particularly to allow investigators to develop possible relationships between individuals that otherwise would be hidden by the mass of data obtained.

Klovdahl (1981) explained the value of creating a visual representation of a sociogram. He says it makes things hitherto unseen become

[2] For more information contact Anacapa Sciences, Inc., P.O. Box 519, 901 Olive Street, Santa Barbara, California 93102.

painfully obvious. It is this benefit that Anacapa charts bring to intelligence analysis.

But note that the intelligent functions still take place in the minds of the analysts or investigators, and they use the chart simply as a pictorial aid. So the predominant locus for the law-enforcement community's intelligent interpretation of networks remains somewhere between the "picture on the wall" (or on the graphics display terminal) and the analyst's brain.

The analyst working to create a chart, from a set of link data, has a number of objectives in mind. First, he or she intends that the relative spatial proximity of two individuals on his or her chart will be crudely representative of their "closeness" within the criminal organization. Hence the common experience among intelligence analysts of having to keep redrawing the graph as new information, altering the apparent proximities, becomes available.

Second, the analyst aims to design the whole chart in such a way that no two connecting lines cross one other. They will often change the spatial positions of individuals, or entities, to prevent such crossings occurring.

Third, the analyst intends that "central" on the picture imply "central" within the organization. Centrality is determined at the outset by counting the total number of known associations for each individual, and picking the person with the highest number. This practice is described by Harper and Harris (1975), and is explicit within both the FBI's "Link Analysis" training manual, and in Metropolitan Police Training manuals on Intelligence Analysis. Formation of an "association matrix" (sociogram, or adjacency matrix) is described in each of these sources as a prerequisite for preparing the chart.

The individual thus designated as central is represented by a circle placed centrally on the chart. The analysts then work outwards, seeking to satisfy the first two objectives (described above) as they go. More broadly, the total number of established links (the "degree" of the node) for any one individual in the network determines whether that individual is to be treated as central or peripheral.

To a network or graph theorist these three objectives might seem a little bizarre. It is an interesting exercise to examine them against the backdrop of developed network theory. The crude representation of some notion of proximity is familiar as the objective of multidimensional scaling (see, for an introduction to the subject, Breiger *et al.*

1975; Kruskal and Wish 1978). But Anacapa charts are automatically two-dimensional. So this first objective can be translated into network theory as an attempt to find a stress-free multidimensional scaling of the network in just two dimensions. With organizations of any complexity there is no reason to believe that such a scaling will exist. There is even less reason to expect that analysts will be able to generate them manually! [3] Of course the one good reason for keeping Anacapa charts two-dimensional (whether manual or computerized), is that use of any other kind of chart as a visual aid is somewhat awkward.

The second objective – keeping the lines from crossing – is, again, sensible from a practical standpoint but theoretically perverse. It keeps the picture uncluttered and easy to follow. But it requires the network to be planar (in the graph theoretic sense)! Again, there is no reason to think that any complex criminal network will satisfy the requirements of planarity, or that it will have a representation which is even close to planar. (See any basic introduction to Graph Theory, (e.g. Wilson 1972) for a discussion of planarity, and Kuratowski's theorem (circa 1930) for a necessary and sufficient condition for planarity.)

The third objective – representing "centrality" within the organization by "centrality" on the chart – is very reasonable. But it employs a most unsophisticated concept of centrality, namely the selection of the point or points of "maximum degree" (those with the most established connections). Moreover the context in which it is applied makes the use of maximum degree potentially misleading: the determination of centrality will depend upon *who you know most about*, rather than *who is central or pivotal in any structural sense*. The danger in this practice is that it may incline an agency to pay closest attention to those it already knows most about, individuals who may not in fact be the principal characters. The practice may therefore serve to perpetuate unfortunate and misleading biases in the initial intelligence collection. Analysts are specifically warned of these dangers during Anacapa's training.

Providing a network-theoretical commentary on Anacapa charts does not, of course, diminish their usefulness in any way. They were a highly significant innovation and will remain a valuable aid to analysts for years to come. The commentary simply places this particular device within the broader landscape of network theoretic concepts, and thereby shows up some of its essential limitations as a network analytic tool.

[3] No Law Enforcement agency known to the author uses any of the available computerized multidimensional scaling algorithms.

2.2. Computerized "link analysis"

Computers are now being used to take some of the laborious manual work out of link charting. There are commercial products available (e.g. "Enhanced Computer Network Analysis Program" [ECNA] from Anacapa Sciences, California) and others under development, which incorporate the facility to lift the traditional link chart off the paper and put it on a graphics display terminal instead. It makes storage, retrieval and amendment of charts relatively speedy and efficient. It also adds the benefits of handling elastic images; images which can be enlarged, stretched, shifted and otherwise manipulated in the many and diverse ways which screens, and mice, make possible.

So computer-aided link analysis has arrived, and is clearly here to stay. It is another very valuable addition to the analyst's toolkit. But, for the most part, it is still not the computer which does much of the analysis. It is the analyst. The computer merely provides a versatile drawing board, complete with the option of burying within the picture references or sections of text (in the style of hypertext), retrievable at the click of a button.

But production of the charts from the database is far from automatic. The analyst is responsible for that, picking and choosing links from the database and asking the computer to show them to him. The analyst then shunts entities around on the screen in an attempt to produce some meaningful, planar, stress-free, centrality-preserving two-dimensional representation of the network. During this process the computer acts as a highly versatile image storage and retrieval device, but performs relatively mundane analytic tasks. Use of modern graphic user-interfaces (with windows, hypertext, and pull-down menus) have thus produced some first-class methods of *showing the results* of link analysis. But the computers are still not *doing the analysis*.

Computer-assisted link analysis significantly speeds up the generation of link charts, and makes updating or expanding them highly efficient. However, it is important for intelligence analysts to understand that use of technology in this way has not yet brought to bear on the structural analysis of criminal organizations either the computational capacity of modern computers or the sophistication of existing network analysis techniques.

2.3. Visual investigative analysis

Several of the more technically sophisticated law enforcement agencies also use, during major crimes enquiries, some form of "Event Flow charting" (Howlett 1980). Computerized versions of Event Flow Charts have been variously called Visual Investigative Analysis, or CAVIA (computer-aided VIA).

The basic concept, just like the Anacapa charts, looks remarkably like a network. In this case events are used as nodes, and events are connected if one either caused the other, or had to happen before it. The "Event Flow Chart" is therefore a pictorial representation of the chronology of all the relevant events surrounding the commission of the crime. Unlike Anacapa charts, it has a time line, traditionally running left to right. Preparation of such charts shows up obvious disparities in witnesses' statements or in their estimates of when things happened, and often reveals potentially fruitful avenues of enquiry.

A description of such systems by the FBI (entitled simply "Visual Investigative Analysis", FBI, US Department of Justice) explains:

> Through the use of a network (flowchart), VIA graphically displays the sequential and concurrent order of events involved in a criminal act.... Leads not ordinarily discernible through file review may become more apparent when the information is chronologically arranged.

This is not commonly considered structural network analysis by network theorists. It is mentioned here because many law enforcement officials (and analysts) think of it as network analysis. From the network theorists' point of view it is interesting mostly in that it employs the concept of "causal links", or of one event depending upon another, in much the same way that PERT (Program Evaluation and Review Technique) charts and CPM (Critical Path Method) analyses do.

2.4. Template matching

Some progress has also been made in the use of computers to perform "template matching", a process which helps the analyst to determine whether or not a particular type of crime is likely to have been

committed, or whether a particular pattern of criminal relationships is in existence.

The FBI's "Big Floyd" prototype is an example of such a system (see Bayse and Morris 1987; and "An algorithm.." 1986). It performs the regular functions of storage and retrieval of link data, encompassing links of many different specified types. It does an excellent job of facilitating the interaction between investigator and the visually displayed network, or selected subgraphs from it. It also has first class facilities to enable the investigator to re-order and interrogate the database.

Significantly "Big Floyd" also introduces a new dimension of analysis – namely the notion of template matching. Essentially, ingredients of a criminal network are superimposed on a model template for particular kinds of deduction (example "Smith is probably guilty of embezzlement"). The template is the encapsulation of an expert investigator's accumulated experience and knowledge about a particular type of offence. If the appropriate combination of linkages exists, the deduction is probably "true". This inferential system is used as a component of an Artificial Intelligence system for investigation of organized crime activities.

2.5. Telephone toll analysis

Another useful, albeit extremely simple, device is pictorial presentation of telephone toll analysis as a network. Telephone numbers are used as nodes. Connecting lines are drawn wherever a call was made from one number to another. And the directed links (directed according to who initiated the call) are assigned a weight, which corresponds to the frequency of calls during some specified time period. It is a useful way to present a summary of call activity, where a criminal organization is known to be using certain telephones. The toll analysis can give some crude clues as to the command structure, and even the social cohesiveness, of the organization being monitored.

2.6. Structural network analysis

Despite the existence, and growing awareness, of tools such as those described above, law enforcement agencies have not pushed the frontiers of *structural* network analysis very far. Some of the available

computerized link analysis packages (e.g. Anacapa Sciences' ECNA) are capable of finding connecting paths of length greater than 2 between specified entities, identifying groups and cliques, and separating large networks into their maximal connected subcomponents. But most agencies have no automated method for performing such rudimentary analyses, and remain largely unaware of the existence of such analytic capacities.

3. The needs of law enforcement

3.1. The need for strategic analysis

Traditionally law enforcement agencies, in attempting to combat the activities of sophisticated criminal organizations, have looked for some initial lead, and then have sought to exploit and develop that lead to its fullest potential. Lupsha (1980) surmised that the "lead-following" approach was not ultimately effective:

> Overall, in these [intelligence] units, there is a great deal of information collection and filing, but there is little analysis beyond the targeting and profiling of individual organized crime figures. In terms of the war against organized crime, this approach has caused some analysts to wonder if individual-oriented prosecutions merely help to open the promotion ladder within organized crime groups, moving new individuals into management positions while the group and the crime matrices they engage in continues.

Some agencies have become highly skilled at making the most of any leads they receive, frequently introducing undercover agents into an organization in order to uncover its entire workings. Some agencies quite deliberately wait, before making arrests or seizures, until they feel ready to close down the entire organization.

The problem is that such operations are difficult, dangerous, time-consuming and expensive. And many law enforcement agencies have far more leads than they have the resources to pursue. Given the fact that crime levels are not diminishing, despite countless "successes" against individual criminal enterprises, investigative agencies are discovering the need to perform strategic analysis of organized crime; that is, to try and grasp the whole picture, and to allocate investigative

resources to the principal vulnerabilities of criminal enterprises and professions.

It was precisely the need to perform *strategic* analysis of the money laundering business, rather than simply follow each available lead to its natural conclusion, that gave rise to the establishment of the U.S. Treasury Department's Financial Crimes Enforcement Network in 1990 (see Kennedy 1990; Sparrow 1990). FinCEN is an intelligence operation dedicated to the analysis of the financing of criminal enterprises whatever their primary criminal activity (drugs, racketeering, vice etc.).

With that focus FinCEN has the capacity and opportunity to ask deep structural questions about trends and practices in modern money laundering techniques. Doing so should, over the long term, facilitate more effective targeting and resource allocation as well as the design of appropriate new financial regulations and controls. FinCEN, along with a few other intelligence agencies, now knows that it needs a whole new generation of sophisticated network analysis tools in order to do its job.

3.2. Characteristics of criminal networks

Much criminal intelligence data, as mentioned above, either appears in link form or is readily convertible to it. It would be enormously gratifying, therefore, if we could simply throw the existing network analysis toolkit at criminal intelligence databases, and come away with a set of valuable new insights. Of course it is not that easy. If it were, it would surely have been done before.

The fact is that most network analysis tools have been developed within the context of retrospective social science investigations, and they are therefore designed for use on networks which are small, static, and with very few distinct types of linkages (generally only one).

It is worth considering the properties of criminal networks, and associated intelligence databases, which present significant challenges to the science of network analysis as it now stands.

Size

First and foremost, criminal intelligence databases can be huge, with many thousands of nodes. The computational ramifications are obvious – mandating the use of sparse matrix techniques or extensive exploita-

tion of parallel processing should any analytical algorithm exceed $O(n^3)$. Some network analysis algorithms do claim to be able to handle very sizeable networks. For example the NEGOPY program (Richards and Rice 1981) claims to handle up to 30,000 links or so. But it contains an unfortunate reliance on a one-dimensional interim stage in the analysis of groups and cliques, which will inevitably render its results suspect when applied to networks of any complexity.

Incompleteness

Criminal network data is also inevitably incomplete; i.e. some existant links or nodes will be unobserved or unrecorded. Little research has been done on the effects of incomplete information on apparent structure. There is some work on the problems of statistical inference from incomplete graphs (Frank 1978), researched using random link samplings from known networks; also on the relationship between network density and structural properties (Friedkin 1981). But the relevance of such work to criminal networks is largely negated by the fact that the incompleteness in the criminal databases will be anything but random – it will be systematic, at least in part, in accordance with the biases introduced by investigative methods and assumptions. The focus of existing intelligence data is determined more by the prior subjective judgments of investigators than by objective reality.

Fuzzy boundaries

The boundaries of any particular criminal web are quite ambiguous. Even organized crime families are often interrelated. And many significant crime figures are significant precisely because they are connected to a number of different criminal organizations. So there is no obvious criterion by which players can be excluded or included in any one network analysis (in stark contrast to the scenarios of Sampson's monastery data or the Bank Wiring Group). Of course criminal networks, like any other, can be split unambiguously into maximal connected subcomponents, but these may still be extensive.

Dynamic

Criminal networks are, for all practical purposes, dynamic, not static. Each contact report, telephone call, or financial transaction has a time and date. The relationship between any two individuals is not merely present or absent (binary), nor is it simply weaker or stronger (ascribed

a static analogue weighting); rather it *has a distribution over time*, waxing and waning from one period to another. Many of the most useful network questions depend heavily on this temporal dimension, begging information about which associations are becoming stronger, or weaker, or extinct.

The problematic absence of research on dynamic networks was echoed by Barnes and Harary (1983). A little work has been done on the evolution of network connections over time in dynamic networks (e.g. Hammer 1979/80), and a little on structural change within networks (Doreian 1980), but little or nothing has been done to develop algorithms for revealing significant network changes over time in the context of networks where each link has a time-dimension coordinate.

3.3. Effects of these properties

It could be argued that these properties are in fact quite typical of real-life networks, and that the discipline of network analysis has not as yet faced up to these broader and more general challenges. On the other hand these properties produce computational nightmares, demand algorithmic complexity, and require substantial advances in methods of statistical inference.

These properties certainly render some existing network theory concepts less useful than others. For example, the fuzzy boundaries render precise global network measures (such as radius, diameter, even density) almost meaningless. With the global measures go some, but not all, measures of centrality.

The remainder of this paper is devoted to an examination of some of the existing concepts of network theory. The intention is to provide an initial assessment of their relevance to, and promise for, criminal intelligence analysis.

4. Relevant concepts from social network analysis

4.1. Centrality

Some notion of centrality is clearly relevant. In seeking to incapacitate criminal organizations one obvious approach is to identify those players who are somehow central, vital, key, or pivotal, and target them for

removal or surveillance. The network centrality, or otherwise, of individuals arrested will determine the extent to which their arrest impedes continued operation of the criminal activity. Thus centrality is an important ingredient (but by no means the only one) in considering the identification of network vulnerabilities.

The network analysis literature contains many different notions of centrality, however. Six of them seem reasonably distinct, and each will be examined briefly. The first three of these six were subject of a "Conceptual Clarification..." by Linton Freeman (Freeman 1979).

(1) Degree. The "degree" of any node of the network is defined as the number of other nodes to which it is directly linked. In the case of directed networks (where links have direction and may be asymmetric) the degree is usually defined as the number of paths coming *from* a node.

(2) Betweenness. The "betweenness" of a node is defined as the number of geodesics (shortest paths between two other nodes) which pass through it. It is a measure of how important any one node might be to effective communication within, or operation of, the network. Removing a node of high "betweenness" will, by definition, lengthen the paths connecting several other nodes, rendering communication or transactions between them less efficient. Precise measures of "betweenness" permit the counting of fractional geodesics in cases where there is a "tie" for shortest path. Also measures of betweenness in non-symmetric networks have been proposed (Gould 1987).

(3) Closeness. The concept of "closeness" picks as central to a network the node which minimizes the maximum of the minimal path-lengths to other nodes in the network. That is, the central node becomes the node of minimum radius, where the radius of a node is defined as the longest of its shortest connecting paths to other nodes.

(4) Euclidean centrality after multidimensional scaling. This idea is seldom made explicit, but is implicit in a great number of presentations. Any plot (two-or three-dimensional) of the results of multidimensional scaling makes this kind of centrality quite apparent. It is analogous-to the "center of gravity" of a network. (See Kruskal and Wish 1978 for a general introduction to MDS.)

(5) Point strength. A node's "point strength" is defined as the increase in the number of maximal connected network subcomponents upon removal of that node. So it is a measure of how much network

fragmentation would be caused by removal of that node. Algorithms for computing point strength have been created (Capobianco and Molluzzo 1979/80).

(6) Business. Finally, there is the notion of the "business" of a node – which is a measure of the local information content when the network is seen as a communications network (Stephenson and Zelen 1989). To obtain some precise numerical scale upon which to measure "business", one can imagine all nodes firing (transmitting) along each of their links once per unit time. Choose some retransmission ratio (between zero and one), whereby every received transmission is retransmitted one period later but with some loss of intensity, by each node. Keep the system firing repeatedly until the total information content of each node and each link reaches equilibrium. This will occur asymptotically and monotonically both for directed and undirected networks. Then measure each node's total transmission intensity per unit time. The equilibrium transmission intensities represent useful relative, but not absolute, indicators of "how busy" each node might be.

4.1.1. Applications of centrality

So, which of these six concepts are most relevant to intelligence analysis? With respect to targeting, it would appear that the second and the sixth (Betweenness and Business) would be useful measures of significance within communication networks. To apply them to large networks, however, would necessitate the addition of some severe distance limiting effects in order to avoid imponderable computational problems.

The third and the fourth (Closeness and Euclidean Centrality) become quite arbitrary if the network has arbitrary or fuzzy boundaries. But, in fact, Euclidean Centrality is probably closest to the reality of the Anacapa chart – where centrality on the chart equates with Euclidean centrality after a manual version of two-dimensional scaling – even though the practical determination of the starting (central) node was initially by its Degree.

The fifth idea, Point Strength, seems particularly important if an agency's objective is fragmentation of a criminal network. But it seems insufficiently general. The Point Strength of a node measures its fragmentation effect when regarded as a cutset of size one. But it is quite practical, and probably useful, to consider larger cutsets. So we should extend the concept of point strength to what could clumsily be

called "Set Strength", being the increase in the number of disconnected components resulting from removal of a set of nodes. The notion of non-trivial cutsets is familiar to graph theorists as the subject of Menger's Theorem (Wilson 1972; Seidman and Foster 1978).

Finding minimal cutsets, or just small cutsets, that effectively sever communications channels or supply lines is a versatile and useful strategy, whether an agency is concerned to halt drug supply from one place to another or to prevent a terrorist organization from acquiring explosives. It is useful both for general network fragmentation objectives, as well as for targeted or specific disconnection objectives. It is also highly relevant to the selection of targets for communications interception, as communications between one group and another must, by definition of a cutset, pass through any cutset that would disconnect them.

In fact the practical task facing many law enforcement agencies, in seeking to rupture criminal supply operations, is to identify not just a manageable cutset, but manageable cutsets *within that agency's jurisdiction*.

Application of these various measures to asymmetric networks may have some relevance too. In drug supply networks drugs essentially flow one way and money flows the other, but the two commodities do not necessarily pass through symmetric channels. Strangling either one of those two flows is enough to put a supply operation out of business. It is therefore better to view the network as the overlay of two directed networks, even in those parts where it appears to be symmetric.

On balance it appears that the second, fifth and sixth notions of centrality (Betweenness, Point Strength, and Business) have greater relevance to the identification of network vulnerabilities than the others (Degree, Closeness, and Euclidean Centrality).

4.2. Equivalence

The disruptive effectiveness of removing one individual or a set of individuals from a network depends not only on their centrality, but also upon some notion of their uniqueness. The more unique, or unusual, their role the harder they will be to replace. The most valuable targets will be both central and difficult to replace.

Once again the network analysis literature offers a variety of concepts of equivalence. There are three in particular that are quite distinct

and which, between them, seem to capture most of the important ideas that have been expressed on the subject. There is, in the literature, an abundance of variations on these three main themes. There has also been an abundance of confusion over the last decade between these ideas, and some rather pointless debate as to which of them is most important. Like all models, they are each important in appropriate contexts and less so in others. In the context of criminal intelligence analysis they turn out to have very different applications.

The three concepts selected for examination here are "Substitutability", "Stochastic Equivalence", and "Role Equivalence". All three types of equivalence are mathematical equivalence relations (being symmetric, transitive and reflexive) and therefore produce exhaustive partitions of any network.

Substitutability. This is the simplest notion of equivalence, and goes under a variety of names, including "interchangeability" and (somewhat misleadingly) "structural equivalence". A clear definition is given by Lorrain and White (1971):

> Objects a, b of a category C are structurally equivalent if, for any morphism M and any object x of C, aMx if and only if bMx, and xMa if and only if xMb. In other words, a is structurally equivalent to b if a relates to every object x of C in exactly the same ways as b does. From the point of view of the logic of the structure, then, a and b are absolutely equivalent, they are substitutable.

For networks, this definition means two nodes are substitutable, or interchangeable, if they are linked to precisely the same set of nodes.

It is important to note that blockmodelling is discerning the "substitutability" of nodes as opposed to any other of the more sophisticated forms of equivalence. (See Breiger *et al.* (1975) regarding the CONCOR algorithm, a hierarchical clustering approach to blockmodelling. Panning (1982) develops goodness of fit measures for blockmodels. See Heil and White (1976) re BLOCKER, and Everett (1982) re EBLOC, a fast blocking algorithm based on a slightly extended definition of structural equivalence.)

Borgatti and Everett (1989) discuss the complete class of automorphic equivalences, and mention that substitutability (which they call structural equivalence) implies role equivalence – which it does. The

converse is obviously not true. Thus substitutability is a (mathematically) stronger condition than role equivalence.

4.2.1. Stochastic equivalence

This is a slightly more sophisticated idea. A clear definition is given in Wasserman and Anderson (1987), definition 4:

> Given a stochastic multigraph X, actors i and j are stochastically equivalent if and only if the probability of any event concerning X is unchanged by an interchanging of actors i and j.

In other words, two network nodes are stochastically equivalent if the probabilities of them being linked to any other particular individual are the same. This notion is closer to an intuitive notion of structural equivalence than simple substitutability. Street-level drug suppliers, working for one particular distribution organization, could be seen as stochastically equivalent if they, as a group, all knew roughly 70 percent of the group, did not mix with street-level dealers from any other organizations, all received their supplies from one person, and all delivered their cash to just one, randomly selected, of three collectors.

This idea of equivalence is importantly *not* the same as substitutability, as it allows for the probabilities in the definition to be something other than zero or one (as exemplified in the hypothetical situation above). Most blockmodelling techniques could only find the equivalence classes under stochastic equivalence if all the underlying probabilities were either very low or very high (because those are the probabilities that will produce discernible zero-blocks and one-blocks in a reordered adjacency matrix).

4.2.2. Role equivalence

In many ways this is the most intuitive idea of equivalence, as it allows two individuals to be counted equivalent if they play the same role in different organizations, even if they have no common acquaintances at all. It has been termed "Regular Equivalence" by some. We should probably have a precise mathematical definition:

> In a network X, a and b are role equivalent if there exists an automorphism f of X which maps a onto b and b onto a, and which is link-preserving. That is, $f(a) = b$ and $f(b) = a$; also $f(c)$ is linked to $f(d)$ if and only if c was linked to d.

The important way in which this differs from substitutability is that it permits permutation of the other nodes of the network. In other words node a can be mapped onto node *b* provided you map *a*'s organization onto *b*'s organization at the same time.

Role equivalence is also intuitively quite different from stochastic equivalence. Individuals in entirely separate organizations, who are thus in disconnected subcomponents of the overall network, can still be role equivalent. They could not possibly be regarded as stochastically equivalent as each has zero probability of knowing anyone in the other's organization.

Despite the intuitive appeal of role equivalence it was not much discussed in the literature until recently. Sailer (1978) described the notion, but the confusion between role equivalence and the other forms of equivalence persisted long after. Role equivalence was mentioned by Doreian (1988) as "regular equivalence" although he did not dwell on its significance. The concept was then spelt out in great detail by Faust (1988).

Some algorithms for finding role similarities have been suggested. Everett and Borgatti (1988) present a method based on determination of a node's orbits by examining the size of successively higher order neighborhoods. Also Burt (1990) describes an approach developed by Hummell and Sodeur for detecting role equivalence by counting and classifying types of triadic relations. This method unfortunately only takes close neighborhoods into account, not any deeper structure.

4.2.3. Applications of equivalence

The concept of substitutability has some ramifications for the assessment of network vulnerabilities. Whether or not a target individual has a substitute has an obvious and direct bearing on the extent to which his or her removal will damage the operation of the network. If another individual exists, who can take over the same role, already having the same connections, then the target individual was not well chosen. To damage the network (assuming the absence of individual capacity constraints) an agency would need to remove or incapacitate not only the target individual, but all other substitutable individuals as well. Individuals who have no available network substitutes would make more worthwhile targets.

The concept of substitutability also has relevance to detecting the use of aliases. The use of an alias by a criminal might also show up in a

network analysis as the presence of two or more substitutable individuals. This is particularly likely if the analysis is performed on aggregated link data, drawn from two or more agencies or investigations. It is conceivable that the same individual could be known to different agencies by different names. In which case, the merged data would show two or more nodes for the same person. But, provided different modes of agency operation did not unduly bias the types of contacts or transactions they were likely to witness, the immediate network neighborhoods of those nodes would be similar or identical. The interchangeability of the nodes would reveal the interchangeability of the names.

There is a simple computational method of discovering such aliases within a network, should they exist. Two alias nodes would have no link joining them directly, but would have a significant number of paths of length two connecting them, one for each member of their immediate neighborhood. Existence of many paths of length two without a direct connection is, otherwise, a most unlikely phenomenon.

The concept of role equivalence is clearly applicable when considering the roles that individuals play within different criminal structures. In some ways the FBI's use of template matching can be regarded as a particular form of a search for role equivalence. The distinguishing characteristic of the template matching approach being the comparison of network individuals with a hypothetical, idealized individual (or template) rather than with another existing network node. The hypothetical individual is constructed by an investigator expert in that particular type of crime, or role.

The same concept might also be useful in performing strategic analysis of various criminal trades. Agencies might choose to focus investigative efforts on some particular, and essential, role in any criminal activity. Identification of vulnerabilities special to one role might lead to a shortage of people able to offer those services to criminal organizations. For instance, targeting courier-recruiters could help reduce cross-border currency or drug traffic, just as targeting drivers could stall armed robbery gangs. Any kind of role uniqueness represents a strategic vulnerability within a criminal profession, not least because insertion of undercover agents within criminal organizations is normally role-specific.

It is illuminating to compare the template-matching methods for role detection with the traditional approaches to blockmodelling, and to

consider how each would cope with an individual who played more than one role. The issue revolves around the treatment of zeroes in the adjacency matrix, or sociogram. Blockmodelling techniques usually require two nodes not only to have a similar pattern of connections before they are deemed equivalent, but also a similar pattern of disconnections (i.e. absent links). As soon as the possibility of multiple roles is acknowledged, then the requirement to match the zeroes as well has to be dropped. Attention has to be paid only to finding the *presence* of a designated set of connections, no matter what other extraneous links are observed.

This raises the possibility of a further field of enquiry. Suppose there were a number of designated roles within a network, and a template of connections had been prepared for each role. Then a useful question might be "which set of roles best explains this individual's aggregate network connections?". The task would then be to find not just the best-fitting template, but the best-fitting set of templates.

4.3. Weak ties

The notion of weak ties (Granovetter 1973) may be of particular interest in finding the vulnerabilities of criminal communication networks. The "cell" structure of the Irish Republican Army fits Granovetter's model exceptionally well. IRA terrorists work together in small, well-established teams (cliques), which makes the organization particularly difficult to infiltrate. Command and control communications directing the operations of individual "cells" use channels that, within the organizational context, look exactly like Granovetter's weak ties. The most valuable communications channels to monitor, therefore, are those which are seldom used and which lie outside the relatively dense clique structures.

It is reasonable to assume more generally that weak ties are the ties which add most to the efficiency of communication within a network. They will be disproportionately represented within the network's geodesics, precisely because of their network spanning properties. Urgent or important network signals are therefore more likely to be detected on the weak ties than on the stronger ones.

Disabling communication channels which are weak ties is also likely to have the greatest effect on the completeness of network transmission, as well as upon its speed.

Note that intelligence analysts have traditionally used the terminology of "strong" and "weak" links in a very different sense – to indicate the reliability of the information rather than the links' structural importance (Harper and Harris 1975). A "strong" link has been, for analysts, one which has been "confirmed" by a second independent source.

5. Summary

Existing concepts from the discipline of network analysis have been shown to be relevant to the analysis of criminal intelligence. These include several different notions of centrality and of equivalence, and the concept of weak ties. There are many other network analysis concepts which might turn out to be useful also. But the object of this paper was not to be exhaustive; merely exploratory and introductory.

The law enforcement community, being largely unaware of the methods and concepts developed within the discipline of network analysis, has not yet had the opportunity to enunciate its needs for more sophisticated tools.

There are consequently two audiences for this paper; and two aims. First, for the network theorists: that it would act as a belated invitation to consider intelligence analysis as a most interesting application for their skills.

Second, for the law enforcement agencies in general and intelligence analysts in particular: that it would familiarize them with the fundamentals of network theory sufficiently to enable them to imagine what is possible, and to enable them to begin the process of specifying exactly what they need by way of additional analytic tools.

References

Barnes, J.A. and F. Harary
1983 "Graph-theory in network analysis." *Social Networks 5* (2): 235–244.
Bayse, W.A. and C.G. Morris
1987 "FBI automation strategy: Development of Al applications for national investigative programs". *Signal Magazine* May.
Borgatti, S.P. and M.G. Everett
1989 "The class of all regular equivalences: Algebraic structure and computation". *Social Networks 11*(1): 65–90.

Breiger, R.L., S.A. Boorman and P. Arabie
1975 "Algorithm for clustering relational data with applications to social network analysis and comparison with multidimensional-scaling". *Journal of Mathematical Psychology 2*(3): 328–383.
Burt, R.S.
1990 "Detecting role equivalence". *Social Networks 12*: 83–97.
Capobianco, M.F. and J.C. Molluzzo
1979/80 "The strength of a graph and its application to organizational structure". *Social Networks 2*: 275–284.
Coady, W.F.
1985 "Automated link analysis – artificial intelligence-based tool for investigators". *Police Chief 52*(9): 22–23.
Davis R.H.
1981 "Social network analysis – an aid in conspiracy investigations". *FBI Law Enforcement Bulletin 50*(12): 11–19.
Doreian, P.
1980 "On the evolution of group and network structure". *Social Networks 2*: 235–252.
1988 "Borgatti toppings on Doreian splits: Reflections on regular equivalence". *Social Networks 10*(3): 273–287.
Everett, M.G.
1982 "A graph theoretic blocking procedure for social networks". *Social Networks 4*(2): 147–168.
Everett, M.G. and S. Borgatti
1988 "Calculating role similarities: An algorithm that helps determine the orbits of a graph". *Social Networks 10*: 77–91.
Faust, K.
1988 "Comparison of methods for positional analysis: Structural and general equivalences". *Social Networks 10*: 313–341.
Frank, Ove.
1978 "Sampling and estimation in large social networks". *Social Networks 1*: 91–101.
Freeman, L.C.
1979 "Centrality in social networks: Conceptual clarification". *Social Networks 1*: 215–240.
Friedkin, N.
1981 "The development of structure in random networks: An analysis of the effects of increasing network density on five measures of structure". *Social Networks 3*: 41–52.
Gould, Roger V.
1987 "Measures of betweenness in non-symmetric networks". *Social Networks 9*: 277–282.
Granovetter, Mark S.
1973 "The strength of weak ties". *American Journal of Sociology 8*: 1360–1380.
Hammer, M.
1979/80 "Predictability of social connections over time". *Social Networks 2*: 165–180.
Harper, W.R. and D.H. Harris
1975 "The application of link analysis to police intelligence". *Human Factors 17*(2): 158.
Heil, G.H. and H.C. White
1976 "An algorithm for finding simultaneous homomorphic correspondences between graphs and their image graphs". *Behavioural Science 21*: 26–45.
Howlett, J.B.
1980 "Analytical investigative techniques: Tools for complex criminal investigations". *Police Chief 47*(12): 42–45.

Kennedy, D.M.
1990 "On the kindness of strangers: The origins and early days of FinCEN". Teaching Case no. C16-90-1000.0 John F. Kennedy School of Government, Harvard University, Cambridge, Mass.
Klovdahl, A.S.
1981 "A note on images of networks". *Social Networks 3*: 197–214.
Kruskal, Joseph B. and Myron Wish
1978 *Multidimensional Scaling*. Beverley Hills, CA: Sage.
Lorrain, F.P. and H.C. White
1971 "Structural equivalence of individuals in networks". *Journal of Mathematical Sociology 1*: 49–80.
Lupsha, P.A.
1980 "Steps toward a strategic analysis of organized crime". *Police Chief*. International Association of Chiefs of Police, Gaithersburg, MD.
Panning, W.H.
1982 "Fitting blockmodels to data". *Social Networks 4*(1): 81–101.
Richards, W.D. and R.E. Rice
1981 "The NEGOPY network analysis program". *Social Networks 3*(3): 215–223.
Sailer, L.D.
1978 "Structural equivalence: Meaning and definition, computation and application". *Social Networks 1*(1): 73–90.
Seidman, S.B. and B.L. Foster
1978 "A note on the potential for genuine cross-fertilization between anthropology and mathematics". *Social Networks 1*: 65–72.
Sparrow, M.K.
1990 "An evaluation of the potential of the U.S. Department of the Treasury's financial crimes enforcement network". Report prepared for U.S. Treasury and Congress.
Stephenson, K. and M. Zelen
1989 "Rethinking centrality: Methods and examples". *Social Networks 11*(1): 1–38.
Wasserman, S. and C. Anderson
1987 "Stochastic a posteriori blockmodels: Construction and assessment". *Social Networks 9*: 1–36.
Wilson, R.J.
1972 *Introduction to Graph Theory*. London: Longman.
1986 "An algorithm for finding patterns of racketeering through subgraph embedding". Informal Technical Notes, Institute for Defence Analysis. 20 August.