Application of Social Network Analysis (SNA) to Terrorist Networks in Jammu & Kashmir

Sudhir Saxena, K. Santhanam, Aparna Basu

Abstract

The paper presents initial results from a limited exercise to apply Social Network Analysis (SNA) methodology to the database on terrorism created in IDSA called Terrorism Tracker (or T2).

T2 has been developed in-house. It uses systematic search for information on terrorist events from open sources including cyberspace, newswires and print reportage. This is followed by application of filters. It interfaces with standard commercial packages for data extraction from cyberspace as well as text mining from both structured and un-structured data. Classification of data into sub-groups for organisation-to-organisation, person-to-organisation and person-to-person network links has been automated. The present paper addresses organisation-to-organisation links of terrorist organisations operating in the Indian State of Jammu & Kashmir, through SNA.

The SNA software package, Visone, developed in Germany, has been used with the T2 generation of "co-occurrence" pairs where organisations are cited together in an event during the period 2000 – 2003. This output was converted into an adjacency matrix to form the input to Visone for analysis and generation of linkage graphs. Among the number of network parameters in the Visone suite, we have studied Degree which is defined as the number of direct connections that a node in the network has with other nodes.

We are of the view that the analysis performed by SNA is in line with our understanding of the phenomenon of terrorism in Jammu and Kashmir. These techniques, additionally, offer effective graphic visualisation of the network. In some cases, e.g., Umma Tameer-e-Nau (a Pak NGO), the Visone output has revealed linkages which have existed in T2 but were not easily obvious in poring over the large volume of data.

These initial findings appear to indicate that SNA or more broadly Graph Theoretic methods could be utilised in understanding terrorist networks.

Strategic Analysis, Vol. 28, No.1, Jan-Mar 2004

They appear, also, to provide a theoretical framework for understanding such networks. We believe that SNA application to terrorism is likely to be of utility to strategic analysts and information scientists; and, possibly, to agencies dealing with national security.

— * —

Introduction

The increase in violent terrorist incidents, in recent years, has heightened the feeling of insecurity in many countries. In this regard, privileged information with government agencies is not available to the public or academic research institutes. However, the IT revolution has opened up a growing volume of information on these incidents. This information in the open domain can be a valuable resource in creating a data base which, after proper capture, cleaning and classification could be useful in strategic analysis. IDSA's "Terrorism Tracker" (abbreviated as T2) system enables automated terrorism-related data capture from cyberspace, newswires and print reportage followed by filtering for creation of a database for using network matrix generation and analysis with a GIS engine.

Terrorist and criminal organisations, generally, organise themselves as secret networks with distributed work share. It is not easy to obtain confirmed information in the open domain on their membership, ties and management. However, terrorists need the oxygen of publicity for their 'cause'; and, hence, open source information[1] becomes relevant in exploring such networks.

Textual analysis with scientometric tools (eg co-occurrence or co-presence in an event) and graph theoretic methods of Social Network Analysis (SNA) are increasingly being applied in a variety of fields[2]. These could be applied to mapping of links between terrorist organisations as well. The existence of textual links in incident reportage may be considered as indicative of their *collaboration* or *association.*

The idea of applying SNA to understand networks is not new[3]. In one form or other, network analysis has been used to uncover unlawful entities and activities. It has been used for evidence-mapping in fraud and criminal conspiracy cases[4,5,6]. A suspect's network can be built through relational information in telephone logs or credit card transactions[7]. More recently, network methods have formed a useful part of intelligence work. Bruce Berkowitz has reported on the use of commercial network analysis packages

by US agencies in understanding Saddam Hussein's large extended family and kinship network prior to his capture[8].

This paper is organised into four sections. The first section gives a brief description of the IDSA's T2 system. The second summarises salient features of the theory behind SNA techniques. The third deals with the application of and output from the SNA software, *Visone*, with data provided from the T2 database on terrorist outfits in Jammu and Kashmir. The fourth presents our analysis of the network graphs generated by this SNA tool.

About the T2 Software and Database

The Terror Tracker (T2) system has been designed and developed at IDSA[9] as an in-house endeavour to archive and analyse open source information on terrorism-related events. The system incorporates modules for automatic capture of open source information (primarily from the Internet), data/text mining, statistical analysis and spatial analysis through a GIS engine. The T2 architecture is given in Annexure-2**.**

'Events' form the core data component of T2 analysis. They constitute information/data elements on the occurrence of terrorist/terrorism-related incidents or activities. In T2 design, event information is captured from structured sources (e.g., online incident databases on the Internet) and from unstructured sources (free flow text in news reportage on the Internet and in print media). The captured raw data is sanitised through redundancy and noise removal filters. Thereafter, the system undertakes an elemental breakdown (data 'tear') followed by reconstruction (data 're-stitch') and text mining leading to data incorporation in the main structured database. Data from structured sources (e.g., Institute for Counter-Terrorism, Israel; Institute for Terrorism and Political Violence, USA) can be directly fed. This structured database is, thereafter, available for analysis.

The T2 database contains data elements like entity name (organisation/ personnel), incident date/time, incident location and incident category. The locational data is dynamically linked to a Geographic Information System (GIS) which is projectable on a vector map backed by raster layers.

Another module undertakes classification of event data into five baskets for first-order links between entities for network analysis: Organisation-to-Organisation, Organisation-to-Personnel, Organisation-to-Location, Personnel-to-Personnel and Personnel-to-Location.

The database has been used at IDSA in profiling militant organisations which have been active in J&K after 1989.[10]

For the research results presented here, the T2 database was mined with co-occurrence as the criterion for Organisation-to-Organisation network links for the years 2000, 2001, 2002 and 2003.

An Overview of Social Network Analysis (SNA)

Social Network Analysis (SNA) is about mapping and measuring of relationships (and flows) between people, groups, organisations or other information/knowledge processing entities. They are represented as nodes in the network alongwith links, with/without attribute weightages. SNA attempts to provide both mathematical analysis and visual representation of relationships in a network. SNA is an interesting cross-fertilisation of sociology and mathematics to the benefit of both.

The location of a node defines its *centrality* which is a measure of importance or prominence in the network. There are many network parameters. Four major ones are:-

- Degree — Number of *direct connections that a node has*
- Betweenness — The number of paths that *connect pairs of nodes that pass through a given node*
- Prestige — A measure of *links to other highly central nodes*
- Closeness — The number of *other nodes that are linked to a given node*

These parameters have calculated indices based on matrix algebra with direct network implications. An entity with a high Degree index means that it is very strongly networked and active. An entity with a high Betweenness index would have a strong 'brokerage' role. A centralised network with a very high Degree index in one or a few nodes can become a single point of failure. A less centralised network would be resilient in the face of collapse or failure; it would experience graceful degradation.

We have used the SNA software package, *Visone*, developed by Ulrik Brandes and Dorothea Wagner of Germany[11] in our study. It works on social networks with graph-theoretic analysis and explores visual network presentation. It is designed to meet two criteria:

- Correct acceptance and mathematical treatment of information manifest in data about a network data; and
- Correct visualisation of the network.

It accepts network information presented through an adjacency matrix generated by T2. *Visone* is also a test-bed for algorithm research.

We have given a list of militant organisations covered in the paper along with abbreviations in Annexure-1 to facilitate understanding of the analysis and the network graphs.

Visone Application to the T2 Database

Unlike in scientific literature, for example, where collaboration is explicitly given in author lists, the 'association' (or nexus) between terrorist outfits has to be inferred differently. We have taken the *co-occurrence* of different organisations in data pertaining to a terrorist incident as indicative of a link between them.

As mentioned earlier, the data set from the T2 database has provided such co-occurrence links in terrorist incidents in Jammu and Kashmir through the Organisation-to-Organisation link classification of events during the 4-year period 2000-2003. These are summarised and presented in Annexure-3 which provides a bar chart version of this dataset. We have chosen to focus on the more active organisations for the purpose of clearly identifying the centers of gravity. Further, data clutter is reduced without loss of context in understanding events of importance in a given year. An *adjacency matrix* was generated from the co-occurrence dataset as input to *Visone*.

A visualisation of these links as multi-nodal *Visone* output graphs for the years 2000, 2001, 2002 and 2003 is given in Annexure-4. These 'circular' graphs indicate *tie strength* of links through line thickness (thick, thin) for a given node. This is directly related to the number of co-occurrences. The *radius of a node* in the network, on the other hand, is based on Degree which is defined as the total number of links to other nodes which pass through the given node.

Analysis of the SNA Graphs

Being engaged in profiling of terrorist activities with traditional methods, the SNA where have to be first understood as visual depictions of trends

from open source data in the T2 database. We set ourselves 3 questions in understanding these graphs:

- Are the results from SNA and their graphic depiction in agreement with insights available through classical organisation profiling?
- Can SNA assist or enhance traditional analysis and understanding of terrorist networks?
- Can SNA help in reducing the time and effort needed to identify new links or predict them?

The answers to these questions, from our exercise, is: Yes. Details behind this affirmative finding are given below.

Correctness of Analysis and Graphic Depiction

Regarding the first question, the first observation is that the results from graph-theory calculations and their presentation through network graphs are in agreement with our understanding of the nodes and network through traditional methods — within the limits of accuracy set by the nature of open source information. The SNA methodology does not have *a priori* knowledge of the terrorism domain in Jammu and Kashmir and is not an 'expert system' in the way such systems are used in specific domains. It has operated on the co-occurrence dataset from the T2 database (and the resulting adjacency matrix) alone.

Under these circumstances, we find this agreement satisfactory. Additionally, graphic visualisation helps to create a mental picture of the network as a whole for a quick grasp of network links.

We have the following additional observations on SNA output, year-wise:-

2000

A total of 7 organisations are depicted, a sort of lean year — in hindsight.

HuM remains the most active organisation followed by JeM and Al Badr. The tie-strength between KJHC and HuM is higher than that between KJHC and JeM. KJHC is the most networked organisation.

A word of explanation seems necessary at this point. KJHC is a conglomerate of 26 'political' organisations which is in a dialogue with the

Indian government on the Kashmir issue. However, many KJHC members have linkages with militant outfits.[12] KJHC is at the centre of coordination activity in the graph.

There is strong co-occurrence between KJHC and HuM which has been a very active militant outfit in Jammu and Kashmir. But *HuM co occurs with KJHC alone*. Further, JeM co occurs with HuJI which is an 'Afghan' outfit created in Pakistan to fight the Soviet presence in Afghanistan; it now operates in Jammu and Kashmir after the Soviets left Kabul in 1988-89.

2001

26 organisations are depicted, a significant rise over the co-occurrence number in 2000.

HuM is the most active outfit followed by Al Badr, JuM, JeM and HuJI. HuM is seen to be strongly linked with HuJI, UTN and SIMI. Except for the link with SIMI, the others are not surprising since HuM operates out of Muzaffarabad in Pakistan-occupied Kashmir (PoK). The high co-occurrence count of HuM in open sources may, additionally, be due to its involvement in peace initiatives with the Indian government during the period. Strong links between HuM and HuJI are depicted in the graph.

Al Badr, JuM and similar small outfits have less co-occurrence counts and the SNA graphs have positioned them rightly. JeM has shown an increase in its links with Al Qaida and international Islamic fundamentalist networks compared to year 2000.

Taliban and Al Qaida show a low but steady presence. Al Qaida is seen as co-occurring with JuM and JKIF.

UTN appears as a surprise entrant in the graph. It came into prominence after 9/11 primarily due to the interrogation/questioning of its founder, Basheeruddin Choudhary and his associates over suspected links of UTN with Al Qaida for transfer of WMD technology. SNA application to the T2 co-occurrence dataset reveals UTN's links with HuM, KJHC, JuM, HuJI, Taliban and ISI. Its links with Taliban, and JeI (Pakistan) and ISI are well known since Basheeruddin Choudhary, a senior engineer with the Pakistan Atomic Energy Commission, has deep fundamentalist inclinations. UTN's co-occurrence with militant outfits operating in Jammu and Kashmir was missed in traditional methods of pattern analysis. The fact that SNA has

thrown up the network links of UTN in this regard is a pointer to the usefulness of SNA while pointing to inadequacies in orthodox approaches to network analysis involving a flood of data.

SIMI, a pan-India organisation, co-occurs with Taliban, JeM, JKLF and UTN.

Sikh militant organisations pop up, surprisingly, in the graphs. It is known that links between them and terrorist outfits existed in the past in the context of the Punjab militancy in India**.** However, co-occurrence of the ISYF in the graph may indicate continuing residue of linkages.

Finally, the appearance of LTTE with ISI and HuM invites a comment. LTTE may not have common cause with ISI or HuM since it has its own, primary, struggle in Sri Lanka. Regardless, LTTE's appearance in the graph would indicate that a non-ideological but commercial role-like facilitation of arms supplies to militant organisations in Jammu and Kashmir is likely.

The complexity of the graph, when compared to that of 2000, hints also at the possible existence of other direct/indirect linkages between most militant organisations operating in Jammu and Kashmir. In this regard, the application of SNA to open source information may, potentially, assist security agencies with access to other privileged information.

2002

24 organisations have appeared in the very active list.

HuM remains the most co-occurring organisation followed by JeM, JuM and KJHC. The trend is essentially similar to that observed in the previous years.

Various coordinating organisations such as Muttahida Jihad Council appear in the lead list indicating the importance of their role in the conduct of militancy operations in J&K. These organisations are prominently linked with outfits like JeM, Al Qaida and JKLF. So is the case with ISI which is clearly linked with Taliban and other militant outfits operating in Jammu and Kashmir.

Al Qaida and Taliban continue to show a relatively low but steady co-occurrence presence in the data.

UTN continues to appear in 2002. This may be due to continued reportage on the follow-up to the questioning of UTN personnel by the Pakistani and US intelligence agencies in 2001.

Sikh militant organisations such as ISYF and KZF continue to appear in low and steady numbers indicating continued, residual linkages in Jammu and Kashmir. Both co-occur with Muslim Jihad Council, another coordinating organisation.

2003

21 organisations are in the very active list.

HuM remains the most co-occurring organisation followed by KJHC, LeT, JKJeI, JuM, JeM, JuM and JKLF. The trend remains largely similar to that of the previous years. However, JKLF emerges as the most networked organisation.

KJHC has a low number of network links. But it has thick co-occurrence link strength with JKJeI. Some of the high co-occurrence visibility of KJHC in 2003 in open sources may also be due to its engagement in the front-end dialogue process with the Indian government and back-end affiliation of its constituent entities with militant outfits. Further, KJHC has been in the news due to a split in the organisation and the formation of two factions, one led by moderate Abbas Ansari and the other by pro-Pakistan hardliner SAS Geelani. The split has also brought JKJeI into prominence as Geelani happens to lead one of the JKJeI factions.

LeT has shown a marked improvement in its co-occurrence visibility. It has been one of the most violent militant outfits in the region. Though its visibility has been poor in terms of co-occurrence in the earlier years, statistically it has continued as the lead contributor for terrorist incidents in Jammu and & Kashmir; it is known to be operating under pseudonyms as well. Higher co-occurrence in 2003 compared to the earlier years indicates enhanced coordination/cooperation with other organisations.

Sikh militant organisations continue to be present. ISYF co-occurs with HuM, JuM and Taliban.

Taliban and Al Qaida continue to present in low but steady numbers.

D Company appears to be an interesting co-occurrence. The organisation

is led by Dawood Ibrahim, a gangster don who is the prime accused in the 1993 Bombay serial bomb blasts. He has been provided sanctuary in Karachi by the ISI. D Company co-occurs with Al Qaida and the combination of terrorism with gangsterism may well portend heady and lethal times head, regionally and internationally.

UTN is absent in the graph, probably due to more prominent developments in Iraq. Also, it is likely to have got a bit submerged in the flood of nuclear revelations after the AQ Khan sluice gates were opened abroad (and in Pakistan).

A general comment about the low co-occurrence of ISI in all the graphs. It is generally known that there is ISI backing for most militant organisations operating in Jammu and Kashmir even though its name may not figure in every event report. In the T2 main structured database, we have incorporated only those reports in which ISI is explicitly mentioned. As such, the number of ISI co-occurrences would be expected to be higher than depicted in the T2 database and in the SNA analysis presented here.

Potential Enhancement of Traditional Analysis with SNA

We would like to mention that, apart from calculations based on in graph theory, SNA software packages provide visualisation the network structure. Some amount of pre-processing is, indeed, required before SNA can be applied to network data collected for orthodox analysis. The T2 System was designed, initially, for traditional analysis. However, it has been quite easy to generate to co-occurrence data and, thereby, the adjacency matrix for *Visone*.

SNA output based on the T2 adjacency matrix appears right. But it is premature, at the present stage of our study, to say whether new links can be identified by SNA. We believe, however, that it would be possible to link T2 with SNA in an interactive manner to graphically flag dormant links that have surfaced or new nodes that have dynamically emerged. We believe that, if node attributes are integrated into SNA as part of continuing research, the depth of analysis could be increased.

Summing Up

The mathematical analysis embedded in SNA tools and the Degree centrality calculations of nodes in a network of our interest appear to be correct. These determine node size and link strength. The use of an adjacency

matrix generated from co-occurrence links in events in the T2 database as the input for SNA, hence, appears to be valid.

Secondly, the network graphs have correctly displayed the linkages between militant organisations operating in Jammu and Kashmir.

Thirdly, the network maps assist in creation of a mental picture of complex linkages. This is useful in supporting pattern analysis through traditional means.

We are of the view that SNA packages, like *Visone*, score over traditional analysis of networks with a large volume of data. SNA could reduce the consequent overload on analysts.

It has not escaped our notice that SNA, duly validated and used with real-time or near real-time information from a multiplicity of databases could have the potential to generate early warning signals of utility in detecting and deterring terrorist attacks. It is necessary, of course, to have 'experts' in the loop.

The results presented here are initial findings based on a limited exercise in exploring the utility of SNA in analysing terrorist networks. We are engaged in exploring other SNA envelopes in our continuing studies. These initial findings on the utility of SNA in the terrorism domain appear to be promising. In a way, SNA can be considered as an enlargement of the systems approach to analysis.

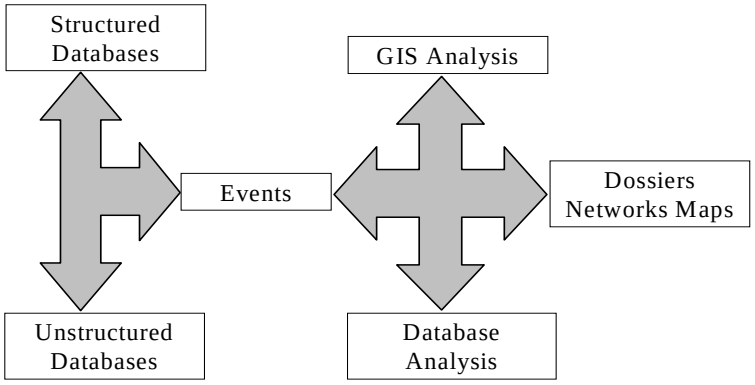
Annexure-1

List of Abbreviations

HuM	:	Hizbul Mujahideen
LeT	:	Lashkar-e-Tayyaba
KJHC	:	Kul Jamaat Hurriyat Conference or All Party Hurriyat Conference
JeM	:	Jaish-e-Mohammad
HuJI	:	Harkat-ul-Jihad Islami
JuM	:	Jamaat-ul-Mujahideen
SIMI	:	Students Islamic Movement of India
ISYF	:	International Sikh Students Federation
JKLF	:	Jammu Kashmir Liberation Front
ISI	:	Inter Services Intelligence (Pakistan)
KZF	:	Khalistan Zindabad Force
JKJeI	:	J&K Jamaat-e-Islami

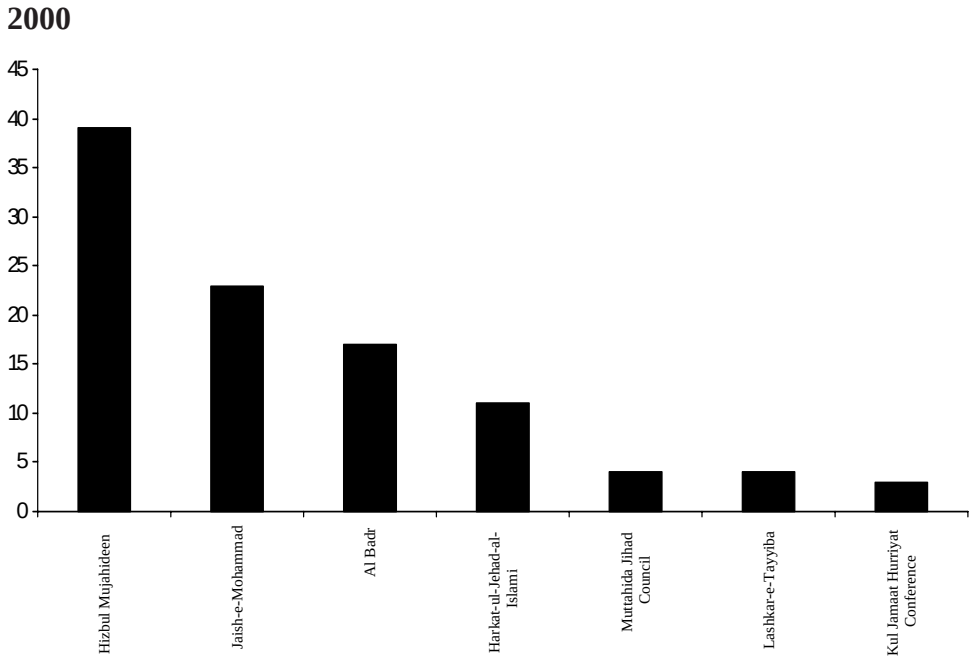
Annexure-2

Macro Structure of T2 System

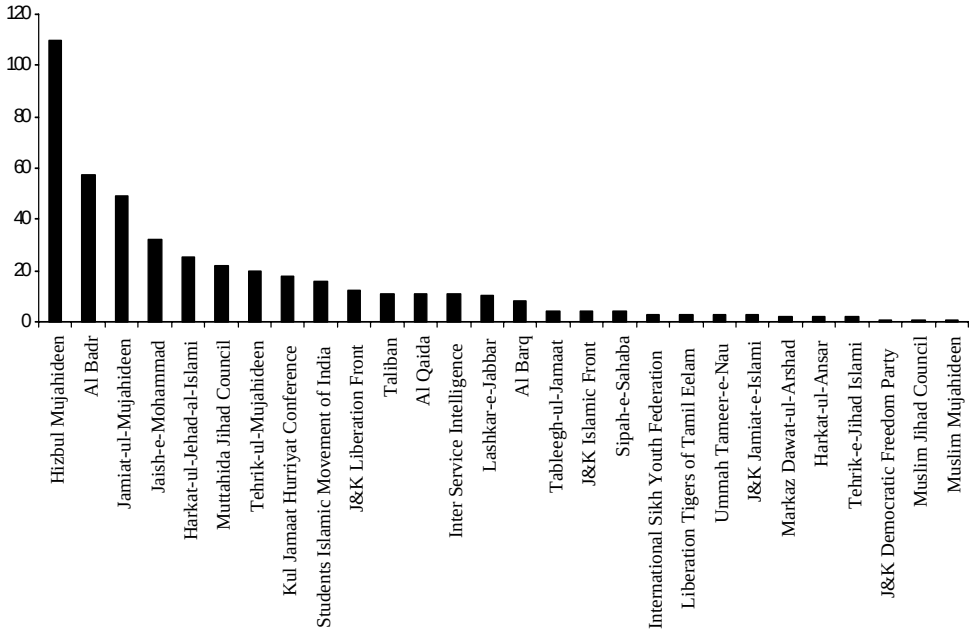


Annexure-3

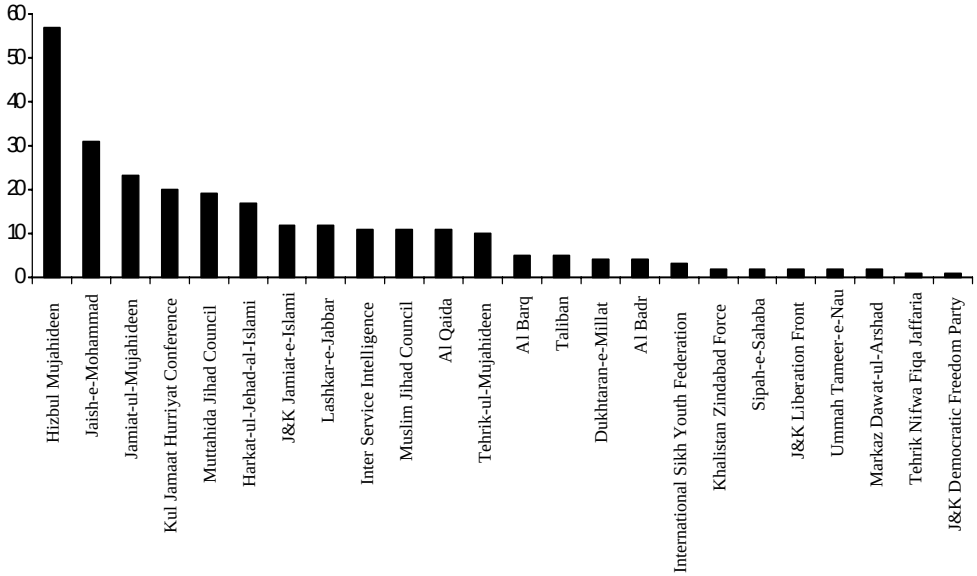
Year-Wise Bar Chart Representation of Co-Occurrence Activity of Terrorist Organisation in J&K



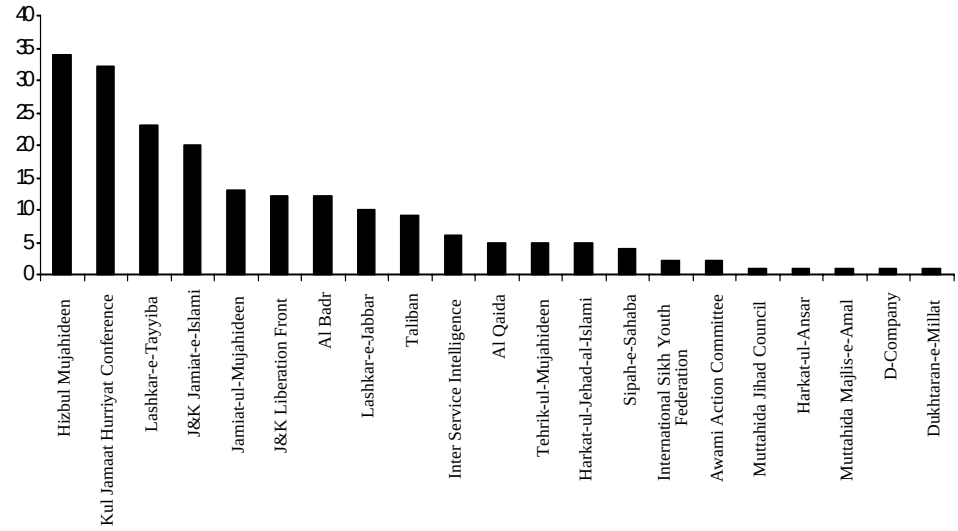
2001 **Annexure-3/**(contd)



2002 **Annexure-3/**(contd)



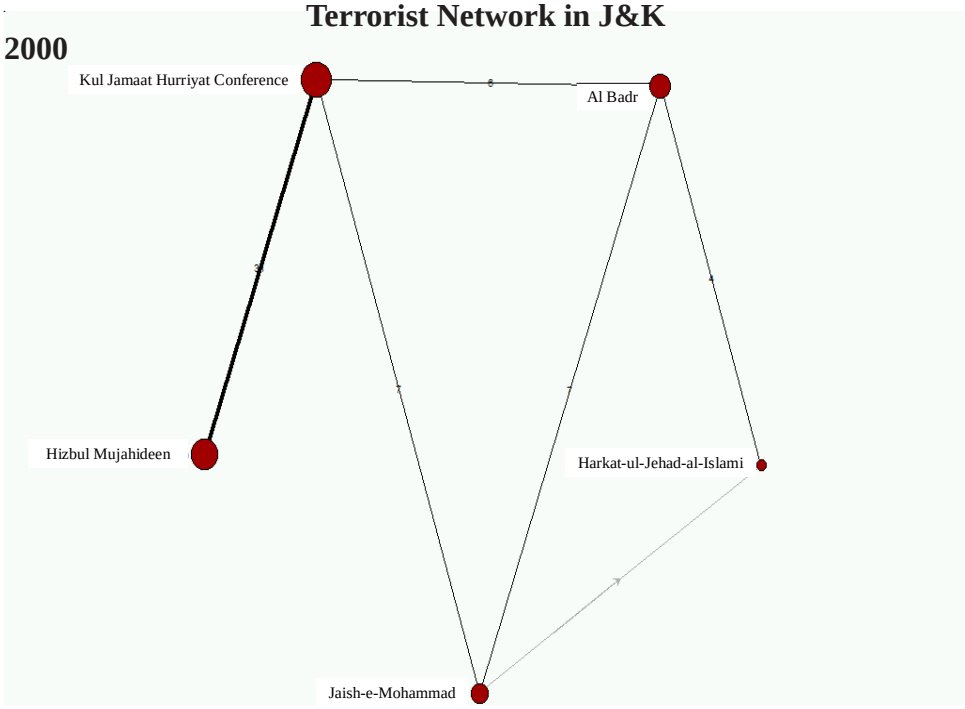
2003 **Annexure-3/**(contd)



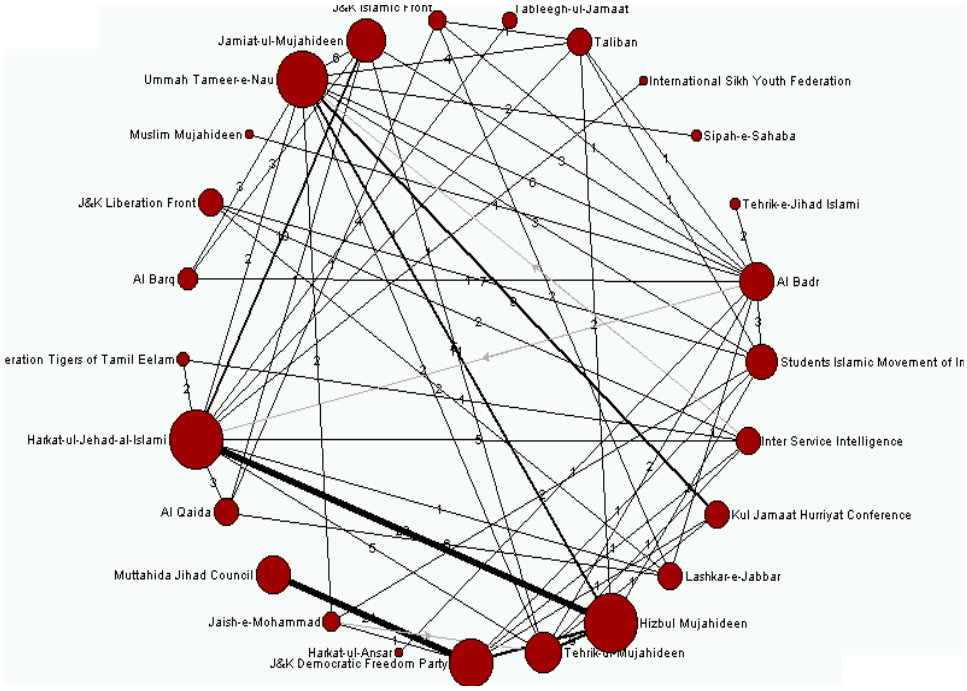
Annexure-4

Year-Wise SNA-Circular Graph Representation of the Terrorist Network in J&K

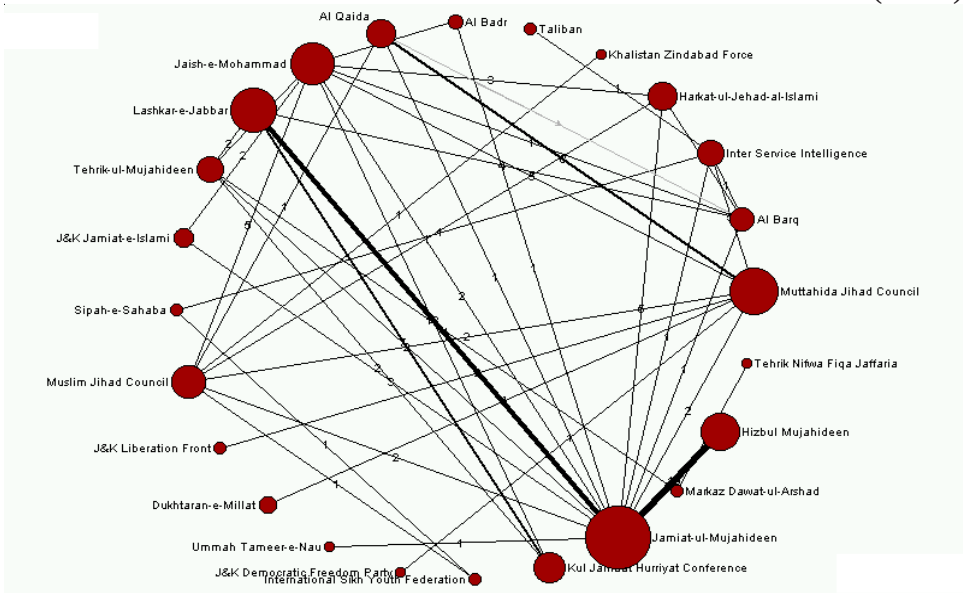
2000

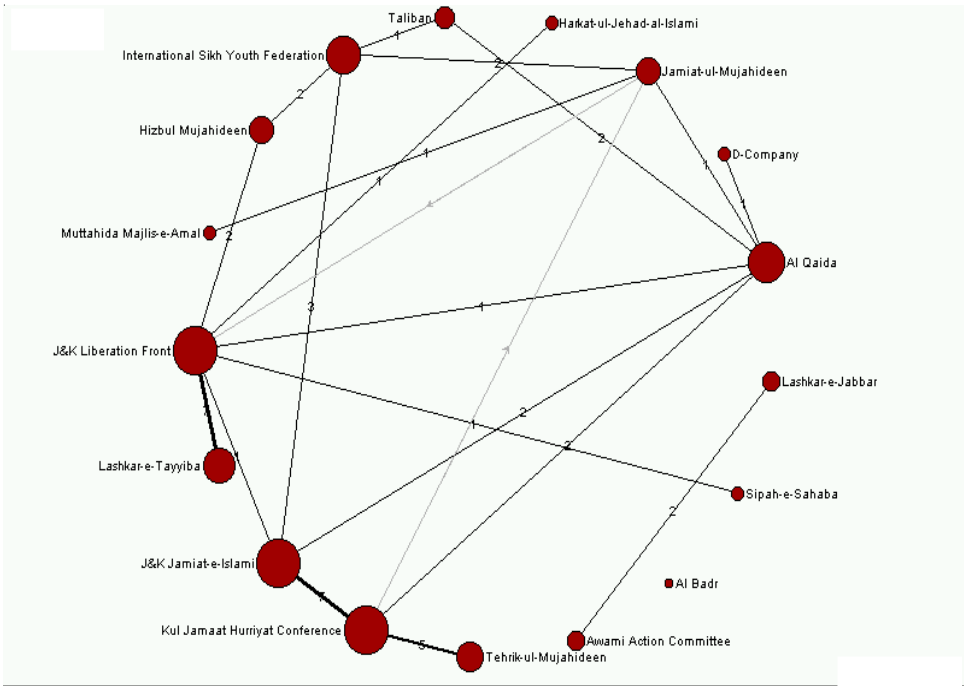


2001 **Annexure-4/**(contd)



2002 **Annexure-4/**(contd)





Acknowledgements

We would like to thank Urlick Brandes and Dorothea Wagner for permission to use the *Visone* software. Also, several serving and retired persons from Track I who saw our intermediate research outputs and felt that they are useful.

We would like to thank, again, many IDSA scholars who have contributed to the evolution of the T2 software and database.

References/ End Notes

1 Most think tanks, including IDSA, conduct research on published, attributable information. Of late, intelligence analysis based on open sources is emerging as a movement. See, for example the website of Open Source Solutions Inc, http://www.oss.net, which offers services at a price.

2 A good starting point is John Scott, Social Network Analysis: A Handbook. 1994. Sage Publications.

3 Sparrow, M.K., The application of network analysis to criminal intelligence: An assessment of the prospects, Social Networks. 1991 **13** 251-274.

4 Baker, W.E. and Faulkner, R.R., The social organization of conspiracy: Illegal networks in the heavy electrical equipment industry, American Sociological Review. 1993, **58** (6) 837-860.

5 Criminal Network Analysis Training Course. Defense Intelligence Agency. 2000.

6 Klerks, P., The network paradigm applied to criminal organizations. *Connections*. 2001, **24** (3)

7 Krebs, Valdis E., Mapping Networks of Terrorist Cells. *Connections*. 2002, **24** (3) 43-52.

8 "Berkowitz, Bruce, Learning to Break the Rules" Commentary. *New York Times*. . December 19, 2003 at http://www.rand.org/commentary/121903NYT.html .

9 Saxena , Sudhir and K. Santhanam, Design Approach to Creating a Terrorism Database with Open Source Information. IDSA Internal Report. July 2001.

10 Santhanam, K., Sreedhar, Sudhir Saxena and Manish, Jihadis in Jammu and Kashmir: A Portrait Gallery. 2003. Sage Publications; New Delhi.

11 Brandes, Ulrik, and Dorothea Wagner: Visone - Analysis and Visualization of Social Networks. *In* Michael Jünger and Petra Mutzel *Eds.*, Graph Drawing Software. 2003. Springer-Verlag pp. 321-340.

12 Saxena, Sudhir and K. Santhanam, no. 9, p 212.

Sqn Ldr Sudhir Saxena is a former Air Force officer, currently working at IDSA as Research Fellow on modeling and simulation issues.

K Santhanam is Director General, IDSA and a physicist.

Dr Aparna Basu has a PhD in plasma physics and has recognised contributions in scientometrics.She proposed the idea of using co-occurrence data in the T2 database to enter SNA.