1. **Program Name:**
   “2008 NGA University Research Initiatives (NURI)”

2. **BAA Number:**
   “HM1582-08-BAA-0003”

3. **Title:**
   Dynamic Purpose aware Graph Models for Composite Networks

4. **Research Topic:**
   TOPIC 4.2: Representation and Reasoning about Complex and Fluid Spatio-temporal Structures

5. **Principal Investigator:**
   Shashi Shekhar, Professor, University of Minnesota, Department of Computer Science and Engineering, 4-192 EE/CS Building 200 Union Street SE, Minneapolis, MN 55455,
   Phone: (612) 625-4012 Fax: (612) 625-0572, Email: shekhar@cs.umn.edu

   **Alternate contact information.**
   Angie Vail, Grants Management Coordinator, Department of Computer Science and Engineering.
   Phone:(612) 625-0785, Fax:(612) 625-0572, Email: vail@cs.umn.edu

6. **Signature of Principal Investigator.**
   ______________________________
   Date____________

7. **Applicant Institution:**
   Regents of the University of Minnesota, DUNS 555917996, TIN 416007513

8. **Administrative Officer:**
   Kevin McKoskey, Senior Grant Manager Sponsored Projects Administration,
   University of Minnesota, 200 Oak St SE, Suite 450, Minneapolis, MN 55455-2070,
   Phone: 612-624-5066; Fax: 612-624-5599, Email: awards@umn.edu

   **Alternate contact information.**
   Sowolo Abdullai, Grant Administrator, Sponsored Projects Administration,
   University of Minnesota, 200 Oak St SE, Suite 450, Minneapolis, MN 55455-2070,
   Phone: 612-626-2724; Fax: 612-624-5599, Email: abdu0092@umn.edu

9. **Signature of Administrative Officer.**
   ______________________________
   Date____________
BAA-4.2.2. ABSTRACT

Geospatial Intelligence Analysis often requires the ability to represent and reason about ad-hoc, fluid network structures in complex, heterogeneous, seemingly disconnected data. Composite GEOINT networks that include infrastructure networks (e.g., transportation networks) as well as non-physical networks (e.g., social networks) may involve complex interactions between component networks, and exhibit spatio-temporal semantics. These networks present unique challenges in terms of their complex and heterogeneous structure, fluid properties, and their time-dependence. Also of significance are the multiple roles that the entities might assume and the fact that each constituent network may serve different purpose at different points of time.

In general, existing techniques in areas such as graph theory, spatial network databases, and time geography do not adequately model the composite nature of the networks, the complex interactions between them, and the multiple roles of entities and purposes of a network. Moreover the existing techniques do not effectively represent the fluid, ad-hoc nature of the networks of relevance in geospatial intelligence analysis, that might undergo frequent changes in their composition, structure, and functions. Finally the current methods do not effectively model and analyze networks that change with time. These limitations present major barriers to GEOINT analysts who must rely on manual labor to query disparate network sources of information and then integrate the information using rules based on manual knowledge and experience.

The following proposed novel approaches attempt to address these limitations by (1) exploring functional decomposition to represent complex networks, (2) modeling the fluid nature of complex networks by assigning “roles” to entities and “purposes” to networks. (3) providing a model based on time aggregation to represent the temporal changes in networks. The proposed research has the potential to make a significant contribution in the area of representation and reasoning about complex and fluid spatio-temporal networks (Topic 4.2). The research attempts to provide a common set of models and operations for improved reasoning on complex, time varying, fluid networks.

Our team, consisting of geographic information scientists with considerable domain knowledge in spatial databases particularly spatial networks and time-geography is capable of carrying out the proposed tasks. They not only have strong track records in G.I.Sc., data management, and analysis for army intelligence but they have also collaborated with various external agencies and individuals from academia over the years. The PI, Dr. Shashi Shekhar, is a leader in the field of Spatial Network Databases particularly modeling of networks that change with time (Time Aggregated Graphs) and storage and efficient access of large networks. Collaborators, Dr. Harvey J Miller, Chair of the Department of Geography at the University of Utah and Dr. May Yuan, Professor of Geography at the University of Oklahoma are renowned experts in the area of time geography. They will participate in periodic discussions to ensure that the proposed computer science ideas are meaningful and effective in the context of time geography. Collaborator Dr. Siva Ravada, director of software development in the Spatial Product development group at Oracle Corporation, USA, has been responsible for adding spatial features to Oracle’s RDBMS and Application Server products. Dr Wee Liang Heng from ESRI has extensive experience in devising innovative techniques for handling large network datasets. Dr. Budhendra Bhaduri is the Research Leader of the Geographic Information Science and Technology group at the Oak Ridge National Laboratory.
1. OBJECTIVES

The overall objective of this proposal is to model complex fluid spatio-temporal networks and provide operators for reasoning on the proposed representation. Several scientific and technical challenges arise when modeling such networks due to the spatio-temporal nature and heterogeneity of component networks. These represent major barriers to progress in the form of limited reasoning in Geospatial Information Science (G.I.Sc.) fields such as spatio-temporal databases, environmental criminology, etc. We propose a model called “Dynamic Purpose-Aware Graph” to represent complex, fluid spatio-temporal networks, and define operations for this model that would provide the means for efficient reasoning on networks.

The current body of knowledge may be categorized into the following main areas: graph theory [17, 11], spatial network databases [2, 6, 8, 9, 10, 19, 14, 13], and Time-Geography [6, 18]. Graph theory provides a set of mathematical concepts (e.g. nodes and edges) and operations (e.g. Min cut, path) for modeling networks. Social network analysis, transportation and evacuation planning applications utilize graphs and their properties for efficient data analysis. Recently concepts from graph theory have been introduced into databases [7, 9, 10], leading to the development of spatial network databases by providing the network data model, query language, storage structure, and indexing methods. One characteristic that distinguishes a spatial network database is the primary focus on the role of connectivity in relationships in addition to the spatial proximity between objects. These databases are the kernel of many important applications, including transportation planning; air traffic control; water, electric and gas utilities, and telephone networks. For example, spatial software libraries also known as cartridge, datablade or spatial option supported by Oracle, ESRI and IBM respectively [13, 14, 19] have existed for some time. The recent release of Oracle (version 10g) includes a network data model to store and maintain connectivity of link-node network and supports basic features such as shortest path [14]. The Network Analyst extension of ArcMap from ESRI supports a network geodatabase and provides basic algorithms (e.g., shortest path, service area, closest facility, etc.) [13]. There have also been considerable research initiatives in the field of query processing for spatial network databases [10]. In time-geography, the process-event [6] model is widely used to describe temporal changes.

Current methodologies have the following limitations. They do not adequately support the representation of complex networks that have heterogeneous components ranging from physical networks to social networks and their interactions. They do not effectively model fluid, ad-hoc changes to the structure, composition and function of complex networks or their components. The state of the art does not provide scalable data modeling techniques for and storage of dynamic complex networks.

Thus, we propose novel approaches to address these limitations by (1) developing a representation for complex networks based on functional decomposition and proposing operations for improved reasoning from the proposed representation. (2) proposing incorporation of “roles” to entities and “purposes” to component networks to address the fluid nature of complex networks and (3) providing a scalable data model based on time aggregation to represent the temporal nature of complex networks as well as operations to reason about the temporal variations.
2. ANTICIPATED RESULTS

Anticipated results of this proposal include mathematical models, computer algorithms, data analysis methods, and analytical tools, e.g. new software or extensions to existing software. Specifically, we anticipate the following results:

- **Mathematical models:** We anticipate developing a graph-based data model to represent the fluid and ad-hoc nature of complex networks. We expect that the proposed dynamic purpose-aware data model would provide an efficient way to represent time-variant composite networks.

- **Computer algorithms:** We expect to design new scalable query processing methods supporting complex queries such as finding an "in-purpose nearest neighbor" and to develop efficient and scalable accessing strategies for large datasets.

- **Data analysis methods:** We aim to extend existing network taxonomies to meet the requirements of dynamic complex networks. We also expect to define new operators for reasoning about the complex networks which would support new abilities such as functional decomposition.

- **Analytical tools:** We expect to develop new software and extend the capabilities of existing software such as Oracle Spatial and ESRI Network Analyst [13, 14] to incorporate newly designed purpose-aware graph models and purpose-aware dynamic graph models.

3. APPLICABILITY

Representation of and reasoning about networks is a crucial challenge in GEOINT analysis. For example, Scott Loomer [1] identified several critical networks of interest to GEOINT. These include Telecommunications, Electrical Power System, Banking and Finance, Transportation, Water Supply Systems, Emergency Systems, Continuity of Government Services, Public Health Services, Defense Industrial Base, Chemical and Manufacturing, etc. Let us now consider the role of a network in a few specific cases including a suspected nuclear site in Syria, money laundering in Manhattan banking facilities, etc.

Figure 1 gives a recent intelligence-related example of the process of a tall square building in Syria appearing, vanishing, and re-appearing again. Notice that at the suspected nuclear site appears (top image of Figure 1) in August '07 and then vanishes (middle image of Figure 1) in October '07 due to

Figure 1: Syria's Suspected Nuclear Facility [5]
(Note: Best Viewed in Color)
Source: New York Times and Digital Globe
a widely speculated Israeli air raid [22]. Then, in January ’08, the New York Times reported “…a tall, square building under construction that appears to closely resemble the original structure …” suggesting the re-appearance of the structure (See the bottom of Figure 1) [5]. Well known methods in image processing may detect changes at a selected site. However, analysts need spatial context including networks to interpret changes for GEOINT analysis, which may account for the underlying complex network structure of the data, that might also include heterogeneous components (e.g., road and river network). Also, the structure of the complex network and roles of the component objects are influenced by the appearance and disappearance of the square building. For example, the snapshots of the site taken in August 2007, October 2007, and January 2007 (as shown in Figure 1) show a water pumping station (see left side of each image for a circled structure) whose role may have changed from irrigation of the surrounding area to potential cooling of the suspected nuclear facility. Thus, traditional image processing methods need to be complemented with a novel network models and reasoning methods to interpret the dynamic nature of the pursuit to establish a possible nuclear facility.

Figure 2 gives another example from the domain of intelligence analysis illustrating a money laundering scheme (also called “micro structuring”) in New York that involved placing more than $111,000 into 112 accounts in various ATM locations within a small neighborhood and withdrawn from Columbia [15]. The suggested paths of the criminals in Figure 2 were created based on the seized deposit slips. Even though the intelligence information shown in Figure 2 may have been gathered by the FBI, GEOINT plays a major role due to several elements found in this example. These elements include geo-spatial networks (e.g. road, subway, and Automatic Teller Machine (ATM) networks) having heterogeneous components and the spatio-temporal nature of the dataset that provides the “glue” across all kinds of intelligence. Also, the function of these networks changes over time through the re-use of the network for criminal purposes. Though traditional methods such as ESRI network engine [13] can model transportation networks, they may not be able to adequately model non-transportation networks involving ATMs and bank accounts in the US and Colombia which may also have been involved in this incident.

Finally, an example of inferring intelligence information from a non-physical network may be found in social networks. Traditional social network theory models a “friend-of” relationship (i.e., edge) among a set of actors (i.e., nodes). It provides operations like centrality to identify
the most important actor, diameter query to identify a subset of actors close to the most important actor, and vulnerability analysis to identify critical actors whose departure may cripple the network. However, traditional social network analysis techniques are challenged by composite GEOINT social networks with adversarial scenarios such as multi-party competitive systems where there exist two types of relationships namely, "Friend of" and "Foe of", i.e., Hindrance networks [4]. However, traditional social networks focusing primarily on only homogeneous relationships in a single graph may not handle interactions between heterogeneous graphs. For example, a common query that may only occur on a composite network may be the formation of allies through the coalition (e.g. Allied nations during the World War II) of parties with a common dominant enemy (e.g. Hitler). This example also challenges diameter queries to capture coalitions via two independent graphs (i.e., friend of and foe of) to determine these relationships which may not be addressed by existing spatial network database methods [9, 10].

4. PROPOSED APPROACH TO ACHIEVE OUR OBJECTIVES
This section presents the scientific (Section 4.1) and management (Section 4.2) approaches.

4.1. SCIENTIFIC APPROACH
Efficient tools to reason over disparate, complex, and heterogeneous network data are crucial for many GEOINT analysis problems. While efficient and effective techniques for network data analysis using graphs have been studied, pertinent challenges in the context of geo-spatial intelligence analysis still need to be addressed. The first challenge arises from the heterogeneous and composite nature of the networks involved. These complex networks can include a variety of components ranging from infrastructure networks such as transportation networks to social networks (e.g., terrorist networks) and interactions between them. Second, the networks encountered in intelligence analysis are often fluid, undergoing ad-hoc changes in their structure and composition, thus making effective modeling of these networks difficult. The third challenge comes from the spatio-temporal nature of the networks. A model for a time-variant network needs to extend beyond a series of snapshots at different instant of time, to be able to answer the queries that might arise in intelligence analysis. This model may provide support for processing frequent queries, while also maintaining storage efficiency and representational adequacy.

This proposal addresses these challenges through an algebraic approach that consists of a representation (domain) and a set of operators. To address the complex nature of the networks, we propose the notion of functional decomposition, which could include discovering a modular structure or a hierarchical nature of a network. This approach is described in detail in Section 4.1.1. To model the fluid nature of the networks, we introduce the ideas of “role” and “purpose”. Every entity in a graph (node, edge) would be assigned a role and a network (or a component network) would be attributed with a purpose. As noted earlier, a pumping station being used for cooling purposes in a nuclear facility or for irrigation in the context of a river network, exemplifies the different ‘roles’ attributed to the node that represents the facility. A formal description of this approach is provided in Section 4.1.1. To model the time dependence of the networks, an approach based on time-aggregation and an extension of the concepts of events and processes are proposed. Here, the time dependence of networks is represented by allowing the properties of edges and nodes to be modeled as time series. This representation is described in Section 4.1.2.

4.1.1. Purpose-aware Graph (PAG) Modeling
Since analyzing complex, fluid networks is a key component of geo-intelligence analysis, there is a critical need for an effective model and a set of analysis tools for such networks. These
composite networks could consist of many component networks, each with its own purposes and every entity in a component network assuming multiple roles. In addition, these networks could display a fluid nature, suggesting the possibility of ad-hoc changes in their nature and structure.

Existing graph based approaches might not be able to model the complex and fluid nature of the graphs. They deal with graphs with singular purposes and assume unique roles for nodes and edges. Further, they might not provide enough support to represent relationships between component graphs. These limitations suggest the need to a model that can support complex, fluid graphs. However, incorporating purposes and roles in a graph raise many challenges. It may require the definition of new alternative semantics for common graph operations (e.g., in purpose neighborhood graphs) and the formulation of new operators such as decomposition and composition. For example, a nearest neighbor in classical spatial graph databases is defined as a set of components, found in a spatial proximity. When the purpose is considered, the nearest neighbor operator (in-purpose-nearest-neighbor) might be defined as the components whose functional uses are similar, in addition to being spatially close.

We propose a model called purpose-aware graphs that can model complex and fluid nature of graphs by taking into account their purposes, the roles of their features (e.g. nodes, edges), and the inter- and intra- functional dependencies. Formally, a purpose-aware graph (PAG) is defined as follows:

\[
PAG = (S =< G_1, G_2, G_3, ..., G_n >, P, h_1, ..., h_n, h : G \rightarrow P),
\]

where \( S \) is a collection of component graphs \( G_1, G_2, G_3, ..., G_n \) and \( P \) is a set of purposes, and \( h \) is the mapping from the component graph \( G_i \) to the set of purposes \( P \). Each component graph can be defined as

\[
G_i = (N, E, R, f_1, ..., f_n, g_1, ..., g_n, f : E \rightarrow R, g : N \rightarrow R),
\]

where \( N \) is the set of nodes, \( E \) is the set of edges, \( R \) is the set of roles that could be assigned to edges and nodes, and \( f \) and \( g \) are the mappings from the set of edges and set of nodes to the set roles respectively.

These concepts can be illustrated using Figure 1 first introduced in Section 3. The suspected nuclear site is composed of several component networks, such as a water supply network, a road network, a set of buildings, and social knowledge transfer networks, each used for a different purpose. In the proposed model, the component graphs, \( G_1, G_2, G_3, ..., G_n \), would represent the component networks and the set of purposes would include the specific intents of each network. In addition, each entity in a network could possibly have different roles in the context of different component networks. For example, the role of a pumping station would be to supply water to fields in an irrigation network or to pump water to cool the reactor in a nuclear facility. The model would represent the pumping station as a node and assign the roles such as irrigation and cooling to it, in the context of different networks.

In a purpose-aware graph, nodes can be either atomic or can include a hierarchy e.g. they may represent another component or composite graph \( G_i \). For example, the location 'Astoria' or midtown east Manhattan, modeled as a node in the top right part of Figure 2 is represented as a network in the bottom right part of the same figure. Edges in higher level graphs may represent relationships among higher level nodes that represent lower level networks (e.g., Astoria and midtown east Manhattan).

The tasks to carry out the proposed research are listed below.
**Task T1: Develop a Conceptual Model for Purpose-aware Graph Concepts:** We propose to develop a conceptual model to represent purpose-aware graph (PAG) properties. We will explore the possibilities of both adapting existing graph data models and developing new models for purpose-aware graphs. Figure 3b depicts an existing graph model from Oracle (version 10g) to model graphs and Figure 3a depicts possible extensions to the model. We will investigate purpose association, purpose manager, and purpose factory in the context of the existing model to associate purposes to the graph components, to maintain purpose ontology and roles in order to generate new semantics for ontology and roles respectively. We will explore a subgraph extension to represent subsets of component and/or composite purpose-aware graphs.

![Figure 3a. Proposed Extension](image)

![Figure 3b. Existing Graph model [21]](image)

**Task T2: Design Data types, Query Operators and Query Language for PAG:** We plan to explore various data types and query operators for reasoning with the purpose-aware graphs. We will design a set of query operators to develop an algebra closed under the proposed operations. Other possible operators that we would explore are those that (i) discover/predict the component of the graph that fulfills a specific purpose (ii) find how a purpose will be fulfilled, and (iii) determine what factors affect a specific purpose. Potential significant operations include composition and decomposition, shortest path queries, diameter queries, connected component labeling, and recursive path queries, in the context of purpose aware graphs. Also, operators specifically affecting the purpose or functionality of components or the networks would be incorporated. Some of the classical graph operators such as shortest path and k-core queries would be revisited to suit the semantics to the purpose aware graphs, namely in-purpose shortest path and in purpose k-core queries. We will study existing query languages to explore possible extensions by adding the required operators to suit purpose aware graphs.

**Task T3: Formulate Query Processing Algorithms:** We would explore the formulation of query processing algorithms to implement selected reasoning operators on the purpose aware graphs. We will investigate physical database issues (e.g., storage and indexing methods for processing strategies/algorithms, cost models and query optimization methods to select most suitable algorithms) in the context of purpose aware graphs.

**Task T4: Develop a taxonomy of rules and purposes:** We plan to investigate existing taxonomies to define purpose-aware graph semantics. We will explore taxonomies for GEOINT purposes to detect the functional relationships of component graphs. Such a classification can be seen in the threat taxonomy [16] in Figure 4. We plan to extend them by proposing new semantics. We also plan to develop new hypotheses by exploring unexpected combinations of
different graphs for different purposes. The goal would be to formalize computational models for roles and purposes in purpose aware graphs.

**Task T5: Validate the proposed model:**

We would evaluate the proposed purpose-aware graph model analytically and experimentally using real-world data such as Lincoln, NE crime dataset involving physical networks (e.g., street network) and logical networks (e.g., gang of criminals with “friend of” and “foe of” relation). We also plan to conduct the validation in consultation with domain experts from GEOINT and Environmental Criminology.

### 4.1.2 Dynamic Purpose-Aware Graph (DPAG) model

Information represented on complex networks encountered in GEOINT analysis may change over time, requiring the purpose aware graphs proposed in the previous section to be dynamic. These changes can be attributed to the restructuring of graphs or variation of roles of nodes and edges or purpose of the network. As shown in the example in Figure 1, the complex network consisting of the river network, roads and the buildings seems to change over time from August 2007 to January 2008. The tall square building represented as a node in the facility network disappears and reappears over time, changing the sets of nodes and edges. In addition, the pumping station appears to assume different roles over time. Addressing the dynamic nature of purpose-aware graphs is critical in answering some important questions in GEOINT analysis. These dynamic roles and purposes can be represented as the “re-purposing” of a network, where the nodes and edges assume new roles and the graph acquires a new purpose.

Traditional graph based approaches might not offer effective ways to model dynamic scenarios. Static graph based methods do not provide adequate support for the new semantics of operations introduced by the addition of time, as a dimension of analysis. Further, an existing approach based on time expansion [2] to represent a dynamic graph relies on replicating the network over the entire set of time instant, leading to excessive storage requirements. Though spatial network access methods such as CCAM [7] provide support for efficient network computations, they do not account for graphs that change over time. Hence representations for dynamic networks need to be investigated and efficient algorithms need to be designed to implement the key operations.

We propose a new extension called dynamic-purpose aware graph to represent time-variance of a purpose-aware graph. Dynamic purpose-aware graph (DPAG) models the time dependence by aggregating the purposes of a network and the roles of nodes and edges into time series attributes. It is formally defined as

$$DPAG = \{DS = <DG_1, DG_2, \ldots, DG_n>, P, T, h_1, \ldots, h_n | h_i : DS \rightarrow P^T\},$$

where $DS$ is a composite dynamic graph, $P$ is the set of purposes, $T$ is the time period, $h$ is the mapping from a composite graph to a time series of purposes, and $P^T$ represents a time series of purposes.

$$DG = \{N, E, T, R, f_1, \ldots, f_h, g_1, \ldots, g_I | f_i : N \rightarrow R^T, g_i : E \rightarrow R^T\}$$
where, $N$ is the set of nodes, $E$ is the set of edges, $T$ is the time period, $R$ is the set of Roles, $f$ and $g$ are the mappings from nodes and edges to time series of roles, and $R_T^T$ represents a time series of roles.

Figure 6a illustrates three snapshots of the suspected nuclear facility network in Figure 1. Nodes N1, N4 and N5 represent pumping station, secondary structure and the square building respectively. Node N2 models the intersection of the facility access road with the major highway along the river and N3 represents a fork in the access road, branching to the square building and the secondary structure. Edges represent the road segments connecting the nodes. As illustrated by Figure 1, node N5 (the square building), and edges N3-N5 and N4-N5 that connect the square building to the network are absent in the second snapshot (Oct '07). A time aggregated representation of this scenario is shown in Figure 6b. Each node and edge has an existence time series across the three snapshots. A value ‘0’ in the time series represents an absence of the associated node (or edge) in the corresponding snapshot whereas the value ‘1’ represents a presence. For example, the time series attribute [1, 0, 1] on node N5 and edges N3-N5 and N4-N5 indicate their absence in the second snapshot and presence in the first and the third snapshots.

The proposed research would be accomplished through the following tasks:

**Task G1: Model Event and Processes:** Concepts of processes and events have been used to model temporal variations in geospatial data [18, 23]. In the example shown in Figure 2, the money laundering activity can be modeled as a process, whereas the bank run where a large amount of money was deposited in a set of banks in close proximity can be called an event. In Figure 1, disappearance and appearance of tall square building may be represented as events in a broader process of Syria trying to acquire nuclear capability. We will explore the possibility of extending the concepts of “Events and Processes” to the conceptual model of DPAG thus making these concepts applicable in the context of complex networks.

**Task G2: Formulate Data types and Query Operators:** We plan to explore various abstract data types for the DPAG. We will also formulate a complete set of query operators to enable users to frame queries to retrieve information from a DPAG, and develop an algebra that is closed under these operations. Examples of operators include $get\_edge\_roles(graph, node1, node2)$ that would return the time series of roles assumed by the edge from node1 to node2. Similarly the operator $get\_purposes(graph, t1, t2)$ would result in the set of purposes of the graph between time instants $t1$ and $t2$.

**Task G3: Design Algorithms for Dynamic Purpose Aware Graphs:** We plan to design query processing algorithms to support efficient reasoning on DPAG. Table 1 lists a set of relevant questions that might arise in the context of purpose aware graphs. For the example shown in
Figure 1, the query 1 b) listed in Table 1 would return the roles namely, [irrigation, cooling] for the node representing the pumping station. We will formulate algorithms for DPAG operators that would help answer similar questions.

<table>
<thead>
<tr>
<th>Purpose-Aware Graph</th>
<th>Dynamic Purpose-Aware Graph</th>
</tr>
</thead>
<tbody>
<tr>
<td>1a) What is the role of a node or an edge?</td>
<td>1 b) How does the role of a node or an edge vary over time?</td>
</tr>
<tr>
<td>2 a) How will a particular purpose be fulfilled; what are the roles of the nodes or edges that are involved?</td>
<td>2 b) When is the best time to achieve a particular purpose?</td>
</tr>
<tr>
<td>3 a) What are the nodes or edges required to identify a specific purpose?</td>
<td>3 b) What are the nodes and edges that causes the re-purposing of a network?</td>
</tr>
<tr>
<td>4 a) Which nodes and edges can be part of a specific purpose?</td>
<td>4 b) Which nodes and edges are part of a series of re-purposing?</td>
</tr>
</tbody>
</table>

Table 1 Graph operations for PAG and DPAG

**Task G4: Develop Storage and Access methods for Dynamic Purpose Aware Graphs and their extensions:** We will develop indexing methods to provide efficient access to the dynamic purpose aware graphs. The extensibility of existing access methods like CCAM [7] will be studied in the context of DPAGs. We will also study and compare the performances of various disk based data structures for the storage of the dynamic purpose-aware graphs. Further, data structures for the storage of dynamic networks, incorporating multiple node types and complicated objects as nodes need to be designed.

**Task G5: Validate Dynamic Purpose-Aware Graphs:** We would validate proposed dynamic purpose-aware graph model using real-world data such as Lincoln, NE crime dataset involving physical networks (e.g., street network) and logical networks (e.g., gang of criminals with “friend of” and “foe of” relation) in consultation with domain experts from GEOINT and Environmental Criminology.

**4.2 MANAGEMENT APPROACH**

We will measure the success of this project in terms of (i) successful G.I.Sc. research resulting in the creation of new dynamic purpose-aware graph models, (ii) the building of tools embodying the new results, and their use by GEOINT analysts, (iii) the success in being able to address questions on complex network structures which were previously intractable.

<table>
<thead>
<tr>
<th>Quarters</th>
<th>Year 1</th>
<th>Year 2</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Q1</td>
<td>Q2</td>
</tr>
<tr>
<td><strong>Scientific Approach</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Purpose aware Graph Models (4.1.1)</td>
<td>T1</td>
<td>T2</td>
</tr>
<tr>
<td>Dynamic purpose aware Graph Models(4.1.2)</td>
<td>G1</td>
<td>G2</td>
</tr>
<tr>
<td><strong>Management Approach</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Progress Monitoring</td>
<td>Quarterly review of goals</td>
<td></td>
</tr>
<tr>
<td>NGA Reports</td>
<td>Six month progress report to NGA</td>
<td></td>
</tr>
</tbody>
</table>

Table 2: Project Task Schedule (Tasks T1…T5,G1…G5,described in Section 4.1)

The main milestones will include: (1) the development of a purpose-aware graph data model for reasoning on complex networks by defining a conceptual model, a set of operators and a set
of algorithms; (2) the development of a dynamic purpose-aware graph data model for reasoning on complex Networks by extending the time aggregated graph model for addressing the temporal semantics. The breakdown of the research tasks is given in Table 2, which lists the milestones for quarterly intervals and project management activities to track the progress of this project. Milestones are categorized based on the 2 sets of tasks proposed in Sections 4.1.1 and 4.1.2. Each work-package is subdivided into well-defined quarterly tasks specified in Section 4.1. The last quarter is reserved for preparation of the final report.

In the third, fourth, and fifth optional years, we will explore the tools and the test bed to handle functional decomposition, hierarchical DPAG, interaction across component networks and ontologies for roles and purposes. Additional research tasks and work-packages will be determined in consultation with NGA NURI program managers and GEO INT Analysts.

The PI, Dr. Shekhar, will provide overall leadership for the project. He has experience in leading similar and larger size research teams. Graduate and undergraduate students will be working on the research and development of the methodologies proposed in the project.

5. FACILITIES AND EQUIPMENT

The department of Computer Science at the University of Minnesota has several workstations and servers available for the proposed research. However, we will need high-resolution monitors on systems equipped with sophisticated video cards and high memory cards for visualization of large and complex Networks and validation of our results. Real datasets provided by Mr. Tom Casady (Chief of Police, Lincoln, NE) and Dr. Ned Levine (CrimeStat Project U.S. Department of Justice) will be used as input data to test our techniques.

6. COLLABORATION

We plan to collaborate with commercial firms, namely, ESRI, Oracle, and Dr. Ned Levine and Associates. This will provide a means to disseminate research results. Dr. Budhendra Bhaduri will provide guidance on data-integration across disparate networks based on his experience in large data-integration projects at ORNL (Oakridge National Laboratory). Dr. Harvey J Miller and Dr. May Yuan are renowned experts in the area of time geography. They will participate in periodic discussions to ensure that the proposed computer science ideas are meaningful and effective in the context of time geography.

7. STUDENT TRAINING AND INSTITUTIONAL IMPROVEMENT

This grant will enhance capacities of University of Minnesota in the G.I.Sc. areas. A direct outcome of this research project will be the training and development of G.I.Sc. graduate students. Two Ph.D. students will be supported by research assistantships. Through this project, they will interact with team members from diverse domains and learn valuable skills of communicating and contributing to G.I.Sc. research. The research will provide G.I.Sc. projects for the Undergraduate Research Opportunity Programs and undergraduate honors theses. The research results will provide two weeks' worth of materials in G.I.Sc. courses. Students will be encouraged to participate at relevant G.I.Sc. Conferences such as annual summer assembly of the University Consortium on Geographic Information Sciences.

8. CURRENT AND PENDING SUPPORT

The research proposed in this proposal has not been and will not be submitted to any other sources of funding during the evaluation period.
AA-4.2.4. RESEARCH TEAM

Table 3 lists the names of all persons for whom financial support is proposed, the planned commitments (in units of a percentage of full-time work year) to the proposed research, and the planned commitments to other work and professional activities.

<table>
<thead>
<tr>
<th>Research Team</th>
<th>Planned Commitments (% full-time work year)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>University of Minnesota</td>
</tr>
<tr>
<td></td>
<td>Faculty</td>
</tr>
<tr>
<td></td>
<td>Graduate Students</td>
</tr>
<tr>
<td>This Proposal</td>
<td></td>
</tr>
<tr>
<td>Shashi Shekhar</td>
<td></td>
</tr>
<tr>
<td>Graduate Students</td>
<td></td>
</tr>
<tr>
<td>Projects</td>
<td></td>
</tr>
<tr>
<td>Dynamic Purpose-aware Graph Data Models for Representing Reasoning About Networks (THIS PROPOSAL)</td>
<td>9%</td>
</tr>
<tr>
<td>1. III-CXT: Spatio-temporal Graph Databases for Transportation Science, National Science Foundation, 08/2007 – 07/2010, $449,993</td>
<td>9%</td>
</tr>
<tr>
<td>2. Spatio-Temporal Pattern Mining for Multi-Temporal Activity Datasets, National Geospatial-Intelligence Agency, 08/14/2007 – 07/30/2009, $300,000.</td>
<td>9%</td>
</tr>
<tr>
<td>3. Modeling and Mining Spatio-temporal Data, Army Corps of Engineers, (student support only) 03/15/2006 - 09/30/2008, $112,000.</td>
<td>5%</td>
</tr>
<tr>
<td>4. CRI:IAD Infrastructure for Research in Spatio-Temporal Context-Aware Systems and Applications, National Science Foundation (Equipment grant), 07/01/2007 – 06/30/2010, $140,403</td>
<td>5%</td>
</tr>
<tr>
<td>Other Work and Professional Activities</td>
<td></td>
</tr>
<tr>
<td>University related teaching and professional services</td>
<td>Balance</td>
</tr>
<tr>
<td>Total</td>
<td></td>
</tr>
<tr>
<td>100%</td>
<td>50%</td>
</tr>
</tbody>
</table>

Table 4: Summary of Research Team and Planned Commitments
REFERENCES

13. ESRI Network Analyst.