NASA/CR-2005-213522

A Comparison of Geographic Information Systems, Complex Networks, and Other Models for Analyzing Transportation Network Topologies

Michael Kuby, Sean Tierney, Tyler Roberts, and Christopher Upchurch
Arizona State University, Tempe, Arizona

February 2005

The NASA STI Program Office . . . in Profile

Since its founding, NASA has been dedicated to the advancement of aeronautics and space science. The NASA Scientific and Technical Information (STI) Program Office plays a key part in helping NASA maintain this important role.

The NASA STI Program Office is operated by Langley Research Center, the lead center for NASA's scientific and technical information. The NASA STI Program Office provides access to the NASA STI Database, the largest collection of aeronautical and space science STI in the world. The Program Office is also NASA's institutional mechanism for disseminating the results of its research and development activities. These results are published by NASA in the NASA STI Report Series, which includes the following report types:

- TECHNICAL PUBLICATION. Reports of completed research or a major significant phase of research that present the results of NASA programs and include extensive data or theoretical analysis. Includes compilations of significant scientific and technical data and information deemed to be of continuing reference value. NASA counterpart of peer-reviewed formal professional papers, but having less stringent limitations on manuscript length and extent of graphic presentations.

- TECHNICAL MEMORANDUM. Scientific and technical findings that are preliminary or of specialized interest, e.g., quick release reports, working papers, and bibliographies that contain minimal annotation. Does not contain extensive analysis.

- CONTRACTOR REPORT. Scientific and technical findings by NASA-sponsored contractors and grantees.

- CONFERENCE PUBLICATION. Collected papers from scientific and technical conferences, symposia, seminars, or other meetings sponsored or co-sponsored by NASA.

- SPECIAL PUBLICATION. Scientific, technical, or historical information from NASA programs, projects, and missions, often concerned with subjects having substantial public interest.

- TECHNICAL TRANSLATION. English-language translations of foreign scientific and technical material pertinent to NASA's mission.

Specialized services that complement the STI Program Office's diverse offerings include creating custom thesauri, building customized databases, organizing and publishing research results ... even providing videos.

For more information about the NASA STI Program Office, see the following:

- Access the NASA STI Program Home Page at ***http://www.sti.nasa.gov***

- E-mail your question via the Internet to help@sti.nasa.gov

- Fax your question to the NASA STI Help Desk at (301) 621-0134

- Phone the NASA STI Help Desk at (301) 621-0390

- Write to:
  NASA STI Help Desk
  NASA Center for AeroSpace Information
  7121 Standard Drive
  Hanover, MD 21076-1320

NASA/CR-2005-213522

A Comparison of Geographic Information Systems, Complex Networks, and Other Models for Analyzing Transportation Network Topologies

Michael Kuby, Sean Tierney, Tyler Roberts, and Christopher Upchurch
Arizona State University, Tempe, Arizona

National Aeronautics and
Space Administration

Langley Research Center
Hampton, Virginia 23681-2199

Prepared for Langley Research Center
under Contract NNL04AA57G

February 2005

Acknowledgments

The authors would like to thank Natalia Alexandrov of NASA Langley Research Center for her support for, and comments on, this report. The characterizations and conclusions expressed in this report, however, are solely those of the authors.

Available from:

NASA Center for AeroSpace Information (CASI)
7121 Standard Drive
Hanover, MD 21076-1320
(301) 621-0390

National Technical Information Service (NTIS)
5285 Port Royal Road
Springfield, VA 22161-2171
(703) 605-6000

Abstract

This report reviews six classes of models that are used for studying transportation network topologies. The report is motivated by two main questions. First, what can the "new science" of complex networks (scale-free and small-world networks) contribute to our understanding of transport network structure, compared with the kinds of research questions answered by more established methods? Second, how can geographic information systems (GIS) contribute to studying transport networks? The report begins by defining terms that can be used to classify different kinds of models by their function, composition, mechanism, spatial and temporal dimensions, certainty, linearity, and resolution. Six broad classes of models for analyzing transport network topologies are then explored: GIS; static graph theory; complex networks; mathematical programming; simulation; and agent-based modeling. Each class of models is defined and classified according to the attributes introduced earlier. The paper identifies some typical types of research questions about network structure that have been addressed by each class of model in the literature. GIS is able to address a variety of different network-related research questions on its own, but it can also provide a spatial database and mapping platform for other methods. Therefore, for each of the other classes of models, we explore the potential for integrating them with GIS in what are known as Spatial Decision Support Systems (SDSS). We also provide, in an appendix, a preliminary list of spatial GIS data layers relating to the air transport network. This report should be considered a starting point for further investigations, as we have only begun to explore the classes of models for studying network topology, the types of research questions addressed within each class, and the specific applications undertaken for each type of question.

1. Introduction

The structure of transportation networks varies greatly by transport mode, by stage of evolution, by level of development, and by technology. New technologies and new business models constantly restructure networks, although the influence of earlier forms often remains indelible. Researchers in fields ranging from economics, geography, business, engineering, urban planning, mathematics, regional science, and others have employed a wide variety of methods to study transport network topologies. Two of the most important recent innovations in transportation network modeling are complex networks and geographic information systems (GIS).

Recently, a "new science of networks" has emerged from outside of the traditional transportation-related disciplines [9]. The new approach to networks, commonly referred to as "complex network theory," focuses on scale-free or small-world networks that display complex adaptive behavior. Complex networks grow and evolve in size and complexity through systemic interactions among network nodes and subsystems, and they display unforeseen (emergent) behavior [3].

A large subset of transportation network analysis and design problems, however, can be handled with traditional methods, such as graph theory, mathematical programming, simulation, and agent-based modeling. One of the foci of this preliminary paper is to establish the domain of traditional approaches to studying network topology vs. the complex-network approach. This will point us toward the missing pieces in methods and tools and identifying the most promising opportunities for scale-free network analysis.

A second focus of this preliminary report is geographic information systems (GIS). While GIS has been growing rapidly for 10-20 years, it has recently been highlighted as one of the most important technologies for the 21st century [40]. On its own, GIS is a valuable tool for studying transportation networks, but it can also provide a spatial database and mapping platform for other methods. The integration of GIS with other methods, known as spatial decision support systems (SDSS), has great advantages for modeling transportation networks. To understand the kinds of network-related research questions within the domain of GIS and what it can contribute to other methods as part of an SDSS, it is necessary to outline the nature of the other methods. The methods covered in this report are:

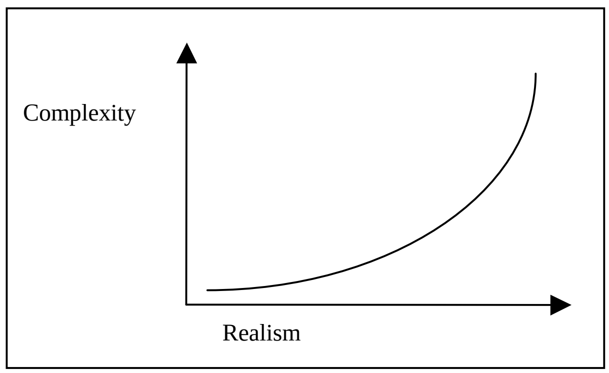
1. GIS
2. graph theory
3. complex networks
4. mathematical programming
5. simulation
6. agent-based modeling

Although we go into greater depth for GIS, the discussion of all methods in this paper should be considered a preliminary overview and no more than a possible starting point for deeper investigation. Furthermore, this list of methods is not all-inclusive—the scope of this report is such that it cannot include all methods that can contribute to the study of transportation network topology.

The report begins by introducing a classification scheme for models. Models are classified according to the dimensions of function, composition, process, spatiality, temporality, certainty, linearity, and resolution. In the following six sections, each broad class of method is first defined and then characterized along the eight different dimensions. Next, we list several types of research questions related to transport networks that the method is able to answer. These lists of research question types are meant to be exemplary, not comprehensive. Likewise, the references cited should be seen as selected examples, not seminal works or comprehensive lists. Finally, each section concludes with a discussion of the benefits of integrating the method with GIS—or in the case of GIS, the benefit of combining it with other methods in a spatial decision support system (SDSS).

2. A Classification System for Models

Models can be defined as "simplified representations of reality" [93], "simplified statements of structural interdependence" [44], or more simply as a way to connect inputs and outputs. Scientists simplify reality to generalize its most fundamental, relevant, or interesting aspects. Models allow scientists to focus on the main signals and filter out the noise. Models cut a problem down to a practical size, and allow scientists to get an answer to a question by applying what they believe to be the general principles to the situation in question. Scientists build models by extracting certain key phenomena for study and suppressing or compensating for incidental details and/or complicating factors. Because of the extraction and suppression that goes into building them, models have a *subjective element*—they are not purely objective. The fundamental tradeoff in model-building is that realism tends to be purchased at the cost of increasing complexity, with diminishing marginal returns (Figure 1). Model transparency may also be sacrificed.



Because reality can be modeled in so many different ways, it is important to develop a terminology for describing them. Models can be classified according to many dimensions. Every model can be classified according to *each* of these dimensions, i.e., each model has a function, something of which it is constructed, a way of treating time and space, etc. Most of the following discussion is our selection and reinterpretation of the ideas originally presented by Haggett and Chorley [44], Harvey [47], Thomas and Huggett [93], and Wilmott and Gaile [100].

1. Function—what does it do?

The most important distinction among models is its function, or literally what does it do? Here it is important to distinguish between what the model actually does and what it is used for. Models can be descriptive, explanatory, predictive, or prescriptive. The function of *descriptive* models such as maps, networks, indices, and diagrams is simply to describe reality. *Explanatory* models go one step beyond description by attempting to explain why. The function of *predictive* models is to predict an outcome or output given some inputs, i.e., they usually answer a "what if" question. In contrast, *prescriptive* (or normative) models answer a "should" question, as in where or how should we do something, or what is the best way to do something.

A subtle but crucial distinction is between the function a model actually performs and the purpose for which it is used. For instance, simulation models may be used for identifying the best network structure, but what the simulation model actually does is predict how different networks might perform—the model itself cannot design the network. Also note that some models have more than one function. Other model purposes mentioned in the literature include psychological, acquisitive, organizational, logical, systematic, constructional, and cognitive, but these can usually be boiled down to the other four.

2. Composition—what is it made of?

We will not dwell on this distinction, as nearly all models of transport network topology are *mathematical* and/or *graphical.* Other models can be *physical* (e.g., a wave tank or wind tunnel), *physical analogue* (composed of a different but analogous physical material), *conceptual/historical analogue* (e.g., A is like B), or *conceptual*.

3. Mechanism—is the cause and effect mechanism included in the model?

The major distinction here is between process-response and black-box models. In a *process-response* model, the components and process of cause and effect are explicitly represented. A *black-box* model represents the relationship between cause and effect but does not explicitly account for the process. A graph of the relationship between traffic volume and average speed on a highway is a black-box relationship, whereas a simulation of individual vehicles and driver behavior is a process-response model. Many models are hybrids of the two pure forms.

Process-response models tend to be more complex, but not necessarily more realistic because process-response models depend on accurate depiction of the underlying mechanisms whereas black-box models can be finely tuned to the outcomes that one is trying to model. On the other hand, black-box models may be less generalizable outside of the context in which they were fitted, whereas process-response models may do far better and understanding interactions in a completely new context for which there are no data yet.

4. Geography—how does it treat space?

In the field of transportation, most models are explicitly *spatial*. Very few, such as supply and demand curves in microeconomics, are *nonspatial* in that they ignore space altogether. The distinguishing factor among transportation models—and a key to understanding the importance of GIS—is how they treat space. Explicitly spatial transportation models can simplify space to be *linear* (e.g., a function of distance from an airport), *network*, *2-dimensional*, or *3-dimensional*. Digital elevation models are recognized as *two-and-a half dimensional* in that every x, y coordinate has one and only one associated z-coordinate (three dimensional models can have multiple z-coordinates, allowing for overhangs and other complex shapes, as well as subsurface features). Furthermore, in each of these types of models, the space can be treated as *continuous* or simplified to a set of *discrete* points. The choices made in modeling space affect the complexity and realism of the model.

5. Time—how does it treat time?

Models can be static or dynamic. *Static* models do not explicitly represent more than one period or instant of time. Supply and demand curves, for instance, do not represent the process of achieving equilibrium over time. *Dynamic* models, on the other hand, explicitly represent more than one period or instant of time. This can be done in several ways. *Continuous* time models allow an event to occur at any instant of time along a time line, for instance, by selecting the time until the next event from a continuous random probability distribution. *Discrete* time models break time into representative periods or instants. These discrete time periods can be at regular intervals (hours, days, weeks, months, years, five-year plans, etc.) or intermittent (drawn from a discrete probability mass function). The choices made in modeling time, like those for modeling space, affect the complexity and realism of the model.

6. Certainty—are the values of the key numbers taken as known or unknown?

The basic distinction here is between deterministic and stochastic models. In a *deterministic* model, all numbers are input to the model as single values. Given the same inputs, a deterministic model will deliver the same outputs every time, except perhaps when breaking ties. In a *stochastic* or *probabilistic* model, certain numbers are considered to be unknown and represented by a probability distribution rather than a single value. Stochastic models will create a range of possible outcomes from the same inputs, and therefore are usually run multiple times.

Inferential statistics models fall in a gray area between deterministic and stochastic, but by our definition would actually be classified as deterministic. Inferential statistics models explicitly acknowledge that one's sample is but one of many that could possibly have been drawn from a population, and they produce an output that quantifies the resulting uncertainty. Yet, given the same sample as an input, a statistical model such as a multiple regression model follows a deterministic process that will always produce the same exact estimates of slope, intercept, error, and significance. By this definition, it is a deterministic model of a stochastic process.

7. Mathematical relationships—is the model linear?

Relationships in mathematical models can be linear or nonlinear. *Linear* functional forms are sometimes adopted because they (a) provide reasonable approximations of the real world and (b) are tractable by certain methods. This distinction is perhaps most important in mathematical programming models. Linear programming is a powerful optimization tool, but is limited by the fact that all variables must be continuous and the objective function and all constraints must be linear, as in $aX_1 + bX_1 + cX_1$…. Extensions to linear programming include integer programming, which can model *binary* (0,1) or *integer* (0, 1, 2, etc.) decision variables, and *nonlinear* programming, which allows for nonlinear terms such as XY or X^2. For modeling transportation networks, binary and integer variables are especially important for modeling decisions such as all-or-nothing infrastructure investment decisions or one-or-the-other passenger mode-choice decisions.

Some problems are linear in nature, and some are not. If the problem is nonlinear in nature, sometimes linear models can approximate the nonlinear effects purely for tractability. Often it is difficult to solve nonlinear problems or linear problems with a large number of discrete variables. For instance, it is relatively easier to formulate and solve a model that approximates network capacity as a linear "brick-wall" limitation that cannot be exceeded, compared with the more accurate, nonlinear representation of capacity as an upward-curving congestion-cost function.

8. Resolution—what are the smallest units of observation?

A major difference among transportation models is between aggregate and disaggregate models. *Aggregate* models typically model the transportation patterns of geographic regions, which include traffic analysis zones (TAZs), census tracts, zip codes, counties, metropolitan areas, states, provinces, and countries. All of the traffic generators or attractors within the spatial unit are aggregated. Data for each region can be in the form of averages (e.g., of income or age), totals (e.g., of jobs or people), or percentages (e.g., of car owners). Each zone is treated as homogeneous, with all individuals assumed to share identical average traits. *Disaggregate* models, on the other hand, represent each individual separately, by location and/or socioeconomic characteristics. Greater realism in this case may be possible at the expense of enormous computational and data requirements and greater privacy concerns. *Partially disaggregated* models relax the assumption that all individuals within a zone are identical and average. They disaggregate the population by family size, income brackets, occupation, or age, and then model each subgroup separately.

The eight aspects of models summarized here are valuable for characterizing the nature of transportation network models. These dimensions affect which kinds of research questions a type of model can address. There are, of course, other ways of categorizing models that were not included on this list.

3. Geographic Information Systems (GIS)

A. Definition

Geographic Information Systems (GIS) have emerged as an essential analytical tool for a variety of human endeavors, including agriculture, anthropology, biology, business, geology, planning, and transportation. GIS has been defined as:

> A system of hardware, software, data, people, organizations and institutional arrangements for collecting, storing, analyzing, and disseminating information about areas of the Earth [31].

The above definition combines the components of a GIS (hardware, software, data, people, organizations and institutional arrangements) and the functions of a GIS (collecting, storing, analyzing, and disseminating information about areas of the earth) in a single summary.

The acronym “GIS” is also used for Geographic Information *Science* [42]. GIScience is the more theoretical study of how to best represent and manipulate spatial data in electronic format. Since the remainder of this section emphasizes practical applications, the “S” in GIS will refer to systems rather than science. Nevertheless, the importance of GIScience should not be understated, since this research continues to expand the capabilities of GIS.

GIS meshes geographic data into a single software environment using data layers. Data layers contain a single geographic element or a class of geographic elements, such as streets, maintenance facilities, topographic elevation, or census tracts. Data layers consist of the following components:

1. the spatial coordinates of each element of the layer;
2. an associated table containing attribute information about each element (e.g., census tract attributes such as population, average age, or median income; or the length, maximum speed, capacity, and surface type of streets); and
3. the visual display of the elements and/or their attributes.

Different data layers can be related to each other by their spatial locations or by one of their attribute fields.

Data layers are at the heart of a GIS. Appendix 1 provides a long but undoubtedly incomplete list of data layers that are potentially useful for planning or studying air transport networks. Data layers can be collected using global positioning systems (GPS) that record latitude, longitude, and altitude, by scanning or digitizing paper maps, from satellite imagery or aerial photography, from other hardware devices such as road sensors or wind gauges, by importing data layers from other GISs, or in tabular database format that can then be “joined” to spatial coordinates through one of the attribute fields. Data layers can be stored on a single computer or shared over a network. Data layers can be analyzed singly or in relation to each other using a wide variety of software tools, some of which are specific to GIS and explicitly spatial, while others are tools originally developed for databases, spreadsheets, or mathematical modeling. Finally, GIS layers can be disseminated to various people and organizations in digital (graphic or tabular) or paper form, and more recently, there is a growing distribution of GIS data with which users can interact over the Internet.

Transportation systems are designed to overcome geographical separation. With increasing amounts of geographic information available in a digital form, GIS offers a platform for a more comprehensive analysis [65]. By combining location data with attribute information, GIS offers visualization capabilities and a data management platform to study and improve transportation systems. Greater efficiency and increased accessibility within transportation demand increasingly flexible and scalable solutions. This flexibility is necessary because transportation decisions are impacted by a variety of social, environmental, and financial constraints. Analytical tools must balance these competing issues. Furthermore, scalability is critical for decision making on local, regional, and national levels. GIS gives its users both flexibility and scalability previously unseen in transportation analysis.

GIS's ability to integrate a variety of different data layers into a software environment in order to manipulate and analyze the *spatial* relationships is of great value for transportation planners. Without the spatial component, GIS is very similar to a database management system, where information is queried based upon some index [91]. This spatial aspect takes advantage of the geo-referencing system that ties objects to unique locations in space.

a. Coordinates, Projections, Scale, and Metadata

In order to be able to relate different data layers in a GIS to each other, they must be defined and displayed with the same coordinate system, projection, and scale. Coordinates, scale, and projection are at the heart of the problems most users have with getting one set of data to conform with a dataset acquired from a different source. There are a variety of coordinate systems in use besides longitude and latitude, such as Universal Transverse Mercator (UTM) or, in the USA, various state plane systems centered on a particular part of a particular US state. Next, the data layers must be projected on the screen using the same units of distance. Finally, they must use the same projection system to translate the spherical surface of the earth into a flat representation on a computer monitor. GIS has tools that can convert both the coordinate system and projection so they are uniform; in some instances the software can handle the conversion on the fly and in other instances the users will have to make the change manually.

While this all-inclusive technique has many benefits, GIS users need to be aware of the data integrity issues that can compromise the accuracy of an analysis. Recall that the components of GIS included organizational and institutional arrangements. While digital geographic data are more ubiquitous than ever, the aggregation and dissemination of this information lacks uniformity. *Metadata* exists to provide information to secondary users of GIS data. Metadata is information about the associated spatial data: the source, projection, date acquired and assembled, coordinate system used, etc. Accurate metadata makes it much easier to deal with the problems of scale, projection and coordinate system mentioned above. More importantly, it allows secondary users to assess the accuracy and precision of GIS data. GIS's ability to store and display location data with great precision can lead to unwarranted precision. While the software may specify a feature's location down to a fraction of an inch, that location may be no more than a rough guess by the layer's original creator. Metadata explaining the origin of a layer's location information allows subsequent users to make informed decisions about how much faith to put in the data's accuracy.

b. Data Implementation Environments: Vector vs. Raster

Within a GIS, information is viewed and stored using two very distinct formats: vector data and raster data. Vector data represents spatial objects as points, lines, or polygons. Attribute information about each point, line, or polygon is stored in a database table. Polygons represent features with an areal extent, such as an entire state. Lines are used for features of negligible width, such as a road or river. Points features have no dimensions, only a location, such as a survey marker. Depending on the scale and purpose of the analysis, the same entity might be modeled as a point or an area. For instance, an airport would probably be a point feature on a layer representing an entire state, but would be a polygon on a city level layer. Similarly, at most scales, an access road would be treated as a line, but for traffic engineering purposes, it would become a polygon feature, with a distinguishable width.

In a raster format, data consist of regularly spaced grid cells of the same shape and size. Raster data layers often contain information that is continuous (present everywhere) or planar in nature, such as elevation, temperature, air pollution, or vegetation. Raster data stores information in predefined cells of an arbitrarily assigned size. Their size, shape, orientation, and starting point may be arbitrary. Defining small cell sizes produces highly accurate and specific data, but results in longer processing time and requires a lot of storage space. Conversely, using larger cell sizes can ease the computational and storage burdens, but information will be lost as the cells become larger and more generalized.

Remotely sensed imagery, from satellite or aerial photos, often dictates the use of raster formats and may determine the cell size. Remote sensing has made enormous contributions to our understanding of the earth's systems at a regional and global level. Analysis of remotely sensed imagery requires an application that can display, interpret, and manipulate this data. GIS has emerged as a leading platform for clarifying the spatial organization of the earth's systems by converting digital images into data that can then be analyzed in a variety of ways.

c. Data Interpretation Models: Object, Network, and Field

It is useful to think of three kinds of spatial structures that can be represented in GIS. Objects are unique features on the earth. Networks are connected systems of points and lines enabling movement. Fields are some kind of signal or force exerted over space.

Objects are simple to represent in a vector GIS using points, lines, and polygons. In a raster GIS, membership in a given object can be defined as a 0-1 condition of each cell. For instance, the object "Virginia" or "wetlands" can be defined by assigning a value of 1 to all qualifying cells.

GIS generally represents transportation *networks* with vector data using points and lines. Points (nodes) correspond to arc endpoints and intersections, while lines (arcs) refer to the links that connect different nodes. In a GIS, however, not every set of points and lines constitutes a network. A network requires not only points (nodes) and lines (arcs), but a table of information describing their connectivity. This attribute table for the arcs will define whether the arc is one-way or two-way, and describe its other attributes such as length, time, cost, impedance, capacity, etc. There can be other layers of points and lines in a vector GIS that are not considered a network. Other points may represent locations of air traffic control towers or major employers, while other lines could represent rivers or other linear features.

Raster data are less commonly used in transportation network analysis, but raster GIS software is usually equipped with some network tools. In a raster GIS, a network can be defined as all cells through which a network link passes. Such cells can be given a low impedance representing the typical time or cost to traverse a network cell. Off-network cells can be given a higher impedance—prohibitively high if the analyst wishes to prevent any and all off-network travel. Finding shortest paths over a raster network is accomplished using a "cost-distance" command that finds the sequence of contiguous cells offering the lowest total impedance from the origin cell to the destination cell. The ability to do network analysis in a GIS context offers several advantages, such as being able to define highly accurate service areas [94] or to relate network decision-making with environmental, land use, and other planar variables.

Spatial relationships that tend to decay over distance or due to environmental factors are best modeled as a field. Fields are important for transportation research in a variety of ways, from radar signals to pollution plumes to attraction to facilities. Fields can be modeled in either raster or vector GIS.

d. Data Storage

Attribute data is any information the user assigns to a particular point, line, polygon, or grid cell. GIS stores these characteristics in a standard database format consisting of records (rows) and attribute fields (columns). It is important to note that attribute data contains more than just geographic coordinates; it includes myriad data types and characteristics about a particular spatial feature that can be stored in the database. The attribute information is viewable in aggregate by opening the data table, or individually by selecting (clicking) the specific geographic object (point, line, or polygon).

e. Joins and Merges

GIS allows a user to combine the attribute data from different databases into a single data layer. For instance, you may have two databases of airport information, each containing different attribute fields. If they share a common field, such as an airport code, the two databases can be joined (a virtual combination of two separate databases) or merged (an actual combination resulting in a single database). Any duplication or inconsistency between the databases' common field can result in inconsistent joining or merging.

B. Model Typology Classification

- *i. Function* – While GIS can be used for descriptive, explanatory, predictive, and prescriptive purposes, the function GIS itself performs—the nature of its output—is essentially descriptive. It can answer very complex and detailed descriptive questions. The analyst can interpret these descriptive answers for explanatory, predictive, and prescriptive purposes. Furthermore, when integrated with other tools such as inferential statistics, ABM, or a shortest path algorithm, GIS can also explain, predict, and prescribe, but it would be the other tools that are actually producing the explanatory, predictive, and prescriptive outputs.
- *ii. Composition* – GIS is a mathematical model.
- *iii. Mechanism* – As a descriptive database, mapping, and spatial analysis tool, GIS in and of itself does not model any processes. However, process-response models from other modeling paradigms can be fully integrated within GIS software.
- *iv. Spatial* – GIS is always spatial, and can handle every variety of spatial representation.
- *v. Temporal* – GIS can handle dynamics by having different data layers for different time periods. Animations of spatial change and change detection over time, are

common applications of GIS. Researchers are making progress towards representing time in GIS as completely as space is represented.

vi. Certainty – Most GIS data are treated deterministically, although it can be an ideal platform for various stochastic models such as random walks, probabilistic spread of disease, and countless other probabilistic simulations. One of the important frontiers of GIS research is spatial statistics. Much of the research focuses on the issue of spatial autocorrelation, which challenges the common assumption in statistical or econometric research that neighboring cases are statistically independent [45].

vii. Mathematical Relationships – Mathematical relationship within GIS can be linear or nonlinear.

viii. Resolution –Any level of aggregation or disaggregation is possible in a GIS. Most importantly, GIS provides the ability to easily change the level of resolution, depending on the resolution of the underlying data.

C. Literature Review: Typical Research Questions

Question Type 1: Where is a particular thing located, or what is at a particular location?

These kinds of simple questions are a basic starting point for GIS. It's a stretch to refer to these as "research" questions, but they are nonetheless a useful task performed by GIS.

Question Type 2: How are different phenomena spatially related?

Using the GIS *overlay* feature, two or more data layers can be visualized simultaneously. Before computers and GIS, overlaying two or more data layers required combing two paper maps. Within a GIS, each layer is its own map, and each layer can be displayed or turned off with one click. Data for further statistical analysis of the relationships among variables can be generated from a GIS.

The ability to perform spatial *queries* separates GIS from other database management systems. Similar to other database systems, GIS can query attribute information, e.g., find all entities meeting certain conditions. GIS, however, can also query based on geographic location. It can find all roads that cross a certain river, all airports of a given type within certain weather conditions and pavement types, or all airline crash sites within 5 miles of a coastline and involving greater than 10 fatalities. Standard database software may be able to find all entities within a given state, but only if there is a field within the table that defines state. A GIS can perform the same query even if the state-identifier field is absent.

Question Type 3: What areas lie within a given travel time, cost, or distance of a point, line, or polygon?

Buffering allows a user to lock a reference object (point, line, or polygon) and subsequently select other objects that lie within a predefined buffer distance, time, or cost. A similar analysis can be performed as a query; the difference is, a buffer operation creates a new data layer, whereas a query would select a subset of a preexisting data layer. The simplest buffers are based on the Euclidean distance around an object, but buffering can also be based on a field effect or on the shortest path on a network.

Question Type 4: How many potential users exist within reach of a particular transportation system?

By combining the buffering and overlay functions, GIS is uniquely well-suited to address questions of accessibility because of its ability to link point and areal data with network information. Detailed spatial data on population, employment, and location of facilities are available from a variety of local, state, and federal agencies. Land use data are available from zoning maps and remotely sensed satellite imagery. Using GIS, point or areal data on the locations of people and firms can be related to the points and lines representing transportation stations and links. Shaw [80] studied the Tri-Rail system of Broward County, Florida. He overlaid data on street networks, transit and feeder bus routes, and Tri-Rail stations with employment data for traffic analysis zones and major employers. Using travel time data for the various modes on different parts of the network, he used GIS to estimate the number of jobs within a 25-minute bus ride and 2000-foot walking distance of each Tri-Rail station.

Question Type 5: What areas have one or more zones in common?

GIS can perform a variety of *spatial operations* on two layers that are overlaid on each other. One layer can be *clipped* to match the extent of the other layer. *Intersection* or *union* operations can be performed on the polygons of two different layers to create new polygon layers that split or combine the boundaries of one layer based on those of the other layer. A *dissolve* operation aggregates features sharing the same value of a particular attribute, which could be used to create a new layer consisting of, say, all segments of a particular interstate highway.

Question Type 6: What are the characteristics of the area surrounding a place, and where are the local maxima and minima?

One needs only the "sort" command of a spreadsheet or database to find the maximum or minimum values of some variable. GIS, however, can find the *local* maxima and minima points of a raster grid of cells using n*eighborhood analysis*. Users specify the size and shape of the spatial neighborhood, such as all cells sharing a border with a cell, or all cells sharing a border or corner, or all cells such that their centroid lies within 5 Euclidean miles. By centering such a neighborhood on each and every cell, neighborhood analysis can calculate, for each cell, the maximum or minimum value for all cells within their neighborhood. By checking if the center cell's value is higher [lower] than the values of all cells in its neighborhood, GIS can determine if the cell is a local maxima [minima] of the surface. Similar analyses can be done using averages, total, or other formulae, to study clustering or other neighborhood effects. The surface could be the topographic surface of the earth, or a spatial distribution such as income or population density.

Question Type 7: What are the best estimates for the missing values of a spatial distribution of data?

Spatial interpolation is the process of filling in missing values of a spatial data layer based on the known data values for a subset of locations. GIS includes some standard functions for spatial interpolation, and provides a platform for designing customized interpolation.

Question Type 8: How have geographical phenomena changed over time?

One of the research frontiers in GIS is how to incorporate time [73]. A standard process, however, is to have data layers for the same phenomena at different times. These can then be overlaid and the change from one time to another can be detected. This *change detection* operation is particularly important for satellite imagery, where analysts commonly look for changes in the environment or land use. Change detection has many applications within transportation, such as searching for new roads, new residences, deterioration of surfaces, increase of usage, etc.

Question Type 9: Where should a new transportation facility be located?

Locating an entire system of related facilities, such as warehouses, fire stations, schools, or maintenance facilities, is a combinatorial problem that is often best handled by methods, such as mathematical programming, that can deal with billions of potential solutions. Locating a single facility within a continuous planar space, however, is often best handled by a GIS [22]. GIS tends to be well-suited to the task if the facility uses a large amount of space, such as does an airport. If the facility has many complex interactions with the surrounding physical and human environment, GIS can be very useful. Evaluation of locations of new airports or expansions of existing airports would have to include all of the following spatial issues, and many others not listed here:

- noise contours;
- surrounding land values;
- surrounding land uses;
- heights of surrounding features vis-à-vis the flight paths;
- effect on surface traffic patterns;
- environmentally sensitive areas, such as wetlands;
- flight paths of migratory birds;
- local weather and climate;
- flood or earthquake zones; and
- accessibility of population and employment;

Question Type 10: What alignment should a new transportation corridor follow to connect two places?

GIS is well suited to the task of planning a road corridor between two locations. As in the previous question, it is not, on its own, equipped to systematically search the infinite solution space consisting of all possible corridors to find the single mathematically optimal solution. Yet it has two advantages over mathematical programming for this kind of research question. First, to formulate such a problem using mixed-integer programming (see Section 6), a technique that can systematically search a huge solution space without enumerating all possible solutions, would require too many 0-1 variables for even the fastest computers to handle. Also, even with an impossibly large number of 0-1 variables, the road corridor would not be spatially smooth and accurate. Secondly, mathematical programming could not match GIS's geographic accuracy.

GIS could address this kind of problem by serving as an expert system platform for drawing, testing, and improving corridor plans. Relevant data layers might include terrain, rock and soil types, sensitive environmental zones, land uses, already built-up areas, politically inhospitable (NIMBY) areas, industrial development zones, intervening populations, rivers to be bridged, mountains to be tunneled, etc. Cost calculations and demographic potential could be automatically calculated for each plan generated.

Question Type 11: What will be the environmental impacts of a transportation project?

Given the ability to simultaneously handle multiple data layers, GIS enables its users to understand and predict the effect of a transportation project on the environment. Environmental shocks caused by a new transportation projects are too numerous to mention, but they must be considered in concert with network benefits and financial costs. GIS can account for these factors and determine if there are more optimal alternatives that better serve both the environment and the users. In one case study, Bachman et al. [6] attempted to balance the mobility needs of a growing population with environmental protection, which presents numerous policy challenges. Within a GIS, they assigned certain ambient polluting activities to certain zones within a city. They were able to spatially contrast areas of engine-start emissions with areas of prolonged enginerunning emissions using specific road segment information within a GIS. As a result, they were able to equip policy makers with tools to monitor both spatial and temporal motor vehicle emission reduction strategies.

Question Type 12: Which areas are suitable for a given activity?

Optimal location analysis is critical in transportation decision-making, and GIS can assist in defining and analyzing the parameters that will lead to specific site consideration or rejection. In the case of a proposed new runway at Minneapolis-St. Paul International Airport, planners needed to consider both flight path and radar interference from surrounding buildings and minimize noise levels in the surrounding residential areas [57]. To display the approach surface around the runways planners used 3D Analyst (a component within GIS that allows users to view spatial data in three dimensions). To assess noise pollution, GIS is used to identify the areas around the proposed runways where the noise levels will be least intrusive. Once flight-track areas that impact the fewest possible residents are identified they are matched up with runways where buildings don't run interference for aircraft or radar; only then could local planners proceed with a recommendation to the Minneapolis airport.

Question Type 13: How do the answers to other research questions depend on the units of spatial analysis?

In GIS-T, analysts use data based on preset geographic boundaries, such as states, counties, census blocks or tracts, air space, etc. Combining these can be done in several ways that lead to different zonation schemes. Other polygons may be created by the agency, such as traffic analysis zones (TAZ), in a variety of ways. Likewise, defining the size, shape, orientation, and starting point of a raster grid of cells can be done in countless different ways. It has long been recognized that definition of the zones is a potential source of error that can affect spatial studies that utilize aggregate data sources [92]. In GIS, this issue is called the *Modifiable Area Unit Problem (MAUP).* The computing power within a GIS raises the level of complexity with respect to zone assignments, yet GIS also enables a user to quickly change the zoning system (thereby aggregating the data differently) and rerunning the analysis (Miller and Shaw 2001). There is no "correct" way to partition transportation zones, but understanding how results differ depending on the zone sizes is an important concept in GIS.

D. Spatial Decision Support Systems

Davis and Elnicki [26] note that "when different people are faced with the same problem, they will adopt a range of decision-making strategies." Spatial decision support systems (SDSS) are explicitly designed to facilitate a flexible decision research environment using geographic information [27]. SDSS empowers analysts and decision-makers to:

- refine the problem definition and data, and
- generate and evaluate alternative solutions rather than a single final answer.

SDSS integrates database management system, analytical models, table and graph generators; map display, and a user interface. SDSS adds prescriptive and predictive functions to the GIS spatial data management and visualization platform. The visualization and reporting capabilities of SDSS allow decision-makers to participate more in the analysis process, while the flexibility of the software system allows analysts and decision-makers to progress towards a better problem definition and solution [18, 27].

By combining a GIS-T with the other methods reviewed in the sections to come, SDSSs can be created that are able to answer many network-related questions. SDSS applications can address overall and nodal connectivity, shortest paths, alternative routing, network flow, network efficiency, network equilibrium, location of a system of facilities, and allocation of places to service facilities. These GIS tools often originate in modeling fields outside of GIS, such as graph theory or mathematical programming, but have been incorporated into most standard GIS software packages.

An SDSS for GIS-T adds real value to network analysis when it combines transportation data with non-transportation data. Referring once again to the overlay functionality that distinguished GIS from other database analysis techniques, network analysis within a GIS can incorporate environmental, economic, and political data layers, which can improve the understanding of things like facility location or land-use interaction, and lead to a more complete grasp of the effect transportation policy has on social welfare. After all, with the increasing interaction between humans and their physical world, transportation analysis has to account for more externalities than ever before.

4. Static Graph Theory

A. Definition

Static graph theory provides a suite of tools researchers can use to analyze the topology of unchanging (static) networks. The fundamental tool of static graph theory is the concept of a network as a graph, or matrix. The matrix representing the network can be manipulated mathematically with a series of network measures [79]. Static graph theory measures networks two ways: 1) by applying a series of *full-network measures* to a matrix in order to understand the structural properties of the network, and 2) by mathematically manipulating the matrices through matrix multiplication, giving a clear picture of the network's "internal spatial structure" [87]. These network measures give a basic picture of the type of network being analyzed. Although graph theory is hardly cutting edge (very little has been added to the literature since the mid-1960s), it is such a basic measure of network structure that any comparison of methods for analyzing network topologies must begin here.

In the matrix representation of a network, rows in the matrix are origin nodes i while columns in the connectivity matrix C are the destination nodes j. An existing link between two nodes is represented by a "1" and all other cells are a "0." This matrix is referred to as a connectivity (C) matrix. Undirected networks are represented by symmetrical matrices, while directed networks (*digraphs*) are asymmetrical. In *planar* networks (such as the interstate highway system) where all edges in the metric are on the same plane, intersections of edges must be accompanied by a vertex. For example, highway intersections are (almost) always accompanied by interchanges. *Non-planar* networks (such as the airline system) allow for edges to cross without intersecting at a vertex. An example of this would be airline network where aircraft are at different altitudes during flight.

B. Model Typology Classification

- *i. Function* – Graph theory describes the structure of networks.
- *ii. Composition* – Graph theory is a mathematical model that translates a spatial network into mathematical terms.
- *iii. Mechanism* – Graph theory is a black box model because cause and effect are not modeled.
- *iv. Spatial* – Graph theory, by definition, is a network-based spatial model.
- *v. Temporal* – Most graph theory is static, although random graph theory allows additions over time.
- *vi. Certainty* – Static graph theory is deterministic; random graph theory is stochastic.
- *vii. Mathematical Relationships* – Can be linear or nonlinear.
- *viii. Resolution* – Networks are broken down into their component parts, but the output is either a disaggregated measure per node or an aggregated measure for the whole network

C. Literature Review: Research Questions

Question Type 1: Which nodes in a network are the most accessible and best connected to other nodes?

A simplistic answer to this question is provided by the *degree* of a node *i*, which is represented by:

$$\sum_{j=1}^{n} c_{ij} \tag{1}$$

and is simply the sum of the links connecting to a node. The ratio of the node degree to the total number of links gives the node's "share"of the network. Taaffe et al. [87] note, however, that putting too much emphasis on a node's degree in terms of analysis can be misleading, because the degree measures only direct linkages. Various methods have been developed to create more meaningful measures of nodal accessibility based on the following principles [87]:

1. Direct and indirect linkages: Methods should recognize that it is possible to move from one node to another even if no direct link between them exists by traveling along indirect routes via other nodes.
2. Attenuation: Methods should discount indirect routes relative to direct routes.
3. Multiple paths: Methods should give credit for the existence of alternative routes.
4. Redundancy: Methods should correct for alternative routes containing meaningless back-and-forth loops.
5. Topological distance vs. length: Methods should be able to measure connectivity to other nodes in terms of the actual length of all the arcs in a route, or in terms of the number of hops or legs.
6. Circuity: This is similar to attenuation, but for length rather than number of links.
7. Unequal node size: Methods should account for the size of nodes.

Indirect linkages and multiple paths can be modeled with a Total Accessibility, or *T*-matrix. The *T*-matrix is produced by matrix multiplication and gives a picture of how many paths exist between vertices v_i and v_j for a given number of edges in the path. Multiplying the connectivity matrix C (referred to here as C^1) by itself yields a C^2 matrix. In the C^2 matrix, the values for each O-D pair represent the number of original paths from v_i to v_j for any two-edge trip in the network. Multiplying C^2 by C^1 gives the number of original paths from any O-D pair for any 3-hop trip in the network. Multiplying the matrix in full, that is, until all connectable O-D pairs have at least one path, gives us the C^n-matrix. In order to find the *T*-matrix, we use the equation:

$$T = C^1 + C^2 + C^3 + \ldots + C^n \tag{2}$$

where n is the topological diameter of the network (the maximum number of hops between any two nodes). Summing across the rows indicates the number of ways v_n can reach all nodes in the network, but does not correct for attenuation, redundancy, length, or node size.

Garrison [35] introduced a scalar weighting system that accounts for attenuation by assigning lower weights to paths with more links. Shimbel [81] eliminated both attenuation and redundancy by generating a D-matrix that records only the number of hops in the shortest topological path between any pair of nodes. An L-matrix of shortest-path lengths between all pairs of nodes can be generated with another type of matrix manipulation; the end result is similar to an optimal shortest path algorithm. Summing across the rows of the D- and L-matrices will yield the total number of hops or the total length to move from that particular node to all other nodes in the network. Kansky [50] developed measures of circuity for individual nodes, based on the difference between Euclidean distances and network shortest paths.

Finally, the topological distances in D or lengths in L can be mathematically combined with variables representing the size of the nodes to generate measures of each node's accessibility to the distributed population in the other nodes of the network [46]. By summing across all destination nodes, one can generate a measures known as the population potential of a node, given by:

$$V_i = \sum_{j=1}^{n} P_j / d_{ij} \qquad (3)$$

where V_i is the "population potential" and P_{ij} is the population of the node.

Question Type 2: How do networks compare in terms of connectivity?

Various measures have been developed to characterize the connectivity of an entire network rather than a single node within it. These methods provide a basis for comparing networks of different sizes, modes, times, and places.

A minimally connected network, or tree, is a network with no circuits, where the removal of any edge would produce two unconnected subgraphs. The equation is given by:

$$e_{\min} = v - 1 \qquad (4)$$

where e represents the number of edges in our network, and v the number of vertices in our graph. Most of the measures that follow are based on the insight that each additional edge added to a tree network produces one additional circuit. A maximally connected network is a network where every vertex is connected to all other vertices. For an undirected nonplanar graph the equation is as follows:

$$e_{\max} = v\,(v - 1) / 2 \qquad (5)$$

For digraphs, in which arcs are not necessarily two-way, the equation becomes:

$$e_{\max} = v\,(v - 1) \qquad (6)$$

The number of edges in a maximally connected planar network is represented by the expression:

$$e_{\max} = 3(v - 2) \qquad (7)$$

The *gamma index* uses (6) or (7) to summarize how relatively connected a network is. Its value ranges from 0 to 1 (using, of course, the proper equation for maximally connected networks for planar and non-planar networks, respectively).

$$\gamma = e/e_{\max} \tag{8}$$

One of several similar measures is the *beta index*, which measures the average number of edges per vertex, and has a minimum value of zero.

$$\beta = e/v \tag{9}$$

Question Type 3: Is the network infrastructure adequate to serve the area or population?

A limitation of all of the full-network measures under Question Type 3 is that they take into account topology only, not scale or density. *Network density* is a measure of the total network length *L* (all links added together) per some base variable *B*.

$$\mathrm{ND} = \mathrm{L/B} \tag{10}$$

The default base variable *B* is the area of the region in question. Development organizations such as the World Bank often use arable land area as the base, or even a nonspatial measure such as millions of inhabitants. The World Bank frequently uses network density to compare networks across countries and to identify needs to invest in more transport infrastructure. Another measure of the scale of a network is the *eta index*:

$$\eta = \text{total network miles/ total number of links (average link length)} \tag{11}$$

Question Type 4: How are movement and economic activity related to network structure and nodal accessibility?

It is a common hypothesis that better connectivity and accessibility are associated with improved economic performance. Researchers have explored the statistical relationships between full-network measures and various national development indicators (Garrison and Marble [36, 37]). Similarly, hubbing (and therefore a higher degree of nodal accessibility) has been correlated with export-oriented development in Southeast Asian countries [17]. Bowen also illustrated the important role of the state in using transportation as an economic development tool, using the SE Asian government's competition for regional hub-status during the recent financial crisis. Chou [21] outlines the impact of deregulation in the US airline industry and economic performance, specifically noting the changes in network structure with increased hubbing and the economic performance of those hub cities. He cautions, however, that the research implies no correlation.

D. Integration with GIS.

Integrating static graph theory with a GIS opens up the opportunity for increased ease of use and data management. Networks are represented visually in a GIS as points and lines, whereas in a more standardized method of network analysis (such as typical graph analysis) use a matrix format that becomes cumbersome with large networks. Within a GIS, links and edges can easily be removed and added, and measures recalculated. Properties of sub-graphs of the larger graph can be explored. Another advantage is that GIS can produce more accurate values for distances between nodes. Finally, GIS can make it very easy to calculate network density in terms of land, arable land, or population.

5. Complex Networks

A. Definition

Complex network theory, or the "new science of networks" [9] is the study of the structure of interaction networks and their evolution based on the assumption that new linkages are formed under a set of probabilistic governing principles based on the distribution of existing linkages [97, 98]. Network formation according to these governing principles has resulted in striking similarities in network topologies found across a broad spectrum of interaction networks: protein interaction networks, the Internet [11], e-mail networks, academic citation webs [12, 63, 74], business networks [88, 99], ecological networks, and sexual and disease networks.

The term "interaction networks" is used here to refer to networks in which a linkage is defined by interaction of some kind between the two nodes. The term "interaction network" helps distinguish these networks from "hardware networks" consisting of physical connections such as roads, railways, or telecommunications wires.

Within the domain of complex network theory lie both small-world theory and scale-free network theory. It is important to note that small-world theory and scale-free theory are not one and the same, although the terms are sometimes used interchangeably. The chief underpinnings of small-world networks are 1) a small network diameter, as in "six degrees of separation" between any two people, and 2) a high degree of clustering [86, 97]. In scale-free networks, on the other hand, there are many nodes with few links and few nodes with many links—in fact the probability distribution of nodes per link follows a power-law functional form. In addition, in scale-free networks, new links are more likely to attach to nodes that already have many links—the so-called preferential attachment rule [2, 8]. These rules, it is thought, capture something fundamental in the way that interaction patterns drive network formation patterns in the real world.

B. Model Typology Classification

- *i. Function* – Complex network models can be descriptive or predictive. The descriptive functions describe the degree of clustering or the shape and steepness of the power law distribution. The predictive functions predict new node attachment.
- *ii. Composition* – Complex network theory translates the network into mathematics.

iii. Mechanism – The process of new node attachment is modeled as a black-box relationship. It is known that an input (higher node degree) usually leads to an output (new nodes attaching), but the mechanism by which this occurs is not explicit in the model.

iv. Spatial – Surprisingly, complex network theory's representation of space is topological but not geographical. The location of the nodes in longitude and latitude, their proximity to each other, and the length of the arcs are all irrelevant.

v. Temporal – The description of complex networks is static, while the prediction of new node attachment is dynamic.

vi. Certainty – The description of complex networks is deterministic, while the prediction of new node attachment is stochastic.

vii. Mathematical Relationships – Most of the relationships modeled in complex network theory are nonlinear.

viii. Resolution – Networks are broken down into their component parts, but the output is either a disaggregated measure per node or an aggregated measure for the whole network

C. Literature Review: Research Questions

Question Type 1: What are the topological characteristics of interaction networks?

Intrinsic in the structure of a complex network, as opposed to a random network, lattice network, or point-to-point network akin to Southwest Airlines, is the power-law distribution of node degree. In a static snapshot of a complex network, we find that a few nodes have a great number of incoming links, while most nodes have only a few or one. This probability distribution of link attachment is described by a power-law node-degree distribution [10]:

$$P(k) \sim k^{-\gamma} \tag{12}$$

where k is the number of links and γ is the scaling exponent. The Barabási-Albert Model gives the value of $\gamma = 2.9 \pm 1$, although networks considered to be scale free have had a range from 2.3 for movie actor networks [10] to $\sim 5 \pm 1$ for various sexual networks [56].

Question Type 2: Where are linkages most likely to be added?

The power-law distribution is explained by the way scale-free networks evolve. In scale-free network theory, the probability of links being added to a node is proportional to the number of links the node already contains. This is known as the Preferential Attachment rule:

$$\Pi(k_i) = \frac{k_i}{\sum_j k_j} \tag{13}$$

Small-world networks are characterized by a high degree of clustering. The clustering coefficient (C) is related to the propensity for two linked nodes to have a common neighbor to which both are linked. (C) can be estimated by:

$$C = \frac{3(\text{number of triangles})}{\text{total possible } k \text{ in the network}} \tag{14}$$

where triangles are three nodes each connected to one another.

Question Type 3: How many linkages are required in paths connecting nodes of the network?

Perhaps the most well-known trademark of small-world and scale-free theory is the property of small network diameter found in these types of networks. Network diameter is the path length between any two nodes in the network. Most studies of small-world networks, dating back to Stanley Milgram's "six degrees of separation" have used the *average* path length definition rather than the *maximum* path length that is more traditional in random graph theory. Regular Erdős-Rényi random graphs have an average diameter that can be approximated by:

$$d \sim \ln N \tag{15}$$

where d is the estimated average diameter of the network and N represents the number of nodes in the network. A recent paper by Cohen and Havlin [23] found that scale-free networks display behavior that is consistent with:

$$d \sim \ln(\ln N) \tag{16}$$

for networks with degree distribution $P(k) \propto k^{-\gamma}$, where $2 \leq \gamma \leq 3$. The Floyd–Warshall shortest-path algorithm can estimate the maximum path length. There is much debate about the significance of small diameter (L) in relation to small-world networks, as a small network diameter can also occur in many random networks of similar size.

Question Type 4: How vulnerable are complex networks to attack and failure?

Perhaps the most useful discovery in research about complex networks is that small-world and scale-free networks illustrate both great resiliency and vulnerability. Compared with networks with a Poisson-shaped node degree distribution, complex networks illustrate a higher resiliency to random attack and a higher vulnerability to a directed attack. Barabási [9] cites a number of instances of vulnerability when scale-free networks, including regional power-grid failures and attacks on online retailers, fell victim to a directed attack on highly connected nodes. The subsequent failures cascaded through the network, leading to near-total temporary collapse of the systems. Other researchers have noted that, had the national airline system not been shutdown by the FAA on September 11th, the nation's airline system would have ground to a halt within a few days anyway due to the closure of airports in New York and Washington, D.C.

The root of this vulnerability to failure and attack is, of course, in the degree distribution. Using simulation models, Albert et al. [9] found that eliminating the top 5% of most-connected nodes can double the network diameter. Removing the top nodes may also disconnect single nodes and smaller clusters of nodes from the rest of the network.

The same power-law degree distribution is behind the resilience of scale-free networks to attacks on *random* nodes. Due to the skewed nature of the node degrees in the network, there is a far greater chance that nodes with very few links will be attacked or experience failure in a random attack.

Question Type 5: What is the best point of attack on a complex network?

This question is very similar to Question Type 4. It simply looks at it from an offensive rather than defensive point of view. In some cases, researchers are looking to disable a complex network, such as a network of hackers; protein interactions in a cancer cell; criminal networks, etc. The approach, quite simply, has been to target the node(s) with the highest degree, i.e., the greatest number of links.

Question Type 6: What are the optimal strategies for dealing with a catastrophic failure or attack?

This research question is closely related to the previous two. The question is from the defensive standpoint after an attack or failure has occurred. Motter and Lai [66] and Motter [67] discussed the cascading failure effects, whereby an attack at a key node can propagate quickly through the network, as with a disease or computer virus. They found that an attack on a 5,000 node network simulation rendered 3,000 nodes inoperable [66]. Motter furthermore showed that the quick removal of key nodes that are topologically "close" to the source of the attack is the best strategy to prevent catastrophic failure of system. This is analogous to an act of quarantine, whereby a few nodes are sacrificed (and possibly the connectivity of the entire graph) in order to save a larger portion of nodes in disparate sub-graphs.

Question Type 7: Which transportation networks have scale-free or small-world properties, and if so, why?

Both the airline network [43] and sections of the ocean-shipping network [76] display properties consistent with scale-free topology. Guimerá et al. [43] contend that the most efficient organization for an airline network from the standpoint of minimizing lag-time between flights is a star-shaped hub pattern. Regional airline networks can be organized in a perfect hub topology, but larger airline networks overcrowd the hub. Given the limited capacity of any airport, a multiple-hub topology evolves in order to deal with the volume of flights, leading to a scale-free distribution. Guimerá et al. [43] empirically illustrated that the international airline network is scale-free. They also show that the most connected nodes (the global hubs) are not necessarily the most centrally located, where centrality is defined by the number of shortest paths that pass through a node.

Preliminary results from an ongoing investigation by [76] indicate that the network of domestic ocean shipping routes in the United States are scale-free and have a degree distribution of $P(k) \sim 3 \pm 0.1$. When broken down by vessel type (barge, container carrier, bulk carrier), the networks have a degree distribution $P(k) \sim 4.5 \pm 0.2$.

D. Integration with GIS

In complex network theory, topological connectivity of nodes matters, but geographic location of nodes does not. Software designed to make topological maps of scale-free networks, such as GraphViz, ignores the node's spatial location. Integrating scale-free network analysis with GIS has the potential to open new and important avenues for research.

For instance, by locating the nodes and links in geographic space, they can be related to other spatial data layers, such as population, topography, climate, political boundaries, rivers, military facilities, environmentally sensitive areas, and countless other phenomena. Node weights or untenable links could be determined by linking these different data sets.

A key idea is the integration of layers into the network, where various sub-graphs (such as the mobility layer, operator layer, and others from Alexandrov et al. [3]) can be separated for analysis from the all-inclusive network file representing the complete graph. Isolating these graphs provides numerous opportunities for analysis .

Links and nodes could be selected based on GIS spatial functions such as spatial intersection with certain polygon features or lying within a certain buffer around a feature. Links and nodes could also be selected based on querying the underlying database for certain attributes of the nodes and links, such as mode, carrier, national affiliation, engine type, length, size, etc.

As discussed in the static graph theory section, the standard benefits of incorporating GIS into other network-analysis methods also apply here. Data management capabilities could be applied to editing nodes and links, joining data sets, storing attributes, aggregating or disaggregating polygons, and visualizing spatial relationships.

6. Mathematical Programming

A. Definition.

This section focuses on mathematical programming approaches to transport network applications. Mathematical programming is the study of methods for defining and solving constrained optimization problems. Mathematical programming problems include linear, integer, mixed-integer, multiobjective, nonlinear, and dynamic programming problems. Linear programming (LP) is the simplest and most prominent type of mathematical programming model, and will be used as the basis of this overview. The other methods, including heuristic algorithms for solving the same kinds of problems, will be discussed as they relate to, or differ from, LP. On a broader level, mathematical programming is part of a field known variously as Operations Research (OR) or Management Science (MS) OR/MS encompasses a wide variety of quantitative decision making methods of which LP is generally the most prominent.

A linear program is composed of an objective function, constraints, and decision variables. The objective function represents a value to be maximized (such as profit or coverage of demand) or minimized (such as cost). Constraints define which solutions are feasible and which are not (a scheduling problem might be constrained by the number of aircraft available). The optimal solution is the best solution that does not violate any of the constraints.

Compared with other prescriptive models, linear programming has the advantage of always finding an optimal solution. Solving a linear program can be very computationally intensive, however. The time and memory space required generally grows exponentially with the size of the program: the number of variables (especially integer variables) and constraints. Even with the speed and power of modern computers, a large program can take days or even longer to solve, or may not solve at all. To overcome this hurdle, researchers often develop heuristic algorithms, based on techniques such as Lagrangean relaxation [25, 75], out-of-kilter algorithms [25], greedy algorithms [25], genetic algorithms [29], tabu search [78, 82], or simulated annealing [33, 53]. While heuristics are not guaranteed to be optimal, they often can find optimal or near-optimal solutions to far larger problems, or solve the same size problems far more quickly.

B. Model Typology Classification

i. Function – Linear programs are always prescriptive, they are designed to help choose a course of action. While the description of a LP problem given above has only a single objective, LP can also be used to solve *multiple objective* problems. Generally, two (or more) objective functions are defined and combined into one by giving them different weights, or converting all but one into constraints. Solving the problem with a succession of different weights or different constraint limits can define a series of Pareto-optimal solutions (solutions for which neither objective can be increased without decreasing the other). One important note—while mathematical programs are always prescriptive, heuristic algorithms can seek an equilibrium rather than an optimum.

ii. Composition – All linear programs are mathematical models.

iii. Mechanism – Mathematical programs can be process-response, black box or hybrid, depending on the intended use, available data and the complexity of the model.

iv. Spatial – Mathematical programs can treat locations or areas as discrete sites or as a network. Heuristic algorithms are better suited for dealing with continuous planar space because they do not necessarily require an input of all possible locations before the model is solved, as does a linear program. Most models applicable to the airline network are, of course, network models, though there are a few exceptions.

v. Temporal –Mathematical programs can be either static or dynamic. Dynamic models usually use discrete timescales. Continuous dynamic problems, in which events can occur at any moment in time, are notoriously difficult to model using LP. Dynamic problems can also be solved with dynamic programming, a method specifically designed to handle decisions that are linked through time. In dynamic programming problems, the problem is divided into stages, with decisions required at each stage. Each stage has a number of states associated with it. *Dynamic programming* makes problems tractable by making the decisions at a given stage dependent only on the current state, not on the previous states or previous decisions. The approach is to solve a multi-period problem by breaking it up into smaller problems that are solved recursively forward or backward.

vi. Certainty – Linear programs are deterministic. The coefficients within the objective function and constraints must be given. There are, however, ways to incorporate probability into deterministic linear programs. *Chance-constrained programming*, for instance, aims to satisfy constraints a certain percentage of the time [20]. The probability function on which this is based, however, is evaluated outside of the linear program, and a deterministic value corresponding to that probability is then input into the LP. Other probabilistic LP models may weight outcomes depending on their likelihood. Heuristic models, on the other hand, can be stochastic in the pure sense of not yielding the same solution every time. *Queueing theory* can be incorporated into models that have come to be known as "simulation-optimization," in which a master program iterates between an optimization module that chooses values of decision variables, and a simulation module that evaluates the probable performance of those decisions [15, 34].

vii. Mathematical Relationships – Implicit in the term, "linear programming" is that all of these functions must be linear. Neither the objective nor any of the constraints can multiply one variable by another or use exponents. There is a related category of *nonlinear programming* that allows nonlinear objectives and constraints. Nonlinear programs are more difficult to solve than linear ones; many of the techniques used to solve linear programs are not applicable. *Integer programming* is a subset of LP. In many problems, one or more of the variables can only logically take on integer values; one cannot build a fraction of a new runway or a fraction of an aircraft and then use that fraction of its capacity. Integer programs (IP) confine their variables to integer values, while mixed integer programs (MIP) use both continuous and integer variables. Ensuring that you find the best all-integer solution can be quite difficult, and generally adds considerably to the processing time. Heuristic methods are often employed for this reason. Programs that restrict a variable's value to 0 or 1 (Boolean or binary variables) are a subset of IP.

viii. Resolution –The resolution of a linear program can be either aggregate or individual, depending on the model. While most hubbing models are aggregate, with a single point representing all travel demand for an entire region, crew-scheduling models generally deal with an individual or very small group.

C. Literature Review: Research Questions

Question Type 1: Where should facilities be located?

Mathematical programming is widely used to optimally locate systems of facilities. Whereas GIS may be well suited to locating a single facility, math programming is better suited to combinatorial problems in which one must locate p facilities out of n nodes. Many location problems are defined to locate facilities either at nodes of a transportation network, or anywhere along the arcs of a network. We refer readers to Daskin [25] or similar texts for the range of possible problems, many of which could be applied to the airline industry.

Question Type 2: How should (or will) traffic flow over a network?

This question is answered by traffic assignment models. The usual inputs are a trip table consisting of the flow demand between all origin-destination pairs, and relationships defining the cost (or travel time) and capacities of different arcs of the network. There are two main versions of this problem. If the traffic flows are dictated by a single managing entity, it is usually formulated as a cost-minimization model. If independent network users (such as drivers) determine the traffic flows, the problem is usually approached as a static equilibrium problem [58]. The usual "Wardropian" equilibrium condition is that no user can reduce his cost or time of travel by switching routes. While this is a predictive rather than a prescriptive problem, the heuristic solution methods bear much in common with other mathematical programming approaches.

Question Type 3: Which links should be added to a network?

This problem, known generally as the *network-design* problem, results from combining a traffic assignment model with a facility location model in which the facilities are new or expanded links. A network design model usually minimizes the combined costs of building and using a network. Magnanti and Wong [61] provide a comprehensive, if somewhat dated, review of the network design literature.

Question Type 4: What is the optimal design for hub-and-spoke network? What is the optimal location of airline hubs?

The location and assignment of hubs in a hub-and-spoke system is a topic of great interest to airlines and overnight delivery companies. Facility location and network design can be treated as two separate questions (see above), but they are often intertwined. Hub location and network design have been the topic of a great number of OR papers. Work on these problems actually predates the adoption of the hub-and-spoke model for air transport under the guise of locating warehouses or other centralized distribution facilities.

One of the aspects of airline/overnight delivery hub-and-spoke networks that makes them more complex than warehousing problems is the interaction between multiple hubs. O'Kelly [69] formulated a general form of the multiple-hub location problem as a non-linear quadratic integer program (QIP). The practical difficulties involved in solving a QIP led Klincewicz [59] and Aykin [5] to present various heuristic algorithms for solving the multiple-hub problem. Linear versions of the multi-hub location problem, which are easier to solve, have been put forward by Cambell [19] as well as Sohn and Park [83] who derived a linear formulation of O'Kelly's QIP.

O'Kelly and Miller [70] recognized that the hubbing problem could be divided into four sub-problems: (1) locating the hub(s); (2) allocating nodes to hubs; (3) determining inter-hub linkages; and (4) routing traffic though the network. Since these four sub-problems are interdependent, a global optimum solution would require solving them simultaneously, which can be difficult and time-consuming for large networks.

Many researchers have concentrated on the second sub-problem. O'Kelly and Lao [71] developed a model for allocating nodes to hubs in a multi-modal express delivery system. Sohn and Park [83] examine both single (hubs are allocated exclusively to a single node) and multiple (hubs are allocated to multiple nodes depending on the origin/destination of the flow) allocation problems. Kuby and Gray [55] developed a model for a single hub network that allows stopovers and feeder service. Barnhart and Schneur [13] developed one of the most comprehensive treatments of assignment and schedule design for a hub-and-spoke system in the OR literature, including stopovers, ground and air feeder service, hub capacity constraints, and multiple aircraft types. A large carrier has adopted this model and integrated it into a decision support system. Models developed with practical applications in mind generally emphasize short solution times in the face of very large problems. The formulation described by Armacost, Barnhart and Ware [4], for instance, can optimize UPS's next day air network in less than two hours.

Although most hub location problems are designed to serve a fixed demand at minimum cost, other objectives are not unknown. Dobson and Lederer [30], for instance, propose a LP-based heuristic that maximizes profit in a competitive environment. Marianov, Serra and ReVelle [62] have a similar model that maximizes an airline's traffic levels rather than profit.

Question Type 5: Which type of network structure is best, and what parameters affect the optimal choice?

With the rise of Southwest Airlines and its greater use of a point-to-point network structure, an obvious question that arises is whether hub-and-spokes or point-to-point (direct) is the best structure. Yang and Kornfeld [104] developed a mixed-integer programming model to design an air freight network, which allowed them to compare hub-and-spokes with other network designs. See also Lederer and Nambimadom [59] for a closed-form analytic model (not a mathematical programming model) for calculating the costs for hub networks, direct networks, a complete tour, and a set of subtours, and for testing the sensitivity of the optimal choice to several parameters of the network, such as the number of cities, the demand, and the distances among cities.

Question Type 6: What is the optimal schedule?

While hub location and network design have been the most popular areas of airline-related LP research, aircraft scheduling problems have also attracted attention of the field. Balakrishnan, Chien and Wong [7] present a model for long-haul airline scheduling. Their model attempts to maximize profit in a network of intermediate stops between predefined origin and destination airports. Teodorović [89] proposes a model for designing a schedule that minimizes total cost while minimizing the likelihood of weather-related delays. Delay costs are calculated using the probability of a delay at a given airport, based on historical data.

Question Type 7: What is the optimum way to recover from a schedule perturbation?

Teodorović and Guberinić [90] developed a nonlinear integer program to minimize the total delay in a schedule after an aircraft becomes unexpectedly unavailable. Since the model is nonlinear, it is difficult to solve quickly. In order to be useful in practice, a schedule perturbation model has to be solvable in a short amount of time. Yan and Tu [102] present a perturbation recovery model that can be solved in a short time for a realistically large problem. This model uses a network to represent time as well as space.

Question Type 8: What is the optimal crew schedule?

Given the complicated work rules that govern airline flight crews, creating a crew schedule that serves all of an airline's flights at the minimum cost is an extremely complex task. Yan and Chang [101] developed an LP model to optimally schedule cockpit crews while Yan and Tu [103] presented a similar model for scheduling cabin crews. Like the Yan and Tu [102] schedule perturbation model, these models use a space-time network to represent both the spatial and temporal dimensions of scheduling.

Question Type 9: What are the optimal flight and crew routings and schedule for an on-demand airline?

Operators of on-demand air-taxi or shared-ownership systems face a complex daily planning problem. Customers request transport between a pair of airports for a given number of people within a particular time window, sometimes only a day ahead. Aircraft and crews must be allocated to these flights, and then reallocated to the next trip from there or somewhere else, or to rest or maintenance. Aircraft must be serviced periodically at maintenance facilities, while crews must receive adequate downtime each day while on the road and at prescribed intervals in their home town. Aircraft should spend as much time as possible on revenue-generating flights. This is the problem faced by Bombardier Aerospace, manufacturer and operator of the FlexJet® on-demand air transport system [48]. Bombardier was named a Finalist for the 2004 Franz Edelman Award for Management Science Achievement, given annually by the Institute for Operations Research and Management Science (INFORMS) for the modeling project with the greatest impacts on its client. Numerous airlines have been Finalists in the Edelman competition (see *Interfaces'* annual Edelman Award issue), but this is the first Finalist from an on-demand airline.

Question Type 10: What are the tradeoffs between multiple objectives of an optimization problem?

Multiobjective optimization problems are those in which there are two or more objectives measured in different units ("apples and oranges"), and no agreed-upon conversion factor exists to convert all criteria into a single metric [24]. In these cases, multiobjective programming provides a set of tools to identify solutions that perform as well as possible on both objectives, and quantify the tradeoffs among the objectives. Multiobjective programming identifies solutions that are *Pareto-optimal*. Technically, a solution is Pareto-optimal if there is no other solution that performs at least as well on every criteria and strictly better on at least one criterion. That is, a Pareto-optimal solution cannot be improved upon without hurting at least one of the criteria. Solutions that are Pareto-optimal are also known in various literatures as *nondominated*, *noninferior*, or *Pareto-efficient.* A solution is *not* Pareto-optimal if one criterion can be improved without degrading any others. These solutions are known as *dominated* or *inferior* solutions. Multiobjective programming does not recommend any single solution as the best overall or best compromise, but it eliminates the dominated solutions and quantifies the tradeoffs involved between the nondominated solutions. A multiobjective model for air transport networks, developed by Flynn and Ratick [32], selects routes to be included in the Essential Air Services program so as to minimizes costs and maximizes coverage of population.

D. Integration with GIS

GIS is primarily used with mathematical programming as an input-output tool. GIS can be used to create networks and process data for use in formulations as well as allowing solutions to be presented visually, rather than in tables and text. GIS with integrated Management Science decision-making tools, including LP, is a type of Spatial Decision Support System (SDSS). Because GIS software can incorporate nonspatial databases, GIS can serve as a bridge between a corporation's or agency's Operations Research department and its other information systems.

Before LP models can be used to optimize decisions in the real world, they must be calibrated to ensure that they produce realistic results. Accurate data on supply, demand, cost, and capacity are essential to accurate models. The aforementioned ability of GIS to incorporate non-GIS databases can allow detailed information to be processed for the LP inputs.

Often, the key to calibration is greater spatial detail. In many transport systems, supply and demand are continuous spatial phenomena, with origins and destinations spread continuously over the landscape. LP models require discrete origins and destinations, which are created by aggregating all of the supply or demand within a polygon and representing that polygon by a centroid. Often, there are an infinite number of ways that continuous space can be subdivided into polygons. Optimal solutions often depend on the way in which space was subdivided. This is commonly known as the Modifiable Areal Unit Problem, and it is not confined to LP applications. The key point, however, is that integrating an LP model with a GIS can enable operations researchers to easily change the areal units when calibrating a model, so that the spatial allocations of each node is more realistic for the whole polygon represented by the node. Miller [64] touts GIS's "unrealized potential for more sophisticated geometric representation in facility location problems," and provides an integrated modeling framework for realizing this potential.

7 . Simulation

A. Definition

A simulation model represents the form or process of a system and changes in its state over time. Simulation goes beyond most forms of mathematical modeling in that it represents all the major elements of the system, not just its inputs and outputs. There are numerous types of simulation models, but in this paper we focus on stochastic, queuing-type simulation models, which are widely used for transportation.[1]

[1]Another class of simulation model is known by its practitioners as *systems dynamics* models, but might be better classified as deterministic, aggregated stock-flow models to distinguish them from stochastic, disaggregated queuing models, which also model dynamic systems. Systems dynamics models are commonly used in economics, ecology, and demography to model the evolution of systems consisting of resource stocks and flows, positive and negative

At the most basic level, queuing-type simulation models are built of entities, resources, and queues. In essence, resources act, while entities are acted upon. If the entity is a passenger, the resources might include check-in counters or security checkpoints. If the entity is an aircraft, resources could be gates, taxiways, runways or airspace. Queues are "lines" of entities, such as people waiting at a security checkpoint or aircraft in the landing pattern. While they need not be physically lined up, they are a set of entities awaiting use of a resource that have some sort of specified service order (first-in-first-out, last-in-first-out, some sort of priority system, or even random choice). While not all transport systems have queues, they are a very common feature, if only because most transport systems will, at some point, have more entities to service than they have available resources.

In this paper, we make a distinction between what we call "systems-level simulations" and "agent-based simulations." This section focuses on the former, while the following section focuses on the latter. Both types of simulation models meet the definition given in the first paragraph. The difference between them is that, in systems-level simulations, the system acts on the entities, whereas in an agent-based model, the entities (agents) act autonomously. Put differently, in the systems-level simulations, there is one set of rules (however complicated) governing the system, whereas in an agent-based simulation, different types of agents can have different sets of rules.

It is important to differentiate between a simulation and a simulator. Both involve representations of a system, but they differ in purpose. The primary objective of a simulation is to represent the system and to test it under various conditions. The primary goal of a simulator is to provide experience and training to an operator of the system. A simulation of the air traffic control system could be used to test changes in the system or its operating policies, while an air traffic control system simulator would be used to train air traffic controllers.

B. Model Typology Classification

- *i. Purpose* – Simulation models are generally predictive, though they can also be used for explanatory or prescriptive purposes. When used predictively, they can test a system's design, and its response to various conditions or operating policies. Simulations can also provide explanations of system behavior, especially in systems too complex to examine without the help of a model. Finally, simulations are also used to test the output of other, prescriptive methods, including optimization, expert judgment, or simple trial-and-error methods.
- *ii. Composition* – Most simulation models are mathematical, but some are physical (e.g., wave tank or wind tunnel) or analog (e.g., entities such as planes or cars represented by electric pulses on a specially designed chip).
- *iii. Mechanism* – Simulations are process-response models. However, simulation models often incorporate black box submodels to represent particular components of the system whose internal functions are not germane to the system as a whole.

---

feedback loops, and delays. Resource stocks and flows are usually modeled as aggregates of individuals and places, and numbers are usually treated as deterministic.

iv. Spatial – Simulation models can be spatial or nonspatial. Spatial simulations can treat space linearly, as a network, in two dimensions or in three dimensions.

v. Temporal – Simulation models are always dynamic. Most simulations use some form of discrete time. Some use regular time intervals, generating a new state of the system at regular intervals. Others use intermittent, event-driven timescales, where the simulation advances to the next random or scheduled event.

vi. Certainty – Queuing-type simulation models are generally stochastic, and are often run multiple times to determine typical outcomes.

vii. Mathematical Relationships – Simulation models can be either linear or nonlinear. They are generally better at handling nonlinearity than other methods, such as mathematical programming.

viii. Resolution – Queuing-type simulation models are usually disaggregate, with each entity representing something that cannot be easily divided into smaller units (such as a person or aircraft).

C. Literature Review: Research Questions

As a general rule, simulation models are used to answer what-if questions.

Question Type 1: What if an airline's schedule is disrupted?

An example in the air transport field is what-if questions about schedule disruptions. Rosenberger et al. [77] developed the SIMAIR model to represent an airline's daily operations. The model stochastically simulates events like weather delays and unscheduled maintenance, allowing users to see what effects these events can have on a schedule. Lee et al. [60] present a revised SIMAIR model, which includes a more modular design. In particular, it incorporates a better post-disruption recovery plan, based in part on OR models such as Teodorović and Guberinić [90]. The modular design also allows airlines to substitute their own recovery plans.

Question Type 2: How will airport capacity be affected if the airside facilities are changed?

Stamatopoulos, Zografos and Odoni [85] present a decision support system for use in airport strategic planning. Their system incorporates several simulation models of airside systems, including runway and apron capacity models. The runway model provides for a variety of configurations, including single, parallel dependent, parallel independent and intersecting runways (configurations with more than two runways can be simulated by combinations of these models). The system incorporates a black-box weather model, or it can "re-play" historic weather conditions.

D. Integration with GIS

Queuing-type simulation models are often integrated into decision support systems (DSS), the non-spatial progenitors of SDSS. Indeed, all of the simulation models referenced above were built as decision support system components. Using GIS to add a spatial context to these systems is an obvious change, especially for transportation system models, which have inherent spatial aspects to begin with.

GIS could be used to track the spatial location of entities with greater precision and incorporate data about their surroundings into the simulation. Similarly resources could also be placed in their spatial context. Perhaps the greatest benefit of integrating spatial data with simulation models could be in more accurate queuing. For instance the model by Stamatopoulos, Zografos and Odoni [85] described above includes a queue between the runway and apron, where planes wait if no apron space is immediately available. If this queue is full, the model allows no further landings until space is available in the buffer. Realistically the space available might depend on where the aircraft exits the runway, which apron stand it is eventually destined for, and the size of each aircraft. All of these are spatial characteristics which could be modeled in a GIS as part of an SDSS.

8. Agent-Based Modeling

A. Definition

Agent-Based Modeling (ABM) is a new simulation modeling paradigm that focuses on modeling disaggregated activities and decisions, rather than modeling the system as a whole. Often called a bottom-up approach, the crux of ABM is that a group of entities (the "agents") behave according to certain rules in the simulation. The agents in the model can be constructed hierarchically such that agents not only act on their own accord, but sets of agents contain common characteristics. An advantage of ABM is that the agents can learn during each iteration, changing their behavior throughout the simulation, hopefully producing a more realistic result [39].

Bonabeau [16] asserts that "ABM is a mindset more than a technology." The mindset of ABM is to model individual agents' behavior, and then let that behavior play out in a simulation that yields the aggregate net results of their interactions. Qualitatively, ABM illustrates how complex, nonlinear phenomena can be replicated from seemingly simple rules and actions by autonomous agents. ABM has been championed in the social sciences [14] for its real-world applicability, based on its ability to take what we know about humans and their preferences, incorporate these preferences into agents, and simulate how agents of different classes will interact. ABMs are well-suited for studying evolutionary processes based on individual behaviors.

B. Model Typology Classification

i. Purpose – Agent-based models themselves are predictive. They predict the aggregate state of a system consisting of multiple interacting agents at any given time. Like other simulation models, however, their results can be interpreted or applied for explanatory or prescriptive purposes. ABMs can also provide explanations of what kinds of individual decision rules produce which kinds of aggregate system results.

ii. *Composition* – Although nearly all agent-based models now are mathematical, some early models were done using human subjects who were instructed to follow certain rules.

iii. *Mechanism* – ABMs are process-response models, in that the processes through which a set of inputs (agents with rules, resource limits, etc.) produce a set of outputs (aggregate results, system performance) are explicitly represented in the model. Agent's interactions with each other and their environment are explicitly modeled in an ABM. In fact, ABMs not only model agent interactions, but also explicitly model the decision process within each agent. Furthermore, agents in ABMs can be programmed to "learn," based on feedback of previous agent experience. As such ABMs explicitly incorporate three different kinds of processes, including agent decision making, agents changing their decision making rules (learning), and agents interacting with other agents and their environment.

iv. *Spatial* – Agent-based models can be spatial or non-spatial. Spatial ABMs can treat space linearly, as a network, as a regular grid of cells, or as planar surface with irregular parcels.

v. *Temporal* – ABMs are always dynamic, and typically generate a new state of the system at regular time intervals.

vi. *Certainty* – ABMs can be deterministic or stochastic, depending on whether an agent's decision rules dictate only one decision for any given set of conditions (deterministic), or a probability of making various decisions for the given conditions.

vii. *Mathematical Relationships* – Agent-based models can be either linear or nonlinear. They are generally better at handling nonlinearity than other methods, such as mathematical programming.

viii. *Resolution* – ABMs are by definition disaggregate. Individual agents are explicitly modeled as independent entities.

C. Literature Review: Research Questions

Question Type 1: How do vehicle operators interact on a network?

Agent-based models are particularly well suited to understanding how agents will interact in a given situation. Traffic flows, traffic congestion, and traffic accidents are the net result of the interactions of vehicle operators on a network. Peeta et al. [72] illustrate this technique for car and truck interactions, based on drivers' characteristics and the propensity for a driver (of either a truck or a car) to become "uncomfortable" in situations encountered on a freeway. The situational factors that affect a driver's discomfort in normal and high-risk situations, for drivers with different socioeconomic characteristics, were based on a stated preference survey. These were extrapolated into propensities or rules governing when a driver will become "uncomfortable" and how they will respond to various car-truck interactions on a highway. The model was run using one-second time intervals with vehicle speeds based on uniform probability distributions between 60 and 70 mph for cars and 55 to 65 mph for trucks. Simulations were run multiple times to get average behavior patterns.

Several ABM studies (e.g., Bonabeau [16]) have outlined the effects of human behavior at bottlenecks, and what ABM has to offer for problems of this nature. One application to transport networks is to model traffic congestion resulting from highway accidents. Another application could be to the loading and unloading of airline passengers and their carry-on luggage. Assigning agents functions of speed, responsiveness, and comfort in crowded situations with limited storage space could offer an appraisal of different embarkation/debarkation strategies. The time may be particularly ripe for this kind of study, given the imminent introduction of the A380, which will carry so many passengers that standard single door/jetway embarkation/debarkation procedures will be exceedingly time consuming.

Question Type 2: What types of trips will be made by which kinds of people at which times?

Neidringhaus [68] has used ABMs to estimate demand for air transport. Using passenger preference surveys, he developed set of trip-making rules for different types of passengers (business, pleasure, or otherwise). He then used an ABM to model the interactions among these different types of passengers in bidding for limited airline seats at different times. The results of such a model could be used to improve airline pricing schemes. Demand for on-demand, disaggregated air transport could be estimated using this method.

Question Type 3: How do people choose their routes to travel on a network?

With "learning" agents, passengers can learn from previous bad experiences in making certain choices. Positive experiences will cause the agent to repeat previous choices, while negative feedback will cause the agent to make a different choice. Dia [28] used ABM to model how drivers select their routes given real-time information about traffic. Similar models could be developed for air passenger routing choices, based on previous experiences with airline delays and limited information about alternative routes.

Question Type 4: How do different sets of agents of different classes within the air transport system interact with each other?

Different individual agents can be combined together to form higher-order classes in an ABM, resulting in a hierarchy of agents in the model. For example, agent classes could include coach passenger class, first-class passengers, frequent flyers, elite frequent flyers, pilots, crew, ground staff, management, and union organizers. These could be combined in a variety of ABMs to answer a wide range of research questions. Gaupp and Hill [38] used ABM to study air force pilot retention. A similar model might shed some light on the ongoing negotiations among Delta Airlines, its pilots, and the pilots union regarding the pension funds, early retirements, and other issues. A hierarchical model could be applied to the process of obtaining volunteers to give up their seats on overbooked flights.

Question Type 5: How do land use and transportation systems co-evolve?

Agent-based modeling has emerged as the leading method for predicting the co-evolution of linked land use and transportation systems. Models such as UrbanSim are used by metropolitan planning organizations to forecast future urban growth. In UrbanSim, households and businesses make residential location choices, developers choose the locations, density, uses, and scale of developments, and governments decides where to build transportation links and other infrastructure investments, and set a variety of other policies. Land use-transportation ABMs produce forecasts for urban growth, transport network evolution, employment and population, land use plans, and policy change. Land use-transportation modeling is most highly developed for road transport: other modes can also have a considerable influence on land use that could be represented using ABM.

D. Integration with GIS

The combination of GIS and ABM is potentially very powerful [41]. The key benefit of combining the two is the ability to model spatially explicit agents. Agents can represent spatial units, such as owners of land parcels. Alternatively, the agents can be mobile but located at a particular location in space at a particular time. In either case, the marriage of ABM and GIS enables agents to interact with each other and their environment based on their location.

Various kinds of spatially explicit data in a GIS can add value to agent-based models:

- Transport demand is highly spatially dependent. An ABM can base the characteristics of its passenger agents on the socioeconomic characteristics of the spatial hinterlands of specific airports in the region.
- Weather phenomena are inherently spatial. An ABM application dealing with pilot responses to inclement weather could use GIS to represent the spatial aspects of weather phenomena with greater accuracy.
- The attractiveness of different places is an important variable for airline crew placement.
- Land use is obviously dependent on surrounding land uses, proximity to transport, physical characteristics of the land, and local/regional/national political conditions, among other factors.
- Because agent behavior will tend to be more influenced by nearby agents than more distant agents, behavioral changes will tend to have a spatial component.

The key attribute of GIS in all of these situations is management of spatial data and the ability to link agent behavior to their location and the various attributes of that location, which will yield more accurate results from the simulation.

9. Conclusions

Table 1 summarizes the preliminary findings of this report. It classifies each kind of model according to the dimensions defined in Chapter 2. In particular, we have attempted to distinguish between what a model actually does (its function) and how researchers use and interpret it. Table 1 also lists some types of network-related research questions that have been addressed by each kind of model. The lists of models, research questions, and references are provided as examples, and are not meant to be comprehensive or exhaustive.

Transportation-related research has been enhanced by the use of GIS, which enables its users to assimilate a variety of data types for simultaneous viewing and analysis. Blending various types of data together is critical for transportation research that must account for the associated socioeconomic and environmental factors. Transportation infrastructure decision-making demands that both network and non-network related data be considered. Furthermore, as the price tag of transportation network infrastructure escalates, spatial decision support systems, which integrate GIS spatial data management and visualization functions with predictive and prescriptive analytical models, provide a flexible decision-making environment to balance the social, financial, and environmental costs with overall network efficiency. GIS is improving safety, efficiency, and information flow to enrich transportation network topology analysis.

As Appendix 1 shows, there are an enormous variety of data layers that are related to different parts of the air transport system. The air transport system, like any other system, consists of the components of the system itself as well as its environment. In the case of air transport, the components exist at scales from local to global, while the environment consists of biological, geological, demographic, political, and economic layers. Many of these layers will come into play in evaluating some aspect or another of on-demand air transport networks, or some new business model not yet envisioned. By developing a spatially referenced GIS database, we would develop a framework for relating any part of the system or the environment to any other. GIS would also provide a platform for supporting almost any type of model and any type of research question by providing spatially accurate inputs and powerful visualization and data management capabilities.

Complex network theory has burst upon the scientific scene over the last five years, and there are reasons to believe it may offer insights for current and future air transport networks. A wide variety of interaction networks appear to share the same scale-free and small-world properties. At the continental scale, the entire passenger air transport network displays scale-free properties. Individual hub-and-spokes airlines, however, do not. Airlines employing the hub-and-spokes business model are threatened by bankruptcies that may be partly due to their network structure. Other airlines that use more of a point-to-point network and rely on less-congested airports on the metropolitan fringes have been gaining market share. Newer business models being pioneered by NetJets® and FlexJet®, based on distributed, on-demand flights by smaller aircraft, are challenging both the hub-and-spoke and point-to-point networks of the major airlines. There is a need for researchers to examine which network topology most closely fits the scale-free assumptions at the carrier, country, or continental levels of aggregation. It is also not yet understood if small network diameter and scale-free properties are desirable characteristics in an airline network.

On-demand air transport, which is most likely to have scale-free and small-world properties, offers several advantages and disadvantages. The smaller aircraft are able to utilize thousands of general-aviation airports across the North American landscape, offering a solution to the looming airport-capacity crisis. With cities in North America sprawling over enormous areas, utilization of a number of general-aviation airports offers greater convenience in terms of direct flights between points much closer to the ultimate origins and destinations, at the exact times the travelers wish to travel. These advantages, however, must be balanced against the loss of economies of scale at larger airports and on larger aircraft. Safety may also be a concern, with thousands of additional aircraft crowding the skies.

There are a number of challenges to overcome before small-world and scale-free network theory can be practically applied to planning or operating an on-demand airline network. First, capacities have yet to be incorporated in complex network theory. Most networks investigated so far have been uncapacitated. There are no limits on the number of hyperlinks to Google® or the number of times an academic paper can be cited. Airports, aircraft, and airspace, on the other hand, have limited capacities. Given that demand for air travel is expected to triple, any planning model must be able to deal with capacity constraints. The hub sites with the greatest room for expansion are not necessarily the places where airlines would like to add more flights.

Second, scale-free network theory is descriptive and predictive. It has prescriptive implications for vulnerability analysis, but it is not clear how it can be used for planning a cost-efficient airline network. One of the most important research papers cited in this report is the paper by Hicks et al. [48] describing how the on-demand airline FlexJet® is using mathematical programming, not complex networks, to plan its flight and crew schedules for the next day. Their study was a Finalist for the prestigious Edelman Award given by the Institute for Operations Research and Management Science (INFORMS) in 2004.

What then can complex network theory contribute to planning efficient airline networks? One possibility might be to use it to predict traffic flows for on-demand airlines. Forecasted traffic flows are a necessary input to any prescriptive modeling. Complex network theory is able to describe and predict the existence of interaction links between nodes. The problem we foresee for forecasting traffic flows, however, is that the existence of an interaction link between two nodes implies only that the two nodes have interacted at some time, not that they will interact on a particular day. With on-demand airlines, the demand will be different every day, and as a result, the routings of aircraft and crews will be different every day. Over an extended period of time, complex network theory might be able to predict whether interaction will occur, but not on a daily basis. Nor is it equipped to predict the *volume* of interactions, which is necessary for planning networks.

The most widely used model for explaining and predicting flow volumes is the gravity model (Howrey [49]).[2] Decades of research into gravity-type models show that flow volumes are usually proportional to the product of the two nodal populations and inversely proportional to the distance or impedance between them:

$$I_{ij} = k\frac{P_i^{\alpha} P_j^{\gamma}}{d_{ij}^{\beta}} \tag{17}$$

where I_{ij} is the flow volume from i to j; P_i and P_j are the populations of i to j (or some other measures of their trip-generating and trip-attracting strength); d_{ij} is the distance (or other impedance factor) between i and j; k is a fitted scaling constant; and α, β, and γ are fitted exponents. Statistical analysis of gravity-type models has typically found R^2 values greater than

[2] Reviewing the gravity model was beyond the scope of this paper because the gravity model deals with spatial flows, whereas this paper deals with network topology.

0.8, and that node size and distance are highly significant variables. Although the gravity model's ability to predict short-distance air travel is problematic because of people's tendency to switch modes, it still seems unlikely that complex network theory, without using node size or link distance, could match the gravity model's predictive fidelity. A potential direction for future research might therefore be to modify scale-free and small-world network models to incorporate distance and/or node size and to predict not only the existence of links but also the volume of interactions. GIS represents the best platform for this effort, thanks to its spatial capabilities, including calculating distance, managing data, performing calculations, visualizing where the model fits well and where it does not, and flexibly improving the method.

References

[1] Albert, Réka, Hawoong Jeong, and Albert-László Barabási. 2000. Error and attack tolerance of complex networks. *Nature*, 406: 378-382.

[2] Albert, Réka, and Albert-László Barabási. 2002. Statistical mechanics of complex networks. *Reviews of Modern Physics*, 74: 47-97.

[3] Alexandrov, Natalia, ed. 2004. *Transportation Network Topologies,* NASA/TM-2004-213259, October 2004.

[4] Armacost, A. P., C. Barnhart, and K. A. Ware. 2002. Composite variable formulations for express shipment service network design. *Transportation Science* Vol. 36, No. 1: 1-20.

[5] Aykin, T. 1994. Lagrangian relaxation based approaches to capacitated hub-and-spoke network design problem. *European Journal of Operational Research* 79:501-523.

[6] Bachman, W., S. Wayne, S. Hallmark, and R. Guensler. 2000. Modeling regional mobile source emissions in a geographic information system framework. *Transportation Research Part C*. 8 (2000) 205-229.

[7] Balakrishnan, A., T. W. Chien and R. T. Wong. 1990. Selecting aircraft routes for long-haul operations: a formulation and solution method. *Transportation Research-B* Vol. 24B, No. 1:52-72.

[8] Barabási, Albert-László and Bonabeau, Eric. 2003. Scale-free methods. *Scientific American* (May): 60-69.

[9] Barabási, Albert-László. 2002. *Linked: the new science of networks*. New York: Plume.

[10] Barabási, Albert-László, and Réka Albert. 1999. Emergence of scaling in random networks. *Science*, 286: 509.

[11] Barabási, Albert-László, Réka Albert, and Hawoong Jeong. 2000. Scale-free characteristics of random networks: the topology of the world-wide web. *Physica A*, 281: 69-77.

[12] Barabási, A.L., H. Jeong, Z. Néda, E. Ravasz, A. Schubert, and T. Vicsek. 2002. Evolution of the social network of scientific collaborations. *Physica A*, 311: 590-614.

[13] Barnhart, C., and R. R. Schneur. 1996. Air network design for express shipment service. *Operations Research* Vol. 44, No. 6: 852-863.

[14] Berry, Brian J. L., L. Douglas Kiel, and Euel Elliott. 2002. Adaptive agents, intelligence, and emergent human organization: Capturing complexity through agent-based modeling. *Proceedings of the National Academy of Sciences*, 99: 7187-7188.

[15] Boesel, J., B. L. Nelson, and N. Ishii. 2003. A framework for simulation optimization. *IIE Transactions* 35: 221-235.

[16] Bonabeau, Eric. 2002. Agent-based modeling: methods and techniques for simulating human systems. *Proceedings of the National Academy of Sciences*, 99: 7280-7287.

[17] Bowen, John. 2000. Airline hubs in Southeast Asia: national economic development and nodal accessibility. *Journal of Transport Geography*, 8: 25-41.

[18] Brill, E.D., J. M. Flach, L. D. Hopkins, and S. Ranjithan. 1990. MGA: a decision support system for complex, incompletely defined problems. *IEEE Transactions: Systems, Man, and Cybernetics.* 20: 745-757.

[19] Campbell, J. F. 1994. Integer programming formulations of discrete hub location problems. *European Journal of Operational Research* 72:387-405.

[20] Charnes, A. and W. Cooper. 1959. Chance-constraint programming. *Management Science* 6: 73-79.

[21] Chou, Yue-Hong. 1993. Airline accessibility and nodal accessibility. *Journal of Transport Geography*, 1(1): 37-46.

[22] Church, Richard L. 1999. Location modeling and GIS. In *Geographical Information Systems, Vol. 1: Principles and Technical Issues* (2nd ed.), ed. P. A. Longley, M. F. Goodchild, D. J. Maguire, and D. W. Rhind, 293-303. New York: John Wiley & Sons.

[23] Cohen, Reuven, and Shlomo Havlin. 2003. Scale-free networks are ultrasmall. *Physical Review Letters*, 90 (5): 058701.

[24] Cohon Jerry L. 1978. *Multiobjective programming and planning*. New York: Academic Press.

[25] Daskin, Mark. 1995. *Network and discrete location.* New York: John Wiley & Sons.

[26] Davis, D.L. & R. A. Elnicki. 1984. User cognitive types for decisions support systems. *Omega.* 12: 601-614.

[27] Densham, Paul J. 1999. Spatial decision support systems. In *Geographical information systems: principles and applications*, ed. D. J. Maguire, M. F. Goodchild, and D. W. Rhind, 403-412. London: Longman.

[28] Dia, Hussein. 2002. An agent-based approach to modeling driver route choice behavior under the influence of real-time information. *Transportation Research Part C*, 10: 331-349.

[29] Diplock, G. and S. Openshaw. 1996. Using simple genetic algorithms to calibrate spatial interaction models. *Geographical Analysi`s* 28: 262-279.

[30] Dobson, G., and P. J. Lederer. 1993. Airline scheduling and routing in a hub-and-spoke system. *Transportation Science* Vol. 27, No. 3: 281-297.

[31] Dueker, K. J. and D. Kjerne. 1989. *Multipurpose cadastre: terms and definitions*. Falls Church, VA: ASPRS and ACSM.

[32] Flynn, John and Samuel Ratick. 1988. A multiobjective hierarchical covering model for the Essential Air Services Program. *Transportation Science* 22: 139-147.

[33] Friesz, Terry L., G. Anandalingam, Nihal J. Mehta, Keesung Nam, Samir J. Shah, and Roger L. Tobin. 1993 The multiobjective equilibrium network design problem revisited: a simulated annealing approach. *European Journal of Operational Research* 65: 44-57.

[34] Fu, M. C. 2001. Optimization for simulation: theory and practice. *INFORMS Journal on Computing* 14: 192-215.

[35] Garrison, William L. 1960. Connectivity of the interstate highway system. *Papers and Proceedings of the Regional Science Association*, 6: 121-137.

[36] Garrison, William L. and Duane F. Marble. 1961. *The structure of transportation networks*. Washington DC: U. S. Department of Commerce, Office of Technical Services.

[37] Garrison, William L. and Duane F. Marble. 1962. *The structure of transportation networks* (DA 44-177-TC-685). Fort Eustis, VA: U. S. Army Transportation Research Command.

[38] Gaupp, Martin P., and Raymond R. Hill. 1999. Using adaptive agents in Java to simulate U.S. Air Force pilot retention. *Proceedings of the 1999 Winter Simulation Conference*, 1152-1159.

[39] Gelenbe, Erol, Esin Şeref, and Zhiguang Xu. 2001. Simulation with learning agents. *Proceedings of the IEEE*, 89(2): 148-157.

[40] Gewin, Virginia. 2004. Mapping opportunities. *Nature* 427: 376-377.

[41] Gimblett, H. Randy (Ed.). 2002. *Integrating geographic information systems and agent-based modeling techniques for simulating social and ecological processes*. New York: Oxford University Press.

[42] Goodchild, Michael F. 1992. Geographical information science. *International Journal of Geographical Information Systems* 6: 31-45.

[43] Guimerá, Roger, Sefano Mossa, Adrian Turtschi, and Luís A. Nunes Amaral. 2004. Structure and efficiency of the world-wide airport network. http:xxx.arxiv.org/abs/cond-mat/0312535.

[44] Haggett, Peter. and Richard Chorley. 1967. Models, paradigms, and the new geography. In *Physical and Information Models in Geography*, ed. by P. Haggett and R.Chorley. London: Methuen, p. 19-41 (read p. 19-26 only)

[45] Haining, Robert. 1990. *Spatial data analysis in the social and environmental sciences.* Cambridge: Cambridge University Press.

[46] Hansen, W. G. 1959. How accessibility shapes land use. *Journal of the American Institute of Planners*, 25: 73-76.

[47] Harvey, David. 1969. *Explanation in geography*. New York: St. Martin's Press, Ch. 4 (p. 30-43), Ch. 10 (p. 141-161).

[48] Hicks, Richard, Richard Madrid, Chris Milligan, Robert Pruneau, Mike Kanaley Yvan Dumas, Benoit Lacroix, Francois Lessard, John Allarie, Francois Soumis, Jacques Desrosiers. 2005. Breakthrough performance improvement in large fractional aircraft ownership operations by Bombardier Flexjet. *Interfaces* (forthcoming, 2005).

[49] Howrey, E.P. 1969. On the choice of forecasting models for air travel. *Journal of Regional Science* 9: 215-224.

[50] Kansky, K. J. 1963. *Structure of transportation networks: relationships between network geometry and regional characteristics*. Research Paper 84, Chicago: Department of Geography, University of Chicago.

[51] Kincaid, Rex K. 2003. *Scale-free graphs for general aviation flight schedules* . NASA/CR-2003-212648.

[52] Kincaid, Rex K. and Natalia Alexandrov. 2004. Scale-free networks: a discrete-event simulation approach. Unpublished manuscript.

[53] Kirkpatrick, S., C.D. Gelatt Jr., and M. P. Vecchi. 1983. Optimization by simulated annealing. *Science* 220: 671-680

[54] Klincewicz, J. G. 1991. Heuristics for the p-hub location problem. *European Journal of Operational Research* 53:23-57.

[55] Kuby, M. J. and R. G. Gray. 1993. The hub network design problem with stopovers and feeders: the case of Federal Express. *Transportation Research-A* Vol. 27A, No. 1:1-12.

[56] Jones, James Holland, and Mark S. Handcock. 2002. Statistical evidence tells tails of human sexual contacts. Working Paper no. 21, University of Washington Center for Statistics and the Social Sciences.

[57] Lang, L. 1999. *Transportation GIS*. Redlands, CA: ESRI Press.

[58] LeBlanc, L. J., E. K. Morlok, and W. P. Pierskalla. 1975. An efficient approach to solving the road network equilibrium traffic assignment problem. *Transportation Research* 9: 309-318.

[59] Lederer, Philip J. and R. S. Nambimadom. 1998. Airline network design. *Operations Research* 46: 785-804.

[60] Lee, L. H., H. C. Hunag, C. Lee, E. P. Chew, W. Jaruphongsa, Y. Y. Yong, Z. Liang, C. H. Leong, Y. P. Tan, K. Namburi, E. Johnson, and J. Banks. 2003. Discrete event simulation model for airline operations: SIMAIR. *Proceedings of the 2003 Winter Simulation Conference*, S. Chick, P. J. Sanchez, D. Ferrin, and D. J. Morrice, eds.

[61] Magnanti, T. L. and R. T. Wong. 1984. Network design and transportation planning: models and algorithms. *Transportation Science* 18: 1-55.

[62] Marianov, V., D. Serra and C. ReVelle. 1999. Location of hubs in a competitive environment. *European Journal of Operational Research* 114:363-371.

[63] Menczer, Filippo. 2004. Evolution of document networks. *Proceedings of the National Academy of Sciences*, in press.

[64] Miller, Harvey J. 1996. GIS and geometric representation in facility location problems. *International Journal of Geographical Information Systems* 10: 791-816.

[65] Miller, Harvey J. and Shih-Lung Shaw. 2001. *Geographic information systems for transportation: principles and applications.* Oxford University Press, Inc.

[66] Motter, Adilson E., and Ying-Chen Lai. 2002a. Cascade-based attacks on complex networks. *Physical Review E*, 66: 065102.

[67] Motter, Adilson E., Takashi Nishikawa, and Ying-Chen Lai. 2002b. Range-based attack on links in scale-free networks: Are long-range links responsible for the small-world phenomenon? *Physical Review E*, 66: 065103.

[68] Neidringhaus, William. 2000. *An agent-based model of the airline industry*. MITRE Corp. http://www.mitrecaasd.org/library/papers/ACSEM.pdf

[69] O'Kelly, Morton E. 1987. A quadratic integer program for the location of interacting hub facilities. *European Journal of Operational Research* 32:393-404.

[70] O'Kelly, Morton E. and Harvey J. Miller. 1994. The hub network design problem. *Journal of Transport Geography* 2(1):31-40.

[71] O'Kelly, Morton E. and Y. Lao. 1991. Mode choice in a hub-and-spoke network: a zero-one linear programming approach. *Geographical Analysis* Vol. 23, No. 4.

[72] Peeta, Srinivas, Pengcheng Zhang, and Weimin Zhou. 2004. Behavior-based analysis of freeway car-truck interactions and related mitigation strategies. *Transportation Research Part B*, forthcoming.

[73] Peuquet, Donna J. and Elizabeth A. Wentz. 1994. An approach for time-based analysis of spatiotemporal data, *Advances in GIS Research* 1: 489-504.

[74] Redner, S. 1998. How popular is your paper? An empirical study of the citation distribution. *European Physical Journal B*, 4 (2): 131-134.

[75] Ríos, M., V. Marianov, and M. Gutierrez. 2000. Survivable capacitated network design problem: new formulation and Lagrangean relaxation. *Journal of the Operational Research Society* 51: 574-582.

[76] Roberts, Tyler D. (in preparation). Complex networks and the US waterborne shipping network. MA Thesis, Arizona State University.

[77] Rosenberger, J. M., A. J. Schafer, D. Goldsman, E. L. Johnson, A. J. Kleywegt, and G. L. Nemhauser. 2000. SIMAIR: A stochastic model of airline operations. *Proceedings of the 2000 Winter Simulation Conference* J. A. Joines, R. R. Barton, K. Kang, and P. A. Fishwick, eds.

[78] Rosing, Kenneth E., Charles S. ReVelle, E. Rolland, David A. Schilling, and John R. Current. 1998. Heuristic concentration and tabu search: a head to head comparison. *European Journal of Operational Research* 104: 93-99.

[79] Schwartz, Jacob T. 2001. *Introduction to matrices and vectors*, Dover: Mineola, NY.

[80] Shaw, Shih-lung. 1991. Urban transit accessibility analysis using a GIS: a case study of Florida's Tri-Rail System. *Southeastern Geographer* 31(1):15-30.

[81] Shimbel, A. 1953. Structural Parameters of Communications Networks. *Bulletin of Mathematical Biophysics*, 15: 501-507.

[82] Skorin-Kapov, D. and J. Skorin-Kapov. 1994. On tabu search for the location of interacting hub facilities. *European Journal of Operational Research* 73: 502-509.

[83] Sohn, J. and S. Park. 1997. A linear program for the two-hub location problem. *European Journal of Operational Research* 100:617-622.

[84] Sohn, J. and S. Park. 1998. Efficient solution procedure and reduced size formulations for p-hub location problems. *European Journal of Operational Research* 108:118-126.

[85] Stamatopoulos, M. A., K. G. Zografos, A. R. Odoni. 2004. A decision support system for airport strategic planning. *Transportation Research Part C* Vol. 12. 91-117.

[86] Strogatz, Steven. 2001. Exploring complex networks. *Nature,* 410: 268-276.

[87] Taaffe, Edward J., Howard L. Gauthier, and Morton E. O'Kelly. 1996. *Geography of transportation*, 2nd Ed. Prentice Hall: Upper Saddle River, New Jersey.

[88] Tassier, Troy, and Filippo Menczer. 2001. Emerging small-world referral networks in evolutionary labor markets. *IEEE Transactions on Evolutionary Computation*, 5 (5): 482-492.

[89] Teodorović, D. 1985. A model for designing the meteorologically most reliable airline schedule. *European Journal of Operational Research* 21:156-164.

[90] Teodorović, D., and S. Guberinić. 1984. Optimal dispatching strategy on an airline network after schedule perturbation. *European Journal of Operational Research* 15:178-182.

[91] Thill, Jean-Claude. 2000. Geographic information systems for transportation in perspective. *Transportation Research Part C*. 8 (2000) 3-12.

[92] Unwin, D. J. 1996. GIS, spatial analysis and spatial statistics. *Progress in Human Geography*. 20 (4) 540-551.

[93] Thomas, R. W. and Huggett, R. J. 1980. *Modeling in geography: a mathematical approach.* Totowa, NJ: Barnes and Noble Books, p. 3-10.

[94] Upchurch, Christopher, Michael Kuby, Michael Zoldak, and Anthony Barranda. 2004. Using GIS to Generate Mutually Exclusive Service Areas Linking Travel On and Off a Network. *Journal of Transport Geography*, 12: 23-33.

[95] Waddell, Paul. 2002. UrbanSim: modeling urban development for land use, transportation and environmental planning. *Journal of the American Planning Association* 68: 297-314.

[96] Waters, Nigel M. 1999. Transportation GIS: GIS-T. In *Geographical information systems, Vol. 2: Management issues and applications* (2nd edition), ed. P. A. Longley, M. F. Goodchild, D. J. Maguire, and D. W. Rhind, 827-843. New York: John Wiley & Sons.

[97] Watts, Duncan J. 1999. Networks, dynamics, and the small-world phenomenon. *American Journal of Sociology*, 105 (2): 493-527.

[98] Watts, Duncan J. 2003. *Six degrees: the science of a connected age*. New York: W.W. Norton.

[99] Wilhite, Allen. 2001. Bilateral trade and 'small-world' networks. *Computational Economics*, 18: 49-64.

[100] Wilmott, Cort. and Gaile, Gary. 1992. Modeling. In *Geography's Inner Worlds*, ed. By R.F. Abler, M.G. Marcus, and J. M. Olson. New Brunswick, NJ: Rutgers University Press, p. 163-186.

[101] Yan, S. and J. Chang. 2002. Airline cockpit crew scheduling. *European Journal of Operational Research* 136:501-511.

[102] Yan, S., and Y. Tu. 1997. Multifleet routing and multistop flight scheduling for schedule perturbation. *European Journal of Operational Research* 103: 155-169.

[103] Yan, S., and Y. Tu. 2002. A network model for airline cabin crew scheduling. *European Journal of Operational Research* 140: 531-540.

[104] Yang, L. and R. Kornfeld. 2003. Examination of the hub-and-spoke network: a case example using overnight package delivery. In *Proceedings of the 41st Aerospace Sciences Meeting and Exhibit*, January 6-9, 2003, Reno, Nevada, AIAA 2003-1334.

Table 1. Summary of Types of Models and the Types of Research Questions They Address

Method	Classification	Type of Research Question
GIS	i. Descriptive ii. Mathematical iii. No process iv. Spatial (any kind) v. Time slices vi. Deterministic (but can support stochastic analysis) vii. Linear or nonlinear viii. Any level of disaggregation	*1: Where is a particular thing located, or what is at a particular location?* *2: How are different phenomena spatially related?* *3: What areas lie within a given travel time, cost, or distance of a point, line, or polygon?* *4: How many potential users exist within reach of a particular transportation system?* *5: What areas have one or more zones in common?* *6: What are the characteristics of the area surrounding a place, and where are the local maxima and minima?* *7: What are the best estimates for the missing values of a spatial distribution of data?* *8: How have geographical phenomena changed over time?* *9: Where should a new transportation facility be located?* *10: What route should a new transportation corridor follow to connect two places?* *11: What will be the environmental impacts of a transportation project?* *12: Which areas are suitable for a given activity?* *13: How do the answers to other research questions depend on the spatial units of analysis?*
Graph Theory	i. Descriptive ii. Mathematical iii. No process iv. Topological and sometimes spatial v. Static, except random graph theory vi. Deterministic, except random graph theory vii. Linear or nonlinear viii. Aggregate or disaggregate	*1: Which nodes in a network are the most accessible and best connected to other nodes?* *2: How do networks compare in terms of connectivity?* *3: Is the network infrastructure adequate to serve the area or population?* *4: How are movement and economic activity related to network structure and nodal accessibility?*
Complex Networks	i. Descriptive and predictive ii. Mathematical iii. Black box iv. Nonspatial but topological v. Dynamic vi. Stochastic vii. Nonlinear viii. Aggregate or disaggregate	*1: What is a complex network topology, and how is it constructed?* *2: How vulnerable are complex networks to attack and failure?* *3: What are the optimal strategies for dealing with a catastrophic failure or attack?* *4: What are some problems with integrating complex network theory with the airline system?* *5: What is the best point of attack on a complex network?* *6: What are the optimal strategies for dealing with a catastrophic failure or attack?* *7: Which transportation networks have scale-free or small-world properties, and if so, why?*

Mathematical Programming	i. Prescriptive ii. Mathematical iii. Black box or process-response iv. Spatial or nonspatial v. Static or dynamic vi. Deterministic or chance-constrained vii. LP is linear, but there are methods for nonlinear or integer. viii. Aggregate or disaggregate	*1: Where should facilities be located?* *2: How should (or will) traffic flow over a network?* *3: Which links should be added to a network?* *4: What is the optimal design for hub-and-spoke network? What is the optimal location of airline hubs?* *5: Which type of network structure is best, and what parameters affect the optimal choice?* *6: What is the optimal schedule?* *7: What is the optimum way to recover from a schedule perturbation?* *8: What is the optimal crew schedule?* *9: What are the optimal flight and crew routings and schedule for an on-demand airline?* *10: What are the tradeoffs between multiple objectives of an optimization problem?*
Systems-Level Queueing-Type Simulations	i. Predictive ii. Mathematical iii. Process-response iv. Spatial or nonspatial v. Dynamic vi. Stochastic vii. Nonlinear viii. Disaggregate	*1: What are the consequences of disruptions of an airline's schedule?* *2: How will airport capacity be affected if the airside facilities are changed?*
Agent Based Models (ABM)	i. Predictive and explanatory ii. Mathematical iii. Process-response iv. Spatial or nonspatial v. Dynamic vi. Stochastic vii. Nonlinear viii. Disaggregate	*1: How do vehicle operators interact on a network?* *2: What types of trips will be made by which kinds of people at which times?* *3: How do people choose their routes to travel on a network?* *4: How do different sets of agents of different classes within the air transport system interact with each other?* *5: How do land use and transportation systems co-evolve?*

Appendix 1

A Preliminary List of GIS Layers for Air Transport

1. Airports
    a. Type
        i. Primary Commercial Service
        ii. Non-primary Commercial Service
        iii. Reliever
        iv. General Aviation
    b. Hub classifications (large, medium, small)
    c. Capacity
        i. Gate space
        ii. Hangers
    d. Aircraft
        i. Types and number of takeoffs
            1. Jets
            2. Turbo-props
            3. Regional Jets
            4. single engine (duel engine planes)
        ii. Number of locally owned aircraft
            1. Takeoffs
    e. Runway Parameters
        i. Number of runways
        ii. Length
        iii. Type of aircraft capable of landing on each runway
        iv. Pavement management
            1. Resurfacing
            2. Failures
            3. Damage caused
    f. Enplanements
        i. Passenger
            1. Destination
                a. Domestic
                b. International
            2. Type of traveler
                a. Connections (for hub airports)
                b. Origin / destination
        ii. Cargo
        iii. Hazardous materials
    g. Airlines at the airport
        i. Passenger
            1. Gates allocated to each airline
            2. Market share
            3. Hub status
            4. Facilities (such as maintenance)

            5. Banks (groups of incoming/outgoing flights, for hub operations)
        ii. Cargo
h. Safety
    i. Accidents
        1. Human error
        2. Weather related
        3. Other
        4. Fatalities
    ii. X-ray machines
    iii. Emergency personnel
        1. Fire trucks / ambulances
    iv. Deicing facilities
    v. Snow removal equipment
    vi. Search and rescue resources
    vii.
i. Accessibility
    i. Public transit
    ii. Parking facilities / lots / spaces
    iii. Local population within “X” miles of the airport
    iv.
j. Personnel
    i. Security (TSA)
    ii. Customs officials
    iii. Emergency personnel
    iv. General employment
k. Layout of existing facilities
l. Future Expansion
    i. Construction projects
    ii. Costs
    iii. Underground Utility lines
        1. Management and repairs
    iv. Land area
    v. Surrounding land uses
m. Environmental
    i. Noise Mitigation
        1. Noise contour data
    ii. Crosswinds
    iii. Air quality
n. Financial Performance
    i. Revenues
        1. Rents, Fees and user taxes
        2. Aeronautical vs. land based
    ii. Expenses
        1. Property taxes, land acquisition
        2. Fuel costs
    iii. Estimated cost for future operations

            iv. Funding
                1. Federal
                2. Local
    o. Connectivity
            i. With every other airport
                1. Number of flights per day
    p. Efficiency
            i. Average turn-around time for aircraft
            ii. Resource utilization (ability to handle traffic redistribution)
            iii. Delays
                1. Peak times
                2. Congestion related
                3. Air Traffic Control related
                4. Weather related
            iv. Lost bags
    q. Pricing
            i. Average fares
            ii. Maximum fares
    r. Competition
            i. Hubbing factor (percentage of passengers changing planes)
            ii. Monopoly/duopoly measures
            iii. Hinterland

2. Airspace/Flight paths (possibly in 3-D)
    a. Capacity
    b. Between airports
    c. Emergency routes
            i. Weather-related detours
            ii. Distance between airports
            iii. Required fuel for all routes
    d. In-tow requirements (the distance a plane must trail the plane in front of it)
    e. Restricted airspace
3. Mobility (Demand)
    a. Passengers or Cargo
    b. Inter-airport (e.g., O'Hare-LaGuardia)
    c. Inter-city (e.g., Chicago-New York)
    d. Seasonality
    e. Growth rate
    f. General aviation
    g. Charter
    h. Other modes
4. Airlines
    a. Hub locations, if any
    b. Route network
            i. Airports served
            ii. Aircraft utilized

            iii. Frequency
    c. Crew scheduling
    d. Headquarters
    e. Maintenance facilities
    f. Passenger volume
    g. Charters
5. Topographic elevation levels
    a. Digital elevation model
    b. Minimum – around metropolitan areas
    c. Maximum – en route

6. Weather
    a. Average annual rain / snow
    b. Effect on air traffic control
    c. Air pressure
    d. Temperature
    e. Density Altitude
7. Air Traffic Control
    a. Available controllers
    b. Tower computer equipment
        i. Recent upgrades
        ii. Number of Failures
8. Air route traffic control centers
    a. Number of employees
    b. Number of flights
        i. Peak hours
        ii. Other
9. Radar Stations
    a. Type
    b. Coverage Area
10. Demographic Data
    a. Population
    b. Growth
    c. Income
    d. Age
    e. Mobility (enplanements per person)
    f. Migration
11. Economic Data
    a. Employment
    b. Types of industries
    c. Growth
    d. Tourism
12. Modal Competition
    a. Rail Network
        i. Standard
            1. Travel time

        2. Cost
        3. Capacity
    ii. High-speed
        1. Travel time
        2. Cost
        3. Capacity
    iii. Freight
        1. Travel time
        2. Cost
        3. Capacity
b. Road Network
    i. Automobile
        1. Travel time
        2. Cost
        3. Capacity
    ii. Bus
        1. Travel time
        2. Cost
        3. Capacity
    iii. Freight
        1. Travel time
        2. Cost
        3. Capacity
c. Water Network
    i. International freight
        1. Travel time
        2. Cost
        3. Capacity

13. NORAD
    a. Bases
        i. Aircraft
            1. Speed
    b. Radar

REPORT DOCUMENTATION PAGE	Form Approved OMB No. 0704-0188

1. REPORT DATE *(DD-MM-YYYY)*	**2. REPORT TYPE**	**3. DATES COVERED** *(From - To)*
01- 02 - 2005	Contractor Report	

4. TITLE AND SUBTITLE
A Comparison of Geographic Information Systems, Complex Networks, and Other Models for Analyzing Transportation Network Topologies

5a. CONTRACT NUMBER
NNL04AA57G

5b. GRANT NUMBER

5c. PROGRAM ELEMENT NUMBER

6. AUTHOR(S)
Kuby, Michael; Tierney, Sean; Roberts, Tyler; and Upchurch, Christopher

5d. PROJECT NUMBER

5e. TASK NUMBER

5f. WORK UNIT NUMBER
23-090-20-15

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)
NASA Langley Research Center, Hampton, VA 23681-2199
Arizona State University, Department of Geography, Tempe, Arizona 85287

8. PERFORMING ORGANIZATION REPORT NUMBER

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)
National Aeronautics and Space Administration
Washington, DC 20546-0001

10. SPONSOR/MONITOR'S ACRONYM(S)
NASA

11. SPONSOR/MONITOR'S REPORT NUMBER(S)
NASA/CR-2005-213522

12. DISTRIBUTION/AVAILABILITY STATEMENT
Unclassified - Unlimited
Subject Category 66
Availability: NASA CASI (301) 621-0390

13. SUPPLEMENTARY NOTES
Langley Technical Monitor: Natalia Alexandrov
An electronic version can be found at http://ntrs.nasa.gov

14. ABSTRACT
This report reviews six classes of models that are used for studying transportation network topologies. The report is motivated by two main questions. First, what can the "new science" of complex networks (scale-free, small-world networks) contribute to our understanding of transport network structure, compared to more traditional methods? Second, how can geographic information systems (GIS) contribute to studying transport networks? The report defines terms that can be used to classify different kinds of models by their function, composition, mechanism, spatial and temporal dimensions, certainty, linearity, and resolution. Six broad classes of models for analyzing transport network topologies are then explored: GIS; static graph theory; complex networks; mathematical programming; simulation; and agent-based modeling. Each class of models is defined and classified according to the attributes introduced earlier. The paper identifies some typical types of research questions about network structure that have been addressed by each class of model in the literature.

15. SUBJECT TERMS
Transportation systems; Transportation networks; Geographical Information Systems; Complex networks

16. SECURITY CLASSIFICATION OF: a. REPORT	b. ABSTRACT	c. THIS PAGE	17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON / 19b. TELEPHONE NUMBER *(Include area code)*
U	U	U	UU	58	STI Help Desk (email: help@sti.nasa.gov) / (301) 621-0390