UNIVERSITY OF CALIFORNIA

Santa Barbara

Analysis of Geographically Embedded Networks

A dissertation submitted in partial satisfaction of the

requirements for the degree Doctor of Philosophy

in Geography

by

John Alan Glennon

Committee in charge:

Professor Michael Goodchild, Chair

Professor Richard Church

Professor Keith Clarke

Professor Shih-Lung Shaw

March 2013

The dissertation of John Alan Glennon is approved.

____
Richard Church

____
Keith Clarke

____
Shih-Lung Shaw

____
Michael Goodchild, Committee Chair

March 2013

Analysis of Geographically Embedded Networks

by

John Alan Glennon

ACKNOWLEDGEMENTS

The completion of this dissertation ends a memorable period for me in Santa Barbara. It was my privilege to progress as a scholar with the support, guidance, and friendship of many of the world's great geographers.

At UCSB, almost every Geography faculty member offered useful advice, interesting lunch discussion, or supportive words at some time during my residence. I am grateful for the keen guidance of my committee of Michael Goodchild, Rick Church, Keith Clarke, and Shih-Lung Shaw. Thanks also go to Helen Couclelis, Sara Fabrikant, James Frew, Don Janelle, Hugo Loaiciga, Dan Montello, Fiona Goodchild, and Waldo Tobler.

Discussions and travels with my spatial@ucsb colleagues Josh Bader, Karl Grossner, and Indy Hurt remain my fondest memories of Santa Barbara. The following students have also made positive contributions to my thinking about the geographic discipline, including Matt Rice, Andrea Nuernburger, Sean Benison, John Gallo, Jorge Sifuentes, Linna Li, Kitty Currier, Reg Archer, Carlos Baez, Tim Niblett, Matt Niblett, Matt Vitale, Julie Dillemuth, Laurel Suter, Wen Wen Li, Shaun Walbridge, Chad Catacchio, Kailen Wright, Eric Davila, Suzanne Foss, Keely Roth, Nate Royal, Jeff Howarth, Ed Pultar, Kerry Halligan, Susan Tran, Tom Pingel, and Drew Dara-Abrams.

I appreciate the UCSB Geography staff for smoothing the path and offering administrative support, particularly Bryan Karaffa, Guylene Gadal, Dylan Parenti, Karen Doehner, Mo Lovegreen, Connie Padilla, Susanna Baumgart, Bernadette Weinberg, Beilei Zhang, and Jose Saleta.

Numerous external scholars offered support and advice in this research including Will McClintock (UCSB Marine Science), Tom Cova (Utah), May Yuan (Oklahoma), Dan Sui (Ohio State), Diana Sinton (Redlands), Karen Kemp (USC), Ben Zhao (UCSB Computer Science), and Barbara Harthorn (UCSB Feminist Studies).

Thanks to Jack and Laura Dangermond for their kind financial support during portions of this study.

The geyser use case study hinged on the contributions of many observers and scientists. I especially appreciate the efforts of Jacob Young, my fellow developer on the geyser eruptions database, mobile app, and web site. Also, thanks to Graham Meech, Tara Cross, Jeff Cross, Lynn Stephens, Mary Beth Schwartz, Scott Bryan, Clark Murray, Vicki Whitledge, David Goldberg, Dean Lohrenz, Jim Scheirer, Kitt Barger, Ralph Taylor and the many other gazers that have contributed data, waited long hours with me for an eruption, and taught me so much.

Finally, a great thanks goes to my family, including Ashley, Suzanne, Merritt, and Christian Glennon. Rhonda's parents, Keith and Arlene Pfaff, offered car and home maintenance on the long bouts when I was studying and writing. Thanks to my parents, Robert and Rita Glennon, for the many years of love and support that allowed me to pursue both geographic exploration and academia.

I dedicate this work to my wife, Rhonda Glennon.

VITA OF JOHN ALAN GLENNON

January 2013

Education

2013 Doctor of Philosophy in Geography, University of California, Santa Barbara. Dissertation: Analysis of Geographically Embedded Networks. Advisor: Dr. Michael Goodchild.

2001 Master of Science in Geoscience, Western Kentucky University. Thesis: Application of Morphometric Relationships to Active Flow Networks within the Mammoth Cave Watershed, Kentucky. Advisor: Dr. Chris Groves.

1994 Bachelor of Science, Recreation, Park and Tourism Sciences, Texas A&M University. Advisor: Dr. Valeen Silvy.

Professional Experience

Principal, Advanced Research on Geographic Information (Arogi) Laboratory, Loma Linda, California, 2012-.

Geospatial Consultant, Center for Spatial Studies, University of California, Santa Barbara, 2008-2011.

Instructor, Geography 12: Maps and Mapping, University of California, Santa Barbara, Department of Geography, 2007.

Graduate Student Researcher, Department of Geography, University of California, Santa Barbara, 2003 – 2007.

Co-Founder and Assistant Director, Hoffman Environmental Research Institute, Western Kentucky University, 1999 – 2003.

Research Hydrologist, Center for Cave and Karst Studies, Western Kentucky University, 1998 – 1999.

Park Ranger, Mammoth Cave National Park, KY, 1994, 1996 – 1998; Jewel Cave National Monument, SD, 1991 – 1993; Great Basin National Park, NV, 1995.

Awards and Honors

Visiting Faculty, LENS Summer Institute, University of Redlands, May 2011.

Most Creative App (Team): Android Developer Camp, PayPal, San Jose, CA, 2011.

Jack and Laura Dangermond Graduate Fellowship, Department of Geography, University of California, Santa Barbara, May 2007-May 2008.

Graduate Science Fellow, Center for Nanotechnology in Society, University of California, Santa Barbara, September 2006-April 2007.

Graduate Student Block Grant Fellowship, Department of Geography, University of California, Santa Barbara, 2003-2004.

John D. Minton Award for Outstanding Graduate Student, Western Kentucky University, 2001.

Special Achievement Award, Interpretation and Education, National Park Service, 1991.

Peer Reviewed Articles

Glennon, J.A., 2010, Creating and Validating Object-Oriented Geographic Data Models: Modeling Flow within GIS. Transactions in GIS. Vol. 14(1), pp.23-42.

M.F. Goodchild and J.A. Glennon, 2010, Crowdsourcing geographic information for disaster response: a research frontier. *International Journal of Digital Earth* 3(3): 231–241. [495]

M.F. Goodchild and A. Glennon, 2008, Representation and computation of geographic dynamics. In K.S. Hornsby and M. Yuan, editors, *Understanding Dynamics of Geographic Domains.* Boca Raton: CRC Press, pp. 13–30. [448]

Glennon, J.A., 2006, Cave map, in Cartography in the Twentieth Century, Volume 6. edited by M. Monmonier.

Glennon, J.A., 2006, Geysers and hot springs, in New Book of Knowledge Encyclopedia. Scholastic: New York.

Glennon, J.A. and R.M. Pfaff, 2005, The operation and geography of carbon-dioxide-driven, cold-water geysers. *GOSA Transactions*, vol. 9, pp. 184-192.

Glennon, J.A. and R. Pfaff, 2003, The extraordinary thermal activity of the El Tatio Geysers, Antofagasta Province, Chile, Transactions of the Geyser Observation and Study Association, vol. 8, pp. 31-78.

Glennon, J.A. and C. Groves, 2002, An examination of perennial stream drainage patterns within the Mammoth Cave watershed. Journal of Cave and Karst Studies, vol. 64, pp. 82-91.

Technical Articles and Invited Talks

Glennon, J.A., 2012, Characterizing Geographically Embedded Networks, University of Redlands Colloquium, Esri, Redlands.

Glennon, J.A., 2011, Panelist, Increasing National Resilience to Hazards and Disaster, NAS Committee Meeting, Beckman Center of the National Academies, Irvine.

Clarke, K., Glennon, J.A., Goodchild, M.F., Hurt, I., and D. Janelle, 2010, Tools for Map Making, invited workshop on spatial technology for the UCSB Interdisciplinary Humanities Center.

Glennon, R.M., and J.A. Glennon, 2007, “Making Better Maps with ArcGIS 9.2.” ArcUser, Summer (July-September) 2007, pp. 59-62.

Glennon, R.M., and J.A. Glennon, 2007, "Building a Groundwater Protection Model." in GIS Best Practices: GIS for Cave and Karst, ESRI: Redlands, pp. 37-49.

Glennon, J.A., 2006, Evolving Capabilities for Virtual Globes, Invited talk for the session “Use of Virtual Globes in the Earth Science,” American Geophysical Union Fall Meeting, San Francisco.

Software Portfolio

Held, M., Gardi, C., Shomali, R., Glennon, J.A., O’Brien, R, and S. Liu, 2011, Android Assassin (Android/Java/Ruby on Rails), location-based tag game.

Glennon, J.A. 2010-present, Geyser Notebook (Android/Java/php), geyser information app available on the Google Play App Market

Glennon, J.A., 2010, Flowpy (Python), http://alanglennon.com/genets

Glennon, J.A., Goodchild, M.F., and W. Tobler, 2008, Flow Data Model Tools (VBA for ArcGIS 9.x), http://alanglennon.com/genets

Fields of Study

Geography and Geographic Information Science

ABSTRACT

Analysis of Geographically Embedded Networks

by

John Alan Glennon

Geographically embedded networks (GENets) are systems of physical and abstract linked relationships contained wholly or partially within geographic space. The purpose of this dissertation is to examine the question, "what are the particular characteristics of networks in geography?" To uncover such properties, a series of network case studies is examined and each translated into computational data structures or workflows. Constrained as computational models, the cases are assessed and compared for patterns and common approaches. The modeling serves, 1) as a methodological template for the discovery of network properties, and 2) to reveal an initial set of characteristics for consideration based on the contemporary state of quantitative geographical analysis. The use cases are examined first by focusing on data structures and second through analytical workflows. The data structure use cases were selected to represent diverse GENet conditions and include Minard's map of Napoleon's March on Moscow, a Census table of state-to-state human migration, and a GIS dataset of the flow routes in

Kentucky's Mammoth Cave karst watershed. The analytical workflow use cases include an optimization problem (shortest path calculation), a process model (stream network generation from an elevation model), a simulation (growth along an urban corridor), and a GIS feature comparison (selection of streets bounding a city block). Commonalities from the data structure cases are distilled into a general model, but the analytical cases cannot be similarly reconciled. In order to make the geographic aspects of the analytical cases explicit, the cases are evaluated against a set of tests for spatial models devised by Goodchild (2012). From the case studies and tests, GENet characteristics can be subdivided into two categories: those that are properties of network with respect to the physical environment and those that are properties of representation. GENets with arcs or nodes that physically exist in the environment may exhibit characteristics of constituent heterogeneity and areal interaction. GENets representations are affected by issues of scale and spatial uncertainty. To evaluate these characteristics in a GIS implementation, an itinerary planning GENet problem, the Geyser Travel Problem, is introduced. The problem offers a practical application of GENet workflows in GIS and an avenue for discussion of associated best practices.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ORGANIZATION OF CHAPTERS

The dissertation is organized into six chapters with references compiled at the end of the document. Chapters and figures are labeled using the format “Chapter.number”.

The preface provides a summary of the dissertation with key contributions to geographic analysis enumerated and briefly described. Chapter 1 offers a historical context of network analysis in geography. After an introduction to network science, a literature review focuses on the geographic domains that commonly implement network analysis, including hydrology, transportation studies, agent-based modeling, and operations research.

Chapter 2 reviews forms of GENet representation, from mathematical forms to computational structures and visualization. Cataloging the available network representations is useful because representations formalize the varying strategies for conceptualizing connectedness.

Chapters 3 and 4 pursue the characterization of GENets via use case analysis, initially involving the modeling of data structures (Chapter 3) and then through analytical workflows (Chapter 4). Chapter 3 describes the creation of a data model for GENets associated with flow. By synthesizing the GENet aspects of Minard’s Map of Napoleon’s March on Moscow, U.S. Census data on state to state migration, and a karst groundwater flow map, a general flow data model is devised and interpreted. Chapter 4 follows four use cases from an analytical perspective. Representative queries from hydrology, transportation, GIS, and simulation were formalized via flowcharts and pseudocode. The general cases are then compared with respect to their data requirements, algorithmic differences, nature of output, and spatial character.

In Chapter 5, a practical GENet analytical application is examined—the Geyser Travel Problem. The problem involves the creation of a travel itinerary that maximizes the quality and quantity of geysers viewed during a visit to Yellowstone National Park. Visitors are constrained to view geysers along a built pathway, beginning and ending the tour at the same location. The case is chosen due to the diversity of techniques and consideration of GENet characteristics for its solution. Three visitor scenarios are implemented, each involving more available information during the itinerary creation process.

Chapter 6 further defines the categories of GENets, considers hybrid cases, and relates GENet categories to the case studies examined in the dissertation. Themes of the study, research contributions, and areas warranting further study and development are discussed.

Data and supplementary materials used in this study are accessible via the author's website: http://alanglennon.com/genets

PREFACE

Geographically embedded networks (GENets) are systems of physical and abstract linked relationships contained wholly or partially within geographic space. Such networks are common and can be seen in the assembly of roads and intersections, geometry of rivers, alignment of electric transmission grids, maps of trade flows, and migration paths of animals. Each of these systems includes links or actors that reside in physical geographic space. Whereas the mathematical discipline of graph theory and allied field of network science have endeavored to uncover the general properties of all networks, an enumeration of the unique properties of geographically-related networks has not been undertaken. Thus, the purpose of this dissertation is to examine the question, “what are the particular characteristics of networks in geography?” Also, knowledge of the specific characteristics of such networks affords their better organization and analysis within the computational environment.

A series of network case studies are examined and each is transferred into computational data structures or workflows. Unified Modeling Language (UML), a general purpose graphical notation language that is often used for object-oriented software engineering, is the dissertation’s standard method for structuring the use cases. So constrained, the case studies can be assessed and compared for patterns and common approaches.

Though the work engages the question of unique GENet properties, this work is not intended to be exhaustive. Rather, modeling via case synthesis is intended as a methodological template for the discovery of such network properties and to reveal an initial set of characteristics for consideration based on the contemporary state of quantitative geographical analysis.

With respect to GENets, the bounds and scales of geographic space have been considered and defined by previous researchers (Montello 2001). Geographic space is the habitable domain of humans, thus for instance, quantum and interplanetary spatial and temporal scales are not included in this study. Due to the ubiquity of networks in scientific work, the terms network and graph are sometimes used interchangeably. In this dissertation, the use of these two terms is separate and deliberate: networks are systems of interconnected entities, and the term *network* will be used to describe conceptual notions surrounding interconnected geographic actors. *Graphs* are the mathematical representation of connected systems composed of a set of edges and vertices.

The dissertation begins with a discussion of existing studies on networks in geography. While the breadth of work is vast, network research in geography stems primarily from hydrology, transportation, regional science, location optimization, agent-based modeling, and GIS. Representations allow the conceptual notions encumbered within networks to be instantiated for manipulation and study. Analytical approaches for each domain differ, but their representations and data structures for computational analysis are similar.

In the dissertation's third chapter, a data model is created for flow networks. The endeavor highlights GENet data organization, modeling several use cases involving flow and then distilling their commonalities into a single model. The cases, selected to represent a wide variety of flow network conditions, are a Minard's map of Napoleon's March on Moscow, a Census table of state-to-state human migration, and flow routes in Kentucky's Mammoth Cave karst watershed. The derived general data model offers insights into conditions of mapped and uncertain channelized movement as well as dynamic flow characteristics through space.

GIS data structures offer little indication as to how a geographic phenomenon would interact with other entities in the environment. Further, such interactions are likely to be illuminating with regard to the fundamental characteristics of GENets. So, four diverse analytical GENet use cases are modeled, diagrammed, and compared in a GIS context. The cases include an optimization problem, a process model, an agent-based simulation, and a GIS feature comparison. Unlike the data model distillation process for the GENet flow data model, combining and reconciling the analytical workflows into a unified model is not possible. The use cases thus are evaluated against Goodchild's tests for spatial models (2012) in order to make the geographic aspects of each case explicit. From these tests, GENet characteristics can be subdivided into two categories: those that are properties of a network with respect to the physical environment and those that are properties of representation—its modeling, abstraction, or visualization.

GENets with arcs or nodes that tangibly exist in the environment possess constituent heterogeneity and areal interaction. Concerning heterogeneity, GENets may possess variable attributes, function, and geometry of constituent parts. For instance, a road may have variable speed limits and different widths along its course. Intersections likewise may be simple or complex. A complex intersection may hold characteristics like turn rules and traffic signals that vary with respect to approach direction. Tangible constituents of a GENet are influenced by their surroundings. A river network is a manifestation of the character of its catchment. The watershed, in turn, is affected by the river: the river may flood, form valleys, deposit material, and so on.

Representations of GENets are affected by scale and spatial uncertainty. Scale defines data granularity and its effects propagate to data collection, storage, analysis, visualization, and interpretation. For example, a road on a map often increases in length with finer scale. The physical road does not get longer, but its increased length arises with the additional detail afforded the fine-scale representation. Spatial uncertainty is another representational characteristic of GENets. Uncertainty arises from numerous sources, including imprecise measurement, missing data (resulting from unknown and unmeasured linkages), and data aggregation. Human migration, for instance, often is represented as a simple from-to link, while in fact, the actual people involved often will have taken numerous varying paths, and at different times and rates of travel.

GENets may be entirely embedded with geographic space—their arcs and nodes both reside in the physical world, or portions of the GENet may be abstracted. A social network or trade flow, for instance, possesses endpoint actors that represent tangible entities in geographic space. Their linkages, social and economic relationships in the aforementioned cases, are conceptual. The use cases in this dissertation emphasize GENets possessing tangible constituents. GENets with arcs representing conceptual relationships also are likely to have a set of unique properties, and with their increased use in computing, particularly via social networks, offer an avenue ripe for further study. The dissertation's discussion explores the continuum of physical and abstract GENets, and considers networks that possess arcs and nodes of differing character.

In the final section of this study, a route itinerary planning problem is presented that incorporates several GENet characteristics and technological design issues. These additional issues include spatiotemporal routing; crowdsourcing; mobile, in-field decision making; and ongoing versus one-time analysis. The chapter's routing problem involves visitors trying to maximize viewing satisfaction of numerous geysers proximal to a trail at Yellowstone National Park. Visitors are constrained to the trail and have limited time. The view quality of an eruption varies by location, and it is possible that more than one geyser could be viewed from a single location. The problem is approached from three perspectives common to Yellowstone visitors. First, the visitor follows the trail without regard to eruptions, and observes geysers only by chance. The second approach allows visitors to use a schedule of geyser eruption predictions as known at the time the tour begins. This approach is quite common among Yellowstone visitors. Visitors often obtain a prediction schedule at the beginning of the day from the Visitor Center, and then have no easy method to obtain updates throughout the day. The third approach for solving the problem allows visitors to simulate the acquisition of continuously updated eruption predictions throughout the day.

The Yellowstone routing problem advances discussion of best practices and issues for GENet workflow implementation with GIS. First, and as one would expect, methods that afford updated data and ongoing, rather than one-time, analysis yield more utility for users. Also, while GIS offers a generic platform for geographic analysis, its feature set focuses on comparison and selection analysis. Functionality, such as ArcGIS ModelBuilder, allows the ability to iterate results and feed the results into a subsequent analytical step, thus offering functionality to handle the basic workflow of process modeling. Future research should be conducted to include similar generic methods into mainstream GIS for temporal comparison and optimization analytics.

CHAPTER 1:
HISTORICAL FOUNDATIONS

Network and graph-like structures have a long history of influential use within the scientific community. The related mathematical discipline, graph theory, traces to the 18th Century with Leonhard Euler's approach to a routing problem over the seven bridges of Prussian Koenigsberg (Agnarsson and Greenlaw 2007). Euler used a graph to model bridges and path intersections to assess whether a route could be created that traversed each bridge only once (Figures 1.1 and 1.2). Euler surmised that such a path would require the number of linkages connecting a node be odd in number for either zero or two nodes. For Euler's Koenigsberg bridge problem, the graph possesses four nodes all with an odd number of associated linkages; thus, traversing all seven bridges would require crossing at least one bridge twice.

Graph theory's explicit treatment as a mathematical subject begins with König's (1936) seminal book on the subject, and since has progressed to develop a rich taxonomy of graph primitives, patterns, properties, and problem sets (Tutte 1984). As a foundation for assessing order within networks, Gilbert (1959) and Erdős and Rényi (1959) endeavored to formulate and characterize the properties of random graphs. Hypothesis testing compares the properties of a dataset, offered as an alternate hypothesis, against those of a corresponding random dataset, a null hypothesis

(Schabenberger and Gotway 2005, page 80). The formulation of random graphs provides a means to create null hypotheses in network studies. Random graphs thus allow for deeper examination of basic graph properties, and a means to compare expectations of order in real world networks.

Figure 1.1 The location of Euler's Koenigsberg bridges
Modified from Giuşcă (2005); basemap imagery from Google

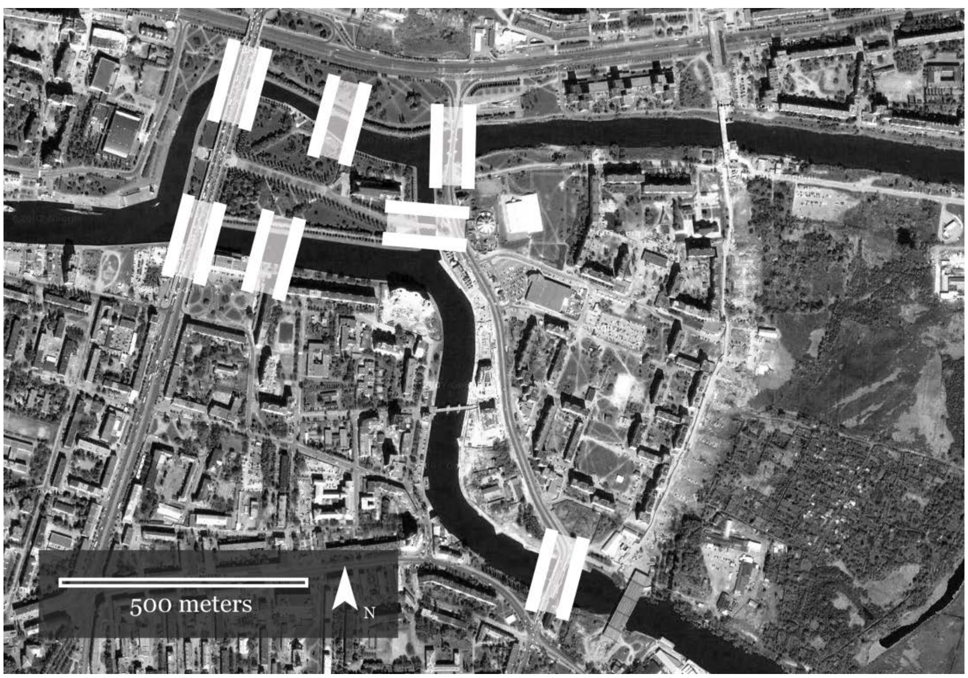
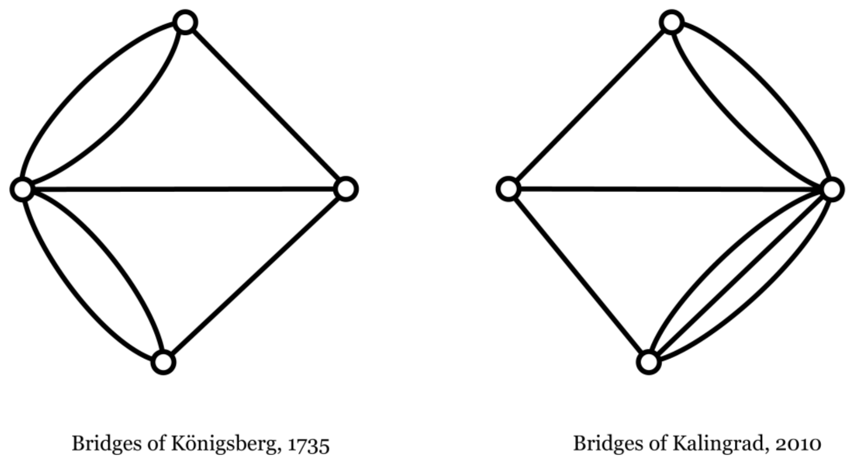
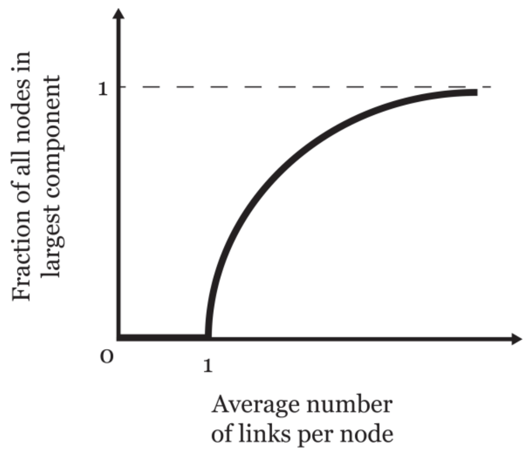


Figure 1.2 Graph depiction of Euler's Koenigsberg bridges
After Martin (2006)



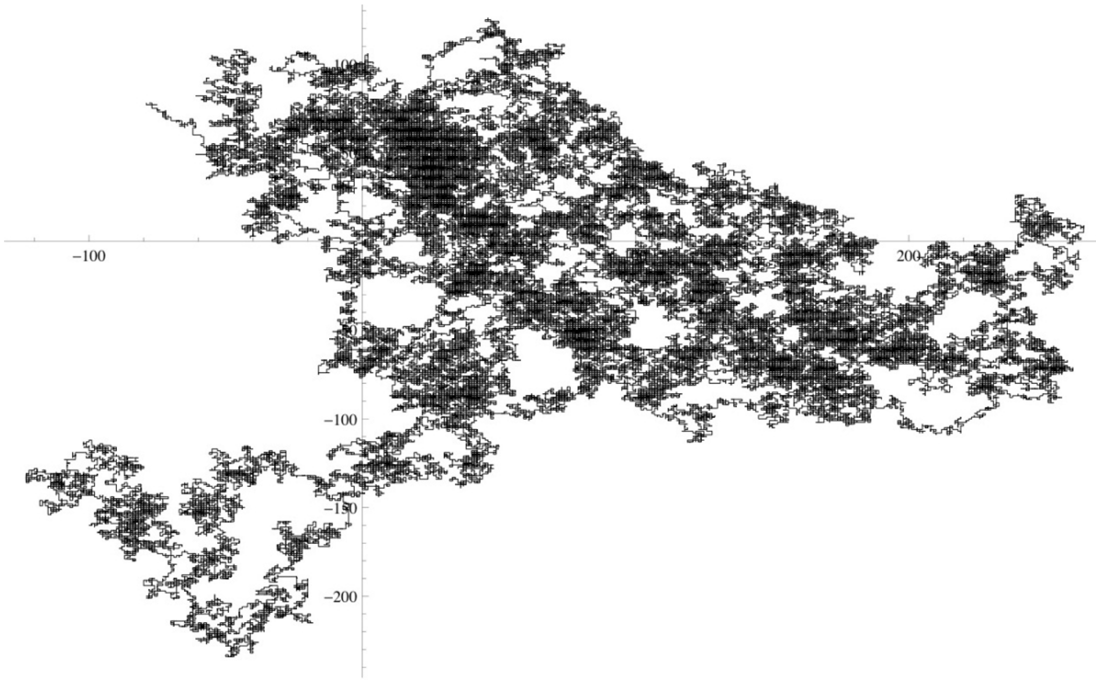
Graph theory also has grown in parallel with increased computing power to model, iterate, and analyze larger networks. Since the late 1990s, coupled with the work on random graphs, research has been conducted on the structure of very large networks and their emergent properties (Barabasi 2003). Ongoing efforts strive to characterize networks possessing heavy-tailed linkage distributions, such as the hub-spoke structure of scientific paper citations or the topology of webpage linkages (Price 1965; Albert, Jeong et al. 1999). Barabási and Albert (1991) note that by preferentially connecting nodes, a graph structure emerges with a degree distribution following a power law—a scale-free network. Such networks appear promising for describing many real-world networks, and work is underway to test the other generative explanations and applications (Dorogovtsev and Mendes 2002). In the creation of large random graphs or scale-free networks, researchers have found that the network structures sometimes yield qualities irreducible to the network's constituent parts (Watts 2003). These emergent properties result from the complexity of internal interactions within the graph itself (Watts 2003; Laughlin 2005). For instance, in a set of random nodes, creating linkages eventually reaches a critical moment at which disparate random graphs exhibit a phase change to a single random graph (Figure 1.3). The critical moment arises when the mean number of links per node reaches one. At that threshold, the fraction of nodes connecting the graph's largest linked component rapidly rises from almost zero to nearly one (Watts 2003, page 45). The rapid rise represents the change from disparate linked node pairs to a large highly connected network.

Figure 1.3 Phase change emergence within a random graph
From Watts (2003)



Emergent network structures also have been described within geography, notably in experiments with stream networks and random walks (Haggett and Chorley 1970). Random walks (Figure 1.4) are a formalization of the trajectory of successive random steps and used in many fields to describe movement and dispersion (Weiss 1994). A notable geographic example of their usage stems from the study of hydrology. Examining the morphometry of stream networks, random walk experimentation was used by Leopold and Langbein (1962) in their investigation of Horton's (1945) stream laws. By in-field measurement, Horton observed that stream lengths of differing orders tend to follow a geometric relationship. Shreve (1967) demonstrated that the approximation of geometric relationships was not necessarily indicative of orderly evolution, but here just as likely a result of random geometric convergence.

Figure 1.4 A random walk with directions seeded by the first 166,000 digits of pi
From Davis (2012)



Leopold and Langbein (1962) constructed simulations of stream networks where randomly originating streams were allowed to freely move stepwise in any of the four cardinal directions. Reverse flows were disallowed, and the walk ended upon converging with an existing path. Leopold and Langbein's results generated a dendritic network that satisfied Hortonian laws. In the case of random walks for streams, stochastic processes may create networks that appear orderly. As noted by Goodchild (1992), the patterns emerging from the random walk serve as a warning for spatial analysis;

careful null and alternate hypotheses must be formulated so as not to create an alternate hypothesis that is equivalent to an experiment's null hypothesis.

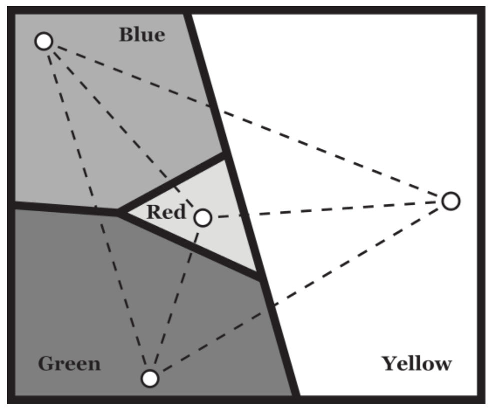
Many scientific fields have contributed to the deep literature on graphs and networks, including mathematics, economics, quantitative sociology, physics, and computer science. Internet-related social graphs have become a popular object of study for these disciplines. As platforms for user communication with friends, family, and other acquaintances, websites such as Facebook and Twitter track linkages and interactions among users. As of March 2012, Facebook cites over 900 million monthly active users (Facebook 2012). In Facebook, users that agree to link to one another are called *friends*. For Facebook users in November 2011, the mean of these friendship linkages is 190 (Backstrom 2011). As of July 2012, Twitter has 140+ million users contributing 340 million posts per day (Twitter 2012).

To promote value for their users, these websites allow third-party developers to program add-in features to the sites. These applications include opt-in affinity groups and multi-user games. The sites require developers to register and track their usage to minimize abuses. As a consequence of such third-party access, portions of the sites' social network database, including not only linkages, but often also attributes such as age, gender, and social preferences, are available for study. Further, the ubiquity of mobile devices with capabilities to collect positional data is likely soon to lead to widespread research correlating location data and social network properties. For instance, using anonymized mobile phone users, Song et al. (2010) tracked movement culled from communication towers and found likely user position to be highly predictable.

1.1 Geographical network analysis

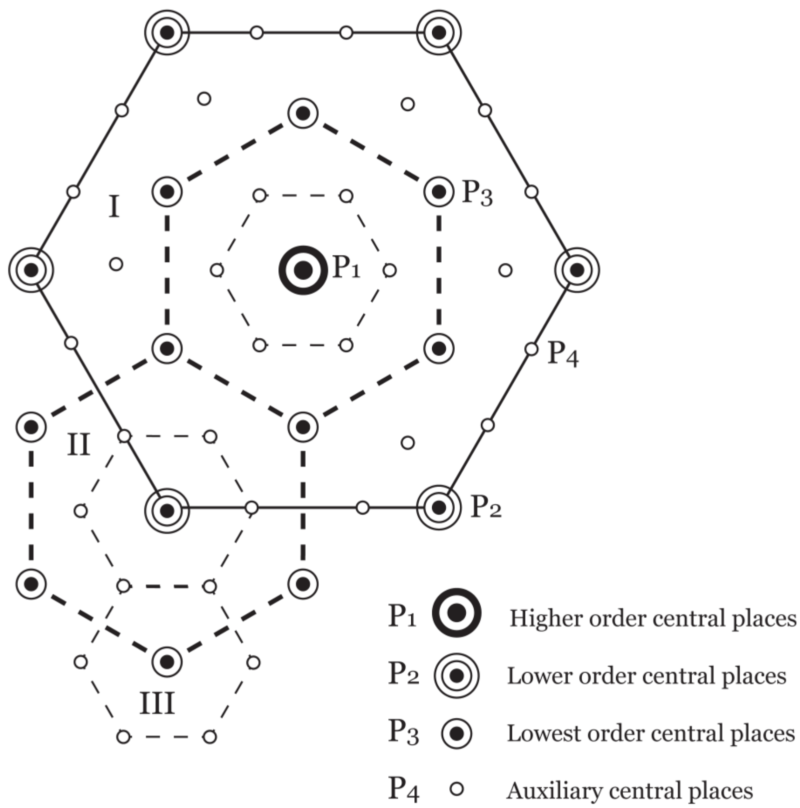
Early network problems, such Euler's Bridges of Koenigsberg and the four color map theorem (Cayley 1879), possess significant geographic components (Figure 1.5). Graphs were used as a modeling framework to solve problems operating upon an existing geographical system. The Central Place Theory was the first time network analysis was used describe the nature of the geographical system itself (Christaller 1933). Modeling the location, size, and quantity of towns, Christaller described regional subdivisions as ordered by functional hierarchy, positing that centrality is an ordering principle.

Figure 1.5 The four color map theorem
After Cayley (1879) and Inductiveload (2007)



Christaller's Central Place is a nodal point that serves its surroundings with services and goods; to describe the model and under the simplifying starting assumption that areas have similar characteristics in all directions, a place's surroundings are represented by a hexagonal coverage scheme (Figure 1.6). The theory hinges on the concepts of threshold and range. *Threshold* is the surrounding market area needed to maintain a goods or service provider, and *range* is the maximum distance that people will travel to obtain a service or goods. As Central Places serve their surroundings, the transportation linkages from a place to its hinterlands develop a network of many linkages but minimize overall length. Christaller's work was a precursor to similarly-themed efforts in regional science that arose twenty years later within the context of geography's quantitative revolution.

Figure 1.6 Central Place Theory
After DHV Consulting Engineers (1979)



Networks were a common object of study in the post-World War II period that brought the quantitative revolution to geography. At the University of Washington in the 1950s, William Garrison and his student-colleagues begin using numeric, statistical, and computational analytic techniques on a variety of geographic problems, including human movement, regional subdivision, and transportation. Their studies, including research by Brush (1953), Taaffe (1956; 1958a; 1958b; 1962), Garrison and Marble (1961), Nystuen and Dacey (1961), Kansky (1963), Taaffe, Morrill, and Gould (1963), and Tobler (1970) among others, provide a foundation for modern geographic network analysis. Brush (1953) describes the regional patterns of Wisconsin as a hierarchy, with respect to the Central Place Theory. Brush notes similarities and differences in settlement patterns in America and Europe, and suggests proceeding with a comparative study of socioeconomics to better weight the contributing factors. In this period, the earliest computer-drawn flow maps were created by the Chicago Area Transportation Study (1959). Nystuen and Dacey (1961) examined ordering of telephone-call flows, and described a hierarchy of dominant nodes. They noted that the phone flow hierarchy is based on a functional linkage, in this case, call activity, and not population. In Washington State telephone flows, dominant nodes were more likely to communicate directly, with smaller nodes connecting to the dominant nodes through medium-sized intermediaries. Kansky (1963) wrote *The Structure of Networks*, perhaps the first geography doctoral dissertation on networks. Taaffe, Morrill, and Gould (1963) constructed a network model for transportation development in undeveloped countries. Drawing upon Newton (1687), Ravenstein (1885; 1889), and Thornwaite (1934), Tobler summarized geographic interactions in his First Law of Geography: "Everything is related to everything else, but near things are more related than distant things" (Tobler 1970). Tobler's work on flows, interactions, and networks often simultaneously incorporated continuous and discretized space.

Emphasizing the need to include individual people and their activities in these quantitative analyses, Hägerstrand (1970) introduced a model of time geography. Hägerstrand argued that studying the time-space of individuals, particularly via constraints and affordances on geographic movement, provides a foundation for understanding human interactions and activities. Some of the conceptual frameworks from the 1950s and 1960s have been re-examined and implemented in modern computational environments; for instance, Hägerstrand's time-geography ideas have been implemented quantitatively, computationally, and in a network context (Miller 1991; Kwan 2000; Shaw 2006). Additional ideas germane to quantitative geographic network analysis stemmed from Cold War military strategic planning. In particular, the field of operations research devised concepts of cellular automata (von Neumann 1966; Gardner 1970) and routing optimization algorithms and heuristics (Dijkstra 1959; Moore 1959).

A comprehensive examination of network analysis in geography was conducted by Haggett and Chorley (1970) that synthesized the preceding decade's research. Their work largely drew upon examples in hydrology, transportation, and regional subdivision. Such cases remain the stalwarts of geographic network analysis, though implementations within digital information systems continue to evolve and augment the analytical landscape. The amount and availability of network data has increased; digital tools now exist to facilitate larger, faster, and more complex analyses; and, there has been the realization that the study of the design of human-created geographic data structures may assist in understanding the associated actual geographic objects or phenomena. The conceptual data structure humans organize for a geographic entity constrains and facilitates the questions that can be asked and how an entity is perceived. This idea is the underpinning for Geographic Information Science, a scientific pursuit standing astride the three domains of physical reality, human perception, and the computing environment (Goodchild 1992).

Theoretical and applied GENet research has amassed on parallel tracks in separate disciplines. Many inventories of geographical network problems and algorithms exist including those for GIS operations (Albrecht 1997; Zhan 1998), transportation (Miller and Shaw 2001), hydrology (Maidment 2002), geographic analysis (Haggett and Chorley 1970), operations research (Ahuja, Magnanti et al. 1993; Du and Pardalos 1993), and computational geometry (Mitchell 2000). No one inventory is all inclusive, but substantial overlap exists within these tracks. The disciplines' analyses reflect different facets of GENet understanding and accordingly rely on disparate data structures and operations. Through an examination of representational forms, data models, and analytical operations, the subsequent chapters attempt to identify the particularly geographic aspects of these approaches so they can be better understood and more effectively used.

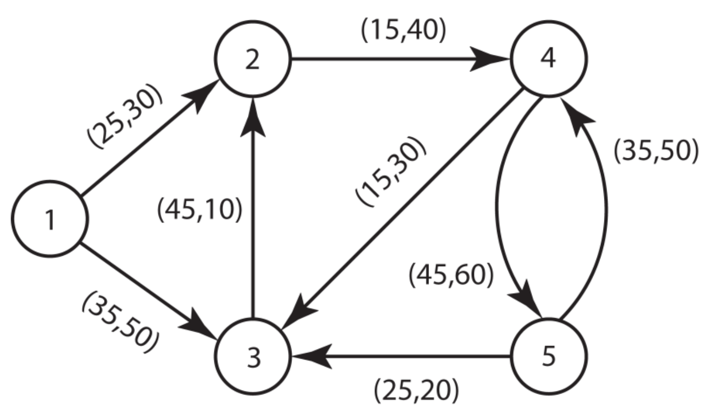
CHAPTER 2:
NETWORK REPRESENTATION

For geographic operations, networks are represented in mathematical, computational or visual form. Due to their shared numerical underpinnings, mathematical and computational representations are tightly coupled. Mathematical representations are used most often to express general or conceptual network properties, and computational forms are implemented when graphs are instantiated with actual, specific data. All but the most basic networks possess numerous arcs and intersections, and specific internal structures and relationships may be difficult for humans to disentangle when encoded as mathematics or computer code. Further, in geographic space, a network possesses not only internal relationships, but also interaction with the area it inhabits. Visual representations tend to be most effectively interpreted by people, but are less useful for automated computational interpretation and analysis.

2.1 Mathematical representation

The most common mathematical representation of a network (Figure 2.1) is the graph, described as $G = (V,E)$ (Agnarsson and Greenlaw 2007). A graph G is comprised of a set of E edges and two-element subsets of vertices V. That is, the edges each relate to two vertex endpoints. Graphs can be further delineated by describing the properties of edges and vertices. For instance, edges may possess directionality, magnitudes, capacity, speed limits, etc., or multiple edges may exist between two vertices. Numerous other graph properties have been mathematically described for specific circumstances. An example would be *signed* edges where one endpoint denotes a positive relationship while the other endpoint represents the negative. Connection between predators and prey could be denoted as a signed graph. Predators benefit and prey are harmed.

Figure 2.1 Flows represented mathematically
After Ahuja, Magnanti et al. (1993)



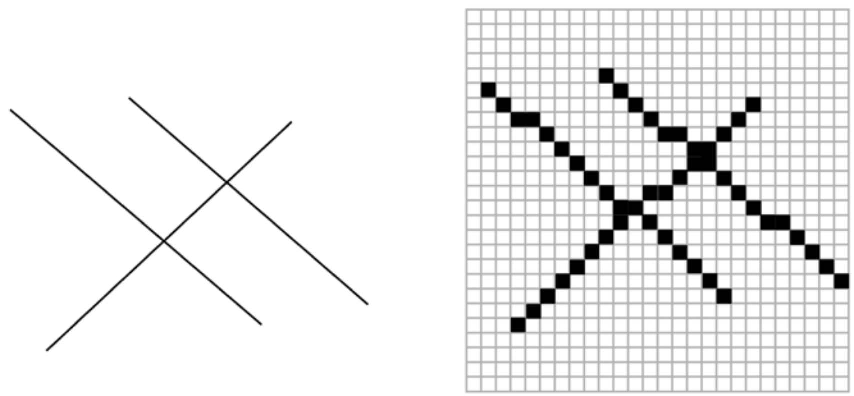
Topological graph theory is the branch of graph theory concerned with graphs embedded in surfaces and topological spaces. An example problem in topological graph theory is the three-cottage problem (Chartrand 1985): Three cottages exist on a two-dimensional plane. If each needs gas, water, and electric connections, is there an arrangement where the nine connections do not cross? No such arrangement exists in a strict, two-dimensional Euclidean plane. Topological graph theory leverages the constraints a surface exhibits on a graph, and thus is particularly relevant to networks embedded in geographic space.

Mathematical representation is often the most appropriate representation for communicating generic graph properties. Graphs instantiated with specific data may be more usefully manipulated with a computational representation.

2.2 Computational representation

Two common approaches exist for digital representation of network data (Figure 2.2), a raster model and an object-relational model (Couclelis 1992; Longley, Goodchild et al. 2005). In GIS, raster data types are those that discretize continuous space into cells, often rectangular or square. The form allows data storage as arrays of either single or multiple dimensions. The value of each cell describes an internal condition, affording comparison and differentiation from neighbors. For instance, cells that participate in a network could be labeled with values of 1, while cells that are not part of the network could be assigned 0. Such assignment for a network does not explicitly identify a network's node and arc features. In some instances, whole or partial conversion of a network represented as a raster to a vector object is possible. For example, considering cells that are labeled as streams (e.g. labeled with a 1), a query could track cells with two stream neighbors as part of an edge and cells with more than two stream neighbors as vertices. Additional logic might be required depending on the nature of the network and granularity of raster data. The vector result would typically be stored as a relational database. The raster form, nevertheless, is well suited for analyses associated with the interaction between the network and its surroundings. For instance, when deriving stream networks from a raster ground elevation model, a straightforward workflow is possible: first, aspects are derived from the elevation model. Then, each cell can be queried with respect to the number of cells uphill from itself, a measure of accumulated flow. Finally, this dataset can be tested against a threshold of accumulated flow. The result can be visualized or saved as a stream network dataset.

Figure 2.2 Vector and raster representations of networks



For the object-relational form, graph-type structures are the primary representational method, and vertex locations must be defined and related to edges. The object-relational form, also called the vector data model, represents geographic phenomena as discrete features. The arrangement of vertices and edges usually are stored within a tabular database or encapsulated within a computational object. The exact data structure varies depending on whether the data will be related explicitly as a computational object with inherent behaviors and relationships, or implicitly with tables or arrays containing key pairs. If key pairs are used, the implicit relationships are programmatically related as needed. With additional defined relationships, these databases also can be extended to include specialized attributes and behaviors (Arctur and Zeiler 2004).

For specific construction of arc, node, and numeric attributes, typical data structures include: node-arc incidence matrix, node-node adjacency matrix, adjacency lists, and forward star and reverse star matrices (Ahuja, Magnanti et al. 1993). The following few paragraphs are a cursory description of each approach. For GIS, each of the object-based representations requires the foundational declaration of node coordinate locations, which are often defined via arrays or in tables.

The Node-Arc Incidence Matrix is a representation composed of *m* columns and *n* rows (Figure 2.3). The data also could be held as a multidimensional array of size *m* x *n*. As such, for this discussion of data structures, the terms table and array will be used somewhat interchangeably. In a Node-Arc Incidence Matrix, the *m* columns represent defined arcs and the *n* rows denote nodes. Each cell of the matrix contains a value 0 for no connection, 1 for a forward connection, and -1 for a reverse connection. Since the rows and columns are dependent, a Node-Arc Incidence Matrix would be inappropriate for operations that frequently add and remove arcs or nodes. Such implementations, at least, would need to be prepared to handle restructuring with each new node change. The structure offers a useful form for the calculation of the minimum cost flow problem, though is not storage efficient in that it explicitly saves the “non-connection” data of all nodes and arcs (Ahuja, Magnanti et al. 1993).

Figure 2.3 Node-arc incidence matrix
From Ahuja, Magnanti et al. (1993)

Node-arc incidence matrix

$$
\begin{array}{c}
\\ 1 \\ 2 \\ 3 \\ 4 \\ 5
\end{array}
\begin{array}{c}
\begin{array}{cccccccc} (1,2) & (1,3) & (2,4) & (3,2) & (4,3) & (4,5) & (5,3) & (5,4) \end{array} \\
\begin{bmatrix}
1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\
-1 & 0 & 1 & -1 & 0 & 0 & 0 & 0 \\
0 & -1 & 0 & 1 & -1 & 0 & -1 & 0 \\
0 & 0 & -1 & 0 & 1 & 1 & 0 & -1 \\
0 & 0 & 0 & 0 & 0 & -1 & 1 & 1
\end{bmatrix}
\end{array}
$$

The Node-Node Adjacency Matrix is a square matrix where all nodes are represented as a row and then again as a column (Figure 2.4). The row represents a “from” condition and the columns a “to” condition. Connectivity is denoted as a 1 and lack of connectivity is denoted as a 0. The relationship between all nodes is stored in the structure, but this yields inefficient data storage if the network is sparsely connected. In contrast to the Node-Arc Incidence Matrix, arcs can be removed and added without change to the size of the matrix. The addition or removal of a node requires resizing the matrix by one row and one column.

Figure 2.4 Node-node adjacency matrix
From Ahuja, Magnanti et al. (1993)

Node-node adjacency matrix

$$\begin{array}{cc}
 & \begin{array}{ccccc} 1 & 2 & 3 & 4 & 5 \end{array} \\
\begin{array}{c} 1 \\ 2 \\ 3 \\ 4 \\ 5 \end{array} & \begin{bmatrix} 0 & 1 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 1 \\ 0 & 0 & 1 & 1 & 0 \end{bmatrix}
\end{array}$$

Adjacency Lists are composed of nested arrays where each node holds an array of its connected vertices (Figure 2.5). That is, each node possesses a list of nodes to which it connects. Additionally, the connecting nodes may hold a label or other attribute data for the arc. As an example, an adjacency list was created by van Rossum (1998) in a web post on defining and using networks in his Python programming language. His following code represents a network instantiated where A connects to node B via an arc and node C via another arc. Node B is connected to node C via and arc and node D via another arc, and so on.

Figure 2.5 Adjacency list
From Ahuja, Magnanti et al. (1993)

Adjacency list

Graph = { ’1’ : [’2’, ’3’] , ‘2’ : [‘4’] , ‘3’ : [‘2’] , ‘4’ : [‘3’ , ‘4’] , ‘5’ : [‘3’ , ‘4’] }

Adjacency lists offer efficient network data storage as only existing arcs are defined and stored. The condensed data storage however must restructure with any changes to arc or node membership. Unlike the previous matrix representations, no simple algebraic properties, like the simple calculation of node degree by row summation, are apparent. While Cherkassky et al. (1997) found that other representations are often more algorithmically efficient for common network analysis, Van Rossum (1998) demonstrates the form’s affordance of simple coding for network operations; Python code to find all paths between two nodes (Figure 2.6).

Figure 2.6 Python code to find all paths between two nodes
From Van Rossum (1998)

```
def find_all_paths(graph, start, end, path=[]):
    path = path + [start]
    if start == end:
        return [path]
    if not graph.has_key(start):
        return [ ]
    paths = [ ]
    for node in graph[start]:
        if node not in path:
            newpaths = find_all_paths(graph,node,end,path)
            for newpath in newpaths:
                paths.append(newpath)
    return paths
```

The Forward and Reverse Star representations possess a row for each network arc. In Forward Star, a row's first element is the identifier for the first node, the second element is the second node, and any subsequent elements are arc attributes. For networks with arcs that have directionality, so called *directed* networks, the first and second elements refer respectively to an arc's *from* and *to* nodes. Related, a Reverse Star Representation stores the arc's tail as the first element and the head as the second element (Figure2.7). Element reordering is simple, so the computational burden for the two representations is equivalent. While Ahuja, Magnanti et al. (1993) assert that the Forward Star representation is more space efficient than Adjacency Lists, conceptually, adjacency lists are more compact that Forward Star in that head nodes need not be redundantly stored. The Forward and Reverse Star representations do offer an advantage in that when changing an arc or node, only a change in the number of rows is required.

Figure 2.7 Reverse star representation
From Ahuja, Magnanti et al. (1993)

Reverse Star

	point
1	1
2	3
3	4
4	5
5	7
6	9

	tail	head	cost	capacity
1	1	2	25	30
2	1	3	35	50
3	2	4	15	40
4	3	2	45	10
5	4	3	15	30
6	4	5	45	60
7	5	3	25	20
8	5	4	35	50

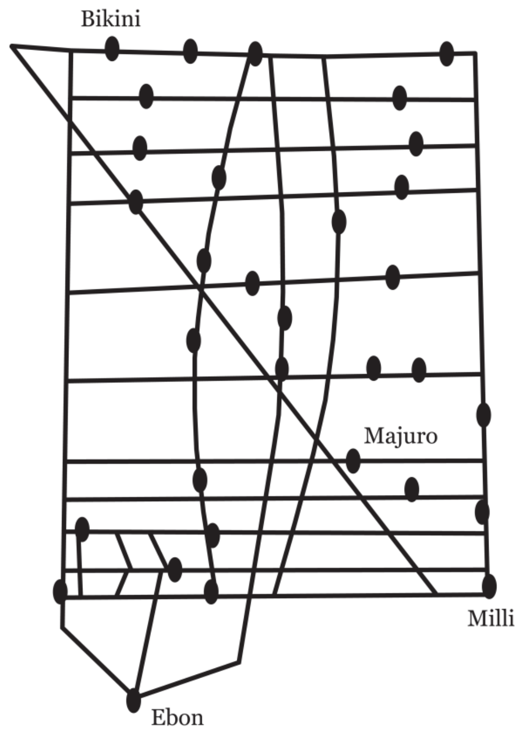
2.3 Visual representation

Visualizations foster human interpretation. The visual display of GENets commonly takes the form of graphs, schematics, overhead imagery, and various types of cartographic maps. A GENet visualization may aid navigation, tell a story or history, convey patterns, speculate on what will or could be, or serve as an artistic, creative expression. Further, visualization may clarify a system's fuzzy or uncertain boundaries or simplify data via aggregation. A trade network map combines numerous interactions from multiple locations and simplifies them into a smaller set of arrows and curves. Similarly, a map of a slot canyon or mine represents a network of "negative" space, the absence of rock creating traversable pathways. There is no way to "see" an entire underground mine complex without the aid of a visualization.

A well-designed visualization's purpose will align with its intended analytical purposes. Network visualization occurs on physical media or in digital form, and can possess static or dynamic characteristics. Visualizations on physical media include articles like paper maps, drawings, carvings, and scale models. Digital visualizations include maps and models viewed on electronic displays.

Visualizations developed on physical media with no moving parts are perhaps most suited for comparison analysis. Using a static, paper map, for

Figure 2.8 Marshall Island stick chart
After Genz, Aucan et al. (2009)



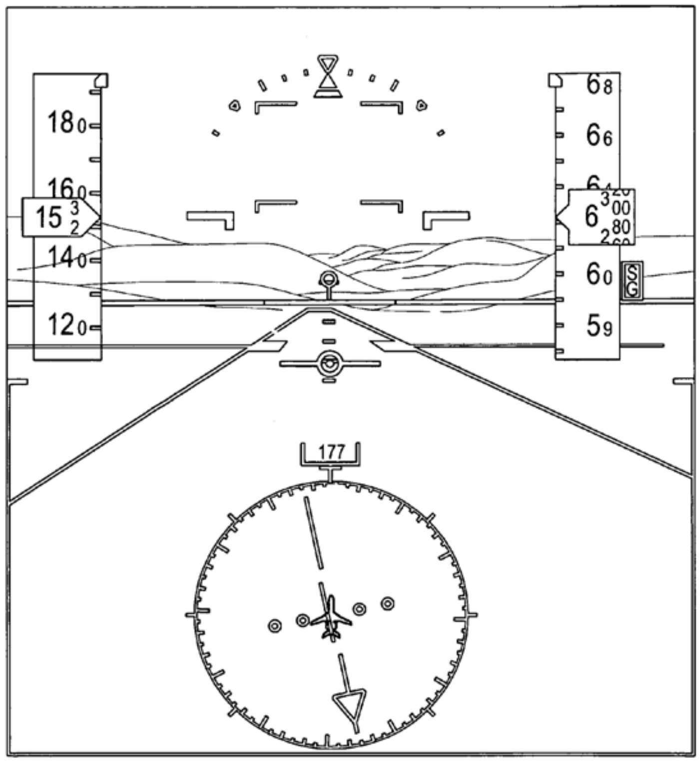
instance, iterative cycles conducive to simulation or optimization are unavailable. While static maps generally are not well suited for process modeling, thoughtful design may convey information about dynamic behavior. For instance, Marshall Island stick charts (Figure 2.8) use the linear and curvilinear arrangements of sticks and shells to signify not just geographic positions and routes, but also characteristics of ocean waves disrupted by islands (Genz, Aucan et al. 2009).

A physical representation does not have to be static. It could be made dynamic via moving parts or with interactive overlays, thus affording analytical operations. Examples include paper maps with translucent overlays and scale models. For instance, from the 1940s to 1980s, the United States Corps of Engineers maintained a 200-acre scale model of the Mississippi River Basin west of Jackson, Mississippi (Cheramie 2011). The abandoned Mississippi model is located at 32.305°N, 90.315°W (Figure2.9). The model acts as both a visual representation and an object for process model experimentation. The model's behavior was proportional to responses within the real watershed.

Figure 2.9 Aerial view of the Mississippi Basin Physical Model
Looking east from Kentucky Lake and Barkley Lake, Kentucky, circa 1965. From U.S. Army Corps of Engineers (2012)

Contemporary computer visualizations occur on thin film transistor liquid crystal displays (TFT-LCD) using raster graphics (Castellano 2006). TFT-LCD monitors discretize the screen area into a rectilinear array of tightly packed picture elements (pixels). The liquid crystal array serves as a polarizing shutter that allows light to pass through red, green, and blue colored filters at controlled intensities. All computer visualizations, no matter whether their conceptual underpinning is vector or raster, must be translated into the TFT-LCD raster model. Fortunately, this data translation is controlled by display rendering libraries and when combined with contemporary fine resolution displays, need not be a significant concern for the visualization author. Vector monitors do exist but are less common. Such screens plot an electron beam across a rapidly fading phosphor. The technology is used in oscilloscopes and some early video games. A few of these games included networks embedded within a geographic-like space. For instance, the 1980 video game Star Castle by Tim Skelly and Scott Boden of Cinematronics used a series of concentric, rotating boundary networks to fortify an enemy base. Due to the ability to create high contrast renderings visible in diverse lighting conditions, even bright sunlight, vector displays also were used in early generation fighter pilots' head-up displays (HUD) (Naish 1964) (Figure 2.10). LCD displays are now ubiquitous in aviation, but cartographic elements such as distinct lines, sparse design, and pale green coloration remain from the vector user interface (UI) aesthetic.

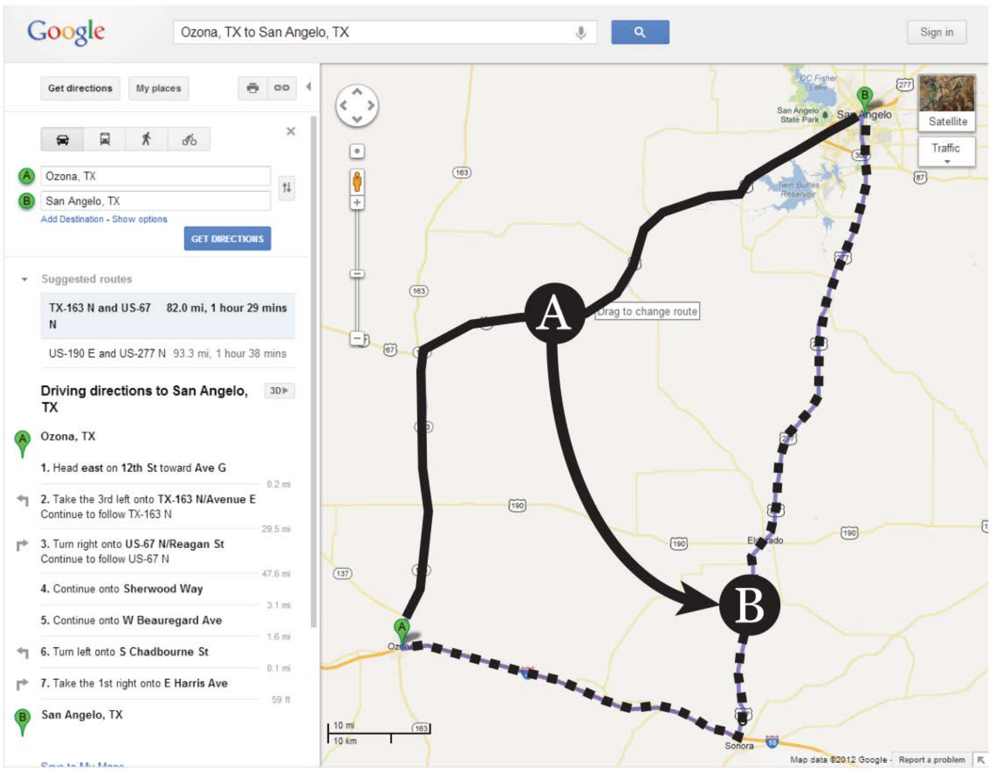
Figure 2.10 Aircraft head-up display
From Nichols and Little (2011)



Similar to paper maps, most computer displays are rectangular. Computerized three-dimensional images can be created by transmitting different images to each eye. Such technology is widely available using either polarizing glasses or limited angle autostereoscopic display. Unlike a physical map, computer visualizations may allow the viewer perspective to be mobile. A moving perspective allows the network to be seen from different angles and is particularly useful in apprehending complex 3D networks.

Dynamic computer visualizations support simulation, optimization, process modeling, and comparison. Dynamics in visualization can occur as the network or perspective changes. Until recently, dynamic visualizations would have been synonymous with animation. However, with more powerful computational platforms, dynamic visualizations now are as likely to be interactive. User interaction can be facilitated via Computer-Human Interaction (CHI) techniques such as pull-down menus, dashboard controls, or direct user manipulation of network arcs and nodes. Navigational visualization is a highly developed area of interactive displays on computational media. Google Maps' Directions functionality offers an illustrative interactive workflow (Google 2012) (Figure 2.11). A common dynamic perspective change is the piece by piece delivery of a road map during navigation with a GPS.

Figure 2.11 Google Maps driving directions
After Google Maps calculates a route (shown by A), a user can drag the route line along the road network to change the intermediate route between nodes (shown by B).



2.4 Visualization components

Graphic components used to create GENet visual display stem from the fields of cartography and data visualization, as well as scientific and engineering domains that have devised workflow-specific symbology. GENet visualizations may include the full breadth of features of any cartographic or spatially enabled rendering. The complexity of cartographic design stems from endless possible graphic elements available and their interplay. Defining the cartographic possibilities for any single map component can be made simpler if features are considered as vector and raster data. At the most basic, visual displays are composed of sets of object-centric primitives like points, lines, and areas, and/or field-centric primitives like grids, images, or textures. That small set of primitives is constrained by the finite set of variable visual characteristics possible on that particular feature. To start, components can be varied by geometry, color, and transparency. Layering and grouping features and variables endows the visualization with meaning.

On physical and computer visualizations, the most common graphic representation of connectivity is two dots connected by a line. Numerous other conventions exist for networks in geography, with most following techniques generalized for line symbols (Brewer 2005). If the relationship is not the same in both directions, like the flow of water down a channel, an arrow symbol commonly is used. The width of the line may denote the magnitude of the relationship. A large river may be symbolized with a wide line, while a small creek represented with a thin line. Temporary or speculative connections are often symbolized by dashed lines, with lighter colors, or as translucent. For accumulating of attenuating magnitudes, the width of the line may change along its length. The drawn line need not be straight. Curves can be used to denote an indirect path, show a generalized route, or promote cartographic clarity. In cases where the link is meaningful but endpoints not, the points may be omitted.

Network visualization types such as graph diagrams, schematics, and maps each facilitate varying interpretive and analytical operations. A graph diagram is symbolized using vertices and curved or straight lines. Their components tend to be unambiguous, and all extraneous information is removed. For instance, partial lines or complex node symbolization are uncommon. In the typical case, vertices are of similar character, and linkages express the same type of relationship. For GENet visualization, nodes tend to be positioned in the physical space, while arcs may represent physical or conceptual links. A graph diagram emphasizes basic relationships. The system's internal patterns and global structure may also be visible. The general purpose of a schematic diagram is to emphasize function or process (Esri 2006). In many schematics, linkages retain their relationship type, while vertices possess varying functions or characteristics. The archetype is the electrical diagram, with its description of an electrical circuit's components (Figure 2.12). Circuit symbols include the abstracted route of the electrical current and linked transformative agents such as voltage sources, resistors, and switches. Schematic visualizations of GENets may distort distances in order to emphasize network topology. For instance, the Tube Map of London (Figure 2.13) distorts scale to highlight the relative positions of stations (Garland 1994; Zhan 2011).

Figure 2.12 A simple electrical circuit

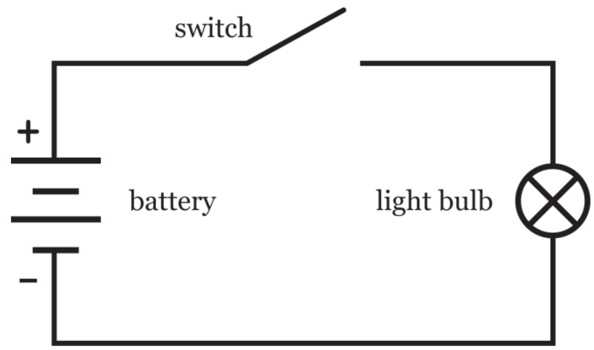
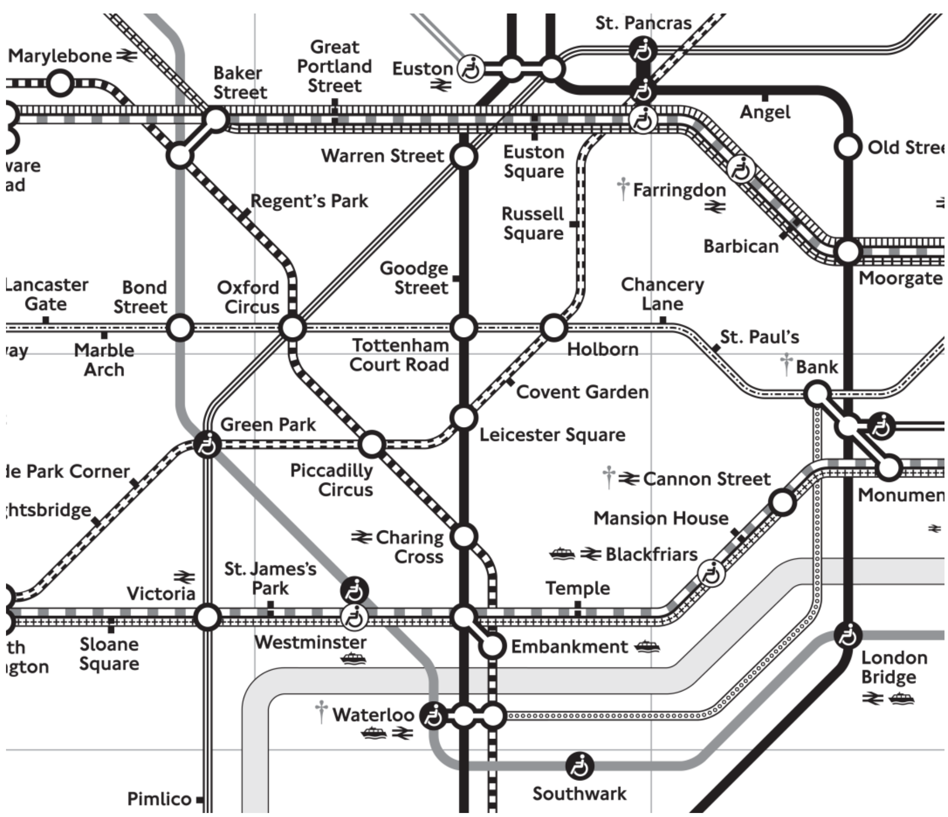


Figure 2.13 Map of the London Underground
From Transport for London (2012)



Cartographic maps use a wide array of visual possibilities to serve myriad purposes. With respect to GENets, maps of roads and rivers are perhaps the most common and have been the object of cartographic art and science for centuries. Network features such as connectivity, magnitude, capacity, and direction are common themes, and as discussed earlier, cartographers strive to communicate such characteristics using techniques like differing line widths, color intensities, patterns, and translucency. A cartographic representation differs from a schematic or graph in its connection to a physical setting. Graphs emphasize structure; schematics emphasize function; and maps emphasize context. On a map, the location of a network matters. Issues like scale, distance, and orientation affect the function and meaning of any mapped network. Also though, a network in geographic space can affect and be affected by its environment's social and physical landscape.

CHAPTER 3:
MODELING GENET FLOW

3.1 Modeling GENet flow

Geographic movement often is concentrated along network paths. In GIS, movement of individuals may be simple and implemented as navigational or tracking maps. Collective movement—flow—is, however, considered more complex and enigmatic. The purpose of this chapter is to develop a GIS data model for GENet flow and interpret its organization with respect to the general properties of GENets. A secondary purpose is to describe a method for creating such data models. As noted in the previous chapter, representational and analytical approaches for GENets are numerous and diverse. Flow has been chosen as an initial situation within a large population of possible GENet variations and circumstances. As its occurrence is widespread and its inner workings not yet well defined, GENet flow offers a good starting point for modeling.

Creating a GIS data model requires reducing a phenomenon to its basic conceptual elements and offers a formalization of a phenomenon's ontology. The model provides meaning to primitive GIS elements such as points, polylines, and polygons; relates their attributes; facilitates data collection and sharing; and affords the building of analytical and representation functionality. The first part of this chapter includes a brief discussion of object-oriented (OO) concepts and the study phenomenon GENet flow. The chapter's second section presents an approach for geographic data model development and follows the creation and validation of a data model for GENet flow. The chapter concludes with a discussion of GENet properties garnered from the model and areas for future research.

Previous researchers have conducted a tremendous amount of work on geographic subjects related to this paper's focus, including object-oriented GIS (Worboys, Hearnshaw et al. 1990; Egenhofer and Frank 1992; Raper and Livingstone 1995; Balram and Dragicevic 2006), database engineering (Ralyté J, Deneckère R et al. 2003; Worboys and Duckham 2004), dynamics and representation (Peuquet 2002; Goodchild, Yuan et al. 2007; Lohfink, Carnduff et al. 2007), ontologies (Couclelis 1992; Smith and Mark 2001; Fonseca, Egenhofer et al. 2002), model normalization (Date 1995; Miller and Shaw 2001), and relational operators (Goodchild, Haining et al. 1992; Albrecht 1997), among others. This paper provides a procedure for developing geographic data models, but it is not intended to discount or replace the aforementioned endeavors. It is a starting point for spatial data modelers to define, organize, and formalize their geographic entities of study so they can engage in those broader efforts. It is further intended that by enumerating an example process of geographic data modeling, this work will foster discussion of opportunities and mechanisms for modeling the world in more useful, possibly innovative, ways.

3.2 Object-oriented concepts and geographic data models

Object-oriented engineering describes the computing paradigm of grouping, organizing, and associating related data and functionality (Jacobson 1993; Weisman 2003). In modern GIS, object-oriented data structures have become prevalent over topological map-based structures. Map-based data structures, like the Environmental Systems Research Institute (Esri) coverage format, allow robust topological queries, but points, polylines, and polygons are constrained to homogenous behavior (Zeiler 1999). It should be noted that Esri now favors object-based formats and its associated shapefile format is widely used as a geographic data standard. The object-based concept allows spatial and other data to possess behavior (encapsulation), acquire common characteristics (inheritance), and differentiate actions based on input (polymorphism) (Wegner 1988; Rumbaugh, Blaha et al. 1991).

For GIS, object-oriented concepts allow users with specific interests to assign meanings, relationships, and behavior to map features. Accordingly, Maidment defines a geographic data model as an OO-based structure for organizing geospatial data and relating it to GIS cartographic elements

(Maidment 2002; Maidment 2003). Geographic data models can 1) assist in identifying how information is organized and stored for a study phenomenon, 2) organize data in a standardized way for specific problem types, 3) provide users with a common database design template that can be populated with their project data, and 4) promote an environment where domain-specific functionality can be more easily integrated (Arctur and Zeiler 2004). When thoughtfully designed, geographic data models facilitate data exchange and promote an understanding of a spatial phenomenon's organization.

Within computer science, an *ontology* is a conceptual scheme that enumerates and describes the relationships and rules for a particular domain (Worboys and Duckham 2004). Thus, the ontology and geographic data model are, or at least should be, tightly coupled. Geographic data model development relies on the ontology for its logic, and the model itself may help illuminate a domain's previously unrealized order and relationships. The ontology and data model both guide relevant, logical data analyses, offer a mechanism to evaluate database capabilities, and facilitate the assessment of structural order.

3.3 Characterizing GENet flow

Geographic change often is manifested by channelized movement, particularly aggregate movement. Flows—the movement of collectives, like people, materials, or ideas—are dynamic, and generally lack standard functionality in contemporary GIS (Tobler 1987; Yuan 2001; Cova and Goodchild 2002; Tobler 2003; Goodchild and Glennon 2008). Network-related collective movement is widespread in the abstraction of the social and natural world: groups of people move from one place to another; money moves from buyer to seller; water molecules move *en masse* down valleys as rivers; and, flocks of birds migrate from place to place. These types of aggregation provide a mechanism to observe trends and motion without tracking every individual in space-time. Due to its widespread use, flow's meanings are broad; for instance, the flow of a river differs in meaning from the movement of ideas. Fortunately, because the flow concept often has a spatial anchoring, many cases can be represented computationally as a georeferenced object, field, or hybrid of the two. Objects represent spatial data as discrete entities, and fields represent data as continuously varying through space. Hybrid field-objects and object-fields for digitally representing spatial dynamics are described by Cova and Goodchild (2002) and Yuan (2001). Goodchild, Yuan et al. (2007) later offer a unified perspective on these geographic representations.

Characterization of flow requires identifying the actor that is moving and a behavioral description. Such a description typically identifies the flow's pathway when known, a beginning location, end position, and a metric of flow magnitude. Further examination of flows yields two prominent conditions: *steady* and *transitory* flow. Steady flows are aggregate movements that have been completed or exist in a stable state. Examples of steady flow could include a database of interstate human migration or a map of a river. As steady flows tend to exist within a single time step or possess a simple temporal scheme, these instances generally fit within regular GIS representational frameworks. Transitory flows exist in a state of unfolding motion (Lowe 1998; Sider 2001; Grenon and Smith 2004). Two examples of transitory flow analyses include prediction between known flow space–time locations (interpolation) or criteria-based anticipated flow reactions (predictive modeling). Transitory flow data typically thus require handling of time. Though less supported in the core functionality of mainstream GIS software, in the GIScience community, ongoing efforts by Peuquet (2002), Raubal, Miller et al. (2004), Yuan (2009), and Miller and Bridwell (2009) discuss storing and querying spatial data's relationships with respect to unfolding time.

The chapter continues with the creation of a model for GENet flow, offering an applied approach for developing geographic data models and enumeration of GENet characteristics apparent from the model. The model development procedure, based on examining and deconstructing the organization of various use cases, is intended to offer a general method for developing data models suited to any spatial phenomenon—tangible or abstract.

3.4 A use case approach to geographic data model design

Describing and analyzing use cases to facilitate system design are common practices within the computer science community (Weisman 2003), and can be readily adopted for creating geographic data models. According to Larman (1998), use cases "are stories...; they illustrate and imply requirements in the stories they tell." In designing a data model, use cases provide a starting point for assessing necessary classes, associations, and methods, and infrastructure requirements for common queries. For example, with respect to GIS, some typical use cases might include the process a taxi driver uses to drive from an origin to destination, assessing a river's vulnerability to pesticide pollution, or identifying hidden enemy gun placements on a battlefield. Maps and cartographic databases also may provide useful cases, as cartographic representations give a rich narrative of components and relationships within a system.

The procedure offered here to design a geographic model is: 1) identify the components of disparate use cases and schematize them into UML; 2) distill the resulting models into a single generic UML structure; and 3) then test and validate the generic result. Throughout the model creation process, the designer should consider the typical queries the model should be able to answer. For GENet flow, example queries might be: which direction is the flow moving; what is the magnitude of flow at a location along a pathway; or what is the difference in magnitude at one location compared to another? Such example questions can offer assistance during the design process by providing a test by which individual decisions can be gauged (for example, if a specific design decision is made, will the model still be able to answer the question regarded as important?).

3.5 Use case descriptions

Three use cases were selected, and effort was taken to maximize their variety. The cases all stem from real world datasets though their data sources are markedly different; data include a simple from–to table, a map, and a more complex GIS database. The types of associated GENets appear different, with both physical and less certain routes. The cases selected were: the twenty largest state-to-state migration flows, 1995-2000 (U.S. Census Bureau 2003), Napoleon's march to Moscow, 1812-1813 (Minard 1869); and karst flow routes near Mammoth Cave, Kentucky (Ray and Currens 1998a; Ray and Currens 1998b; Glennon and Groves 2002).

The human migration case data originate from a 2000 U.S. Census table of the 20 largest net state-to-state migrations (U.S. Census Bureau 2003) (Table 3.1). The case exhibits flow with a magnitude that remains constant along its time step. The flows' origin and destinations are known, but route geometries are uncertain. Cartographers often represent such

Table 3.1 The 20 largest state-to-state migration flows: 1995-2000
From U.S. Census Bureau (2003)

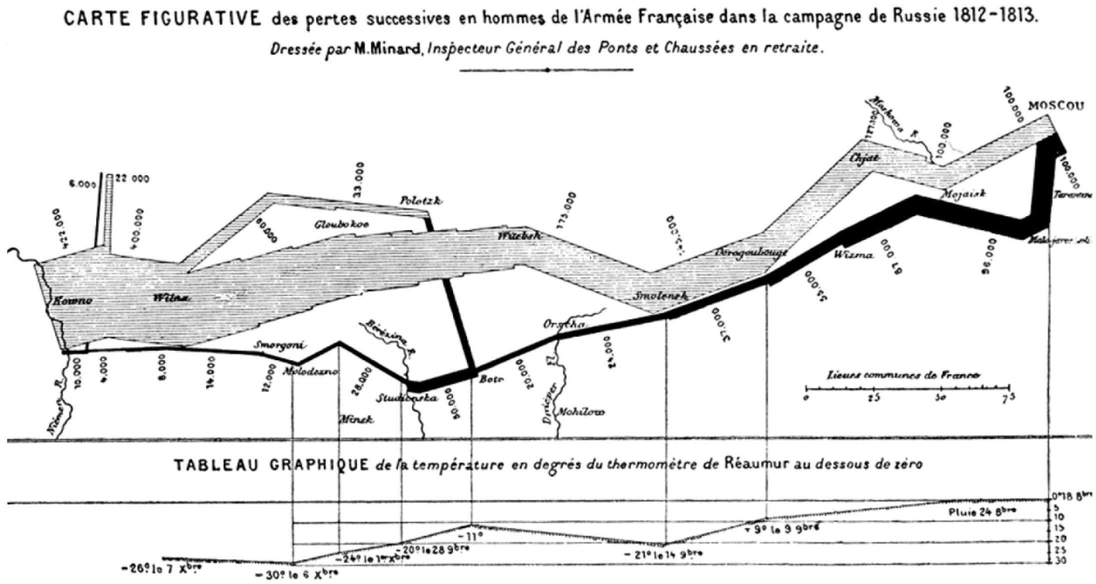
State of Origin	State of Destination	Migration Flow	Reverse Flow
New York	Florida	308,230	70,218
New York	New Jersey	206,979	97,584
California	Nevada	199,125	60,488
California	Arizona	186,151	92,452
California	Texas	182,789	115,929
Florida	Georgia	157,423	99,225
California	Washington	155,577	95,469
California	Oregon	131,836	67,642
New Jersey	Florida	118,905	34,896
Texas	California	115,929	182,789
New York	Pennsylvania	112,214	67,213
California	Colorado	111,322	56,050
New Jersey	Pennsylvania	110,436	88,202
New York	North Carolina	100,727	20,262
Georgia	Florida	99,225	157,423
New Jersey	New York	97,584	206,979
Florida	North Carolina	96,255	57,564
New York	California	95,952	65,160
Washington	California	95,469	155,577
California	Florida	94,265	65,211

note: numbers are U.S. Census estimates from sample data.

interaction tables as maps of lines and arrows to represent flow, but such interpretation is not inherent in the source data. Other implicit data within the dataset include direction and gross and net magnitudes. Tobler (Center for Spatially Integrated Social Science 2005) and Glennon (2010) developed software to create flow arrow maps from tabular interaction matrix data and calculate some of these implicit characteristics, and Maidment (2002) has developed supporting software for associated cases in hydrology.

For the second use case, the data source is a cartographic map: Charles Minard's map of Napoleon's 1812–1813 Moscow campaign. The source was selected for its concise display of a large amount of data (Minard 1869) (Figure 3.1), and because it has been lauded for its elegance in the visual display of quantitative information (Tufte 2001). Minard shows an army of 422,000 withering in battle, inhospitable weather, and sickness, then returning with only 10,000 troops. The map graphically represents the army's attenuation in two ways: 1) as the march progresses, the line representing the army's size proportionally decreases in width, and 2) at several space-time locations, an abrupt change in line width represents a large number of casualties from battles or harsh weather. Along the route, portions of the army detach and later reattach to the main column. Minard chose a binary shading scheme to represent the army's differing paths in advance and retreat. Along the march, the army does retrace its steps on one city-to-city stretch. The retreating portion of the campaign is accompanied by a timeline with the location and corresponding temperature. Several cities, rivers, and longitude lines provide spatial reference. Unlike the human migration case, Napoleon's march path and geometry are known per the locations provided on the map.

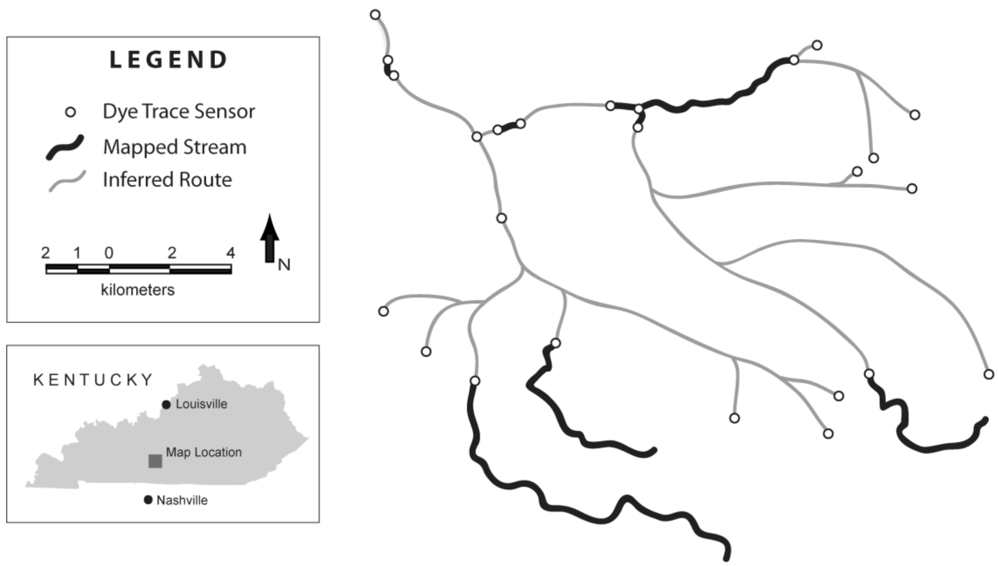
Figure 3.1 Napoleon's march to Moscow, 1812-1813
From Minard (1869)



The third use case uses a GIS database of stream networks within a portion of the Mammoth Cave watershed in south central Kentucky—the Turnhole Karst Watershed. A karst watershed is characterized by caves, sinkholes, and sinking streams (Ford and Williams 1989); the stream network is comprised of interconnecting surface and subsurface streams within caves. In the 245 km^2 Turnhole Karst Watershed, numerous surface streams disappear into the subsurface, feeding an underground stream network (Quinlan and Ray 1989; White and White 1989). In the subsurface, these streams converge and flow toward a spring outlet on the surface. While some of these subsurface streams have been explored and mapped inside caves, other routes are uncertain. Classification schemes to describe stream segment's participation and relationship to the entire network have been developed by Howard and others (Howard 1971; Glennon 2001; White 2003). Where routes have not been mapped, hydrogeologists use a technique called dye tracing to determine the fate of the groundwater. When an underground stream follows a course that is impassable by humans (following a cave passage too small or too dangerous), non-toxic dye is poured into the water. Then, nearby springs, caves, and water bodies are monitored for the dye.

When the dye is detected, it reveals a link between the input and output locations, though the exact flow path geometry is not known. Several hundred of these dye traces have been conducted in the Turnhole Karst Watershed, allowing hydrogeologists to construct a map of the subsurface network geometry revealing a branching flow pattern (Quinlan and Ray 1989; Ray and Currens 1998a; Ray and Currens 1998b). Such dye trace linkages provide only a point-to-point schematic of the uncertain flow paths. For the karst use case, the database includes both mapped stream networks and the dye traces' less-certain paths (Figure 3.2). Data come from the Kentucky Geological Survey's karst watershed hydrologic data for the Campbellsville and Beaver Dam 30 x 60 quadrangles (Ray and Currens 1998a; Ray and Currens 1998b). Surveyed subsurface stream data was also used (Glennon and Groves 2002). The data portray both mapped and dye-traced flow routes within the karst watershed.

Figure 3.2 Karst flow routes at Mammoth Cave, Kentucky
From Glennon and Groves (2002); modified from Ray and Currens (1998a); Ray and Currens (1998b)

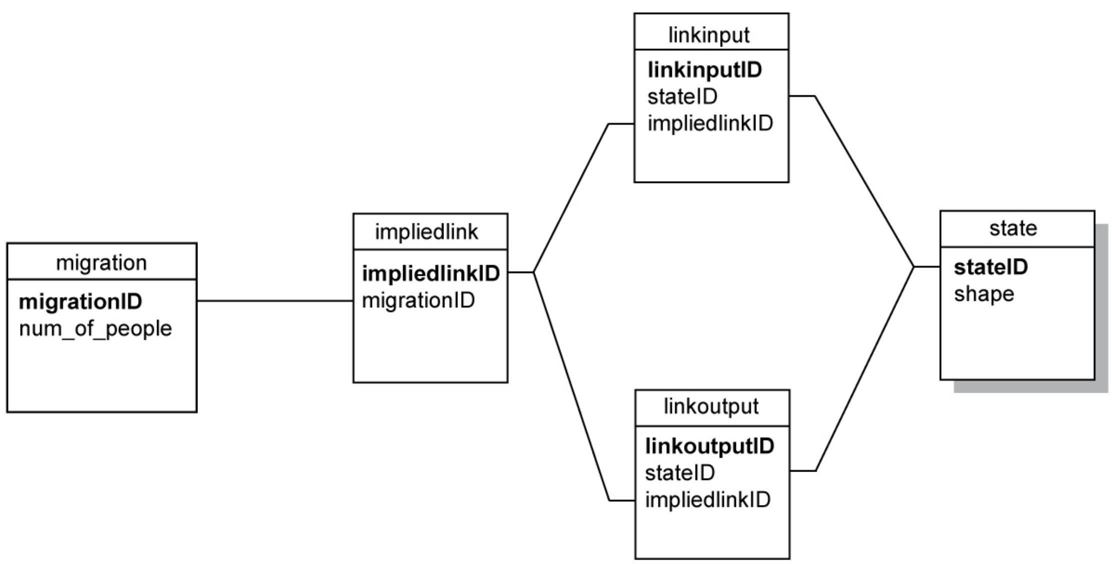


3.6 Developing UML for the use cases

A useful UML diagram should be able to hold a schematic of all the necessary data and relationships to re-create the case within a GIS. The point-of-entry for the model design process is not necessarily obvious, and a useful starting point often is to catalog all map features (i.e., points, polylines, and polygons) and list their associated meanings within the use case. This operation can be followed with an inventory of feature attribute data and consideration of how these attributes relate to their associated map features. The model designer can use abstract relationship classes and linkages to define more complex associations. Methods, operators, specifications, and examples for delineation of classes and relationships can be found in UML specification documentation by Alexander (2002), Ambler (2003), Fowler (2003), Arctur and Zeiler (2004), and others. Constituents and relationships were examined and a corresponding UML diagram developed for each of the GENet flow cases.

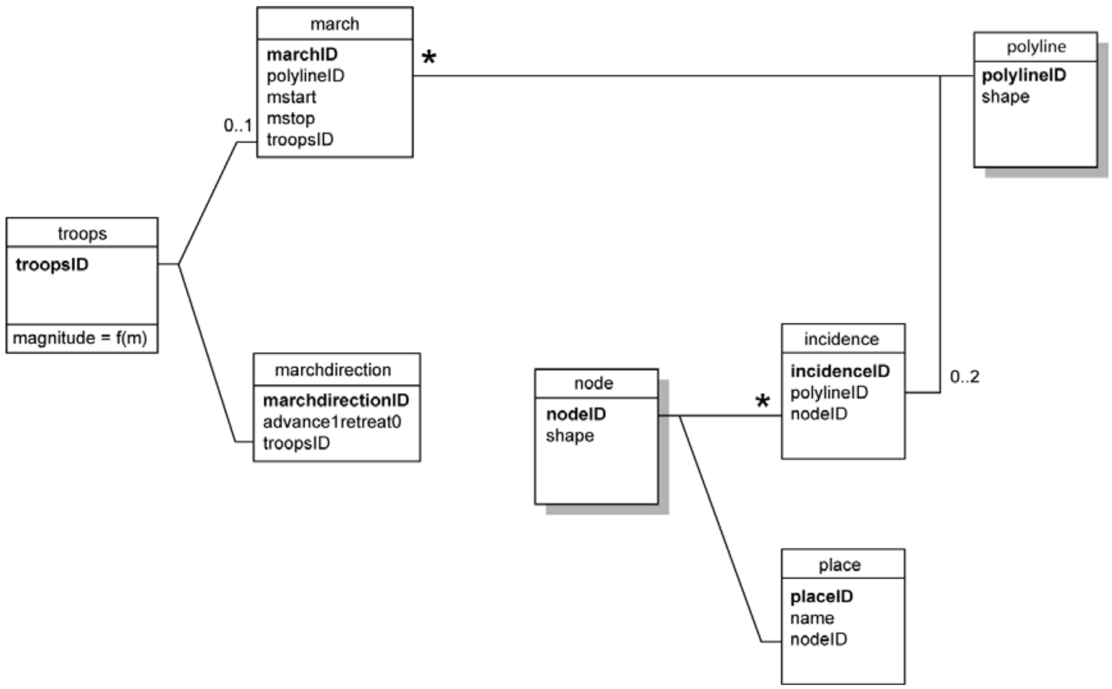
For the state-to-state migration use case, classes to describe the number of people migrating, *from* and *to* state locations were created (Figure 3.3). For simplicity, the states were represented as point locations. Since the migration table does not provide the exact path of the flow or even unambiguous begin or end locations, the case UML does not account for path geometry. Several relationship classes were developed to hold information about a `state`'s association to either a *from* or a *to* location. The relationship classes developed were: `impliedlink`, `linkinput`, `linkoutput`, and `migration`. Each of these classes possesses its own numeric identifier. Other than this identifier, `migration` holds a measure of the number of people migrating. The `impliedlink` class is the connector between *from* and *to* locations and the flow magnitude. `Impliedlink` refers to the flow class, which holds the flow magnitude. The `linkinput` and `linkoutput` classes each associate with a node and an `impliedlink`. Together these classes form a model able to contain all the data for the use case: a flow magnitude, input locations, and output locations.

Figure 3.3 Human migration data model
From Glennon (2010)



For the Minard map, a similar inventory of case components was undertaken (Figure 3.4). However, unlike the migration case, the source raw data were unavailable. Relevant components and data thus were culled and interpreted from the map. The map shows Napoleon's physical routes, so network geometry is a necessary part of the model. Also, some army units detached and reattached during the course of the campaign, so path intersections must be possible. During the march, army strength exhibited both gradual and abrupt attenuation. Cold weather and conditions led to an ongoing loss of men, and individual battles caused sharp declines. Place names are associated with some locations and a binary attribute for advance and retreat is part of the use case. To address these situations, two feature classes and five relationship classes were developed for the model. The geometric components of the use case consisted of `nodes` and `polylines`. The `polylines` represent the army's marching path; the `nodes` provide locations for cities, detachment/reattachment points, and locations where abrupt changes in magnitude occur. In the model, `polylines` possess zero, one, or two associated `nodes`, and the `nodes` only occur at the end points of the `polyline`. The `incidence` relationship class associates the node locations with their appropriate polylines. The `troops` class handles the magnitude as a *method* defined as a function of length along the polyline—denoted as m.

Figure 3.4 Napoleon's march data model
From Glennon (2010)



Held in the `march` class, the *m* value refers to a measured value of the length of the polyline (Equation 3.1). Using these *m* values, the flow magnitude method can be defined as changing along the length of the flow. For instance, for linear interpolation, magnitude at a location, *x*, within the polyline representing flow can be calculated by:

Equation 3.1 Measuring flow volume

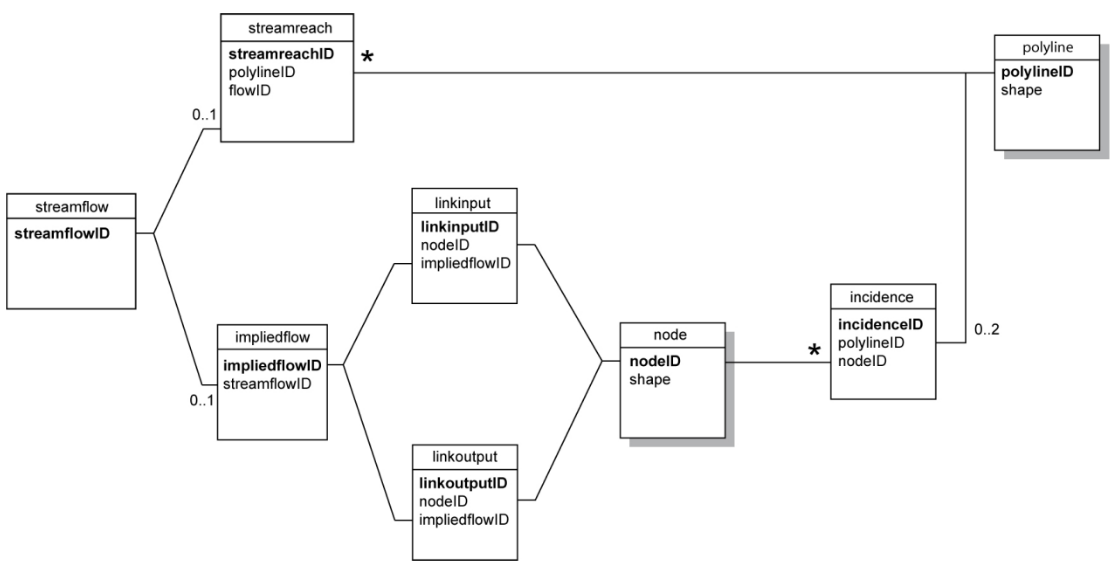
$$magnitude_x = \left[magnitude_{start} - \left[\left(magnitude_{start} - magnitude_{stop} \right) \times \left(\frac{m_x}{m_{stop}} \right) \right] \right]$$

where:

$magnitude_x$ is the measure of flow volume at location x;
$magnitude_{start}$ is the measure of flow volume at the start of the associated polyline;
$magnitude_{stop}$ is the measure of flow volume at the end of the associated polyline;
m_x is a measure of distance along the associated polyline at location x;
m_{stop} is the measure of distance at the end of the associated polyline.

The karst use case possessed two primary flow types: mapped and uncertain network paths (Figure 3.5). The network geometry used nodes and polylines: the `polyline` class represents reaches of the mapped stream segments, and the `node` class represents *from* and *to* locations for the uncertain flow paths, as well as junctions for converging routes. A `polyline` may relate with zero, one, or two `nodes`; the `nodes` can only be located on the endpoints of the polyline. The `streamreach` class relates the `polyline` to `streamflow` along mapped routes. However, magnitudes are not given, and thus are undefined in the use case. As a result, the `streamflow` class

Figure 3.5 Karst flow network data model
From Glennon (2010)

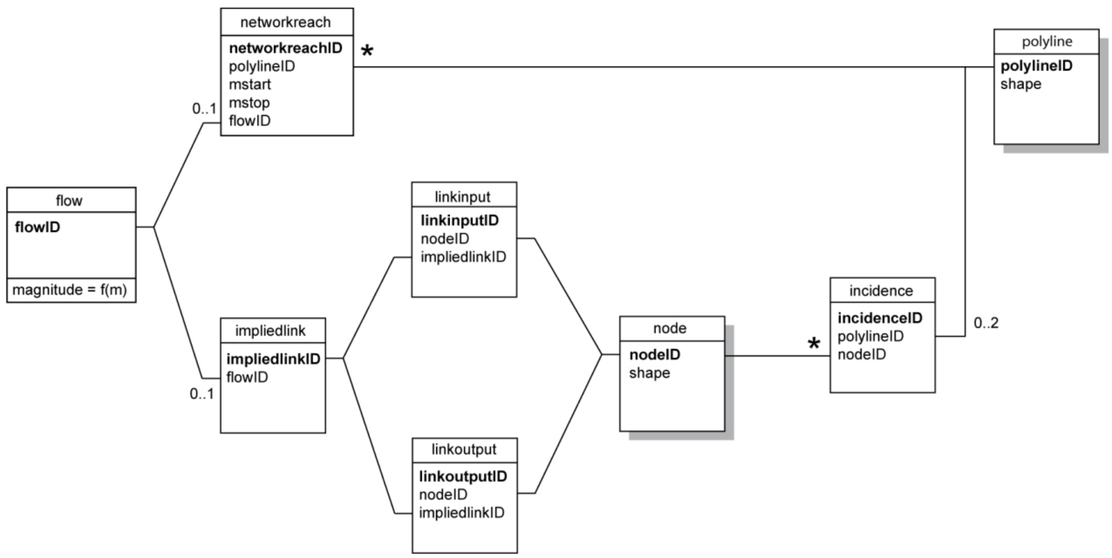


exists only as an identifier of flow existing along a given reach. Likewise, in the use case, magnitude does not change as a function of position, so *m*-values are not used. The relationship classes `impliedflow`, `linkinput`, and `linkoutput` associate input and output locations for flows along unmapped paths. Concerning the uncertain subsurface routings, the dye traces are treated as point-to-point links.

3.7 Distilling the use cases into a general data model

With each of the use cases described as UML, the next step in developing the data model was to assess the use cases' UML for commonalities and critical components related to GENet flow. The draft of the generic data model was developed by distilling and combining the commonalities into a single UML. The draft model development required identifying all the parts from the use case UMLs that were relevant and critical to the cases' generic concept of GENet flow. Components not critical to the generic flow concept, such as specific place names, were stripped from the use cases' UML diagrams. Use case specific attribute and class names were replaced with generic counterparts; for instance, the class names "migration", "troops", and "streamflow" were relabeled "flow".

Figure 3.6 GENet flow data model
From Glennon (2010)



Simplified to core components, the three UML diagrams reflected a more-generalized concept of flow. From examining the generically modified UML, two different types of flow were apparent within the GENet: flow along a mapped route, and flow where the input and output locations are known but the connecting path is not. Also, while the magnitude of flow was static over its time step in two of the cases, in one diagram, the flow magnitude attenuated and accumulated. To create the draft model, portions of the three cases' UML diagrams that were assessed as general to the concept of GENet flow were distilled and combined into a single diagram. The resulting draft data model is shown in Figure 3.6. An explanation of its classes and attributes is provided in Table 3.2.

Table 3.2 Components of the GENet flow data model
From Glennon (2010)

Flow Data Model Classes		Attributes	
	Feature		
polyline	a feature representing a line containing one or more line segments	polylineID	a unique numeric identifier for an individual polyline
		shape	geometric information for the polyline
node	a single coordinate point (typically a x,y pair)	nodeID	a unique numeric identifier for an individual node
		shape	geometric information for the node
	Relationship		
flow	holds the quantity of the flow. This class's association with a networkreach or impliedlink provide a flow's length and direction	flowID	a unique numeric identifier for an individual flow
		magnitude = f(m) (*a method*)	*magnitude* is a method that calculates a quantity based on a function of the polyline's *m*. The magnitude is the amount of aggregate movement; allowing magnitude to change along the length of the flow allows accumulation and attenuation. Conversely, the magnitude can instead be assigned a constant.
networkreach	refers to the geometry information for flows along a network. A networkreach is able to follow the entire length or a portion of a polyline's length	networkreachID	a unique numeric identifier for an individual network reach
		polylineID	a single polyline along which the flow travels
		mstart	the location along the polyline where the flow begins
		mstop	location along the polyline where the flow ends
		flowID	refers to a single flow associated with the above attributes
impliedlink	contains information for flows that travel uncertain routes between an input and an output node. An impliedlink stores flows that are not along networks	impliedlinkID	a unique numeric identifier for an individual impliedlink
		flowID	refers to a single flow associated with the impliedlink
linkinput	associates a single node with the input of an impliedlink	linkinputID	a unique numeric identifier for an individual linkinput
		nodeID	refers to a single node
		impliedlinkID	refers to a single impliedlink object
linkoutput	associates a single node with the output of an impliedlink	linkoutputID	unique numeric identifier for an individual linkoutput
		nodeID	refers to a single node that acts an input to the non-network flow
		impliedlinkID	refers to a single impliedlink object
incidence	associates polylines and nodes. An incidence refers to only one polyline and one node. When a node is linked to more than one polyline via incidences, the node becomes an intersection of polylines. Incidence provides a means of creating a variable number of intersections without changing the number of polyline or node attributes	incidenceID	a unique numeric identifier for an individual incidence
		polylineID	refers to a single polyline to be associated with a single node
		nodeID	refers to a single node to be associated with a single polyline

3.8 Testing and validation of the geographic data model

To be considered complete, the model must fulfill the following several requirements, 1) have structural locations to hold all data required to recreate each use case; 2) be able to logically describe the linkages and components of the case; 3) provide adequate logical infrastructure for associated queries to be successfully performed; and 4) minimize redundancies and dependencies.

Testing the first two requirements can be accomplished by instantiating the use cases with the new generic model. The third requirement requires practical queries against the model, and their successful calculation further reinforces the integrity of the first and second requirements. For the fourth requirement, a multi-step normalization process is described by Miller and Shaw (2001). The validation process refines a geographic data model to minimize associated storage space, reconcile database inconsistencies, and maximize structure stability. The normalization process becomes increasingly important with large, complex databases and those subject to ongoing revision. For rapid prototyping and simple models, unnecessary redundancies and dependencies will be significantly reduced if the first three requirements are satisfied.

The generic GENet flow data model is assessed by instantiating the use cases and determining whether the database possessed locations for all core flow data. An unsuccessful implementation of the cases with the generic model would necessitate revisiting the use case model creation and distillation process, and making necessary modifications to accommodate for unaccounted data. The process of revising the original use case models and distillation would need to be iterated until adequate data locations exist to implement the model in a GIS. The data model designer should carefully consider the entity being modeled to avoid creating an overabundance of specialty classes that go beyond the model's original intent.

Conversely, each use case need not use all classes in the domain's geographic data model. With the GENet flow data model, it follows that when implemented within a GIS, some classes and attributes may be used while others are dormant. For instance, the human migration case does not use the `networkreach` or `polyline` components of the model (Figures 3.7 and 3.8). The Minard map does not use the relationship classes associated with unmapped routes (Figures 3.9 and 3.10). The karst case uses all of the generic model's relationship classes, but attributes associated with *m* values remain inactive (Figures 3.11 and 3.12).

Figure 3.7 Active model components in the human migration use case
From Glennon (2010)

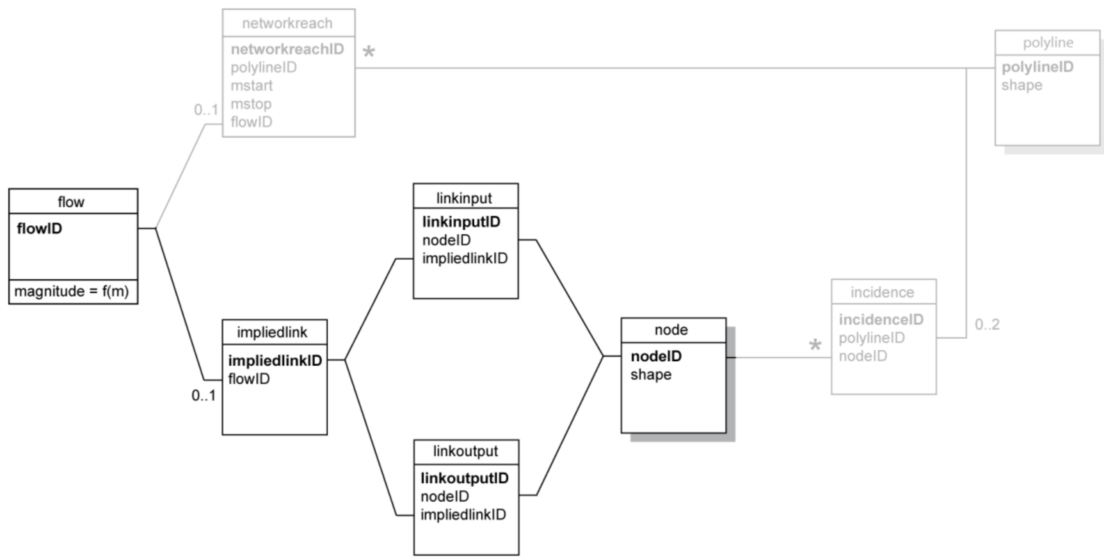


Figure 3.8 ArcGIS implementation of the human migration use case
From Glennon (2010)

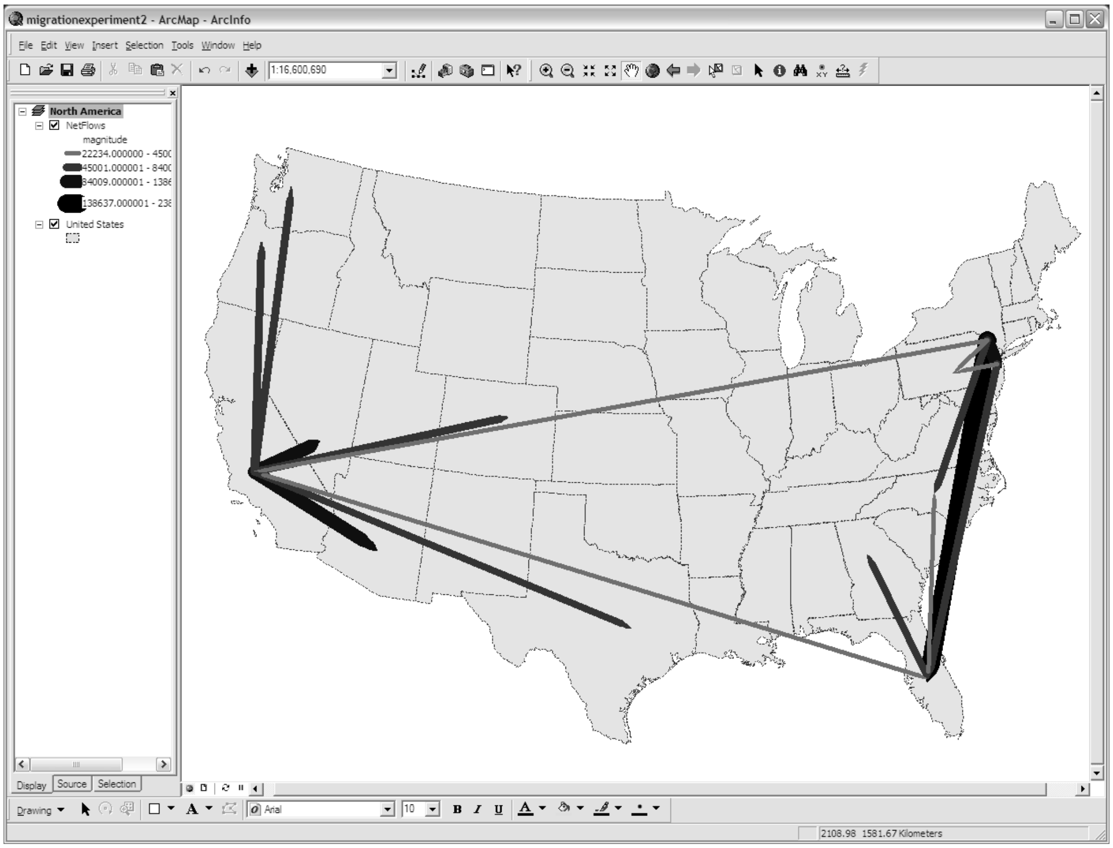


Figure 3.9 Active model components of the Minard map use case
From Glennon (2010)

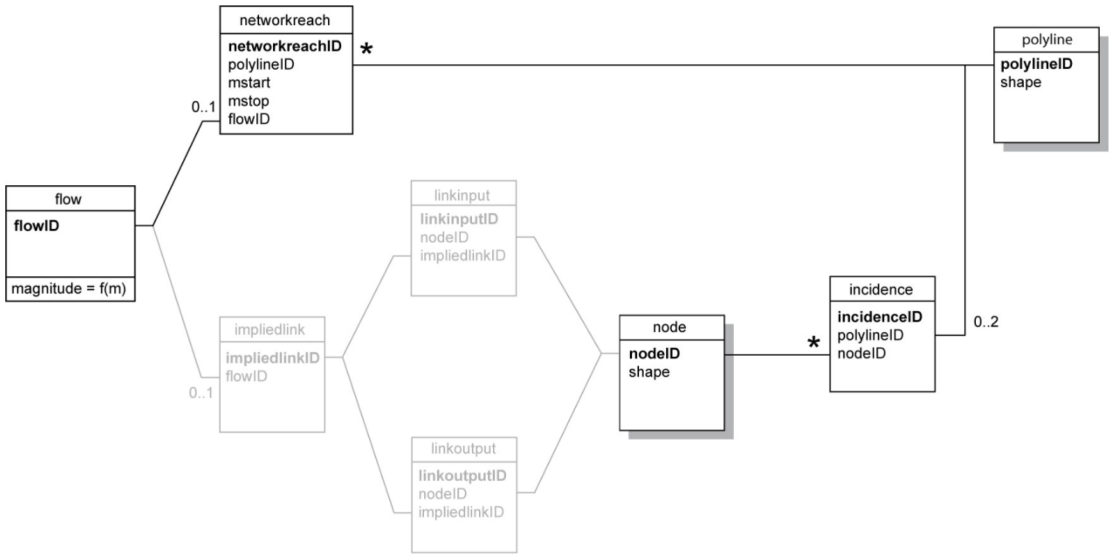


Figure 3.10 ArcGIS implementation of Napoleon's march on Moscow
From Glennon (2010)

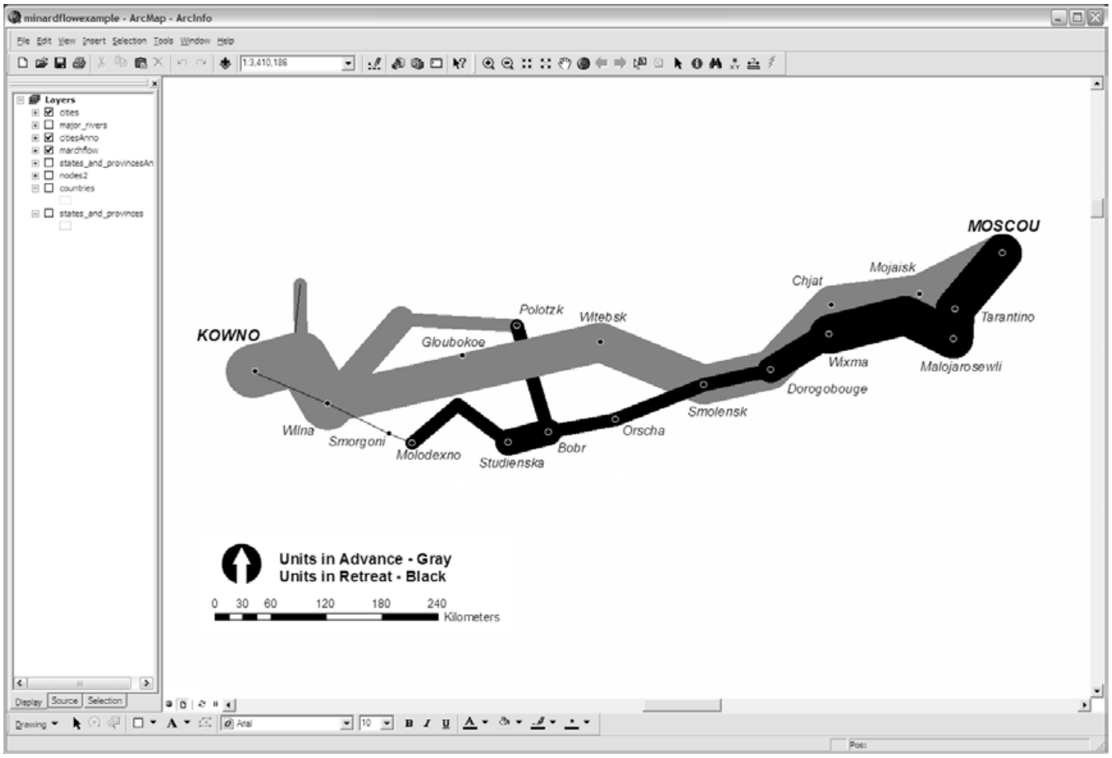


Figure 3.11 Active model components in the karst flow route use case
From Glennon (2010)

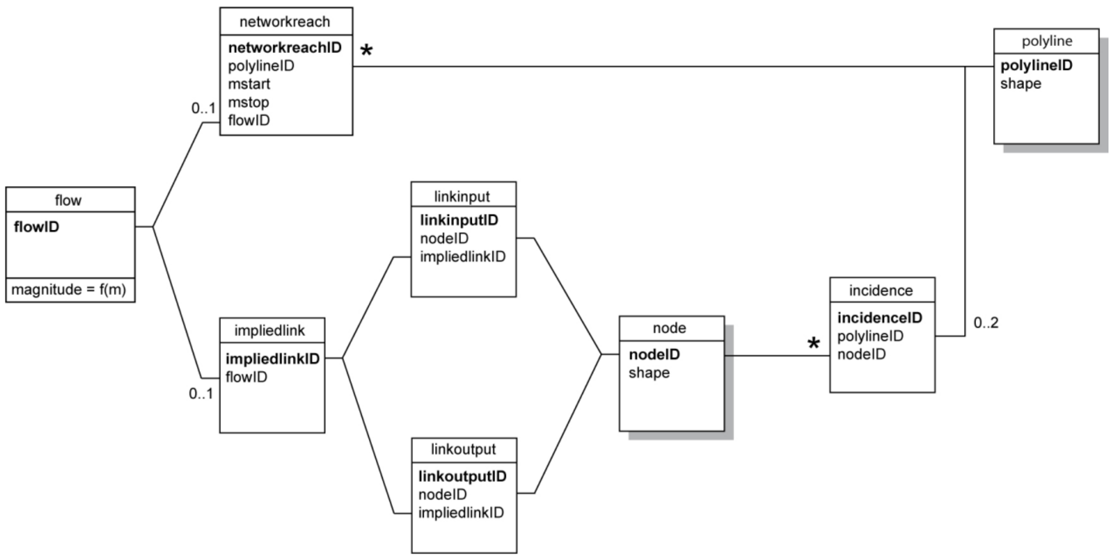
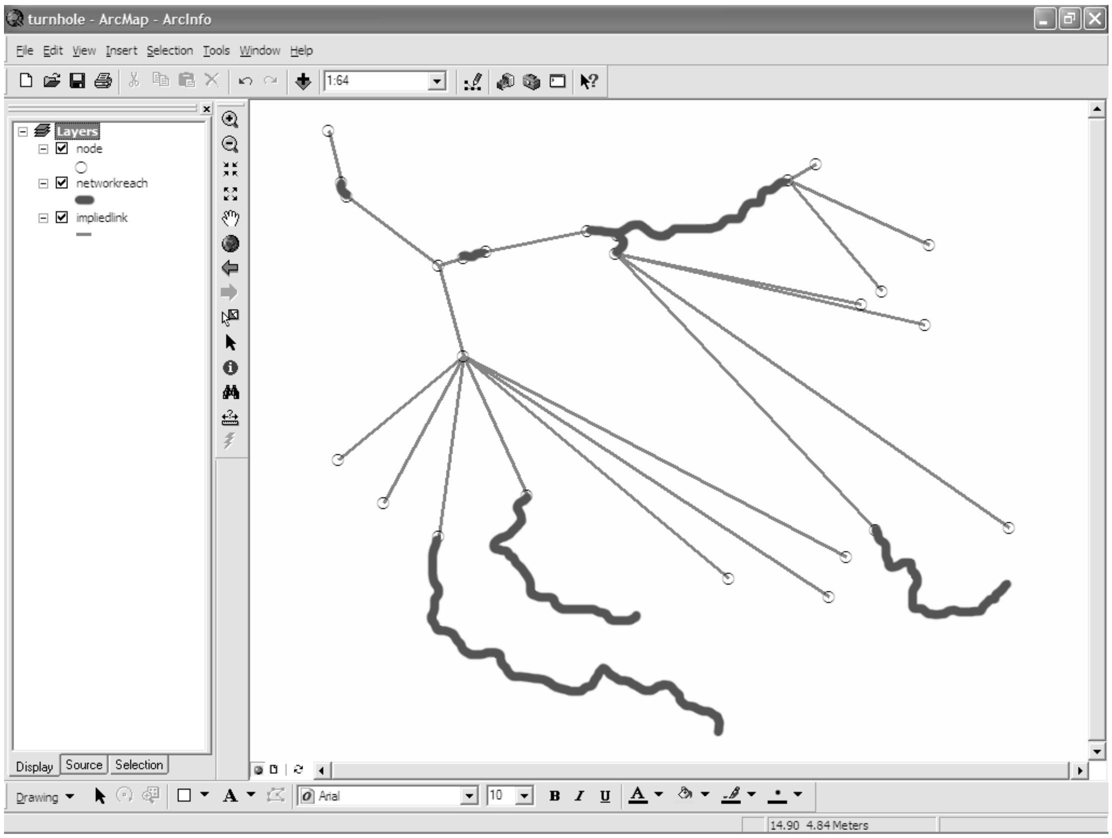


Figure 3.12 ArcGIS implementation of the karst flow use case
From Glennon (2010)



Besides building the use cases within a GIS using the generic model, performing domain queries against instantiated databases affords testing model completion. Considering such queries during the design process is likely to enhance the model's utility. To maximize the flexibility of a data model, formulating queries on the widest variety of relational algebraic operators is a suggested approach. Date (1995) enumerates the relational operators as: Select, Project, Join, Union, Intersection, Difference, and Product (Table 3.3). Egenhofer (1992) and Miller and Shaw (2001) discuss the opportunities and limitations of relational database queries with respect to geographic data models. As an example of a test query against the flow data model, Figure 3.13 shows the procedure to calculate a flow magnitude interpolated along a network path on Minard's map.

Table 3.3 Relational algebraic operations
From Date (1995) and Miller and Shaw (2001)

Select: Select certain tuples within a relation based on a state criterion

Project: Select certain fields within a relation

Join: Logical linking of tuples across different relations based on the matching of specific fields

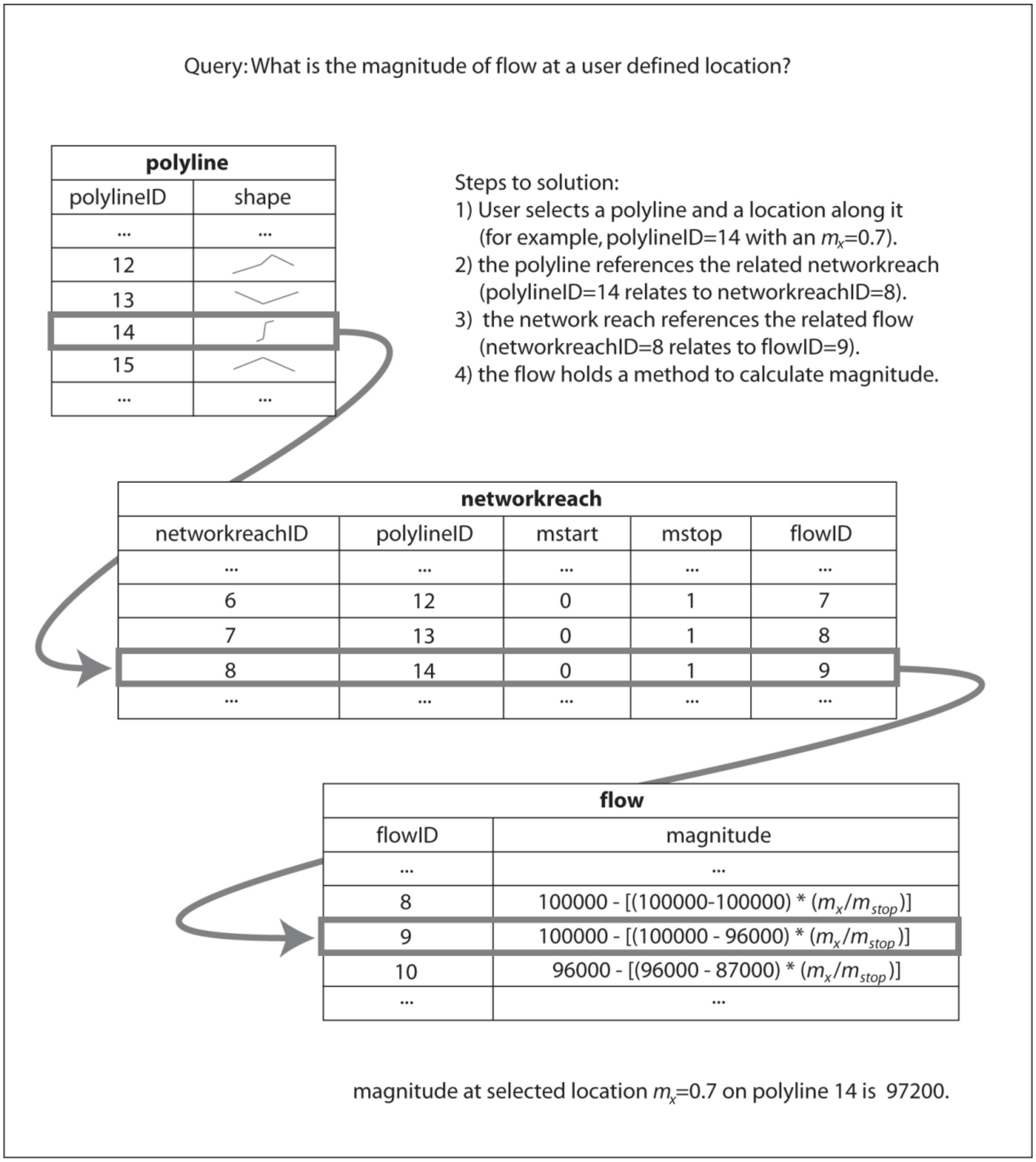
Union: Combination of all tuples of two relations into a single relation

Intersection: Find the tuples that are identical across two relations

Difference: Generate all tuples that are not also in another relation

Product: Generate all possible combinations of tuples from two relations (more precisely, generate a set of tuples that correspond to the ordered pairs that result from taking each tuple in the first relation and combining it with every tuple from the second relation)

Figure 3.13 Example linear interpolation query against the GENet flow data model
From Glennon (2010)



3.9 Discussion

Model development may provide insights into the internal organization and external relationships of geographic phenomena. In this regard, the purpose of this chapter is to study the characteristics of GENets by examining their data structures in a GIS setting.

3.9.1 GIS models and software

A geographic data model provides only data structure and relationships. For models instantiated with actual specific data, software is required for their storage, analysis, and rendering. For example, data within the GENet flow model can be queried to calculate net and gross flows, but a raw dump of any stored data would not explicitly provide that information. An analytical operation would be necessary. Similarly, beyond basic point, polyline, and polygon objects, a model does not contain accompanying cartography, so visualization relies on user and software interpretation of model attributes and associations. Fortunately, within geographic domains where symbology is more standardized, proprietary and open source developers have begun to leverage visualization templates to deliver expected cartographic outputs from data models (Davis 2007; Glennon and Glennon 2007). Since instantiated data models depend on compatible software tools for interrogation and visualization, a useful geographic data model is likely to require the modeler to create both a model and software tools to work with it. As such, an associated example set of open source software tools were developed by the author for the GENet flow data model. The tools include Flow Data Model Tools, a Visual Basic for Applications (VBA) script for Esri ArcGIS 9.x. This module performs basic import, export, query, and visualization operations on the flow data model. As VBA is being phased out in ArcGIS, an abbreviated Python implementation called Flowpy has also been developed. Flowpy requires the widely available, open source GDAL library to operate, but otherwise Flowpy is GIS platform agnostic. Flow Data Model Tools and Flowpy can be downloaded at http://alanglennon.com/genets.

3.9.2 Model purpose

Due to the complexity of the geographic world, there are no preset limits or guidelines for the number of use cases necessary to model a particular subject. When new data cannot be adequately addressed, a geographic model can be modified or augmented by developing a use case model for the new data, distilling it, and restructuring the generic model. However, modelers should be cautious about arbitrarily adding new cases. A model can be expanded until its original meaning, and possibly utility, is lost. The dangers of poorly specifying a model and selecting inappropriate use cases can be reduced by clearly defining the model's purpose. This purpose should be closely coupled with and rely on the queries and interrogations that the modeler plans to perform. In this chapter, the model was continuously held to questions such as: where are the flow routes; what is the direction of flow at a location on the network; how large is the flow at a given location; and, what is the net, gross, or two-way flow magnitude at a given location on the network? Because a model is created for specific goals and objectives, a model of a similar or even identical entity may be markedly different. Thus, modelers and users should carefully consider, and, as feasible, test assertions about a modeled entity's behavior within other contexts.

3.9.3 Limitations of geographic data modeling

For a geographic use case to be modeled with a computer, the case's fundamental aspects must be unambiguously encoded. Unfortunately, ambiguity is pervasive in human interpretation of geographic systems (Chrisman 1987; Taylor 1990; Pickles 1995). For instance, the attributes of a GENet arc may take many meanings. Consider the descriptions of a river that might be given by residents of a city along its banks. The residents might characterize the river as beneficial, harmful, relaxing, terrifying, or any number of dissonant descriptions. Further, each individual may view the river as having several of these attributes simultaneously and at different gradations. Encoding these attributes and relationships into a database may be possible, but the model is likely to become so cumbersome that it loses utility. The modeler, thus, must balance the violence that simplification imposes on the geographic case against the intent of the model. Further, the modeler should enumerate aspects of the case that are not included and evaluate their implications. Guidance for such assessment may be found in the literature associated with critical GIS (e.g., Harvey, Kwan et al. 2005; Sheppard 2005; Elwood 2008), introductory GIScience (e.g., Duckham, Goodchild et al. 2003) and uncertainty in geographic information (e.g., Zhang and Goodchild 2002).

3.9.4 Characteristics of GENets

The completed model provides both a data structure and a formalization of the phenomenon's ontology. The data structure facilitates organization and association of meaning to basic GIS geometries, creating opportunities for constructing queries, streamlining organizational tasks, assessing data completeness, and sharing data. For instance, the GENet flow data model suggests that in order to store, represent, share, and analyze common flow data, a user needs at least: 1) knowledge on whether the flow follows a mapped or uncertain path, 2) the start and end location of a flow link, and 3) the magnitude as a function of location within the flow network. Though encumbered by unique identifiers and rigid associations necessary in the computing environment, the general organization of a geographic data model serves as an ontological description.

The general data model highlights two conditions relevant to the characterization of GENets. First, the model describes flows along a known route and flow where the intermediate geometry is less certain. That is, GENets are characterized by arcs that are physical, abstract, or a hybrid of both. Further, the abstraction of physical pathways is common. From this study's use cases, the unmapped cave routes possess unknown geometry, so their pathways are simplified as a from-to description. The state-to-state migration case however is different. It abstracts numerous routes into a single from-to description to make the data more accessible. Abstract GENet arcs need not be reflections of underlying physical pathways. The flow of money or social linkages can be entirely conceptual.

A second GENet characteristic uncovered in the data modeling process is that GENet linkages may possess heterogeneous properties along their course. The accumulation and attenuation of soldiers along the path of Napoleon's march is an example. This GENet characteristic has been described by numerous GIS and transportation scientists with respect to road and intersection characteristics (Miller and Shaw 2001). Roads have varying widths, speed limits, and built characteristics. Advanced transportation modeling systems routinely incorporate these characteristics. In GIS, such characteristics along a network reach may be subdivided using dynamic segmentation functionality. Such functionality is currently constrained to simple operations within mainstream GIS, and no standard method exists to handle dynamic attributes such as Napoleon's losses and gains.

3.9.5 Generalized networks

In operations research (OR), generalized flow network models incorporate magnitude multipliers for each edge (Ahuja, Magnanti et al. 1993, page 566). These can be implemented in GIS using object-oriented methods and scripting to handle dynamic attribute situations. Through such means, capacity analyses involving accumulating and attenuating flow can be modeled. For instance, network capacity operations, like the Ford-Fulkerson (1956) algorithm, may require single sources, sinks, and circuits for their calculation. If a network originally had multiple sources and sinks, a typical OR approach to solve the problem would be to create an abstract single source (a supersource) and single sink (a supersink), overlay the virtual supersource and supersink into the problem, and attach them to the original network via generalized network edges. That is, each edge would hold an attribute multiplier that assigns the appropriate proportions from the supersource to the original sources.

In addition capacity problems, a generalized flow network model also can be used for transformational purposes (Ahuja, Magnanti et al. 1993, page 568). For example, a multiplier can convert one type of entity into another. In a trade network, an edge may represent both the geographic flow and the exchange of currency between types, such as the movement of United States dollars to Japanese yen.

Generalized flow network models and their analyses are not well addressed in contemporary GIS. The development of shared tools and best practices can foster such approaches.

3.10 Future work

The creation of a geographic data model for GENet flow illuminates two areas that are not well addressed by contemporary GIS: 1) functionality to allow spatially enabled software objects that possess path heterogeneity and dynamic attributes and; 2) methods to create, analyze, and visualize GENet implicit data and abstracted links. Addressing dynamics, particularly spatiotemporal databases, is a significant focus in GIScience (Raubal, Miller et al. 2004; Goodchild and Glennon 2008; Yuan 2009). These ongoing efforts are likely to possess useful techniques to address heterogeneous path characteristics and dynamic attributes. Extending standard GIS functionality through scripting and external software libraries also can allow classes within the data model to possess behavior. For instance, the GENet flow model stores attenuating and accumulating flow along certain and uncertain paths using an Object Oriented Programming (OOP) *method*. Within a class, a method can define an operational function, such as a relationship based on time or the object's distance away from another object in space. Methods offer a technique to integrate situations like changing geometry, location, or internal structure into object models. Unfortunately, despite their promise, geographic data model classes that explicitly include methods are neither widely used nor supported. The use of methods extends the standard concept of GIS data from static data to a more advanced feature that changes based on context. Given the increased ability to use programming languages to customize GIS implementations, related functionality and generic best practices should be explored.

Tools for manipulating GENet flows are largely unavailable in mainstream GIS. Gross and net flow calculations are simple, but not included in the core of Esri ArcGIS, GRASS, QGIS, Google Earth, or any other mainstream GIS package. Such functionality must be added ad hoc via scripting by the user. Also, abstract GENet linkages often are computationally stored as from–to data, implying a straight line link, yet mapmakers often represent these pathways as curves. Esri ArcGIS possesses some relevant cartographic workflows for grouping linear features and simplifying paths with respect to scale. A useful function would be to allow a network link to curve in reaction to a continuous cost surface. The function in particular would be useful for abstract routes that are affected by landscape, ocean currents, or population.

The use cases and final model possess conditions of GENet flow along mapped routes and location-to-location flow. Some flows, however, such as wind, ocean currents, and porous media groundwater flow, are continuous, unchannelized, not easily discretized, and thus, largely incompatible within the object-oriented representational framework (Couclelis 1992; Peuquet, Smith et al. 1998). As a proposed, preliminary solution to this object-oriented modeling issue, Goodchild and Glennon (2008) describe a UML equivalent for continuous fields. Nevertheless, overall, efforts to seamlessly integrate continuous field data with object oriented design remain at an early stage.

This chapter creates a GIS data model for GENet flow and offers a use case-based method for developing specialized geographic data models. By creating individual domain models based on a common theme and distilling them into a final product, a procedure is outlined for grounding geographic phenomena to the bounds of a class and relationship organizational scheme. This design process also facilitates the ontological description of the study domain, and provides insight into the character of GENets.

CHAPTER 4:
ANALYTICAL OPERATIONS ON GENETS

4.1 Introduction

A geographic feature's data structure offers little insight into how the feature interacts with the environment. In GIS, analytical operations can evoke such interactions, and with respect to this dissertation, offer a mechanism to uncover more on the nature of GENets. Formalizing analytical workflows using UML and programming code, operations can be interrogated and compared against their query types, database inputs, algorithms, and outputs. The purpose of this chapter is to uncover and describe the properties of GENets as they are affected by geographic analysis.

For this study of GENets, a set of representative queries is taken from hydrology, transportation science, GIS, and simulation. Since at least the 1970s, quantitative analysis of GENets has coevolved within these geographic subdisciplines. While the domains share many analytical techniques, their interest in different facets of GENets yields diverse approaches: hydrology implements process modeling to study inputs and reactions; transportation science uses optimizations to determine best pathways; GIS leverages spatial measurement and comparison; and, geographic simulation conducts rules-based iteration with agents or cells.

An example use case is chosen from each of these subdisciplines and their analytical workflow is formalized via UML or pseudocode. To allow comparisons of the workflows, each is subjected to a set of questions modified from Mitchell (2000) (Table 4.1). The use cases are: from hydrology, deriving a stream network from a Digital Elevation Model (DEM); from transportation science, calculating the shortest path between two points in a network; from GIS, identifying streets bounding a city block; and from simulation, modeling urban growth near roads. The four cases do not represent all possible operations performed on networks in geography. The cases do, however, reflect a wide range of geographic analyses balanced against the diversity of their specific algorithmic approaches. Analysis often is a transformational process, and in this regard, the cases also are selected to offer diverse outcomes. The transportation and GIS cases yield solutions that are subsets of an existing network. The hydrology case use a region's attributes to derive a new network, and the simulation case uses a network to seed a non-network solution.

Table 4.1 Parameters for characterizing GENet operations
Modified from Mitchell (2000)

Objective

What is the problem objective?
What are the problem constraints?

Data

How is the network represented?
Does the data support uncertain, fuzzy, or missing data?

Algorithm

Are operational rules deterministic or stochastic?
Does the algorithm require iteration?
    If so, what is its nature (finite, continuous, dynamic feedback)?
What dimensions does the algorithm consider?
Is the environment static or dynamic?
Does the algorithm use exact or approximate methods?
How does the algorithm handle uncertain or missing data?

Output

What is the nature of the solution? For instance, is it vector, raster, numeric, descriptive?
Is the solution a subset of existing data or something newly derived?

Process

Is scale a consideration? For instance, does the answer change with varying scale?
Is the algorithm designed for single or repeated use?
Are the results repeatable?
Is the process reversible without data loss?
How are distance units or length addressed?

The chapter begins with a description of each use case and its associated workflow. Then, similar to the geographic model creation process described in the previous chapter and by Glennon (2010), an attempt is made to reconcile the workflows into a general GENet analysis model. Further, each workflow has been chosen for its geographic nature. The use cases thus are evaluated against a set of four tests by Goodchild (2012) to identify their specific geographic components (Table 4.2). This study of workflows and their geographic nature intends to elicit a set of properties associated with their common element, the GENet. The chapter concludes with a summary and description of GENet properties.

Table 4.2 Tests of a spatially explicit model
From Goodchild (2012)

Invariance test
spatially explicit models are variant under relocation of the objects of study

Representation test
spatially explicit models include representation of location in their implementations

Formulation test
concepts such as location or distance (if we include scale too) appear directly in the model

Outcome test
spatial structures of inputs and outputs are different
it modifies the landscape on which it operates

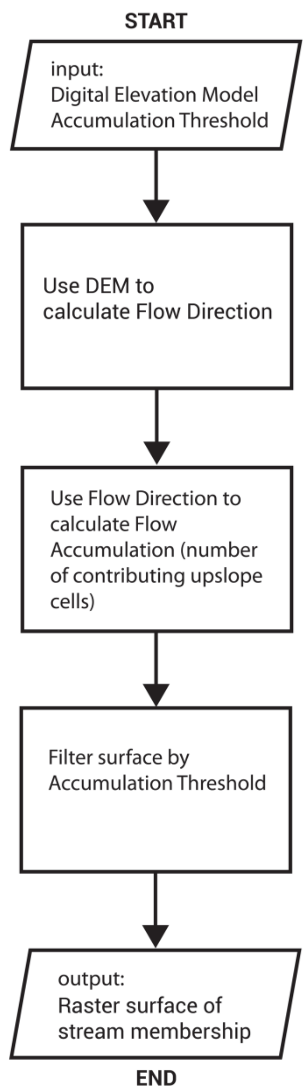
4.2 Workflow descriptions

4.2.1 Hydrology: construction of a stream network from a digital elevation model (DEM)

The standard workflow for stream network construction from a DEM follows a multistep sequence (de Smith, Goodchild et al. 2007). The operation requires two primary data inputs: a DEM and a flow accumulation filter value. A DEM is a surface of height values; for this case, the source data is a raster grid of cells with each element possessing an elevation value. In the stream network construction, a cell is considered a member of the network based on a predefined threshold of cells draining into it. This predefined threshold, set by the user, is the flow accumulation filter value.

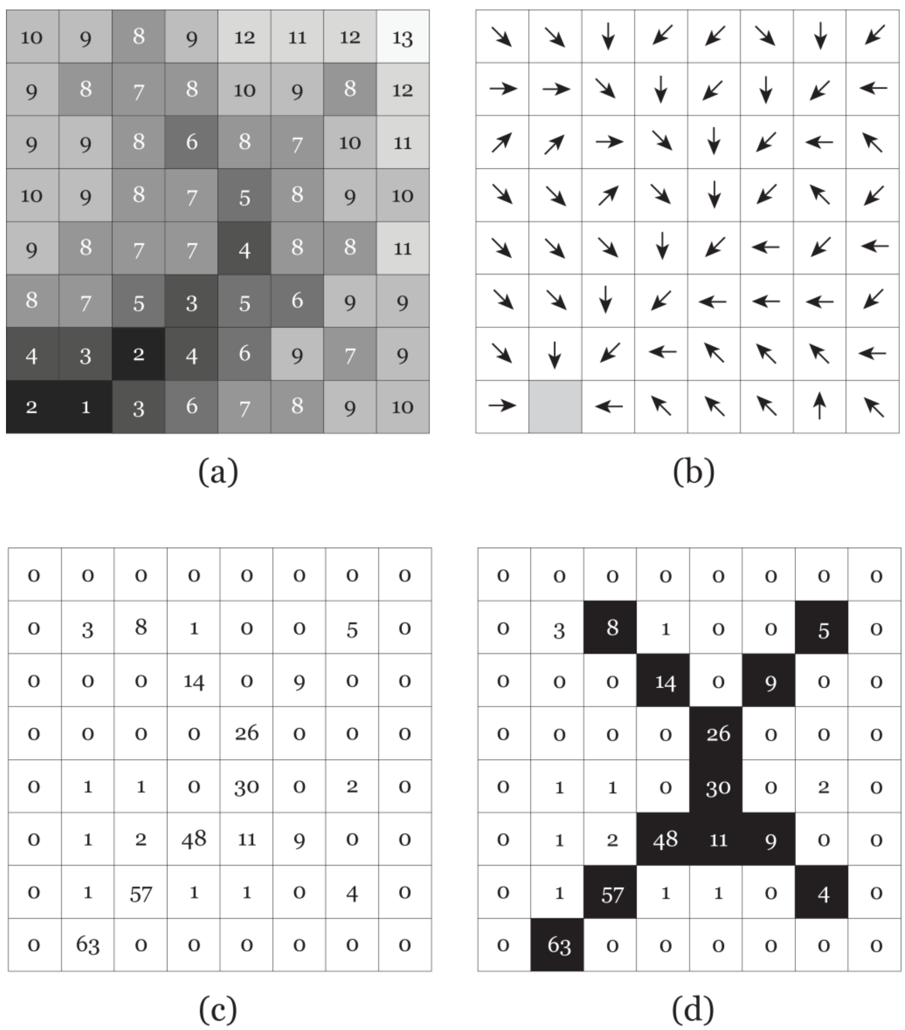
To construct the stream network, first, the aspect of each DEM cell is calculated (Figure 4.1). The process is iterative for each cell using data from its neighbors, usually from the eight surrounding cells or the four cells at cardinal directions. For the second step, using aspect, the number of cells upstream from a cell can be determined. The resulting collection of cell values is the flow accumulation surface. Similar to the calculation of aspect, assigning flow accumulation values iterates across each cell of the surface. Also, though, computing the number of flow contributing cells will vary with each cell depending on the number and geometry of its uphill cells. The flow accumulation surface is then queried against the flow accumulation filter value. Cells equal or higher than the value are defined as members of the stream network. The output is a raster of cells, usually composed of zeroes and ones to denote stream network membership. Areas with internal sinks or flat regions may cause unexpected results or processing ambiguity and thus require additional consideration to correct, like filling or proximity assignment rules.

Figure 4.1 Workflow for constructing a stream network from a DEM



The workflow for stream construction is a deterministic sequence. Output from prior steps feeds as input to the next, but iteration to evaluate cell values occurs entirely within a step (Figure 4.2).

Figure 4.2 Example of constructing a stream network from a DEM



4.2.2 Transportation: calculation of the shortest path

Numerous methods exist to calculate the shortest path between two points on a network (Ahuja, Magnanti et al. 1993; de Smith, Goodchild et al. 2007). Dijkstra's Algorithm (1959) is perhaps the most common and is the method used in this paper (Figure 4.3). To calculate the shortest path, the Algorithm tracks the length of possible paths for each node in the network, compares them, and retains the minimum. For data input requirements, Dijkstra's Algorithm requires a defined network topology, distance attributes, and from-to vertices. The method supports network topology defined through standard representations, such as Node-Node, Node-Arc, Forward Star, and Adjacency Lists (Ahuja, Magnanti et al. 1993).

The operation initializes by setting the path distance to the start node to zero and other nodes to infinity. The nodes are all labeled *unvisited*, and the start node is labeled *current* node; all nodes except the start are included in this *unvisited* set. For the *current* node, calculate the *temporary* path distance to all *unvisited* adjacent nodes. If a *temporary* path exists to a node, retain the lowest path distance value and note the node from which it came. When all of the *current* node's neighbors have been evaluated, label the current node as *visited* and remove it from the *unvisited* set. The distance to this node is now minimal and the node does not need to be reevaluated. This condition is labeled *permanent*. When the destination node is labeled *visited*, the algorithm is finished. Otherwise, select the *unvisited* node with the shortest *temporary* distance as the new *current* node. The procedure now loops back to the step with the calculation of *temporary* path distances to neighboring nodes.

For this study's implementation, the shortest path's vertices are retained as an array. The output is the shortest path length and an associated array of vertices (Figure 4.4).

Figure 4.3 Workflow for Dijkstra shortest path calculation

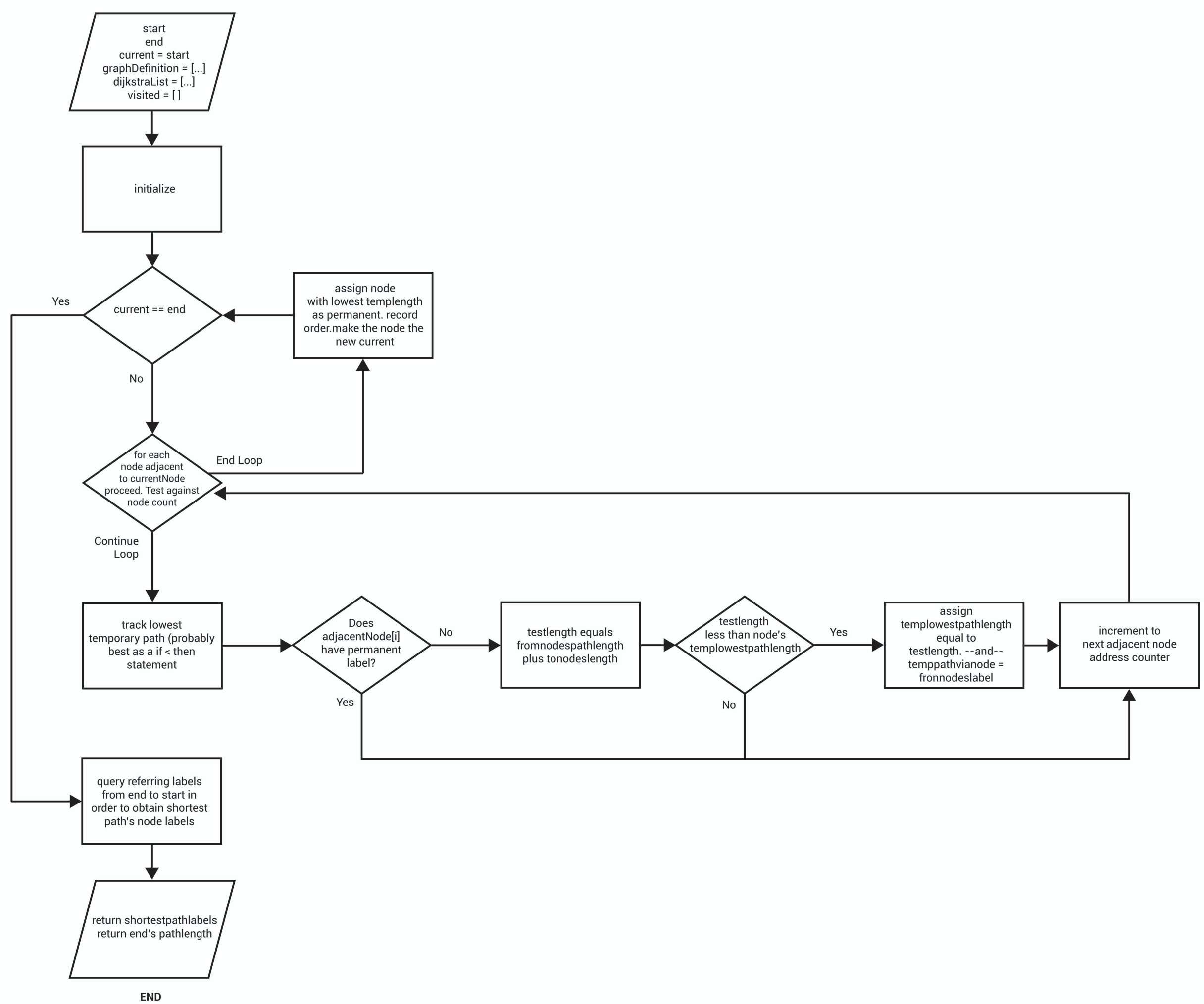
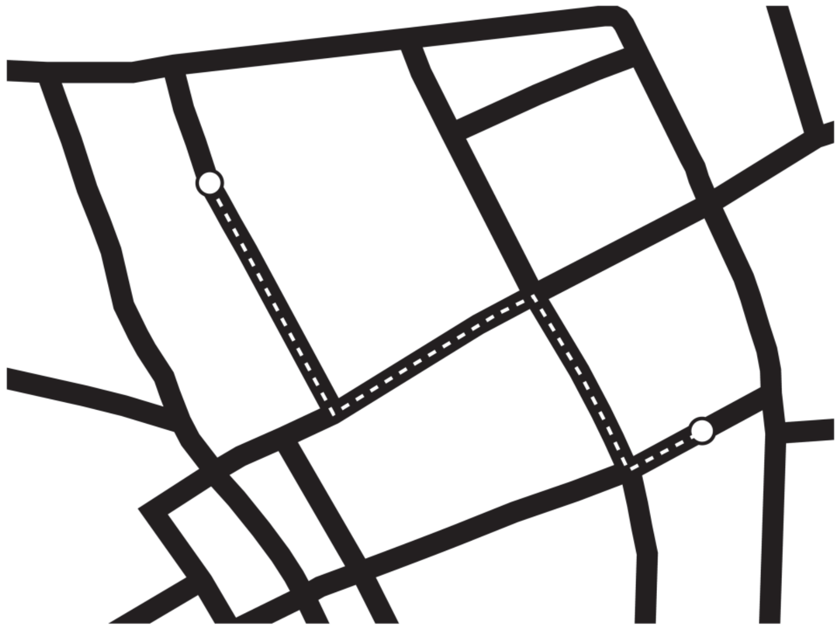


Figure 4.4 Example of calculating a shortest path



4.2.3 GIS: identification of streets bounding a city block

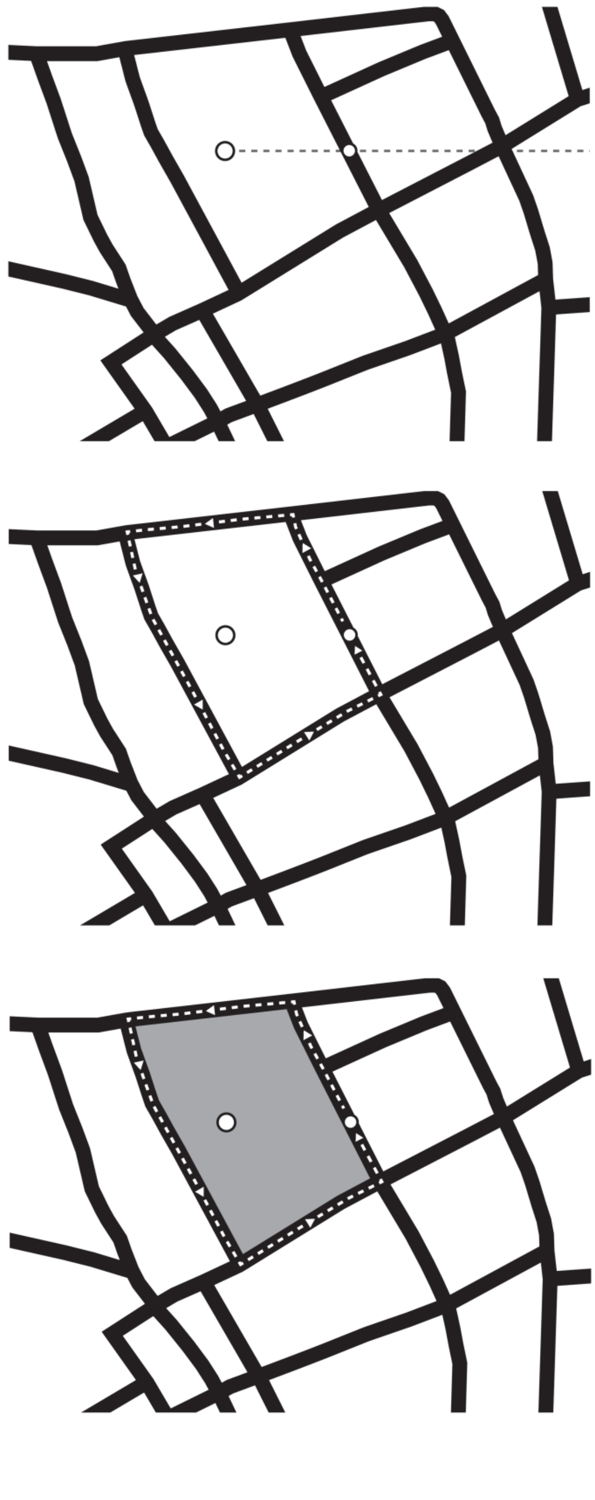
Spatial selection and comparison are standard operations in GIS. In this respect, the identification of streets bounding an area is a representative GENet analytical case. Real world street networks can be complex with cases like overpasses, directional rules, and noncontiguous reaches and circuits. To emphasize the core GENet selection operation, this use case holds to a simple two-dimensional network where all street networks possess clear intersections, two-way travel is allowed, and no dead end streets exist. Implemented with GIS, polylines represent the street network and the city blocks are the interstitial space. The operation is initiated by the selection of a point within one of the open regions. From that point, the spatial selection identifies the subset of the polyline network bounding the point using a procedure described by Worboys and Duckham (2004) (Figure 4.5,).

Figure 4.5 Computing the counterclockwise sequence of arcs surrounding a selected location
From Worboys and Duckham (2004)

```
Input: Node n
find some arc x which is incident with n
arc <-- x
repeat
     store arc in sequence s
     if begin_node(arc) = n then
          arc <-- previous_arc(arc)
     else
          arc <-- next_arc(arc)
until arc = x
Output: Counterclockwise sequence of arcs s
```

With a point selected within the city block area, an arc is extended any direction until it intersects the surrounding network. The intersection point is noted and becomes the starting point of the selection operation. Moving from this intersection in either direction, the operation tracks along the polyline until it reaches the next intersection. The choice of direction, clockwise or counterclockwise, is recorded, and all tracked arcs’ identifier information is stored for later retrieval. For this use case, the arbitrary direction chosen is counterclockwise. Upon arriving at an intersection, the operation continues tracking on the left-most path until arriving at the starting point. If the clockwise direction had been chosen, every right-most path would be selected. Upon arriving at the original intersection, the bounding polylines have been identified, and the output is the retrieved list of arcs (Figure4.6).

Figure 4.6 Example of workflow for identifying a city block

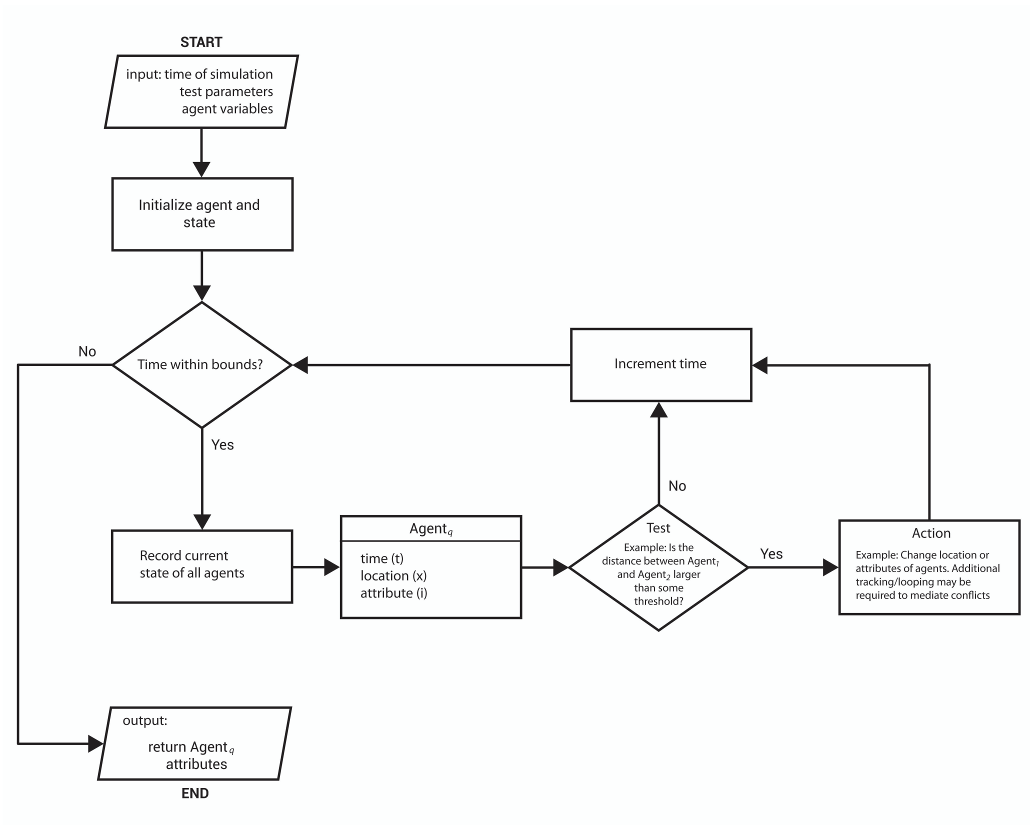


4.2.4 Simulation: modeling urban growth around a road network

Cellular automata and agent-based modeling offer powerful tools for modeling geographic interactions (Figure 4.7). These approaches entail the notion that simple individual activity, in aggregate, may yield unforeseen, complex systems. Similar to many geospatial operations, analogous techniques are available using different representations, like raster or vector. Generally, agent-based simulation is associated with vector and object-based representations, while cellular automata tend to be associated with raster space. In the previous use cases, two have used object-based representations and the hydrology case worked in a raster. As such, this use case uses a raster simulation approach to maintain the balance. This use case simulates urban growth near a road network using cellular automata.

Transportation corridors promote development, and the associated dynamics have been widely studied in geography (Christaller 1933; Hägerstrand 1967; Herold 2003). Urban growth tends to occur adjacent to other urbanized areas. From those premises, a highly simplified process for urbanization is undertaken with respect to a GENet, in this case, a road network. Network data are presented as a rectilinear grid raster, with cells included in the network holding a value of one and cells outside the network zero. For this use case, the road is considered an urban area and it thus becomes the seed for adjacent development. The model iterates by time step

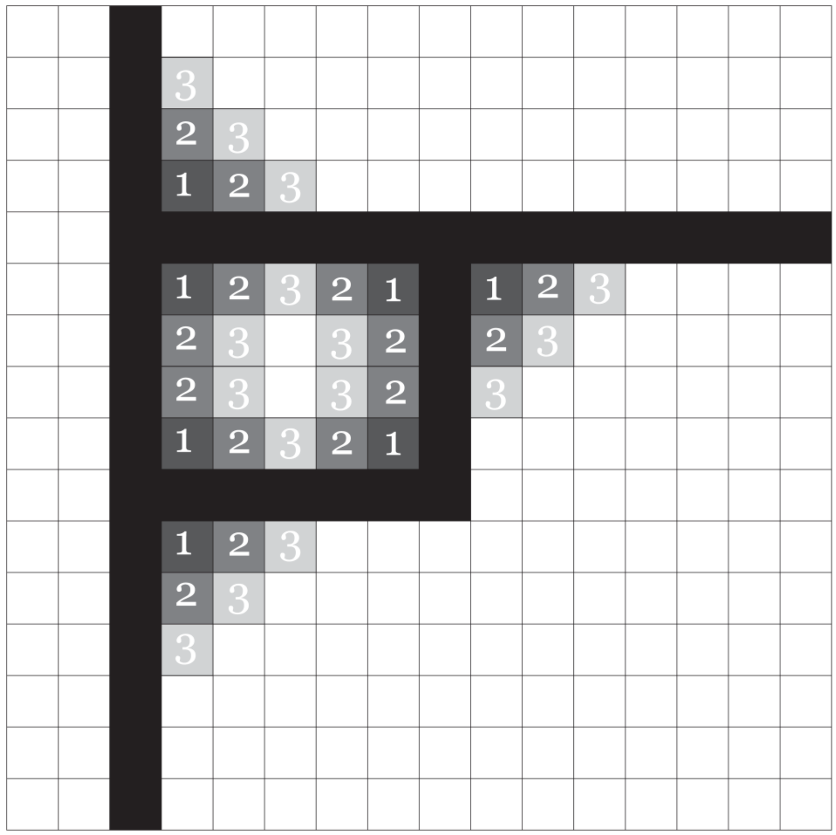
Figure 4.7 Generic geographic simulation workflow



and reclassifies cells as urban or not following the rule: if four or more of a cell's eight neighbors are urban, the cell becomes urban. Otherwise, the cell remains in its current state. Highly advanced geographic cellular automata simulations and associated theory have been developed by other geographers (White and Engelen 1993; Itami 1994; Couclelis 1997; Clarke and Gaydos 1998; Onsted and Clarke 2011). This use case simulation has been reduced to basic components to emphasize the cellular automata workflow and its interaction with a GENet.

The input for the model is a raster data set of a road network, the classification rules (e.g., reclassify a cell as urban if four or more of its eight neighbors are urban), and number of time steps to perform. In a full, realistic urban growth model, the time period represented by each iteration and the cellular reclassification rules would be informed by thorough study of a region's physiographic and socioeconomic setting (Herold 2003). This use case performs three iterations, and these successive generations are labeled as 1, 2, and 3 (Figure 4.8). Each generation creates a new raster that becomes the new state of comparison for the next iteration. All of these generations can be stored for later analysis or overwritten if not needed. The output of this case's model is a raster representing urbanized areas (Figure 4.8).

Figure 4.8 Example of agent-based modeling of urban development
Cells labeled 1 represent the first generation of model run, 2 are the second run, and 3 are the third generation.



4.3 Use case comparison

In a survey of network shortest path algorithms, Mitchell (2000) offers a set of parameters that characterize such problems (Table 4.3). To make those parameters applicable to a wider set of GENet operations, Mitchell's parameters have been reorganized into the categories: Objective, Data, Algorithm, Output, and Process (Table 4.1). Each use case is interrogated against these parameters (Appendices A1-4). The problem parameters and workflows are then evaluated for commonalities. To aid the comparison, problem parameters are translated to their general problem type. For instance, the transportation case's path minimization is described as "optimization". Similarly, specific constraints have been changed to descriptive characteristics like "proximity" or "topology".

Table 4.3 Parameters for geometric shortest path and network optimization problems
Modified from Mitchell (2000)

problem objective:
How do we measure the "length" of a path?

path constraints:
Are we simply to get from two points, or must we also visit other points or regions along a path or cycle?

input geometry:
What types of "obstacles" or other entities are specified in the input map?

dimension of the problem:
Are we in 2-space, 3-space, or higher dimensions?

type of moving object:
Are we moving a single point along the path or something else?

single shot vs. repetitive mode queries:
Do we want to build an effective data structure for efficient queries?

static vs. dynamic environments:
Do we allow obstacles to be inserted or deleted, or do we allow obstacles to be moving?

exact vs. approximate algorithms:
Are we content with an answer that is guaranteed to be within some small factor of optimal?

known vs. unknown map:
Is the complete geometry of the map known in advance, or is it discovered online, using some kind of sensor?

The four use cases possess widely varying network operations: a process model, optimization, comparison, and simulation. The hydrologic problem transforms an elevation surface to a stream network via a process model; the transportation case uses Dijkstra's Algorithm to calculate an optimum path; the GIS problem performs a series of spatial comparisons to identify the arcs bounding a point; and, the simulation problem uses cellular automata to iterate and build an urbanized surface.

The operations possess different strategies for considering constraints, including proximity, attributes, and topology. For instance, the hydrologic and simulation problems consider the attributes and adjacency of neighboring cells to calculate their solutions, the transportation problem relies on network topology and arc attributes, and the GIS problem leverages topology and the properties of the two-dimensional plane. No consistent or general pattern is recognizable across the problems in the relationship between objectives and constraints.

The use case representations were chosen to maximize diversity: two of the problems use a raster representation and two use vector. In each case, alternative representations are possible. None of the problems' data structures include missing or uncertain data. However, each of the problems could be restructured to manage and reconcile such data. The addition of missing or uncertain data likely would require different algorithms and yield different results, but would not change the operations' geographic character and associated GENet interactions.

All of the operations follow deterministic methods. The common deterministic property of each operation was the result of use case selection, and not the operation itself. Each of the use cases could be re-created to incorporate stochastic methods. Since randomness is the scientific benchmark against which order is measured, future work should consider incorporating analytical operations that use stochastic methods.

The use cases each exhibit different types of iteration. The hydrology problem follows through several compartmentalized steps: aspect calculation, flow accumulation calculation, and stream construction. Each step contains internal loops, and these are independent of loops in other steps. The internal loops mostly iterate through each cell of its raster surface and perform a calculation. Some of these calculations require additional iterations themselves, such as when elevation values are retrieved from neighboring cells. The output surface from each compartmentalized step becomes the input into the next step. The next process then has its own internal iterations and calculations. The transportation case performs an optimization requiring multiple interdependent iterations. These nested loops required an internal loop to process before continuing. Further, Dijkstra's Algorithm considers non-optimal solutions as it runs, and discards them in order to reduce computational burden. Thus, looping becomes more efficient as the algorithm moves closer to the solution. The GIS problem iterates as it traces the problem's bounding arcs. In this case, looping is used to facilitate the decision making process with respect to each intersection that is encountered (e.g. choosing the left-most path) until arriving at the starting point. The amount of looping therefore is related to the complexity of feature topology. For the simulation problem, iterations are used in a similar manner as the hydrology problem but with the addition of time. In both cases, looping is compartmentalized. That is, iterations occur internally with a generation, and an output state is created. The next iteration cannot begin until the previous has completed. The output of a process becomes the input of the next.

The urban growth simulation involves a dynamic environment, while the other cases exist in a static state. With increasing computing power, device mobility, and real time result expectations, GENet implementations increasingly are likely to include movement and other dynamic behavior. Though further study is necessary, no general strictly time-dependent GENet properties are obvious.

A property common to all the selected use cases, but not a property of all GENet analysis, is that outputs cannot be queried, or reverse engineered, to obtain the problem's original input data. The outputs from the use cases are lossy reflections of the source data. For GENet analysis, lossy output appears typical, but not universal.

The use cases chosen for this study represent a wide diversity of GENet analysis. The common characteristic among operations is the use of comparison and iteration, but a shared GENet approach to problem solving is not apparent. With respect to GENet analysis in GIS, this is not a particularly interesting result, because comparison and iteration is fundamental to all computing. The comparison of the use case parameters offers inconclusive results, especially when considering aspects distinct to GENets (Table 4.4). The workflows reflect different facets of GENet operations. Therefore, the creation of a single, meaningful, generic workflow model for GENet analysis is not possible.

Though a generic model is not possible, the use case workflows still offer insights on the GENet interactions not elicited by data structures alone. A systematic evaluation of these characteristics can be accomplished by identifying the spatial properties of the use cases with respect to networks.

4.4 Geographic characteristics of the use cases

The respective geographic characteristics of the use cases can be identified using a series of tests devised by Goodchild (2012). These tests define a geographic model as spatially explicit if it varies with relocation, includes spatial concepts—like distance, area, or location—within its representation or model, or creates a new spatial structure as output (Table 4.2). With focus on GENets, each use case is evaluated against the tests and tentative general properties identified. The properties are then collected and discussed in the chapter's final section.

Table 4.4 Comparison of generalized use case parameters

	Hydrology	Transportation	GIS	Simulation
Objective				
Objective	Transformation	Optimization	Comparison	Simulation
Constraints	Distance, proximity attributes	Topology, attributes	Topology	Distance, proximity, attributes
Data				
Representation	Raster	Vector	Vector	Raster
Handling Uncertainty	No	No	No	No
Algorithm				
Deterministic/ Stochastic	Deterministic	Deterministic	Deterministic	Deterministic
Iteration	Multiple, independent loops	Multiple, dependent loops	Looping for feature comparison	Independent loops for feature comparison and time
Dimensions	2D	Not spatially dependent	2D	2D
Environment (static/dynamic)	Static	Static	Static	Dynamic
Exact or approximate methods	Exact	Exact	Arbitrary from point to network, then exact	Exact
Address uncertain or missing data	No	No	No	No
Output				
Type	Raster	Vector, numeric	Vector	Raster
Subset of existing or new	New derived form	Subset of existing	Subset of existing	New derived form
Process				
Scale considered	No	No	No	No
Repeatable Results	Yes	Yes	Yes	Yes
Reversible	No	No	No	No
How are distance or length addressed?	Via cell units	Via attribute	Via 2D coordinate system	Via cell units

In the hydrologic use case, the input data represents the land surface, and a change to a different location would alter the landscape's elevation values—passing Goodchild's Invariance Test. The input data cells possess and rely on fixed area, a spatial concept—passing the Representation Test. The analytical model explicitly references the spatial notion of adjacency in its evaluation of neighboring cell values—passing the Formulation Test. Further, the output is a raster representing a stream network in geographic space. Overall these tests highlight conditions in which the character of a region influences the location of a network. In this case, the network is defined by elevation and the function of the network. The hydrologic network occurs where it does because of the physical properties of water—it flows downhill. An indicative GENet property appears to be close interaction, if not dependence, on its surroundings.

Extending Goodchild's Invariance Test to include scale, different scale input data would affect the use case's output. The constructed stream geometry would vary with different granularity of elevation data, either through a change in cell sizes or in elevation measurement resolution. Scale appears to affect all GENets representations. Related, the analysis does not address the likely attributes of the resulting stream network. The resulting stream would be larger, perhaps wider, based on the size of its watershed. GENets may possess heterogeneity—for instance, differing widths, directionality, and capacity—with respect to their constituent parts.

In the transportation use case, moving to another location (Goodchild's Invariance Test) would yield different results as it is assumed the GENet's location is fixed. Given a network of identical topology and distance attributes at the new location, though, the analysis would yield the same shortest path result. The input data for the use case includes network topology, passing the Representation Test. In the problem, distances are stored as attributes of the arcs. The model considers distance, but as a generic cost rather than a spatial concept. A similar problem could be constructed where the shortest path followed a non-spatial concept between nodes, like financial cost or time. GENets reaches can exhibit physical and abstract properties. Dijkstra's Algorithm yields a new spatial structure as output, a subset of the network with associated distance attributes.

The GIS problem would yield different results for different starting points. The GIS case is embedded in two-dimensional space, and thus the input data includes spatial aspects within its representation. The operation leverages the two-dimensional plane to constrain an arc to intersect the bounding streets. The case creates a GENet of streets surrounding the select point, and relies on the properties of geographic space interacting with the network. In defining its topology, the GENet relies both on the relationship of its parts and the surrounding space. Such context allows the user-selected point to represent the bounded region.

Relocating the study area of the simulation case would alter the output of the analysis. Results from agent-based and cellular automata are highly sensitive to the initial state (Wolfram 1984). Similar to the stream construction problem, data input for the use case includes the spatial concept of area—the raster cells are defined to have fixed area and dimension. The case and GENet are thus subject to effects of scale. Similar to the earlier hydrologic use case, the simulation uses spatial concepts in its analysis for evaluating adjacency. The simulation use case yields a new spatial structure, a raster of urbanized areas, as output. As represented in GIS, all the use case GENets are affected by spatial uncertainty, arising from causes such as measurement accuracy, positional errors, and data granularity.

4.5 Properties of GENets

Several GENet characteristics are noted in the dissection of the analytical workflow use cases. GENets may exhibit a mix of tangible and conceptual constituents, and these are influenced by their embedded setting in different ways. As expected, GENet arcs and nodes that tangibly exist in physical space impact and are impacted by their environment. Such GENets also routinely possess constituent heterogeneity in space and time. All GENets modeled within GIS are affected by representational issues of scale and uncertainty. Inventorying and organizing such characteristics can be used to create more useful and relevant models of networks in geography.

4.6 Physical and Abstract GENets

Based on its constituent arcs and nodes, a GENet can be subdivided into two types, Physical and Abstract. Physical GENets are composed of pathways and intersections that tangibly exist in the physical environment, and Abstract GENets possess a mix of geographic and conceptual elements. Examples of Physical GENets include rivers, roads, and electrical grids. Abstract GENet examples include economic trade maps, airline route maps, and georeferenced social networks. Abstract relationships may be entirely intangible, like the movement of ideas, or partially grounded in geographic space, like the aggregated representation of many airplanes' travels into a single arc. Thus, some Abstract GENets are more susceptible to the influences of the physical environment than others. For instance, the GPS track of an individual through the environment is conceptual—no tangible trail exists. The individual would, of course, choose her movements with respect to the barriers and opportunities of her geographic surroundings. However, if total time of travel is represented as a single time step, the trail might be considered tangible as the individual's presence becomes a smear along the distance of the entire track. Tracking individual movement is routine, but the logic necessary to define the path as tangible is peculiar, and accordingly, such GENets are placed in the Abstract category. Such a track could be considered a hybrid arc—neither tangible nor independent of its geographic embedding.

Categorizing GENets by the nature of their constituent arcs and nodes is discussed in Section 6.1.2; in the section, hybrid, physical, and abstract arcs are defined.

Abstract GENets may have less interaction with the environment than a Physical GENet. In these cases, tangible nodes are embedded in geographic space, but the arcs are conceptual and non-spatial. The nodes thus are subject to the First Law of Geography—near things are more related than distant things (Tobler 1970). The cognitive meaning and perceived strength of conceptual versus physical relationships is complex, and may not be intuitive. In a study of spatialized linkages, for instance, Fabrikant, Montello et al. (2004) found that people interpreted links as more influential than the space between nodes.

The categorization of GENets as Physical or Abstract serves as a starting point for delineating the utility, behavior, and meaning of network linkages. Geographic embedding may be reflected in arc attributes such as cost, utility, or effort along a link, as well as whether an arc can be meaningfully subdivided. Presumably, abstract arcs, like a friendship link between georeferenced friends, lose meaning if their links are broken into two or more segments. However, a road blocked by several intermediate barriers—a physical arc—is still traversable within the noncontiguous sections. Whether Physical or Abstract, all GENets modeled within GIS are affected by representational issues of scale and spatial uncertainty. Further, GENets with tangible arcs and nodes possess at least two additional properties, those of constituent heterogeneity and areal dependence.

4.7 Characteristics of GENets with respect to representation

4.7.1 Scale

GENet representations are affected by scale—the resolution of the units in which they are measured, modeled, and displayed. Scale issues influence both raster and vector representations of GENets. The cell values of a raster are either an aggregation of an area's values or the measure at one point. In either case, an area with many attribute values is distilled into one. The larger the cell size, the more heterogeneity within the cell is likely, and the less the single selected value appropriately represents the area. Vector GENet representations possess similar scale issues. Except for a perfectly straight geographic pathway, the finer the unit of measure, the longer the measure of the pathway becomes. The classic description of this problem is the example by Mandelbrot (1967) calculating the length of coastlines. Measuring the coastline of Britain using different measurement lengths, a measurement unit of 100km yields a coastline length of approximately 2,800km, while a 50km unit measure results in a coastline length of approximately 3,400km. This effect is exacerbated in modern online computer maps, as not only length increases but also the number of features. For instance, viewing the same highway intersection at multiple scales yields not only more length, but more roads. At coarse scales, ancillary details are bundled into a simplified representation, but as the scale becomes finer, the map unpacks the details and more roads are made visible.

4.7.2 Spatial uncertainty

Spatial uncertainty with respect to GENets arises throughout the GIS workflow, and can stem from myriad sources such as measurement accuracy, measurement error, incomplete data, and data aggregation. For example, in the data collection process, measures are constrained by the physical limits of the measuring instrument. As mentioned previously, unit granularity may profoundly affect the measurement of GENet distances. Modeling the real world as a digital system always requires simplification and compromises involving which details to include and exclude. Features may be aggregated for the purpose of a specific implementation, but at the cost of losing specificity with respect to individual items. Data entered and stored in GIS are subject to the limitations of the system and architecture. For instance, computerized input typically requires a single measurement value and possesses no ability to handle associated information about a measurement's confidence. During analysis and for other data manipulations, iterative calculations can cause minor measurement imprecisions to compound. Output, particularly visual representation, may vary in resolution, either as a hardware limitation or as a result of decisions made by the system designers.

Specific to GENet representations, incomplete data and knowledge are common. A network representation may be missing vertices or linkages because they are not known to exist or not yet measured, and uncovering such relationships is a fundamental objective of science. Incomplete data may be the result of being unable to measure or follow a linkage. Inaccessible GENet pathways may occur as an artifact of time. That is, a historical path may no longer exist or a planned pathway may not yet exist. Some arcs may be unmeasured due to physical obstructions or unsafe locations—like enemy trails, underground stream networks, or active lava tubes. These cases all disallow certain knowledge of at least a portion of the GENet.

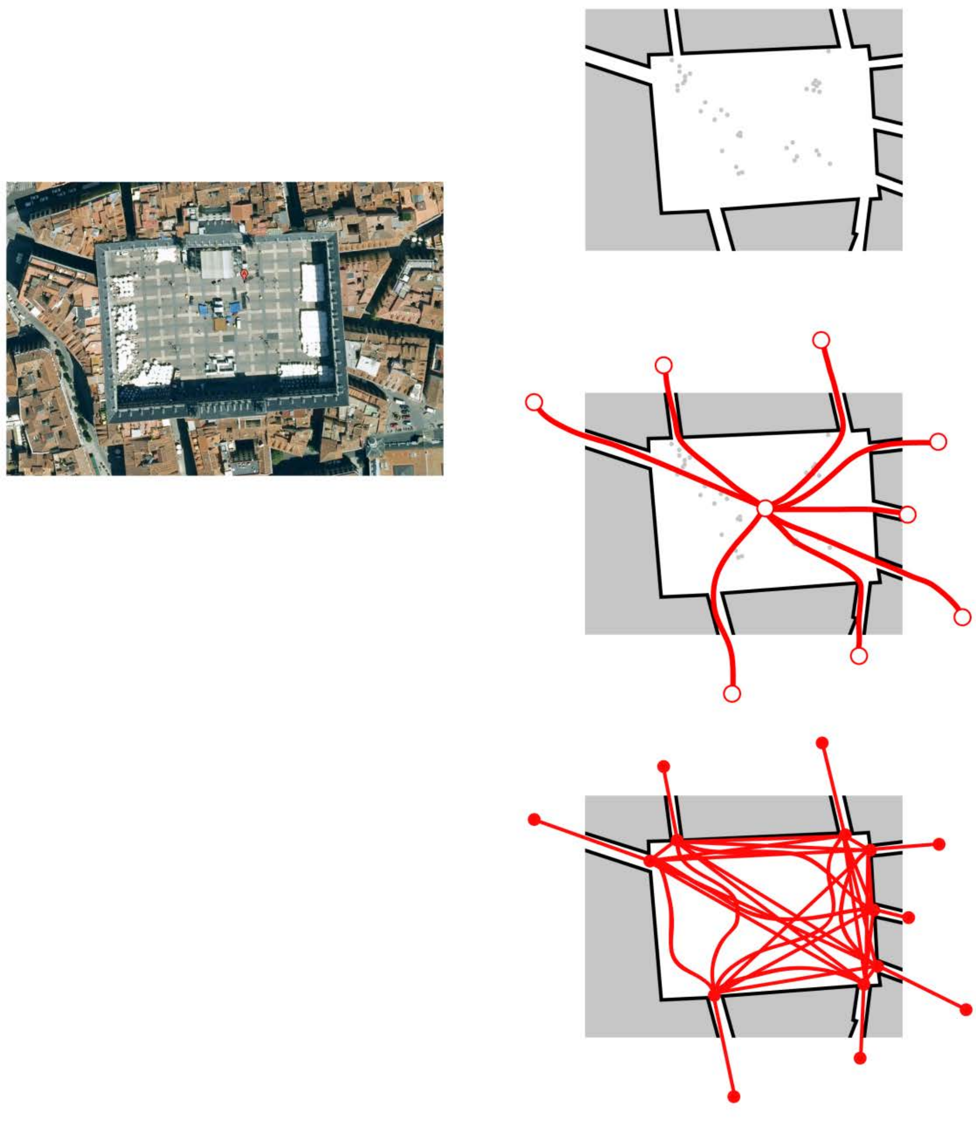
Data aggregation can cause positional uncertainty from scale issues or through the creation of pathways that cannot be directly measured. Multiple physical pathways can be aggregated to be represented as a single conceptual arc. Considering the example of human migration, people traveling from one place to another often take different routes. The aggregated path is the transformation of these numerous pathways into a single abstract route.

4.8 Characteristics of physical GENets

4.8.1 Constituent heterogeneity

GENets often possess arcs and nodes with heterogeneous properties (Miller and Shaw 2001; Maidment 2002; Longley, Goodchild et al. 2005). Changes in GENet character occur across space and time. A GENet existing at a single time or unchanging in time may be geometrically heterogeneous. A river, for example, possesses differing channel widths and depths along its course. The widening river may blend into a lake or reservoir. In this regard, GENet heterogeneity may reflect change in function of a network or reach. A river acts to transport water and the lake to store it, but a distinct boundary or intersection between the two does not always exist. Related, GENet arcs and nodes may possess multiple functions, like a road that also serves as a walking or cycling path. Geometric heterogeneity perhaps implies that it only considers the shape of the network's boundary with neighboring space. However, physical arcs and nodes may also possess internal geometric and attribute heterogeneity. A few examples include variations in flow capacity along a reach, roads with variable number of lanes and variable attributes—like fast and slow lanes, changing speed limits along a reach, turning rules, and internally complex intersections. For instance, Madrid's Plaza Mayor (Figure 4.9), a pedestrian open space, is a complex intersection that serves multiple feeder pathways while also offering multiple internal destinations that currently include restaurants, street performances, and market shops.

Figure 4.9 Plaza Mayor, Madrid, Spain
Basemap image from Google



Geometric, functional, and internal heterogeneity also may change with respect to time. Rivers oscillate between low and high flows, and the change affects the location and geometry of the banks. A road widening project offers a narrow road before and a wider one afterwards. Physical GENets may vary in function through time. For instance, an abandoned train track may become remade as a bike or hiking trail. Similarly physical or social factors may change the functioning of a network. When a river is not flowing, the dry riverbed may double as an animal migration corridor. Social factors can facilitate change of GENet function, such as rules to modify a road's speed limit, capacity, or direction. Temporal internal heterogeneity of network constituents can be seen, for example, in the use of automatic signal lights at an intersection to govern traffic.

A Physical GENet may change its location or move in and out of existence. Because Physical GENets exist as tangible features, the implications of locational changes are more pronounced than for Abstract GENets. Energy is required to move a road or a river, and unlike a change to an abstract arc, the move directly impacts its environment.

4.8.2 Areal dependence

GENets, both natural and socially constructed, interact with their surroundings. In the process model, GIS, and simulation use cases, interaction between the network and adjoining space facilitate the analytical result. Only the transportation case's shortest path calculation did not leverage the properties of the surrounding landscape. In the general case, such a shortest path calculation over a street network would rely on the landscape, as each arc inherits an associated distance from the geographic setting. The close communication of a network with its setting perhaps is illustrated best in the relationship between a river and its watershed. The river shapes the surface of the watershed and the watershed supplies the river with water. Neither would exist without the other. For networks created by people, access and utility drive the interaction. Networks like roads and electrical grids are designed to serve regions. The regions in turn make the roads and electrical grids useful.

4.9 Discussion

Beyond the purpose of uncovering novel properties of GENets and their analysis, translating this study's cases into GIS facilitates the identification of areas for future software tool development and study. GIS modelers should devise and deliberate upon best practices for modeling and representing the function of GENets. Domains like hydrology and operations research, for instance, use superimposed secondary networks to assist analytical operations and the modeling of network function. In hydrology, river networks are sometimes represented with two data sets—a set of connected polylines to represent the flow network and then another set of polygons to represent channel width and lakes. The Maidment (2002) hydrology model delineates the function and interaction among multiple GENet data sets. Ideas from such domains should be evaluated with respect to representing GENet functionality.

Given the commonalities in GENets visual representations, a coherent set of network display tools are needed. Further, to promote sharing of useful techniques and the reuse of GENet data, a set of consistent practices for handling dynamic properties and visualization with respect to GENets should be developed. Related, network representation communicates meaning and affects interpretation. GIScientists should investigate the principles of spatial cognition with respect to GENets and inventory associated naïve perspectives.

For this study, the comparison of GENet analytical operations is undertaken to facilitate a better understanding of linked geographic relationships. Using computational formalizations such as UML and pseudocode, the use cases are constrained and reduced to comparable elements. The exercise affords evaluation of the operations and serves as a method to delineate the properties of GENets. Representational issues of scale and uncertainty impact all GENets. Constituent heterogeneity and areal interaction affect network features residing in geographic space.

CHAPTER 5:
A GENET ANALYTICAL APPLICATION: THE GEYSER TRAVEL PROBLEM

5.1 Introduction

This dissertation emphasizes the use of computing and computational models to better understand networks in geography. GENet problem solving approaches and the associated use cases are investigated not only as an end to the individual queries they involve, but as a means to uncover the underlying interactions of networks and geographic phenomena. Further, ubiquitous computing, including smartphones and laptop computers, has increased the ability to perform locally-aware, real-time analysis. Associated geographic analytical approaches are accelerating in relevance. This chapter introduces a practical application, the Geyser Travel Problem (GTP), which combines several disparate GENet operations with location-based, real-time analysis. The purpose of the chapter is to implement the GTP in GIS and evaluate the impact of GENet properties on the analyses and solutions.

The GTP describes the issue of people trying to make best use of their day while visiting Yellowstone National Park. Their travel is limited to a walking path and the visitors try to plan and follow an itinerary that maximizes the likelihood of seeing the park's various geysers. Unfortunately, geyser eruption times are not always available, nor are the forecasts precise. When available, the predictions are offered as time windows. For instance, Old Faithful Geyser may be predicted to erupt at 09:00 plus or minus 10 minutes, or Grand Geyser may be forecast to erupt between 08:00 and 12:00. Visitors try to arrange their day to maximize what they see, minimize wait times, uncertainty, and the amount of energy spent walking.

Three GIS implementations of the GTP are evaluated: a random walk, an itinerary created prior to travel, and an itinerary created and continuously updated with respect to incoming data. The body of the chapter outlines the methods, implementation, and results for each case. The chapter ends with a discussion of the GIS implementations with respect to the properties of GENets.

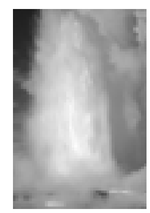
5.2 The setting of the GTP

The Upper Geyser Basin (UGB) in Yellowstone National Park, Wyoming, USA, holds approximately 200 geysers—20 percent of the world's total and more than any other single location (Bryan 2008). The basin is home to Old Faithful Geyser and several other large, predictable geysers.

Park visitors travel through the UGB by traversing foot paths of an established trail network. Cycling is allowed on a two-kilometer paved portion of the trail, but that mode of travel is relatively uncommon and not considered in the present study. For safety and protection of the hydrothermal features, visitors must remain on the designated trails. At least 19 scalding deaths have occurred from Yellowstone hot springs as a result of careless travel (Whittlesey 1995). The portion of the UGB trail network in the study is 6.87 kilometers long and is based upon National Park Service (NPS) data collected by global positioning system (Karplus 2004).

The NPS calculates general predictions for five major UGB geysers by considering recent observations and behavior, including Castle, Daisy, Grand, Old Faithful, and Riverside geysers (Table 5.1). The time from an eruption to the next is called the *period*. Eruption *duration* is the eruption's length in time, and the geyser's *interval* is the length of quiescence until the start of the next eruption. Thus, a geyser's period is equal to its duration and interval. For geysers with consistent behavior, a geyser's distribution of periods with respect to previous eruptions can be transposed into the future using a recent eruption (Rinehart 1980); the result is a probability distribution—a prediction model—of the next, future eruption. The span of the distribution is presented to visitors as a homogeneous prediction time window. The NPS-predicted times are posted for tourists at the visitor center and are updated as new data are available. In addition, in the basin, each of the predicted geysers has a

Table 5.1 NPS-predicted geysers in the Upper Geyser Basin

GEYSER	BEHAVIOR	PHOTO
Castle	Erupts from a tall, castle-shaped cone. It has both major and minor eruptions; unpredictable after a minor. Prediction window: +/- 60 minutes Height: 10-35 meters Interval: 13 hours Duration: 60 minutes	
Daisy	Erupts as an angular stream of water from a low cone. Splashes and overflows prior to an eruption. Prediction window: +/- 30 minutes Height: 25 meters Interval: 2.5 hours Duration: 4 minutes	
Grand	Erupts as a bursting fountain from a large pool, often in concert with several neighboring geysers. World's tallest predicted geyser. Prediction window: +/- 120 minutes Height: 45-60 meters Interval: 7.5 hours Duration: 14 minutes	
Old Faithful	Erupts from a mound and builds to a graceful column of water. Prediction window: +/- 10 minutes Height: 30-55 meters Interval: 90 minutes Duration: 5 minutes	
Riverside	Erupts from a large cone on the opposite side of the river from the viewing area. Overflows and splashes for an hour or more prior to an eruption. Prediction window: +/- 30 minutes Height: 20 meters Interval: 6 hours Duration: 20 minutes	

changeable sign; a grease marker is used by rangers to post the predicted eruption window. These signs, however, are inconsistently updated.

Solving the GTP considers visitor itinerary planning and travel as they integrate their knowledge of the environment and balance it against their preferences and abilities. Strategies for itinerary creation and travel behavior vary by visitor, and associated motivations may be complex. Visitor experience is influenced by satisfaction, benefits, experience, and meaning (Borrie and Birzell 2001). Thus, visitor behavior can range from deliberate, goal-oriented activities to passively allowing outcomes to emerge (Patterson, Watson et al. 1998; Rademaker 2008). In the UGB, some visitors walk through the area without any preparation, while others utilize as much information as possible, including NPS predictions and additional sources of decision support. For example, supplementary information includes knowledge individuals bring to the site (such as printed and online guides), evidence collected via personal experience, and information shared by other guests. The reasons for visiting the park without an itinerary may be due to lack of preparation or value on less goal oriented experiences. Some park visitors value free time and exploring with the family over goal-oriented activities. Social interactions at Yellowstone have been found to deepen the understanding of the geysers (Brody, Tomkiewicz et al. 2002). Despite differing motivations and goals at the basin, a common objective is to experience the most or the highest quality geyser eruptions possible.

5.2.1 Visitor types

This study describes three classes of visitors to the UGB, each with higher levels of engagement with respect to itinerary planning; these are categorized as Naive, Informed, and Expert.

The Naive visitor enters the UGB with little care or knowledge of the spatial arrangement of visitor paths or locations of geysers (objectives). The visitor may truly be unprepared or may find satisfaction in less goal-oriented endeavors. This visitor explores the network path ad hoc and sees geyser craters, hot springs, and scenery without deliberate planning. As the visitor travels along the path, he or she may fortuitously witness eruptions nearby and in the distance; all without prior knowledge of the event.

Informed visitors have knowledge of UGB trail geometry, geyser locations, and basic geyser behavior, including the likely intervals between eruptions. This visitor starts the day at the visitor center and obtains prediction windows for the Castle, Daisy, Grand, Old Faithful, and Riverside geysers. The visitor makes no return trips to the visitor center for updated information and uses only basic knowledge to estimate eruption past the given predictions. Using his or her knowledge of the trails, viewing locations, and prediction windows, the visitor creates a single itinerary before departing. The itinerary offers a route that endeavors to view the most geysers possible during the visit. If windows conflict, the user must make decisions of which prediction takes precedence.

The Expert visitor enters the UGB with knowledge of the trail, NPS predictions, and also the ability to receive and evaluate real time observations in the field. The Expert starts with a tentative route and schedule and revises them as incoming data predict new eruption windows and afford new opportunities. While the Informed visitor creates a single itinerary once before departure, the Expert visitor itinerary is reevaluated with each new observation.

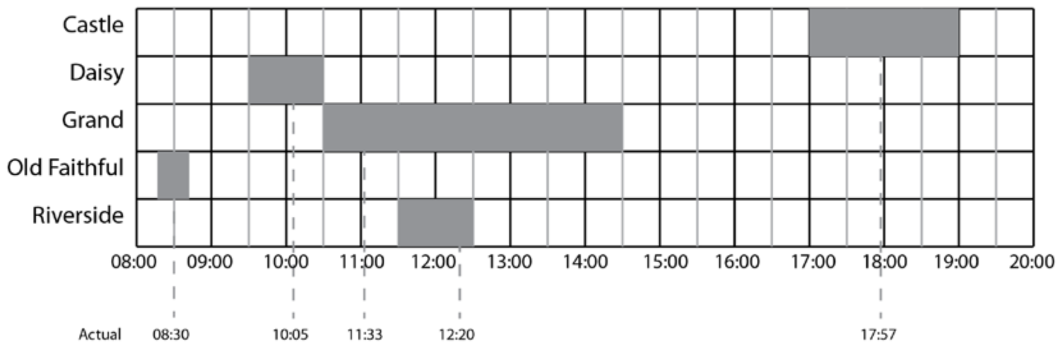
5.3 Methods

Of the visitor cases—Naive, Informed, and Expert—each subsequent case possesses more information about geyser behavior and path characteristics. To assess how this information affects user outcomes, the cases are implemented and evaluated in GIS. Results from each case then are evaluated and compared for efficacy.

The visitor cases utilize common data, but the decision making approaches differ. Common data include time, the network, network viewshed attributes, geyser behavior, and start and end location. Time for each of the cases is a 12-hour visit (720 minutes). For the GIS implementations, the time is divided into two minute increments, so each case has 360 steps of two minutes each. The visitors also travel on the same network. The UGB foot path consists of several connected circuits, totaling approximately five kilometers. For the implementation, the path has been subdivided into 160-meter long subsections in which each arc is bounded by a node. With a 4.8 kilometer per hour walking rate, each node is two minutes of travel time apart. Therefore, the time steps of each case match the travel time between nodes.

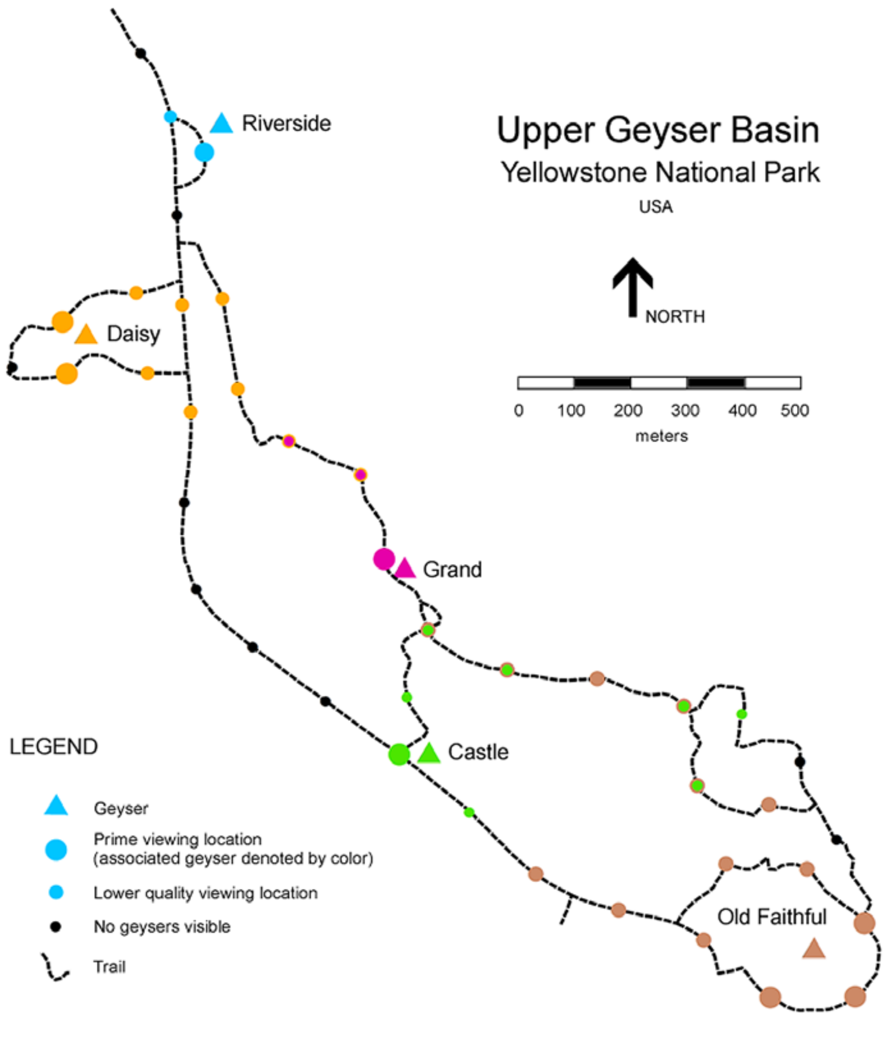
The cases share the same geysers and associated eruptive activity. Each use case includes the five geysers in the UGB with prediction windows provided by the NPS: Castle, Daisy, Grand, Old Faithful, and Riverside geysers. The behaviors of these geysers are well-studied, and a knowledgeable park visitor could couple past and current behavior to make predictions comparable to those from the NPS. Each of the GTP cases is tested using eruption data for July 4, 2012 (Figure 5.1; Appendix B). The run begins at 08:00 and ends at 20:00 Mountain Daylight Time (GMT -6 hours), corresponding to the hours the Old Faithful Visitor Education Center is open and issuing geyser predictions. This day is chosen because it maximizes information gathering capabilities; the summer daylight hours are long and the park is crowded with amateur geyser observers on the Independence Day holiday. The dataset is thus known to be complete with respect to the time period and study geysers.

Figure 5.1 Prediction windows and actual eruption times for July 4, 2012



All cases are implemented as tours, and start and end their days at the Old Faithful Visitor Education Center—node 6 in the GIS data set. The cases also share the same viewsheds, and thus, what is visible from the path. Viewshed polygons are digitized in Esri ArcGIS 10.1 by running viewshed analysis tools on a LiDAR dataset (National Center for Airborne Laser Mapping 2008) from the point representing the centroid of the geyser. The resulting viewshed raster is converted to vector polygons, and these outlines are reconciled against the vantages of a set of georeferenced eruption photographs taken from the trail. To create the viewshed data for the problem, the eruption viewsheds are overlaid on the network. From the overlay, each node is assigned a viewing quality associated with each geyser (Figure 5.2). That is, each node has five viewing quality attributes—one for each geyser. If a geyser is not visible from a node, its associated viewing quality value is zero. If a node possesses a prime viewing location, then its viewing quality is one. Suboptimal viewing is given the value of 0.5. All other GIS datasets, including geyser locations and trails geometries, were administered in ArcGIS 10.1.

Figure 5.2 Viewing locations for UGB predicted geysers



5.3.1 Naive visitor case methods

The Naive visitor case is instituted as a random walk. This visitor is considered to have no knowledge of the network or the eruptive behavior of the geysers. Thus, such information is not used to make travel decisions. For each time step, the visitor chooses either to travel along a path or to wait. If the visitor is on a path, he or she can choose to go forward, back, or stay in place. At an intersection, the visitor could choose any of the available directions or wait. Each of these potential actions is equally probable in the simulation. While allowing equal probability of waiting would be unusual in a typical GENet problem such as a shortest travel path, the UGB path traverses near hot springs, crosses a stream, and is surrounded by interesting scenery and wildlife. Visitors often stop to view the environment.

The Naive case is coded and simulated in the Python 3.3 programming language, incorporating network topology and time steps to create a 12-hour travel itinerary for the visitor. The itinerary includes time steps from 1 to 360, corresponding nodes, and standard time—the day spanning from 08:00 to 20:00 (8 a.m. to 8 p.m.) subdivided into two minute increments.

The use case compares the generated travel itinerary against two eruption data streams: a dataset of actual geyser eruptions from 08:00 to 20:00 on July 4, 2012, and a synthetic eruption dataset. Synthetic eruption data are derived from the behavior of the five geyser’s activities throughout 2011. While geyser data have been collected at Yellowstone since its creation as a national park in 1872, these data generally consist of eruption times only. Most predictions estimates are not archived. Synthetic data afford the creation of prediction windows and eruption times, including temporal schedules that have not occurred yet, but may happen in the future. The ability to create simulated eruption data also offers a mechanism to evaluate the normal or unusual nature of the July 4, 2012 real world data.

To determine the geysers observed by the naive visitor, these two eruption datasets are matched against the naive visitor's itinerary. Random walks are simulated one million times against the July 4, 2012 eruption data, and also one million times against the synthetic eruption data set. The resulting two simulations yield the average number of observations per geyser and average viewing quality.

5.3.2 Informed visitor case methods

The Informed visitor use case is implemented in ArcGIS 10.1 and Python 3.3. In the case, the visitor retrieves a prediction schedule at the Old Faithful Visitor Education Center. Using these data, knowledge about the trail network, and optimum eruption viewing locations, the visitor attempts to arrange an itinerary that puts them at the proper locations to see as many of the predicted geysers as possible.

Because prediction windows overlap and a visitor’s motivation is to decrease distance traveled and wait time, intuitive optimum solutions are rarely possible. Further, some geysers may be prioritized over others. Old Faithful Geyser is often the top priority for visitors, and rare, less frequent, geysers often are more desirable than frequent spouters.

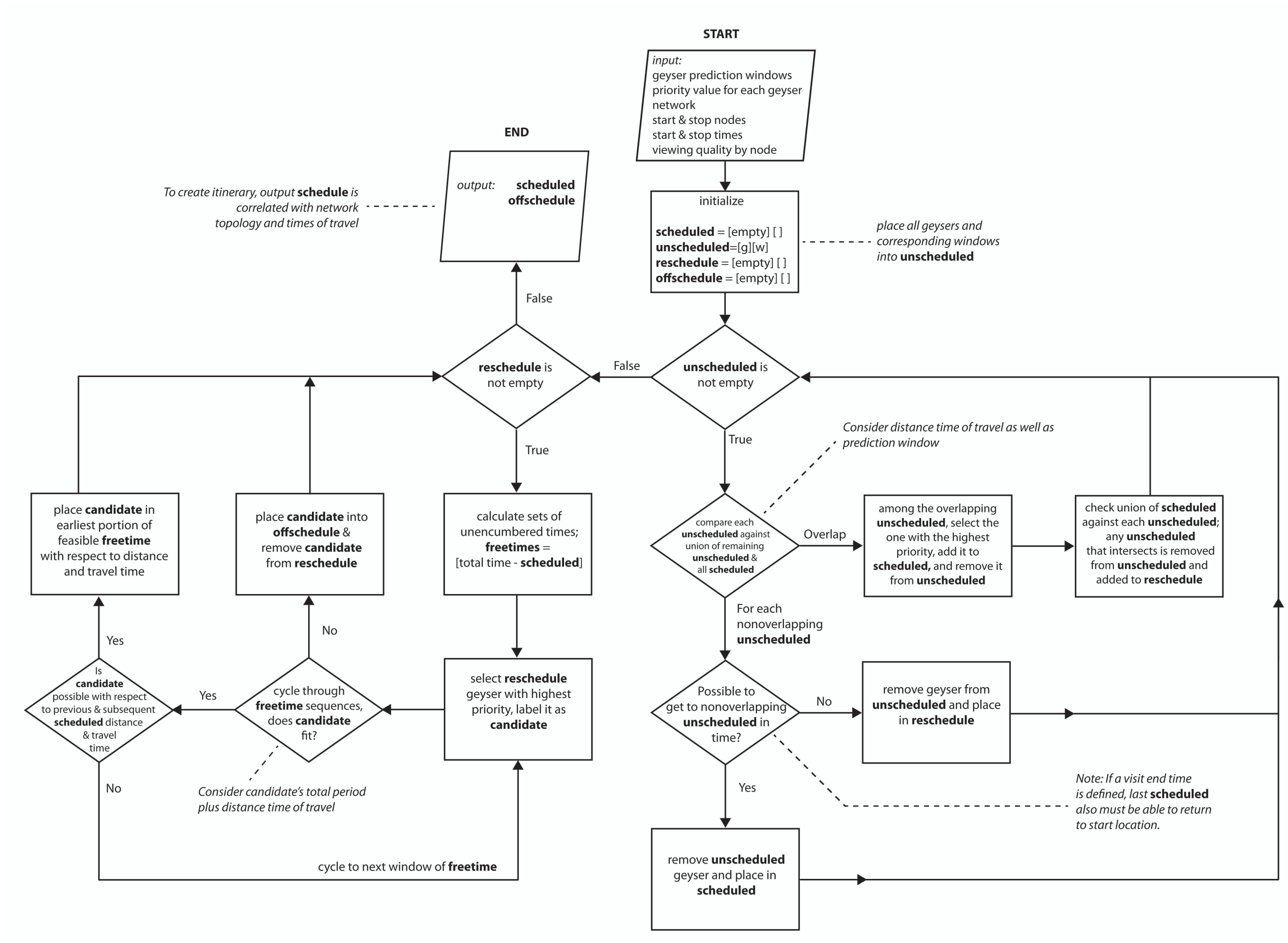
The Informed visitor makes use of NPS predictions, but does not collect additional data or develop personal eruption projections. The Informed visitor's itinerary is created at the beginning of travel and does not change.

To solve the Informed visitor's scheduling problem using GIS, a heuristic workflow is implemented. The workflow uses common GIS elements, particularly distance comparisons. The case also relies heavily on temporal comparisons; much of the workflow is spent making decisions about which geyser should be traveled to among geysers with overlapping prediction windows. Network and viewshed data remains the same as the Naive case, as it is for all the GTP cases. The GIS implementation does not offer a quantifiably optimum solution, but seeks to closely mimic the types of decisions made by interested and informed visitors to Yellowstone.

The heuristic created for the case is given the name the *Informed Scheduler* (Figure 5.3). The Informed Scheduler is a greedy heuristic; it finds an immediate optimal solution, but the global result (e.g., the completed itinerary) may be less than optimal. The Informed Scheduler requires an input of the geyser prediction windows, priority value for each geyser, the network, node viewing attributes, start and stop times, and start and stop nodes. The logic of the Informed Scheduler relies on a labeling mechanism for the sets of geyser prediction windows as they are queried in the workflow. For brevity, these groupings of geyser windows and itineraries are referred to as buckets.

To start, the problem creates four buckets. Three buckets begin empty; they are called *scheduled*, *reschedule*, and *offschedule*. Geysers prediction windows placed in the *scheduled* bucket are added to the visitor's itinerary. *Reschedule* is used for prioritizing geyser waiting times that possess travel time (too distant) conflicts or overlap a window of a higher priority geyser. Geyser windows are placed in *offschedule* if they are irreconcilable with other conflicts. Geyser windows associated with *offschedule* are not visited. The fourth bucket is *unscheduled*. All geyser wait windows are held in this bucket while they are being considered to be included in *scheduled, reschedule,* or *offschedule*.

Figure 5.3 Informed Scheduler heuristic



The Informed Scheduler works to compare all *unscheduled* geysers. If a prediction window possesses no time or distance conflicts, the wait is placed into *scheduled*. When conflicts do exist, the Scheduler selects the most desirable choice and places it on the itinerary. Remaining windows are then placed in the *reschedule* bucket. All of the Informed Scheduler comparisons consider time and distance.

The *reschedule* bucket exists to reconsider lower priority geysers in the free time remaining between *scheduled* waits. These geysers no longer fall in their known prediction windows, but a wait that is equal to the total duration of a geyser's period affords certainty that it can be observed. For instance, if Old Faithful Geyser is known to erupt every 90 minutes, plus or minus 10 minutes, a 100-minute wait would be certain to afford a viewing of an eruption. Geyser *reschedule* windows that do not fit due to distance or time constraints are removed from consideration and placed in *offschedule*.

The Scheduler continues as long as *unscheduled* and *reschedule* possess potential geyser windows to consider for the itinerary. When *unscheduled* and *reschedule* are empty, the Scheduler is finished. At that time, geysers are either on the itinerary (*scheduled*) or not (*offschedule*).

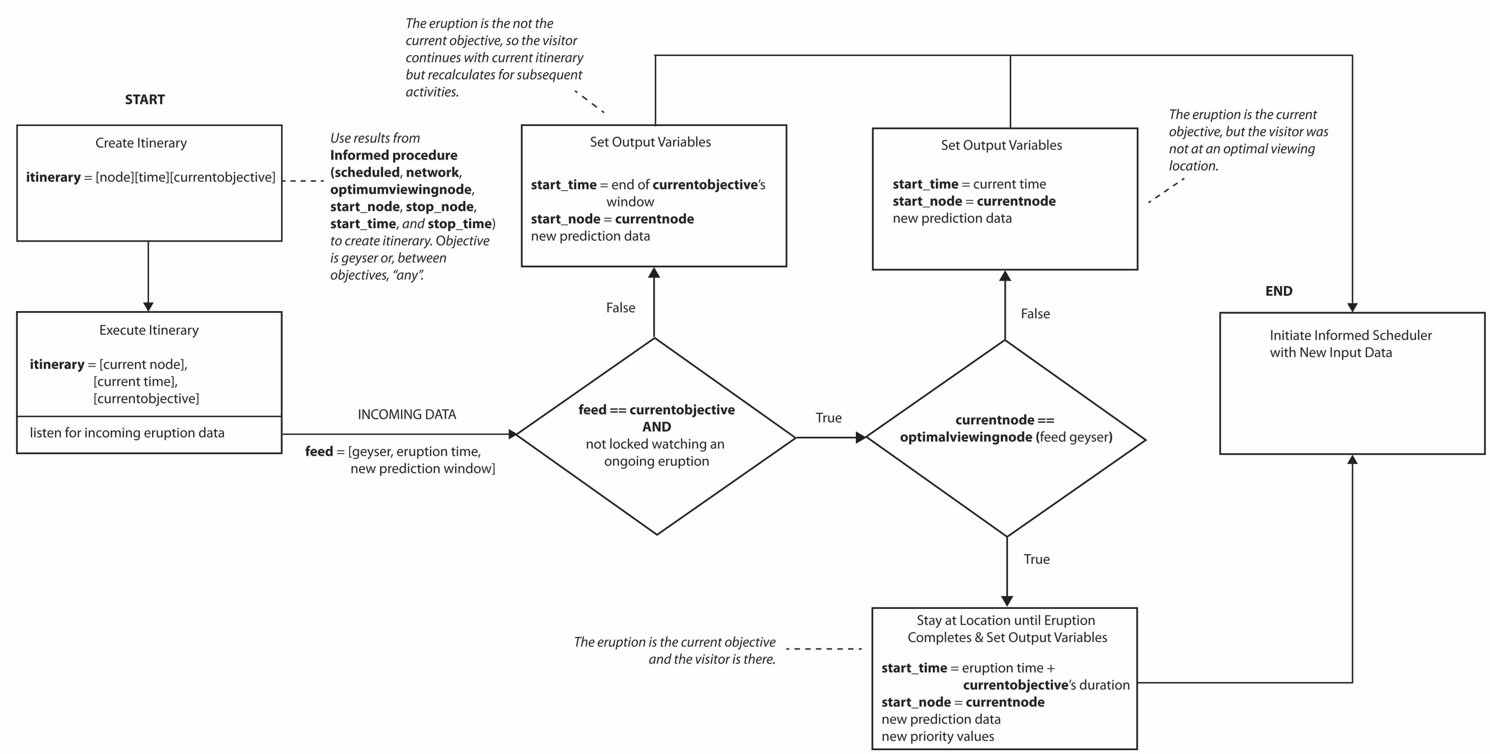
The Informed Scheduler is implemented against NPS eruption prediction windows beginning at 08:00 to 20:00 on July 4, 2012. The resulting travel itinerary is compared to geyser activity for the day, yielding a list of geysers observed, eruption times, observation locations, and associated viewing quality.

5.3.3 Expert visitor case methods

The Expert case represents visitors who collect and share in-field observations, and transform the data into actionable information—eruption predictions. To facilitate the case, an Android smartphone and web application were developed by the author to afford data sharing and to automatically calculate predictions (Glennon 2011). The Android application, Geyser Notebook, is available at no cost on the Google Play Store, and the web implementation is accessible at http://geysers.net/mobile. The associated open-source data collection project also includes Jake Young's http://geysertimes.org. These applications allow amateur observers to record eruptions on mobile devices. The data are immediately shared with other community members, and the members can rate the reliability of the observations. Incoming eruption data are compared with a geyser's previous behavior to project a future eruption window. Users calculate an in-field travel itinerary using these predictions and their knowledge of the trail network travel times. To date, such itinerary planning has been manual and ad hoc. This study's implementation of the Expert case occurs within ArcGIS 10.1, and uses Python 3.3 for additional calculations.

The Expert visitor case is tightly coupled with the Informed case, but with the addition of a trigger to handle analysis of incoming data (Figure 5.4). The Expert user begins by using the Informed Scheduler to create a travel itinerary. For each node and time on the itinerary, an attribute assigns the current objective. In periods of unencumbered time, the attribute is considered an *openobjective*. As visitors proceed following the itinerary, they continually listen for incoming data and likewise are ready to disseminate any personal observations. If no data arrive, they continue with the existing itinerary. Any incoming data are immediately evaluated against the user's current objective and the user's current location. If the current objective erupts, then the Expert watches the eruption and the Scheduler creates an updated itinerary starting at the eruption's completion. If the current objective erupts but the user is not present to see it, the Informed Scheduler is reinitiated, and new objectives are immediately set. If incoming data involve a geyser that is not the current objective, the day's remaining itinerary is rescheduled starting at the end of the current objective's prediction window. The Expert process continues until a predefined visitation end time occurs. A realistic addition to the workflow would be to end the process once all priority geysers have been observed. The Scheduler allows priority rankings, and the Expert workflow could be made to stop when the priority values were assigned the value of zero.

Figure 5.4 Expert Visitor scheduling heuristic



The Expert case is tested against NPS and crowd-sourced prediction windows for 08:00 to 20:00 on July 4, 2012. When available, NPS predictions are given precedence over community data. The travel itinerary is recalculated throughout the visit, and a single, permanent itinerary is known only when the visit ends. The finalized travel itinerary is compared to geyser activity for the day, yielding a list of geysers observed, eruption times, observation locations, and associated viewing quality.

5.4 Geyser Travel Problem results

Each visitor case yields two main types of data: geyser observations and travel characteristics (Figure 5.5). Concerning the geyser data, the visitor cases present the number of geysers observed and a quality measure. For the Naive case, quantity count is given as the mean number of a geyser's eruptions observed per simulated day. For the Informed and Expert cases, the number of observations is the count of observations for each geyser. Mean quality for the cases represents the accumulated total of view quality values for a geyser divided by the total number of observations for that geyser. Concerning travel characteristics, the case outputs include the amount of distance covered, time in motion, and number of minutes spent waiting. For the Naive case, these numbers are given as averages.

Figure 5.5 Geyser observation results from the Visitor use cases

	Naive with synthetic data	Naive	Informed	Expert
Geyser				
	Mean Observations			
Castle	0 (0.2)	0 (0.2)	1	1
Daisy	1 (1.0)	0 (0.8)	2	2
Grand	0 (0.1)	0 (0.1)	1	1
Old Faithful	4 (4.2)	3 (3.9)	2	3
Riverside	0 (0.1)	0 (0.1)	0	0
	Mean Quality			
Castle	0.6	0.6	1.0	1.0
Daisy	0.6	0.6	0.5	1.0
Grand	0.7	0.7	1.0	1.0
Old Faithful	0.6	0.6	1.0	1.0
Riverside	0.7	0.7	0.0	0.0

5.4.1 Naive Visitor observation results

The Naive case is simulated in two rounds: one using field data from July 4, 2012, and another using a synthetic eruption dataset. For the July 4 results, of the five geysers, only Old Faithful Geyser is routinely seen; Old Faithful is seen several times per day on average. Unfortunately for the Naive

Visitor, the quality of the observation is low. When Old Faithful erupts, it is not seen from a prime viewing location.

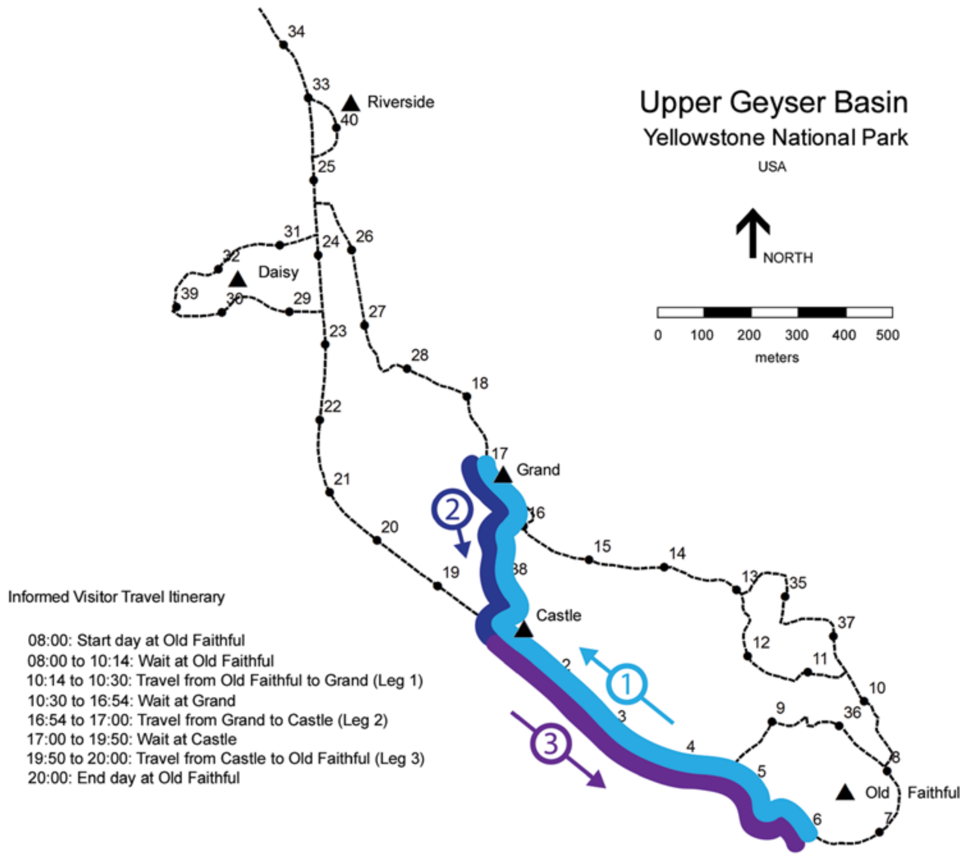
The synthetic dataset of the Naive case is run in order to assess whether the July 4 data reflects a typical viewing day. From the synthetic Naive simulations, the geysers observed are slightly different than those of July 4. While Old Faithful's quantity of observed eruptions remains similar, Daisy Geyser is now seen, but Castle, Grand, and Riverside remain unseen by most Naive Visitor simulations. The mean viewing quality for Daisy and Old Faithful are low. These geysers normally are viewed by the Naive Visitor from afar. The synthetic data show that the July 4 data offers one or two fewer geyser eruption viewing opportunities than expected. The ability to compare the planning strategies of the Naive, Informed, and Expert cases is unaffected.

5.4.2 Informed Visitor observation results

With NPS predictions and a travel itinerary (Figure 5.6) planned at the beginning of the day, the Informed Visitor case yields a higher quantity of geysers observed and a higher quality of viewing than for those eruptions of the Naive visitor. Castle, Daisy, Grand, and Old Faithful eruptions are observed, while Riverside remains unseen. Four of the six eruptions are experienced through deliberate planning, and Daisy Geyser is seen twice in the distance while waiting at Grand. Accordingly, the viewing quality of the

Castle, Grand, and Old Faithful observations hold a mean value of one, the highest possible, because all are seen from primary viewing locations. Both Daisy eruptions are observed from a distance, and its mean viewing quality reflects secondary vantages. A total of six eruptions are observed: Castle and Grand are viewed once, and Daisy and Old Faithful twice.

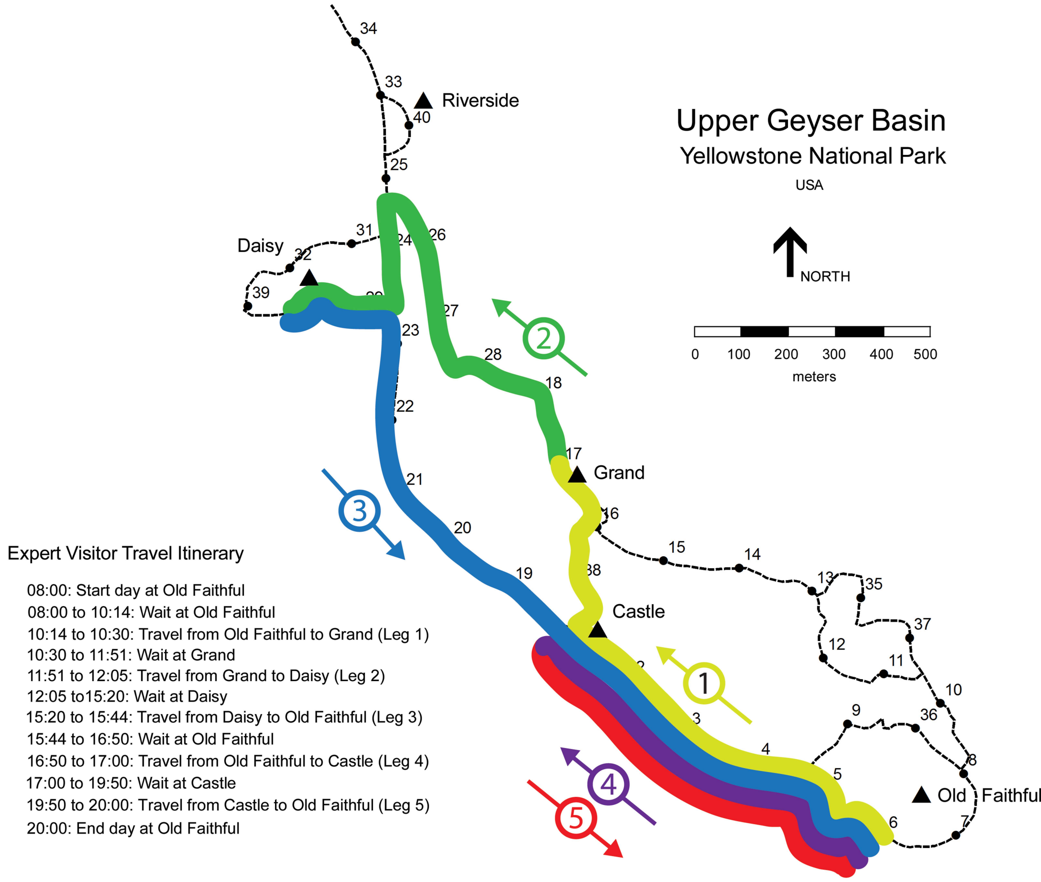
Figure 5.6 Informed Visitor travel itinerary



5.4.3 Expert Visitor observation results

The Expert Visitor continuously updates the travel itinerary (Figure 5.7) to account for new eruptions and predictions. Similar to the Informed Visitor case, four of the five geysers are seen; Riverside is not observed. However, the Expert Visitor still is able to see more eruptions and all from the highest quality viewing locations. Seven eruptions are witnessed, all through intentional planning, and every eruption observed is from a primary viewing location. Castle is viewed once, Daisy twice, Grand once, and Old Faithful three times.

Figure 5.7 Expert Visitor travel itinerary



5.4.4 Travel results

Each of the visitor cases has 12-hour visit, and the entirety of the time is spent either in transit or stopped. Based on pathway geometry, the mean amount of travel and wait time for the Naive Visitor are estimated as 69 percent in motion and 31 percent waiting. Exact paths for each of the combined two million simulation runs are not captured. These travel values slightly underestimate actual time in motion, as a higher density of path intersections reside near the start location and longer uninterrupted stretches exist farther afield. For a three-way intersection, the visitor has a choice between one of the three directions or to wait. Thus, waiting possesses a 25 percent chance of being selected. At a node along an uninterrupted reach, the visitor has fewer choices: move forward, back, or wait. Here, waiting has a 33 percent chance of occurring. Naive Visitor travel would be expected to cluster around the visitor center starting point due to its higher density of intersections. With that caveat, the Naive Visitor spends 496 minutes in motion and 224 minutes stopped. Thus, on average, the Naive visitor travels 39.7 km.

Unlike the simulation cases, the Informed and Expert Visitors follow an itinerary that affords simple recording of their travel times. The Informed Visitor walks for 32 minutes, totaling 2.6 km, and spends 688 minutes waiting. Travel occurs in three legs of 16 minutes (1.2 km), 6 minutes (480 m), and 10 minutes (800 m). Compared to the Naive traveler, the Informed Visitor's preplanned itinerary reduces the amount of travel by an order of magnitude, but triples waiting time.

The Expert Visitor synthesizes incoming data and revises plans accordingly. As might be expected, this new actionable information increases the amount of travel in comparison to the Informed Visitor itinerary. The Expert Visitor travels a total of 5.9 km and 74 minutes—a time and distance double that of the Informed case. Travel occurs in five legs of 16 minutes (1.2 km), 14 minutes (1.1km), 24 minutes (1.9 km), 10 minutes (800 m), and 10 minutes (800 m). The visitor spends 646 minutes waiting.

5.5 Discussion

The GTP implementations serve two purposes: 1) as solutions to a problem in geyser data collection and itinerary planning, and 2) as a practical application of several GENet analytical approaches within a single problem. The first part of the discussion focuses on the results and details of the application, and the second section examines the problem's general issues associated with GENets.

5.5.1 Efficacy of visitor strategies

While the GTP involves itinerary creation around geysers, strategies for its solution hold general relevance to a variety of geographic travel problems. In this regard, the terms *geyser* and *eruption* could be replaced with *objective* and *event* and often remain meaningful.

The results of the cases confirm the intuitive notion that, with respect to travel itinerary creation, consideration of timely information yields a better result. Taking action on better information leads to a higher quantity and quality of objectives achieved. Nevertheless, with the additional opportunities, there are costs in terms of increased time and distance of travel. The problem output offers measures of geyser observed, quality of eruption, distance traveled, and time waited. The heuristics implemented maximize observation quantity and quality while travel distance is considered only with respect to accessibility (e.g. can a geyser be reached in time?). An operations research optimization approach would afford strategies to balance geyser observation maximization with distance traveled, length of waits, path retracing, geyser diversity, and so on. Until these considerations are added to itinerary creation, a visitor is likely to use the Expert workflow, but only follow its schedule as it suits their personal preferences.

Increased efficacy of itinerary planning creates a peculiar problem. If everyone chooses the same best itinerary, key observations to create a later, best itinerary may be missed. This problem is actually common in viewing Yellowstone geysers. When a desirable geyser is in its prediction window, knowledgeable observers congregate, and other geysers are neglected.

Data and sharing for effective itinerary planning comes at a cost. Submitting field data is work; also, computing infrastructure must be developed and maintained, and the observer community must be continuously cultivated and encouraged. The July 4, 2012 data used by the analysis were carefully collected in order to ensure completeness. On many days, including the weeks the park is closed each winter, very few geysers are observed or recorded.

5.5.2 Implementation strategies, issues, and opportunities

These initial cases offer starting points for future, more realistic models of itinerary planning on a GENet. The cases each afford an opportunity to inventory and evaluate the additional decision making considerations based on travel preferences, uncertain time windows, and real-time data.

The Naive case, for instance, may best be framed as an agent-based simulation. The random walk is a beginning toward modeling an uninformed visitor, and future models should consider the rationality of movement. A few agent rules might include limits of retraced steps, only stopping at points of interest, and making navigation decisions mostly at major intersections and not in the middle of an uninterrupted stretch of trail. A visitor, or an agent, is also likely to congregate where other visitors are congregated. Though often mentioned in jest, a good indicator of whether Old Faithful is due for eruption is that a large crowd is gathered. Immediately after an eruption, the crowd disperses and gradually accumulates over the next 90 minutes for the next eruption. Likewise, the agent might be attracted to path characteristics. When known, and personnel are available to perform the duty, the NPS will write the next predicted time on a small sign near the geyser. That information can be accumulated and used to create ad hoc itineraries (e.g., stay and wait or come back later). Other decision making factors and behaviors exist. These should be inventoried, evaluated, and integrated in future models of visitor behavior.

For the Informed and Expert cases, geyser priority is considered in making itinerary decisions. For this study, priority is assigned based on period length (such as rare geysers are prioritized higher and frequent performers lower) and remains the same for each iteration of the Informed Scheduler. In practice, these priority values are likely to vary through the day. In particular, after a geyser has been viewed once, subsequent viewings may offer less utility. The Expert Visitor is able to change the priority values, but the most appropriate practice for doing so should be evaluated.

The GTP implementations focus on the five UGB geyser predicted by the NPS. The basins holds hundreds of geysers, and data are collected for any observed eruption. Reasonable prediction windows can be calculated in near real time by the Geyser Notebook Android application for approximately a dozen geysers. Many geysers are unpredictable, but these still can affect visitor activities. Many geysers possess interesting rock formations, strange runoff channel coloration, and beautiful pools, and these locations are points of interest even when not erupting. Unpredictable geysers and geysers with no current prediction often are included on a knowledgeable visitor's itinerary; the geyser's period may be short and waiting is likely to yield an eruption, or the visitor just hopes a rare performer will erupt. Some geysers, both predicted and not, possess known pre-eruptive behavior (Rinehart 1980). Grand Geyser, for instance, often erupts in concert with a neighboring, smaller geyser called Turban (Whitledge and Taylor 2008). When Grand Geyser is due, Turban erupts in cycles of 15 to 20 minutes. Grand is most likely to erupt at the beginning of Turban's cycle. Riverside and Daisy Geysers both overflow their craters before erupting (Bryan 2008). If the geysers are not overflowing, then the visitor is in for a long wait. To increase its utility, an ideal workflow would consider user preferences regarding scenic locations, unpredictable geysers, and pre-eruptive activity.

Improvements to the models also should include variable visit lengths. Visitors could create better suited itineraries, but also, the functionality could be used as a tool for visitor and natural resource management. Concerning visitor management, allowing variable lengths could answer questions such as: what is a visitor likely to see if they have one hour, two hours, or another specified time at the basin?; and, what is the average amount of time and travel distance needed to see an eruption of Old Faithful and one other significant geyser?

Similarly, considering observations over a variable time length could be used to evaluate the minimum amount of observation time required to assess a basin's level of activity and natural resource health. These times should also be calibrated with a basin's physiographic features, such as deep, maintained runoff channels; state of geyser cones; pool levels; discharge amounts; ground and water temperatures; and areas of dry and wet barren ground. Comparing all of these items can be used to answer the question: for a period of observation, is the activity normal? Geyser fields are imperiled and destroyed due to mischaracterization and poor management (Glennon and Pfaff 2003; Barrick 2007; Bryan 2008). Better tools for evaluating geyser field health are needed.

5.6 GENets and the GTP

5.6.1 Similar problems

The GTP is presented as a practical application that requires a broad set of GENet analytical operations for its solution. It is perhaps most closely related to series of problems in network operations research; these include the Weighted Benefit Maximal Covering Problem (Church and Roberts 1983), the Maximal Covering, Shortest Path Problem (Current, Re Velle et al. 1985), Vehicle Routing with Time Windows (Desrochers, Lenstra et al. 1988), the Median Tour and Maximal Covering Problems (Current and Schilling 1994), and the Vehicle Routing Problems with Soft Time Windows (Calvete, Galé et al. 2004).

Of this previous work, the GTP shares several attributes, but offers at least one additional consideration. Previous work considers tours along networks, both for existing networks and in the creation of networks to satisfy sets of conditions. The previous work considers time and the achievement of objectives in windows of time. Minimization of distance and path retracing are well-studied constraints. Less covered in operations research, but common in transportation science, is consideration of multiple visits to a location. In the transportation science literature, the characteristic of satiation describes the lessened utility achieved from subsequent visitation (Bhat, Goulias et al. 2012). The characteristic of the GTP not fully addressed in the existing operations research literature is the ability to achieve multiple objectives at a single location. From the GTP, more than one geyser can be viewed from one place. This characteristic stems from the GENet property of areal interaction. The area around the network affects the operation of the network. In this case, viewsheds may overlap and the associated vantages propagate to the GENet.

The Expert case of the GTP also considers real time, incoming data; the workflow listens for data, parses it, evaluates it, and acts on it. In computing, such an ongoing process is called a *daemon*. With the increase in mobile and locationally aware computing, the utility of such daemons in GIS are likely to be common parts of the workflow. As such, the use and nature of daemons for geographic analysis should be studied further.

In particular, the GTP cases involve a mix of travel and waiting. For the Informed case, the visitor waits for the entire duration of the window. Even after the geyser erupts, the window encumbers the visitor’s schedule. In the Expert case, however, if a geyser erupts, the visitor watches the eruption and then is released to continue with other objectives. The allowance for a user participating in a task of uncertain duration to be released upon the completion is not well addressed in spatiotemporal GIS.

The novel characteristics of the GTP cases include handling uncertain task length, achieving multiple objectives at a single location, and considering recurrent visits. GTP-inspired approaches could be used to improve automated itinerary creation; if a task takes shorter or longer than expected, daily activities can be accordingly adjusted. Such an approach could be used in activities where a traveler targets the occurrence of an uncertain event—for example, police patrols and criminal activity, fishermen and fish, and vehicles and traffic jams.

5.6.2 GENet analytical approaches

This dissertation focuses on several GENet analytical operations, including process modeling, simulation, comparison, and optimization. All are engaged in the problem. Further, embedding and intertwining the operations with each other occurs without significant difficulty.

In the context of this dissertation, process modeling is framed as a transformative operation. In the GTP cases, itinerary planning is conducted in such a manner. A scheduling process model combines the network geometry, viewing priority, and prediction windows to create a travel itinerary

For the Naive case, simulation and comparison are used; the itinerary of a random walk on a GENet is compared against the day's geyser eruptions.

As programmed, the case conducts the walk then performs a comparison for the number and quality of observations from the route. A rearrangement of code would allow the comparison to occur in step with the random walk. The simulation also could be modified into an agent-based model by incorporating deliberate decisions during the run.

Comparison associated with the problem's GENet includes the overlay of the viewshed with the network. From the overlay, viewshed properties are inherited by the network nodes. Solutions of the GTP also require the evaluation of accessibility via comparison of a traveler's current location against travel distance and access time to a spatiotemporal objective.

The GTP is closely related to several network optimization approaches in operations research. In this study, the Informed and Expert cases use a heuristic instead of a mathematically rigorous linear programming optimization. Using a linear programming optimization, success of the cases with respect to the balance of objectives can be more readily evaluated. In this regard, this chapter offers a starting point in enumerating the objectives and constraints of an associated linear programming (LP) optimization formulation. An optimization could be integrated into the Informed and Expert cases. An LP optimization would replace the Informed Scheduler, and find the ideal balance based on visitor preferences of eruption priorities, travel distance, and wait time. An LP optimization also could weigh issues not easily handled by the existing workflows. For instance, suppose a geyser has a one hour prediction window. Considering a real time schedule, arriving as late as possible is ideal in that it means the least amount of waiting. However, with a late arrival, it increases the risk of missing the eruption. An LP approach could balance such risk against the other opportunities available in that time period.

5.7 GENet characteristics and the GTP

Scale and spatial uncertainty are representational issues that affect GENet analysis. Likewise, physical characteristics of GENets can influence operations and results. GIS implementations may be impacted by all four of these factors, and thus modelers should consider GENet properties when designing data collection strategies, storage, analytical operations, and output.

5.7.1 Scale

Spatial and temporal scales are leveraged in the GTP implementations to harmonize network characteristics, distance, and travel times. For example, the network is subdivided into sections of equal length, so travel time along the trail is the same between any two adjacent nodes. Given the geometry of the path and spans between points of interest, a larger distance between nodes would decrease the utility of the model. Geyser observations typically arrive at a one minute resolution, and that increment would perhaps be the most natural scale for the data. However, such environmental resolutions are balanced against the associated computational burden. In this regard, 40 nodes are more tractable than 80, particularly with respect to running numerous iterations with synthetic data. The project data are also intended for experimentation and future use within a linear programming optimization, so reducing computational load is a priority.

5.7.2 Spatial uncertainty

With respect to issues of spatial uncertainty in the problem, the nature of network nodes and geyser observational data are concerns. First, the GTP cases' network nodes are abstract. In the field, these nodes do not exist, and in the problem, these nodes reflect the characteristics of the area around them. Each node possesses attributes for the viewing quality of each geyser at their location. Despite the distinct labeling of a node's viewing quality for a geyser as values of 0, 0.5, and 1, reality is more nuanced. There is no clear delineation between primary and secondary viewing locations. This leads to ambiguity of what can be achieved at the node's location and where one needs to be located to receive its attributed utility.

The GTP cases are crafted to avoid an incomplete dataset for the study period's reporting. However, incomplete data for many geysers is the norm. Old Faithful is the only geyser with a record of active, continuous monitoring. For itinerary planning, incomplete data lead to missed opportunities and higher costs for achieving objectives (i.e., a longer wait). Further, real time data must not only be collected, but it also must be shared and evaluated within a relevant time period. Data exceeding one or more of a geyser's periods offers insight on the behavior of a feature, but it is not otherwise actionable. Erroneous reporting via misidentified geysers and improper times is common. Such errors propagate into eruption projections and itinerary planning. With respect to incomplete data, it is sometimes possible to devise a prediction on a double period. The Geyser Notebook app will perform such a calculation when consistent behavioral data exist. With a double-period prediction, the eruption window becomes much longer, but the additional data nevertheless can be considered in itinerary planning. So far, erroneous data have been detected manually by the community of observers. Geyser Notebook and geyser websites include the ability to flag observations as suspicious. If the number of flags exceeds the number of observers for an eruption, its data are excluded from the behavioral and prediction model. Automated evaluation of observations with respect to behavioral expectations is under development. Despite these efforts, geysers sometimes significantly alter their behavior without warning. These cases necessitate rapid identification and consideration of such changes into the behavioral model before they affect the navigation of large numbers of visitors.

5.7.3 Areal interaction

The trail and the geysers interact in the GIS implementation of the GTP. Foremost, the geyser eruption viewsheds are areas that affect the utility of the network. At the intersection of the viewshed and GENet, the network's nodes inherit attributes associated with the view. In this respect, the trails have been built by the NPS to serve the viewsheds. For the five study geysers, their viewsheds have remained stable for at least a decade. However, Yellowstone's geyser fields are dynamic landscapes, and sometimes activity at one location wanes while a nearby area increases. These changes can affect the utility of the trail, and in several instances, the pathway has had to be moved to accommodate a new hot spring or geyser. Other than the indirect social effect of litter and debris, the trail network itself does not have any apparent effect on the operation of the geysers. However, this is perhaps due to active efforts to protect the geysers. On a road approximately 10 kilometers north of Old Faithful, road builders cut through rock to maintain the level road. The 0.5 meter cut exposed some small parts of nearby Pink Cone Geyser's plumbing system. When Pink Cone now erupts, a few holes on the roadside sputter.

Specific to the GTP, itinerary success can be affected by spatial correlation amongst geysers. Grand, in particular, is affected by its neighbors (Bryan 2008; Whitledge and Taylor 2008). It erupts in concert with adjacent Turban Geyser, but can be delayed by several of its other neighbors. A likely explanation for the regularity of Old Faithful Geyser is that it, being alone on a hill, possesses plumbing distant and separate from the influence of other geysers. Observers have noted many trends, but a systematic geostatistical examination of geyser interactions has not been undertaken. Such a study would afford better solutions to the GTP and reveal new information about Yellowstone geyser operation.

5.7.4 Constituent heterogeneity

GTP implementations are affected by the functional and physical heterogeneity of the network. The existence of different objectives—the various geysers—along the network is one such form of internal variation. A distinctive characteristic of GENet within the GTP cases, is the ability to achieve multiple objectives at the same location, including the same time; that is, a visitor can view more than one geyser from a single vantage. Further, the utility value of objectives varies along the network: the viewing quality for each geyser differs from node to node. Such heterogeneous function is considered in this study's GTP implementations.

The real world naturally is more complex than a GIS model, and network heterogeneity characteristics may emerge via innumerable geographic interactions. For example, vantage quality varies in response to factors such as weather, sun angle, and crowds. Watching a geyser involves directional viewing from an observer's location to the geyser. The quality of the line of sight, for example, can be diminished or enhanced by relative sun position: reduced when staring into bright sunlight and improved when positioned properly behind creating a rainbow off the geyser spray. Viewing location recommendations that consider not only current geyser predictions, but also weather and lighting conditions, are a popular feature request for the Geyser Notebook application.

The UGB pathways vary in terms of visitor affordances. Some of the trails are wide, some stretches are narrow, and other locations possess benches for waiting. Each of these conditions implies intended path function: a wide trail is better for slower and two-way travel; narrow stretches are suited for single direction traffic; and, benches are placed to facilitate comfortable stop locations. These conditions reinforce activities that occur on the trail. Rapid movement necessitates concentration on navigating the trail itself, while slower travel and waiting allow focusing outward, away from the trail, and into the surrounding area. Such function and the consideration of the variable speeds of travel will enhance the visitor itinerary creation process. Likewise, travel decisions tend to occur at intersections and waiting locations, rather than in the middle of straight pathway stretches. Integrating such heterogeneous properties of the network can improve analytical results.

CHAPTER 6:
DISCUSSION AND OUTCOMES

6.1 The continuum of Physical and Abstract GENets

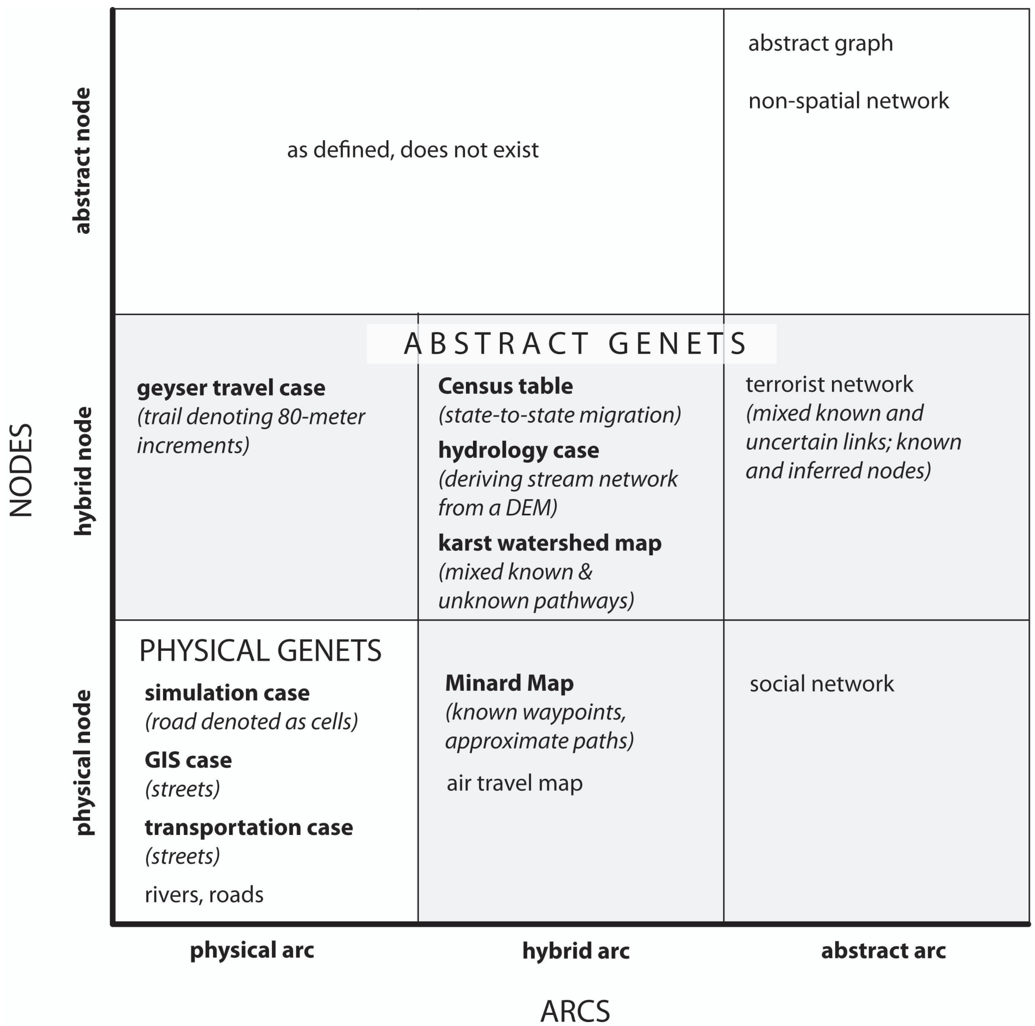
In this dissertation, GENets are categorized as either Physical or Abstract. Physical GENets are composed of pathways and intersections that tangibly exist in the environment. Abstract GENets allow both conceptual and physical elements. The categories were created to help evaluate appropriate analytical operations and the effects of geographic considerations on network constituents. This section endeavors to clarify the categories, discuss associated issues, and suggest areas for future research.

Networks in geography can be considered from two perspectives: 1) GENets as they reside in physical space—a tangible existence perspective; and 2) GENets as they are abstracted and modeled—a representational perspective. Whether a GENet is Physical or Abstract can be confusing because, despite these perspectives, all GENets must exist as representations in order to be analyzed. Thus, a useful approach for distinguishing between Physical and Abstract GENets is to test for physical existence; that is, could the network phenomena exist in physical space without abstraction or modeling. Networks that pass this test are Physical GENets—likely to affect and be affected by the space they inhabit. All remaining networks are

Abstract; they might not be divorced from geographic influence, however, and numerous circumstances exist that blur the lines between physical and abstract presence. Hybrid cases occur, for instance, when a network's nodes physically exist and its arcs are conceptual, or vice versa (e.g., tangible arcs with conceptual nodes).

Separately examining a network's constituents, nodes from arcs, may help delineate the bounds of abstract, physical, and hybrid network cases. Within the framework of nodes and arcs of abstract, physical, or hybrid types, nine potential arc-node combinations are possible (Figure 6.1). These combinations form networks that can be categorized as Physical GENets, Abstract GENets, or non-spatial networks. The subsequent section defines abstract, physical, and hybrid nodes and arcs. The definitions follow the logic that abstract constituents are conceptual and independent from geographic considerations; physical constituents are embedded in geographic space; and hybrid constituents are dependent on geographic space but possess conceptual aspects. The definition of hybrid nodes and arcs attempt to describe several of the circumstances under which such constituents may arise. The list is intended to offer initial guidance on the constituent's likely interaction with geographic space, and may not be exhaustive.

Figure 6.1 Characterization of networks based on arc and node type



6.1.1 Abstract, Physical, and Hybrid nodes

Abstract nodes are conceptual labels. With respect to Goodchild's spatial tests (Table 4.2, 2012), such nodes do not vary upon relocation, do not possess concepts of location in their representation, and do not modify the landscape upon which they reside.

Physical nodes are tangible, focal areas. While use of nodes in GIS implies a precise point, the term *node* derives from the Latin word for *knot*. With respect to this meaning, Physical nodes have dimension, and may be objects or the intersection of tangible arcs.

Hybrid nodes are conceptual focal areas that are dependent on geographic space. The representation of an areal expanse as a single node is one circumstance that yields such nodes. For instance, regions, such as cities and states, sometimes are represented as single points. The geographic location of the point retains meaning—though altered and condensed—particularly relative to other regions that have been similarly represented. A Hybrid node may also be associated with a conceptual, geographically embedded arc—such as a political boundary or movement track.

6.1.2 Abstract, Physical, and Hybrid arcs

Abstract arcs are non-spatial, conceptual linkages. Such links do not change character upon relocation or with differing scale, and do not interact with the physical environment. Abstract arcs may offer conceptual linkages between geographic actors, such as the case of social links between located individuals.

Physical arcs exist in tangible space. Such arcs possess geographic location and change in character with relocation. Common examples of physical arcs are rivers, roads, and sidewalks.

Hybrid arcs are embedded in geographic space, but are not clearly tangible. Hybrid arcs occur in at least four situations: conceptual arcs for social or analytical purposes, routes of individual movement, aggregations of other geographically embedded arcs, and tangible arcs with speculative or uncertain locations. First, geographically embedded, intangible arcs can be created to serve social or analytical purposes. Examples of such conceptual arcs include survey transits, property lines, and political boundaries. Such arcs do not communicate directly with their surroundings, but may drive interaction through social activities related to their purpose. These types of Hybrid arcs are subject to scale and spatial uncertainty issues in their representation, and spatial concepts such as the measurement of distance over their length are meaningful. Second, though tracks of individual movement are not tangible, particularly in the same manner as a river or sidewalk, they are embedded in geographic space. The traveler interacts with the environment at all moments of travel and the final route over its time window reflects these accumulated relations. Third, the aggregations of arcs embedded in geographic space may be represented as a simplified arc that retains many of the embedded properties. The product aggregation, however, does not necessarily respond to the original landscape in the same manner as a Physical arc traversing the same path. Fourth, Hybrid arcs may represent pathways that are embedded in geographic space but have uncertain or speculative locations in space or time. In the karst watershed use case (Figure 3.2), some cave streams were mapped and others were inferred. The inferred routes exist, but their exact pathways, and thus the exact locations of the embedding are not known. Physical arcs that no longer exist, such as abandoned roads or rerouted rivers, may be considered Hybrid arcs. Speculative network paths that would be Physical arcs if realized may be also considered Hybrid arcs. For instance, in the planning process, a proposed road would be an actual road if built. Operations, representations, and analysis germane to a Physical arc would be appropriate.

6.1.3 Arc-node combinations

The defined arcs and nodes create nine potential network combinations (Figure 6.1). Six of these are GENets, two of the combinations cannot exist per the constraints of abstract nodes, and the remaining combination is non-spatial.

As defined, abstract nodes are non-spatial, conceptual labels. Combined with an abstract arc, abstract nodes and arcs create non-spatial networks. A social network with no geographic underpinning is an example, as well as a graph linking movies by their shared actors. Networks that combine abstract nodes with physical or hybrid arcs do not exist. If an abstract node were associated with a physical or hybrid arc, the node would inherit the arc's location and cease to be non-spatial.

Physical GENets are composed of the combination of physical nodes and arcs. These networks are likely to interact with their surroundings and possess internal heterogeneity. They also are subject to geographic representation issues of scale and spatial uncertainty. Among the eight cases examined in this dissertation, three possess physical arcs and nodes and can be classified as Physical GENets: transportation case (Figure 4.4), the GIS case (Figure 4.6), and the simulation case (Figure 4.8). In Figure 6.1, use cases examined in this dissertation are highlighted in bold.

With respect to constituent node and arc types, Abstract GENets are composed of the combinations: physical nodes-hybrid arcs, physical nodes-abstract arcs, hybrid nodes-physical arcs, hybrid nodes-hybrid arcs, and hybrid nodes-abstract arcs. These combinations represent the numerous cases where a GENet possesses a conceptual constituent. Five of the dissertation's eight use cases are built upon Abstract GENets, including the Census migration table (Table 3.1), Minard's Map of Napoleon's March on Moscow (Figure 3.1), the karst watershed case (Figure 3.2), the hydrology process model case (Figure 4.2), and the geyser travel case (Chapter 5).

Of the analysis cases described in Chapter 4, the three that begin with Physical GENet inputs all yield outputs of Abstract GENets—with hybrid arcs and nodes. The transformation of GENets to a conceptual form during analytical operations should be expected. Modeling requires abstraction, and many, if not most, solutions are speculative and intangible. The geyser travel case input includes the physical arcs of a visitor trail and hybrid nodes (locations marked at 80-meter increments). The itinerary creation process for the geyser case transforms the network's arc type from physical to hybrid: the output proposes a potential path itinerary of hybrid arcs and a set of associated hybrid nodes. Finally, the process model case—deriving a stream network from a DEM—possesses no input network, and its output is an Abstract GENet comprised of hybrid arcs and nodes.

Future research

Via use case studies, this dissertation has examined several Physical and Abstract GENets. All have been composed of physical or hybrid arcs, and no cases have incorporated abstract arcs. In this regard, GENets of abstract arcs require further study with respect to the circumstances in which they are used, their associated representations, and their appropriate operations. Any two geographically embedded entities can be related through abstract arcs, and as such, uses are basic, widespread, and of general importance. For example, Abstract GENets with abstract arcs encompass a wide variety of communication cases, including person-to-person information transfer. Associated cases also include spatially aware social networks—an area of increasing importance with the spread of mobile computing and online social networks.

Given the variety of hybrid network circumstances, further work is needed to develop an exhaustive inventory of conditions that define hybrid arcs and nodes. Abstract GENets derived from hybrid arcs and nodes represent the vast majority of networks in geography. A thorough classification will assist in determining appropriate GENet operations and representations. A related effort should endeavor to weigh the influence of areal interaction, network constituent heterogeneity, scale changes, and spatial uncertainty under different GENet conditions.

6.2 Themes

GENets are affected by the representational issues of scale and spatial uncertainty. Scale influences the granularity of analysis and the operations and manipulations that can be performed on a network. Spatial uncertainty may alter the contextual meaning of data, and affect the way GENet data are stored, displayed, and interpreted. GENets that exist in tangible space, like roads and rivers, influence their surrounding areas and their neighborhoods influence with them. Such GENEts are heterogeneous in terms of attributes, function, and the geometry of constituent parts. The purpose of this dissertation is to uncover such GENet characteristics, and in that pursuit, several recurring themes appear, including data models, case studies, and analytical operations.

6.2.1 Data models

Geographic data models facilitate the assignment of meaning to vector geometric primitives—points, polylines, and polygons, in GIS. In a geographic data model, relationships are drawn between an entity's geometric primitives and descriptive classes. Such a model can be as simple as assigning a polyline an attribute or class that gives it meaning as a road. More often, it serves to relate associations of multiple geometries and meanings. For example, a geographic data model of a watershed might include a river confluence (point) existing when more than one river (polylines) intersect within a catchment (polygon). The formalization of such relationships often occurs in UML. With respect to geography, UML operators, specifications, and examples for delineation of classes and relationships are found in Alexander (2002) and Arctur and Zeiler (2004).

6.2.2 Case studies

Case studies are used throughout this study for examining GENet data structures and analytical operations. Examining use cases to facilitate system design is a common practice within computer science (Weisman 2003), and is adopted in this dissertation to evaluate GENet organization. For this, the GENet cases are deliberately chosen to be as diverse as possible. The intent is to select cases that have no or few common elements except the study phenomenon. A formal model is created for each case in UML. Aspects relevant to GENets then are inducted from each specific model into a general model. Any superfluous details of the specific case studies are discarded. The new model offers the combined, distilled generic aspects of the study phenomenon. Success of the process relies on the careful selection and reconcilability of the case studies. When complete, the general model offers a more sophisticated and revealing object for study and manipulation than the distillation of relevant components of any single case.

6.2.3 Analytical operations

Geographic analysis of networks focuses on process modeling, simulation, comparison, and optimization. Analytical operations elicit interactions between a network and its environment. The process may also offer insights into the network's internal workings. While geographic data models offer meaning, they may conceal implicit abilities. For instance in the GENet flow model, magnitude exists as an attribute of flow. Besides this raw data value, gross and net magnitudes may be garnered using an analytical operation—combining or subtracting flows from the opposing direction. Analysis thus offers a complementary tool to data models for the understanding of GENets and other geographic phenomena.

6.2.4 Computing to constrain complexity

For the examination of case studies and analytical operations, this dissertation takes a computational approach. Computing and computational models are rigid in their requirements—input must be clear and precise. A computing platform demands order. The domain of geography meanwhile fosters ambiguity, complexity, and fluidity. To analyze its workings, geography's social and physical landscape must be simplified. Computation offers an appropriate foundation in this regard. Despite the strict input requirements, the computing environment affords modeling of structure, behavior, and operations.

A geographic model within a computational environment is suited to store only a limited amount about an entity's properties and relationships. The modeler must make choices about their entity's most fundamental attributes and important relationships. These chosen aspects are included in the model and made available for investigation. Some, and hopefully many, of the entity's relationships are defined. Through such computational modeling, a portion of the geographic world is made accessible for study and manipulation. The choices left out of the model need not be forgotten. Though these remaining characteristics cannot be operated upon with the same rigor as the computational model, these portions can be evaluated with respect to model realism and completeness. The act of making such choices may also serve to define an entity's internal and external character. The violence done upon the entity to fit it into the strict constraints of a computational model often offers insights into the nature of an entity.

Defining a model purpose, making choices about the most important aspects, enumerating those choices formally by creating a model, considering the conditions not included in the models, and evaluating each of these steps, is a recurring approach used in this dissertation.

6.3 Undercurrents

6.3.1 Software

Modern analytical geography necessitates the use of computational modeling and software development. Associated accessible software is fundamental to facilitating thoughtful review, and a wide community must possess the skills to not only use, but evaluate such content. Theoretical and practical geographers must be able to engage, understand, and develop programming code. Science relies on verifiability through reproducibility and the improvement of methods via experimentation. As such, techniques and software created by scientific geographers should be open source and made accessible through suitable distribution channels. In *Nature*, Ince, Hatton, et al. (2012) argue that in respect to scientific research, “anything less than the release of source programs is intolerable that depend on computation”. Adopting this philosophy is required not only to promote vibrancy within the discipline, but it is also demanded for interaction with the broader scientific community.

In support of this dissertation, the following open software tools and technological platforms, accessible at: http://alanglennon.com/genets, were created:

- Code in Python 2.7 for calculating two-way, gross, and net flow from an origin-destination matrix, and yielding associated GIS feature data
- Code in Python 3.3 for simulating geyser eruption data based on previous behaviors
- The first mobile crowdsourced geyser observation platform; developed in Android Java and php
- The first publicly-available system for automated geyser prediction
- With colleague Jake Young, an open source geyser eruption data archive at http://geysers.net/mobile and http://geysertimes.org

6.3.2 Human thinking and GENets

The breadth of scientific research on networks is so vast that no single work can synthesize it all. This dissertation charts a narrow course through this wide area of knowledge. To constrain the complexities of social and physical reality, computational analogies are emphasized, and thus, the ways humans consider and manipulate GENets largely are unaddressed.

The computational requirements for operations on a network database—cryptic tabular or numeric arrays—are poorly suited for human understanding and manipulation. In this regard, a map or drawing offers a concrete object for interpretation, experimentation, and communication. By design or through interpretation errors, such maps hold potential to mislead

(Monmonier 1996). The following is an initial inventory of circumstances that may lead to misinterpretation in network maps. The list is non-exhaustive and intended as guidance for future research. As Egenhofer and Mark (1995) have described more general cases of intuitive geographic thinking, this inventory extends their work with regard network-related map reading. The situations outlined should be considered not only as an opportunity to avoid potential map design problems; the issues also offer possibilities to reinforce proper messaging via intuition. That is, simply because an issue may lead to misinterpretation sometimes, does not mean the interpretation is always incorrect. In cases where intuition and analytical results agree, conclusions will resonate.

Networks may distort metric space

Some GENets are represented as schematics, not preserving all aspects of metric space and distance. Many such maps, like Beck's London Tube map, are celebrated for their elegant design. Londoners have deep affection for the Tube Map and protested vigorously when a change was made in 2009 that removed the Thames River. London's Mayor Boris Johnson was reportedly furious about the change and upon learning of it during a foreign visit declared: "Can't believe that the Thames disappeared off the tube map whilst I was out the country! It will be reinstated". The Tube Map distorts distance and shape of subway routes to emphasize route topology. Using travel behavior data from 1995-2005, Guo (2011) found that people trust the schematic more than their own experiences, routinely taking longer routes to their destination in correlation to the distortions of the map. Guo's finding reiterates Egenhofer and Mark's (1995) assertion about intuitive geographic thinking: "maps are more real than experience."

Networks imply regions

The connection between networks and areal coverage has been well established in geography and parallel disciplines, including hydrology (Haggett and Chorley 1970), regional studies (Christaller 1933), transportation (Garrison and Marble 1961; Miller and Shaw 2001), operations research (Church and Roberts 1983), and GIS (Worboys and Duckham 2004). Imagination facilitates creative representations of the environment, and networks may reflect not only the connection among specific geographic entities, but also the totality of the region they inhabit. Visual representations of networks may imply regions—whether intended or not.

Networks imply systematic order

While a network may have the appearance of an overarching order, the pattern may be the product of random processes. Common morphometric measures from hydrology have been found to be similar whether the stream was created by a deterministic or stochastic process (Kirchner 1993). In complex geographic systems including networks, Goodchild (1992, page 150) states that the action of underlying processes may be more clearly confirmed by deviations from the expected orderly arrangement.

The sum of network parts is less than the whole

Dillemuth (2009) found that map readers exposed to only a portion of a road network on a small display interpreted with less accuracy than readers exposed to the full map. However, the smaller map had no effect on the participants' confidence in their performance.

People make assumptions about incomplete networks

The complete extent of a network is not always known, and it is common to represent only portions of a system on a map. The complete network may be unknown due to untracked changes, incomplete knowledge, representational simplification due to scale, or other compromises made during the modeling process. Egenhofer and Mark (1995) assert that people are accustomed to incomplete geographic data and have incorporated uncertainty into their spatially-related decision making. Concerning navigation, Dillemuth (2009) found that incomplete information impedes map interpretation and recall with no effect on participants' perceived confidence in their performance. People make assumptions about incomplete networks, and how the task is performed is likely to differ by person and context.

Network distance may not equal metric distance

In an experiment by Fabrikant, Montello, et al. (2004), the researchers look at the distance-similarity metaphor in a graph representation. They describe several notions of network distance, including direct metric distance (straight line between nodes without regard to the network), network metric distance (shortest path between two nodes along the network), and network topology (number of intermediate nodes or other measures between two points of interest). The study found that when a graph representation is used, the distance-similarity analogy holds strongest with a network metric distance. That is, when reading a network map, closer things along the network are interpreted as more similar than features with equivalent direct metric distance.

Network links show connectivity or the exact opposite

The three traditional geographic subtypes of network—branching, circuit, and barrier— afford different functions (Haggett and Chorley 1970). While branching and circuit networks are most often associated with pathways and connectivity, barrier networks denote separation. The archetype barrier network is the political boundary. When a mix of network types is represented on the same map, interpretation speed and accuracy may be affected. Thoughtful cartography strives for each map symbol to have clear meaning, but in many contexts, network function ambiguity effects will persist.

6.4 Research contributions

The contributions of this dissertation are:

- The definition of GENets with respect to their physical and abstract properties
- The development of a geographic data model for GENet flow
- The identification of an initial set of GENet characteristics and description of methods for uncovering more such properties
- The description of a method for developing geographic data models
- The creation of a GIS workflow and heuristic for addressing real time data flows in spatiotemporal path itinerary creation
- The development of a GIS workflow and heuristic for network path itinerary creation that addresses: location-objective replenishment allowing beneficial recurrent visits along a GENet; achievement of utility from multiple objectives at a single location; and activities with uncertain completion times

6.5 Areas for future research

This dissertation identifies several areas of GENet modeling and analysis that merit consideration for additional research. Techniques described in the dissertation, particularly geographic data model creation and case study analysis, are intended to be used by others toward their own specific domains. The techniques also may be used, modified, and improved to reveal additional fundamental properties of GENets.

6.5.1 Optimization and time comparison

As GIS is a platform for performing spatial operations, its algorithms and data models are objects for scientific consideration (Goodchild 1992; Mark 2003). To encourage thinking about new approaches to GENet analysis and spatiotemporal operations, two specific functions are suggested to be incorporated into GIS: an optimization solver and time comparison.

Among the GENet analytical approaches emphasized in this dissertation, each has been made as accessible as general toolkits within GIS. Comparison operations are included as core functionality for most GIS platforms. Simulation and process modeling are possible in tools such as Esri ModelBuilder. Optimization, however, has only been integrated with respect to specific domains, like transportation cases within Esri Network Analyst. To introduce users to optimization operations, functionality might solve a limited set of linear programming cases that parallel the functions of familiar GIS overlay operations, such as identifying the best location among several similar choices.

Another simple software function that would promote temporal thinking with respect to GIS is time comparison. Existing GIS temporal functionality emphasizes visualization, and rarely offers analytical output. While there are many types of times in GIS (Frank 1998), operations that evaluate whether one well-defined event occurs before, after, or within another exist in high level programming languages and likely would be simple to incorporate into existing GIS code bases. Such comparisons also offer an entry into more sophisticated temporally-enabled GIS operations. For example, in the GTP implementations, the Python 3.3 commands *in* and *intersection* allow the comparison of two data arrays with respect to whether and when time windows overlap. Adding optimization and time comparison functions would afford a large and interesting new set of problems to be considered in GIS.

6.5.2 Egocentric geography

GIS implementations of the GTP require spatiotemporal data to be evaluated in real time in the field. Starting with efforts such as NCGIA Project

Battuta in 2001, the workflows and impacts of ubiquitous computing continue to be an active area of GIScience research. Also, the notion of what actions can be performed within a spatiotemporal context—Gibson's concept of affordance—is relevant (Gibson 1977; Jordan, Raubal et al. 1998; Howarth 2008). The combination of ubiquitous computing, spatial analysis, and affordance yields a question of import in the future of GIScience; that is "what is possible here?" Third person type perspectives initiated the development of GIS, and now egocentric geographic questions are inspiring new techniques, requiring the translation of existing techniques, and offering areas for new geographic discovery.

6.5.3 Networks in geography

The study of networks, and in particular GENets, is worthy of ongoing, comprehensive review by the discipline of geography. Modern surveys concerning networks in geography tend to focus on specific domains, like transportation networks (Miller and Shaw 2001); network algorithms (de Smith, Goodchild et al. 2007); or emerging topics from outside the discipline, like network evolution (Batty 2005). These are welcome additions to the GENet literature, but a multi-researcher, interdisciplinary inquiry is overdue. The most recent comprehensive work on networks in geography is Haggett and Chorley (1970). A modern GENet synthesis should span not just the geographical approaches relating to transportation, hydrology, and GIS, but also solicit input from the natural, physical, and social sciences, as well as the arts and humanities. An obvious platform for such an effort is the coordination and creation of a GENet encyclopedia.

APPENDIX A:

ANALYSIS USE CASES

A.1

HYDROLOGY PROBLEM PARAMETERS

Objective

What is the problem objective?

The purpose is to derive a stream network from an elevation surface.

What are the problem constraints?

A flow accumulation threshold is required. A cell's neighborhood is constrained, often to the eight adjoining or four cardinal neighbors.

Data

How is the network represented?

Input data form a rectangular grid. No network is defined as input data.

Does the data support uncertain, fuzzy, or missing data?

Such data must be reconciled before performing the operation.

Algorithm

Are operational rules deterministic or stochastic?

The process is deterministic.

Does the algorithm require iteration? If so, what is its nature (finite, continuous, dynamic feedback)?

Iteration occurs as each cell is individually processed.

What dimensions does the algorithm consider?

The operation requires two-dimensional space.

Is the environment static or dynamic?

The environment is static. Changing input elevation would require a recalculation of the full operation.

Does the algorithm use exact or approximate methods?

The algorithm uses exact methods, though the results vary with data granularity.

How does the algorithm handle uncertain or missing data?

The basic algorithm does not handle uncertain or missing data. Generally, input data would be process as to remove such data. An advanced iteration of the algorithm could be modified to handle such data by polling values of neighboring cells.

Output

What is the nature of the solution? For instance, is it vector, raster, numeric, descriptive?

The output is a raster with cells denoting their status as a member of the stream network (e.g. 0 or 1).

Is the solution a subset of existing data or newly derived?

The output stream network is a newly derived product of the operation.

Process

Is scale a consideration? For instance, does the answer change with varying scale?

Varying cell size would change the output.

Are the results repeatable?

Any single source dataset would yield identical results.

Is the process reversible without data loss?

In general, the derived stream would be unable to reproduce its contributing elevation surface.

How are distance units or length addressed?

Distance and scale are implicit with cell spacing. The derived stream network does not have explicit length, but it may be calculated by querying its component cells.

A.2

TRANSPORTATION CASE PARAMETERS

Objective

What is the problem objective?

> The objective is defined as a function to minimize the path between two points on a network.

What are the problem constraints?

> The constraints of the problem are the topology and distance attributes of the network, as well as, the starting and destination points. The objective function and constraints comprise all data for the problem.

Data

How is the network represented?

> The network is represented as an Adjacency List. The data are most closely associated with a relational, vector-based space.

Does the data support uncertain, fuzzy, or missing data?

> For this implementation of the problem, uncertain, fuzzy, or missing data are not considered.

Algorithm

Are operational rules deterministic or stochastic?

> In this implementation, the operational rules are deterministic. Similar problems commonly use heuristics, including stochastic methods, to reduce computational burden.

Does the algorithm require iteration? If so, what is its nature (finite, continuous, dynamic feedback)?

> The problem requires several loops, including one to track nodes, one to track adjacent nodes and path attributes, and another to track and compare potential paths.

What dimensions does the algorithm consider?

> In this implementation, the location of nodes is not considered. The shortest path calculation uses the network's connectivity and attributes. For the problem, distance is the attribute, but a non-spatial attribute, like cost, could perform a similar function.

Is the environment static or dynamic?

> This standard implementation of Dijkstra's Algorithm assumes a static network.

Does the algorithm use exact or approximate methods?

The algorithm uses exact methods.

How does the algorithm handle uncertain or missing data?

The algorithm does not consider uncertain or missing data.

Output

What is the nature of the solution? For instance, is it vector, raster, numeric, descriptive?

The solution is the calculated length and point array of the shortest distance path.

Is the solution a subset of existing data or something newly derived?

The solution is a subset of the input network.

Process

Is scale a consideration? For instance, does the answer change with varying scale?

Scale is not considered. The answer would only change if the network topology and distance attributes were modified.

Are the results repeatable?

Dijkstra's Algorithm will yield the same result with every run.

Is the process reversible without data loss?

If all data were retained from the entire algorithm run, the full input dataset could be reconstructed. In general though, the shortest path could not be queried to reconstruct the full network.

How are distance units or length addressed?

Distance units and length are stored explicitly as arc attributes.

A.3

GIS PROBLEM PARAMETERS

Objective

What is the problem objective?

> The purpose is to identify the arcs bounding a chosen point.

What are the problem constraints?

> The problem requires all data to be on the same two-dimensional plane. A feasible solution must exist in order for the algorithm to operate.

Data

How is the network represented?

> Network data are represented as vector polylines in two-dimensional space.

Does the data support uncertain, fuzzy, or missing data?

> The operation requires full, precise data to obtain a solution.

Algorithm

Are operational rules deterministic or stochastic?

> The operation chooses an arbitrary arc to intersect part of the bounding network. The nature of that arc is arbitrary. Also, the algorithm uses an arbitrary direction, counterclockwise, to perform its work.

Does the algorithm require iteration? If so, what is its nature (finite, continuous, dynamic feedback)?

> The operation iterates along bounding arcs continues until returning to its origin.

What dimensions does the algorithm consider?

> The algorithm operates in two-dimensional space.

Is the environment static or dynamic?

> The operation is intended for a static environment, but may tolerate a dynamic network if the changes are minimal.

Does the algorithm use exact or approximate methods?

The operation requires the creation of an arbitrary arc and a defined direction. Otherwise, the operation uses exact methods at every evaluation or decision point.

How does the algorithm handle uncertain or missing data?

The algorithm does not address uncertain or missing data.

Output

What is the nature of the solution? For instance, is it vector, raster, numeric, descriptive?

The output is a collection of vector arcs.

Is the solution a subset of existing data or something newly derived?

The solution is a subset of the existing network.

Process

Is scale a consideration? For instance, does the answer change with varying scale?

The user first selects a point in order to identify its bounding arcs. The point will not vary with changes in scale. Typically, but not always, the bounding arcs would vary with changing scale.

Are the results repeatable?

Given the same network and same origin point, the solution set will be the same with every run.

Is the process reversible without data loss?

Unless all data were retained through the process, the operation would not be reversible. That is, given the arcs bounding a region, the user's exact point could not be identified. However, the possible locations for the point are bounded by the arcs.

How are distance units or length addressed?

The operation relies on the network's connectivity, and distance is not addressed.

A.4

SIMULATION PROBLEM PARAMETERS

Objective

What is the problem objective?

The operation objective is to simulate urban development around a road network.

What are the problem constraints?

The operation occurs over iterative time steps. For each time step, cells that are urbanized remain urbanized. Cells that do not neighbor at least two other urban cells remain undeveloped.

Data

How is the network represented?

The network is represented as a rectangular raster.

Does the data support uncertain, fuzzy, or missing data?

The operation ignores missing data. Advanced implementations of the case could allow more nuanced interpretation of cell values, like partial urbanization or probability of development.

Algorithm

Are operational rules deterministic or stochastic?

The operation is deterministic. However, more realistic models could add random characteristics.

Does the algorithm require iteration? If so, what is its nature (finite, continuous, dynamic feedback)?

The operation iterates on two levels. First, each cell is evaluated as to whether it meets the condition to be reclassified. Second, the process is repeated for each time step. Also, the output of each time step creates the input state for the next time step.

What dimensions does the algorithm consider?

The operation is two-dimensional, with the addition of time. The operation could be made three-dimensional without significant alteration of the process.

Is the environment static or dynamic?

The algorithm is dynamic. Each iteration creates a new state for evaluation.

Does the algorithm use exact or approximate methods?

The algorithm uses an exact comparison method. However, approximate or random methods could be incorporated into the reclassification scheme.

How does the algorithm handle uncertain or missing data?

The operation does not consider uncertain or missing data. Since the reclassification scheme is composed of a small set of rules, these could be extended to handle uncertain, fuzzy, and missing data cases.

Output

What is the nature of the solution? For instance, is it vector, raster, numeric, descriptive?

The output is a raster dataset of urbanized, with a cell value of one, and undeveloped areas with a cell value of zero.

Is the solution a subset of existing data or something newly derived?

The input data raster creates the bounds of the study area, but the output raster cell values are newly derived. The output of this problem is not a network, but represents areas affected by a nearby network.

Process

Is scale a consideration? For instance, does the answer change with varying scale?

Changing the scale would change the geometry of the input dataset and thus the problem output.

Are the results repeatable?

For the problem as defined, the problem would yield identical results for each run.

Is the process reversible without data loss?

Generally, the operation cannot be reversed. It would be possible to reverse the process if each urbanized cell retained a list of cells that led to its creation.

How are distance units or length addressed?

Distance is implicit in the spacing of the raster cells and via the definition of a cell's neighbors.

APPENDIX B:
UPPER GEYSER BASIN PREDICTION AND ERUPTION TIMES

Geyser	Prediction window start	Prediction window end	Actual eruption time
Old Faithful (1)	8:21	8:41	8:30
Old Faithful (2)	9:50	10:10	10:04
Daisy (1)	9:30	10:30	10:05
Old Faithful (3)	11:24	11:44	11:32
Grand (1)	10:30	14:30	11:33
Riverside (1)	11:30	12:30	12:20
Daisy (2)	12:05	13:05	12:32
Old Faithful (4)	12:53	13:13	12:59
Old Faithful (5)	14:20	14:40	14:25
Daisy (3)	14:30	15:30	14:55
Old Faithful (6)	15:44	16:04	16:08
Daisy (4)	17:00	18:00	17:25
Old Faithful (7)	17:29	17:49	17:39
Grand (2)	17:00	21:00	17:46
Castle (1)	17:00	19:00	17:57
Riverside (2)	17:40	18:40	18:28
Old Faithful (8)	18:59	19:19	19:16
Daisy (5)	19:30	20:30	20:02

REFERENCES

Agnarsson, G. and R. Greenlaw (2007). Graph theory : modeling, applications, and algorithms. Upper Saddle River, N.J., Pearson/Prentice Hall.

Ahuja, R. K., T. L. Magnanti, et al. (1993). Network flows : theory, algorithms, and applications. Englewood Cliffs, N.J., Prentice Hall.

Albert, R., H. Jeong, et al. (1999). "The Diameter of the WWW." Nature **401**: 130-131.

Albrecht, J. (1997). Universal analytical GIS operations: a task oriented systemization of data structure-independent GIS functionality. Geographic Information Research: Transatlantic Perspectives. M. Craglia and H. Onsrud. London, Taylor and Francis: 577-591.

Alexander, N. S. (2002). An object-oriented approach to spatial data capture. Santa Barbara, California, University of California, Santa Barbara.

Ambler, S. (2003). Agile Database Techniques: Effective Strategies for the Agile Software Developer, John Wiley & Sons, Inc.

Arctur, D. and M. Zeiler (2004). Designing geodatabases : case studies in GIS data modeling. Redlands, Calif., ESRI Press.

Backstrom, L. (2011, November 21, 2011). "Anatomy of Facebook." Retrieved July 4, 2012, 2012, from http://www.facebook.com/note.php?note_id=10150388519243859.

Balram, S. and S. Dragicevic (2006). "Modeling collaborative GIS processes using soft systems theory, UML, and object oriented design." Transactions in GIS **10**(2): 199-218.

Barabasi, A.-L. (2003). Linked : how everything is connected to everything else and what it means for business, science, and everyday life. New York, Plume.

Barabasi, A. L. and R. Albert (1999). "Emergence of scaling in random networks." Science **286**(5439): 509-512.

Barrick, K. A. (2007). "Geyser decline and extinction in New Zealand: energy development impacts and implications for environmental management." Environmental Management **39**(6): 783-805.

Batty, M. (2005). Network geography: relations, interactions, scaling, and spatial processes. Re-presenting GIS. P. Fisher and D. Unwin. West Sussex: England, John Wiley and Sons**:** 149-169.

Bhat, C. R., K. G. Goulias, et al. (2012). A Household-Level Activity Pattern Generation Model for the Simulator of Activities, Greenhouse Emissions, Networks, and Travel (SimAGENT) System in Southern California. 91st Annual Transportation Research Board Meeting, Washington, D.C.

Borrie, W. T. and R. Birzell (2001). Approaches to measuring quality of wilderness experiences. USDA Forest Service Proceedings RMRS-P-20. 2001. W. A. Freimund and D. N. Cole. Fort Collins, Colorado, US Department of Agriculture, Forest Service Rocky Mountain Research Station**:** 29-38.

Brewer, C. A. (2005). Designing Better Maps: A Guide for GIS Users. Redlands, Esri Press.

Brody, M., W. Tomkiewicz, et al. (2002). "Park visitors' understandings, values and beliefs related to their experience at Midway Geyser Basin, Yellowstone National Park, USA." International Journal of Science Education **24**(11): 1119-1141.

Brush, J. E. (1953). "The Hierarchy of Central Places in Southwestern Wisconsin." Geographical Review **43**(3): 380-402.

Bryan, T. S. (2008). The geysers of Yellowstone. Boulder, University Press of Colorado.

Buttenfield, B. P. and R. B. McMaster, Eds. (1991). Map Generalization: Making Rules for Knowledge Representation. New York, Wiley.

Calvete, H. I., C. Galé, et al. (2004). "Vehicle routing problems with soft time windows: An optimization based approach." Monografías del Seminario Matemático García de Galdeano **31**: 295-304.

Castellano, J. (2006). "Modifying Light." American Scientist **94**(5): 438.

Cayley, P. (1879). "On the colouring of maps." Proceedings of the Royal Geographical Society and monthly record of geography **1**(4).

Center for Spatially Integrated Social Science (2005). "2005 Tobler's Flow Mapper." from http://csiss.ncgia.ucsb.edu/clearinghouse/FlowMapper.

Chartrand, G. (1985). The Three Houses and Three Utilities Problem: An Introduction to Planar Graphs. Introductory Graph Theory. New York, Dover**:** 191-202.

Cheramie, K. D. (2011). "The Scale of Nature: Modeling the Mississippi River." Places: Forum of Design for the Public Realm.

Chicago Area Transportation Study (1959). Final Report. Study conducted under the sponsorship of the State of Illinois, the County of Cook, and the City of Chicago in Cooperation with the U.S. Department of Commerce Bureau of Public Roads. Chicago. **1, Survey Findings:** 96-99.

Chrisman, N. R. (1987). "Design of geographic information systems based on social and cultural goals." Photogrammetric Engineering and Remote Sensing **53**(10): 1367-1370.

Christaller, W. (1933). Die zentralen Orte in Suddeutschland. Jena, Germany, Gustav Fischer.

Church, R. L. and K. L. Roberts (1983). "Generalized coverage models and public facility location." Papers of the Regional Science Association **53**(1): 117-135.

Clarke, K. C. and L. J. Gaydos (1998). "Loose-coupling a cellular automaton model and GIS: long-term urban growth prediction for San Francisco and Washington/Baltimore." International Journal of Geographic Information Systems **2**(7): 699-714.

Couclelis, H. (1992). People manipulate objects (but cultivate fields): beyond the raster vector debate in GIS. Theories and methods of spatio-temporal reasoning in geographic space. A. U. Frank and I. Campari. Berlin, Springer-Verlag. **639:** 65-77.

Couclelis, H. (1997). "From cellular automata to urban models: new principles for model development and implementation." Environment and Planning B: Planning and Design **24**(2): 165-174.

Cova, T. J. and M. F. Goodchild (2002). "Extending geographical representation to include fields of spatial objects." International Journal of Geographical Information Science **16**: 509-532.

Current, J. R., C. S. Re Velle, et al. (1985). "The maximum covering/shortest path problem: A multiobjective network design and routing formulation." European Journal of Operational Research **21**: 189-199.

Current, J. R. and D. A. Schilling (1994). "The median tour and maximal covering tour problems: Formulations and heuristics." European Journal of Operational Research **73**(1): 114-126.

Date, C. J. (1995). An Introduction to Database Systems. Reading, MA, Addison-Wesley.

Davis, G. E. (2012). A visual window on the apparently random digits of π. Republic of Mathematics. **2012**.

Davis, S. (2007). GIS for web developers : adding where to your web applications. Raleigh, NC, Pragmatic Bookshlf.

de Smith, M. J., M. F. Goodchild, et al. (2007). Geospatial Analysis: A Comprehensive Guide to Principles, Techniques and Software Tools. Leicester, United Kingdom, Troubador Publishing Ltd.

Desrochers, M., J. K. Lenstra, et al. (1988). Vehicle routing with time windows: optimization and approximation. Vehicle Routing: Methods and Studies. B. L. Golden and A. A. Assad, Elsevier Science Publishers: 65-84.

DHV Consulting Engineers (1979). Guidelines for rural centre planning. New York, United Nations Economic and Social Commission for Asia and the Pacific.

Dijkstra, E. W. (1959). "A note on two problems in connexion with graphs." Numerische Mathematik **1**: 269–271.

Dillemuth, J. A. (2009). "Navigation Tasks with Small-Display Maps: The Sum of the Parts Does Not Equal the Whole." Cartographica: The International Journal for Geographic Information and Geovisualization **44**(3): 187-200.

Dorogovtsev, S. N. and J. F. F. Mendes (2002). "Evolution of networks." Advances in Physics **51**(4).

Du, D. and P. M. Pardalos (1993). Network optimization problems : algorithms, applications, and complexity. Singapore ; River Edge, NJ, World Scientific.

Duckham, M., M. F. Goodchild, et al., Eds. (2003). Foundations of Geographic Information Science. London, Taylor & Francis.

Egenhofer, M. and D. Mark (1995). Naive geography. Spatial Information Theory: a theoretical basis for GIS. A. Frank and W. Kuhn. Berlin, Springer-Verlag**:** 1-16.

Egenhofer, M. J. (1992). "Why not SQL! ." International Journal of Geographic Information Systems **18**: 230-245.

Egenhofer, M. J. and A. U. Frank (1992). "Object-oriented modeling for GIS." URISA Journal **4**(2): 3-19.

Elwood, S. (2008). "Volunteered geographic information: key questions, concepts and methods to guide emerging research and practice." GeoJournal **72**: 133-135.

Erdős, P. and A. Rényi (1959). "On Random Graphs. I." Publicationes Mathematicae **6**: 290-297.

Esri (2006). ArcGIS Schematics: Dealing with Connectivity. ESRI White Paper. Redlands**:** 23.

Fabrikant, S. I., D. R. Montello, et al. (2004). "The Distance-Similarity Metaphor in Network-Display Spatializations." Cartography and Geographic Information Science **31**(4): 237-252.

Facebook (2012). "Key Facts." Retrieved July 4, 2012, 2012, from http://newsroom.fb.com/content/default.aspx?NewsAreaId=22.

Fisk, H. N. (1944). Geological Investigation of the Alluvial Valley of the Lower Mississippi River. M. R. C. U.S. Corps of Engineers. Vicksburg, MS, U.S. Corps of Engineers, Mississippi River Commission**:** 78.

Fonseca, F. T., M. J. Egenhofer, et al. (2002). " Using ontologies for integrated Geographic Information Systems." Transactions in GIS **6**(3): 231-257.

Ford, D. C. and P. Williams (1989). Karst Geomorphology and Hydrology. London, Unwin Hyman.

Ford, L. R. and D. R. Fulkerson (1956). "Maximal flow through a network." Canadian Journal of Mathematics **8**: 399-404.

Fowler, M. (2003). UML Distilled: A Brief Guide to the Standard Object Modeling Language, Third Edition. Boston, Massachusetts, Addison-Wesley.

Frank, A. U., Ed. (1998). Different Types of "Times" in GIS. Spatial and Temporal Reasoning in Geographic Information Systems. Spatial Information Systems. New York, Oxford University Press.

Gardner, M. (1970). The fantastic combinations of John Conway's new solitaire game "life". Scientific American. **223:** 120-123.

Garland, K. (1994). Mr. Beck's Underground Map. London, Capital Transport Publishing.

Garrison, W. L. and D. F. Marble, Eds. (1961). The structure of transportation networks. Evanston, IL, Tranportation Center at Northwestern University.

Genz, J., J. Aucan, et al. (2009). "Wave navigation in the Marshall Islands." Oceanography **22**(2): 234-245.

Gibson, J. J. (1977). The concept of affordances. Perceiving, acting, and knowing: toward an ecological psychology. R. Shaw and J. Bransford. Minnesota, Lawrence Erlbaum**:** 67-82.

Gilbert, E. N. (1959). "Random graphs." The Annals of Mathematical Statistics **30**(4).

Giuşcă, B. (2005). The problem of the seven bridges of Koenigsberg, Wikimedia Foundation.

Glennon, A. (2010). "Creating and Validating Object-Oriented Geographic Data Models: Modeling Flow within GIS." Transactions in GIS **14**(1): 23-42.

Glennon, A. and C. Groves (2002). "An examination of perennial stream drainage patterns within the Mammoth Cave watershed." Journal of Cave and Karst Studies **64**: 82-91.

Glennon, J. A. (2001). Application of Morphometric Relationships to Active Flow Networks within the Mammoth Cave Watershed Bowling Green, Kentucky, Western Kentucky University.

Glennon, J. A. (2011). "Geyser Notebook." from http://bit.ly/Geysers.

Glennon, J. A. and R. M. Pfaff (2003). "The Extraordinary Thermal Activity of El Tatio Geyser Field, Chile." Geyser Observation and Study Association Transactions **8**: 31-78.

Glennon, R. M. and J. A. Glennon (2007). "Making better maps with ArcGIS 9.2." ArcUser (July-September 2007): 59-62.

Goodchild, M. F. (1992). "Geographical information science." International Journal of Geographical Information Systems **6**(1): 31-45.

Goodchild, M. F. (1992). Spatial Analysis. Geography's inner worlds : pervasive themes in contemporary American geography. R. Abler and M. G. Marcus. New Brunswick, N.J., Rutgers University Press.

Goodchild, M. F. (2012). Geographic Intelligence. University of Redlands/Esri Colloquium, Redlands, California.

Goodchild, M. F. and A. Glennon (2008). Representation and computation of geographic dynamics. Understanding Dynamics of Geographic Domains. K. S. Hornsby and M. Yuan. Boca Raton, FL, CRC Press: 13-29.

Goodchild, M. F., R. P. Haining, et al. (1992). "Integrating GIS and Spatial Data Analysis: Problems and Possibilities." International Journal of Geographical Information Systems **6**(5): 407-424.

Goodchild, M. F. and J. Proctor (1997). "Scale in a digital geographic world." Geographical and Environmental Modeling **1**: 5-23.

Goodchild, M. F., M. Yuan, et al. (2007). "Towards a general theory of geographic representation in GIS." International Journal of Geographical Information Science **21**(3): 239-260.

Google (2012). "Google Maps Directions." Retrieved June 2, 2012, from http://maps.google.com.

Grenon, P. and B. Smith (2004). "SNAP and SPAN: Towards Dynamic Spatial Ontology." Spatial Cognition and Computation **4**(1): 69-104.

Guo, Z. (2011). "Mind the Map! The Impact of Transit Maps on Path Choice in Public Transit." Transportation Research Part A: Policy and Practice **45**(7): 625-639.

Hägerstrand, T. (1967). Innovation diffusion as a spatial process. Chicago, University of Chicago Press.

Hägerstrand, T. (1970). "What about people in regional science?" Papers of the Regional Science Association **24**: 7-21.

Haggett, P. and R. J. Chorley (1970). Network Analysis in Geography. New York, St. Martin's Press.

Harvey, F., M.-P. Kwan, et al. (2005). "Introduction: Critical GIS." Cartographica: The International Journal for Geographic Information and Geovisualization **40**(4): 1-3.

Herold, M., Goldstein, Noah C., Clarke, Keith C. (2003). "The spatiotemporal form of urban growth: measurement, analysis and modeling." Remote Sensing of Environment **86**(3): 286-302.

Hirtle, S. and J. Jonides (1985). "Evidence of Hierarchies in Cognitive Maps." Memory and Cognition **13**(3): 208-217.

Horton, R. E. (1945). "Erosional development of streams and their drainage basins: hydrophysical approach to quantitative geomorphology." Bulletin of the Geological Society of America **56**: 275-370.

Howard, A. D. (1971). "Quantification measures of cave patterns." Caverns and Karst **13**: 1-7.

Howarth, J. (2008). Landscape and Purpose: Modeling the Functional and Spatial Organization of the Land. Department of Geography. Santa Barbara, University of California, Santa Barbara. **Doctor of Philosophy**.

Ince, D. C., L. Hatton, et al. (2012). "The case for open computer programs." Nature **482**: 485-488.

Inductiveload (2007, February 16, 2007). "Four Colour Planar Graph." Retrieved March 10, 2010, from http://en.wikipedia.org/wiki/File:Four_Colour_Planar_Graph.svg.

Itami, R. (1994). "Simulating spatial dynamics: cellular automata theory." Landscape and Urban Planning **30**(1-2): 27-47.

Jacobson, I. (1993). Object-oriented software engineering : a use case driven approach. Wokingham, England and Reading, Massachusetts, Addison-Wesley Publishing Company

Jordan, R., M. Raubal, et al. (1998). An affordance-based model of place in GIS. Eighth International Symposium on Spatial Data Handling, Vancouver, Canada.

Kansky, K. J. (1963). Structure of transportation networks: relationships between network geometry and regional characteristics. Chicago,, University of Chicago.**:** x, 155 p.

Karplus, D. (2004). Trail Survey of Yellowstone National Park, Wyoming, Montana, Idaho. Yellowstone National Park, Spatial Analysis Center, National Park Service.

Kirchner, J. W. (1993). "Statistical inevitability of Horton's laws and the apparent randomness of stream channel networks." Geology **21**(7): 591-594.

König, D. (1936). Theorie der endlichen und unendlichen Graphen. Leipzig, Akad. Verlagsgesell.

Kwan, M.-P. (2000). "Interactive geovisualization of activity-travel patterns using three-dimensional geographical information systems: a methodological exploration with a large data set." Transportation Research Part C **8**: 185-203.

Larman, C. (1998). Applying UML and Patterns: An Introduction to Object-Oriented Analysis and Design. Upper Saddle River, New Jersey, Prentice Hall.

Laughlin, R. B. (2005). A different universe : reinventing physics from the bottom down. New York, Basic Books.

Leopold, L. B. and W. B. Langbein (1962). The concept of entropy in landscape evolution. Interior. Washington, D.C., USGPO.

Lohfink, A., T. Carnduff, et al. (2007). An object-oriented approach to the representation of spatiotemporal geographic features. Proceedings of the 15th International Symposium on Advances in Geographic Information Systems Seattle, Washington, ACM.

Longley, P., M. F. Goodchild, et al. (2005). Geographical information systems and science. Chichester ; Hoboken, NJ, Wiley.

Lowe, E. J. (1998). The Possibility of Metaphysics: Substance, Identity, and Time. Oxford, UK, Oxford University Press.

Maidment, D. R. (2002). Arc hydro : GIS for water resources. Redlands, Calif., ESRI Press.

Maidment, D. R. (2003). Modeling and GIS Workshop. Redlands, California.

Mandelbrot, B. B. (1967). "How long is the coast of Britain? Statistical self-similarity and fractional dimension." Science **156**: 636-638.

Mark, D. M. (2003). Geographic information science: defining the field. Foundations of Geographic Information Science. M. Duckham, M. F. Goodchild and M. F. Worboys. New York, Taylor and Francis**:** 1-18.

Martin, C. (2006). Abstract graph corresponding to bridges of Koenigsberg, Wikimedia Foundation.

Miller, H. (1991). "Modeling accessibility using space-time prism concepts within geographical information systems." International Journal of Geographical Information Systems **5**: 287-301.

Miller, H. J. and S. A. Bridwell (2009). "A field-based theory for time geography." Annals of the Association of American Geographers **99**: 44-75.

Miller, H. J. and S.-L. Shaw (2001). Geographic information systems for transportation : principles and applications.

Minard, C. J. (1869). Carte Figurative des pertes successives en hommes de l'Armée Française dans la campagne de Russie 1812-1813. Paris, Réginer et Dourdet.

Mitchell, J. S. B. (2000). Geometric shortest paths and network optimization. Handbook of Computational Geometry. J. Urrutia, Elsevier**:** 633-701.

Monmonier, M. (1996). How to lie with maps. Chicago, University of Chicago Press.

Montello, D. R. (2001). Scale, in geography. International Encyclopedia of the Social & Behavioral Sciences N. J. Smelser and P. B. Baltes. Oxford, Pergamon Press**:** 13501-13504.

Moore, E. F. (1959). The shortest path through a maze. International Symposium on the Theory of Switching, Cambridge, Massachusetts, Harvard University Press.

Naish, J. M. (1964). "Combination of Information in Superimposed Visual Fields." Nature **202**(4933): 641-646.

National Center for Airborne Laser Mapping (2008). EarthScope Intermountain Seismic Belt LiDAR Project. Yellowstone National Park, UNAVCO for EarthScope

Newton, I. (1687). Philosophiæ naturalis principia mathematica. Londini,, Jussu Societatis Regiæ ac Typis Josephi Streater. Prostat apud plures Bibliopolas.

Nichols, T. and M. C. Little (2011). Aircraft Systems and Methods for Displaying a Touchdown Point. United States Patent and Trademark Office. United States, Honeywell.

Nystuen, J. D. and M. F. Dacey (1961). "A graph theory interpretation of nodel regions." Papers of the Regional Science Association **7**: 29-42.

Onsted, J. A. and K. C. Clarke (2011). "Forecasting enrollment in differential assessment programs using cellular automata." Environment and Planning B: Planning and Design **38**(5): 829-849.

Palmer, A. N. (1991). "Origin and morphology of limestone caves." Geological Society of America Bulletin **103**(1): 1-21.

Patterson, M. E., A. E. Watson, et al. (1998). "An Hermeneutic Approach to Studying the Nature of Wilderness Experiences." Journal of Leisure Research **30**(4): 423-452.

Peuquet, D., B. Smith, et al. (1998). The Ontology of Fields: Report of a Specialist Meeting Held under the Auspices of the Varenius Project, NCGIA**:** 42.

Peuquet, D. J. (2002). Representations of space and time. New York, Guilford Press.

Pickles, J., Ed. (1995). Ground Truth: The Social Implications of Geographic Information Systems. New York, Guilford Press.

Price, D. J. (1965). "Networks of Scientific Papers." Science **149**: 510-515.

Quinlan, J. F. and J. A. Ray (1989). Groundwater basins in the Mammoth Cave region, Kentucky Friends of the Karst Occasional Publication. **1**.

Rademaker, L. G. (2008). Interpretive technology in parks a study of visitor experience with portable multimedia devices, University of Montana. **M.S.:** 102.

Ralyté J, Deneckère R, et al. (2003). Towards a Generic Model for Situational Method Engineering. Heidelberg, Springer Berlin.

Raper, J. and D. Livingstone (1995). "Development of a geomorphological spatial model using object-oriented design." International Journal of Geographical Information Systems **9**(4): 359-383.

Raubal, M., H. J. Miller, et al. (2004). "User-centered time geography for location-based services " Geografiska Annaler-B **86**: 245-265.

Ravenstein, E. (1885). "The Laws of Migration." Journal of the Statistical Society **46**: 167-235.

Ravenstein, E. (1889). "The Laws of Migration: Second Paper." Journal of the Statistical Society **52**: 241-305.

Ray, J. A. and J. C. Currens (1998a). Mapped Karst Ground-water Basins in the Beaver Dam 30 x 60 Minute Quadrangle. Lexington, KY, Kentucky Geological Survey.

Ray, J. A. and J. C. Currens (1998b). Mapped Karst Ground-water Basins in the Campbellsville 30 x 60 Minute Quadrangle. Lexington, KY, Kentucky Geological Survey.

Rinehart, J. S. (1980). Geysers and geothermal energy. New York, Springer-Verlag.

Rumbaugh, J., M. Blaha, et al. (1991). Object-Oriented Modeling and Design. Englewood Cliffs, New Jersey, Prentice Hall.

Schabenberger, O. and C. A. Gotway (2005). Statistical methods for spatial data analysis. Boca Raton, Chapman & Hall/CRC.

Shaw, S.-L. (2006). "What about "time" in transportation geography?" Journal of Transport Geography **14**(3): 237-240.

Sheppard, E. (2005). "Knowledge production through critical GIS: Genealogy and prospects." Cartographica: The International Journal for Geographic Information and Geovisualization **40**(4): 5-21.

Shreve, R. L. (1967). "Infinite topologically random channel networks." Journal of Geology **75**: 178-186.

Sider, T. (2001). Four-Dimensionalism: An Ontology of Persistence and Time. Oxford, United Kingdom, Clarendon Press.

Smith, B. and D. M. Mark (2001). "Geographical categories: an ontological investigation." International Journal of Geographical Information Science **15**(7): 591-612.

Song, C., Z. Qu, et al. (2010). "Limits of Predictability in Human Mobility." Science **327**(5968): 1018-1021.

Taaffe, E. J. (1956). "Air transportation and the United States Urban Distribution." The Geographical Review **46**: 219-239.

Taaffe, E. J. (1958). "A Map Analysis of United State Airline Competition--Part 2. Competition and Growth." The Journal of Air Law and Commerce **Autumn**: 402-429.

Taaffe, E. J. (1958). "A Map Analysis of United States Airline Competition--Part 1. The Development of Competition." The Journal of Air Law and Commerce(Spring): 121-147.

Taaffe, E. J. (1962). "The Urban Hierarchy: An Air Passenger Definition." Economic Geography **38**: 1-14.

Taaffe, E. J., R. Morrill, et al. (1963). "Transport Expansion in Underdeveloped Countries: A Comparative Analysis." The Geographical Review **53**: 503-529.

Taylor, P. J. (1990). "GKS." Political Geography Quarterly **9**(3): 211-212.

Thornthwaite, W. C. (1934). Internal Migration in the United States, Study of Population Redistribution, Philadelphia, University of Pennsylvania Press.

Tobler, W. (2003). "Movement Mapping." from http://csiss.ncgia.ucsb.edu/clearinghouse/FlowMapper/MovementMapping.pdf.

Tobler, W. R. (1970). "COMPUTER MOVIE SIMULATING URBAN GROWTH IN DETROIT REGION." Economic Geography **46**(2): 234-240.

Tobler, W. R. (1987). "EXPERIMENTS IN MIGRATION MAPPING BY COMPUTER." American Cartographer **14**(2): 155-163.

Transport for London (2012). London Underground: Colour large print Tube map. London, United Kingdom, Transport for London.

Tufte, E. (2001). The Visual Display of Geographic Information, Second Edition. Cheshire, Connecticut, Graphic Press.

Tutte, W. T. (1984). Graph theory. Menlo Park, Calif., Addison-Wesley Pub. Co., Advanced Book Program.

Twitter (2012). "What is Twitter?". Retrieved July 4, 2012, 2012, from https://business.twitter.com/en/basics/what-is-twitter/.

U.S. Army Corps of Engineers (2012). "Aerial view of the Mississippi Basin Physical Model." Retrieved 10/7, 2012, from http://chl.erdc.usace.army.mil/chl.aspx?p=FullSizeImage&a=ImageFile_ID;2094.

U.S. Census Bureau (2003). State to State Migration Flows: 1995-2000, CENSR-8. U.S. Department of Commerce. Washington, D.C.

Van Rossum, G. (1998). Python Patterns - Implementing Graphs. Python Documentation, Python Essays, Python Software Foundation. **2012**.

von Neumann, J. (1966). Theory of self-producing automata. A. W. Burks. Urbana, University of Illinois Press.

Watts, D. J. (2003). Six Degrees: the Science of a Connected Age. New York, NY, Norton.

Wegner, P. (1988). The object-oriented classification paradigm. research directions. Object-Oriented Programming. B. Shriver and P. Wegner. Cambridge, Massachusetts, MIT Press.

Weisman, M. (2003). The Object-Oriented Thought Process, Second Edition. Indianapolis, Indiana, Sams.

Weiss, G. H. (1994). Aspects and applications of the random walk. Amsterdam [The Netherlands] ; New York, North-Holland.

White, R. and G. Engelen (1993). "Cellular automata and fractal urban form: a cellular modelling approach to the evolution of urban land-use patterns." Environment and Planning A **25**(8): 1175-1199

White, W. B. (2003). Speleogenesis and Evolution of Karst Aquifers. Karst Modeling: Special Publication 5. A. N. Palmer, M. V. Palmer and I. D. Sasowsky. Charles Town, West Virginia, The Karst Waters Institute.

White, W. B. and E. L. White (1989). Karst Hydrology: Concepts from the Mammoth Cave Area. New York, Van Nostrand Reinhold.

Whitledge, V. M. and R. Taylor (2008). "The Behavior of the Grand Group During the Summers of 2005 and 2006." Geyser Observation and Study Association Transactions **10**.

Whittlesey, L. H. (1995). Death in Yellowstone : accidents and foolhardiness in the first national park. Boulder, Colo., Roberts Rinehart Publishers.

Wolfram, S. (1984). "Cellular automata as models of complexity." Nature **311**: 419-424.

Worboys, M. and M. Duckham (2004). GIS : a computing perspective. Boca Raton, Fla., CRC Press.

Worboys, M. F., H. M. Hearnshaw, et al. (1990). "Object-oriented data modeling for spatial databases." International Journal of Geographical Information Systems **4**(4): 369-383.

Yuan, M. (2001). "Representing Complex Geographic Phenomena in GIS." Cartography and Geographic Information Science **28**: 83-96.

Yuan, M. (2009). Challenges and critical issues for temporal GIS research and technologies. Handbook of Research on Geoinformatics. H. Karimi. Hershey, PA.

Zeiler, M. (1999). Modeling Our World: The ESRI Guide to Geodatabase Design. Redlands, California, ESRI Press.

Zhan, F. B. (1998, December 23, 1998). "Representing networks." NCGIA Core Curriculum in GIScience. Retrieved May 28, 2012, from http://www.ncgia.ucsb.edu/giscc/units/u064/u064.html.

Zhan, G. (2011). "Mind the map! The impact of transit maps on path choice in public transit." Transportation Research Part A: Policy and Practice **45**(7): 625-639.

Zhang, J. and M. F. Goodchild (2002). Uncertainty in Geographical Information. London, Taylor & Francis.