

DYNAMIC VISUALIZATION OF GEOGRAPHIC NETWORKS USING  
SURFACE DEFORMATIONS

By  
BAŐAK ALPER

Submitted to the Graduate School of Engineering and Natural Sciences  
in partial fulfillment of  
the requirements for the degree of  
Master of Science

SABANCI UNIVERSITY  
Spring 2006

DYNAMIC VISUALIZATION OF GEOGRAPHIC NETWORKS USING  
SURFACE DEFORMATIONS

APPROVED BY:

Assist. Prof Selim Balcısoy  
(Dissertation Advisor)

\_\_\_\_\_

Assoc. Prof Berrin Yanıkođlu

\_\_\_\_\_

Assist. Prof Hüsnu Yenigün

\_\_\_\_\_

Assist. Prof Gürdal Ertek

\_\_\_\_\_

Instructor Murat Germen

\_\_\_\_\_

DATE OF APPROVAL:

26.07.2006

© Bařak Alper 2006  
ALL RIGHTS RESERVED

# DYNAMIC VISUALIZATION OF GEOGRAPHIC NETWORKS USING SURFACE DEFORMATIONS

Başak Alper

EECS, M.Sc. Thesis, 2006

Thesis Supervisor: Assist. Prof. Selim Balcısoy

Keywords: Information Visualization, Network Visualization, Information Interfaces,  
Virtual Environments, Surface Deformations

## **ABSTRACT**

Visualization techniques for geographic data show vast variations which are well-developed over centuries. While most of the known techniques are sound for low dimensional data sets, few techniques exist for visualization of high dimensional data within the geographic framework. This thesis investigates visualization of temporal, high dimensional network data within the geographic context. The resulting visualization system employs network visualization techniques in conjunction with cartographic visualization methods for providing a qualitative feel for the data, while conventional methods are employed for detailed examination. In turn, the visualization facilitates comprehension of non-spatial variables with respect to the geographic context.

# COĞRAFİ AĞ VERİLERİNİN YÜZEY DEFORMASYONLARI İLE DİNAMİK GÖRSELLEŞTİRİLMESİ

Başak Alper

EECS, Yüksek Lisans Tezi, 2006

Tez Danışmanı: Yar. Doç. Selim Balcısoy

Anahtar Kelimeler: Bilgi Görselleştirme, Ağ Görselleştirme, Bilgi Arayüzleri, Sanal Ortamlar, Yüzey Deformasyonları

## ÖZET

Coğrafi bilgi görselleştirme teknikleri oldukça uzun bir süreçte geliştirilen geniş bir çeşitliliğe sahiptir. Bilinen tekniklerin çoğu az boyutlu coğrafi verilerin görselleştirme için yetkin olsalar da, çok boyutlu verilerin coğrafi bağlam içerisinde görselleştirilmesine dair oldukça az sayıda metod önerilmiştir. Bu tez, coğrafi bileşenleri olan ve zamana bağlı değişen ağ verilerinin coğrafi çerçeve içerisinde görselleştirilmesi problemini araştırmaktadır. Tezin sonucu olarak önerilen görselleştirme sistemi, verinin niteliksel özelliklerini görselleştirmek için çok boyutlu veri görselleştirme teknikleri ile harita görselleştirme tekniklerini bir arada kullanırken, istatistiksel nicel değerlerin incelenmesi için çubuk ve yay gibi klasik ağ görselleştirme tekniklerini kullanmaktadır. Ortaya çıkan görselleştirme sistemi veri içerisindeki uzaysal olmayan bileşenlerin uzaysal bileşenlere göre değişiminin algılanmasına yardımcı olmaktadır.

## ACKNOWLEDGEMENTS

I wish to express my deepest gratitude to my supervisor Selim Balcısoy for his valuable advice and guidance of this work. I am grateful to him not only for the completion of this thesis, but also for his unconditional support from the beginning.

I am greatly indebted to Selçuk Sumengen, who cooperated through out this work and assisted implementing implicit integration solution for spring embedders and the surface generation methodology.

I would like to thank all my friends and colleagues, particularly to Can Özmen, Ceren Kayalar and Ekrem Serin in the Computer Graphics Lab, for their friendship and assistance.

I would like to thank my family, particularly to my mother for her unlimited support and trust that made everything possible for me.

This research is supported by TÜBİTAK.

## TABLE OF CONTENTS

1	INTRODUCTION .....	1
1.1	Information Visualization: an Overview .....	1
1.2	An Approach for Visualizing Geographic Networks .....	2
1.3	Summary of Contributions .....	5
1.4	Thesis Outline .....	5
2	MOTIVATION AND RELATED WORK .....	7
2.1	Historical Background and Definitions .....	7
2.2	Visualization and Human Visual Perception .....	10
2.3	Qualitative Representation of Information .....	12
2.4	Interactive Environments for Visualization .....	13
2.5	Visualization of Multidimensional Data .....	16
2.5.1	MDS .....	19
2.5.2	Network Visualization .....	19
2.6	Map Visualizations .....	23
2.6.1	Area Cartograms .....	25
2.6.2	3D Virtual Environments in Geographic Data Visualizations .....	28
2.6.3	Visualization of Geographic Networks .....	29
2.7	Spring Embedders in Visualization .....	33
3	VISUALIZATION SYSTEM .....	36
3.1	The Global Context Visualization Technique .....	39
3.1.1	Input Data Structure .....	39
3.1.2	Spring-Embedder Model .....	40
3.1.2.1	Force Model .....	41
3.1.2.2	Neighboring Heuristic .....	42
3.1.2.3	Integration .....	43
3.1.3	Surface Generation .....	45
3.1.4	Data and Program Flow .....	47
3.1.5	Efficiency .....	48
3.2	Thorough Data Investigation with Analytical Tools .....	49

4	CASE STUDIES .....	51
4.1	Flight Data .....	51
4.2	Diplomatic Exchange Data .....	52
5	ANALYSIS .....	54
5.1	Results .....	54
5.2	Discussion and Future Work .....	56
	REFERENCES .....	58

## LIST OF FIGURES

- Figure 1.1 Snapshots from the *global context visualization* mode. First figure is taken from the diplomatic exchange data visualization, and the second figure is taken from the US domestic air flights data visualization. Details about these data sets can be found in Chapter 4. .... 3
- Figure 1.2 Figures show the analytical tools that appear on demand of the user. The left figure shows loads of all nodes as height bars and all the links between them as arcs. The right figure only shows connections from the user selected node. Statistical data about selected node appears on the bottom left corner of the display. .... 4
- Figure 2.1 Minard’s visualization of Napoleon’s Russia campaign. The graph depicts route of the army and its change in size over time. It differentiates the retreat with black color. Locations that the army passes through and dates are labeled on the route. The change in temperature is demonstrated with a graph and related with the route using scan lines. .... 8
- Figure 2.2 Accuracy ranking of perceptual tasks in visualization of quantitative information. Cleveland and MacGill empirically verified the basics of this ranking. .... 11
- Figure 2.3 Ranking of perceptual tasks for quantitative, ordinal and nominal data values. MacKinlay developed this ranking based on existing psychophysical results and various analysis of perceptual tasks, but it has not been verified empirically. Tasks in gray boxes are not relevant for these data types. .... 12
- Figure 2.4 MarketMap visualization of SmartMoney.com. The snapshot gives an overview of the rises and falls of stock prices for past 26 weeks’ activity. .... 13
- Figure 2.5 Film Finder dynamic query tool of Ahlberg and Shneiderman. Two axes of the display show years and popularity of the movies. Their genres are indicated by colors. Queries are refined by adjusting sliders on the right of the display. .... 14

Figure 2.6 Selective dynamic interaction modes proposed by Chuah et al. The bars colored in green are a selected subset. They can be scaled or translated while their original positions and scales are shown as white shell objects.....	15
Figure 2.7 Multidimensional data visualization using parallel coordinates method exemplified in Yang et al. First figure displays a 4 dimensional data while second figure displays 42 dimensional data.....	17
Figure 2.8 Dynamic grand tour software developed by Yang [60]. Yang limited number of projections to follow specified clusters and principal components in the data. .	18
Figure 2.9 One focal and bi-focal fisheye views implemented by Carpendale et al. [45] .....	21
Figure 2.10 Lamping and Rao’s hyperbolic geometry network visualization. The hyperbolic space is projected on a 2D plane.....	22
Figure 2.11 Tamara Munzner’s network visualization based on 3D hyperbolic geometry. As clusters of nodes move away from the focal they appear smaller and with less detail. ....	22
Figure 2.12 Dot map showing birth places of the 3005 Ming poets in China between 1368-1644. See Chen-Cheng Siang, “A Historical and Cultural Atlas of China”, map 62. Reproduced from Tufte.....	24
Figure 2.13 Grid-square map showing population density of Japan. The map is divided into equal sized small rectangles and than each rectangle is colored to represent population of that area. See Hidenori Kimura, “Grid Square Statistics for the Distribution and Mobility of Population in Japan”, Statistics Bureau, Tokyo. Reproduced from Tufte.....	25
Figure 2.14 Gillihan’s cartogram made with plasticine and a rolling pin. The cartogram shows distribution of Smallpox incidences in California between 1915 and 1924. Regions are scaled proportional to the number of disease incidences.....	26
Figure 2.15 Gastner’s density equalizing map showing 2000 election results in US. ...	27
Figure 2.16 Lokuge et al.’s visualization of tourist attraction sites in Boston. Popularity of sites is shown with height bars over the 3D map. ....	28
Figure 2.17 NFSNET T1 internet traffic backbone visualization of NCSA. ....	30
Figure 2.18 Munzner’s visualization of MBone, multicast backbone of internet.....	30
Figure 2.19 SeeNet network visualization system. First figure displays connections emanating from a single node with lines. Half lines code the overload by direction. Second figure shows the matrix representation of the same network data. The	

vertical or horizontal ordering of nodes can be modified to ease readability of patterns. In third snapshot only half-lines between nodes are drawn to eliminate clutter problem. Fourth figure shows the aggregate load of each node encoded by dimensions and color of each rectangle.....	31
Figure 2.20 SeeNet3D network visualization system. First figure shows color coded network connections over 3D map. Second figure shows a helix representation of connections from a single node. Third figure shows all connections on a planar map. Final figure is another drill down network view showing connections from a single node and their load is coded with the size of spheres representing nodes. ..	32
Figure 2.21 A graph layout optimization with Kamada-Kawai's force-directed placement algorithm. The algorithm effectively eliminates long edges which would clutter the display.....	34
Figure 3.1 Geographic distribution of airports in Figure 3.2 .....	37
Figure 3.2 Air traffic data among 15 airports is visualized using PERMAP, an MDS visualization software. Geographic distribution of nodes in Figure 3.1 is significantly lost.....	37
Figure 3.3 Visualization system diagram.....	38
Figure 3.4 Global context visualization for two different data sets. The map surface is deformed to fit the underlying graph layout.....	39
Figure 3.5 Simplified force diagram for a node with two neighbors and a single data relation.....	42
Figure 3.6 (a) 2D distribution of nodes (b) closest node to the current is chosen to be the first neighbor (c) second neighbor is chosen to be the node with lowest $f(g_i)$ given first neighbor (d) second neighbor is chosen to be the node with lowest $f(g_i)$ given first two neighbors (e) resulting neighbor selection.....	43
Figure 3.7 Multi resolution grid deformation: (a) Low resolution grid constructed with multiple network node attaching to a single grid node, (b) Mid resolution grid eliminates multiple attachments, (c) High resolution grid smoothly interpolates between even close nodes.....	46
Figure 3.8 Data and program flow diagram for global context visualization. Separate control flow for the display module and for the data update enable the viewers to navigate at interactive speeds.....	47
Figure 3.10 Analytical tools for thorough data analysis. First figure shows height bar animations, second figure shows arcs connecting nodes and bottom left figure	

displays both. When a node is selected as in bottom right figure, only arcs connected to that node are shown. Aggregated load of the node is shown as a height bar. ....	49
Figure 4.1 Snapshots from US domestic air flights data visualization. Image on the left is visualization of July 2001. Middle image displays smoothing of the surface on September 9 <sup>th</sup> 2001. Right image is a closer view for July 2001. ....	51
Figure 4.2 Visualization system using diplomatic exchange data is exhibited at TECHNE digital performance platform. ....	52
Figure 5.1 Same data visualized with height bars and arcs versus global context visualization. ....	55

## TABLE OF SYMBOLS

$G=(V,E)$	A graph with vertices $V$ and edges $E$ .
$O()$	Big O notation for computational complexity.

## **TABLE OF ABBREVIATIONS**

VR	Virtual Reality
VE	Virtual Environment
MDS	Multi Dimensional Scaling

# 1 INTRODUCTION

## 1.1 Information Visualization: an overview

Assessments of structure, patterns, change and dynamics within large bodies of data is the concern of information visualization. The goal is to amplify one's ability to process large amounts of data by assisting reasoning, hypothesis generation and cognition. Visualization augments cognition by exploiting human visual perception as an external aid for memory. A picture is worth ten thousand words as they say, because cognition of a visual representation is supported by a large number of perceptual inferences that are extremely easy for humans (e.g. recognizing shapes and colors) [1]. This is why a complex problem is solved faster with the aid of a diagram.

Even though data sets may contain hundreds of dimensions, only few significant features are sufficient to understand, describe and summarize them in a global context. Visual representations are mostly efficient for their ability to emphasize these characteristic features of a data set with high speed and accuracy.

With the aid of computers, visualization has become more than a static visual representation of data. Information visualization systems have become user interfaces that allow manipulating objects on a computer screen, hence exhibit dynamically changing representations of data. They are useful because they can:

- reduce search activity by grouping related information together,
- spatially index data to provide rapid access,
- enable hierarchical search at different zoom and context levels,
- represent huge data sets within small displays while allowing drilling down details on demand, which enable to read data at micro and macro levels.

Despite the variety of available methods, visualization tools have not kept pace with the diversity and increase in data volumes. In our information rich society, there is a myriad of highly complex information systems which we lack proper visualization tools. Diversity of data sets necessitates design of specialized visualization tools for each data set having a character of its own.

Being multidimensional datasets, networks often have a spatial component. Effective use of this spatial component in visualization facilitates recognition and comprehension of the network data. Political relations, monetary transactions, transportation, telecommunication and migration data sets are few examples of information that can be made more comprehensible when visualized in a geographic framework. However, high dimensional networks with geographic components call for dedicated visualization tools which are able to convey characteristics of the data in a spatial framework. Maintaining the geographic context of data in visualization is imperative for enabling semantic interpretation and understanding related human behavior. To approach this problem, this thesis proposes a novel visualization environment for visualization of geographic networks.

The proposed visualization system is developed through the study of multi dimensional data visualization and cartographic visualization methods. In particular, examining how graph drawing algorithms optimize the layout to permit intuitive reading of the data and how geographic context is conveyed with thematic map visualizations gave insight into how to design a visualization environment that augments the perception of the geography related high dimensional data.

## **1.2 An Approach for Visualizing Geographic Networks**

Visualizing spatial networks in a geographic context is a difficult task. The difficulty lies in the struggle to convey both low level details of the large data set and high level contextual information. This thesis proposes a visualization system for geographic network data sets, which tries to achieve both by combining different visualization approaches. The *global context visualization* represents time-series spatial

network data within geographic context through an animation of map deformation, thus maintains spatial framework while reflecting only the general trends and anomalies in the data. The visualization system is extended by a set of *analytical tools* which will reveal accurate statistical details of the data on demand.

In general, methods for displaying multidimensional datasets such as matrix representations, node and link displays, multidimensional scaling or graph drawing algorithms do not represent data within the geographic context.

The presented visualization technique employs a modified graph drawing algorithm based on spring-embedders which positions network nodes according to the time-series data being fed to it. Our contribution lies in the geographic constraints applied. These constraints limit variation of network nodes by favoring inherent geographic distribution of nodes. The graph optimization solution is reached using an implicit integration scheme and allows the system visualize data in real-time.

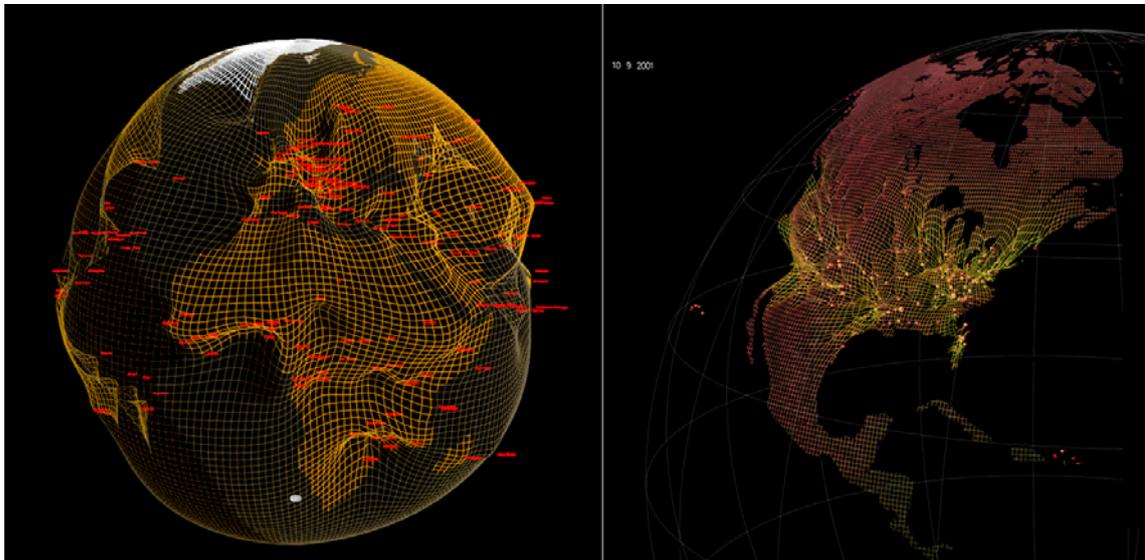


Figure 1.1 Snapshots from the *global context visualization* mode. First figure is taken from the diplomatic exchange data visualization, and the second figure is taken from the US domestic air flights data visualization. Details about these data sets can be found in Chapter 4.

To convey geographic context of the graph animation, the nodes are covered with a surface on which a map is projected. As position of network nodes change, the surface deforms and enables viewers to read the data variation as a map deformation. This representation gives a strong qualitative impression and enables viewers to summarize nature of the data. Engaging quality of the visualization emanates from its visual appeal and undemanding readability of maps.

Instead of representing the complex network data with thousands of arcs each displaying individual connections, the *global context visualization* represents cumulative relations between nodes as deformations on the map surface. Each network node moves towards the other nodes that it is more related. The viewer is able to glean dominant tendencies for each node without examining the relations separately.

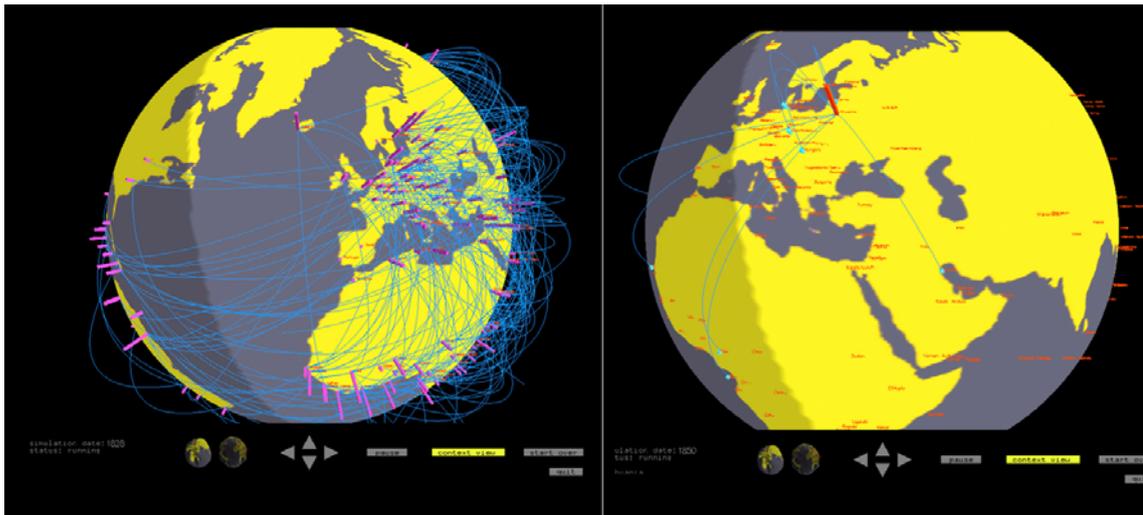


Figure 1.2 Figures show the analytical tools that appear on demand of the user. The left figure shows loads of all nodes as height bars and all the links between them as arcs. The right figure only shows connections from the user selected display. Statistical data about selected node appears on the bottom left corner of the display.

In addition to the global context visualization, the visualization system incorporates a set of analytical tools which enable to examine elements of interest in detail. Interactively responding to the user, these features expand informative quality of the visualization through direct manipulation of visualization paradigms. Users are able to view network data as height bar animations over 3D map or as arcs showing

connections. They are able to filter data through selecting nodes or range of the displayed data.

Providing a *global context view* and *analytical tools* simultaneously, our visualization is able to communicate characteristics of underlying data at micro and macro levels. We experiment the proposed method with two different data sets. First data is the domestic air flights of US among 231 airports between 1991 and 2004. Second data set is the diplomatic exchange data among 128 nations through years 1815 to 1966. Details about these data sets can be found in Section 4.

### **1.3 Summary of Contributions**

The proposed visualization system tightly integrates several visualization techniques in a novel manner. The main contributions of this research are:

- The development of a modified graph drawing algorithm that maintains geographic distribution of nodes to some degree.
- Introducing implicit integration solutions for graph optimization problems which enable to visualize time-series data in real-time.
- Combining map morphing and graph drawing methods to convey geographic context.
- Providing a highly interactive visualization environment which visualizes data at different context levels.

### **1.4 Thesis Outline**

This chapter briefly points out the usefulness of information visualization in analysis and comprehension of complex data sets. It also exposes the need for visualization tools for spatial networks that are capable of conveying data in a geographic context. Finally, it succinctly describes our approach to solve the stated problem. Subsequent chapters contribute to the thesis as the following.

*Motivation And Related Work-* The second chapter summarizes historical background of information visualization and provides definitions to clarify its scope and context. It exemplifies existing methods ordered from general to specific.

*Visualization System-* The third chapter describes the proposed method in detail and elucidates on the technical details. It also points out the differences of the proposed visualization system from the existing methods described in the previous section.

*Case Studies-* The fourth chapter demonstrates two experiments that are examples of visualizations implemented with our method.

*Analysis-* The final chapter describes successes and shortcomings of the work. These themes point to the future work and possible improvements to the initial model presented.

## 2 MOTIVATION AND RELATED WORK

### 2.1 Historical Background and Definitions

Information Visualization as a sub-field of mathematics, statistics and computer graphics has not been recognized as a self-standing research field until late 80's. However, having roots in well established fields like cartography, cognitive science and data graphics (or information design); information visualization has a vast background in literature.

Visualization as an aid for thinking has historical roots. It has been employed in maps and scientific drawings since the age of Renaissance. Mathematical advances in logarithm, calculus, conformation of probability theory and invention of Cartesian coordinates in the 17th century paved the way to the more sophisticated data graphics. Many early examples of such works can be found in Tufte's *Envisioning Information* [3].

By the end of the 18th century, William Playfair (1759-1823) had developed fundamental graphical designs that seek to replace tables of numbers with systematic visual representations. In his books *Commercial and Political Atlas* from 1786 and the *Statistical Breviary*, statistical data is demonstrated with various charts that we are familiar today. Following Playfair, various forms of charts, graphs and thematic maps are used and developed throughout 19th century. A classical example for a very effective data graphics is Joseph Minard's 1812 dated famous map of Napoleon's Russia campaign. This graph is referred by Tufte as one of the best statistical charts ever produced. The piece successfully exploits the spatial visualization and conveys other information densely which is a quality that is sought in every data graphics [3]. Besides the path of the army, its change in size and temperature are conveyed by utilizing additional graphic elements.

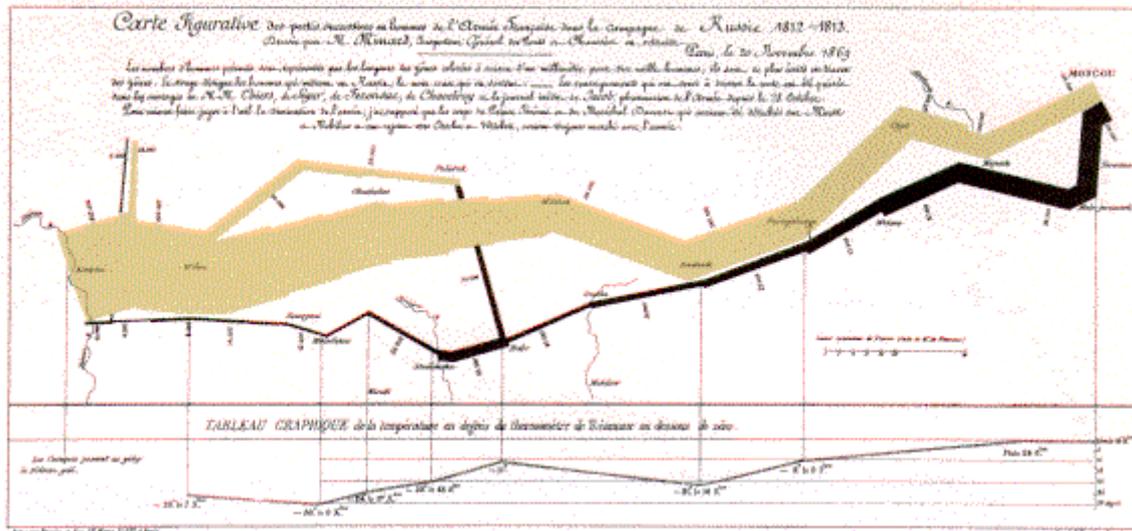


Figure 2.1 Minard’s visualization of Napoleon’s Russia campaign. The graph depicts route of the army and its change in size over time. It differentiates the retreat with black color. Locations that the army passes through and dates are labeled on the route. The change in temperature is demonstrated with a graph and related with the route using scan lines.

In 1967, Bertin prepared the ground for systematic visualization [5]. In his *Semiology of Graphics* he identified basic elements of diagrams and described a framework for their design. In 1983, Tufte published a theory of data graphics that emphasized maximization of useful information [3]. Referred in numerous works, both Bertin and Tufte have been influential figures in the shaping of information visualization as a discipline.

Charts, diagrams, graphs, tables, maps of all sorts are indispensable inventions that we use everyday and as Bertin would say we are accustomed to *use vision to think*. What draws the present attention to visualization is the evolution of computers. Within last two decades, the advances in computer graphics along with the increase of rendering power invoked computer assisted visualization as an active research area. Computers enabled processing thousands of elements for creating a single picture out of large bodies of data. The new medium allowed graphic depictions with changing focus and context, convey information with adjustable parameters at varying resolutions. In turn, these interaction modes created new methods that give insight about the nature of the data and augment cognition.

Computer aided visualization proved its usefulness first in scientific visualizations where real physical phenomena are simulated. Following fruitful works in visualization of fluid dynamics, atmospheric activity, force fields, and etc., 1987 issue of Computer Graphics journal devoted solely on Visualization in Scientific Computing. Today visualization has a broader application area including business, economics, psychology, social sciences, education and large document spaces.

Given the interwoven structure of the field it is worth to define the term visualization in its present sense. Card et al. [1] gives a concise definition as the following:

*Visualization: The use of computer-supported, interactive, visual representations of data to amplify cognition [1].*

Although it is not a necessity by its definition, visualization deals with non-abstract data. Visualization dealing with non-abstract data does not imply that abstractions are not used at any level. However, the abstractions are derived from the real physical space and what is visualized has always a physical analogue. Simulations of fluid dynamics or atmospheric activity typify this kind of visualization, yet they incorporate abstract elements like color or vectors.

By contrast, information visualization deals with abstract information with no obvious spatial mapping. Rendering the non-spatial abstract data into effective visual form is the endeavor of information visualization.

*Information Visualization: The use of computer-supported, interactive visual representations of abstract data to amplify cognition [1].*

Card et al. also provides the following hierarchic table of definitions which clarifies associations between concepts related to information visualization.

Information Design	Design of external representations to amplify cognition.
Data Graphics	Use of abstract, non-representational visual representations of data to amplify cognition.
Visualization	Use of computer-based, interactive visual representations of data to amplify cognition.
(i) Scientific Visualization	Use of interactive visual representations of scientific data, typically physically based, to amplify cognition.
(ii) Information Visualization	Use of interactive visual representations of abstract, non-physical data to amplify cognition.

Table 1 Definitions of concepts related to information visualization provided in Card et al. [1]

As a final remark, it is worth to note that most of the concepts utilized in the making of conventional data graphics are directly incorporated in information visualization, although their scope and methodology show substantial differences.

## 2.2 Visualization and Human Visual Perception

Perception and interpretation of the data is significantly influenced by its representation method. A visual representation is often the most effective way to describe, explore and summarize large sets of numbers. Yet, generating a well-designed representation of data that communicates and facilitates comprehension of statistical information is a challenging task, especially when the complexity, temporality and high dimensionality of the data sets are considered.

If the purpose of information visualization is to make the latent structure in the abstract data visible, then the question is making which characteristics visible and in what ways?

Card et al. gives an alternative definition for visualization as “adjustable mappings from data to visual form to the human perceiver” [1]. When data elements have only three variables, the mapping between the data and the image is quite straight forward. Images provide at most three dimensions, and therefore they are suitable for direct mapping of only three variables. On the other hand, information visualization often

deals with data elements with more than three variables. In such cases, other graphic elements like color, shape, density, etc. are used to characterize individual data elements. Deciding about which attribute of the data will be represented with which graphic element is a non-trivial task. Effectiveness of a graphic representation depends mostly on those decisions.

Although there is no verified theory of human perceptual capabilities that can be used to evaluate effectiveness of graphic representations, Cleveland and McGill provided an observational study which stated that people accomplish perceptual tasks associated with interpretation of graphical representations with different accuracy levels [6].

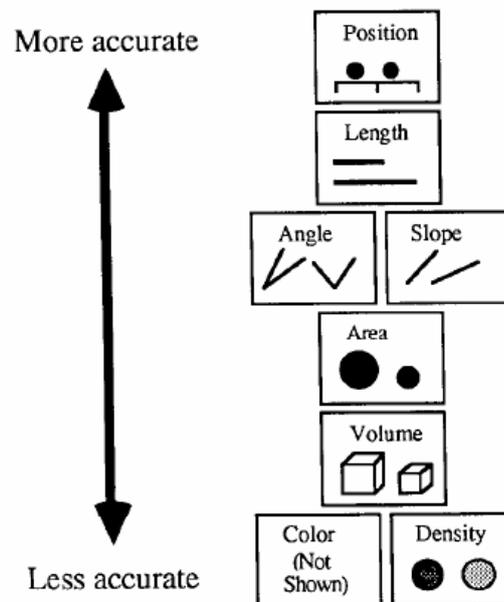


Figure 2.2 Accuracy ranking of perceptual tasks in visualization of quantitative information. Cleveland and MacGill empirically verified the basics of this ranking.

Their study focused on visualization of quantitative information visualization. They identified and ranked the tasks shown in Figure 2.2 for evaluating effectiveness of visualization. MacKinlay extended their work by suggesting different rankings for ordinal and nominal data [7]. His study revealed that, contrary to quantitative data, for ordinal and nominal data color has a very strong perceptual accuracy. These findings have to be taken into account when designing visualizations that is expected to emphasize significant features in a data set effectively.

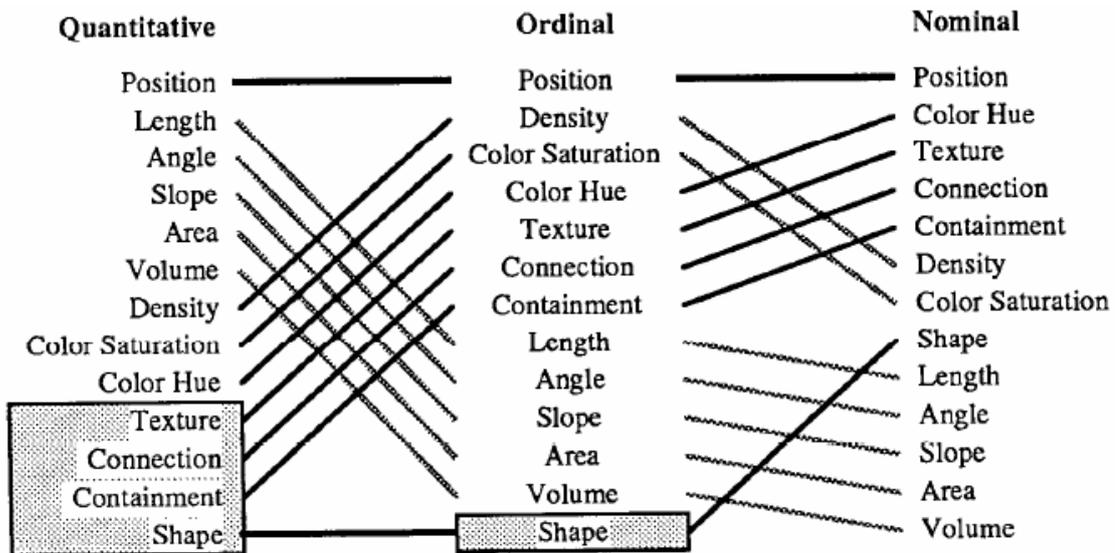


Figure 2.3 Ranking of perceptual tasks for quantitative, ordinal and nominal data values. MacKinlay developed this ranking based on existing psychophysical results and various analysis of perceptual tasks, but it has not been verified empirically. Tasks in gray boxes are not relevant for these data types.

Another issue that requires deep consideration when evaluating effectiveness of visualization is its visual clarity and refinement. Especially in displays where high dimensional data with numerous elements are conveyed, use of color, text and problems like occlusion becomes extremely important. A detailed discussion of these issues is out of scope of this thesis. Interested readers may refer to Tufte [3].

### 2.3 Qualitative Representation of Information

Qualitative representation of information implies conveying key features of a data set rather than the statistical details [33]. For a significant portion of public viewers, a thorough analysis is both unnecessary and extensively laborious especially when they are only interested in capturing a meaningful overview about the data. To answer the need for gleaning general trends out of the data rapidly and intuitively, a qualitative representation exploits visual cognition of humans.



Figure 2.4 MarketMap visualization of SmartMoney.com. The snapshot gives an overview of the rises and falls of stock prices for past 26 weeks' activity.

A convincing example of qualitative representation is MarketMap of SmartMoney.com [52]. Contrary to conventional graphs and charts used in financial data visualization, MarketMap utilizes a spatial metaphor relating volume of each stock to the area of rectangle that represents it on 2D map. Besides, stocks from same market sectors are placed adjacently in separate regions to convey their cumulative trend. The rectangles are colored in shades of green and red for identifying gaining and losing stocks. The dominant color of the map summarizes overall trend for the whole exchanges and for individual sectors. Although the information revealed in one snapshot might not be sufficient for investors, the visualization successfully communicates basic features of the data even for non-expert viewers.

## 2.4 Interactive Environments for Visualization

Difficulty in understanding and analyzing a data set arises from not knowing the degree and nature of relation between its elements. Interactive visualization environments assist viewers to understand the imperfectly known interrelations among data elements by providing means to alternate representation. Real-time interaction gives user free reign over the data set and enables different assemblies through adjustment of parameters used for depiction of the data.

Interactive environments are utilized in dynamic query tools for databases. Shneidermann et al. worked on dynamic query interfaces like Film Finder or Home Finder where users query an underlying database by adjusting some sliders or using other interface tools like buttons [10][11]. Result of the specified query is visualized as a scatter plot in 2D. The direct manipulation of search criteria generates visualizations capable of showing only elements of interest.

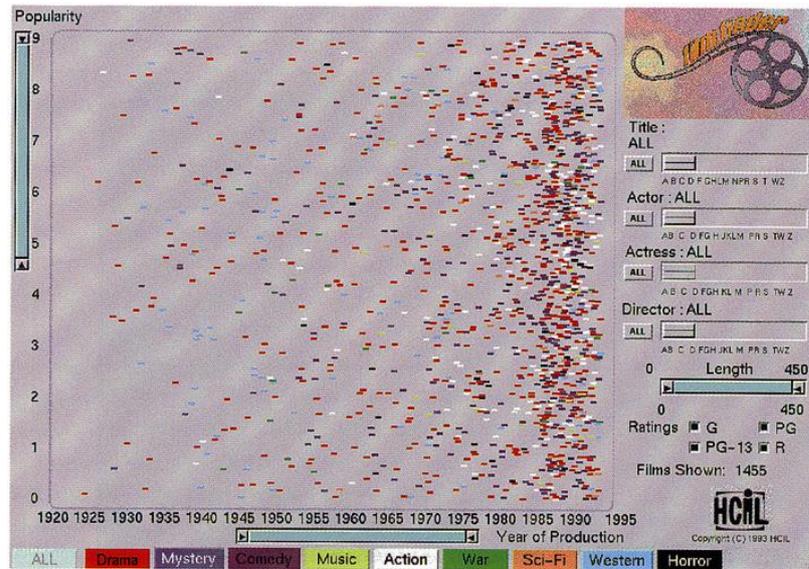


Figure 2.5 Film Finder dynamic query tool of Ahlberg and Shneiderman. Two axes of the display show years and popularity of the movies. Their genres are indicated by colors. Queries are refined by adjusting sliders on the right of the display.

Film Finder interface utilizes simple sliders for manipulating quantitative values, but nominal binary values like film rating are modified through radio buttons. Each of the query components operates as a filter reducing the number of items left in the resulting display. Effectiveness of such a dynamic query interface depends on the design of tools for manipulating the query and more complex queries require more elaborate interface controls. A detailed discussion of slider design for dynamic query interfaces can be found in Eick [12].

Dynamic query tools are significant for demonstrating capabilities of an interactive visualization environment. Showing and hiding elements with specific attributes, or zooming into region for detailed examination are properties offered uniquely by interactive visualization environments. Shneidermann pointed out also the engaging quality of learning through interactive environments and stated that “the

enthusiasm users have for dynamic queries emanates from the sense of control they gain over the database.”[10]

3D interactive environments offer an extra dimension for encoding information. However, the profound advantage of a 3D display is the use of spatial metaphors for visualization of abstract information. Significance of spatial metaphors lies in the intuitive comprehension they provide by taking advantage of human perception and cognition as developed to deal with the physical world.

However, 3D representations bring extra challenges for visualization. Lighting, shadow, texturing, six degree of freedom navigation are problematic issues that are need to be addressed in an effective 3D representation. For representation of quantitative information, 3D displays bring the extra difficulty in making precise comparisons between data elements. Items to be compared must be brought together, yet they have to be kept in the context with the remaining data. Distortion caused by perspective and occlusion of background elements in 3D representations significantly complicates interpretation of quantitative data.

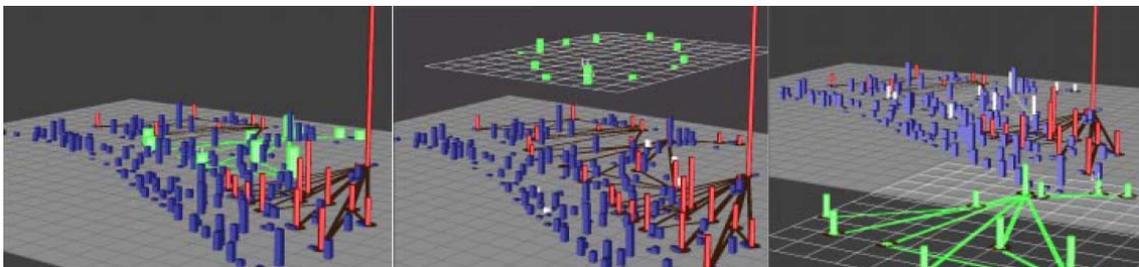


Figure 2.6 Selective dynamic interaction modes proposed by Chuah et al. The bars colored in green are a selected subset. They can be scaled or translated while their original positions and scales are shown as white shell objects.

Chuah et al. proposes a suit of interactive manipulation techniques which users can combine to solve a wide variety of problems related to 3D representation of quantitative information [13]. These techniques include creating a subset of objects dynamically with selection tools, temporally changing scale of a subset, minimizing scales of objects out of interest, elevating or translating a subset, assigning colors to selected objects to indicate similarities.

Main drawback of these extensively flexible controls is loose of context which can lead to misinterpretation of data. Chuah et al. also developed a set of constraints and a feedback mechanism. One significant component in the feedback mechanism is the use of shell objects indicating original scales and positions.

Interactivity also provides navigation tools that enable different views from the same representation of data as well as instant focus and context changes at different zoom levels. If seamless transitions are guaranteed between these views, users' ability to comprehend unresolved relations in data will be augmented.

## 2.5 Visualization of Multidimensional Data

Data sets with hundreds of dimensions are becoming commonplace in an increasing number of areas from bioinformatics to finance. Visualization of multidimensional data is a fundamental task for allowing human observers to perceive outliers, groupings or other regularities in the data.

A multidimensional data consist of vector elements with  $n$ -tuple variables, where  $n > 3$ . If these variables are dependent, then the term multivariate is used instead. In the following, two terms are used interchangeably since multidimensional data is generally multivariate.

Visualizing multidimensional data with matrices representing data elements in one axis, and characteristics in the other axis is a natural approach. In order to glean patterns from the matrix representation, Bertin developed *permutation matrices* [4]. For reorderable elements and characteristics, permutation matrices cluster similar objects by swapping rows or columns, and provide a comprehensible reading. However, permutation matrices are not suitable for visualizing large data sets with more than hundreds of elements.

Second natural choice for visualizing multidimensional data is constructing a matrix of scatter plots for subsets of variables where each variable is projected on one

axis of the display. However, generating multiple views from a single data set sacrifices overall relationship among data elements and complicates gleaning a qualitative feel.

Parallel coordinates, is an alternative technique that involves projecting each variable on a separate axis and positioning them consecutively in 2D. Line segments passing through each axis characterize individual elements. Clusters for each variable can be recognized clearly when lines pass through specific regions on an axis.

Systematic development of parallel coordinates started with the works of Insalberg [8]. He also provided improvements for interactively selecting few variables and defining a hyper surface boundary for eliminated variables. In doing so, the system enables ordering of variables in their predictive power and defining a subset of variables which summarize the data without loss of significant information.

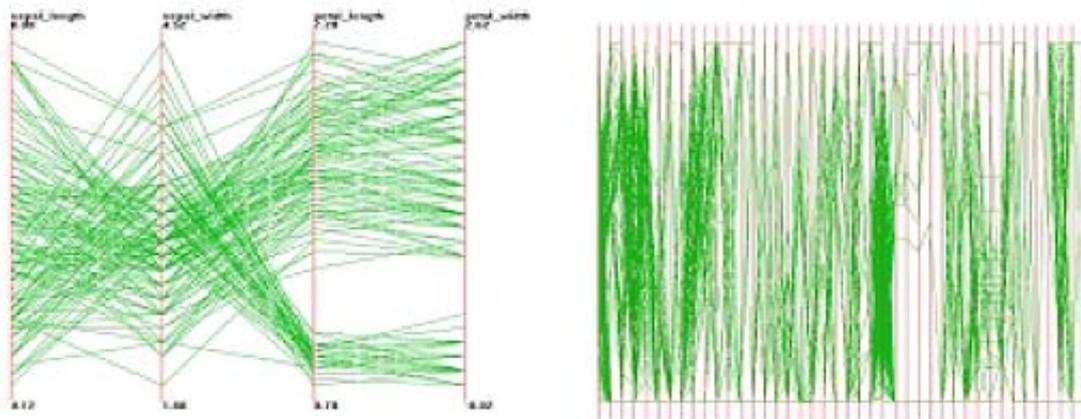


Figure 2.7 Multidimensional data visualization using parallel coordinates method exemplified in Yang et al. First figure displays a 4 dimensional data while second figure displays 42 dimensional data.

Yang et al. revealed drawback of parallel planes by providing visualizations in Figure 2.7 [42]. The first figure shows a data set with 4 dimensions and 150 data items. Individual elements and clusters can be seen clearly from the display. Second figure displays another data set with 42 dimensions and 200 data items. While the number of data items is comparable in both displays, individual elements can not be identified in the second.

The clutter problem in parallel coordinates method is addressed also in Fanea et al.[59] They propose using 3D parallel planes instead of 2D parallel coordinates to overcome the cluttering.

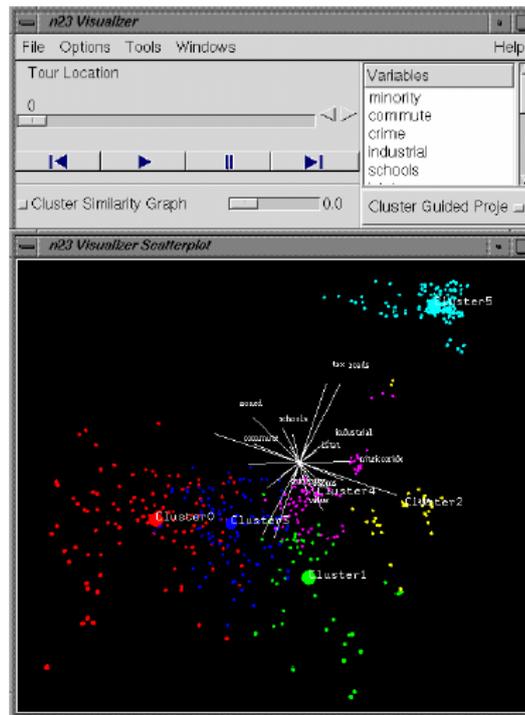


Figure 2.8 Dynamic grand tour software developed by Yang [60]. Yang limited number of projections to follow specified clusters and principal components in the data.

Another influential idea about visualization of multidimensional data is proposed by Asimov [43]. The *grand tour* technique animates iterative projections of variables from a multivariate data to orthogonal axes in 2D or 3D. Buja et al. developed interactive controls for grand tour, which enable users to select more intuitive projections among many [44]. Yang developed an interactively controlled grand tour projection environment [60]. Instead of showing all possible projections, his technique refines number of projections to follow specified clusters of objects or principal components in the data.

An important approach for generating layouts of the multivariate data is reducing dimensionality of the given data, and then projecting it on 2D or 3D display. There are several techniques used for reducing dimension of a multidimensional data including principal component analysis, Kohonen's self organizing maps, and multidimensional scaling (MDS).

### 2.5.1 MDS

MDS is a set of related statistical techniques used for mapping a multidimensional data onto a lower dimensional Euclidian space, suitable for graphing [39] or scatter plot visualization. Main objective of MDS is to preserve existing relationships within data and revealing latent structures intuitively, while reducing its dimensions [50][49].

An MDS algorithm quantifies similarities between individual elements of a data set usually by defining a distance relationship within high dimensional space. Resulting item-item similarity (or dissimilarity) matrix is used for assigning a location for each item in a low dimensional Euclidian space. The output configuration of elements is expected to approximate high dimensional inter relations among data elements.

The classical MDS method, developed by Torgerson, operates by eigenvector analysis of the item-item dissimilarity matrix and produces a layout based on linear combination of dimensions [61]. However, Torgerson method is not suitable for creating real-time interactive visualizations because the procedure has  $O(N^3)$  complexity, and even single alteration in the original data requires entire recalculation.

Iterative techniques are proposed to overcome these difficulties which are referred as non-metric MDS [62][50]. Non-metric MDS operates by iteratively minimizing an error or loss function which evaluates how well the derived configuration fits the given dissimilarities. Typical error function in MDS is proportional to difference in distances among each pair of objects in high and low dimensional space. Kruskal's non-metric MDS method is discussed in section 2.6.

### 2.5.2 Network Visualization

A great deal of multidimensional data encountered can be visualized by networks with nodes corresponding to data elements and links representing relationships among them. Conventional node and link displays are effective for visualization of small sparse networks with tens to hundreds of nodes. However, larger network visualizations

encounter three major problems. Clutter of the displays, positioning of nodes to permit a meaningful interpretation and encoding additional information are problems to be addressed in effective network visualizations [16].

Limited viewing area of displays is the main reason of clutter problem in visualizations of multidimensional large datasets. This limitation also brings forth a tension between showing low level details and high level context information simultaneously and generally users need to access both [2].

Several methods are proposed for providing access to local details without losing the global information which helps users to stay oriented. Chen summarizes those methods under three categories: (i) overview + detail views- displaying overview and detailed information in multiple views, (ii) zoomable views- displaying objects on multiple scales, (iii) focus + context views – displaying local detail and global context in integrated but geometrically distorted views [2].

Major drawback related to multiple view displays is the discrete nature of the visualization which prevents an integrated comprehension of the data as whole. Although zoomable views provide a seamless transition between different contexts, global information is lost temporally. If implemented carefully, focus + context views attained by distorted geometries are most effective for achieving an integrated view in multiple scales. Fisheye views and hyperbolic geometry displays are the most widely accepted examples of distorted displays.

Founded by Furnas [48], fisheye views display objects in focal area with a finer detail compared to objects out of focal area. Furnas formalized the method by defining a metric, termed as degree of interest (DOI), which determines the scale of the object in the display relative to its distance from the focal area. DOI of an object is the summation of an a priori measure of importance and distance from the focal. Elements of interest are made more visible by assigning them higher DOI scores. In doing so, the visualization reflects significant information about the nature of data more intuitively.

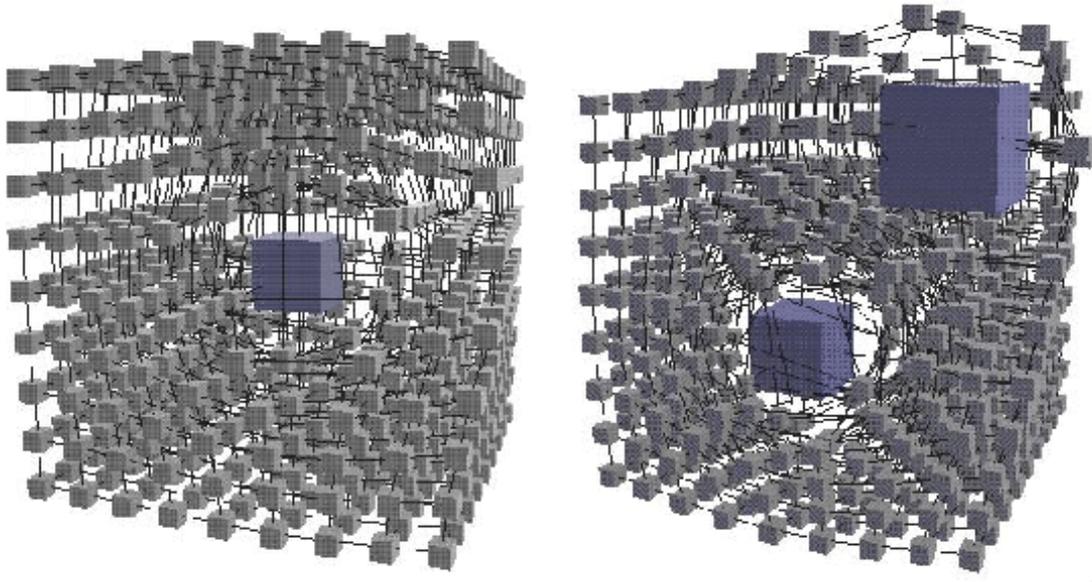


Figure 2.9 One focal and bi-focal fisheye views implemented by Carpendale et al. [45]

Figure 2.9 depicts application of a single and bi-focal fisheye view to 3D network visualization by Carpendale [45]. The neighborhood of each focal point has much lower density of nodes, this gives them higher readability.

*SemNet* visualization system of Fairchild et al. utilizes fish eye view to give a useful balance of local details and surrounding context [64]. Aside from that, *SemNet* visualization system enables users to examine spatial subsets in detail by utilizing different positioning strategies of nodes. These strategies include MDS, mapping functions and simulated annealing combined with interactively set user choices.

Lamping and Rao proposed using non-Euclidian or hyperbolic geometry to distort space itself for giving a natural focus + context character to the network visualization [46]. Hyperbolic geometry is based on a space with exponentially increasing coordinates. Therefore, a hyperbolic geometry display depicts objects out of focal more densely. For low level detailed view, the user drags part of the visualization to the center area giving the maximum magnification. Lamping and Rao use a projection of hyperbolic geometry into 2D space. Their work extended by Munzner who used hyperbolic geometry in 3D space.

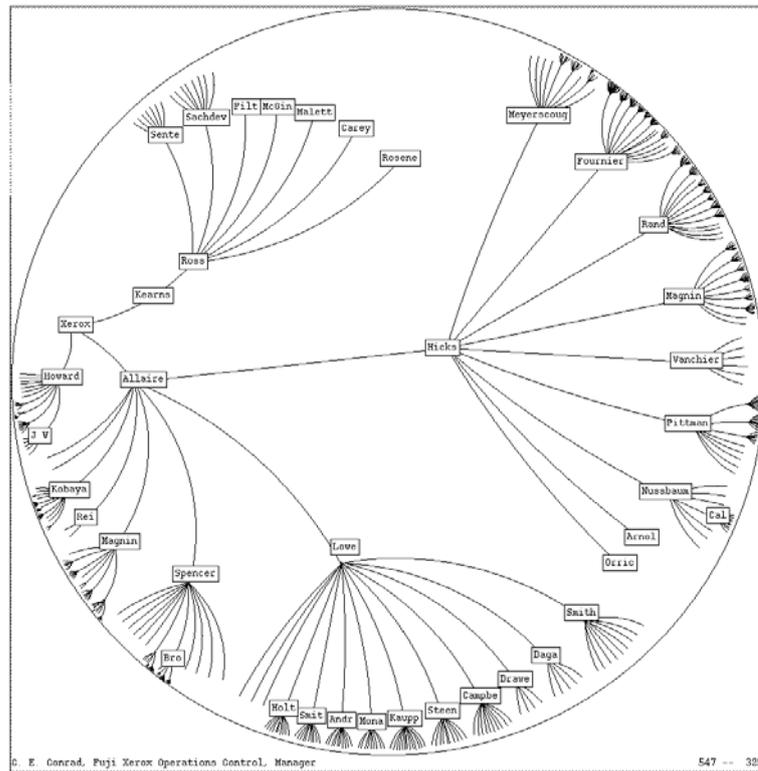


Figure 2.10 Lamping and Rao's hyperbolic geometry network visualization. The hyperbolic space is projected on a 2D plane.

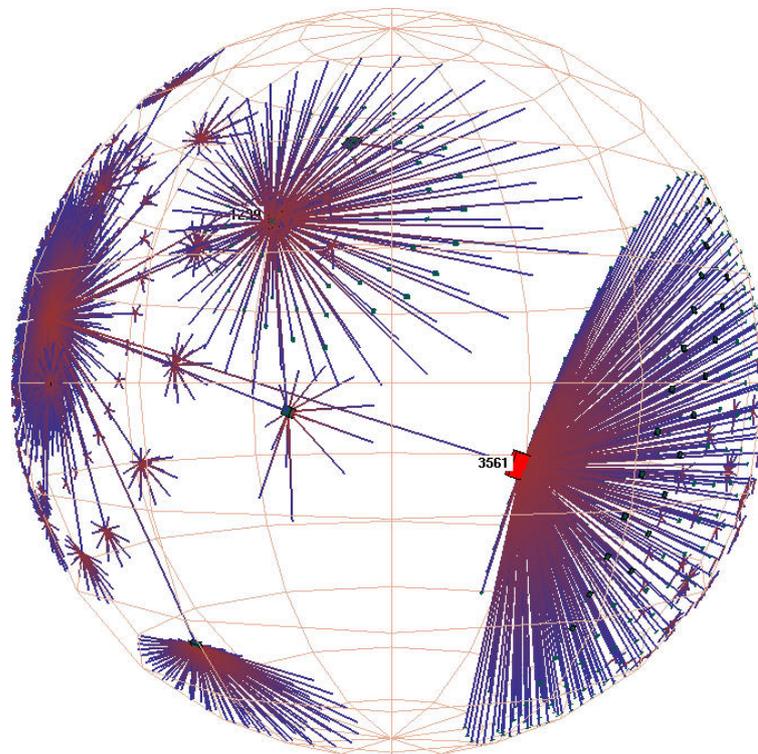


Figure 2.11 Tamara Munzner's network visualization based on 3D hyperbolic geometry. As clusters of nodes move away from the focal they appear smaller and with less detail.

A significant research field on displaying network data is graph drawing by force-directed placement. Graph drawing techniques deal with the node positioning problem and seek a layout that conveys the meaning of the diagram quickly and clearly. Battista et al. formulated the basic aesthetic criteria of an effective layout that determine readability as symmetry, minimization of edge crossings, uniform edge lengths and uniform distribution of nodes [34]. When these criteria are met, resulting layout will be displaying more related nodes closer, and get rid off the long edges and dense regions which clutter the display. Force-directed placement is a family of iterative methods which proposed to meet these aesthetic criteria through utilization of spring-embedders. The idea evolved from a VLSI technique whose aim is to optimize the layout of a circuit with the least number of line crossings. Variations of the force-directed methods are discussed in detail in section 2.6.

## 2.6 Map Visualizations

Cartography, the study of map making, is an ancient practice dating back to 7<sup>th</sup> millennium BC. Because of the expertise gained over centuries about creating and reading them, maps are effective graphic representation tools that we are accustomed with. However; the strongest aspect of visualizing information in the form of maps is the utilization spatio-cognitive skills of humans as developed to deal with the physical world. *By virtue of their spatio-cognitive abilities, humans are able to navigate through geographic space as well as meaningfully communicate geographic information represented in cartographic form* [23]. Therefore, maps are exploited in visualizing all sorts of geography related data in the form of thematic maps [20].

A thematic map displays additional information about distribution of a qualitative (e.g. geographic sites) or quantitative (e.g. population) data on a standard geographic map. Thematic maps show vast variations but few significant methodologies are discussed below. Choropleth maps convey statistical information by color coding areas defined by geographic or political boundaries. As a result, large areas receive highest visual emphasis even though they might be less significant with respect to the visualized data. Dot-scatter maps or pin maps show the existence or occurrence of a data attribute

by placing icons on top of the geographic map. These maps successfully reveal spatial patterns about the distribution of data within geographic context.



Figure 2.12 Dot map showing birth places of the 3005 Ming poets in China between 1368-1644. See Chen-Cheng Siang, “A Historical and Cultural Atlas of China”, map 62. Reproduced from Tufte.

The dot map in Figure 2.12 shows the geographic distribution of birthplaces of Ming poets in China. One significant problem is the occlusion encountered in regions where data is densely distributed.

Grid-square maps, on the other hand, handle the occlusion problem by dividing up the map into equal-sized small units to be color coded. Grid-square statistics for distribution of population in Japan in Figure 2.13 is a well-implemented instance.

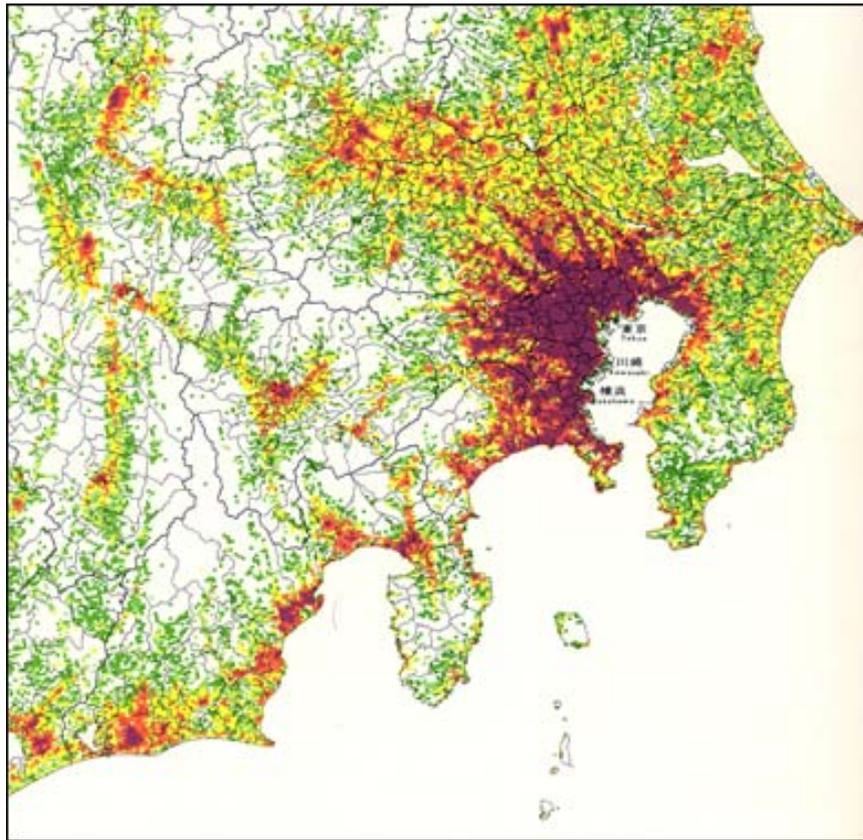


Figure 2.13 Grid-square map showing population density of Japan. The map is divided into equal sized small rectangles and then each rectangle is colored to represent population of that area. See Hidenori Kimura, “Grid Square Statistics for the Distribution and Mobility of Population in Japan”, Statistics Bureau, Tokyo. Reproduced from Tufte.

### 2.6.1 Area Cartograms

Area cartograms are special thematic maps which associate quantitative, spatial data with area on a distorted geographic map. These deformed maps are effective graphic representations for conveying data intuitively in a geographic framework and have been powerfully utilized to display census results, election returns, disease incidences and many other geography-related datasets [26].



Figure 2.14 Gillihan’s cartogram made with plasticine and a rolling pin. The cartogram shows distribution of Smallpox incidences in California between 1915 and 1924. Regions are scaled proportional to the number of disease incidences.

The idea is to scale regions of the map proportional to the data distribution while maintaining geographic accuracy to some extent. Interpretation of a deformed geographic map mostly relies on a priori knowledge of the viewer about the physically accurate version of the same map. Therefore, maintaining basic properties such as shape, topology, orientation and contiguity is necessary to guarantee intuitive recognition of a cartogram [32].

Previous works incorporate different constraints to reach an optimal compromise for trading shape and area adjustments. Tobler developed a Pseudo-Cartogram method that creates an equal density approximation by compressing or expanding lines of latitude and longitude until a least square error solution is obtained [27].

Gastner et al. adopted linear diffusion process method to create *density equalizing* maps which show equal density distribution of a statistical data [31]. They produced a cartogram visualizing US presidential election of 2000. In the map in Figure 2.15 individual states are easily recognizable while each state is scaled to a size proportional to the number its electors. In doing so, the map eliminates unnecessary information about exact geographic areas of individual states and replaces it with the distribution of electoral votes. Color coding of the map on vote percentages enable viewers to see dominant tendency in each state.

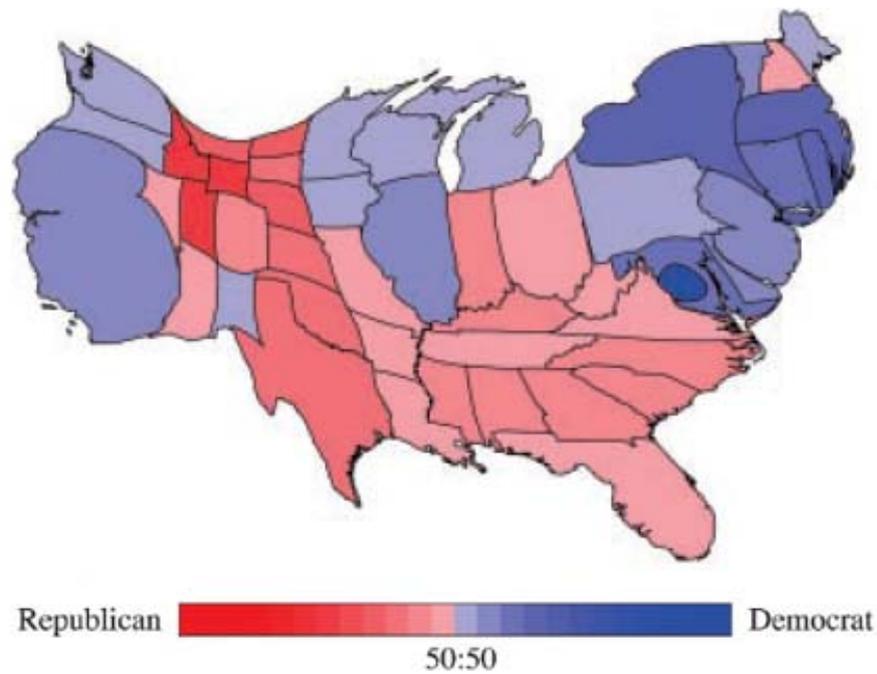


Figure 2.15 Gastner's density equalizing map showing 2000 election results in US.

Cauvin and Schneider created cartograms by deforming an underlying regular grid [29]. Grid mesh structure is formed by dividing geographic space into a set of finite elements. Thematic data and spatial distribution are reduced to load pressures applied on the mesh. Both data components are legibly displayed on the resulting cartogram.

Kocmoud and House used a spring model acting on a map with constraints to maintain certain topographic features such as angles or lengths which aid in preserving essential cues for the recognition of region shapes [30]. Although their results are better than those derived from most other methods, their method proposes a complex optimization algorithm with a prohibitively high execution time.

In addition to shape and topology preserving issues, high time-complexity of optimization algorithms restricts use of area cartograms to static applications. As a fast algorithm for generating continuous cartograms, Keim et al. proposed medial-axis based optimization, where the vertices of the polygon mesh are incrementally repositioned along medial axis segments used as scanlines [32]. In this way they reduced cartogram generation time significantly.

As defined by Tobler, an area cartogram is an equal area map projection of a particular kind. Unlike general thematic maps where visualized data is transcribed by symbols, an area cartogram is the picture resulting from the interaction between the data and the spatial structure [29][28].

### 2.6.2 3D Virtual Environments in Geographic Data Visualizations

Computers have revolutionized the process of making maps as seen in the advancement of area cartograms. However, 3D navigable worlds integrating cartographic methods with other data analysis tools have a more profound impact on computer aided visualization of geographic data.



Figure 2.16 Lokuge et al.'s visualization of tourist attraction sites in Boston. Popularity of sites is shown with height bars over the 3D map.

Stacked bars over 3D maps are a natural extension of cartographic visualization techniques to the 3D virtual environments. These visualizations simply superimpose any type of information on the natural geographic layout familiar to the user. Work of Lokuge et al. displays tourist attraction data of Boston with height bars indicating popularity of the geographic sites [19]. Lodha classifies similar works as three dimensional versions of pin maps and defines their major advantage as the reduction of clutter by the extra axis provided for encoding additional information [18].

Visualizing abstract data overlaid on physical geographic space provides an immersive environment which exploits human sensory and cognitive systems at the highest level. When used in conjunction with interactive data analysis and navigation tools, such environments provide visualizations that harness visible and non-visible characteristics of a real physical space.

MacEachren et al. investigated effectiveness of 3D virtual environments in visualization of geography related data [24]. Specifically for geographic visualizations, they define the term *spatially iconic geo-visualization* as 3D virtual environments mapping three dimensions of the physical space directly to the three dimensions of the display. According to their study, naturalness and immersion offered with this direct mapping of the physical world should not limit geographic visualization to be only the representation of physical reality. A geo-visualization does not require being spatially iconic in all respects or dimensions, *different insights can be gained by using one or more of the virtual environment axes to depict a non-geographic variable* [24].

Interactive map distortion techniques are significant for their quality of redefining physical space for achieving focus + context views, fisheye views or for providing seamless transitions between two different thematic maps. Zanella et al. provided a user study investigating human comprehension of map distortions [22]. Their study demonstrated need for appropriate visual cues (e.g. grid, shadow) to ensure that distortion does not affect an individual's internal model of space.

### **2.6.3 Visualization of Geographic Networks**

Visualization of spatial networks is significant in analyzing transportation, telecommunication, migration data sets and understanding related human behavior in a geographic framework. Network visualization techniques discussed in section 2.5.2 focus on displaying and interpreting the structure of the network itself rather than the underlying data. However, spatial network visualizations require further elaboration because of the necessity of analyzing both the network data and its geographic context.



Figure 2.17 NFSNET T1 internet traffic backbone visualization of NCSA.

Geographic network visualizations began to flourish with attempts to visualize internet traffic in late 80's. Figure 2.17 depicts the visualization of network traffic in the NFSNET T1 backbone in 1991 by NCSA. The color coding of lines represent volume of the traffic, ranging from zero bytes (purple) to 100 billion bytes (white). Projecting the planar map in 3D space, namely using a 2.5D display, enables a clear distinction over the crossings of lines. The layering of the white lines on top of the colored lines enables reader to comprehend the main flow and major nodes instantly.

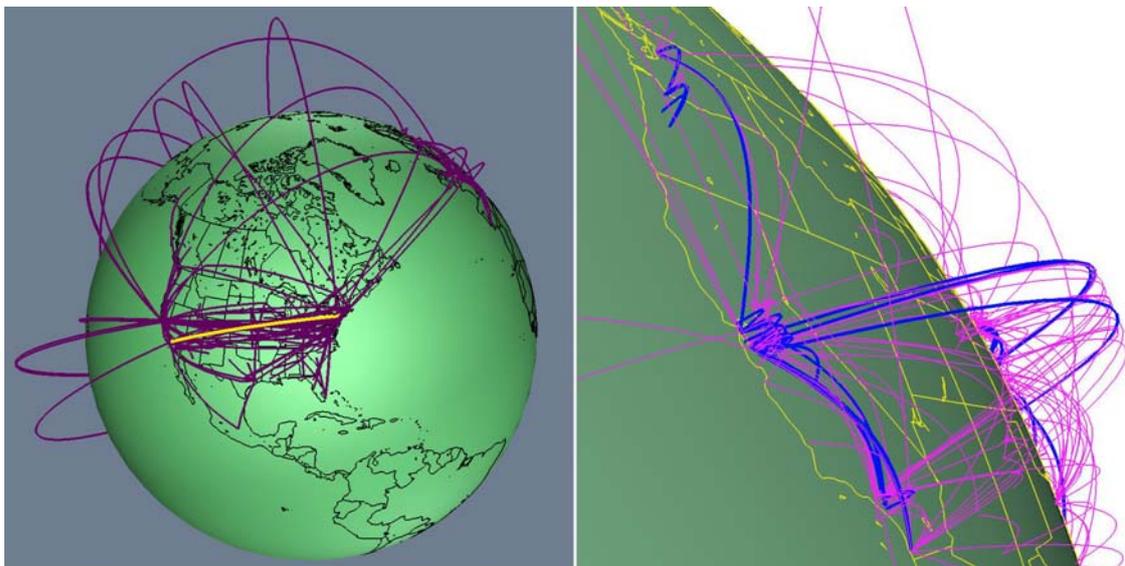


Figure 2.18 Munzner's visualization of MBone, multicast backbone of internet.

Munzner et al. visualized MBone, multicast backbone of internet, by arcs over a globe [63]. The interactive 3D representation permits further analysis of data by use of data analysis techniques like grouping and thresholding. They distributed animations of live data feed using VRML, which allow viewers to analyze data more effectively than would be possible with still pictures or pre-made videos.

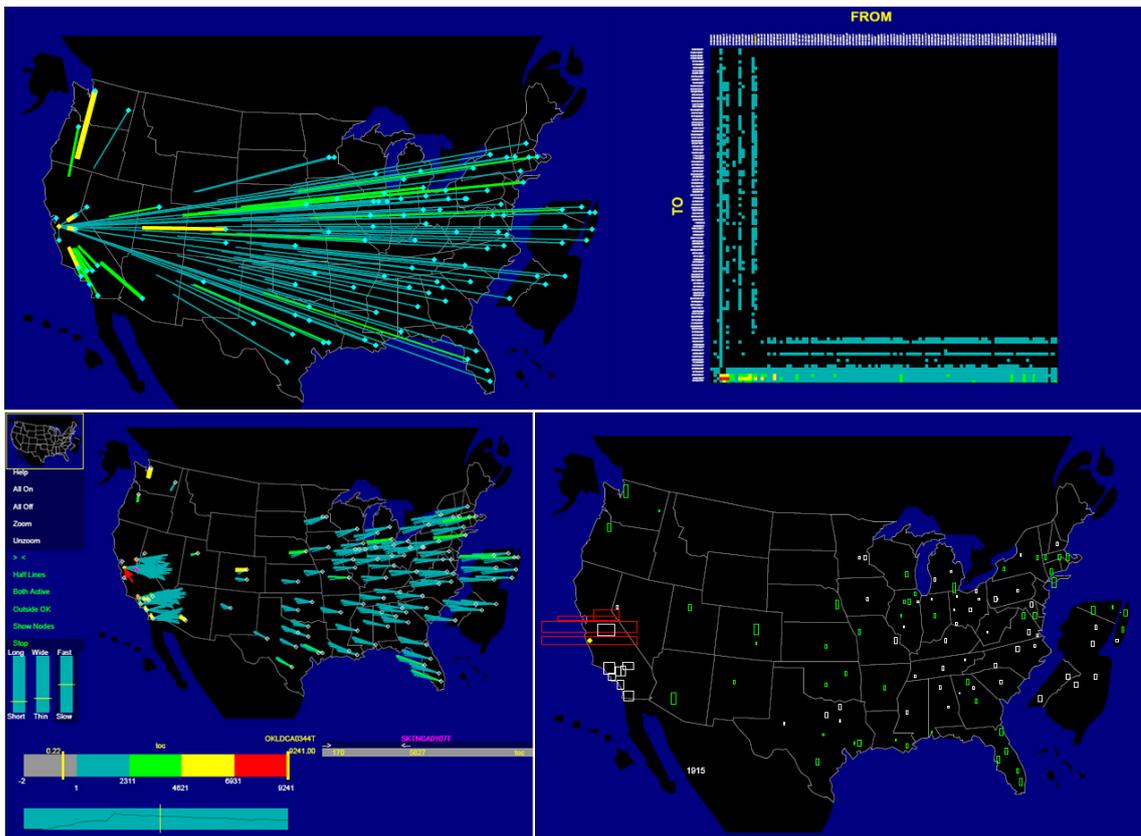


Figure 2.19 *SeeNet* network visualization system. First figure displays connections emanating from a single node with lines. Half lines code the overload by direction. Second figure shows the matrix representation of the same network data. The vertical or horizontal ordering of nodes can be modified to ease readability of patterns. In third snapshot only half-lines between nodes are drawn to eliminate clutter problem. Fourth figure shows the aggregate load of each node encoded by dimensions and color of each rectangle.

A classic and extensive work on visualizing spatial networks is provided by Becker, Eick and Wilks [14]. They presented a highly interactive visualization environment, *SeeNet*, involving static displays and animation with three different display modes. Link, node and matrix displays are provided with a suit of proper interaction tools. Viewers are able to focus on sub-regions of the map, identify elements

of interest by a mouse click, decide which data to be displayed or suppressed, and adjust parameters like time intervals or aggregation.

Their work is extended by Cox et al. as *SeeNet3D* with similar interactive controls [15]. The tool visualizes data on a 3D globe combined with drill-down display modes displaying all links emanating from a designated focal node. *SeeNet3D* also provides a seamless transition between globe map and flat map. One significant contribution of restricting the network display to a sphere is the ease of navigation. In general 3D network displays, navigation is a problematic issue because users may easily lose a sense of overall context. Users interactively rotating around a globe have little chance of becoming disoriented [15][16].

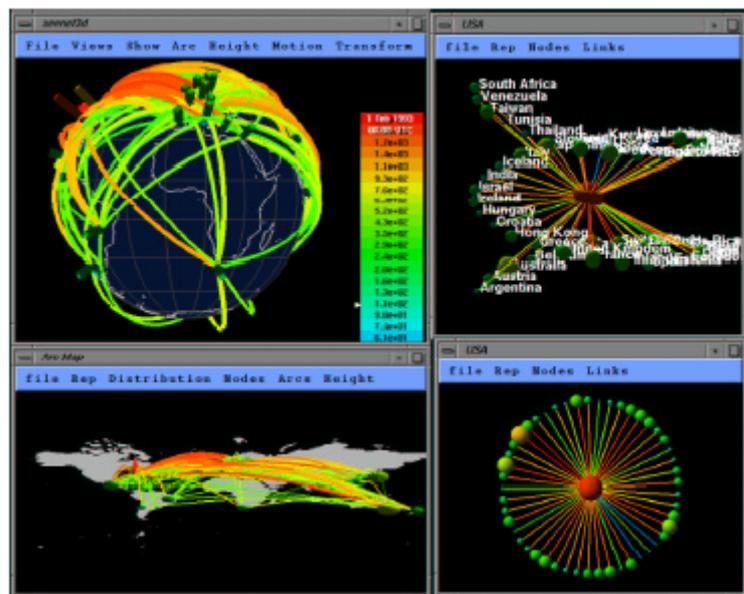


Figure 2.20 *SeeNet3D* network visualization system. First figure shows color coded network connections over 3D map. Second figure shows a helix representation of connections from a single node. Third figure shows all connections on a planar map. Final figure is another drill down network view showing connections from a single node and their load is coded with the size of spheres representing nodes.

## 2.7 Spring Embedders in Visualization

Spring-embedders are utilized in visualization to solve highly complex optimization problems iteratively. Compared to other methods such as finite elements or simulated annealing, spring-embedder methods offer an intuitive physical analogy to be followed easily and robust numeric solutions are provided for their computation. Spring-embedder implementations in force-directed placement and non-metric MDS techniques are in particular relevant for this thesis.

The spring-embedder model for graph optimization originally proposed by Eades as a heuristic approach to achieve two aesthetic criteria: uniform edge lengths and symmetry whenever possible [36]. In his model nodes of a graph correspond to a set of steel rings and edges correspond to a set of springs. Two categories of forces are calculated (i) repulsive forces that are calculated between every pair of nodes, (ii) attractive forces that are calculated only between connected nodes. Since attractive forces are calculated only for connected nodes, the time complexity is reduced to  $O(|E|)$ , although the repulsive force calculation is still  $O(|V|^2)$ , where  $E$  stands for the edges and  $V$  stands for the vertices in the graph. His equations for attractive and repulsive forces were modeled as:

$$f_a(d) = k_a \log d \quad \text{and} \quad f_r = k_r / d^2,$$

where  $f_a$  is the attractive and  $f_r$  is the repulsive force and  $d$  is the distance between two nodes. When the system is “let go”, the attractive and repulsive forces move the system to equilibrium, optimistically where forces on springs are minimized.

Fruchterman and Reingold refined this approach by adding slightly different constraints on a similar system [35]. Their constraints were evenly distributed nodes, minimized edge crossings and uniform edge lengths. Their force model is similar to Eades’ but rejected his formula for  $f_a$  since it was inefficient to compute. In their model attractive and repulsive forces are defined as:

$$f_a = d^2 / k \quad \text{and} \quad f_r = -k^2 / d .$$

Kamada and Kawai variant of force-directed placement follows real physical model of spring-embedders and adopts Hooke’s law for the force model [37]. Additionally, their method seeks ideal distance constraint between every pair of nodes,

which is proportional to the length of the shortest path between the nodes. To achieve this constraint, relaxed lengths of springs between nodes are initially defined as proportional to the shortest path between them. For low dimensionality graphs, the procedure yields a layout where screen distances approximate to edge lengths.

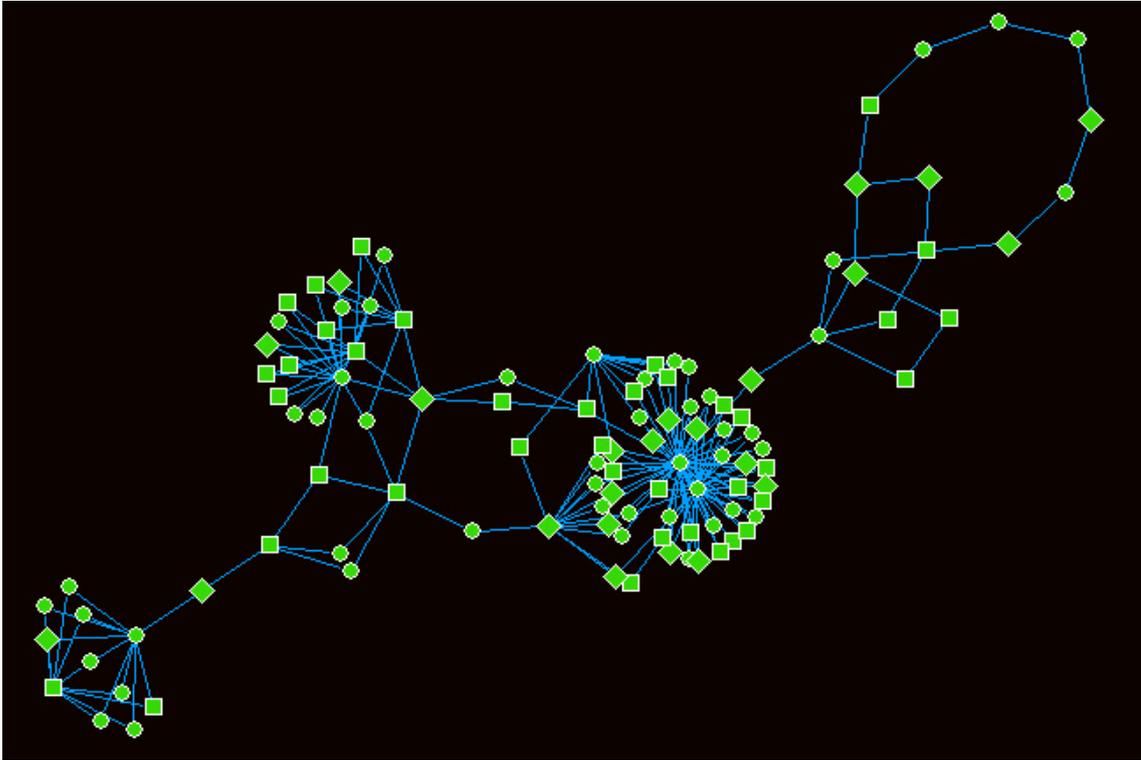


Figure 2.21 A graph layout optimization with Kamada-Kawai's force-directed placement algorithm. The algorithm effectively eliminates long edges which would clutter the display.

Kamada and Kawai approached the graph drawing problem as an energy minimization process. Nodes start from user-defined initial positions, and are iteratively repositioned to minimize the overall "energy" of the spring system. They formulated total energy in the system as:

$$\sum_{i \leq j \leq V} k_{ij} ( |n_i - n_j| - d_{ij} )^2,$$

where  $n_i$  and  $n_j$  corresponds to the positions of  $i^{\text{th}}$  and  $j^{\text{th}}$  nodes and  $d_{ij}$  is the optimum distance between these nodes. For each node, a partial differential equation is solved to find a new position which minimizes energy of springs connected to that node. Repositioning of each node is repeated until the energy goes under a preset threshold. At each iteration only one node is repositioned; therefore the inner loop only needs to recalculate the contribution of that node to the energy of the system taking  $O(|V|)$  time.

Kruskal's non-metric MDS technique employs an error minimization method for visualization of high dimensional data sets in a low dimensional space [50]. The error, or *stress*, is analogous to the summation of forces acting on springs in Kamada Kawai method. Stress is proportional to the difference of inter-object distances in high and low dimensional space. The distance measure between object  $x_i$  and  $x_j$  is defined as:

$$d_L(x_i, x_j) = \left( \sum_{k=1}^m |x_{i,k} - x_{j,k}|^L \right)^{1/L}$$

where  $m$  is the dimension of the data objects and  $L$  is a parameter that takes value between interval  $[1, \infty]$ . Kruskal defined  $L = 2$  which reduces the distance equation to the Euclidian distance function. Given the distance function, *stress* is defined as below:

$$Stress = \frac{\sum_{i < j} (d_{ij} - g_{ij})^2}{\sum_{i < j} g_{ij}^2}$$

where  $d_{ij}$  denotes distance in high dimensional space and  $g_{ij}$  denotes low dimensional layout distance. For minimizing the *stress*, a steepest-descent algorithm starting from an initial configuration and, at each iteration, moving some or all of the nodes along the gradient of *stress* is employed.

### 3 VISUALIZATION SYSTEM

Motivated by a variety of techniques discussed in section 2, this thesis proposes an interactive visualization environment for time-series spatial network data within geographic framework. It incorporates two different approaches for achieving both high-level global context visualization for qualitative representation of data and low-level detailed visualization for thorough data analysis. The first approach tightly integrates graph visualization and cartographic visualization techniques in a novel manner, whereas the latter employs conventional geographic network visualization techniques as referred in section 2.6.3.

The global context visualization employs spring-embedders for drawing and animating a graph in which each geographic location corresponds to a node and non-spatial data components correspond to relations between these nodes. The graph visualization algorithm positions more related nodes closer, according to the time-series input data. In this sense, the proposed technique can be compared to force-directed placement methods. However, the technique does not follow force-directed placement in any precise sense, but instead exploits its key features. The single most important distinction lies in the geographic constraints applied on the system. Connecting each node to its original geographic location and to its geographic neighbors with additional sets of springs, the system reaches a configuration where geographic layout is preserved to a degree for assuring intuitive recognition.

The proposed method might be compared to MDS, in the sense that both convert high dimensional data relations to Euclidian distance relations on a 3D display. Given both geographical distance relations and data relations (similarity) matrix, the suggested method aims to find an optimal solution for placement of nodes where both geographic context and non-spatial relations are reflected. Whereas, an MDS algorithm considers only latter and positions nodes to minimize error between input similarity measures and output distance measures.

To illustrate the difference between the proposed method and MDS, a small portion of US domestic air flights data set is visualized using PERMAP [51], an MDS visualization software in Figure 3.2. The visualized data consists of number of flights among 15 airports and their geographic locations. The similarity matrix used in MDS is formed taking into account both number of flights and geographic distances between each pair of airports. Although geographic distance relations are considered, original geographic layout of airports displayed in Figure 3.1 is significantly lost. On the other hand, our approach limits variation of nodes and displays data deliberately less accurate for the sake of permitting intuitive recognition of the geographic framework.

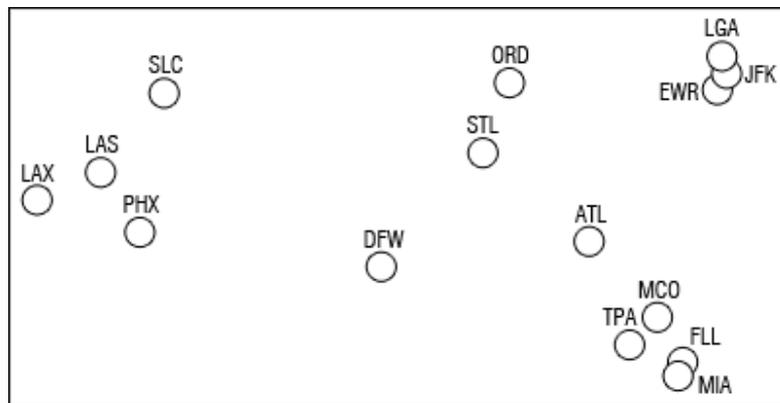


Figure 3.1 Geographic distribution of airports in Figure 3.2

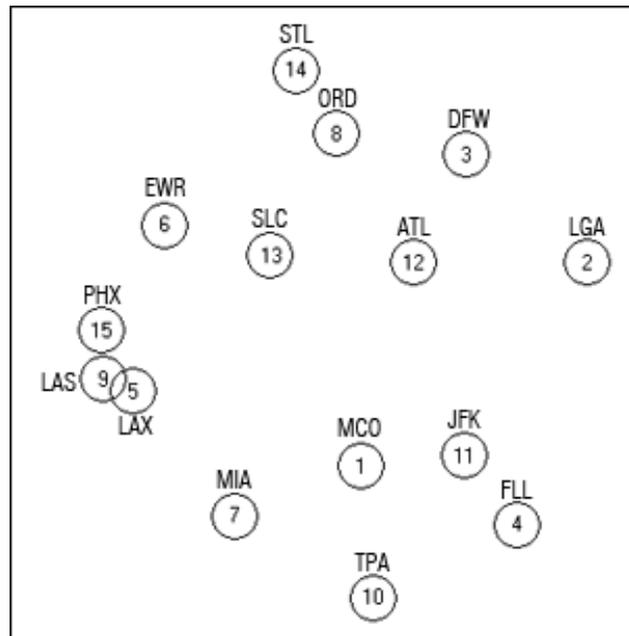


Figure 3.2 Air traffic data among 15 airports is visualized using PERMAP, an MDS visualization software. Geographic distribution of nodes in Figure 3.1 is significantly lost.

However, communicating geographic context of the data requires additional visual cues. Therefore, a map morphing technique is employed to visualize graph animation within a geographic context. Once graph nodes are positioned, the mesh surface covering these nodes is updated to fit exactly their modified positions. The geographic map projected on the surface deforms as the position of graph nodes change according to the time-series data being fed. Resulting map animation highlights variations in the data by exploiting a priori knowledge of the viewer about physically accurate version of the map. Hence, the map animation facilitates comprehension of non-spatial data with respect to the geographic framework.

The global context view provides a perceptive interface to be observed by a diversity of viewers. However, interested viewers may need to drill down to see statistical details about the data. The visualization system provides a suit of analytical tools and interactive widgets to examine data in detail. Viewers are able to observe the data as an animation of height bars over 3D map and/or view arcs showing connections between nodes. They are able to select a node and view connections emanating from that node while information about the selected node is displayed bottom left of the display. Also, time interval for the animation steps is an adjustable parameter. Animation is the most natural technique to analyze time-series data. However, when necessary, users can freeze the animation by clicking the pause button and analyze a snapshot image for a specific time-interval.

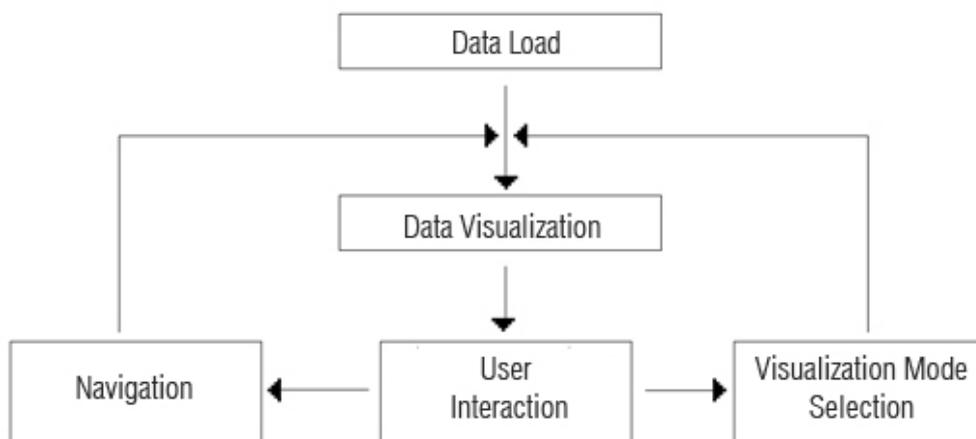


Figure 3.3 Visualization system diagram.

Interactive navigation enables viewers to glean information from different point of views and zooming levels. Visualizing data on a sphere eliminates significant problems related to navigation like loosing the context and getting disoriented [15][16]. The data visualization process is summarized in the diagram of Figure 3.3.

### 3.1 The Global Context Visualization Technique

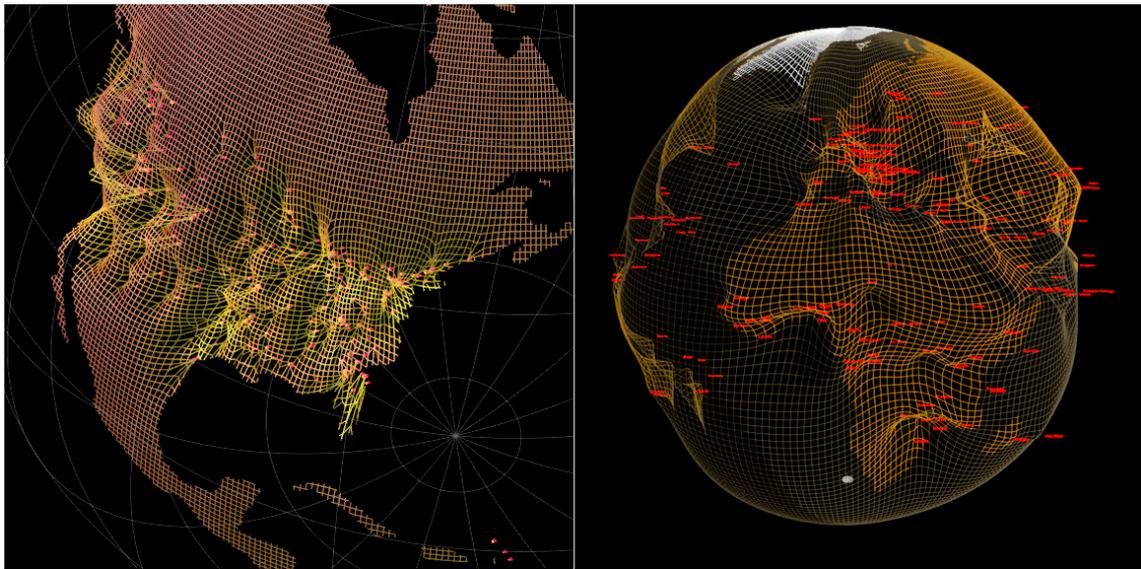


Figure 3.4 Global context visualization for two different data sets. The map surface is deformed to fit the underlying graph layout.

#### 3.1.1 Input Data Structure

The proposed visualization technique can be applied to a class of network datasets where each node has a spatial component and a non-spatial proximity measure with other nodes. Where every node,  $g_i = (x_i, y_i, z_i)$  is a point in 3D Euclidian Space, and  $d_{ij}$  is the Euclidian distance from  $g_i$  to  $g_j$ , the geographic distance relations matrix  $D$  is as follows:

$$D = \begin{bmatrix} d_{11} & d_{12} & \dots & d_{1n} \\ d_{21} & d_{22} & \dots & d_{2n} \\ \dots & \dots & \dots & \dots \\ d_{n1} & d_{n2} & \dots & d_{nn} \end{bmatrix} \quad \begin{array}{l} d_{ii} = 0 \text{ for all } i, \\ d_{ij} = d_{ji} \text{ for all } i \text{ and } j \text{ (symmetric)}. \end{array}$$

The proximity measures or data relations between each node are given by the matrix below:

$$\Delta = \begin{bmatrix} \delta_{11} & \delta_{12} & \dots & \delta_{1n} \\ \delta_{21} & \delta_{22} & \dots & \delta_{2n} \\ \dots & \dots & \dots & \dots \\ \delta_{n1} & \delta_{n2} & \dots & \delta_{nn} \end{bmatrix} \quad \begin{array}{l} \delta_{ii} = 0 \text{ for all } i, \\ \delta_{ij} \in R^n. \end{array}$$

The values of matrix  $\Delta$  vary over time. For a time interval between  $t_1$  and  $t_2$ ,  $\sum_{t_1 \leq i < t_2} \Delta_i$  is computed. Given above two matrices, modified Cartesian coordinates of every object in 3D is retrieved using spring-embedders for each animation step.

### 3.1.2 Spring-Embedder Model

The spring-embedder model utilized for graph drawing is a modified version of Kamada and Kawai's force-directed placement method. The single most important distinction between the two is the additional connectivity introduced for maintaining geographic context.

In the proposed method, each node is connected to its original geographic position with a zero length spring (anchor spring). An anchor spring functions to limit variation of the node and allows a tolerable deviation from its original position.

Furthermore, geographically close nodes are connected to each other with uniformly distributed springs (neighboring springs) with a heuristic method described in section 3.1.2.2. Neighboring springs aid to preserve geographic contiguity and orientation.

Initial lengths of neighboring springs are proportional to the Euclidian distance between the nodes they connect. As a result, a force applied on a node effects other

geographically close nodes, inversely proportional to the Euclidian distance between them. Although this behavior is undesirable in general graph drawing methods, it reflects a meaningful implication for geographic networks. Because of the geographic availability, human activities in a geographic location do not have a discrete effect, but rather have a spreading nature. Neighboring springs reveal this spreading nature of the geographic data.

The anchor and neighboring springs are solely based on geographic data and they exert forces on the nodes to maintain original geographic layout. Besides, each input from the data, which defines a relation between two nodes, is reduced to a three dimensional vector exerting a force on the spring connecting these nodes. These springs are initiated between nodes when a relation is detected and their relaxed length is defined proportional to geographic distance between the nodes. This is similar to Kamada and Kawai's method where relaxed length of a spring is defined proportional to the shortest path between the nodes it connects. As a result, each input from the data deforms the original geographic layout and positions related nodes closer.

### 3.1.2.1 Force Model

The physical behavior of springs is based on Hooke's law. Therefore, the force exerted by each spring is strictly proportional to the difference of its current and pre-defined relaxed length. The simple force equation is as follows:

$$f = k \left( |n_i - n_j| - d_{ij} \right),$$

where  $n_i$  and  $n_j$  is the current position of nodes and  $d_{ij}$  is the geographic distance between them and  $k$  is some constant.

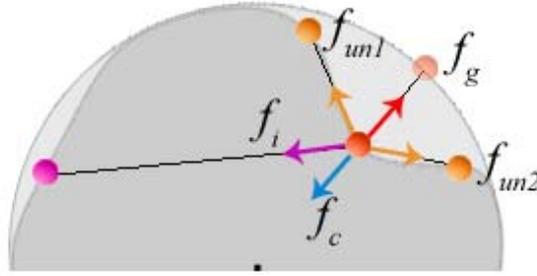


Figure 3.5 Simplified force diagram for a node with two neighbors and a single data relation.

Based on the spring connectivity described above section, there are four sets of forces acting on a single node in each animation step: (i) forces exerted by anchor springs  $f_g$  (ii) forces exerted by neighboring springs  $f_{un}$ , (iii) forces based on relation data  $f_i$  and (iv) central force  $f_c$ .

$$f(n) = f_g(n) + \sum_{u \in N} f_{un}(n) + \sum_{i \in F} f_c(n) + \sum_{i \in F} f_i(n)$$

Total force acting on a single node  $n$  is the summation of above forces for the sets  $F$  and  $N$ , where  $F$  is the set of data inputs in a single animation step, and  $N$  is the set of connected neighbors.

The central attraction force is exerted by each input from the data on regarding node, and is directed towards the center of the Earth. This force has equal magnitude for each input and is required to relate total relation capacity of each node with a visual effect, which in this case is the closeness to the center of the Earth. The forces exerted on a single node and their directions are illustrated in Figure 3.4 for a simplified case with two neighbors and single relation data.

### 3.1.2.2 Neighboring Heuristic

The connectivity of the geographic neighbors is required to satisfy uniform spring distribution. Otherwise, the clustering of springs at some regions would result in stiff and loose areas in the system, which would falsify the effect of applied forces and therefore create misleading visualizations.

In order to distribute neighboring springs in different directions as evenly as possible, the node neighboring heuristic chooses the candidate which is closest to the current node while favoring larger angular distinction from previous neighbors. The heuristic of neighbor preference is defined by,

$$f(g_i) = \arg \min_{g_j} \left( \sum_k \left( \|g_{ik} - g_i\| \cdot \|g_j - g_i\| + C \right) \times \|g_j - g_i\| \right),$$

where  $g_i$  is the current node,  $g_j$  is the candidate node and  $g_{ik}$  is  $k^{\text{th}}$  previous neighbor of node  $g_i$ . The constant  $C$  weighs dot product between vectors  $(g_{ik} - g_i)$  and  $(g_j - g_i)$  and has a value in range  $(1, \infty)$ . Larger values of  $C$  favor angular distinction over shorter distance. For the current system case  $C$  is equal to 2.

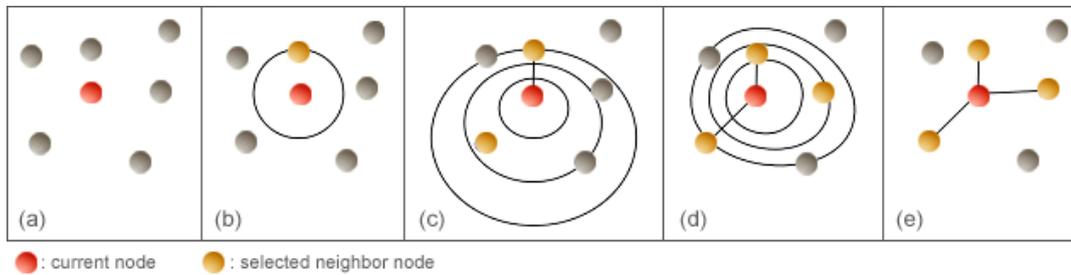


Figure 3.6 (a) 2D distribution of nodes (b) closest node to the current is chosen to be the first neighbor (c) second neighbor is chosen to be the node with lowest  $f(g_i)$  given first neighbor (d) second neighbor is chosen to be the node with lowest  $f(g_i)$  given first two neighbors (e) resulting neighbor selection.

The number of neighbors is an adjustable parameter and for the current applications we connect each node to three neighbor nodes. For a two dimensional case where  $C=2$ , the neighboring heuristic is illustrated in the Figure 3.6. Each curve indicates equal value of  $f(g_i)$ . Inner curves have lower  $f(g_i)$  values, thus in directions with higher angular distinction, the area spanned by a curve gets wider.

### 3.1.2.3 Integration

For each time interval, the spring solver algorithm seeks for a final configuration of nodal positions, where the total forces are balanced; hence summations of forces on each node are equal to infinitesimal values. Solution of system is reached iteratively using an implicit integration scheme and adaptive time steps. Implicit integration

schemes are stable and large steps can be taken in comparison to explicit methods [56][57].

In each iteration,  $\Delta X = \left( \frac{1}{h} I - F'(X_t) \right)^{-1} F(X_t)$  is evaluated to find a displacement vector  $\Delta X$ , where  $h$  is the step size and  $F(X_t)$  is the function of total forces on each node given the current positions. Evaluating this equation is computationally expensive compared to the computations in a single step of explicit methods; however, overall cost of reaching to an approximate solution is decreased by reducing total number of iterations with larger step sizes.  $F'(X_t)$ , is a function expressed by a matrix containing directional derivatives of the forces on each node. The derivative of the force on the node  $i$ , exerted from the spring between the node  $i$  and  $j$  is expressed as:

$$\frac{\partial f_i^j(x_j)}{\partial x_j} = K_{ij} \frac{\|x_j - x_i\| - l_{ij}^o}{\|x_j - x_i\|^3} I_3 + K_{ij} \frac{l_{ij}^o}{\|x_j - x_i\|^2} I_3 \left( \frac{(x_j - x_i)(x_j - x_i)^T}{\|x_j - x_i\|^2} \right),$$

where  $K_{ij}$  is the spring constant,  $I_3$  is the 3x3 identity matrix and  $l_{ij}^o$  is the initial length difference between nodes  $i$  and  $j$  [58]. Evaluated for each spring, the derived equation is inserted into proper location of  $3n \times 3n$  matrix expressing  $F'(X_t)$ .

Step size at each iteration, is adjusted by considering the error of force approximation made in previous iteration.  $F_{app.} = F(X_{prev.}) + F'(X_{prev.})\Delta X_{prev.}$ , is the force approximated in the previous step, and  $F_{curr.} = F(X_{curr.})$  is the current force, so the error ratio can be defined as:

$$F_{error} = \left| \frac{F_{app.} - F_{curr.}}{F_{curr.}} \right|.$$

Larger step sizes can be taken whenever error is below a predetermined threshold. At each iteration,  $F_{error}$  is compared with a predetermined threshold  $t_{error1}$ . If  $t_{error1}$  is not reached, the inner loop of the algorithm calculates the displacement using previous estimation. When  $F_{error}$  reaches  $t_{error1}$ , step size is adjusted accordingly in the outer loop of next iteration. This mechanism reduces the execution amount of matrix inversion routine, which is a very expensive task.

The simplified version of algorithm running in every simulation step is as follows:

```

FUNCTION Step_Solver()
IF (First Run) OR (Nodes Introduced or Deleted)
CALL AssignNeighbors()
Fcurr. = F(Xcurr.)
DELTA = Infinitesimal value;
WHILE ( |Fcurr.| > DELTA )

    F'curr. = F'(Xcurr.)
    Finverse = (1/h(I3) - F'(Xcurr.))-1
    ΔX = Finverse × F(Xcurr.)
    Xcurr. = Xcurr. + ΔX
    Fapp. = F(Xcurr.) + F'(Xcurr.)ΔX
    Fcurr. = F(Xcurr.)
    Ferror = |(Fapp. - Fcurr.) / Fcurr.|
    WHILE (Ferror < terror1)
        ΔX = Finverse × F(Xcurr.)
        Xcurr. = Xcurr. + ΔX
        Fapp. = F(Xcurr.) + F'(Xcurr.)ΔX
        Fcurr. = F(Xcurr.)
        Ferror = |(Fapp. - Fcurr.) / Fcurr.|
    END WHILE
    h /= (Ferror/terror2)
END WHILE
END FUNCTION

```

### 3.1.3 Surface Generation

Once the nodes are positioned, the surface covering these nodes is deformed to fit modified positions of the nodes. The surface on which a geographic map projected is expected to be smooth and fitting all nodes exactly. Initially built as a regular grid, the surface is deformed using a linear interpolation technique based on taking a weighted average of nodal positions within a specified region.

Starting from a high resolution grid creates a surface that fits to the graph nodes, but not smooth. Therefore, a multi resolution grid approach is introduced: Initially the surface is built as a low resolution (30x30) uniform grid. At this stage, each node is attached to a single grid node that is matched by closest distance relation. Positions of grid nodes that are matched with one or more graph nodes are modified by taking

average of corresponding network node(s). A discrete linear interpolation method is applied to approximate positions of grid nodes that are not modified in the previous step.

A higher resolution (120x120) grid is constructed out of the previous surface, and node matching process is performed again. As the resolution of the grid increases, number of graph nodes that are matched with the same grid node decreases. A final refinement to the grid is applied switching to a higher resolution (360x360) that eliminates matching of multiple network nodes to same grid node. The interpolation process in this step causes local variations on the surface, so the local fluctuations become visible.

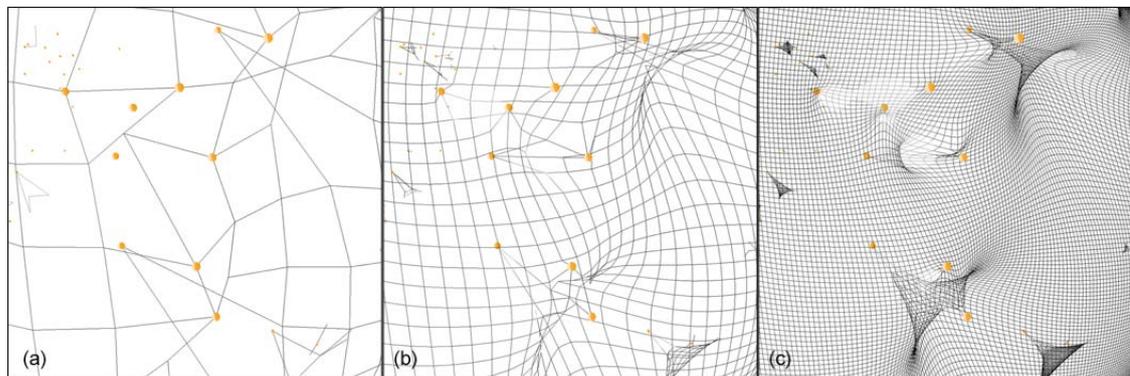


Figure 3.7 Multi resolution grid deformation: (a) Low resolution grid constructed with multiple network node attaching to a single grid node, (b) Mid resolution grid eliminates multiple attachments, (c) High resolution grid smoothly interpolates between even close nodes.

Figure 3.7 displays steps of the grid formation process. Another contribution of the surface generation method is the topology preservation. Matching of graph nodes with grid nodes are performed using initial coordinates of the graph nodes. Since the initial configuration guarantees the topology, this matching will stretch the surface without violating its continuity.

### 3.1.4 Data and Program Flow

Achieving the complex visualization method described above while enabling navigation at interactive speeds requires an elaborate software architecture. The data and program flow of the visualization system is illustrated in Figure 3.8.

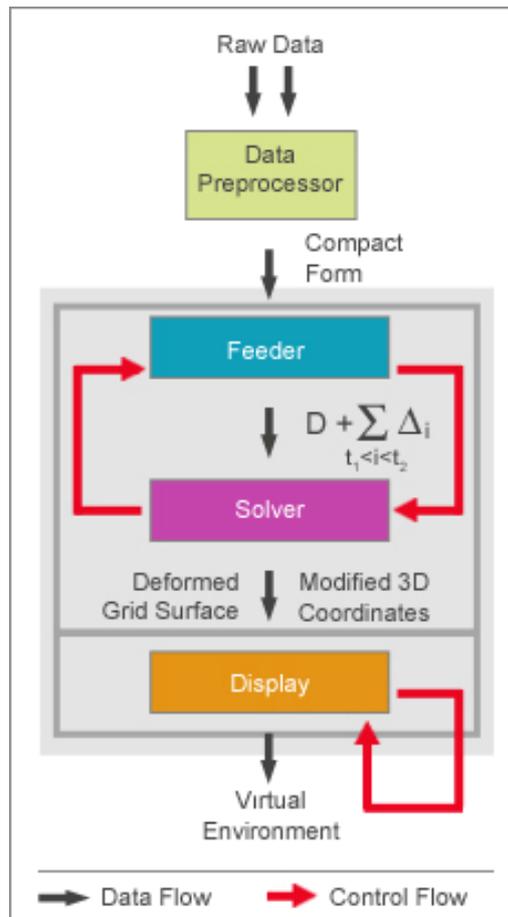


Figure 3.8 Data and program flow diagram for global context visualization. Separate control flow for the display module and for the data update enable the viewers to navigate at interactive speeds.

Any given spatial network data is pre-processed, sorted by time and put in the format specified in section 3.1.1. This offline stage does not introduce any calculations or aggregations.

The data feeder module is responsible for feeding data into spring solver. It loads the data into memory and outputs node positions and proximity measures matrices.

Components of proximity measure matrix can be reformed for a given time interval for visualization in different time resolutions.

The next module, spring solver, initiates springs between nodes and solves the spring-embedder system iteratively by using implicit integration with adaptive time steps. The output modified node positions matrix is then used to deform the grid surface.

The display module renders the deformed grid surface at interactive frame rates, enabling users to navigate through the virtual environment. The data and control flow diagram of the computation pipeline is displayed in the figure below.

### **3.1.5 Efficiency**

Graph optimization solutions based on spring-embedders are generally not suitable for real-time applications because of their high time-complexity. For a graph  $G=(V, E)$ , Kamada and Kawai [37] force-directed placement algorithm has  $O(V^2)$  complexity and Chalmers [65] provided an iterative MDS algorithm capable of producing a layout in time proportional to  $O(V^2)$ .

Our algorithm, on the other hand, has  $O(V+E)$  complexity since it does not calculate any forces between every pair of nodes, but only between connected nodes and geographic neighbors which are in constant number for every node. Therefore, in our method the complexity for graph optimization is significantly less than referred methods. However, surface generation phase of the method brings extra computation.

The proposed visualization system is tested on an Intel Dual Xeon 3.4 GHz with 4 GB of Memory. A data set consisting of 128 by 128 relations matrix with 25% load (approximately 4000 connections per time slice) is used for measurements. Results are shown in the graph in Figure 3.9. On the average, for each time slice, graph optimization with implicit integration takes 0,32 seconds and construction of grid takes 0,48 seconds. Therefore the deformed grid is updated in 0,8 seconds interval. Whereas

users can navigate in the virtual environment at interactive speeds around 15 fps. to 60 fps., depending on the grid resolution.

### 3.2 Thorough Data Investigation with Analytical Tools

The visualization environment provides a set of interaction tools which enable users to drill down the statistical details of the underlying data by manipulating the visualization while continuously providing visual feedback. These features of the visualization system expand its usefulness.

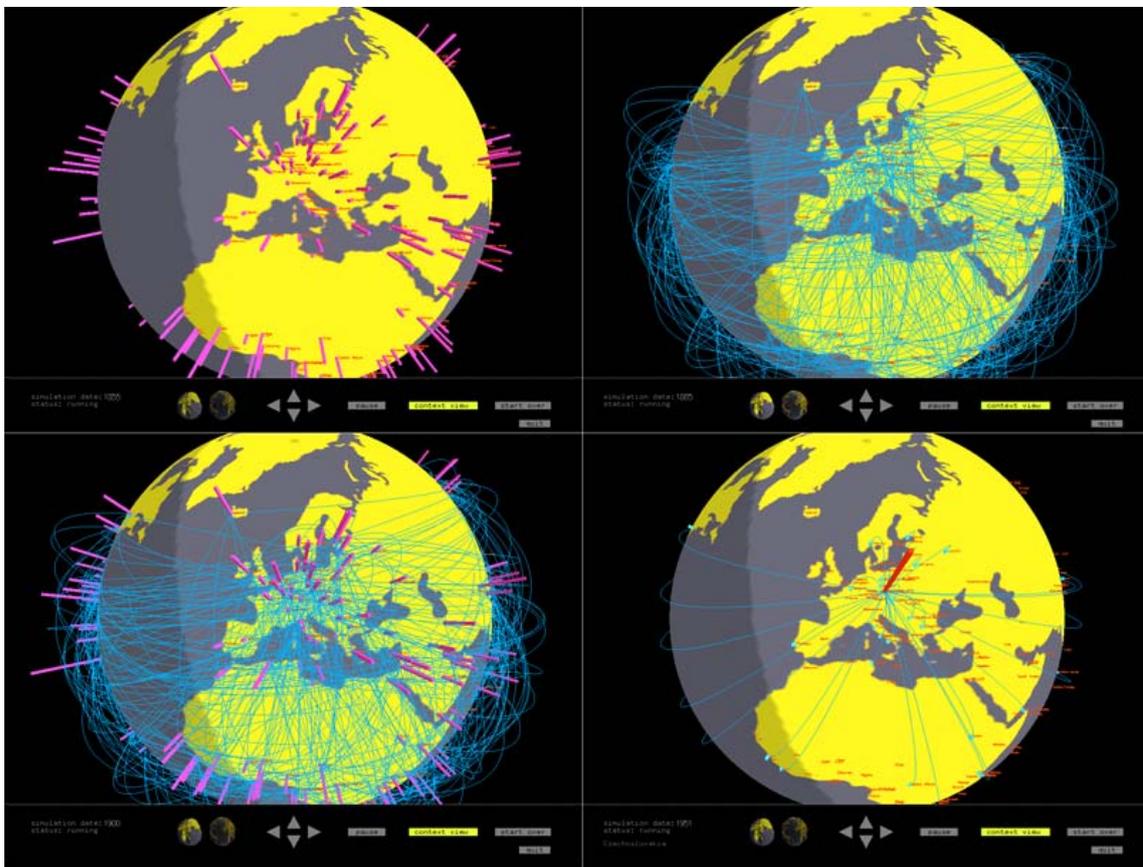


Figure 3.9 Analytical tools for thorough data analysis. First figure shows height bar animations, second figure shows arcs connecting nodes and bottom left figure displays both. When a node is selected as in bottom right figure, only arcs connected to that node are shown. Aggregated load of the node is shown as a height bar.

Any interested node can be identified with a mouse click, which causes the node to become active and turn a different color. Selecting a node also causes to display its name and its total relation capacity. The user is able to view all links originating from

that node by clicking a button provided in the menu. When this visualization mode is selected, the deformed map interpolates to its original spherical form. Otherwise deformed surface would clutter the display and occlude some of the nodes. Selecting any of the linked nodes aside the previously selected node displays their direct link statistics in the right hand corner of the screen.

Another button permits users to view the time series data as an animation of height bars over 3D map. When used in conjunction with the node identifying methods, the visualization only displays height bars for the selected nodes. Thickness of height bars is an adjustable parameter which can be manipulated with a slider.

Unusual patterns are color coded in height bars visualization mode. Whenever the total relation capacity of a node changes abruptly during the animation, the color of that node is highlighted.

## 4 CASE STUDIES

### 4.1 Flight Data

The proposed method is experimented using two different data sets. The first dataset is the records of domestic flights of United States between 1990 and 2004. The data consists of over 80 million flights with ten thousands of different routes among 231 airports. It is freely available at United States Bureau of Transportation Statistics [54].

The output visualization is a real-time, interactive, 3D virtual environment enabling viewers to compare between original geographic distances and perceived distances among these airports by frequent travelers. Modified geographic distances between airports can be thought as the ratio of accurate geographic distance to number of flights between these directions.



Figure 4.1 Snapshots from US domestic air flights data visualization. Image on the left is visualization of July 2001. Middle image displays smoothing of the surface on September 9<sup>th</sup> 2001. Right image is a closer view for July 2001.

At a glance, the convergence of East and West Coasts, especially around Los Angeles in the west and New York in the east, effortlessly conveys information about frequent flights among these locations. The higher curvature of West Coast compared to relatively moderate curvature of the East suggests less domestic flights in that region compared to the East.

Displaying the changes and anomalies over time in the data as a bending, twisting, and relaxing surface; this method utilizes motion as a significant component in the visualization. The effectiveness of motion is best realized when viewing anomalies in the data, such as the smoothing of the surface on 9/11. This abrupt change creates a visualization which aids the viewer to dig out a semantic meaning out of the abstract data such as the decrease in perceived distances. That is to say, it conveys the decrease in availability of places from other places that are originally getting closer in normal air traffic trends.

Users are able to interact with the visualization, navigate around the globe, freeze the animation, and select different time resolutions to view daily, weekly or monthly patterns. They are also able to select between wire-frame and opaque textured mode.

## 4.2 Diplomatic Exchange Data

The second data set, the diplomatic exchange data, is obtained from ICPSR database. [24] The data provides information for a total of 128 nations, for the period 1815-1970, the presence or absence of a diplomatic mission from every other member in the set at five year intervals. The visualization displays political affinities among geographic locations as a physical convergence on the morphing map.

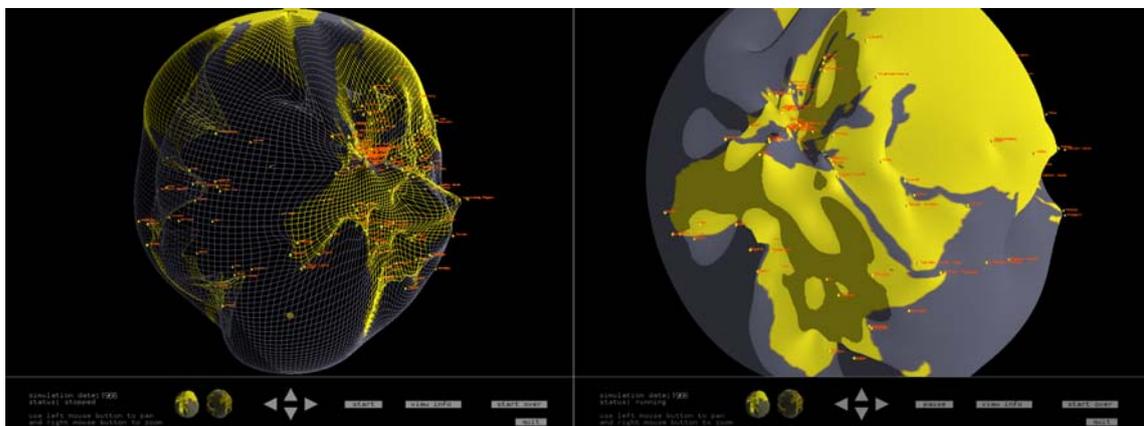


Figure 4.2 Visualization system using diplomatic exchange data is exhibited at TECHNE digital performance platform.

The visualization is exhibited using the diplomatic exchange data in TECHNE digital performance platform for over a week [53]. Users' positive reactions emanate from the visual appeal and intuitive recognition of the world map. Developed interactive navigation tools amplify engaging quality of the visualization. With the aid of brief description of the visualized data and the basics of the visualization provided in the help menu, users' were able to arrive at simple but valuable semantic conclusions about the nature of the data. The shrink of the surface around Europe, the immobility of unpopulated areas, the decrease of world volume with the increase of political activities are instantly captured qualities of the data. When examined more closely, it is possible to compare between political activities of different states. For instance, it is clearly visible that before 1920's, Canada had much close political relations with Europe, than US had.

## 5 ANALYSIS

### 5.1 Results

Visualizing spatial network data within the geographic context is a key factor in making such datasets more readable and meaningful for the viewer. *Global context visualization* permits viewers to arrive at semantically meaningful interpretations about general trends and anomalies in the data. However, there is a tension between displaying the data accurately and displaying it in a geographic context. *Global context visualization* provides a human-centered interface for spatial networks which displays data deliberately less accurate, for the sake of augmenting its comprehension by a diversity of viewers.

The advantage of the *global context visualization* over the raw data representation is obvious. Reading and comprehending thousands of numbers is an impossible, and more importantly a redundant task for viewers who are only interested in capturing the general trends and outliers in the data. The *global context visualization* provides a map animation that exhibits only the qualitative nature of the data by demonstrating its interesting and/or dominant features which constitute only a small portion of the underlying quantitative data.

Existing visualization techniques has major drawbacks in visualization of spatial networks. Multi dimensional data visualization techniques do not convey data within the geographic context. On the other hand, conventional map visualization methods maintain geographic context, but they are not suitable for visualizing high dimensional data. The proposed method integrates significant aspects of these methods for achieving a visualization that communicates general trends in the high dimensional data within the geographic context.

Real-time performing capability of the visualization system enables effective user interaction and navigation. Navigation provides an advanced evaluation environment for the viewers' where they can compare between different data objects more

effectively. Interaction with the visualization permits viewers to change its paradigms and switch from one visualization mode to another with seamless transitions. Hence, the visualization system is capable to communicate data at various levels of detail, while it constantly provides a qualitative feel for its global context.

Moreover, 3D and navigable nature of the visualization provides an immersive environment which exploits human sensory and cognitive systems at the highest level. In doing so, the visualization does not only depend on the visual cognition abilities of humans, but also utilize others like spatio-cognitive skills.

Compared to geographic network visualization methods discussed in section 2.6, the proposed visualization system offers an extra visualization mode which replaces abstract visual elements (e.g. nodes, arcs) by distortion of geographic space. Utilizing a physical phenomenon for representation, the visualization communicates abstract data more intuitively and effortlessly. That is to say, the proposed visualization method decreases abstraction level of the representation.

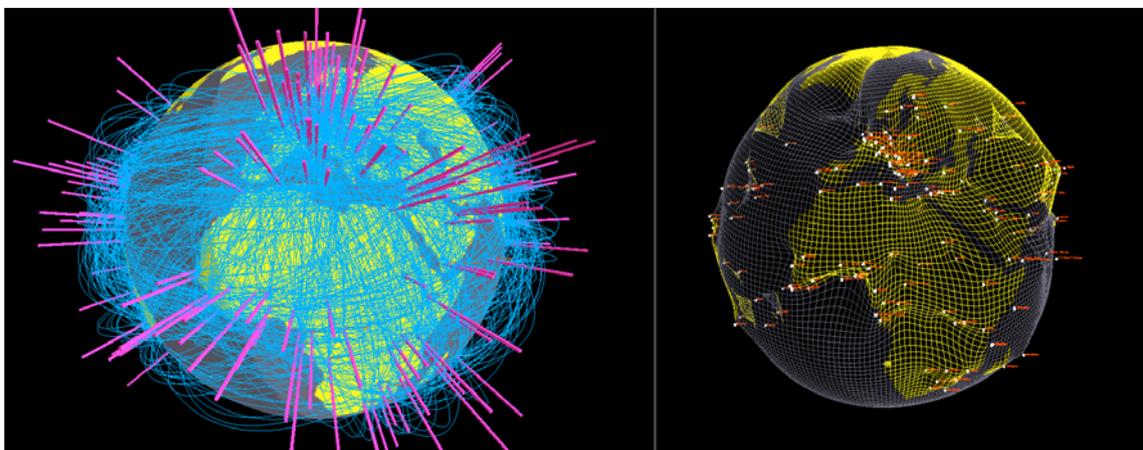


Figure 5.1 Same data visualized with height bars and arcs versus global context visualization.

Analytical tools for thorough data analysis are more abstract in nature and mostly rely on visual cognition abilities for perception of the data. However, utilizing those conventional tools is necessary to expand informative quality of the visualization. These tools engage interested viewers for longer time periods and enable them to dig out underlying quantitative data.

Although with height bars and arcs, it is possible to communicate statistical data as it is, the contribution of global context visualization is best realized when volume of the data visualized in one step gets extremely large. As in Figure 5.1, with arcs and height bars it is almost impossible to differentiate characteristics of a single network node in the midst of hundreds of line crossings, whereas in global context visualization viewers are able to evaluate load of network nodes by looking at their deformation amount.

The best strategy for utilizing proposed visualization system is to observe general trends in the global context visualization and utilize arcs and height bars to view characteristics of a specific node in detail whenever an interesting feature is detected.

## **5.2 Discussion and Future Work**

There are several issues requiring further study and elaboration. Improving the poor labeling of geographic locations in global context view will be a significant contribution. Cartographic labeling methods or focus + context views are possible candidates to develop a more readable and visually satisfying labeling method.

Projecting an opaque or even a semi-transparent map on the deformed surface occludes position of the nodes and decreases perception. For that reason wire-frame is preferred as the default display mode since the grid provides the necessary visual cue to the mental model of the map. To overcome occlusion problem and display geographic map simultaneously, interactive widgets are provided to switch between the opaque map and wire-frame mode.

Although the exhibited work is found to be captivating and informative by the viewers, evaluating its effectiveness requires an elaborate user study to reveal its strong and weak aspects more definitely. A collaborative study with cognitive scientists is targeted to examine perceptive quality and usability of the method in deep.

Increasing the variety of interaction modes and providing more analytical tools will obviously augment capabilities of this visualization system. However, the work presented in this thesis addresses public viewers rather than analysts or scientists. Therefore, the focus of the work will remain to provide the meaningful big picture of the data to a diversity of viewers. Complex and less intuitive visualizations addressing expert viewers are out of scope of this thesis.

In near future, our goal is to create an installation that works with live data feed, i.e. transportation or telecommunication data, which benefits public viewers to understand global context of the data instantly.

## REFERENCES

- [1] S. K. Card, J. D. Mackinlay and B. Shneiderman, "Readings in Information Visualization: Using Vision to Think", *Morgan Kaufmann*, CA, 1999.
- [2] C. Chen, "Information Visualization: Beyond The Horizon", *Springer - Verlag*, London, 2004.
- [3] E. R. Tufte, "Envisioning Information", *Graphics Press*, Connecticut, 1995.
- [4] J. Bertin, "Graphics and Graphic Information Processing", *de Gruyter*, 1981. Translated by W. J. Berg and P. Scott.
- [5] J. Bertin, "Semiology of Graphics: Diagrams, Networks, Maps", *University of Wisconsin Press*, 1983.
- [6] W. S. Cleveland and R. McGill, "Graphical Perception: Theory, experimentation and application to the development of graphical methods", *Journal of American Statistics Association*, vol. 79, no. 387, pp. 531-554, Sept. 1984.
- [7] J. Mackinlay, "Automating the design of graphical presentations of relational information", *ACM Transactions on Graphics*, vol. 5, no. 2, pp. 110-141, 1986.
- [8] A. Insalberg, "Multidimensional Detective", *Proceedings, IEEE Symposium on Information Visualization (InfoVis '97)*, p. 100, 1997.
- [9] K. M. Fairchild, S. E. Poltrock and G. W. Furnas, "SemNet: Three-dimensional graphic representation of large knowledge bases", *In Cognitive Science and its Application for Human-Computer Interface*, *Erlbaum, Hillsdale, NJ.*, pp. 201-233, 1988.
- [10] B. Shneiderman, "Dynamic Queries for Visual Information Seeking", *IEEE Software*, vol. 11, no. 6, pp. 70 - 77, 1994.
- [11] C. Ahlberg and B. Shneiderman, "Visual Information Seeking: Tight Coupling of Dynamic Query Filters with Starfield Displays", *Proceedings CHI'94, Human Factors in Computing System*, 1994.
- [12] S. Eick, "Data visualization sliders," *Proceedings, UIST'94*, pp. 119-120, ACM, New York, 1994.
- [13] M. C. Chuah, S. F. Roth, J. Mattis and J. Kolojejchick, "SDM: Selective Dynamic Manipulation of Visualizations", *ACM Symposium on User Interface Software and Technology*, 1995.

- [14] R. A. Becker, S.G. Eick, and G.J. Wills, "Visualizing network data," *IEEE Trans. on Visualizations and Graphics*, vol. 1, pp. 16-28, 1995.
- [15] K. C. Cox, S. G. Eick and T. He, "3D Geographic Network Displays", *SIGMOD Record*, vol. 25, no. 4, pp. 50-54, 1996.
- [16] S. G. Eick, "Aspects of Network Visualization", *IEEE Computer Graphics and Applications*, vol. 16, no. 2, pp. 69 - 72, 1996.
- [17] E. E. Koutsofios, S. C. North and D. A. Keim, "Visualization Blackboard", *IEEE Computer Graphics and Applications*, pp. 16-19, Mai/June 1999.
- [18] S. K. Lodha and A. Verma, "Spatio-temporal Visualization of Urban Crimes on a GIS Grid," *Proceedings, 8th ACM International Symposium on Advances in Geographic Information Systems, ACM Press, NY*, pp. 174-179, 2000.
- [19] I. Lokuge, S.A. Gilbert and W. Richards, "Visualizing Information Spaces: A Tour Through Boston," *European Conference on Visual Perception '95, Ttbingen. Abstract in Perception*, 24 Supplement, p. 76, 1995.
- [20] A. M. MacEachren, "How Maps Work: Presentation, Visualization, and Design", *The Guilford Press, NY*, 1995.
- [21] D. F. Reilly and K. M. Inkpen, "Map Morphing: Making Sense of Incongruent Maps", *Proceedings, Graphics Interface 2004, in the Canadian Human-Computer Communications Society, London, Ontario*, 2004.
- [22] A. Zanella, S. Carpendale and M. Rounding, "On the effects of viewing cues in comprehending distortions", *Proceedings of NordCHI'02.* , pp. 119 - 128, Aarhus, Denmark, 2002.
- [23] A. Skupin, "From Metaphor to Method: Cartographic Perspectives on Information Visualization", *Proceedings, InfoVis 2000, Salt Lake City*, 2000.
- [24] A. M. MacEachren, R. Edsall, D. Haug, R. Baxter, G. Otto, R. Masters, S. Fuhrman and L. Qian, "Virtual Environments for Geographic Visualization: Potential and Challenges." *Proceedings, Workshop on New Paradigms in Information Visualization and Manipulation in Conjunction with the Eighth ACM International Conference on Information and Knowledge Management*, 1999
- [25] A. M. MacEachren, F. P. Boscoe, D. Haug, and L. W. Pickle, "Geographic Visualization: Designing Manipulable Maps for Exploring Temporally Varying Georeferenced Statistics," *Proceedings, Information Visualization '98, Reliegh-Durham, NC*, Oct. 19-20, 1998.

- [26] W. R. Tobler, "Thirty Five Years of Computer Cartograms," *Annals, Assoc. American Geographers*, vol. 94, pp. 58 - 73, 2004.
- [27] W. R. Tobler, "Pseudo-Cartograms", *The American Cartographer*, vol. 13, no. 1, pp. 43-40, 1986.
- [28] D. Dorling, "Area Cartograms: Their Use and Creation", *first ed. Dept. of Geography, Univ. of Bristol*, England, 1996.
- [29] C. Cauvin, C. Schneider, and G. Cherrier, "Cartographic Transformations and the Piezopleth Method", *The Cartographic Journal*, vol. 26, no. 2, pp. 96-104, Dec. 1989.
- [30] C. J. Kocmoud and D. H. House, "Continuous Cartogram Construction", *Proceedings, IEEE Visualization, IEEE CS Press*, pp. 197-204, 1998.
- [31] M. T. Gastner and M. E. J. Newman, "Diffusion-Based Method for Producing Density Equalizing Maps", *Proceedings, National Academy of Sciences of USA.*, vol. 101, no. 20 pp. 7499-7504, 2004.
- [32] D. A. Keim, C. Panse and S. C. North, "Medial-Axis-Based Cartograms", *IEEE Computer Graphics and Applications*, vol. 25, no. 3, pp. 60 - 68, 2005.
- [33] B. Fry, "Organic Information Design", *Master's Thesis, MIT Media Lab*, Boston, MA, 2000.
- [34] G. Di Battista, P.D. Eades, R. Tamassia and I.G. Tollis, "Graph Drawing: Algorithms for the Visualization of Graphs", *Prentice-Hall*, Englewood Cliffs, NJ, 1999.
- [35] T. Fruchterman and E. Reingold, "Graph drawing by force-directed placement", *Software - Practice and Experience*, vol. 21, pp. 1129-1164, 1991.
- [36] P.D. Eades, "A heuristic for graph drawing", *Congr. Numer.* vol. 42, pp.149-160, 1984.
- [37] T. Kamada and S. Kawai, "An Algorithm for Drawing General Undirected Graphs", *Inform. Process. Lett.* vol. 31, pp. 7-15, 1989.
- [38] I. Herman, G. Melancon and M. S. Marshall, "Graph visualization and navigation in information visualization: A survey", *IEEE Transactions on Visualization and Computer Graphics*, vol. 6, pp. 24-43, 2000.
- [39] J. D. Cohen, "Drawing Graphs to Convey Proximity: An Incremental Arrangement Method", *ACM Transactions On Computer-Human Interaction*, vol. 4, no. 3, pp. 197-229, 1997.

- [40] A. Morrison, G. Ross and M. Chalmers, "Fast Multidimensional Scaling through Sampling, Springs and Interpolation", *Information Visualization*, vol. 2, no. 1, pp. 68-77, 2003.
- [41] W. Muller and H. Schumann, "Visualization Methods for Time Dependant Data: an Overview", *Proceedings, Winter Simulation Conference*, 2003.
- [42] J. Yang, M. Ward, E. Rundensteiner, and S. Huang, "Visual Hierarchical Dimension Reduction for Exploration of High Dimensional Datasets", *Proceedings, Joint IEEE TCVG-EG Symposium on Visualization*, vol. 4, pp. 19-28, 2003.
- [43] D. Asimov, "The grand tour: a tool for viewing multidimensional data", *SIAM Journal on Scientific and Statistical Computing*, vol. 6, no. 1, pp. 128-143, 1985.
- [44] D. Cook, A. Buja, J. Cabrera and C. Hurley, "Grand Tour and Projection Pursuit Journal of Computational and Graphical Statistics", vol. 4, no. 3, pp. 155 -172, 1995.
- [45] M. S. T. Carpendale, D. J. Cowperthwaite and F. D.Fracchia, "Extending Distortion Viewing from 2D to 3D", *IEEE Computer Graphics & Applications*, vol. 17, no. 4, pp. 42-51, 1997.
- [46] J. Lamping and R. Rao, "The Hyperbolic Browser: A Focus+Context Technique for Visualizing Large Hierachies" *Journal of Visual Languages and Computing*, vol. 6, no. 4, 1995.
- [47] T. Munzner, "Exploring Large Graphs in 3D Hyperbolic Space", *IEEE Computer Graphics and Applications*, vol. 18, no. 4, pp. 18-23, July/Aug. 1998.
- [48] G. W. Furnas, "The FISHEYE View: A New Look at Structured Files", *Bell Laboratories Technical Memorandum*, 1981.
- [49] C. Chatfield and A. J. Collins, *Introduction to Multivariate Analysis*. Chapman & Hall, London, 1980.
- [50] J. B. Kruskal, "Non-metric multidimensional scaling: A numerical method", *Psychometrika*, vol. 29, pp. 115-129, 1964.
- [51] R. B. Heady, "PERMAP: Perceptual Mapping Software", retrieved at July 2006, from: <http://www.ucs.louisiana.edu/~rbh8900/>
- [52] SmartMoney.com, retrieved at July 2006, from: <http://www.smartmoney.com>
- [53] TECHNE, International Istanbul Digital Perfomance Platform, retrieved at June 2006, from: <http://www.techneplatform.org/about.html>
- [54] Bureau of Transportation Statistics, retrieved at June 2005, from: <http://www.bts.gov>

- [55] ICPSR, Inter-University Consortium for Political and Social Research, retrieved at January 2006 from: <http://www.icpsr.umich.edu>
- [56] D. Baraff and A. Witkin, "Large Steps In Cloth Simulation. Computer Graphics", *Proceedings, SIGGRAPH*, 1998.
- [57] M. Desbrun , P. Schröder and A. Barr, "Interactive animation of structured deformable objects", *Proceedings, Conference on Graphics Interface '99*, pp.1-8, Ontario, Canada, 1999.
- [58] D. Baraff and A. Witkin, "Physically-Based Modeling", *ACM SIGGRAPH Course Notes*, 2003.
- [59] E. Fanea, S. Carpendale, T. Isenberg, "An interactive 3D integration of parallel coordinates and star glyphs", *IEEE Symposium on Information Visualization*, 2005. INFOVIS 2005, pp. 149- 156, 2005.
- [60] L. Yang, "Interactive exploration of very large relational datasets through 3D dynamic projections", *Proceedings, 6th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pp. 236-243, MA, 2000.
- [61] W. S. Torgerson, "Multidimensional scaling: 1. Theory and method", *Psychometrika*, vol. 17, pp. 401-419, 1952.
- [62] R. N. Shepard, "Analysis of proximities: multidimensional scaling with an unknown distance function I & II", *Psychometrika*, vol.27, pp.125-139, 219-246, 1962.
- [63] T. Munzner, "Visualizing the Global Topology of the mBone", *Proceedings, IEEE Symposium on Information Visualization '96*, pp. 85-92, Oct. 28-29 1996, San Francisco, CA, 1996.
- [64] K. M. Fairchild, S. E. Poltrock and G. W. Furnas, "SemNet: Three Dimensional Representations of Large Knowledge Bases", in R. Guindon (ed.), *Cognitive Science and Its Applications for Human-Computer Interaction*, Lawrence Erlbaum Associates, Hillsdale, NJ, pp. 201-233, 1988.
- [65] M. Chalmers, "A Linear Iteration Time Layout Algorithm for Visualizing High-Dimensional Data", *Proceedings, IEEE Visualization '96*, pp.127-132, 1996.