8-2000

Towards A Spatial Model of Rurality

Gillian AvRuskin

TOWARDS A SPATIAL MODEL OF RURALITY

By

Gillian AvRuskin

B.A. McGill University, Canada, 1996

A THESIS

Submitted in Partial Fulfillment of the

Requirements for the Degree of

Master of Science

(in Spatial Information Science and Engineering)

The Graduate School

The University of Maine

August, 2000

Advisory Committee:

M. Kate Beard, Associate Professor of Spatial Information Science and Engineering, Thesis Advisor

Max J. Egenhofer, Professor of Spatial Information Science and Engineering

Deirdre M. Mageean, Associate Professor of Resource Economics and Policy

TOWARDS A SPATIAL MODEL OF RURALITY

By Gillian AvRuskin

Thesis Advisor: Dr. M. Kate Beard

An Abstract of the Thesis Presented
in Partial Fulfillment of the Requirements for the
Degree of Master of Science
(in Spatial Information Science and Engineering)
August, 2000

The term rural is used to describe people, places, traditions, and spaces. It is often employed as a setting for study as well as an object of study. People's perceptions of rural are confused and differ considerably. For over a century researchers have attempted to define more precisely this term using social, economic, and or ecological components. However, problems of interpreting official definitions and measurements exist. These definitions require extensions in order to capture a more objective meaning of the word.

This thesis presents the foundations of a new approach to measuring and defining rurality. A spatial based approach is taken in which explicitly spatial data instead of social or economic data are collected and indexed. The index is divided into two clusters, a connectivity cluster and an access-to-service cluster. The indicators in the clusters are chosen based on a list of criteria taken from the Institute for International Development. The model employs mathematical foundations of both topology and metrics. The use of fuzzy measures to determine a degree of rurality, instead of classical set theory, enhances the model. A degree of connectivity, a degree of accessibility, and an overall degree of rurality is determined. The model also incorporates scale. The granularity of an indicator depends on a user-required level of detail. The data are manipulated and analyzed in a

GIS. The spatial index is tested on a number of towns throughout Maine. A graphical user interface illustrates the results in an easy to understand format.

The results of this thesis show that a spatial approach to defining rural extends formal definitions to capture a different facet of rurality, a degree of rurality. Furthermore, spatial, temporal and attribute queries are possible enabling users a choice given a particular task.

Acknowledgments

I am grateful to all of my thesis committee members, Dr. Kate Beard, Dr. Deirdre Mageean, and Dr. Max Egenhofer for their support and guidance. I would especially like to acknowledge my advisor Dr. Kate Beard for her encouragement from before the beginning and throughout the thesis writing process.

A special thank you to my friends in the department and the international and not so international crew from other departments for their encouragement, guidance, and ability to talk about anything at anytime. I would also like to acknowledge the staff at Eastern Maine Development Corporation for their words of wisdom.

Finally, I am especially grateful to my family for their love and support at all times.

Table of Contents

List of Tables

List of Figures

Chapter 1

Introduction

Many people use the term rural, and many different concepts are associated with it. It can be both a noun and an adjective and bring to focus pictures of the countryside or of the people living there. Rural describes people, places, things, traditions, and spaces. Sociologists, economists, geographers, government bodies, and laypersons all employ the word. It is often used as if it refers to some important, unique and singular phenomenon, or it may imply a setting for study rather than an object of study. The Encyclopedia Britannica defines rural as “of or relating to the country, country people or life, or agriculture”(Encyclopedia Britannica, 2000). Nevertheless the term is confusing, and it is difficult to know its precise meaning (how much it relates to the country), a particular connotation and context dependencies.

Various scenarios exist in which it is important to define rural. Suppose the town manager of Town A is wondering if Town A has become more or less rural over time, or suppose a development district is wondering how much rural area in its region has been lost (or gained) over the past five years. Therefore, a definition of rural must have a geographic extent and be able to capture temporal change. Another scenario may arise when Town A is compared with Town B to determine which town is more rural or Region A is compared with Region B. This scenario raises two additional issues. One is that a definition of rural must address scale effects. As the size of the geographic areas under consideration increases, is the definition still appropriate and does it still produce consistent results? The second issue is the ability to compare one town or region with another. The definition should be sensitive to regional differences such that it is possible to compare a region in the Midwest with one in the Northeast. Financial aid is awarded to areas based on rural or urban classifications. Having one definition of rural, used nationwide, would clarify any ambiguous definitions fabricated for the purpose of winning funds. It is clear that such scenarios necessitate a scalable and temporally sensitive definition of rural. However, it is also important to remember that a prototypical definition of rurality is difficult to establish. Keeping this in mind, the model used in this thesis is adaptable. The result is multiple valid measurements of rurality that are of different complexity and detail and can be tailored to user needs.

For almost a century now researchers have tried to define the term rural more precisely. Most academics agree that rural is used to designate characteristics of physical areas or attributes of individual persons, and to refer to at least three different substantive aspects. These aspects include an ecological facet, an occupational dimension and a socio-cultural component (Willits and Bealer, 1967). However, they have also discovered that no single aspect captures the meaning of the word precisely. A composite definition is needed.

Initially Galpin (1915) sought to define the meaning of rural by studying town and country relationships (Wilkinson, 1991). It was proven that his discrete definition did not suffice and it was necessary to derive a composite approach. One fairly versatile approach to a composite definition is based on the assumption that people and places differ from one another in degree of rurality (Willits and Bealer 1967). This notion expresses the concept of an urban-rural continuum. There have been several approaches to defining an urban –rural continuum (Pahl 1966, Butler 1990), however they have also not proven satisfactory thus far. This paper focuses on a new approach to the definition, involving a more explicit spatial component. Previous definitions (Willits and Bealer 1967, Pahl 1966) have been aspatial. Reasons for this may include the fact that sociologists are mostly concerned with social and economic data, not explicitly spatial issues; technology has only recently been able to handle this type of spatial data effectively and efficiently; and previous definitions are taken for granted. Another factor encouraging a spatial approach is that a spatial component instead of a human survey proves to be an easier task. As has been demonstrated by the Bureau of the Census, it is time and cost consuming to track down people not filling out information on their census forms (Bureau of Census, 1999), while spatial data are easier to capture because they are stationary. Even though most people accept the Census definition of rural the definition itself changes every five years depending on other statistical definitions. With the vague and confusing definitions that already exist it is clear that a new dimension is needed to elucidate the meaning of the word rural. By using GIS technology and readily available data, spatial aspects of rurality can be defined and quantified. The result is a more robust definition spatially, statistically, and over time. Our proposed definition captures:

- Time dependencies: by examining how spatial attributes of an area change over time
- Scale dependencies: by developing a hierarchical weighting system based on spatial attributes
- Socio economic dependencies: by first examining spatial components.

By studying spatial components of place we can capture ecological, occupational, and socio-cultural factors. Instead of first defining a threshold in the attribute, such as the

Census definition which uses a minimum population of 2500, we investigate the basic geographical concept of place (Curry, 1996). For instance, we choose to study Town A. By investigating spatial components of this area, we determine its rurality. We know that Town A has a demographic characteristic. If we know Town A is rural and we know its demographic characteristic, we have captured the demographic characteristics of the place independently of its designation. In comparing the rurality of towns we can thus describe their demographic and economic trends. Applying a spatial dimension that can capture the ecological, occupational, and socio-cultural dimensions is more easily accomplished with recent geographic information systems (GIS) technology.

The use of GIS in rural sociology has had few applications. Luloof and Befort (1989) introduced it in an article that promised new aerial photography and mapping techniques would enable rural sociologists to integrate and analyze spatial data. While the potential of GIS to be used in rural sociology has been suggested in a number of places, few applications of the technology have been reported in journals and other academic outlets (Bradshaw and Muller, 1998). Rural sociologists have much to gain from involving GIS analysis in their research. Although GIS's were initially created for natural resource and infrastructure planning, in the past decade, GIS has become an in-office essential throughout America. GIS technology is employed by major industries to facilitate economic development, to enhance environmental management, and to model climate change. But smaller agencies have also benefited from GIS. Cities utilize GIS for tax assessment, zoning policies and transportation design (Jacob and Luloff, 1995). They are powerful systems capable of managing, integrating, manipulating, analyzing, and storing spatial data in all fields of academia. Engineers, developers, economists, and sociologists can equally benefit from a GIS. As a result of increased user demands and decreased costs they are becoming more accessible to these fields.

Up until now, sociologists and others have relied on qualitative descriptive methods and informal quantitative methods to collect and analyze data. Their research involved numerous questionnaires to be filled out by people of the community. This is a qualitative and subjective approach. These questionnaires were not uniform, where in one area some questions were asked while these questions were left out in other areas (BEA, 1941). However, within a GIS this information can easily and quickly be tabulated, simulated, and analyzed. The techniques available in a GIS can be employed to produce and evaluate the significance of a spatially infused definition of rurality.

1.1 Hypothesis and Research Questions

A new definition of rural can be established with the help of a GIS. A GIS supports a more statistically and spatially robust result by relying on computer-driven calculations and analysis. These results are easily replicable using a GIS and are more exact than a sociologist can be in the field. Eliminating the social and economic components that heavily rely on human subjects and adding a more objective spatial component to traditional methods will produce a different measurement of rurality. Our hypothesis is:

A definition of rural based on spatial relations of connectivity and accessibility is consistent with previous definitions of rural, but is also operational over time, practical over varying scales, and effective in comparing different regions.

This thesis attempts to answer questions such as:

- What constitutes a spatial index?
- What is the benefit of a spatial index of rurality?
- Does this definition capture traditional components of rurality, such as rural demographics?
- Are time dependencies represented in this definition?
- Are scale effects represented in this definition?

The goal of this thesis is to develop a different approach to measure rurality and test its ability while addressing these questions. Our final model is termed the SRI, the Spatial Rurality Index. The objective of the SRI is to distinguish rural from urban but not to define prototypical rural. Therefore, this model can capture rurality over such spatial scales as the Midwest or New England. However, to extend the SRI globally with places such as sub-Saharan Africa and China is beyond the work of this thesis. The SRI also allows one linguistic term to represent multiple conceptualizations of rural. Again, a prototypical rural is not defined however different representations of what that term characterizes is determined.

1.2 Approach

In order to generate a spatial definition of rural, a review of previous approaches to defining rural is essential. The goal of Chapter 2 is to introduce earlier implications of the word. A brief historical overview of these definitions and measurements is presented. Shortcomings of these approaches are outlined. From the strengths and weaknesses of the analysis, the need for a different approach becomes apparent. The need for a new method of measuring and defining rural is documented.

Chapter 3 introduces the importance of developing a composite index for a definition of rural. Significant characteristics of a spatial index such as indicators and clusters are established. Two clusters are deemed necessary in this investigation. The first defines connectivity and is based on link and node topology and node degree. The second cluster defines access measured in terms of distance. By defining rural in terms of spatial dimensions, demographic and economic dimensions can be evaluated independently as characteristics of a "rural area." We argue that this definition is spatial, because the spatial relations of connectivity and distance are central to its development. Overall, this chapter identifies and explains the variables for the new model of rurality.

The fourth chapter delineates the framework for the SRI. It outlines issues relating to spatial data models. Once the data have been analyzed and categorized into the measures proposed in Chapter 3, the variables are weighted and treated according to the model's framework. The concept of "fuzzy measures" (Zadeh 1965) is reviewed and discussed as it applies to the model. By using fuzzy measures, which are mathematical techniques used to represent vagueness in everyday life, we can further investigate a degree of rurality, instead of a crisp Boolean rural category that prevails in present definitions. Finally issues of scale and granularity are introduced and handled as they pertain to the model.

A visual representation of the model is demonstrated in Chapter 5. The need for direct manipulation and other interface techniques is explained. Finally, the model is shown in three different scenarios. Maine has been described as "tranquil," "beautiful," "awe inspiring," "natural," and "peaceful (www.mainetourism.com)." Yet its proximity to Boston invokes images of traffic, crime, and overpopulation. From these images

Maine and its counties and towns provide an ideal test for this new model of measuring rurality.

Chapter 6 summarizes the findings of this study. As is the case in scientific work, further analysis is always beneficial. The model can be adjusted and different variables added. In analyzing over ten towns in Maine we conclude that for a start, this approach is satisfying and an improvement on other models.

1.3 Outcome

What does this thesis hope to accomplish? The final outcomes include:

1. *A new definition of rurality based on spatial relations*. Especially after considering previous attempts to define the word, this approach will not be limited to traditional methods. It captures previous components (demographic and economic) by examining spatial ones.

2. *A definition that is independent of demographic characteristics.* The indicators included in this definition are not socio economic but spatial.

3. *Development of a model that supports expressions of degrees of rurality*. This model employs GIS methods of analysis and informative displays. Fuzzy categories are discussed and applied to the concept of an urban-rural continuum.

Chapter 2

Summary of Perspectives

The definition of rural is an elusive one, involving concepts that have emerged from the fields of geography, economics, and sociology. When "rural" is spoken of, it can take on numerous meanings including one of place, one of people, or one of lifestyle to name a few. Many authors (Christaller 1935, Pahl 1966, Willits and Bealer 1967, Gilg 1985) agree that most understandings of the word involve the use of ecological, occupational, or cultural dimensions (Gilg 1985). However, all approaches suffer from major drawbacks when used independently. This chapter examines historical definitions and measurements of rural and discusses their limitations. The chapter explores the historical underpinnings of the word and explains how the word has evolved theoretically from Christaller's (1933) Central Place Theory to its economic resurgence in the 1970s. In effect, this chapter delivers an historical and conceptual framework of rurality. The conclusion is that a new dimension is needed and now possible to more adequately describe the idea of rural. No single theory can successfully express how rural a place is and that a new method must be applied, one that has not been given considerable attention so far.

2.1 Stage 1 (1915s – 1930s): The Historical Perspective

As early as 1915 academics felt it necessary to define what was meant by rural. Charles Galpin's (1915) study of town and country, the first of a series of studies that took place in Walworth County, Wisconsin, USA included important questions: (1) Is there such a thing as a rural community and (2) if so, what are its characteristics? Clearly, in this study Galpin sought both a definition of a rural community and a way of measuring it, by exploring specific attributes. He concluded that rural areas in the town and country are not separate communities. Rather they work jointly to provide service zones. He noted that the village and farm residents formed a “rurban” community. Detailed analyses of the different service zones show the communities of the country to be composite ones, possessing characteristics of real life (Wilkinson 1991). Thus a fine line exists between town and country that cannot be exactly defined, and in many cases the two cannot be separated.

Galpin’s study initiated a large number of other descriptive studies throughout the 1920s and 1930s by social scientists. Kolb (1923, 1925) studied service relationships between farm and village residents. Brunner (1927) plotted rural urban boundaries, and Loomis and Beegle (1957) tested other methods using empirical studies. These articles continued Galpin’s work and focused on rural-urban relationships.

However, these studies all tended towards determining geographic boundaries of rural-urban communities. They did not examine components of an urban or rural community. Lowry Nelson’s (1935) analysis of the Mormon village in the 1930s expanded somewhat on this topic. He studied the effects of the interplay of cultural and ecological factors in community formation and change.

Yet again, these studies did not explicitly provide a framework for determining what is rural or what is urban. Instead the terms rural and urban provided a setting for study, rather than a subject of study.

2.2 Stage 2 1933: The Economic Perspective

In the early 1930s Christaller (1933) proposed a theory aimed to anticipate the interdependencies between town and country the central place theory. Even though his theory does not catalogue components of rurality, it provides a framework useful in interpreting settlement patterns, in explaining the decline of small villages, in planning the location of new settlements, and in analyzing the social structures of rural communities. It is most useful in this thesis, as it is a foundation for the spatial measures proposed later. Distance from central place is one of the spatial measures used in the Access to Services Cluster (Section 3.5). Without this brief introduction into the theory behind central place, a variable of such importance might be misunderstood; therefore, it requires an explanation at this time.

In economic terms, the urban place provides the market center for the farmers (King 1984). Thus a functional interdependence between a town and the surrounding rural area exists. This is the foundation of Christaller's central place theory. Although, this idea was not original, Christaller proposed a completely new framework for the study of settlement geography. His major task was to define a central place with its central goods and services and explain its mutual dependence on the country.

In order to focus on the economic interrelationships between central places and rural areas it was critical to make some assumptions. First he assumed that there was a homogenous plain with soil fertility and other natural resources being the same in all parts. Furthermore, settlers had equal levels of income and the same demand for goods and services. Travel was equally possible and transporting goods was a function only of the distance traveled (King 1984). Finally, Christaller assumed that the farmers as consumers and the businesspersons in the urban places as the producers of goods and services were rational individuals who would seek to maximize profit and minimize costs.

With these assumptions, a theory of central place was proposed. This theory is based on the notion of a *range*. A range is divided into two parts, an *upper limit* and a *lower limit*. The upper limit is the farthest distance the dispersed population is willing to go in order to buy a good offered at a place a central place. The more expensive the good the greater the willingness to travel longer distances and the larger the upper limit. The more frequently demanded the goods or the less expensive the goods, the smaller the upper limit range. The lower limit is determined by a threshold value. This is the measure of the minimum level of demand needed to ensure that the offering of a good or service will be profitable. The key concept of Christaller's theory is the upper limit to the range. There are two terms associated with the upper limit, the ideal and the real. The ideal upper limit is the maximum distance over which a good is demanded; but in the case where there is another central place nearby that offers the same good, there is a point at which it becomes cheaper for the purchaser to go to this other center. That point defines the real range of a good. Figure 2.1 illustrates this concept more clearly, where *a* represents the ideal upper limit, while *b* represents the real upper limit. People, therefore, will most likely travel from A to B providing the central place of B is closer and is offering the same good.

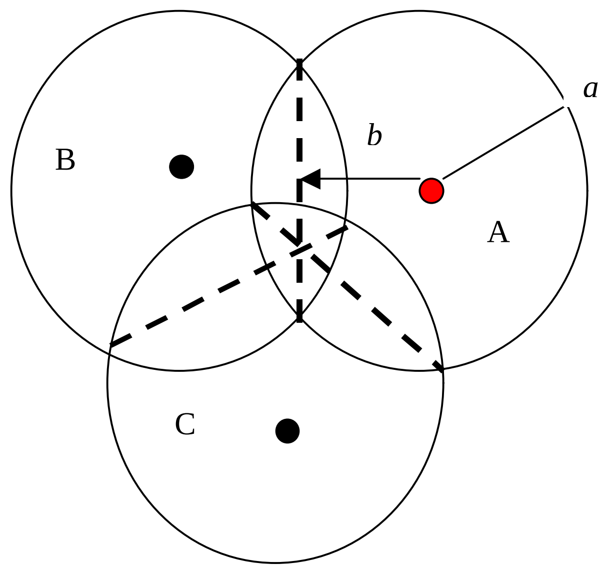

Figure 2.1: Ideal and Real Ranges of a Central Place Function (King 1984).

The spatial result is that centers of equivalent service structures will be located at equal distances from each other. Centers supplying higher order services will be located further from each other than centers having only lower order services, because they are fewer in number and need larger trade areas. The smaller centers are distributed as satellites around larger centers, in a hexagonal shape. The end result is a hierarchy of urban places differentiated not only by their size but also by the number and order of the functions offered by them (King 1984). In this economic perspective, Christaller is actually defining spatial organizations of “urban” places based on services and goods offered in that place. In this manner, he is also attempting to define rural places and people based on the lack of services and goods. Table 2.1 can numerically describe this result.

Type of Place	Number of Tributary Areas	Distance Between Places (km)
M	729	4.0
A	243	6.9
K	81	12.0
B	27	20.7
G	9	36.0
P	3	62.1
L	1	108.0

Table 2.1: The Central Place System of Christaller's Southern Germany (King 1984).

All places in the first column are a subset of L, where L contains all of the goods and services possible and is, therefore, the most urban. For example, L contains the largest tributary made up of 3 P regions, or 27 B regions, and so on. This one L city provides all the functions of the smaller tributaries plus more. Therefore, all the goods and services provided by L cannot be found in P, and all the goods and services in P cannot be found in G. M would be the most rural place in this example and L would be the most urban place. In fact, this could theoretically represent part of an urban-rural continuum, where L is the most urban and M is the most rural. All other places fall somewhere along the continuum. This same concept is illustrated in Figure 2.2 and is referred to as the hierarchical markets system.

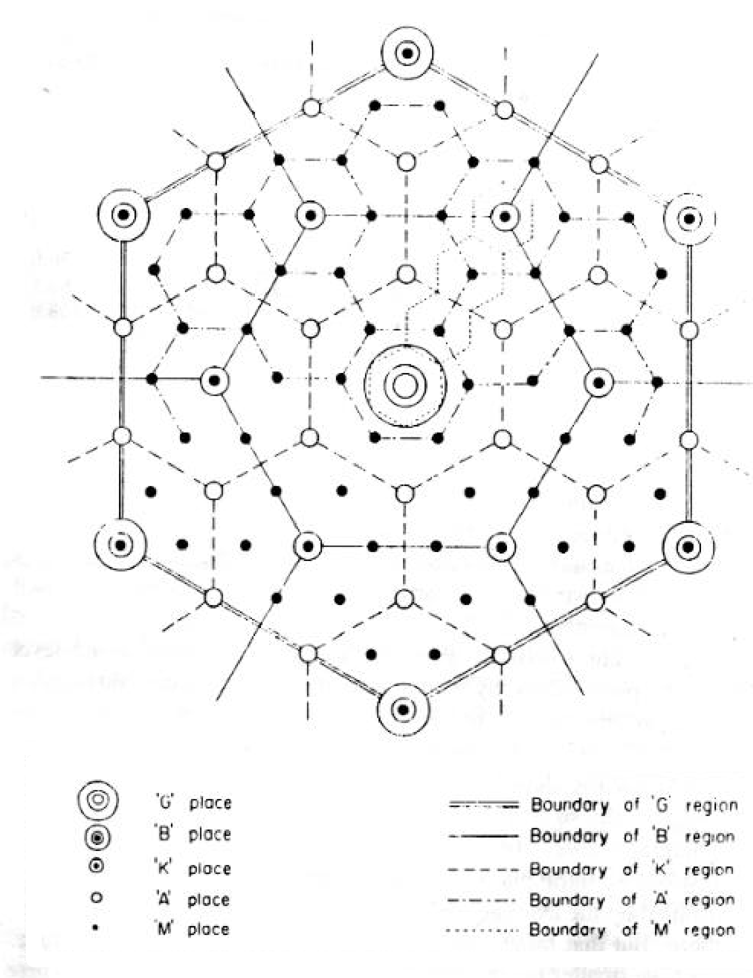

Figure 2.2: Hierarchical Market System of Central Places (King 1984).

Christaller's work remains an influential model for urban geographers. However, there are major drawbacks to his study. First, his assumptions are not plausible in the real world. There is no real place that satisfies his assumptions; each square mile of land is different from the next all over the world. Second, there are other forces that may distort his hexagonal patterning of central places. Transportation was mentioned in his assumptions, but traffic routing was not. Christaller did not test his theory according to the position of central places with respect to traffic routes. His towns may be equidistant from each other; however, one central place may be on an interstate while another may be on a minor collector, ultimately upsetting the real and ideal upper limit of the range. Christaller was an economist. His spatial representation of urban and rural places is a crucial element in any study of rurality; however, his assumption of a homogeneous spatial place is not. Also, he does not provide a sociological or ecological theory of urban versus rural. His theory is purely an economic perspective.

2.3 Stage 3 (1960s): The Sociological Perspective

Up until the 1960s in the sociology field there was no single definition to identify rural. There had been attempts and their outcomes produced a polar dichotomy between urban and rural, yet no definition of rural existed. However, even these theories began to deteriorate with the spread of communications and telecommunications. Instead of two ends of a spectrum, a single culture began to emerge with regional differences being much reduced. The two diverse societies “Gemeinschaft” (rural) and “Gesellschaft” (urban) became a concept of the past and a continuum emerged based on the degree of urbanization experienced in an area. This rural urban continuum theory began in the 1960s with a paper published by Pahl (1966), conceptualizing a completely different method of categorizing rural. In fact, this paper outlined a continuum from urban to rural, where no distinct boundaries occurred from one entity to the next. Instead he described a constant line where the two extremes, urban and rural, were denoted and represented as end nodes along it.

Crucial to this idea of an urban rural continuum was the context of study. In Pahl’s paper *The Urban Rural Continuum* (1966), he admits (p. 301):

> I am not concerned here with the ranking of various settlements according to service function and their size and spacing; these questions may be of interest to geographers and agricultural economists but their sociological relevance has to be demonstrated.

Pahl sought to categorize the people choosing the urban or rural life rather than the economics explaining a rural or urban way of life. He, therefore, conceptualized the social relations and social organizations that were fixed in space (Lobao 1996) rather than the economic relations and organizations that were expressed in Christaller's theory. Pahl maintains that class is the most sensitive index of peoples' ability to choose, and that stage in the life cycle determines the area of choice, which is most likely. A new rural population is delineated in this regard. By considering the physical surroundings and the spatial constraints as 'simply confusing variables' he distinguishes major groupings based on two assumptions. The village should appear to be socially heterogeneous and, because by definition most chief earners commute to work in surrounding towns and this spatial restraint operates differentially, the amount of choice becomes more limited further down the social scale (Pahl 1966). With this spatial constraint, Pahl defines six categories of "people".

1. Large Property Owners. This group includes wealthy landowners who are tied locally by tradition and property but who have financial and other interests elsewhere.
2. The Salariat. This group includes business and professional people who have defined for themselves a village-in-the-mind and whose place of residence is

subjectively an important aspect in the style of life to which they aspire.

3. The Retired Urban Workers With Some Capital. This group includes those who come to the settlement to buy or build a house for retirement.

4. Urban Workers With Limited Capital/Income. This group may not want to live in the settlement but are forced to due to the high price of urban land. These are the reluctant commuters and are perhaps the most important immigrant element in many newly expanded commuter villages.

5. Rural Working-Class Commuters. This group includes those people who have inherited or have other qualifications for a house in the village but who are obliged to seek employment elsewhere.

6. Traditional Ruralities. This group includes a small minority element of local tradesmen, agricultural workers, and so on whose residence and employment are both local.

These classifications may serve some purpose from a sociological perspective. However, the process of collecting adequate and sufficient data involves subjective classifications, which do not meet present scientific standards. In present research disciplines, there are many problems associated with classifying people into these categories. Not only is fieldwork tedious, but also the above categorical definitions do not stand the test of time and consistency between groups. For example, city housing is sometimes less expensive than rural land today, disputing the fourth group's existence namely the Urban Workers with Limited Capital/Income. However sufficient these groups were at the time Pahl's paper was written, they do not broadly and objectively distinguish rural from urban.

The second half of Pahl's paper discusses choice. He clearly states that choice is a way of life. In today's society, specifically in North America and other first world countries, it would be impractical to interview everyone living in a rural society and ask them why they are living there. The motivations of people to live or not live in a rural area have undoubtedly changed since the 1960s. It is not a simple formula. People cannot be placed into six simplified categories. There are numerous factors associated with one household's move to an area. These cannot be scientifically categorized into six groups, probably not even ten groups.

Pahl addresses this issue in his conclusion. He suggests limiting this type of study to one type of worker, the manual worker. However, in some areas this type of worker only accounts for ten percent of all the workers. How can this adequately exemplify the whole community? What about the other ninety percent of the people living there? Today less than two percent of America's labor force is engaged in farming, while other manual industries such as mining and fishing (and some lumber, milling and paper manufacturing) account for less than six percent of the total U.S. labor force (Luloff and Nord 1992). Clearly, using Pahl's purely social urban rural definitions does not provide sufficient detail for those in search of an all-encompassing definition of rurality.

The 1960s and 1970s provided advances to the notion of an urban-rural continuum in the field of sociology. For instance, Rogers and Burdge (1972) took Pahl's notion of a continuum and created their own continuum using variables including population size, population density, and the degree to which the community members observe rural or urban norms, the latter presumably a subjective approach (Figure 2.3).

Figure 2.3: A Linear Representation of Rogers and Burdge's Urban-Rural Continuum.

Where A= rural neighborhood
B= agricultural village
C= small Town
D= rural urban fringe community
E= suburban community
F= small city

Every place falls somewhere along this continuum where A is the most rural and F is the most urban place.

The notion of a rural-urban continuum arose in reaction against the polar type dichotomies of urban and rural. However, as Pahl mentions, there "are equal dangers in over-readily accepting false continuity" (Pahl 1966). There are of course sharp discontinuities in this continuum particularly when applying it to different scales. The point at which a community is more properly described as urban rather than rural is, therefore, not easily determined. In fact it is discontinuous over space as countries around the world use different population sizes to describe what is urban and what is not. This approach would seem to limit the ability to compare regions across scales. For places designated rural it is impossible to determine which is more rural because of differing threshold values of rural in countries around the world. This justifies a definition that captures demographic characteristics, but focuses its attention on spatial components that are universal.

2.4 Stage 4: The Government's Perspective

In the United States the greatest activity in community research in rural sociology occurred in the late 1930s and early 1940s. The Bureau of Economic Analysis (BEA) paved the way with its decision to sponsor a series of analytical and empirical studies. Led by Taylor (1941), the BEA organized a national project to describe and compare small communities in various regions of the country with similar methods of data collection and analysis to be used in all of the cases. To understand the social and economic conditions in the United States, six communities showing great range differences among American communities were selected (Wilkinson 1991). However, referees of the study felt that attributes about the communities were collected without a clear framework for investigation. The study was informal and subjective. The data collectors differed from county to county and definitions of the attributes also varied. It was suggested that had there been an explicitly comparative approach, the contribution of the study would have been greater to the field of sociology (Wilkinson 1991). But, government agencies did not admit defeat. Their efforts continued and their definition of rural continues to be the most widely used in the United States.

Thus, for the official definition of rural, researchers turn to the government. Unfortunately, the government labels rural in terms of what it is not, rather than what it is. Essentially, what is not metropolitan in America is rural (Fitchen 1991), a definition of exclusion rather than inclusion.

> The Census Bureau defines "urban" for the 1990 census as comprising all territory, population, and housing units in urbanized areas and in places of 2,500 or more persons outside urbanized areas. Territory, population, and

> housing units not classified as urban constitute "rural." In the 100-percent data products, "rural" is divided into "places of less than 2,500" and "not in places." The "not in places" category comprises "rural" outside incorporated and census designated places and the rural portions of extended cities. In many data products, the term "other rural" is used; "other rural" is a residual category specific to the classification of the rural in each data product. In the sample data products, rural population and housing units are subdivided into "rural farm" and "rural nonfarm." "Rural farm" comprises all rural households and housing units on farms (places from which $1,000 or more of agricultural products were sold in 1989); "rural nonfarm" comprises the remaining rural. (www.census.gov, 2000).

According to the U.S. Bureau of the Census a “metropolitan statistical area” is a central city of at least 50 000 people or an urbanized area consisting of 50 000 or more people in a city and the surrounding counties that are economically tied to it. Consequently “non metropolitan” are all those places that are not included in the definition for metropolitan. The term “rural” technically refers to “the population outside incorporated or unincorporated places with more than 2, 500 people and or outside urbanized areas” (Fuguitt et al, 1989). Generally, people use rural and non metropolitan interchangeably in the United States. Rural America, then, is officially just a residual from urban or metropolitan, leaving it less than clear what rural really is. The very existence of a rural America is thus contingent upon an urban America (Fitchen 1991).

Even the official Census definition is, therefore, ambiguous in that an urban definition is needed in order for a rural one.

Although the Census Bureau provides a definition, other official groups have determined their own classifications of rural and urban. Official definitions occur not only at the federal level, but also at the state level. For instance, the New York legislature has designated 44 counties as rural. From this, seventeen percent of the state's population is rural. However, according to the Census definition, only nine and a half percent of the population of New York is rural. The result is 1.3 million people who live in limbo; they are classified as either rural or not rural depending on which official designation is being used at that moment.

Defining rural only in demographic terms, only as a residual category, and for a specific agency context has significant shortcomings. If over one million people can be seen as an either-or category it is clear that an improved all-inclusive definition is needed. Defining rural as residual has led some to believe rural areas are undervalued and, therefore, are treated as residual areas.

> Rural America the residual space of the nation, is increasingly becoming the place for the minor and low paying manufacturing enterprises, for the prisons, the landfills, incinerators ash, and nuclear waste. The "rural as residue" problem hurts rural places, then, not just in the funding that they don't receive but also in the "goods" that they are asked to accept. (Fitchen 1991, 248)

By defining rural spatially, demographic characteristics are captured and this focus on residual space may be alleviated.

2.5 Stage 5: A New Approach

The most confusing aspects of rural are the variations in the characteristics on which it is based, such as the three dimensions discussed to this point. Stages 1 to 4 have clearly shown that economists have used this term in their applications, sociologists have attempted to define this term over the years by categorizing the people who live in these areas, and government bodies have struggled to put people into assemblages based on this term. Yet no single definition has proven to be satisfactory. Furthermore, each dimension cannot stand-alone. Butte and Flinn (1977) found very weak support for the hypothesis that ruralism was more strongly associated with environmental (ecological) concerns rather than agrarianism (Gilg 1985). Even with a resurgence of the economic foundations of rurality in the 1970s, the focus was still limited to a few sectors, mainly farming (Lobao 1996). Thus the occupational approach has been nullified with the loss of agricultural employment. This leaves the cultural or sociological dimension. However, as Bealer *et al.* argue (1965) “A single dimension … would probably not receive widespread acceptance. A composite definition has more overwhelming appeal” (from Gilg 1985).

It is only recently (late 1980s), that space has come to play a more prominent role in a definition of rural. The “new rural sociology” or “rural restructuring” is the most current stage showing broadening concern with spatial issues. In 1981 and specifically in 1983 two separate government agencies attempted to redefine rural areas using geographic boundaries. In the 1981 Census an effort was made to physically define urban areas by using a threshold of one thousand people and 20 hectares of land. In 1983 the U.S. Department of Agriculture subdivided the Census metro and non metro categories to form the Rural-Urban Continuum Codes. The classification method distinguishes metropolitan counties by size, and non metropolitan counties by degree of urbanization or proximity to metro areas (Butler 1990). Altogether it includes 714 metro counties and 2, 383 non metro counties. Table 2.2 shows the classification scheme for the northeast region of the United States. This definition has recently been revised. In the 2000 Census new categories will be implemented.

Code	U.S.	Northeast
Metro	714	117
0	54	15
1	173	29
2	289	62
3	198	11
Non metro	2 383	100
4	137	25
5	151	6
6	552	31
7	757	24
8	229	9
9	557	5
Total	3 097	217

Table 2.2: Regional Distribution of Metro and Non metro Counties (Butler 1990).

The Department of Agriculture has brought together two concepts – population density and proximity to place. However, purely empirical methods for analysis only establish comparisons between places, but do not enlighten researchers on the meaning of rural (Jacob and Luloff 1995).

2.6 Summary

Researchers are left with unclear methods for measuring and defining rural. The three concepts –ecological, occupational and socio cultural - have proven to be insufficient. The statistical analysis used by the Census Bureau and Department of Agriculture is only a slight improvement. Further clarification of the concept is necessary if it is to be meaningfully utilized in scientific work. The following chapter explores a new theory to measure and define rural.

Chapter 3

Indexes, Indicators, and Clusters

In order to derive a possible model of rurality it is first necessary to examine the contributing characteristics. The aim of this chapter is to establish and explain the variables that are included in the model described in Chapter 4 and to categorize them into two "clusters." As presented in Chapter 2, previous definitions of rural focused on social and economic clusters. This chapter emphasizes spatial dimensions and develops two clusters or categories based on spatial relations. The objective of this chapter is to first examine the spatial manifestations that affect an area's rurality, both topologically and by degree of isolation, next to categorize these components, and finally to formulate an index for the model.

Terminology must be clarified before any grouping is discussed. The next three sections provide an overall explanation of the terms **index**, **indicator**, and **cluster** and how they relate to this model. With these terms defined a formal index of rurality is developed.

3.1 Determining an Index for Rurality

Indexes are measures that combine indicators to describe the performance of an institution, region, or economic sector. Measurements help the decision-makers and the public to define goals, to link them to clear objectives and targets, and to assess progress toward meeting those targets. An index provides an empirical and numerical basis for categorizing, for evaluating performance, for calculating the impact of activities on the environment and society, and for connecting past and present activities to attain future goals (http://iisd.ca/measure/default.htm).

The medical field is very interested in developing their own "index of rurality." Their index is finely tuned to include indicators specific to the medical profession. This seems very helpful to that particular field for medical practitioners, patient transfers, and funding; however, a broad index must be developed in order to benefit many disciplines on a national and global level. Thus, an instrument for measuring the rurality of a place is needed to provide a standard of comparison that can be used by researchers, educators, and administrators (Leduc 1997). A requirement for this measurement of rurality is to distinguish an area's relative rurality from the rest of society and express it in relational terms. The following collection of rural indicators, once combined mathematically or aggregated (the process of which will be explained in Chapter 4), will result in a number. This number will represent a specific region's rurality. Another region can be chosen and a different number might be calculated. The two numbers can then be compared with each other. The index is designed to provide a method to measure and compare the rurality of two or more places.

Another characteristic of this index is that it can measure change in rurality. This is one feature that previous indices did not include. With sufficient data, this index can reflect change over time within an area. It can answer questions such as, has a town become more or less rural? And, if so, how has this been achieved?

Many indices are widely accepted today. The Human Development Index (HDI) (United Nations Development Program, 1998) measures the quality of life in a nation. It uses life expectancy, adult literacy, and Gross National Product per capita as its indicators. By combining these three elements and by judging each nation's indicators against 'the best,' the result is a worldwide HDI. Additional indices include the Sustainable Process Index (Krotscheck, 1998) and the Ecological Footprint (Wackernagel and Ree, 1996). Both of these evaluate an individuals' influence on the environment. All indices are used to simplify complex systems to just one number, which ultimately can be useful to decision-makers and others on regional, national, and international scales. The "index of rurality" is also used to simplify a number of complex factors into just one number.

3.2 Determining Indicators of Rurality

The measures that constitute an index are generally referred to as indicators. Indicators are presentations of measurements. They are pieces of information that, when put together, summarize a system or indicate the status of a system. Indicators simplify complex phenomena, making it possible to gauge the general status of a system. The dials on a car's dashboard or the financial reports in the business section of a newspaper are examples of indicators. Indicators are found everywhere. The General Practice Rurality Index used by the Canadian Medical profession referred to in Section 3.1 uses indicators such as drawing population, remoteness from a basic referral center, number of general practitioners, number of specialists and presence of an acute care hospital. However, in this index and many others, the spatial indicators of what constitutes rural have been deficient.

Previous indices have not recognized the ability of spatial relations to help classify an area's rurality. The focus of the SRI is, therefore, on indicators that have inherent spatial relations that can be examined. This type of data is designed to enable specific geographic features and phenomena to be managed, manipulated, and analyzed easily and flexibly to meet a wide range of needs. The International Institute for Sustainable Development has prepared a list of important qualities needed for a good indicator. We have slightly modified this list. Table 3.1 shows qualities sought in good indicators.

Quality	*Explanation*
Policy	Can the indicator be associated with one or several issues around which key policies are formulated? Unless the indicator can be linked by readers to critical decisions and policies, it is unlikely to motivate action.
Simplicity	Can the information be presented in an easily understandable, appealing way to the target audience? Even complex issues and calculations should eventually yield clearly presentable information that the public understands.
Validity	Is the indicator a true reflection of the facts? Were the data collected using scientifically defensible measurement techniques? Is the indicator verifiable and reproducible? Methodological rigor is needed to make the data credible for both experts and laypeople.
Time-series	Is time-series data available, reflecting the trend of the indicator over time? If based on only one or two data points, it is not possible to visualize the direction the community may be going in the near future.
Availability of affordable and good quality data	Are good quality data available at a reasonable cost or is it feasible to initiate a monitoring process that will make it available in the future? Information tends to cost money, or at least time and effort from many volunteers.

Ability to aggregate information	Is the indicator about a very narrow or a broader issue? For practical reasons, indicators that aggregate information on broader issues should be preferred. For example, forest canopy temperature is a useful indicator of forest health and is preferable to measuring many other potential indicators to come to the same conclusion.
Sensitivity	Can the indicator detect a small change in the system? We need to determine beforehand if small or large changes are relevant for monitoring.
Reliability	Will you arrive at the same result if you make two or more measurements of the same indicator? Would two different researchers arrive at the same conclusions?
Consistency	Is the indicator consistent with what is already known? Will the indicator provide results that do not complement real world observations?

Table 3.1: Selection Criteria for Good Spatial Indicators.

These criteria are used as guidelines to identify a set of feasible spatial indicators.

3.3 Determining Clusters of Rurality

Clusters are groups of indicators. The use of clusters is important to most indices, because clusters can broaden the focus of a measurement to include a balance of many signals. This index concentrates on two groups of indicators both referred to as spatial clusters. Each cluster is made up of components suitable for that cluster. With clusters of indicators, separate components can still be emphasized while their combined effects are also revealed. For example, the importance of isolation characterized by one cluster is revealed, while the importance of access to services (or lack of) is revealed separately in another cluster. Thus an area may be characterized as less rural or less isolated by showing a high degree of connectivity but this does not suggest anything about that area's access to services. The SRI also includes the aggregation of the two clusters to provide a general picture of the area's rurality.

The SRI includes two clusters. The first is called the Connectivity Cluster. It examines how isolated an area is based on a network of connections such as links and nodes. A network is a system used to move people, transmit resources, and communicate across distances. Counting the number of links and nodes and weighting their attributes establishes a degree of connectivity. The network system applies to infrastructure such as telecommunications, utilities, or transportation. We expect less connected areas to be more rural. Access to Services is the second cluster in this index. While the connectivity cluster focuses on connections available through links and nodes, this cluster is concerned with access to services as a measure of distance or presence and absence from a service. Some services considered include police, fire, schools, and health care facilities. We expect rural areas to have less access to services and to use government or publicly subsidized services as representative measures. Table 3.2 shows the importance and differences of these two clusters.

CONNECTIVITY CLUSTER	*ACCESS TO SERVICE CLUSTER*
Used as a measure of degree of isolation	Used as a measure of degree of accessibility
Measures links and nodes and number of connections	Measures access to a particular service and distance to that service
Coverage format: arcs, nodes, polylines	Coverage format: points and polygons
Hierarchical ranking based on quality and quantity of attribute- count of links and nodes	Hierarchical ranking based on function of distance and quality of attribute

Table 3.2: Differences between Connectivity and Access to Service Cluster.

Thus a spatial index, not reliant on demographic or economic components, is proposed. The spatial data are stored by the geometric location of geographic features, along with attribute information describing what these features represent, thus enabling analysis in a GIS to take place. The contribution of the indicators to the two clusters and to the final index is shown in Figure 3.1. The final index can be decomposed into two distinct clusters. These clusters can then be decomposed again into the indicators chosen for each cluster.

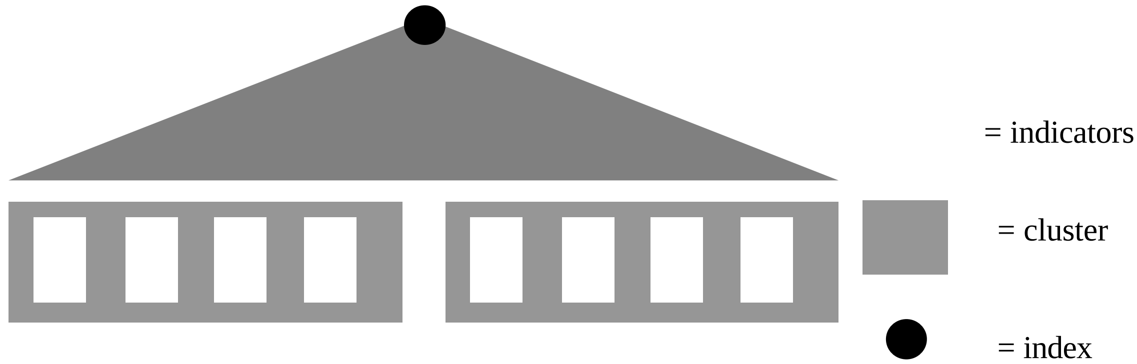

Figure 3.1: The Separate and Combined Components Making up the Index of Rurality.

3.4 Theoretical Examination of the Connectivity Cluster

The word connectivity can be interpreted in various ways. Connect means to join or fasten together; to link; to unite. In a community context connectivity implies an emphasis on utility, transportation, and communication infrastructure. In one definition it is the meeting of various means of transportation for the transfer of passengers. In another it is the line of communication between two points in a telephone or similarly wired system. And in yet another it describes the coverage of sewer or water lines in a community. Cleland (1995) defines connectedness as, “having ties to people in positions of responsibility over resources to conduct one’s activities most effectively for the benefit of self, family and community.” For these reasons, it is essential to clearly define a context for the connectivity cluster. The common thread for all of the above definitions is that connectivity is a relationship between one area and another and measures an area’s level of participation within a larger community or infrastructure. Areas or places can refer to any arbitrary partitioning of geographic space, but often predefined areas are used such as municipalities or counties. Assuming space can be partitioned into any set of regions, connectivity is a measure of the number of nodes for a particular network and or node degree. We predict that rural areas will have fewer or lower-level connections to other areas.

The connectivity between areas is typically materialized as infrastructure. The infrastructures considered for developing indicators in this thesis include utility, transportation, and communication networks. To be counted, a town must have a node despite having links. A link may pass through a town but if there is no node on the link, the link is useless. For instance, an interstate may pass through a town, but if there is no exit ramp, then the town is not considered connected to the interstate. The fewer the number and types of links and nodes connecting a community to another generates a more rural community. Figure 3.2 illustrates this scenario.

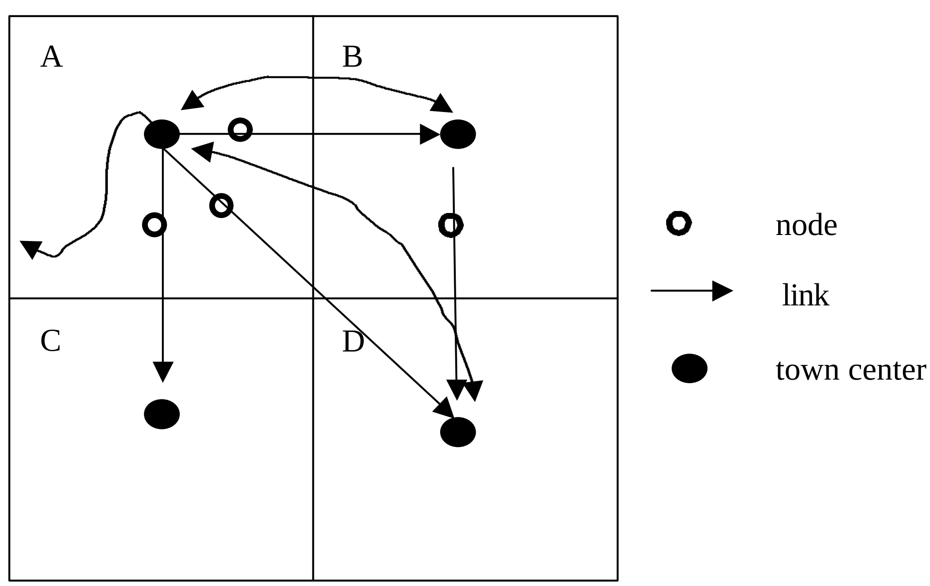

Figure 3.2: Network of Links and Nodes.

In this example Town A is the best connected and least isolated, therefore the least rural. There are six links to other towns and three nodes to a type of infrastructure. Town C is, therefore, the most rural, because it is not connected to any towns by links or nodes.

Similar to graph theory (Crump, 1980), the number of points or nodes is important in determining the total possible number of connections. This type of attribute data is essential in defining a degree of connectedness or isolation. Table 3.3 summarizes the indicators fitting into this cluster. A brief explanation of the importance of each indicator follows.

INFRASTRUCTURE TYPE	**INDICATOR**	**DEGREE MEASUREMENT**
Utility	Water	Size of pipe
	Sewer	Size of pipe
Transportation	Roads	# Of exits, type of road
	Railroads	# Of rail lines
	Airports	# Of airlines, classification of airport (eg. municipal, int'l)
Communication	Internet Connections	Speed of connection, type of connection

Table 3.3: Connectivity Cluster.

3.4.1 Utility

The purpose of most utility companies is to provide safe and affordable water, sewer, and electrical service to the citizens of its district. Rural areas suffer from the competition generated from urban utility companies. Some rural areas do not even have a choice of company, or no company is even represented. These people (potential customers) must resort to their own techniques for a safe and affordable environment, such as depending on well water or on-site septic systems. Rural areas are therefore generally characterized by independence from utilities like sewer and water.

We have chosen sewer and water as utility indicators, because they are policy relevant as towns vie for better and competitive utility companies. They are easy to understand, because either a town has a water system or it does not. It is a valid indicator, because it not only shows connectivity, it also shows degree of connectivity by attribute values such as type of connection, size of pump, and number of customers reached. This information changes yearly as new water mains are put into the system and more (or less) customers become part of the system. In this way the utility is also sensitive to change. Each town is responsible for reporting annually this type of utility information. It is made public, it is affordable, and it is of good quality. Finally, this group of utility indicators is consistent. An area that has a supply of water through a main is considered less rural than an area that pumps water through a well. Clearly, the utility indicators are good quality indicators as they meet all of the criteria listed in Table 3.1.

Water. The Mission of the Maine Water Utilities Association is to enhance public health, safety and welfare by advocating safe drinking water through the advancement of knowledge of the design, construction, operation, maintenance and management of water works through education, development and promotion of legislation, standards and policies and an exchange of information and experience. Thus, connection to water is an important issue. Over the years the federal government has been urging private companies to provide services to rural areas, however many areas still rely on private water or wells. The topological component to consider is: is the community connected to a public water system? This is also the coarsest level of detail of the model, where the entire water system of a town is represented as one node. If more detail is required (a discriminatory power), the indicator can include measures of number of links and end nodes such as how many customers it can supply (how many end nodes), and the strength of the node (what type of pipe transmits the water and what size). These factors help determine a degree of connectedness based on topological link and node relationships.

Sewer. Connection to sewer is similar to connection to water, because it too is an important issue pertaining to health. Following, the water indicator a connection to a public or private sewer system must be established. The coverage of links, the number of nodes, and the quality of a node allow us further examination into a degree of connectedness for the indicator.

3.4.2 Transportation

Rural America accounts for a small and dispersed portion of the nation's population, yet it encompasses a significant portion of the transportation system. Rural areas account for 80% of the total U.S. road mileage and 40% of the vehicle miles traveled (NADO, 1999). Consequently, the rural traveler has similar transportation needs as her urban counterpart, though the priority of these needs differ. These differences reflect the rural environment of long distances on secondary or unpaved roads, relatively low traffic volumes, travelers unfamiliar with the surroundings, and rugged terrain in remote areas. This transportation framework can also be applied to the railroad network.

We have chosen roads, railroads, and airports as indicators for many reasons. Like utility systems, transportation system issues are of major policy relevance. In many areas (including Maine, Arkansas, Iowa, Indiana, Texas, and Pennsylvania) state and federal highway acts such as the Transportation Equity Act (TEA-21) are being rewritten as money is unfairly distributed throughout urban not rural areas (Brown 1995). Transportation data are reliable and affordable. Both the Department of Transportation and GIS affiliated organizations such as the Maine Office of GIS provide free, good quality transportation data. Road classification data, railroad, and airport data are simple to understand, valid, and easy to aggregate to higher spatial levels. Links and nodes are countable and consistent with what is already known. Finally time series analysis is viable since the number of airlines servicing airports in 1990 is known as well as the numbers fifty years ago. Such an indicator is consistent, since the more links and nodes the less rural the place is. Our transportation indicators are good quality indicators that pass the criteria suggested in Table 3.1.

Road coverage. Roads are considered the best routes for getting from one place to another and usually interstates and primary roads are the paths most frequently chosen to travel. This indicator measures connectivity to some other place (town, city, county) via components and attributes of the road network. Degree of connectivity is measured by the number of nodes within an area and the importance of the nodes. The level of the node is determined by the highest-level link incident at the node. The hierarchy of the road system conveys a hierarchy to the indicator allowing it to scale well across regions. Thus an interstate/interstate node is the highest possible node, followed by interstate/primary, and so on. Therefore, Portland with three interstate/interstate nodes is considered less rural than Bangor with two interstate/interstate nodes. First, the highest-level link will be determined within an area by whether the area contains interstate/interstate connections and by the number of exit/on ramps. Urban areas generally contain at least one interstate exit ramp, while rural areas do not. The number of exits within the area accounts for a degree of connectedness. Second, the other levels found within the selected area are determined. An interstate road will be weighted differently than a primary or secondary classification; this is described in more detail in Section 4.1.2. Dennis Brown (1995) claims, "Transportation infrastructure deficiencies are also evident in some poor rural communities whose lack of sufficient revenue for road maintenance limits the communities' economic development potential." One would expect that roads of inferior quality be found in communities with insufficient revenue, like some rural communities. It is the case too that rural areas incur high per capita highway costs, because their roads and bridges serve scattered populations of smaller communities. For these reasons it is assumed that unimproved roads are very likely to be found in rural areas, which translates into lower level nodes found in rural areas. Again, it is assumed that the more urban the place, the more likely a link of the interstate or primary road type is found there, and the greater the number of high-level nodes (such as interstate exit ramps). In a rural area we expect to find a small number of nodes at the highest level and more links and nodes at the lower levels such as secondary or unimproved.

Railroads. Railroads are still considered a fast and efficient means of traveling and moving freight from one place to the next. Customer and freight stations can be found in areas where people enjoy receiving other people or packages. Railroad lines are important links to connect to the outside world for both freight and passengers. They also encourage tourism and trade to an area. Connectivity is measured by containment of nodes (stations) and node degree-the number of links incident at the node. If an area only contains a link, but not a node, it is assumed disconnected because the train will not stop at that area. The number of rail lines available at the node will determine the degree of connectivity for this indicator. It is assumed that a smaller number of rail lines serve a more rural area. Or a smaller ratio of passenger to freight services.

Airport. The placement of commercial airports infers centers of activities or common destinations for travelers or items of trade. The presence of a commercial airport within an area is a measure of connectivity. Evaluating node degree refines this measure. The node degree is the number of airlines serving the node or the number of daily flights. It is assumed that a smaller number of airlines and daily flights serve a more rural area.

3.4.3 Communication

Research suggests that there is a lack of rural telecommunications infrastructure, while there are a growing number of telecommunications businesses in urban areas. It is necessary to first address particular needs of rural areas. Following this, a task force should be established to provide valuable assistance in identifying the issues unique to rural carriers. Without adequate connections to advanced telecommunications infrastructure and services, rural communities may not be able to fully participate in the emerging information economy.

In the past decade considerable legislation has been updated and passed pertaining to rural telecommunications (Telecommunications Act of 1996 and S 1587, ‘Encrypted Communications Privacy Act of 1996’). It is a major concern for policy makers thus meeting the policy relevance criteria. The indicator is also reliable and affordable. Internet coverage is a valid and simple indicator. Type of Internet line is a simple measurement of rurality and the result is consistent across spatial scales. We assume rural areas will not have quality nodes with fiber connections. We also assume slower Internet connectivity in rural areas. Time series analysis is key to communication. Internet services and tele-video conferencing sites are appearing rapidly in more urban areas. This indicator is thus a good quality indicator for measuring connectivity.

Internet connection. Internet connections are becoming influential to educational research, fast and effective business, and an easy way to access information. Connection to the Internet is becoming an important government initiative. A survey of rural businesses and residents conducted by the Rural Policy Research Institute found that 69 percent of the rural community respondents regularly use fax machines, 46 percent use computers, 25 percent use computer modems, 15 percent use e-mail, and 6 percent use the Internet (Byers 1996). The report also found that rural households lag behind urban households in their access to information technologies. The Internet coverage measurement counts the number of Internet providers in an area and type of Internet node. Again nodes and links are essential components to capture a degree of isolation or connectivity. Generally, the more urban an area the larger the number of Internet providers. The quality of the node establishes a degree of connectivity as well, where access to fiber, IP, or quick Integrated Services Digital Network (ISDN) connections are assumed to be found in less rural places and phone line connections (XTB) in more rural places.

3.5 Theoretical Examination of the Access to Services Cluster

The first measure focused on connectivity by determining the number of nodes and node degree. This cluster measures the access to a particular service deemed significant for convenience. Many communities in rural areas do not have access to what urban centers consider basic services, such as hospital facilities or police departments. By examining these factors, both topologically and descriptively, rural areas can be captured. Figure 3.3 illustrates this scenario over a large and small spatial scale.

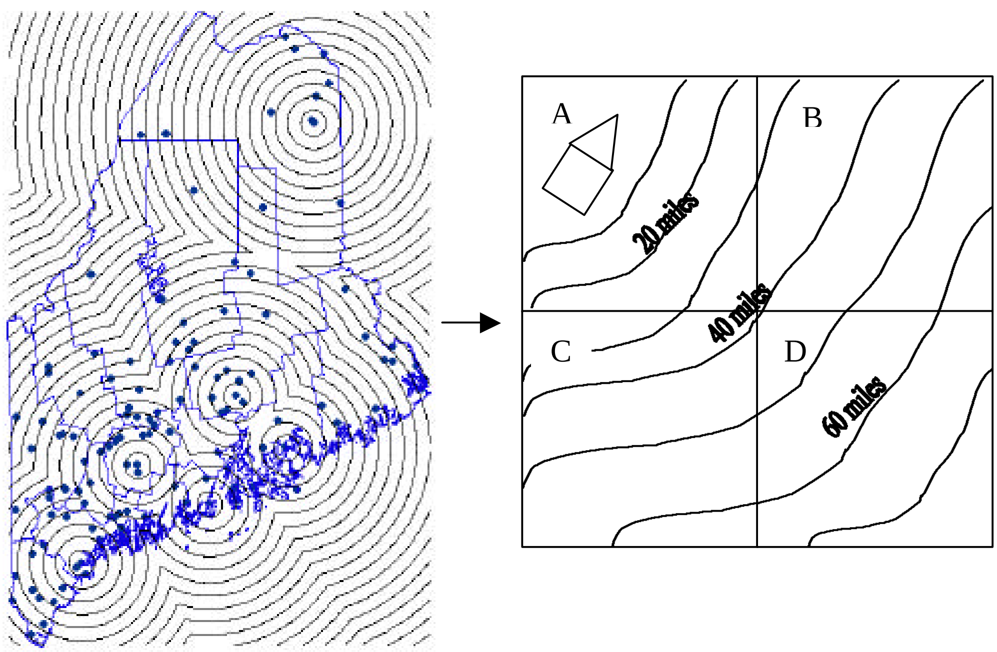

Figure 3.3: An Example of Large Scale and Small Scale Access to Service.

In this example, Town A in County A contains the service. Town B and C are closer than town D to this service. Therefore, Town D, is most rural because it does not contain a service and it is the farthest town away from the closest service. Table 3.4 summarizes the indicators fitting for this cluster (not in any particular order). A brief explanation of the importance of each indicator follows.

ACCESS TYPE	**INDICATOR**	**DEGREE MEASUREMENT**
Health Care	Hospital Facilities	Distance, # of medical services
Education	Schools	Distance, level of education
Safety	Fire Departments	Distance, # of fire trucks
	Police Departments/ Sheriff's Offices	Distance, # of police or sheriff cars
Telephone Service	Exchange Based in a Service Center	Distance, long distance charges

Table 3.4: Access to Service Cluster.

Health Care. There is considerable literature specifying the need for appropriate health facilities in rural areas. It has been proven (Rourke 1997) that the practice of medicine becomes more challenging as distances from urban areas and isolation increase. The medical professions themselves find it necessary to define a general practice rural index. It demonstrates the need to connect the isolated rural dwellers to a facility that can provide immediate and required health care. Similar to the Connectivity Cluster, first a topological relationship of containment is determined. Next a raster overlay is created by a distance function from service points (health care facilities). By overlaying this distance from service layer with the centroids of an area to be measured, a simple distance relation is determined. However, measuring distance along the road network can make a more realistic measure, because people must use the road network to get to a service. Given a set of service centers as starting points, instead of measuring the Euclidean distance (ie., bird's eye view), a center claims all road links closer to it than to any other center. Therefore, links record the increasing distance from a center and are realistic. To study the degree of accessibility, other attributes are required. Number of hospital beds and type of services are significant issues considering access to medical care. Hospital care is a fundamental necessity. Key government initiatives are focused around this indicator. Data are affordable and simple to understand. A major hospital clearly has more services than a county clinic, for example. A small change in a system, such as the addition of a wing or hospital staff helps with time series analysis and sensitivity. This indicator is also a good indicator because of its consistency. For example, services cannot be found in rural areas that are found in urban areas. This is consistent with our notion of accessibility, where more medical services are offered in less rural areas.

Education. Education is a service provided to every child. The Mission of the Maine State Board of Education is to provide statewide leadership by advocating, promoting, and improving education policy and life-long learning for all Maine people, particularly its children. The Board offers direction to the Executive and Legislative branches of state government; thus, fulfilling its legislative requirement. It is the responsibility of the state to provide this essential learning experience to its citizens. We investigate access to education in two levels namely: (1) access to a high school and (2) access to higher education. Distance from each of these point locations is determined for a degree of accessibility. Universities and higher education centers are more likely to be found in less rural areas, (although in some cases locations of land grant universities and their satellites offset this). Thus access to this service is limited in rural areas. The education indicator is a good quality indicator in that it is reliable and policy relevant. The state selects a board of members to run the Department of Education and notes of the meeting are available to the public. Thus information is accessible and affordable. It is also a simple and appealing indicator in that everyone understands the importance of higher education and that access to education leads to clear and effective communicators, self-directed life-long learners, creative and practical problem solvers, integrative and informed thinkers, responsible and involved citizens, and collaborative and quality workers. Finally, events such as opening, closing, expanding, or contracting of schools are available time series data that can capture subtle shifts in rurality. As an example, if a university is built in a rural area, the number of people relocating, other services coming into the area, and money generated from this construction clearly change this rural area into something more urban.

Safety. The Department of Public Safety serves the people by protecting their lives, rights, and properties. This is accomplished through criminal justice, law enforcement, fire safety, and emergency response services. Public safety is a fundamental indicator and everyone is entitled to this right. Access to public safety is a critical issue especially for rural areas. Some towns do not even have their own municipal police office. They rely on the sheriff's office to provide safe and secure streets. We examine access by first establishing whether an area contains a fire station, police or sheriff's office. The latter two safety offices is scale dependent and is explained in more detail in section 4.3.1.

Measures of distance to safety centers allow degrees of accessibility to be established. Other attributes can be added to the model such as number of safety vehicles (police, fire). We assume that rural areas do not have their own fire or police stations, and where they do they are small and lacking essential equipment. Access to public safety is a simple indictor to understand and attribute data are available to the public. It is certainly a policy relevant indicator and reliable. Time series analysis is important and a change in the rurality of a place can be witnessed by an increase in the number of fire or police vehicles. Therefore this is a good quality indicator that is sensitive to change and effective in measuring the rurality of a place.

Telephone service. Present telecommunication literature claims that rural communities face two types of barriers--barriers to access (physical, technological, economic) and barriers to use (Beyers 1996). The higher costs and lower return on investment involved in servicing rural areas discourages many telecommunications providers from expanded or expanding services to rural customers. Urban markets are able to support multiple service providers for inexpensive prices while many rural markets may not be able to support even one single provider. The services offered to rural areas are far less efficient than those offered to urban areas. In this study, we examine the area covered by each telephone exchange company. The indicator is a measure of whether the place is part of an exchange with a service center or not. The classification of a service center is not dependent or related to the size or shape of a telephone exchange center.

The Maine State Planning Office defines a service center by identifying factors that make a strong, vibrant regional center. This definition is independent of work relating to telephone services. More than 25 factors fall into four categories: growing community, vibrant economy, quality of housing and infrastructure, and community well being. All share three attributes: a) they are job centers -- importing workers, b) they are retail centers -- with sales exceeding the needs of the local population, and/or c) they offer an array of social, cultural, health and financial services to the surrounding region. (Maine State Planning Office 1998). First we determine whether or not a place is a service center. To determine a degree of accessibility we calculate a distance to the service center through the telephone lines. This captures whether a call is long distance or not. Generally, more rural areas are charged long distance calls to these service centers. This indicator actually combines link and node topology with distance metrics to form a good quality indicator. It is policy relevant as issues of long distance calls and the area of phone exchanges contributes to state and federal government initiatives. It is a simple and valid indicator because it is a true reflection of the facts. Either a place is charged a long distance or local call to access services within a service center. Service centers change over time. Thus, an area may lose or gain a service center. This indicator captures these changes. Finally data are consistent over spatial scales. We can look at long distance charges (quality of the node), speed of connection and distance from a service center (quality and quantity of the link) to determine a degree of accessibility. We also utilize node data such as number of customers (quantity of nodes) and type of phone program (again quality of the node). We assume that the more rural areas suffer from higher prices, lack of good phone options, and slower connections due to far reaching links to major centers of activity.

3.6 Summary

This chapter develops the method used to determine an index for measuring the rurality of an area. It presents significant characteristics of any index and examines the need for two clusters. The chapter also explains the importance of each indicator in the cluster. These indicators are chosen based on an extensive literature review. In emphasizing two spatial clusters, we hope to capture the socio economic attributes usually associated with rurality, instead of having to measure them independently. By including hierarchies of both topological relationships and a distance metric for analysis, we can study a degree of rurality, instead of a Boolean dichotomous relationship. Segregating the two clusters is necessary in order to examine both a degree of connectivity and a degree of accessibility. The next chapter develops the concepts of degree of rurality in terms of fuzzy measures (Zedah, 1965). The concepts of scale and indicator granularity will be introduced and developed as they pertain to the model.

Chapter 4

Defining Degrees of Rurality–Fuzzy Components and Scale

The goal of this chapter is to create a model to use as an informative index of rurality. The previous chapter discussed the indicators and indicator clusters valuable for this index. Here, a mathematical framework is employed on each indicator. This includes a statistical analysis of the variables using conventional and fuzzy set theory (Zadeh 1965). Fuzzy set theory takes everyday language and applies mathematical reasoning to it. Thus, terms such as not very rural, barely rural, and more or less rural are represented linguistically yet based on a mathematical model. These terms are applied to regions, counties, and towns. The concepts of scale and granularity are reviewed. The importance of these concepts as they apply to the model is also discussed.

First, the indicators must be accurately transformed into a meaningful expression or degree of rurality. The problem with the most widely used definitions of rural (Bureau of the Census, Department of Agriculture) is that the term corresponds to two-valued logic: is or isn't (on or off, black or white, 0 or 1). Town A is rural or Town A is not rural. Using fuzzy models, a degree of rurality can be established mathematically. This satisfies the need for comparative definitions where Town A can be compared with Town B in degree of rurality.

People using this model may be interested in information at the town level, or at a different spatial level. A town clerk might want to know the town's rural index while a county development agency might want to know the county's rural index. A succinct method for measuring how rural an area is must be applicable to all levels (town, county, region). Another example might be a town clerk interested in a very general picture of a town or a more detailed description of a town. Therefore the concept of indicator granularity is discussed and formalized in Section 4.3.2.

4.1 Selected Spatial Models

The techniques used in this thesis are similar to others used in spatial data models. These models use concepts from mathematics including, order, topology, and metrics as a foundation for analytical operations (Hornsby 1999). A GIS uses order relations to determine perspectives such as left/right or front/behind. Topology is used to determine coincidence, connectivity and inclusion. Topological relations are spatial relations that are preserved under transformations such as rotation, scaling, and rubber sheeting. Finally, metrics are used in a GIS to compute quantitative values such as distances between features, or distance to the nearest facility. A metric spatial data model consists of a set of elements, such that for each pair of elements, it is possible to associate a distance subject to certain mathematical conditions.

4.1.1 Topology

A topological spatial relation exists when objects and their fundamental geometric properties are considered. Topology is a mathematical approach for explicitly defining spatial relationships. The access to service cluster considers a set of areas (those to be tested for rurality) and a set of points representing services. The relevant set of spatial relations is one between points and simple regions. Based on Egenhofer's 9-intersection model where spatial representations are simple regions without holes embedded in R^2, eight distinct topological relationships are possible. They are *disjoint*, *contains*, *inside*, *equal*, *meet*, *covers*, *covered-by*, and o*verlap* (Egenhofer and Franzosa, 1991) (Figure 4.1).

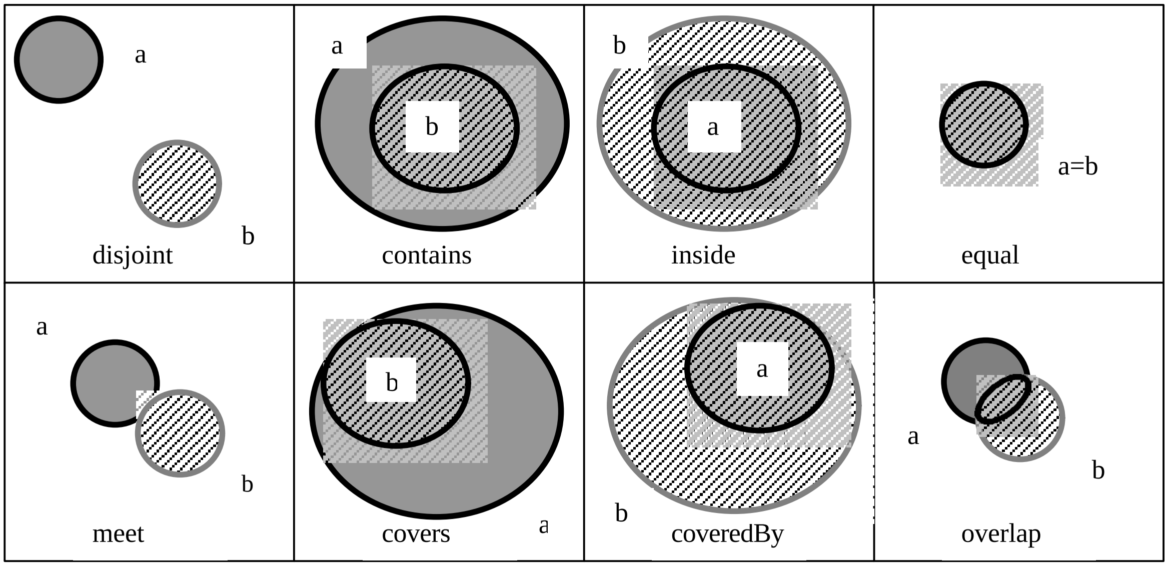

Figure 4.1: The 9-Intersection Model.

However, the SRI is only interested in a *disjoint* or not *disjoint* relationship. Therefore the seven other relationships are collapsed to form everything but a *disjoint* relationship. For this thesis the remaining 7 topological relationships is a *contains+* relationship. An area thus *contains+* a service center or is *disjoint* from a service center. These relationships are Boolean in nature, either a place (*P*) *is disjoint* from an indicator (*I*), or it is not. It follows that if a place is *disjoint* from say a hospital it cannot *contain+* a hospital.

Network or link/node topology is the basis for the connectivity cluster. Graphs are standard representations of this topology. A graph is defined as a finite non-empty set of nodes together with a set of unordered pairs of distinct nodes (edges) (Worboys, 1997). If x and y are nodes of the graph, and $e = \{x,y\}$ is an edge then e is said to join x to y. A graph is a highly abstracted model of spatial relationships, and represents only connectedness between elements of the space. However, such a model is very useful if we allow some extensions. In this thesis we are particularly interested with a labeled graph, where each edge or node is assigned a label denoting an attribute of the node. Based on Egenhofer's nine-intersection model that has been extended for two simple lines (non-branching, no self-intersections) 33 different topological relations are realized (Shariff, 1996). For a line and a region, 19 different situations are found. However, we are not interested in edges or lines unless they end in nodes that are contained within our area of interest. Therefore the 19 different situations can again be collapsed into a *contains+* or a *disjoint* relationship. An area thus *contains+* a link attached to a node or is *disjoint* from a link and its node.

At all spatial levels

If *P contains+ I* then R = 0

where R is the rural nature of P (4.1)

Or

If *P* is disjoint from *I* then R = 1

where R is the rural nature of P (4.2)

For example in our connectivity cluster we have an indicator called Airports. Before determining information about the type of node or airport, we first must identify the correct topological relationship, such as if *P* actually *contains+* an airport, *I*. If *R* = 0 for this chosen indicator, then we can proceed with an examination of the node attributes or node degree data.

Once we have determined the existence of a spatial relationship, we can continue our analysis with a more descriptive one such as node degree to determine the degree of connection. A topological spatial relation is important for the connectivity cluster because links and nodes are counted based on topological relationships and from that a degree of rurality is determined based on node degree. However distance is also an integral part of this model and is defined in terms of metrics.

4.1.2 Distance Metrics

A metric is the measurement of a particular characteristic of an indicator's performance or efficiency. There are four levels of measurements used to classify and distinguish data - nominal, ordinal, interval and ratio (Stevens 1946). A nominal measurement is employed when concerned with only qualitative data. We do not use nominal measurements in our analysis. Ordinal scaling involves nominal classification however it differentiates within a class of data on the basis of rank. The order of the data categories is given, but not any definition of the numerical values. The roads indicator is an ordinal metric in the analysis of type of road (interstate, primary). Interval scaling adds the information of distance between ranks to the description of class and rank. To employ interval ranking, there must exist some standard unit that is expressed by the amount of difference in terms of that unit. Distance from a place is an interval metric utilized often in our index that can be extended to the last measurement. Finally, a ratio measurement requires the employment of an interval scale in which the intervals begin at a zero point that is not arbitrary as are the zeros of the Fahrenheit and Celsius temperature scales. Thus, distance is considered a ratio metric.

Ordinal Indicators. The ordinal indicators in the SRI include: type of road, type of Internet connection, and level of education. The number of ranks is determined *a priori* for each indicator associated with an ordinal metric. For instance, there are 6 ranks in the Internet indicator– cable (1), T3 (2), T1 (3), DSL (4), ISDN (5) and Phone line (6). For each indicator that has an ordinal attribute equation 4.3 is utilized.

$$W_{(i)} = \frac{i}{n+1} \tag{4.3}$$

where W is the weight given to a particular ordinal rank, i is the ordinal rank developed *a priori* for each indicator, and n is the total number of ranks in the indicator, or the highest value of i. Table 4.1 describes this process for the Internet indicator. In this example if i equals 2 (a T3 connection), and n equals six, then W equals two sevenths.

Type of Connection	Indicator Rank	W
Cable	1	1/7
T3	2	2/7
T1	3	3/7
DSL	4	4/7
ISDN	5	5/7
Phone Line	6	6/7

Table 4.1: A Representation of W for The Ordinal Internet Indicator.

The ranking indicated in equation 4.3 is also used as the weight for P in our detailed model of rurality. An area using cable more frequently as its node of connectivity is likely to be less rural than an area with a phone line connection as its node of connectivity. Section 4.3.2 explains how weights and counts are normalized for a refined level of indicator granularity.

Ratio Indicators. Most measures pertaining to length, area, and volume are ratio measures. The ratio indicators in the SRI include: water (number of customers), sewer (number of customers), road (number of type of nodes), air (number of airlines and number of flights), rail (number of rail lines), health care (number of medical services), safety (number of emergency vehicles (police, sheriff, fire)), and telephone (rate of long distance charges). All indicators with a distance metric are also ratio indicators. The GIS program employs a Euclidean distance function to identify the distance from each point to the closest source point. The Euclidean distance is calculated from the center of the source cell to the center of each of the surrounding cells for a raster coverage and by the centroid of the polygon if vector based; for each cell or polygon, the distance to each source cell or polygon is determined by calculating the hypotenuse with the x_max and y_max as the other two legs of the triangle, which derives the true Euclidean, rather than the cell or polygonal distance (Figure 4.2).

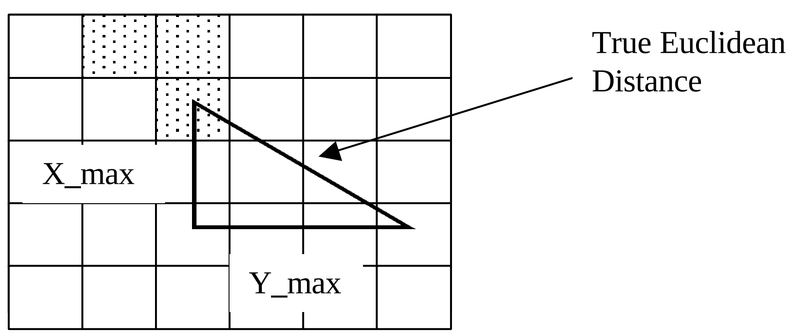

Figure 4.2: Measuring True Euclidean Distance in a GIS.

The result is a table with the minimum, maximum, and mean distance from the source cell to the surrounding cells. We use the mean distance for our analysis. The nearest *P* (place) to *I* (indicator) will have the lowest value, while the farthest *P* to *I* will have the highest value (Figure 4.3). We use a similar equation to the one employed for ordinal indicators. The numbers are normalized to a range between {0,1} and are then put into the model.

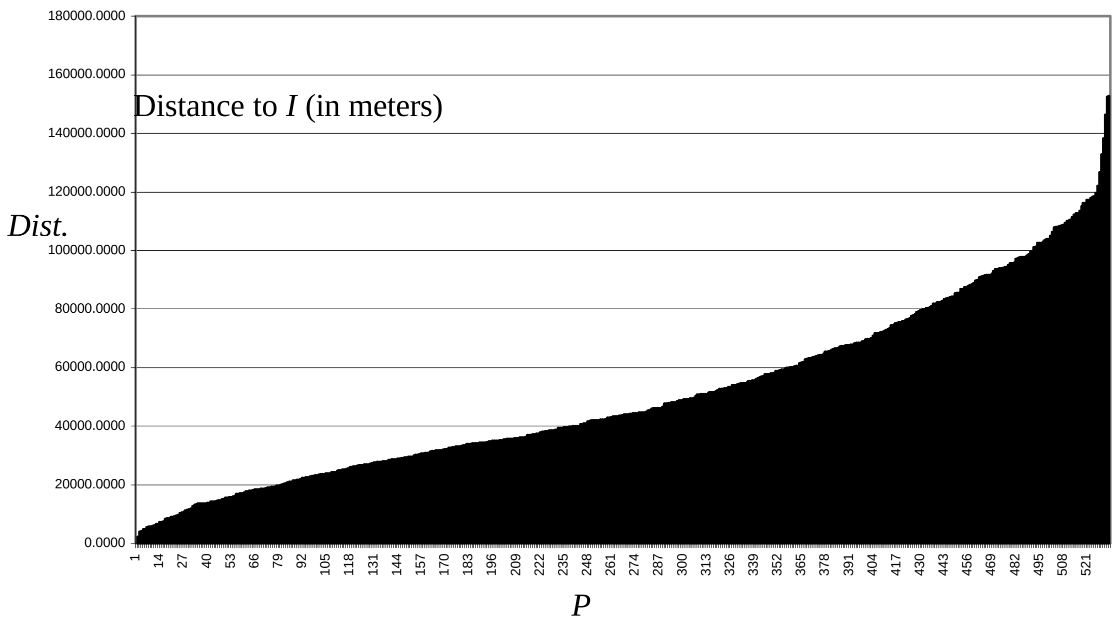

Figure 4.3: Mean Distance of *I* from *P*.

We also use a network distance function for some of our access to service indicators. This identifies the closest facilities and displays the best way to get to or from them. To get these results, a location on a line and the name of the point representing the facility is specified. The centroid of each area (ie. town) is first calculated. The services or facilities are represented (ie. hospitals) and through the road network (the line theme) the closest hospital to the center of the town is found. The length of each link is totaled allowing the final result to be the distance from the centroid of the area to the closest service. Figure

4.4 shows the network distance from St. Joseph's Hospital in Bangor to surrounding municipalities. The length of each line segment is given. From this a total distance is calculated.

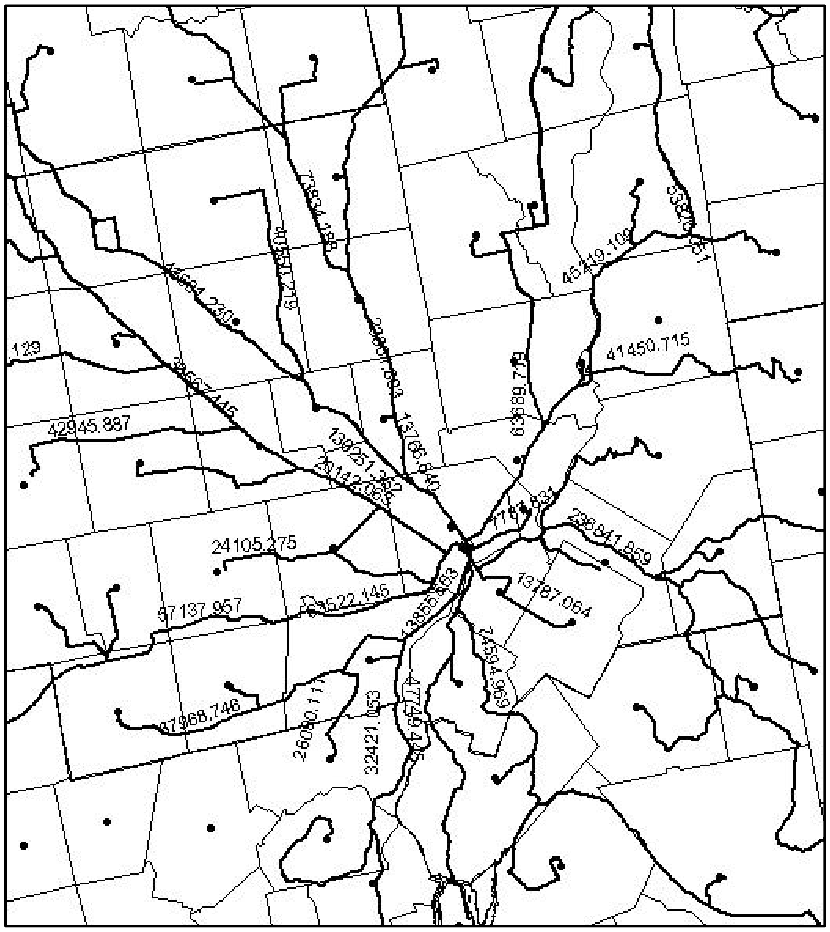

Figure 4.4: Network Distances and Lengths of Links.

From these examples a very precise value results. However, how can we determine a threshold value that makes *P* rural or not. We cannot use conventional mathematical theory and say for example, if *P* is 43549.990 meters away from *I* than *P* is rural, otherwise it is not. To address this we determine a degree of connectivity, and a degree of accessibility, thus a measurement that assigns a degree of rurality based on how well connected or how far away an *I* is to a *P*. Fuzzy models are introduced to generalize discrete levels to a continuous form. Instead of an exact distance or level of connectedness being a threshold value, we can determine a degree of rurality based on a degree of distance (accessibility) or degree of connectedness (connectivity).

4.2 Fuzzy Models

Inexactness and context dependency is an integral component of human cognition and of the human decision making process (Beard, 1994). Current GISs are limited to absolute and exact values and cannot handle inexact terms such as “near”. The concept of a fuzzy set was introduced by Zadeh (1965) to describe imprecision that is characteristic of much of human reasoning. Fuzzy sets are a generalization of conventional set theory and mathematically represent vagueness in everyday life (Bezdek 1993). A primary difference between fuzzy set theory and classical set theory concerns a membership function. In classical set theory each element is either a member of a set or is not, whereas fuzzy set theory allows for grades of membership (Woodcock and Gopal 1999). The benefit is flexibility beyond that of classical set theory.

4.2.1 Fuzzy Models - Theory

The ability to summarize information plays an essential role in the characterization of complex phenomena. In the case of humans, the ability to summarize information finds its most pronounced manifestation in the use of natural languages (Zadeh 1973). Zadeh states, “each word *x* in a natural language *L* may be viewed as a summarized description of a fuzzy subset *F(x)* of a universe of discourse *X*, with *F(x)* representing the meaning of *x*.” In this way the language as a whole can be regarded as a system for assigning atomic and composite (words, phrases and sentences) labels to the fuzzy subsets of *X*. Values of a variable describing a fuzzy set may be atomic or linguistic. The values of the variable *color* are atomic (red, blue, green), while the values of the variable *setting* are composite. To illustrate, the values of the fuzzy variable *setting* might be expressed as *not rural*, *very rural*, *somewhat rural*, *rural but not very rural*, *more or less rural*. The values in question are sentences formed from the label *rural*, the negation *not*, the connectives *and* and *but*, and the hedges *very*, *somewhat*, and *more or less*. In this way, the variable *setting* as defined above is a linguistic variable whose labels are sentences in our everyday language. At the index level, this thesis attempts to summarize information about Town A's *setting* by investing the label *rural*. We also summarize information at the cluster level and at the indicator level. We use a linguistic variable to represent distance (very far, far, near, very near) and connectedness (well connected, moderately connected, barely connected).

4.2.2 Fuzzy Models – Notation

Fuzzy logic is actually a superset of conventional (Boolean) logic that has been extended to handle the concept of partial truth, or degree of membership. Conventional sets contain objects that satisfy precise properties required for membership. For example, the set of *H* integers from 6 to 8 is crisp (6, 7.5, 8). In fuzzy notation *H* is described by its membership function (MF), $m_H : X \rightarrow \{0,1\}$, defined as

$$m_H(y) = \{1;\ 6 \leq y \leq 8\}$$

$$\{0;\ \text{otherwise}\}$$

The crisp set H and the graph of m_H is shown in Figure 4.6A and 4.6B, where every real number is in H or is not. In logic values of m_H are called truth values with reference to the question, "Is x in H?". The answer is yes if and only if $m_H(y)$ =1; otherwise, no. In fuzzy sets there is not a unique membership function, but every function [0,1] is part of a fuzzy set. Fuzzy sets are always functions from a universe of objects defined by X, into [0.1]. This is shown in Figure 4.5, which illustrates that the fuzzy set is the function m_f that carries X into [0,1].

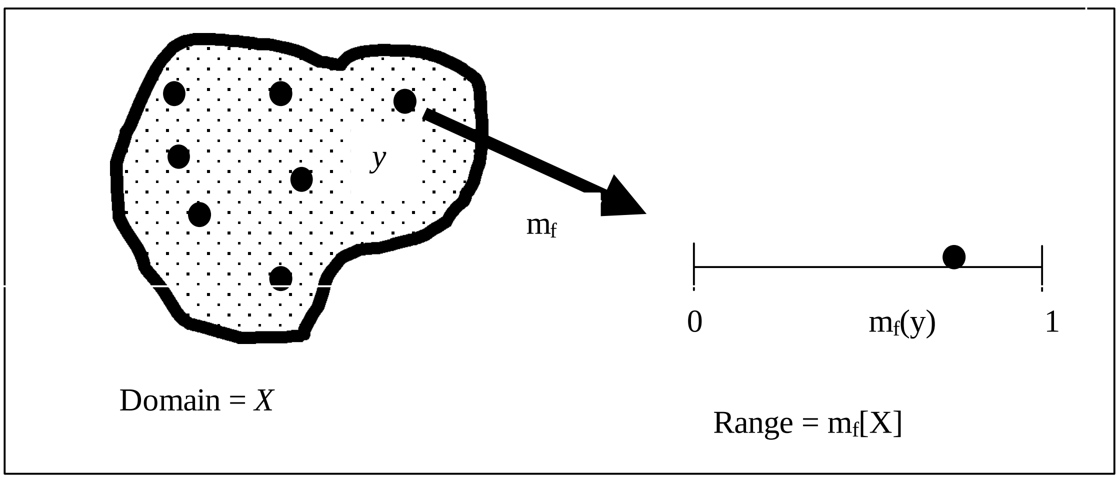

Figure 4.5: Fuzzy Sets are Membership Functions (Bezdek, 1993).

A fuzzy subset F is characterized by a MF, $m_f : X \rightarrow [0,1]$, which associates with each element, y, of X a number $m_f(y)$ in the interval [0,1], $m_f(y)$ represents the grade of membership of y in F. The *support* of F is the set of points in X at which $m_f(y)$ is positive. The *crossover* point is an element of X whose grade of membership in F is 0.5.

The modeler must decide based on the application and properties desired for a set F of real numbers what m_f should be. In this example we chose F to be a set of real numbers close to 7. Properties that might seem plausible for this *F* include (a) normality ($m_f(7)$ = 1), (b) monotonicity (the closer *y* is to 7 the closer $m_f(y)$ is to 1), and (c) symmetry (numbers equally far left and right of 7 should have equal memberships). Figure 4.6C and 4.6D demonstrate these instances respectively.

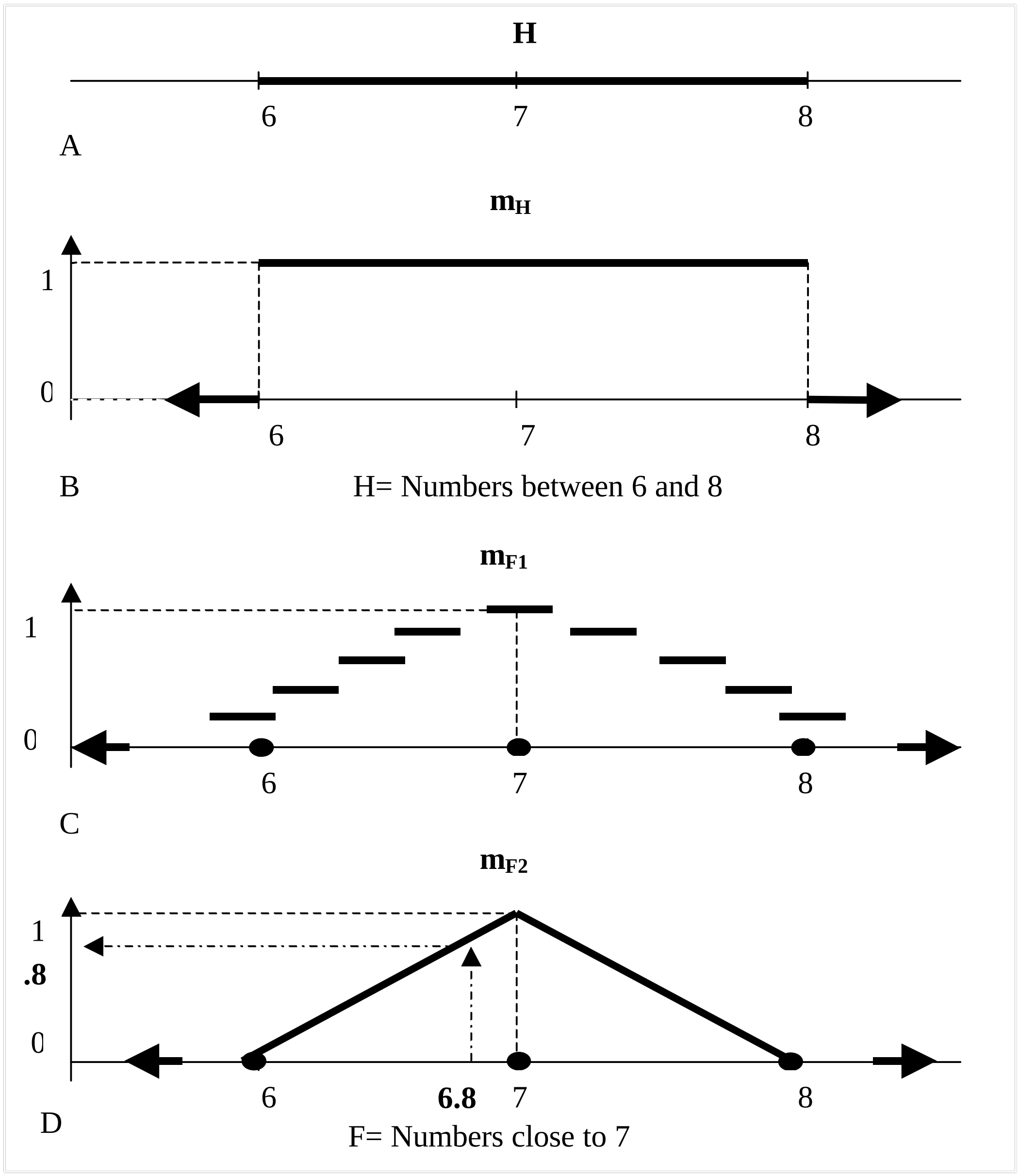

Figure 4.6A, 4.6B, 4.6C, 4.6D: Membership Functions for Hard and Fuzzy Subsets.

Membership functions for linguistic variables are modeled using the same concept of membership functions for other fuzzy subsets. An example of a linguistic variable may demonstrate its structure. Let us define *Distance*. The set of natural language expressions that *Distance* can take for this example is {‘near’, ‘moderate’, ‘far’}. These in turn are names of the following fuzzy sets:

‘near’ = {1/0.1, .0.8/0.2, 0.6/0.3, 0.4/0.4}
‘moderate’ = {0.3/0.3, 0.6/0.4, 1/0.5,1/0.6, 0.6/0.7, 0.3/0.8}
‘far’ = {0.4/0.7, 0.6/0.8, 0.8/0.9, 1/1}

The numerator represents the degree of membership and the denominator represents the element *y* of the set. Figure 4.7 is a pictorial representation for the linguistic variable Distance.

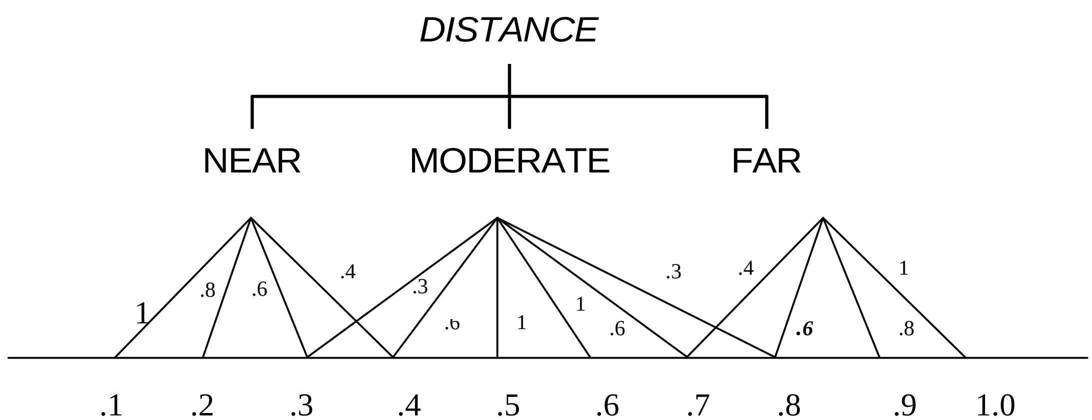

Figure 4.7: The Linguistic Variables Describing Distance.

If the user would like a linguistic representation of the distance metric, these are the linguistic variables and corresponding rules that apply in our model. These membership functions can be extended to accessibility as well, where distance equals accessibility. The linguistic variables become ‘very accessible’, ‘moderately accessible’, and ‘barely accessible’, instead of ‘near’, ‘moderate’, and ‘far’.

The distance metric is used for the access to service cluster. We can also generate a fuzzy set of linguistic variables to describe the degree of connectedness of a place. The set of natural language expressions that *Connectivity* can take for this example is {'well connected' 'moderately connected', 'barely connected'}. These in turn are names of the following fuzzy sets:

'well connected'= {1/0.1, .0.8/0.2, 0.6/0.3, 0.4/0.4}
'moderately connected'= {0.3/0.3, 0.6/0.4, 1/0.5,1/0.6, 0.6/0.7, 0.3/0.8}
'barely connected'= {0.4/0.7, 0.6/0.8, 0.8/0.9, 1/1}

The numerator represents the degree of membership and the denominator represents the element *y* of the set. Figure 4.8 is a pictorial representation for the linguistic variable *Connectivity*.

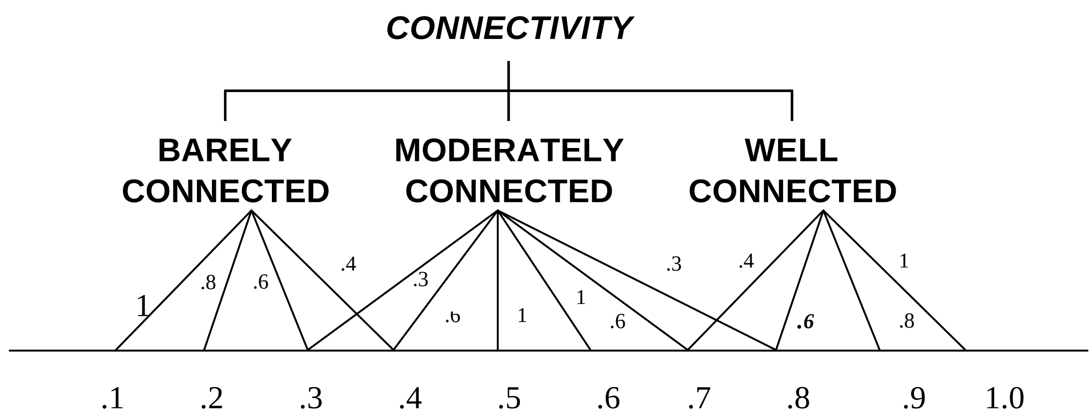

Figure 4.8: The Linguistic Variables Describing Connectivity.

If the user would like a linguistic representation of the detailed node and link topology, these are the linguistic variables and corresponding rules that apply in our model.

4.2.3 Fuzzy Rurality

To define rurality using fuzzy sets we must first chose a set of linguistic variables that appropriately describes the index. The set of natural language expressions in which the linguistic variable takes its values is not an unrestricted set of English phrases. The set can be a large set and requires a set of rules to accompany it. The most natural form of the rule uses Backus-Naur Form. Table 4.2 shows a small set of such rules. The terms <Hedge>, <Primary>, and <Fuzzifier> (non terminals of the grammar) take the roles that 'subject', 'verb', and 'object' do in the construction of sets of English sentences. The primary terms are the notions from which all other elements of the set are built and the hedges allow for the fine tunig of these terms. In this way linguistic variables can be put together to form a fuzzy phrase, or a linguistic expression. Just as 'subjects' and 'verbs' have certain roles to fill in the construction of English sentences, so too the primary terms and hedges fulfill certain functional roles in the construction of the set of possible natural language expressions that a linguistic variable can assume as its values.

```
<Rating>::=            (<Hedged Primary>|<Range Phrase>) - <Confidence>
<Confidence>::=        <Fuzzifier> CONFIDENT
<Range Phrase>::=      <Hedged Primary> TO <Hedged Primary>
<Hedged Primary>::=    <Hedged><Primary>|<Primary>
<Hedge>::=             NOT| VERY|FAIRLY| SLIGHTLY|
<Primary>::=           LOW| HIGH | MEDUIM
<Fuzzifier>::=         REASONABLY| BARELY| null
```

Table 4.2: BNF Notation for a Simple Set of Natural Language Expression (Schmucker 1984).

The set of hedges that the system designer has to pick from is almost endless. This model uses five linguistic variables, four of which contain hedges, and one a negation. Thus we have:

VERY RURAL
FAIRLY RURAL
MORE OR LESS RURAL
BARELY RURAL
NOT VERY RURAL

These five phrases appropriately describe our index of rurality, where very rural is the most rural and not very rural is the least rural on the index. Despite the fact that these variables and their rules are subjective choices made by the system designers a priori (Schmuchker 1984), they follow the rules of mathematical theory and logic.

A specific operation is performed on each hedge, which affects the membership values of some hedges. These operations include *concentration* (this reduces the degree of membership of all elements that are only partly in the set), *intensification* (this raises the degree of membership for elements greater than 0.5 and lowers the degree of elements lower than 0.5, thus modifying the steepness of the degree of membership curve), and *normalization* (this reduces fuzzy sets to the same base and insures that at least one element of the set has a degree of membership of one). Figure 4.9 shows how each hedge in the SRI assumes a separate and distinct curve.

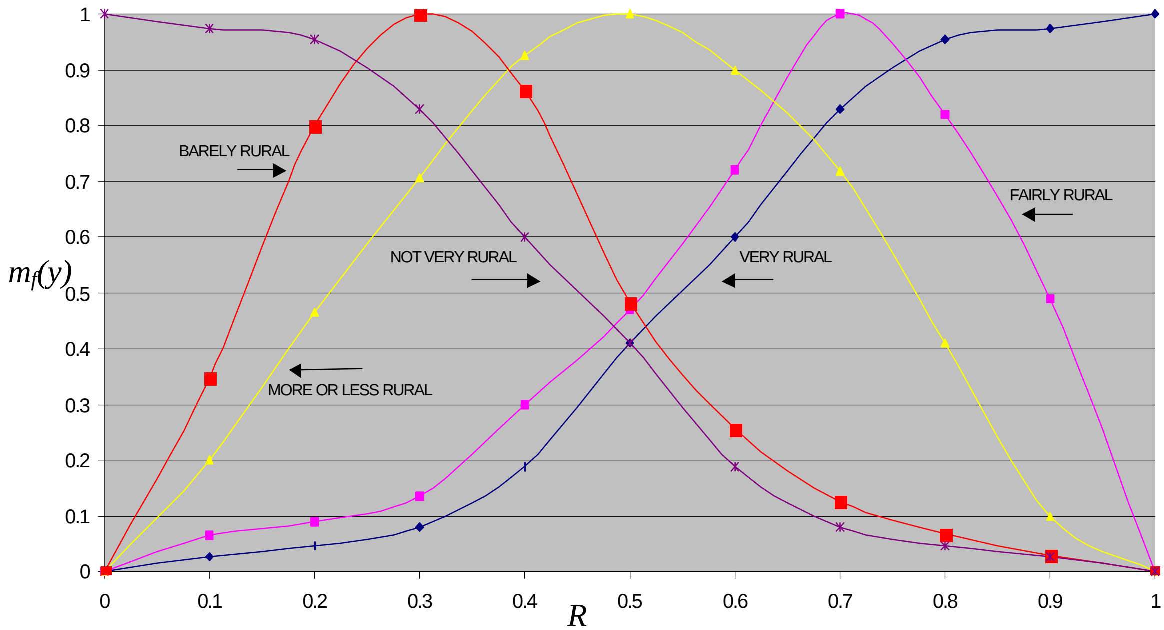

Figure 4.9: Hedges Acting on Rural.

Figure 4.9 represents the linguistic variables suitable for our final index of rurality. All of these results are possible depending on the chosen resolution. This process is explained in the following section.

4.3 Scale Issues

Anything we depict on a map results from a decision about the portion of obtainable information to be represented; how much area (e.g. a state), what range within the data (e.g. only airports with passenger service), and over what time span (e.g. 1900-2000). Similarly, we must specify how much detail to portray in space (e.g. divisions at the county level), attributes (e.g. high, medium, low percentage), and time (e.g. annually) (MacEachren 1995). An inherent property of objects in the world is that they only exist as meaningful entities over certain ranges of spatial scales. Thus, measuring the topological relation of a town and a police station is significant at the town level but may not be significant at the county or regional level, or counting the number of universities at the town level seems inappropriate (because it is not likely that every municipality has one) but might seem reasonable at the county or regional level. Changing spatial scales without first understanding the effects of such action can result in the representation of processes or patterns that are different from those intended. Before data can be integrated for problem solving, scale issues must be addressed. The challenge is to articulate the conditions under which scale-imposed constraints are systematic and to develop geographic models that are sensitive to scale-based variation.

Granularity is used in the field of artificial intelligence to express the idea that people observe the world by different grain sizes or granules (Hobbs 1990). Given a particular task, only certain objects will be of interest. This concept is important in our vision of defining rurality over different spatial scales. People may be interested in how rural a town is or how rural a certain region is. Perhaps a water company is looking to merge with another. The first company might be interested in number of customers and length of water line, but perhaps not the type of water pipe of the second company. Granularity depends on who is collecting the information and the intended purposes of it.

There are two important scales issues with respect to the SRI. The first is the size of the spatial unit to which the analysis is applied. The second is the discriminating power of the rurality indicators. When we talk about the spatial aspect of the SRI we refer to spatial resolution. Attribute data are discussed in terms of indicator granularity. This is the first time such a model has been proposed, one that enables a user to choose a spatial resolution, and an indicator level or granularity. The Census definition, because it is population based, does not let a user decide his or her particular indicators of interest. The user is bound by predefined political borders such as town, county, or city. The SRI lets a user choose any spatial configuration of an area and thus political borders do not bind the user.

4.3.1 Spatial Resolution

The SRI considers three spatial scales (*G1*, *G2*, *G3*), roughly associated with predefined government units such as town, county, and region (Figure 4.10). The user can chose the scale with respect to the purpose. A town clerk might be interested in one town's rural growth, while a regional manager might be interested in the region's rural growth. The user is responsible for the scale choice. However, there is a minimum spatial unit that must exist. For example, an area of 10 square meters is not an appropriate scale of study for the SRI. It is too small an area. Therefore a minimum spatial unit is required in order for the SRI to succeed.

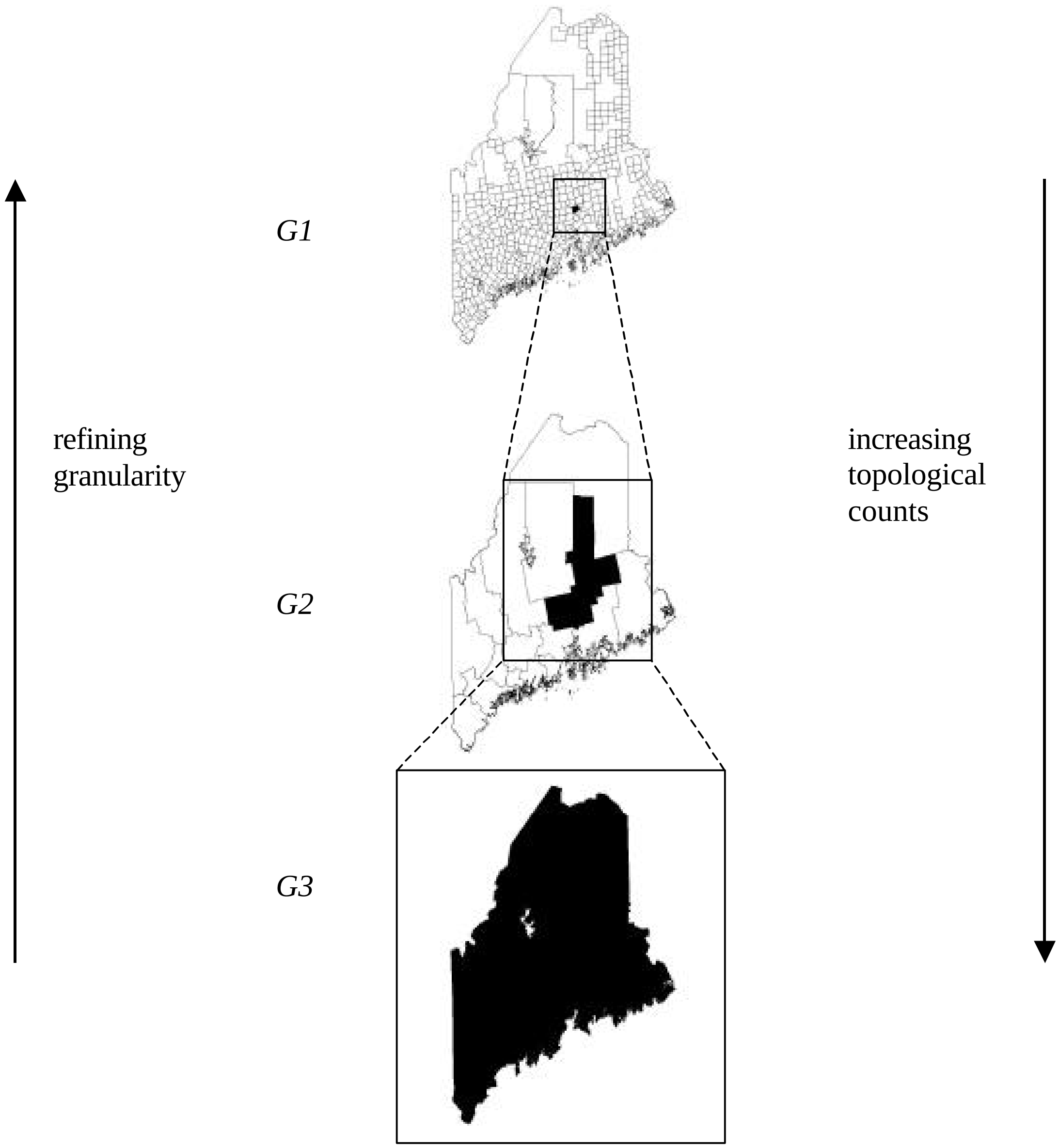

Figure 4.10: Levels of Granularity.

It is also possible in the SRI to aggregate at the *G1*, *G2*, or *G3* level. This function lets a user decide his or her spatial scale for more than one place and not have it be dictated by a political jurisdiction. A number of towns or counties can be aggregated or areas can be grouped together. Thus, Maine can be divided into any spatial partition (Figure 4.11). Because the indicators are topologically dependent and the geometric configuration and spatial location is known, the GIS is capable of aggregating the data under any spatial configuration. Therefore, if the town clerk is interested in a group of towns the SRI can capture the rurality of the grouping or if a circle with a radius of 20 miles is drawn around an area, this area is considered a spatial query. It is expected that as the spatial scale increases, the number of relationships increases, as seen in Figure 4.10. For example, we might expect one hospital to serve an area of 60 miles, therefore two hospitals to serve an area of 120 miles (if based on a linear relationship). If more than one count of a hospital is found over a small spatial scale, that place is rated less rural than if only one hospital is found over a large spatial scale.

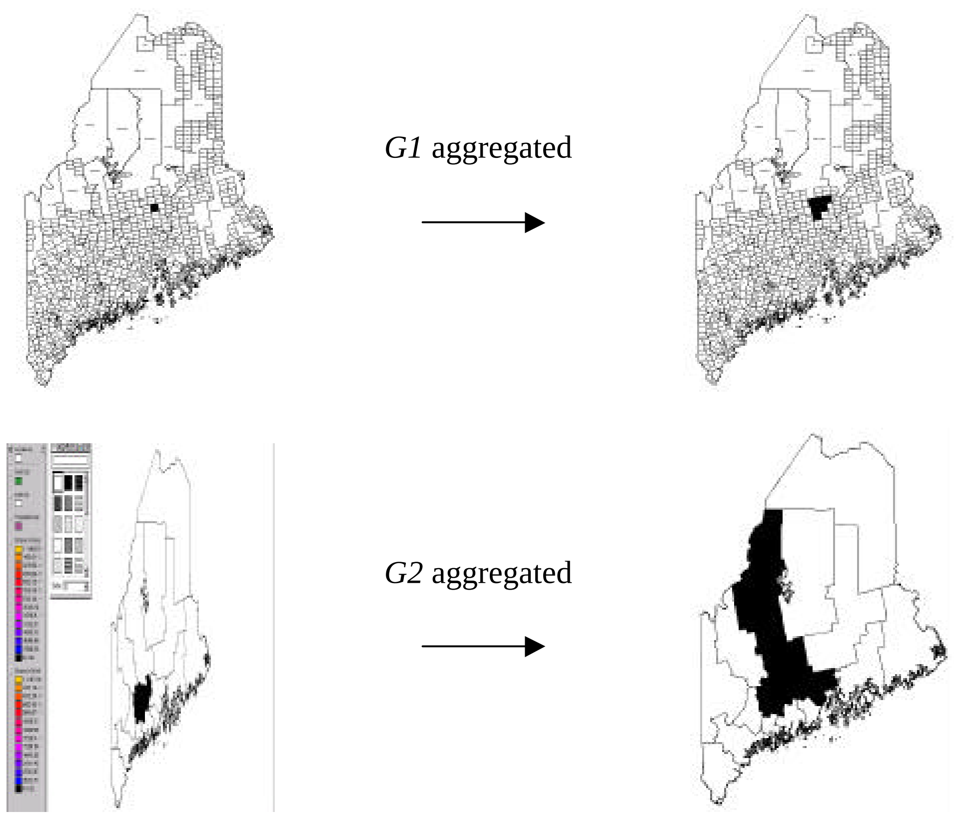

Figure 4.11: Aggregation of Town and County Level.

After the appropriate scale is chosen, the user must decide an appropriate attribute granularity. Is the town clerk interested in a general picture or a detailed analysis of the town? The following section explains the choice of granularity for our index.

4.3.2 Indicator Granularity

The importance of *I* (a user determined indicator) must be considered. Allowing the user to specify indicators of interest and the indicator granularity is similar to an approach of Hobbs' (1990)-- the simplification mapping function. This function maps the objects at one-grain size to a simpler set of equivalence classes of objects at a coarser grain size. For example, in our model a user can choose police offices or sheriff's offices depending on the desired spatial scale. We also use Hobbs' theory of articulation, which refers to the way that different granularities link to each other. For example instead of using the network distance to a hospital, we include different attributes of the indicator such as number of hospital beds and types of medical services. In this respect, the manner in which attributes of the indicator are implemented becomes important.

We offer the user a choice of indicator granularity. The simplest or coarsest level of granularity of the indicator is used to compare places on the basis of topology; either the place contains the indicator or it does not. The more detailed model determines the degree of connectivity based on link and node counts and node degree and accessibility based on the distance function. The ability to choose the indicator granularity is useful depending on the task at hand and fuzzy sets are employed at all levels of granularity to deliver a linguistic variable to the user.

The method used to calculate rurality for the coarsest indicator granularity is explained by equations 4.1 and 4.2 in Section 4.1.1. Either place, *P*, contains an indicator or it does not. This means that the distance to a hospital or the number of medical services offered by the closest hospital is not relevant. Only the spatial relationship of place to hospital is considered at this coarse level of indicator granularity.

But suppose we have two towns such as Portland and Bangor. We know that they both contain hospitals, education facilities, and connections to water and airports. They will be indexed equally using a coarse level of granularity. However, more discriminating power is achieved by increasing the granularity of the indicator. By examining Portland and Bangor with a finer level of granularity the result will be a different index measure of rurality for each place. It is at this point where the ordinal and ratio rankings play a key role. Recall equation 4.3 where the weights of the ordinal indicators were determined. Equation 4.4 expresses how the values of the weights are used for each indicator to represent a degree of rurality.

$$I = \frac{\sum_{i=1}^{n} N_i * W_i}{\sum_{i=1}^{n} N_i} \qquad (4.4)$$

where *I* equals the degree of rurality for the specified indicator, *N* is the number of counts for the indicator, *W* is the weight assigned to the type of indicator, *n* is the number of rankings of the indicator, and *i* is the rank of the indicator. Table 4.3 is an example using the Internet indicator.

	Indicator Rank	W	N	N*W
Cable	1	1/7	3	0.43
T3	2	2/7	7	2.00
T1	3	3/7	10	4.29
DSL	4	4/7	5	2.86
ISDN	5	5/7	7	5.00
Phone Line	6	6/7	75	64.29
Total			**107**	**78.9**

Table 4.3: Degree of Rurality Using the Internet Indicator.

Therefore the degree of rurality based on the Internet indicator for this town is equal to 0.73698. Equation 4.4 is applied to all of the ordinal indicators. The value for the ratio indicators is determined by assuming a threshold value (user defined) and normalizing the data between 0 and 1. For instance, at the town level we assume a distance from a hospital of 230 kilometers to be very rural. Data for a particular place is normalized against this threshold value. Table 4.4 illustrates the normalization process.

TOWN	Real Distance (meters)	Normalized
Brewer	6954.515	0.03
Orono	9426.680	0.04
Hampden	13855.563	0.06
Winterport	26090.111	0.11
Prospect	32421.053	0.14
East Millinocket	142650.938	0.62
East Machias	143492.078	0.62

Table 4.4: Normalizing Distance Data.

Degrees of accessibility fall between 0 and 1. This can then be fitted with the fuzzy graph to show how rural this particular indicator is for a particular town.

A finer level of granularity in the indicator results in a more descriptive characterization of rurality. Indicator data associated with node degree, and distance thus affects the level of granularity of our index. The user has the ability to choose the indicators and clusters and how they are going to be employed in the model. In this thesis we combine clusters using an additive approach. This method does not account for the interdependence between indicators. For example, we expect good access to schools to be a function of a good road network. The SRI adds the cluster results together without first examining spatial autocorrelation. This work does not examine other ways to combine the cluster, however, other approaches may be taken in the future.

4.4 Summary

This chapter has introduced the importance and characteristics of some spatial data models. A model based on spatial scale and indicator granularity has been identified and formalized. Fuzzy models contribute to our linguistic approach of describing rurality. Instead of using a number, we explain how rural a place is in semantics used everyday such as very rural, barely rural, or not very rural. Finally we identify the need to be able to apply the SRI over different spatial scales and levels of granularity. The user can choose a coarse level of indicator granularity built around topological relationships or a fine level of granularity founded on degrees of connectivity and accessibility.

An interface is developed in the next chapter to present the index, clusters and indicators. The components of the display are discussed and the results of three queries are represented. A number of examples conclude that a degree of rurality can be measured using good indicators.

Chapter 5

A User Interface Describing Rurality

Modern data graphics can do much more than substitute for statistical tables. Graphics are instruments for reasoning about quantitative information. Sometimes the most effective way to explore, describe, and summarize phenomena is to look at pictures. Visualization is an important component of any effort to understand, analyze, and explain phenomena on the surface of the earth. It is a method used to interpret data entered into a computer and to generate images from multi-dimensional datasets. This chapter introduces a *user interface* to represent our findings. This interface shows the contributions of each indicator cluster as well as the total aggregation of both clusters. The additional benefit to visualizing information in an interface is that other concepts can be represented such as time and scale. The interface allows time trends and spatial domains to be queried by the user.

We will define each component of the interface. Two complete examples will be given. The first uses a coarse level of granularity and a fine spatial resolution. The second uses a more refined level of granularity and the same spatial scale, the town level. Both examples will reveal the mathematical and the visual components of the SRI. A number of index results will also be given, without the interface.

5.1 Visualization

Growth in the field of visualization has gained critical attention in the fields of SIS, computer science, and image processing over the past decade. Visualization is an important component of any effort to understand, analyze, and explain phenomena on the surface of the earth (Buttenfield and Mackaness 1991). Geographical data and patterns are complex and change with time, resolution, and sampling strategy. The physical appearance of a forested area will vary according to seasons or traffic patterns will vary according to a hold up on a country road caused by construction or congestion on a well-traveled highway section. In these examples visualization can be used to identify and understand spatial, temporal and spectral pattern.

Highly graphical and interactive user interfaces have recently become popular. The design of graphical or visual displays forms an important component in understanding the role of visualization. A graphical user interface is developed in order to visually substantiate the mathematical findings presented in Chapter 4. The design employs interface metaphors, mappings from familiar settings onto an unfamiliar target domain (Bruns and Egenhofer 1997). Our design is similar to that of a “car dashboard”. We apply small multiples, a series of graphics showing the same combination of variables indexed by changes in another variable (Tufte 1984) as a visual enhancer. And we make use of direct manipulation, an interaction by which users see and manipulate objects. Our visual display incorporates spatial, temporal, and attribute data.

5.2 Overall Design

The interface for our index of rurality is divided into three components- spatial, temporal, and indicator. As a first step the user must choose a spatial scale or resolution. Next a time frame is chosen based on three given intervals a 10, 25, or 50 year time period. This allows the user to determine how rural an area was 10, 25, or 50 years prior to the present date of analysis. Finally the user can choose the indicators to be employed in the model (Figure 5.1). At this point the indicator granularity must also be chosen (Figure 5.2). All of these interactions require direct manipulation. Once the parameters have been set the model is run and the interface displays the result.

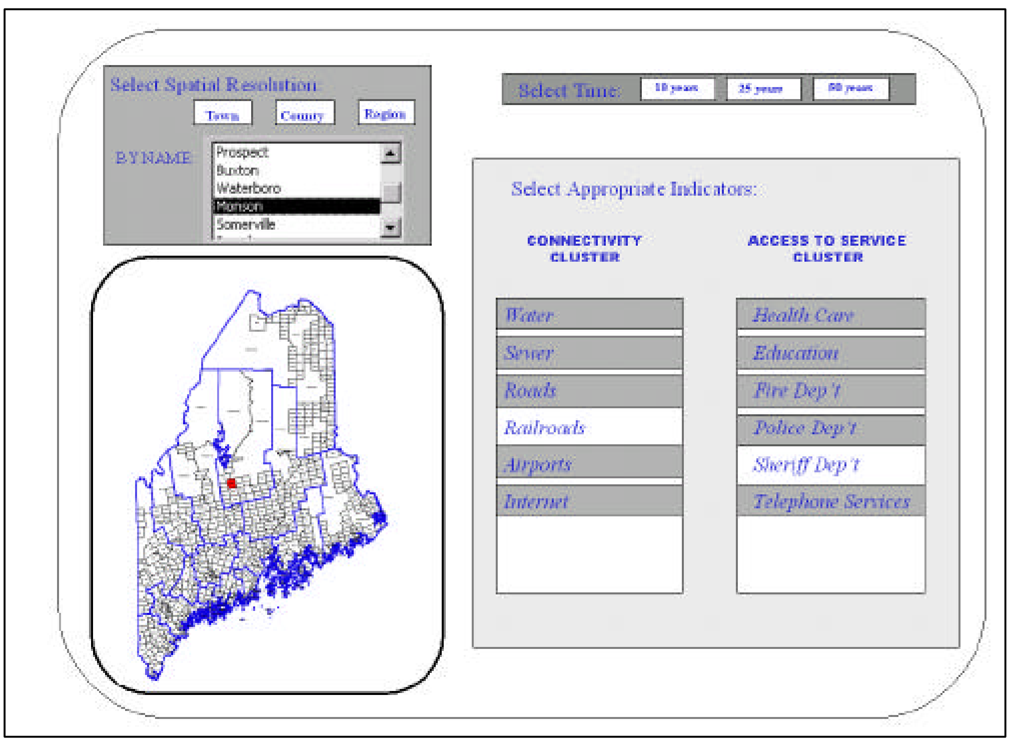

Figure 5.1: The Initial Interface Prior to Results.

5.3 The Spatial Query

The left side of the interface presents the spatial scale of the model. The user first must choose the desired spatial units. Predefined units (eg. towns, counties) are selectable from a drop down menu. A map (of the state) appears in the bottom portion of the interface. The user can then chose by name or by map the area of interest. The user can also choose to draw a circle around the area of interest, such that the query is not confined to political units. Once the area is chosen it is highlighted on the map and in the drop down list. In this example the town of Bangor is chosen, the scale button referring to town, the town name, and the area of the town on the map are all highlighted.

5.4 The Temporal Query

At the top of the right side of the interface is the temporal query. The user is given a choice of three time frames, 10, 25 or 50 years. We are providing a limited set of choices to the user for this query for two reasons - one, to help the user make a quick and simple decision and two because 100 year trends are too general (and time intensive) and less then 10 year trends are too specific and patterns are not yet recognizable. In this example the 50-year interval is chosen, and is highlighted.

5.5 The Indicator Level

The user also has the option to choose which indicators are relevant to the task. Section 4.3.2 described the reason for this. After the selected indicators have been chosen another menu appears (Figure 5.2). The level of granularity is listed and the user has the ability to choose which is of most interest for each indicator. The user can choose the coarsest level of granularity by selecting the button with the "1". In this case, the *contains* relationship is initiated. However, the user can refine the level of granularity by choosing node degree, distance, or other attributes of the indicator, all of which are sorted into predefined levels. Choosing a finer level of granularity in the indicator presents a more detailed picture of rurality. The technique for calculating indicators at different levels of granularity was explained in Section 4.3.2.

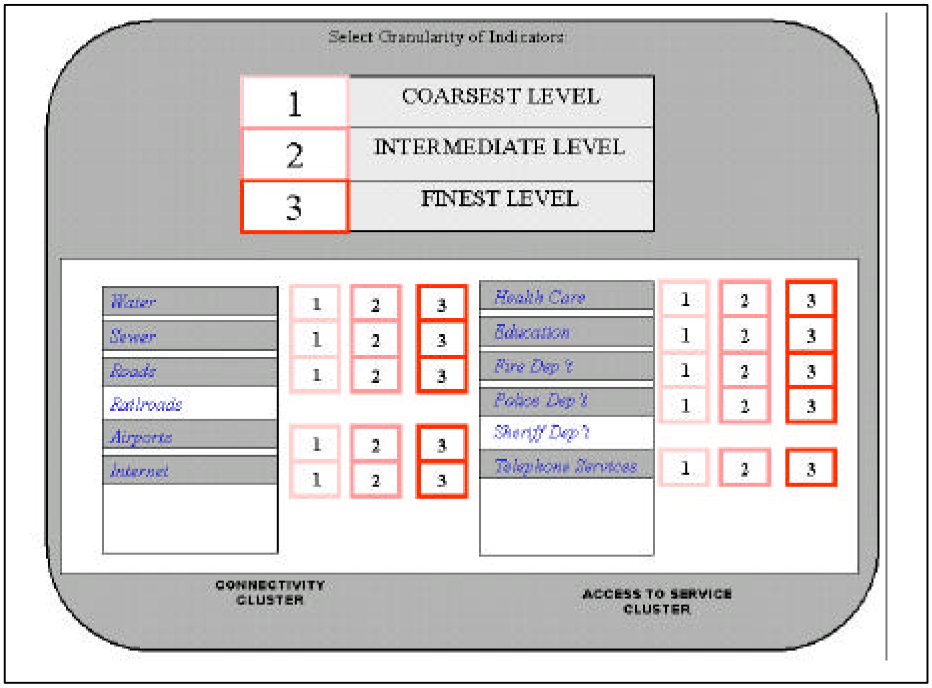

Figure 5.2: Indicator Granularity.

5.6 The Result

Finally, the actual result of the queries is displayed in a dashboard interface (Figure 5.4). We use this analogy in the following manner. It may be helpful to think of the clusters as the instruments in a car, signaling the overall performance and the performance of specific instruments. The dashboard has instruments such as an oil pressure gauge to monitor engine performance and a fuel gauge to show the resources available. These are similar to our two clusters, connectivity and access to service. It is most important for the driver to know what is wrong, so that corrective action can be taken in time, just as it is important for a town clerk or member of a steering committee to know what areas need attention so that appropriate grants can be acquired. The signals are often aggregated to avoid overwhelming the driver with information, but any problem can be traced back through the detailed instrument displays to identify specific information. Our indicators are aggregated for the same reason. We have borrowed this concept from the International Institute for Sustainable Development.

The SRI incorporates four components to the display (Figures 5.4 and 5.5):

1. The first is the overall rurality index. This shows the aggregated value given to the place after the cluster measurements are combined. A linguistic description appears at the top of the interface. The user can click on the name of the place chosen to see why this linguistic variable has been chosen. The fuzzy graph appears and shows the membership function of the chosen place. Color coding reinforces where the place falls on the index. As the color bar approaches green (to the right) it shows an area becoming more rural. The less rural the area the more gray it gets. The green is characteristic of rural environments depicting forest and farm spaces, while the gray represents colors of pavement and buildings typically found in urban environments. The outline of the boxes shows the overall temporal query results. A thick line shows the most recent result, a thinner line shows the rural index of *X* years prior, and a dashed line shows the result of *2X* years prior, where *X* is the resolution of the temporal query.

2. The heart of the model is the two displays, corresponding to the clusters of indicators that measure the connectivity and the access to services characteristics of the place. The design remains constant through the two frames, so that attention is devoted to shifts in the data, a defining characteristic of a small multiple. The needle approach is easily understood - a needle points to the value that reflects the current performance of that system. When the needle points to the green (far right) the combination of the indicators shows a completely rural area. If the needle points straight up then the area is more or less rural. Finally, if the needle is falling to the left, the area is characterized as not very rural.
3. Under each cluster is a list of the indicators associated with it (not shown in the examples). By clicking the indicator another graph can be displayed showing the fuzzy values assigned to that indicator, if the indicator granularity permits. This is important for any person with the desire to decompose the cluster and see the contribution of one indicator to the rurality of the place.
4. Two graphs reflecting the change in the clusters over time are the third element of the model. Under each cluster a graph shows the peaks and troughs of the cluster overtime. A rise of a cluster demonstrates that area's approach to rurality. The graphs are associated with the temporal query.

5.7 Examples

Two complete examples are chosen to demonstrate the model and interface. The first uses a coarse level of indicator granularity and a town query from the state of Maine. The second uses a more refined level of indicator granularity and maintains the town level spatial resolution (*G1*). The examples will reveal both the mathematical and the visual components of our model. Other scenarios are also discussed.

5.7.1 Query 1

A citizen is interested in a quick picture of how rural *Monson* is. She is not concerned with a detailed analysis; a simple yet formal description is key to her project. She chooses, a 50-year time interval, and the indicator granularity shown in Figure 5.3.

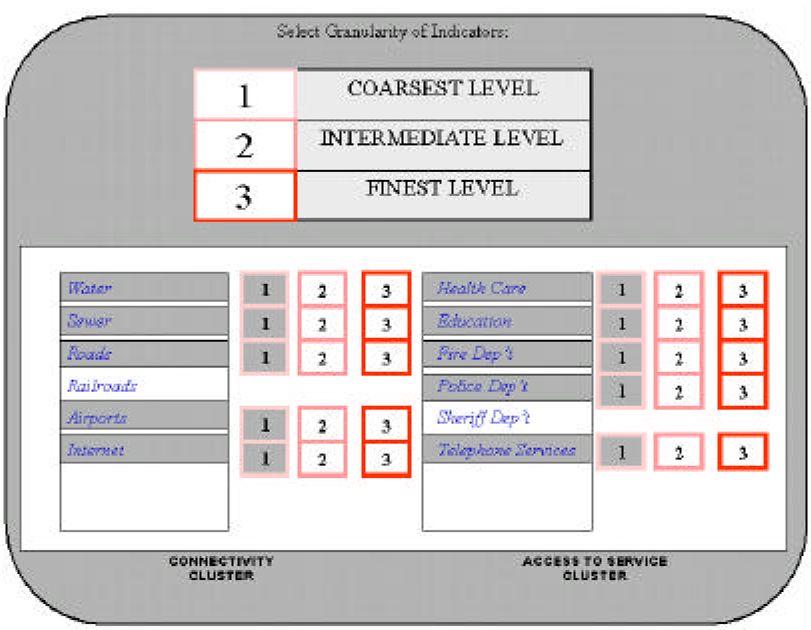

Figure 5.3: Monson Indicator Level of Granularity.

Because the citizen has chosen the coarsest granularity for all indicators, the SRI calculates a *contains* relationship for each indicator. If Monson contains connectivity to water then it is given a value of zero, if it does not a value of one is given. The same is true for sewer. If Monson is connected to an interstate (the highest rank for that indicator) it is given zero, otherwise one. The same is true for airports and Internet connectivity. The municipality of Monson does not have access to a hospital, a higher level of education (such as a university), a fire department, a police department, or a service center. Therefore the value given is one to all indicators in the Access to Service Cluster. Table 5.1 shows the contribution of each indicator to the overall index.

Indicator	Cluster	Value
Water	Connectivity	0
Sewer	Connectivity	1
Roads	Connectivity	1
Airports	Connectivity	1
Internet	Connectivity	0
TOTAL		**0.6**
Health Care	Access to Service	1
Education	Access to Service	1
Fire Dep't	Access to Service	1
Police Dep't	Access to Service	1
Telephone Services	Access to Service	1
TOTAL		**1**
GRAND TOTAL		**0.8**

Table 5.1: Index Results.

The fuzzy graph (Figure 5.5) shows the degree of membership of *R* for all hedges. It is clear that *Monson* is very rural. The degree of membership for the value 0.8 is almost 1, the largest for any of our linguistic variables. The final interface (Figure 5.4) displays the contribution of each cluster, the index value for *R* on a color bar, the time graphs and the spatial extent.

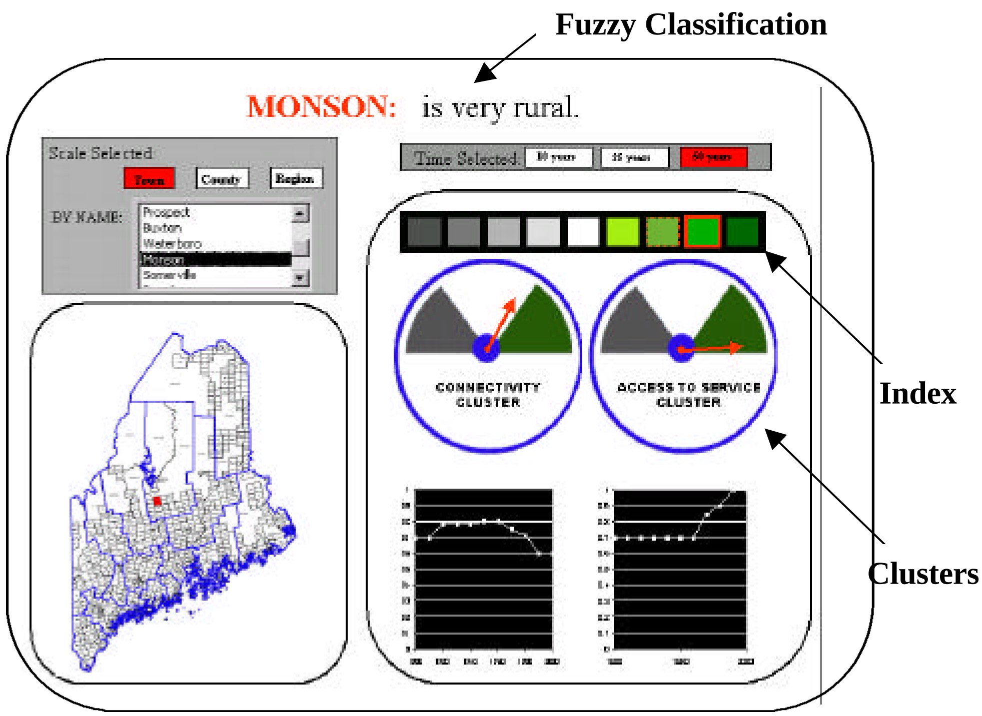

Figure 5.4: Query 1 Results.

5.7.2 Query 2

Bangor Hydro (BH) is looking to expand their business. They are in strict competition with the Portland gas company. A coarse level of granularity will prove insufficient for their purpose because it is too general. BH is looking for a detailed picture of how rural Bangor is, which they can then compare to Portland. BH chooses the finest level of granularity (level 3), a 10-year time period, and the indicators important to the task. The final result is shown in Figure 5.5. The total of the connectivity cluster is 0.072 and the total of the access to service cluster is 0.03, leaving a total of 0.051. From the fuzzy graph (Figure 5.6) we can conclude that Bangor is not very rural. It is also possible to see how quickly Bangor is becoming less rural from the two time trend graphs. BH might want to take the results of the time query into consideration when developing plans for the future.

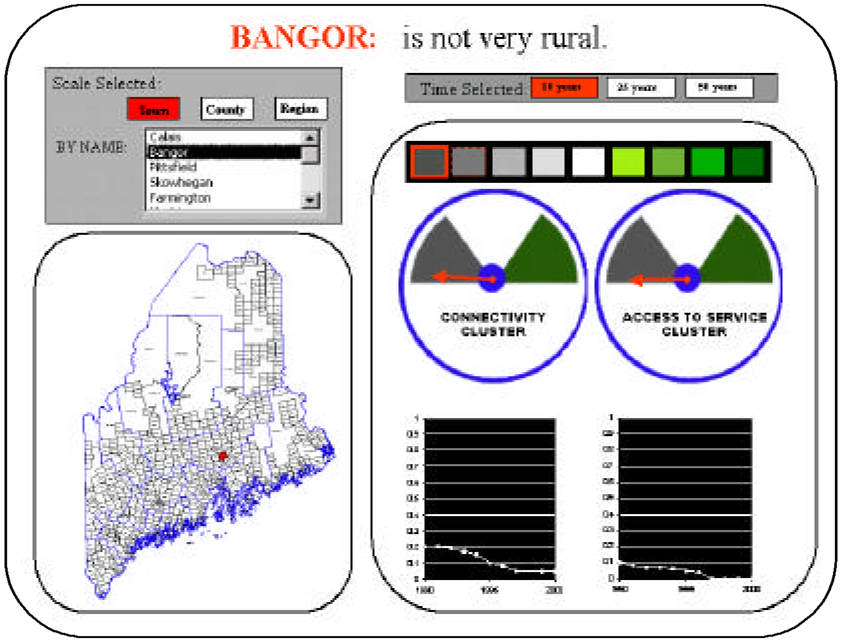

Figure 5.5: Query 2 Results.

Figure 5.6 shows the results of Query 1 and Query 2 on a fuzzy graph. The membership functions clearly describe why each town is linguistically described the way it is. One can also interpret what another index result might produce as a linguistic variable. For instance, if the final query result of Monson was 0.5 of a degree less, its *very rural* membership function would diminish to 0.9, and it would be termed *fairly rural* instead, with a membership function of 0.95. A dashed line shows the calculated value for both Bangor and Monson, while a dotted line shows what might happen if the result was 0.5 of a degree less for Monson.

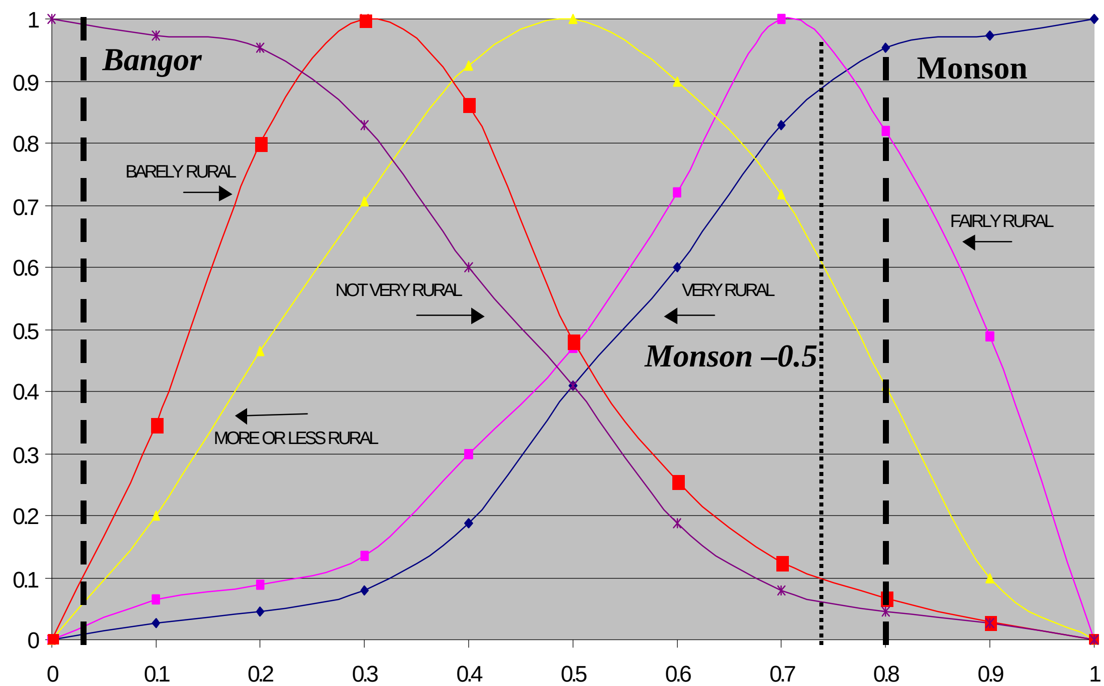

Figure 5.6: Fuzzy Rurality.

Table 5.2 is a list of many towns in Maine. Granularity level two was run on all towns (the intermediate level). This table shows how a degree of rurality can be determined for a spatial extent such as a town. This information is attained using the same descriptors, such that a service center for one town has the same attributes as a service center for another town. Degree of connectivity and degree of accessibility can also be used to compare these towns. For instance, Greenville has better access to service than Rockport, but in total Rockport is less rural than Greenville. These types of comparisons are necessary and have not previously been examined. Figure 5.7 shows the spatial distribution of these towns in Maine, against a backdrop of the interstate (a polyline), service centers (points), and those towns with high schools (flags).

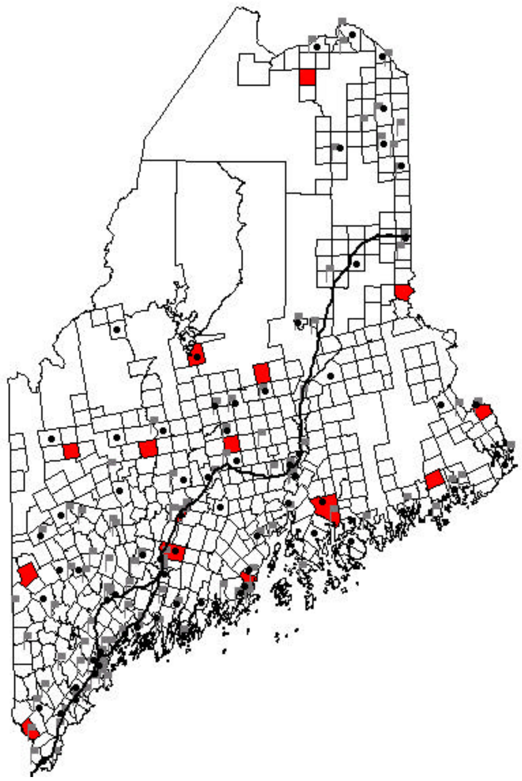

Figure 5.7: Map of Selected Towns and Distribution of Selected Indicators.

Town	Connectivity Cluster	Access to Service Cluster	Total
Portland	0.005	0.005	0.005
Augusta	0.153	0.090	0.122
Waterville	0.275	0.173	0.224
Ellsworth	0.567	0.255	0.411
Rockport	0.395	0.507	0.451
Calais	0.732	0.320	0.526
Greenville	0.871	0.328	0.600
Corinna	0.980	0.433	0.707
Centerville	0.980	0.664	0.822
Brownville	0.801	0.873	0.837
Berwick	0.940	0.943	0.942
Orient	0.980	0.925	0.953
Eagle Lake	0.980	0.970	0.975
Embden	0.980	0.980	0.980
Lovell	0.980	0.980	0.980
Sandy River Plt	0.980	0.980	0.980

Table 5.2: A Comparison of Ruralities Across Maine.

Finally, we can compare our values with other rural indices such as the Census. Table 5.3 presents our results and the results of other indices. It is clear that while we can differentiate between levels of rurality, the other indices cannot. In the Census either a town is rural or it is not. This table also presents median income of our chosen municipalities. These results support our hypothesis that demographic components can be captured using explicitly spatial indicators. We can see that trends in degree of rurality follow demographic and socio economic trends. For example, as population decreases the degree of rurality increases, or as median income increases rurality decreases. (One drawback of the median income measurement however, is that data are skewed for sparse populations. For instance, Centerville's median income is unusually high. This might be attributed to a few members of the community in very high prestige jobs, instead of to a large number of the community in medium to high prestige jobs.)

Town	Our Total	Census Definition	Population (1997)	Median Income ($) (1997)
Portland	0.005	**Urban**	62,239	$31,891.00
Augusta	0.122	**Urban**	19,544	$27,260.00
Waterville	0.224	**Urban**	15,815	$23,839.00
Ellsworth	0.411	**Urban**	6,301	$28,091.00
Rockport	0.451	**Urban**	3,025	$41,516.00
Calais	0.526	**Urban**	4,038	$27,373.00
Greenville	0.600	**Rural**	1,897	$20,114.00
Corinna	0.707	**Rural**	2,122	$26,306.00
Centerville	0.822	**Rural**	30	$43,750.00
Brownville	0.837	**Rural**	1,491	$22,386.00
Berwick	0.942	**Urban**	6,334	$40,135.00
Orient	0.953	**Rural**	161	$10,625.00
Eagle Lake	0.975	**Rural**	794	$18,611.00
Embden	0.980	**Rural**	765	$23,958.00
Lovell	0.980	**Rural**	990	$27,490.00
Sandy River Plt	0.980	**Rural**	63	$26,667.00

Table 5.3: A Comparison of Definitions and Demographic Data.

5.7.3 Other Spatial Scales

The SRI can be extended to incorporate county and regional units as well. For instance, it is time for a development agency to apply for housing grants from the government. The agency must first provide a detailed description of the county applying for the grant. This rural index is used to show how the county has become more rural over time. If this is proven, the result is more funding for better homes for the citizens of the county. The agency would like all of the indicators to be used for this analysis and a 50-year time period, ultimately showing trends towards rurality that occurred prior to their last requested grant.

Another example might set the Boston region against the Augusta region. Two capital cities are vying for a Northeastern regional competition. The winner will be the area that proves to have increased its connectivity cluster over the past 25 years. The ability to choose a time resolution and a spatial scale is crucial to this bid.

5.8 Summary

This chapter has extended the model to include a visual representation of the findings from Chapter 4. Where Chapter 4 provided the mathematical foundations of the model, this chapter emphasizes the need for a visual display of our results. An interactive user interface is designed in order for the user to choose indicator granularities appropriate to the task at hand. The user must make three decisions before the results are displayed. The first task is to decide the spatial scale, the second the temporal range, and finally the number and granularity of the indicators must be determined. The final result of the queries produces a display characterizing the rurality of the selected spatial unit. The components of the display work together to inform the user of the results. Further analysis of each component is possible. The fuzzy graphs and corresponding linguistic variables will reveal the underlying metrics, if needed. We demonstrate our interface with two examples at the town level. A case scenario at a different level of granularity is presented to establish feasibility of the model at all scales.

The next and final chapter of the thesis summarizes the need for a spatial definition of rurality and our approach. It highlights the major findings of the thesis and includes topics for future work.

Chapter 6

Conclusions and Future Work

People's perceptions of rural areas differ considerably. Problems of interpreting official definitions and measurements of population in an increasingly mobile society – such as challenges posed by second homes, seasonal migration, and so forth affect previous social definitions of rurality. It is clear that a different dimension is needed; one that explicitly defines factors affecting the rurality of a place. As Tickamyer (1993) sees it, the usual rural-urban approaches are valuable for descriptions of variation, but less satisfactory as explanations for why rural differs from urban or how they are connected. We have followed her recommendation to incorporate the reconceptualization of space to a new representation of rurality. We have based our approach on spatial dimensions of rurality instead of traditional social and economic measures.

6.1 Summary of Thesis

As areas become more or less rural the importance of being able to capture this change in rurality without changing the definition of rural increases. The index proposed in this thesis is an initial step towards this goal. Allowing spatial dimensions of rurality to define the index, instead of relying on social factors, decreases the chances of changing definitions.

A model of rurality is proposed based on the ability to synthesize spatial structures of a place. We have systematically aggregated a variety of spatial indicators into two clusters using the technologies of a GIS. A final index is introduced, one that is useful to decision-makers and others specifically on regional and national scales. Building on previous definitions that designated places as either rural or not rural, the SRI is extended to incorporate fuzzy categories of rurality. This more comprehensive approach allows an area to be assigned a degree of rurality. Linguistic variables represent fuzzy categories with underlying numerical rankings. The SRI incorporates spatial, temporal, and attribute shifts in granularity something other definitions fail to appreciate. Finally a simple user interface is developed. The ability for a layperson to quickly and methodically answer questions such as how rural is an area, how has it changed over *X* number of years, and how does it compare to another town is key to the SRI formulation.

6.2 Results and Major Findings

The major results of this thesis are:

- *Traditional definitions of rural are not satisfactory. Furthermore, studies of rurality take for granted these definitions and do not attempt to understand the implicit meaning of rurality.*

Traditionally "rural" has been defined on the basis of place of residence of its population. However, we may get a very different picture of social and economic conditions of a person who lives in a rural area but commutes to the periphery of a large metropolitan area. Do we compare this person to her neighbors or to her work colleagues? Defining rural based on social and occupational conditions has become ineffective in our increasingly mobile society. Definitions of rurality are also context dependent meaning one agency, the Census Bureau, has a slightly different method of determining rural areas as the Department of Agriculture, another agency. Choosing one definition over the other can leave over one million people in a residual category. Finally previous definitions are not static. These definitions, being population dependent, change over time as the global population increases exponentially.

There is an enormous body of work on issues relating to the rural arena (journals, task forces, conferences), since rural is already deemed the area of study; anything that takes place within this region is fair game for study. This includes studies of rural community families, organizations, and institutions as well as rural socialization, education, politics, religion, and rural demography. This research ranges from community studies concerned with particular locales to large-scale national activities. Finally, there are rural-urban comparative studies in which the purpose is to delineate differences in any of the preceding topics. All of these approaches result in inventories of information about places already defined as rural. But since they are predicated on a notion of rurality, they do not offer much guidance for understanding the nature of rurality, places that are rural, or how these components of human society construct rural life and are constructed by it.

Recent workshops (1990s) discussed the need for more relevant concepts and for better measures of conditions defining rural areas (Killian 1993). By considering previous studies and recognizing the need for a new dimension to be studied we have developed a new approach to defining rurality. We can thus construct rural life around our findings.

- *An area does not have to be either urban or rural. We can describe an area in degree of rurality by using the notion of fuzzy sets.*

Former definitions of rural described an area as being either rural or urban, either a metropolitan area or a non metropolitan area. The SRI takes away these Boolean sets and employs fuzzy sets to describe an area. Understanding rurality by degree of rurality allows comparisons to be made within rural areas. Where previous definitions do not make distinctions between one rural area and the next, the SRI captures possible subtle differences in rurality. We can examine a specific indicator's contribution or one of the cluster's contributions to an area's rurality. The result is an extension to classical set theory. We utilize the notion of linguistic variables characteristic of fuzzy sets to help describe how rural an area is. By using fuzzy measures to represent vagueness in everyday life, we can further investigate a degree of rurality.

- *An index of rurality can support spatial, temporal, and attribute based queries enabling a user to decide the granularity of her model.*

Other models of rurality do not capture granularity. We have proposed a method that permits the user to choose her level of analysis, spatially, temporally, or with attribute data. The choice of scale depends on the portion of obtainable information to be included, how much area (e.g. a town), what range within the data (e.g. only airports with passenger service), and over what time span (e.g. 1900-2000). These issues were not addressed in previous definitions. The SRI enables the user to choose what granularity best fits her particular task.

- *A visual interface is beneficial in measuring rurality. It enables a user to select her model, and to realize the results in a graphical display.*

It is important to keep in mind who will be using this model. It is not proposed for statisticians or mathematicians, it is for the layperson, the policy maker, and the town clerk. This group of people is not interested in the mathematical formulas behind the model but rely on the outcome for planning, policy, and curiosity. Thus, a simple model allowing querying capabilities is important. A visual interface makes the index easy to understand by displaying the composite pieces necessary to determine the rurality of a place, especially for the layperson.

6.3 Future Work

This thesis has contributed a new definition to existing ones of rurality. We have explained the need for a spatial dimension to replace social and economic based components. However additional key areas can be investigated.

1. Despite the urgency of a satisfactory explanation for rural spatial development, the disciplines and theories charged with dealing with it have only begun to spatialize the study of rural development (Tickamyer 1993). Now that the need has been iterated it is the responsibility of these disciplines to eagerly embrace the idea of a spatial rurality. Rural sociologists, mathematicians, government bodies, and spatial engineers should work together to further develop the ideas of this thesis. All of these perspectives are instrumental to a well-defined theory of rurality. Figure 6.1 charts out dominant fields that should realize their responsibility in initiating a universal spatial definition of rurality.

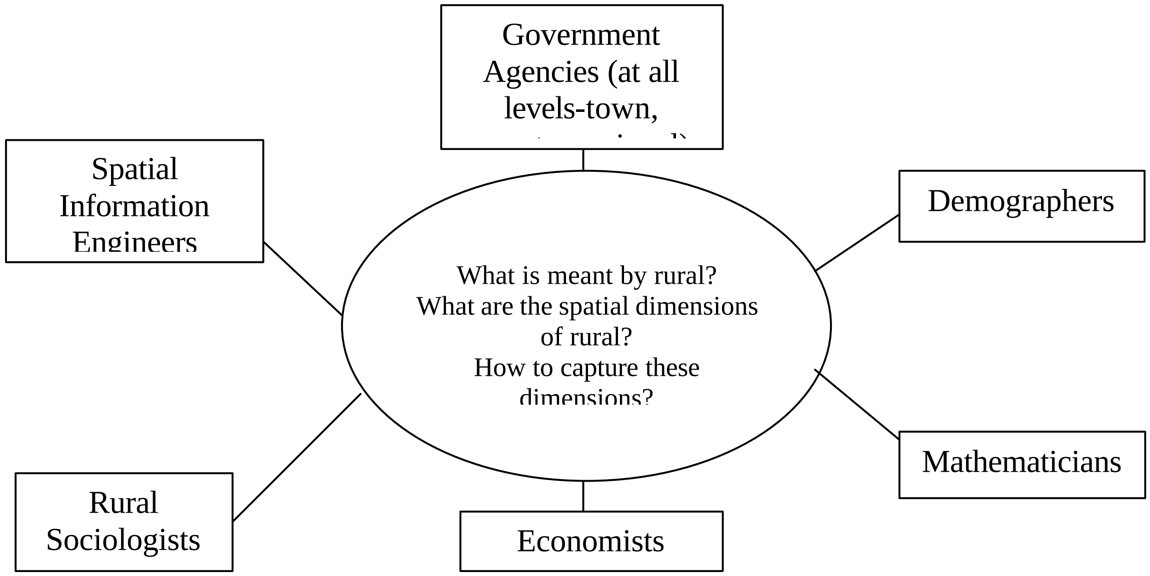

Figure 6.1: Key Disciplines Needed for an Extended Spatial Definition of Rurality

Once a universal working analytical tool to explain rurality is developed, policies and programs to alleviate ensuing problems must be instigated. This is a daunting task that is beyond the scope of this thesis.

2. The second question is whether the ways we think of spatial organization (such as town, county, region) remain adequate for capturing the emerging geography of rurality. New concepts such as edge cities and new delineations such as labor market areas have been developed in the past decade that may be of considerable use in our attempt to understand rurality. The issue of data collection and accuracy becomes significant if these types of spatial organizations grow to a dominant level. Already the labor market area delineation is employed across the country. The SRI must be

extendable to these different concepts of spatial organizations. The way our indicators are aggregated and our querying capabilities deserve further investigation, however, our topological or metric based model can serve as a basis for modeling the rurality of these places.

3. Our model focused on information from the state level, specifically Maine. An extended model, federally, or even globally deserves further exploration. It would prove beneficial to agencies such as the United Nations if we could compare the rurality of two countries such as Australia and Italy, or towns within those countries using spatial components. Other indexes are used globally to compare sustainability, environmental issues, and gross national product. To be able to compare the rural nature of two countries is beyond the scope of this thesis, but an important issue of globalization.
4. We have proposed a number of indicators in our two clusters. These are the basic components needed for a spatial definition of rurality. However other important spatial components exist. These might include – acreage of farmland, cellular phone coverage, access to cable lines, newspaper delivery routes, and garbage collection coverage. Data availability and collection plays a key role in determining indicator viability.

These are all important topics that deserve further analysis but are beyond the scope of this thesis. The spatial index of rurality provides the basis for further investigations. Because current definitions of rural are based on social and economic factors they are lacking in their spatial approach. The index proposed in this thesis is an important first step towards a spatial definition of rurality.

Bibliography

Bezdek, J. (1993). Editorial – Fuzzy Models What are They, and Why? *IEEE Transactions on Fuzzy Systems.* Vol 1 (1) pp 1-5.

Beard, M.K. 1994. Accommodating Uncertainty in Query Response. Proceedings 6th International Symposium on Spatial Data Handling, Edinburgh, Scotland p. 240-253.

Bradshaw, T.K. and B. Muller (1998). Impacts of Rapid Urban Growth on Farmland Conversion: Application of New Regional Land Use Policy Models and Geographical Information Systems. *Rural Sociology*. Vol 63 (1) pp 1-25.

Brown, D. (1998). Will Increased Highway Funding Help Rural Areas? *Food and Rural Economics Division, Economic Research Service, U.S. Department of Agriculture. Agriculture Information Bulletin Number 753.*

Bruns, T. and M. Egenhofer (1997). User Interfaces for Map Algebra. *Journal of the Urban and Regional Information Systems Association*. Vol 9 (1) pp 44-54.

Buttenfield, B. P. and W.A. Mackaness. (1991). Visualization in Longley, P.A., Goodchild, M.F., Maguire, D.J., and D.W. Rhind (Eds) *Geographical Information Systems Volume 1 Principles and Technical Issues*. pp 427-443. John Wiley and Sons, Inc. New York, NY.

Byers, A. (1996). Communities Address Barriers to Connectivity. *Rural Clearinghouse Digest on Rural Telecommunications*. Vol 3 (1) pp 1-7.

Cleland, C.L. (1995). A Measure of Rurality. *Papers and Proceedings of Applied Geography Conferences.* Vol 18 (183), 183-189.

Curry, M. (1996). *The Work in The World: Geographical Practice and the Written Word*. University of Minnesota Press, USA.

Curry, M. and M. Eagles (1998). *Place and Identity in an Age of Technologically Regulated Movement.* http://www.ncgia.ucsb.edu/varenius/place/description.html

Densham, P.J. and G. Rushton. (1996). Providing Spatial Decision Support for Rural Public Service Facilities that Require a Minimum Workload. *Environment and Planning B: Planning and Design*. Vol 23 pp 553-574.

Dickinson, G.C. (1973). *Statistical Mapping and the Presentation of Statistics*. Edward Arnold, London, UK.

Egenhofer, M.J., Clementini E., and P. di Felice (1994). Research Paper. Topological relations between Regions with Holes. *International Journal of Geographical Information* Systems. Vol 8 (2) pp 129-143.

Fitchen, J.M. (1991*). Endangered Spaces, Enduring Places. Change, Identity, and Survival in Rural America*. Westview Press, Inc. Boulder, CO.

Fuguitt, Glenn V., (1985). The Nonmetropolitan Population Turnaround. *Annual Review of Sociology*. Vol 11 pp 259-80.

Fuguitt, Glenn V., et al. (1989). *Rural and Small Town America*. New York: Russell Sage Foundation.

Goodchild, M.F. (1996). Geographic Information Systems and Spatial Analysis in the Social Sciences. In: Aldenderfer, M. and H. D.G. Maschner (Eds*) Anthropology, Space, and Geographic Information Systems* pp 251-241. Oxford University Press, Inc. Oxford, UK.

Hornsby, K. (1999). *Identity-Based Reasoning About Spatial-Temporal Change*. Ph.D Dissertation, Univerity of Maine.

International Institute for Sustainable Development (2000). *Measurement and Indicators for Sustainable Development.* http://iisd.ca/measure/default.htm

Jacob, S. and A.E. Luloff (1995). Exploring the Meaning of Rural Through Cognitive Maps. *Rural Sociology*. Vol 60 (2) pp 260-273.

Jordan, T. et al (1998). An Affordance- Based Model of Place in GIS. In. Poiker, T. and N. Chrisman (Eds) *Proceedings 8th International Symposium on Spatial Data Handling*. Simon Fraser University, BC, Canada.

Keddie, P. and A. Alasia (1999*). The Socio-Economic Diversity of Rural and Small Town Southern Ontario: A Comparitive Profile*. Sustainable Rural Communities Research Program. University of Guelph – OMAFRA.

Killian, M.S. (1993). Demographic Aspects of the Changing Rural Labor Force: Report from the Work Groups in: Swanson, L.L., and D.L. Brown (Eds.), *Population Change and the Future of Rural America A Conference Proceedings*. pp 29-44, Agriculture and Rural Economy Division, Economic Research Service, U.S. Department of Agriculture. Staff Report No. AGES 9324. Washington D.C.

King, L.J. (1984). Central Place Theory. Sage Publication, Inc. Beverly Hills, CA.

Leduc, E. (1997). Defining Rurality: A General Practice Rurality Index for Canada. *Canadian Journal of Rural Medicine*. Vol 2 (2) pp 125-136.

Liggett, H. and D.C. Perry (1995). Spatial Practices: An Introduction. In: Liggett, H. and D.C. Perry (Eds) *Spatial Practices* pp 1- 32. Sage Publications, Inc. New York, NY.

Lobao, L. (1996). A Sociology of the Periphery Versus a Peripheral Sociology: Rural Sociology and the Dimension of Space. *Rural Sociology*. Vol 61 (1) pp 77-102.

MacEachren, A. M. and C.A. Brewer. (1998). Visualizing georeferenced data: representing reliability of health statistics. *Environment and Planning A.* Vol 30 pp 1546-1561.

MacEachren, A.M. (1995). *How Maps Work Representation, Visualization, and Design.* The Guilford Press, Inc. New York, NY.

Maine State Planning Office. (1998*). Reviving Service Centers Volume 1. Report of the Task Force on Service Center Communities* pp 1-29.

National Association of Development Organizations (NADO). 1999. "Focus on Rural America." *Economic Development Digest* 10(6), April 1999.

Okoye, J.C. (1991). Spatial Manifestations of Rural Development in Nigeria. *Journal of Environmental Management.* Vol 32 pp 93-102.

Pahl, R.E. (1966). The Urban-Rural Continuum. *Sociologia Ruralis*. Vol 6 pp 299-327.

Robinson, A., Sale, R. and J. Morrison. (1978). *Elements of Cartography*. John Wiley and Sons, New York, NY.

Rodriguez, Maria Andrea. (2000). *Assessing Semantic Similarity Among Spatial Entity Classes.* Ph D Dissertation, University of Maine.

Rourke, J. (1997). In Search of a Definition of “Rural”. *Canadian Journal of Rural Medicine.* Vol 2 (3) pp 113-117.

Rural Profile of Arkansas 1997*: A Look at Economic and Social Trends Affecting Rural Arkansas.* University of Arkansas – Agricultural Experiment Station and Cooperative Extension Service – and the Office of Rural Advocacy.

Sarbit, L.A., and B. Greer- Wootten. (1980). Spatial Aspects of Structural Change in Central Place Systems: Southern Manitoba 1961- 1971. *Geographical Monographs, No. 4.* York University, Canada.

Schmucker, K. (1984). *Fuzzy Sets, Natural Language Computations, and Risk Analysis*. Computer Science Press Inc. Rockville, ML.

Strover, S. (1999). *Rural Internet Connectivity P99-13.* . Rural Policy Research Institute. Iowa State University.

Strover, S. and L. Berquist. (1999). *Telecommunications Infrastructure Development: The State and Local Role P99-12*. Rural Policy Research Institute. Iowa State University.

Tickamyer, A. (1993). The Uneven Spatial Development of Rural Sociology in: Swanson, L.L., and D.L. Brown (Eds.), *Population Change and the Future of Rural America A Conference Proceedings*. pp 29-44, Agriculture and Rural Economy Division, Economic Research Service, U.S. Department of Agriculture. Staff Report No. AGES 9324. Washington D.C.

Tufte, E. (1983). *The Visual Display of Quantitative Information*. Graphic Press, Cheshire, CN.

Urban and Rural Definitions and Data (1990). *1990 Census of Population and Housing*, "Population and Housing Unit Counts," CPH-2-1.) http://census.gov

Wilkinson, K. P. (1991). *The Community in Rural America*. Under the Auspices of the Rural Sociological Society. Contribution in Sociology, Number 95. Greenwood Press. London, UK.

Willits, F. K. and R.C. Bealer (1967). An Evaluation of a Composite Definition of "Rurality". *Rural Sociology*. Vol 32 (2) pp 165- 177.

Woodcock, C.E. and S. Gopal (2000). Fuzzy set theory and thematic maps: Accuracy assessment and area estimation. *International Journal of Geographical Information Science.* Vol 14 (2) pp 153-173.

Worboys, M. (1995). *GIS A Computing Perspective*. Taylor and Francis Ltd. Bristol, PA.

Wrigley, N., T. Holt, D. Steel, and M. Tranmer (1996). Analyzing, modeling, and resolving the ecological fallacy in: P. Longley and M. Batty (Eds.), *Spatial Analysis: Modeling in a GIS Environment*. pp 25-41, John Wiley and Sons, New York, NY.

Zadeh, L. (1973). Outline of a New Approach to the Analysis of Complex Systems and Decision Processes. *IEEE Transactions on Systems, Man, and Cybernetics*. Vol 3 (1) pp.28-44.

Biography

Gillian AvRuskin was born and raised in Toronto, Canada. She completed all 5 years (a Canadian tradition) of high school at Branksome Hall in Toronto. She obtained a B.A. in Geography with a minor in Anthropology at McGill University in Montreal, Quebec. Before arriving in Maine, Gillian took a year away from academia to travel in Southeast Asia, Australia, and New Zealand. She began her Masters at the University of Maine in the fall of 1998.

Gillian AvRuskin is a candidate for the Master of Science degree in the Department of Spatial Information and Engineering, from the University of Maine May 2000.