

Geo-Coding Survey Truck Route Data: GIS Analytical Applications

Tosmai Puenpatom
tpuenpat@mail.wsu.edu
Ph.D. Candidate

Eric Jessup
Eric_jessup@wsu.edu
Assistant Professor

**School of Economic Sciences
Washington State University
Pullman, WA 99164-6210
Tel: (509) 335-5558
Fax: (509) 335-1173**

Abstract

The application of Geographic Information Systems (GIS) has been widely accepted in a number of areas, i.e. transportation, utility network, land-use, city planning, and policy planning. Transportation policy planners and transportation network analysts, for instance, use the GIS software to analyze highway traffic and congestion problems in urban areas as well as planning for new highway infrastructure and maintenance. This paper is the first attempt in the freight transportation area to develop an automatic algorithm which allows identifying specific segments of highways used by each observed truck using the application of the GIS software.

This paper utilizes data collected from the Strategic Freight Transportation Analysis (SFTA) truck survey which examined truck movements in the state of Washington. The survey includes approximately 24,000 observations of trucks traveling within (or through) the state of Washington with information on origin city and state, destination city and state, truck type, truck weight, type of commodity on board and the list of highways utilized. With known origins, destinations and the list of highways used, Arc-Info may be utilized to specifically identify unique route segments used by each truck and assign geographic attributes to the route data.

Utilizing several different software features within ArcMap, including the Utility Network Analysis tool, data from a traditional relational database is input into the GIS software where several logical conditions are examined to identify unique route attributes associated with each survey response. The logical conditions examined include many attributes collected on the roadside questionnaire and help minimize error associated with the geo-coding process. This process is then automated and applied to the entire database.

Many different analytical capabilities related to freight highway flows are then possible and applicable once specific routes have geographic properties, several of which are provided in this paper. These include specific highway segment properties and traffic profile, including upstream/downstream contribution of freight flows by different highways and commodities throughout the highway network.

Introduction

Geographic Information Systems (GIS) have become a primary analytical tool for conducting research related to subjects with some geographical attribute, including among others, natural resource management (forests, soils, water, etc.) population demographics and/or transportation policy planning. The popularity of GIS is mainly due to the ability to evaluate/analyze multiple layers of geographical information that possess some common relationship amongst the different layers. For example, policy planners may be interested to know the relationship between specific highway segments throughout the transportation network and the degree to which different businesses rely or depend upon the highway infrastructure. Many statewide transportation departments routinely compile traffic counts on specific highway corridors but have limited information pertaining to the type of cargo or commodities moving throughout the highway network on freight vehicles. However, in states where origin-destination surveys have been implemented and data related to route and trip attribute details compiled, a GIS platform is an ideal tool for understanding the relationship between the highway infrastructure and the businesses which rely up the transportation infrastructure.

Truck traffic estimation, forecasting and freight commodity analysis has been an important component of transportation planning agencies for quite some time. This has been especially true in Washington State where resources have been allocated to improve freight mobility, via a multi-modal, statewide research and implementation effort known as the "Strategic Freight Transportation Study (SFTA)"¹. One aspect of this research effort is focused on developing statewide freight and commodity flow data and also analytical methods to evaluate different transportation infrastructure investment alternatives. This research paper develops the analytical capability of geo-coding truck route information using origin-destination survey data and then utilizing this tool to evaluate different empirical questions.

Objective

The primary objective of this research paper is to specifically describe the process for geo-coding route data and also providing examples of how this information may be utilized to evaluate different freight transportation issues. The paper is organized as follows. The next section describes the specific process and details regarding the Geo-Coding procedure. Following this section are applications and extensions of geocoding process and applications related to transportation planning and modeling. The final section concludes with summary findings and research extensions.

The Geo-Coding Process

This section begins with an outline of GIS applications in the transportation research area. It also familiarizes readers to GIS by introducing some technical terms used in the later parts of this section. This includes geodatabase and network characteristics. In addition, with a robust application of the Visual Basic Application (VBA) for GIS, we also disseminate some details of VBA to readers. Explanation of steps and algorithms of the geo-coding process are provided in more detail after these introductions. For interested

¹ For a full description of this research effort, go to www.sfta.wsu.edu.

readers, some parts in this section might be referred to additional references for more detailed information.

Greaves (2004) describes geo-coding as the process by which location information is assigned to an address so that it can be spatially referenced. The process requires at least a map of interested area, i.e. city, county, state, or country level, and information of either events or activities with spatial references. Typically, the spatial references are represented by a co-ordinate system, for example, (x, y) plane or latitude-longitude system. However, survey data with route information recorded in a sequence of highway segments is referenced as name of highways used in one trip and the name of an origin or a destination city. Generally, working with survey data is challenging because of the variation in interviewing procedure and errors in data input such as incorrect city name, unmatched city and state, and highway names not following consistent sequence patterns. Therefore, the simple references from survey data must be extremely carefully applied.

The first requirement is to acquire the highway network to be evaluated which is created from the highway network of a highway map layer. This can be done through geodatabase with the utilization of ArcCatalog. Geodatabase is an assembly of different map layers used in a GIS project. Under the geodatabase, different layers with different co-ordinate systems can be unified and related spatially. Consequently, adding new layers and editing current layers are easily completed.

The established highway network, the focal point of this analysis, consists of two features named *edge* and *junction*. An edge has two junctions at each end and a junction can be connected to any number of edges. A single highway, such as I90, is represented by more than one edge and many junctions which connect edges together. For instance, Figure 1 illustrates one segment of I90, from Vintage to McDonald, consisting of 4 edges called by ID as 167, 12256, 12257, and 12251. With edge IDs, the identification of a truck route is more accurate than by only identifying highway names.

FIGURE 1

A collection of more than 24,000 observations from the SFTA roadside survey makes a manual geo-coding process significantly inefficient. The use of VBA programming to automate the geo-coding process is applied effectively. VBA has become a competent alternative for the Arc-Macro Language (AML) in recent years. With the user-friendlier way of coding and the ability to apply with wide-ranges of software, VBA is remarkably well-known among GIS and non-GIS users/developers. Statements and interfaces are common types of analytical commands within VBA code. Statements, such as *if...then* and *do...loop*, make the coding task more convenient. Interfaces basically are functions carried out in GIS software.

The highlighted interface in this paper is *Utility Network Analysis Interface*, which fulfills a major role in finding and selecting specific paths for truck routes. The interface allows GIS users and developers to operate different tasks such as tracing the shortest path between two points (origin and destination), calculating total cost of a trip and alternatives, assigning direction of traffic flows, and enabling or disabling a highway

partially or thoroughly. Practically, the result from path tracing is not necessary to be the same as the route reported in the survey. To be more accurate, we have imposed constraints on the path tracing part by allowing the tracing on selected highways only. This can be accomplished because all features participating in a network have an *enabled/disabled state*, where enabled (disabled) state allows (disallows) tracing task to be done. Therefore, we enable only survey-reported highways and disable the rest. The result of the path tracing for one truck route is a set of edge IDs which are recorded into MS Access for the future analysis.

From the above synopsis, there are three main steps in the Geo-Coding process; constructing and preparing the highway network, writing VBA codes, and data challenges. The rest of this section presents these three steps with more details.

A. Highway Network Construction.

A user is required to create his/her geodatabase, or personal geodatabase, before creating the highway network. ArcCatalog is the software that supports this task. After obtaining the personal geodatabase, the user is able to import and store relevant map layers into the personal geodatabase. Moreover, the user's geodatabase assists any editing task on current map layers contained inside and adding new layers. In this study, we import a layer of Washington cities, a layer of Washington highways, and a layer of Washington zip code areas. Storing both the city layer and the highway layer in the same geodatabase unifies the spatially reference of both layers. Consequently, it is possible to utilize the *selection by location*, which is an important tool in finding the path between a starting point and an ending point.

The Washington highway layer, along with the GIS ArcInfo software by ESRI, is a predecessor of the highway network.¹ After importing this layer into the personal geodatabase, the user operates ArcCatalog software through the *Build Geometric Network Wizard* to transform the ordinary map layer into a network feature.² The newly created Washington highway network comprises of a great number of edges and junctions; 10,219 of edges and 11,505 of junctions for the state of Washington. Each highway is formed by edges, interchangeably with segments for the rest of this paper. For example, highway I5 inside the state of Washington (connecting the north border from British Columbia with the south border to Oregon) is formed by 72 edges, while highway I90 within Washington (connecting the east border from Idaho to Seattle) is formed by 45 edges. Each edge and each junction are identified by its Object ID, which can be used for future references.

The Utility Network Analyst Interface and the enabled/disabled property of network features are two key components for an accurate result of the path finding task. The first component refers to an operation within GIS ArcInfo designed to operate with network flows, especially water and/or electricity systems. This operation is equipped with several tasks, i.e. find path, find an upstream path to the source, trace downstream, find loops, etc. At this point, we implement only the find path task to this paper. By indicating a starting point and an ending point, the application of the find path task results in a shortest route between those two points. Figure 2 displays an example of using the find path task manually to the trip that originates from Idaho moving toward

Seattle on I90. However, results from the find path task are not necessarily the same as the route reported in the survey.

FIGURE 2

The path finding task requires that a user has two reference points on the network. Nevertheless, the information provided by the SFTA survey provides us only the origin city and the destination city which may or may not be on the Washington transportation network. We cannot perform the task directly without assistance from the *selection by location* function. We perform the function in order to find a starting segment and an ending segment on the network where the first is closest to the origin city and the latter is closest to the destination city. Therefore, we are able to operate the path finding task to trace truck routes.

Some highways' segments created from the network construction may not be appropriate to use for a starting or an ending point. Figure 3, for instance, shows that the segment is too long which overstates the result of the route tracing. If City A is the destination city, then the segment XZ implies that the trip stop at somewhere between X and Z. We solve this problem by manually editing the segment. We divide the XZ segment into XY and YZ segment, thus the route tracing is more accurate.

FIGURE 3

B. Writing VBA Codes.

The path finding performed manually in Figure 2 is operationally unfeasible when working with large route data set of more than 24,000 observations. Thus, automating this manual process becomes paramount when working with a large number of observations. This approach is to write VBA code that automatically traces truck routes from the survey data. The following are algorithms and logical ideas for the VBA codes.

Algorithms

- (i) **Acquiring information from survey data.** VBA for GIS has the ability to access data and information stored in Access tables through the *Table Interface*. The interface reads and stores records' city/state of origin, city/state of destination, highway names used in a given route. Accessing this information is very important for the following steps and algorithms.
- (ii) **Classify each record into 4 special route cases.** The trip route for Washington freight traffic may include routes other than those beginning at Washington cities for an origin and/or destination. On the contrary, the trip from outside, to outside the state of Washington, or the transit trip do not have an origin, or destination or both inside Washington. This becomes a problem for the route tracing task because there is a missing reference point or two. Hence, the solution for each case is not the same. For convenience, we write a condition to differentiate among those four cases by using *If...then Statement* provided by VBA. This will allow the implementation of different treatments for

each case separately. The basic idea of this algorithm is to separate trips into four cases corresponding to state of origin field and state of destination field in the survey table. If both fields report 'Washington', the software classifies this record as *Case 1*. If the state of origin field reports 'Washington' while the other reports something else, it is categorized as *Case 2*. If the state of origin field reports something else while the state of destination reports 'Washington', it is grouped into *Case 3*. The record is grouped into *Case 4* when both fields do not report 'Washington'.

(iii) **Identify starting and ending points.** This algorithm describes the way to specify an origin and a destination point for each case mentioned above.

- **Case 1.** The starting and ending point in this case are simply the origin city and destination city indicated on the survey record. From a number of cities on the map, we can select only two of them for each case through the *Queryfilter Interface*. However, cities are not directly on the highway network created from the highway layer. It is not possible to use cities directly as a starting or ending point. Fortunately, we have information on a list of highways used by a truck driver in each record. We know from how the data was input that the first-listed highway is where the trip starts and those listed subsequently follow the sequential pattern of the trip route. Our strategy is to assign the first-listed highway's nearest edge to the origin city as the starting point. Among several edges on the same highway, the nearest edge to the selected origin city can be identified through the *Spatialfilter Interface*, the same as *selection by location* mentioned above. Similarly, we apply the same strategy to the destination reference of the trip. The last-listed highway in the record is where the trip stops and we assign the last-listed highways nearest segment to the destination city as the ending point. Therefore, this strategy provides both starting and ending point on the highway network for *Case 1*.
- **Case 2.** When the state of origin field reports 'Washington', the same strategy executed above is performed here in order to indicate the starting point on the highway network. Nonetheless, we cannot implement the same strategy here in order to identify the ending point because the destination is outside the state of Washington. Alternatively, we use the information from the state of destination field to indicate the direction of this truck trip, for instance, the truck moving toward Oregon is going south while the truck moving toward British Columbia is heading north. Moreover, we assume that the last-listed highway in each record under this case implies that the truck exit Washington border via that highway. We apply the *If...then Statement* again here to specify the last-listed highway's the edge at the border as the destination point.
- **Case 3.** When the state of destination field reports 'Washington', the same strategy in *Case 1* can be applied here. Since the trip in this case originates outside Washington, we point out the starting point for this case by indicating the first-listed highway's segment at the border as a starting point on the highway network.
- **Case 4.** In the case of a transit trip, there is neither an origin or a destination in Washington. We designate the trip's starting point and

ending point at the border of Washington. The starting point of the trip is the first-listed highway's segment at the border, while the ending point of the trip is the last-listed highway's segment at the border.

- (iv) **Identify highway segments used in each trip and enable them for path tracing.** After we obtain the origin and the destination of the trip from the previous algorithm, we can perform the Utility Network Analyst Interface to trace the truck route. However, the result from the tracing task may or may not be the same as the truck driver reported since there are many possible route combinations. In fact, the Utility Network Analyst Interface performs the trace task in order to find the shortest route; accordingly, local streets are included in the route. On the other hand, the truck driver may prefer traveling on highways to traveling on local or county roads or his/her shipment/delivery schedule requires something other than shortest distance. To be more accurate, we write restricted VBA codes to prohibit the trace task on highways rather than the reported ones. With the enabled/disabled state of highways' segments, we disable all highways and roads first. Next, we enabled only the reported highways. Hence, the result of the trace task is the same as that describe and highlighted on the survey document.
- (v) **Record segment IDs into the MS-Access table for the later analysis.** This algorithm is straightforward. We operate *Table Interface* in order to record IDs of highway's segments into a MS Access table for the future analysis.

C. Data Challenges.

Survey data always has some errors, i.e. incorrect name, case sensitive, non-sequential highway names. Readers might recognize that these names are essential for the geo-coding process described above. Unfortunately the only way to address data input errors is to manually correct them prior to the geo-coding process. Fortunately, the problem of different case sensitivity does not interrupt the geo-coding process because the Queryfilter Interface and the Spatialfilter Interface are designed from the SQL language, which recognizes both upper and lower cases. The other problem occurs when highway names are not in sequential order. As we assume in the geocoding process that the first-listed highway name is where the journey begins and the last-listed highway name is where the trip stops. The result from the route tracing activity may turn out to be different from what the truck driver reported on the survey document.

Applications and Extensions

Once the origin-destination route information has been geo-coded, there are many different types of applications and analyses that may be performed and evaluated. What makes this possible is the bridging of information between what was provided from the truck driver related to trip characteristics (cargo commodity, vehicle type, origin-destination facility type, origin-destination location, etc.) and the highway network infrastructure.

Evaluation of individual or selective commodity flows, by highway segment or corridor is easily performed from these geo-coded data. For example, if one specific industry is known to be a significant contribution to the overall state or regional economy, and state policy planners are focused on identifying those specific highway segments that support

this key industry, it is very straightforward to highlight all origin-destination routes from trucks transporting commodities from this primary industry and geographically illustrate those components of the highway network that are critical to that industry. Highway infrastructure investments may be more efficiently allocated utilizing this analytical capability.

Alternatively, state planners may be more concerned with directional freight flows and identifying the composition of freight traffic at any individual location throughout the highway network. By utilizing geo-coded route information, this type of analytical capability is efficiently available in addition to the upstream contribution or downstream distribution of freight traffic (by freight commodity, truck type, origin-destination category, etc) from each highway tributary throughout the infrastructure network. Having access to this type of information for the entire highway infrastructure opens many research avenues and subsequent improved infrastructure management.

Another application of the geo-coded route information is the relationship between the unique pavement characteristics throughout the transportation infrastructure and the specific commodity movements moving over each highway segment. There may be unique traffic or truck characteristics for specific commodity movement that may accelerate or reduce pavement life expectancy due to unique attributes associated with different types of commodity movements and the unique pavement subsurface materials. Improved pavement management programs may result from utilizing this type of information and analytical method.

Overlaying this freight origin-destination route information with other transportation data (passenger traffic flows, site-specific traffic accident information, weather, etc.) may also open a plethora of research opportunities that may improve how state/federal agencies manage and optimize current transportation infrastructure.

Summary

This research paper offers a step-by-step description of how origin-destination freight route data (truck driver interview data) may be assigned geographic references within a geographic information system. This process, known as geo-coding route data, offers many different analytical advantages over analyses using traditional relational databases of origin-destination information. The primary advantages are the spatial coverage, accuracy and level of detail available on statewide freight flows once route information from all statewide observations have been assigned geographical coordinates and the bridge between information related to the highway infrastructure to that captured from the roadside origin-destination questionnaires.

End Notes:

1. The Washington highway layer incorporates all types of highway; State highway, US highway, and Interstate.
2. The geometric network and the logical network refer to a dual representation of a linear system in the geodatabase. They are two sides of a coin. The geometric network is the set of edge and junction features that participate in a linear system. The logical network, on the other hand, stores the connectivity information of a network. The former has co-ordinate values, but the latter does not. Both of them come along when a user creates the geometric network.

References:

- Arctur, David and Zeiler, Michael (2004). *Designing Geodatabases: Case Studies in GIS Data Modeling*. Environmental Systems Research Institutes (ESRI) Press.
- Chang, Kang-Tsung (2005). *Programming ArcObjects with VBA; A Task-Oriented Approach*. CRC Press.
- MacDonald, Andrew (2001). *Building a Geodatabase*. Environmental Systems Research Institutes (ESRI) Press.
- Minami, Michael (2000). *Using ArcMap*. Environmental Systems Research Institutes (ESRI) Press.
- Razavi, Amir H. (2002). *ArcGIS Developer's Guide for VBA*. OnWord Press.
- Zeiler, Michael (1999). *Modeling Our World: The ESRI Guide to Geodatabase Design*. Environmental Systems Research Institutes (ESRI) Press.
- Zeiler, Michael ,ed. (2001). *Exploring ArcObjects Vol. I: Applications and Cartography*. Environmental Systems Research Institutes (ESRI) Press.
- Zeiler, Michael ,ed. (2001). *Exploring ArcObjects Vol. II: Geographic Data Management*. Environmental Systems Research Institutes (ESRI) Press.

APPENDIX

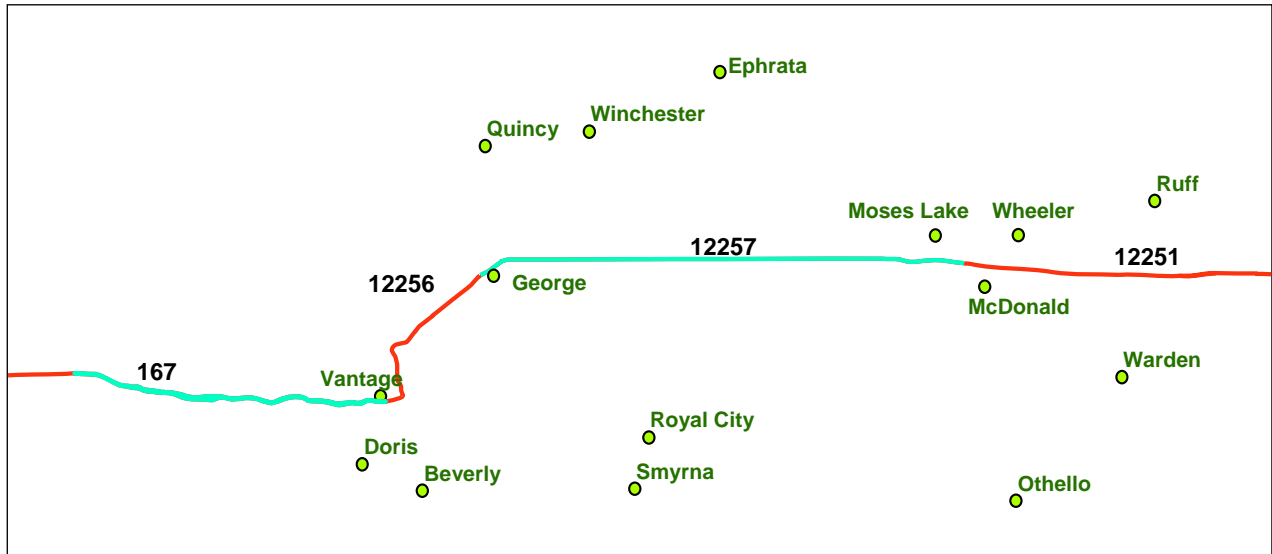


Figure 1. Different edges, with IDs, representing I90

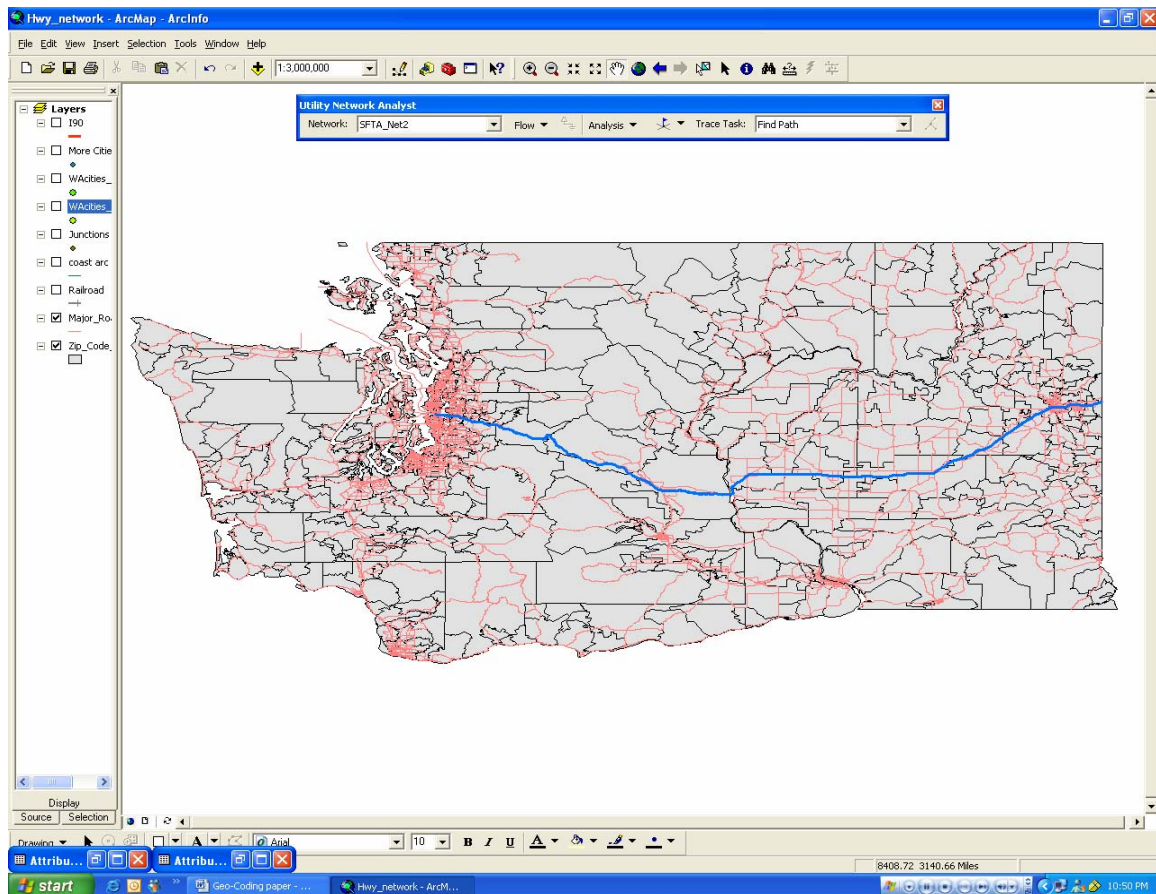


Figure 2. The Utility Network Analyst Example

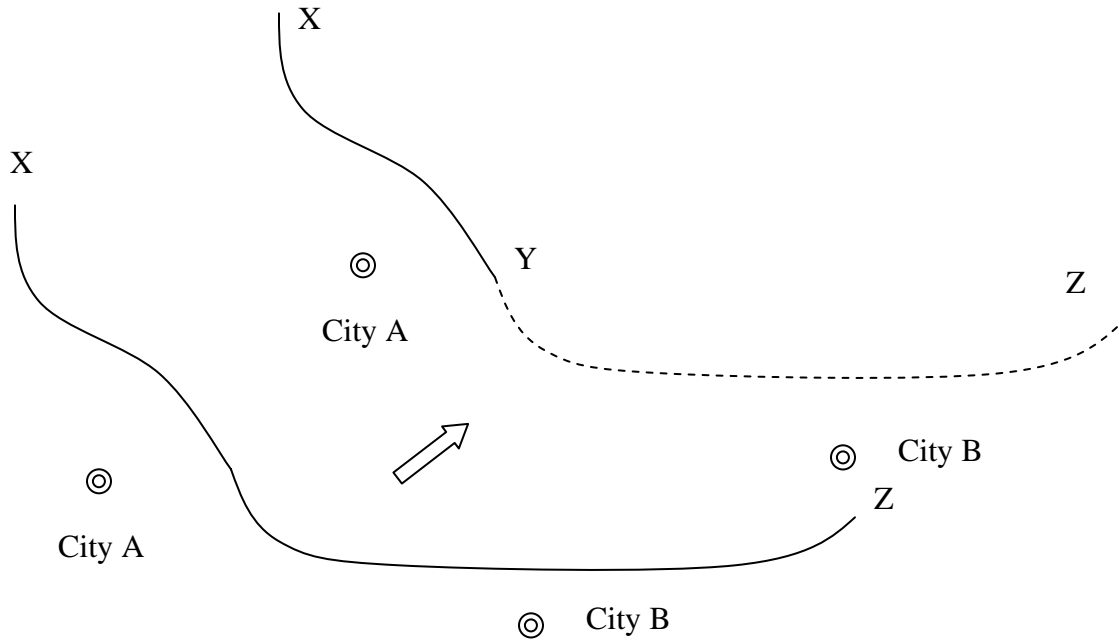


Figure 3. Editing an edge for more accurate results