Transactions in GIS, 2008, 12(1): 61–82

Research Article

Identifying Critical Locations in a Spatial Network with Graph Theory

Urška Demšar
National Centre for Geocomputation
National University of Ireland – Maynooth

Olga Špatenková
Laboratory of Geoinformation and Positioning Technology
Helsinki University of Technology

Kirsi Virrantaus
Laboratory of Geoinformation and Positioning Technology
Helsinki University of Technology

Abstract
Effective management of infrastructural networks in the case of a crisis requires a prior analysis of the vulnerability of spatial networks and identification of critical locations where an interdiction would cause damage and disruption. This article presents a mathematical method for modelling the vulnerability risk of network elements which can be used for identification of critical locations in a spatial network. The method combines dual graph modelling with connectivity analysis and topological measures and has been tested on the street network of the Helsinki Metropolitan Area in Finland. Based on the results of this test the vulnerability risk of the network elements was experimentally defined. Further developments are currently under consideration for eventually developing a risk model not only for one but for a group of co-located spatial networks.

1 Introduction

Protecting the safety of the people and ensuring security of the vital functions of the society are two core tasks of governments in today's world. Every developed society needs to be prepared for crises. Potential threats include disturbances of the functions of various infrastructures (information and energy), transportation infrastructures (roads and streets, airways or waterways), catastrophes caused by terrorism or other

Address for correspondence: Urška Demšar, National Centre for Geocomputation, John Hume Building, National University of Ireland – Maynooth, Maynooth, County Kildare, Ireland. E-mail: urska.demsar@nuim.ie

© 2008 The Authors. Journal compilation © 2008 Blackwell Publishing Ltd

criminal groups, natural disasters and last but not of course least, the use of military power. Most of the crises, however, are caused by unfriendly civilian groups or unpredictable forces of nature. In today's society, both natural disasters as well as terrorism engender acute crises in places which were traditionally safe and outside of the areas of any potential enemy attacks (YETT 2006).

Modern society depends on uninterrupted functioning of its infrastructural spatial networks, such as communication networks, energy networks, water supply and transportation networks. A common problem in the vulnerability analysis of these spatial networks is interdiction, where network elements are intentionally or otherwise disabled, which disrupts the flow through the network (Murray et al. 2007). What are the consequences of a major failure in the electrical network? What happens if water pipes are destroyed and water supply to a city is disrupted? What are the consequences if the transportation does not function? These are important questions, which need to be answered before the actual crisis occurs.

This article introduces a mathematical method for modelling the vulnerability risk of network elements, which can be used for identification of critical locations in the infrastructure. A critical location is defined as an object or an area whose removal or destruction changes the structure of the network in terms of flow and connectedness. Destructing or damaging objects in such locations either disconnects large areas of networks from each other or causes a rerouting of the flow from one area of the network to another through a longer detour path. Such locations are therefore vulnerable and should be protected against attack or damage and should be taken into account in the contingency planning.

The work presented in this article is a pilot experiment which uses the street network of the city of Helsinki as an example of a transportation resource for citizens. Which elements of such a network are critical depends upon several criteria: some are critical because they represent the only connection between separate subparts of the network (examples include bridges and tunnels). Other elements are critical because they are located on several best routes in the network (examples of these are main roads). Additionally, some parts of the network are denser than other parts (city centres are often more densely networked than peripheries) and are vulnerable for that reason.

The vulnerability risk of the elements of the street network of Helsinki was estimated with an approach based on the principles of graph theory and social network analysis. The data used for this project came from Digiroad – The National Road and Street Database of Finland (Finnish Road Administration 2007). While this dataset covers the entire country and includes all roads and streets together with associated attribute information, the network used in this experiment was limited to the Helsinki Metropolitan Area. The aim of the pilot experiment was to identify the critical elements in this network. Our approach uses only the topology of the network elements, even though attribute information could be integrated in the estimation of the realistic risk at some later stage of method development.

At the time of writing, the method uses only one spatial network. Eventually the goal is to develop a method that would identify critical locations that are vital not only in one, but in several co-located networks simultaneously. An example of such locations are bridges with several networks – a road, a railway track, a subway track, electrical and gas lines might all be located on a single bridge. The destruction of such a bridge would disconnect the areas on each side of the bridge and disrupt the flow in all these networks.

© 2008 The Authors. Journal compilation © 2008 Blackwell Publishing Ltd

The remainder of the article is organised as follows: section 2 gives an overview of mathematical terms, presents the linkage between graph theory and spatial networks and gives a theoretical description of our methodology. The testing of the method on the Helsinki street network and the experimental development of the vulnerability risk measure is presented in section 3. Section 4 presents some conclusions and our plans on how to proceed. Ideas for future research include incorporating attribute information into the method, estimating the risk for several co-located spatial networks and potential applications outside the scope of crisis management.

2 Methodology

This section presents the mathematical method for identification of critical locations in a spatial network. The approach is based on graph theory and uses the topology of the network (i.e. how the segments in the network are connected to each other) in order to identify problematic elements. The geometry of the network elements (meaning the actual spatial location of the streets) has been used for visualisation purposes, but the method does not use any attribute information. The method combines dual graph modelling with connectivity analysis and two topological graph measures: 'betweenness' and 'clustering coefficient'. The first part of this section explains these graph theoretic notions, the second one gives an overview of structural network analysis in GI Science, while the third one presents our method for the identification of critical locations in a spatial network using the mathematics presented in the first part.

2.1 Basic Definitions from Graph Theory

The definitions in this subsection were taken from Jungnickel (2005) but can be found in any textbook on graph theory.

A *graph* $G = G(V, E)$ is a pair of two disjoint finite sets V and E, where E is a subset of $V \times V$, which means that it is a set of two-element subsets of V. The elements of V are *vertices*, the elements of E are *edges*. An edge $e = uv$ from the set E connects vertices u and v. When representing this structure graphically, the vertices are usually drawn as dots and edges as lines connecting each two respective vertices. Two vertices x, y of G are *adjacent* or *neighbours* if xy is an edge of G. Two edges of G are *adjacent* if they share a common end vertex. If no direction is specified on the edges (i.e. the edge uv is considered the same as the edge vu), then the graph is *undirected*. If the direction of the edge is important, then the graph is *directed*. The vertices and the edges can also have numerical or other values assigned – this is *a weighted graph* and these values are called the *weights*. At this point in the development of our method we are only interested in undirected and unweighted graphs even though the method could be eventually extended to graphs that are either weighted or directed or both.

When properties of edges need to be translated into properties of vertices, a so-called line graph is usually constructed. Given a graph G, the *line graph* $L(G)$ takes the edges of G as its vertices, i.e. $V(L(G)) = E(G)$. Two vertices e and f in the line graph are connected if and only if the respective edges e and f are adjacent in G (i.e. they share a common end vertex). The line graph is sometimes called an edge-dual of its original graph and the modelling done on the line graph is referred to as dual graph modelling. Figure 1 shows an example of a graph G and its line graph $L(G)$.

© 2008 The Authors. Journal compilation © 2008 Blackwell Publishing Ltd

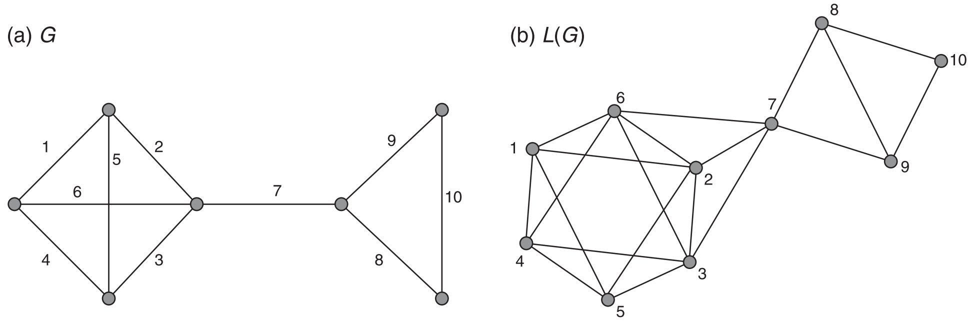

Figure 1 (a) Graph G and (b) its line graph $L(G)$. The edges of the graph G are labelled with the same numbers as the corresponding vertices in its line graph $L(G)$

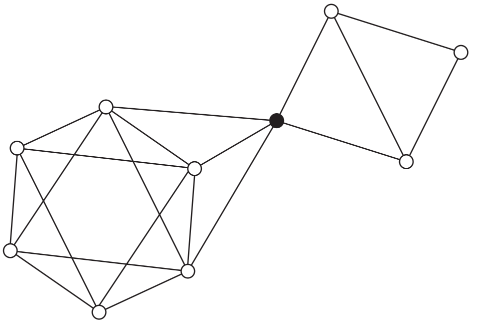

Figure 2 Cut vertex (shown in black) separates the line graph into two biconnected components

A *walk* in a graph is an alternate sequence of vertices and edges, v_1-e_1-v_2-e_2- . . . -v_n, where each edge e_i connects vertices v_i and v_{i+1}. A *trail* is a walk in which no edge is repeated. A *path* is a trail in which no vertex is visited more than once. The *length* of a walk, a trail or a path is defined as the number of edges it contains.

Two vertices u and v of an undirected graph G are *connected* if G contains a path from u to v. A graph is *connected* (or *vertex-connected*) if every two vertices of the graph are connected (i.e. there exists a path between every two vertices in the graph). Obviously not every graph is connected, but every graph consists of *connected components*, which are maximal connected subgraphs of G.

A vertex is a *cut vertex* (or an *articulation vertex*) if its deletion increases the number of connected components in a graph. An example of the cut vertex is shown in Figure 2 (coloured in black). If this vertex is deleted, the graph in Figure 2 falls apart into two pieces.

A set of vertices whose removal turns a connected graph G into an unconnected graph is a *cut* (or a *vertex-cut*). *Vertex connectivity* $\kappa(G)$ of the graph G is the size of the smallest vertex cut. Vertex connectivity is usually referred to as simply *connectivity*. A graph is *k-connected* if its connectivity is greater than or equal to *k*.

© 2008 The Authors. Journal compilation © 2008 Blackwell Publishing Ltd

Similar definitions can be made for the edges of a graph. An edge is a *bridge* if its deletion increases the number of connected components in a graph. An *edge cut* of a graph G is a set of edges whose removal causes the graph to become disconnected. The *edge-connectivity* $\kappa'(G)$ is the size of the smallest edge cut and the graph is *k-edge-connected* if its edge connectivity is larger or equal to k. Edge-connectivity of a graph G corresponds to (vertex) connectivity of its line graph $L(G)$: $\kappa'(G) = \kappa(L(G))$. A bridge in a graph G corresponds to a cut vertex in $L(G)$.

Vertices u and v are *biconnected* if they are connected by two paths that do not share any common internal vertex (i.e. two *independent paths*). If every two vertices of a graph are biconnected, then the graph is *biconnected*. Such a graph has $\kappa(G) = 2$. Every connected graph has a unique decomposition into *biconnected components*, which form a tree structure. Each two biconnected components are either disjoint or have at most one vertex in their intersection. Every vertex located in the intersection of two biconnected components is a cut vertex. Illustrating this is the cut vertex in Figure 2. This graph has two biconnected components: each of them is a union of the cut vertex with one of the two pieces of the graph without the cut vertex. The cut vertex is the only common element of the two components.

Centrality measures are values that describe the structural importance of each vertex in the graph – vertices with higher centrality have a larger impact on other vertices. The three commonly used measures are degree, closeness and betweenness and were first introduced in social network analysis (Freeman 1979).

The *degree* of a vertex v, $d(v)$, is the number of its neighbours. The *closeness* of a vertex v, $cl(v)$, is the shortest distance of v to all other vertices, where distance is measured as the length of the shortest path from v to every other vertex. Closeness and degree are radial measures as they assess properties that emanate from a given vertex. Centrality of a given vertex in the graph can also be described by a measure based on the number of paths that pass through it – this is the *betweenness*, $b(v)$, which is defined as the proportion of the shortest paths between every pair of vertices that pass through the given vertex v towards all the shortest paths. More precisely,

$$b(v) = \sum_{s \neq t \neq v} \frac{\sigma_v(s, t)}{\sigma(s, t)} \quad (1)$$

where s and t are two distinct vertices of G not equal to v, $\sigma_v(s, t)$ is the number of shortest paths from s to t that pass through v (there can exist several wholly or partially parallel shortest paths from s to t that are of the same length), and $\sigma(s, t)$ the total number of shortest paths from s to t. Vertices with the highest betweenness are those that are located on many shortest paths between other vertices. Girvan and Newman (2003) proposed an equivalent definition of betweenness for edges, the so-called *edge-betweenness*, which is defined as the proportion of the shortest paths between each pair of vertices that pass through the given edge e towards all the shortest paths. The edge-betweenness of a graph G corresponds to the betweenness of its line graph $L(G)$.

The importance of the vertex in its immediate neighbourhood can be measured by the *clustering coefficient*, $cc(v)$. This coefficient is defined as the number of edges between the vertices within the immediate neighbourhood of vertex v divided by the number of all possible edges between them. More precisely:

$$cc(v) = \frac{e(v)}{\binom{d(v)}{2}} = \frac{2 \cdot e(v)}{d(v) \cdot (d(v) - 1)} \quad (2)$$

© 2008 The Authors. Journal compilation © 2008 Blackwell Publishing Ltd

where $e(v)$ is the number of edges between neighbours of v and $d(v)$ is the degree of vertex v (and so the expression $\binom{d(v)}{2} = d(v)(d(v) - 1)/2$, which is the number of combinations of size 2 selected from the $d(v)$ items, equals the number of all possible edges between the $d(v)$ neighbours of vertex v – to get an edge we need to arbitrarily select two different neighbours of vertex v). The average clustering coefficient is used together with another measure, the average shortest path, to determine if a graph is a random graph, a regular graph or a small-world network. In mathematics the topology of the network is often assumed to be either completely regular or completely random. Many real-world networks are somewhere between these two extremes: they have a small average path length like random networks and high clustering coefficient like regular graphs – these two properties define the small-world network (Watts and Strogatz 1998).

2.2 What Does Graph Theory Have To Do With Spatial Networks?

Linear geographic phenomena can be modelled as spatial networks. Digital representations of such networks in Geographic Information Systems (GIS) are graphs: the intersections and bending points of linear features are represented as vertices and the features themselves as edges that connect the vertices (Worboys and Duckham 2004). Using graph theory for the analysis of these phenomena is therefore a logical choice. Much has been in done in this area, in particular in transportation (Thill 2000, Miller and Shaw 2001), but many of the GIS applications only use traditional methods, such as finding the shortest path, calculating the flow capabilities or analysing connectivity of a network with the aim to correct the data if the network turns out to be unconnected (Curtin 2007). More complex analyses using tools from other areas of graph theory, such as probabilistic graph theory, complex networks research, algebraic/spectral graph theory and structural graph theory are rarely used. Advanced applications of graph theory are therefore not as common in GI Science as in other disciplines, where graphs represent various phenomena, including organic molecules in chemistry, protein-receptor interaction networks in medicine, genealogies, Internet networks, citation networks and diffusion networks of diseases (Batagelj and Mrvar 2003). Our approach combines dual graph modelling with measures from structural graph theory, probabilistic graph theory and complex networks research. This combination of several advanced graph theoretic concepts presents a novel approach towards the analysis of spatial networks.

One of the oldest fields of application is social network analysis (de Nooy et al. 2005), where the three centrality measures were first defined (Freeman 1979). In social networks vertices represent persons or institutions and edges relationships between them. The degree and closeness centralities in social network analysis describe the reachability of a person in the network: how easily can information reach that person. Betweenness describes the extent to which the person is needed as a link in the chains of contacts that facilitate the spread of information in the network. If a person with high betweenness is removed from the network, many flows of information are disrupted or must take longer detours (de Nooy et al. 2005). Besides these classical three centralities, a number of other centrality measures can be used for analysing the structure of a social network (see Borgatti and Everett 2006 for a review of various centralities in social network analysis) and these principles can also be used for spatial phenomena (Besussi 2006).

© 2008 The Authors. Journal compilation © 2008 Blackwell Publishing Ltd

Another area where graphs are used for describing the space is space syntax (Hillier and Hanson 1984). Space syntax describes relationships between different urban spaces and interactions between space and society based on the structure of the urban environment. The urban space is modelled as a connectivity graph, which is derived from a selected space representation in a similar manner as the line graph is derived from the original graph in graph theory. Several space representations are possible: convex spaces (Hillier and Hanson 1984), axial lines (Hillier and Hanson 1984), isovists (Batty and Rana 2002) or named streets in a street network (Jiang and Claramunt 2004a, b; Rosvall et al. 2005). In all these approaches, the importance of spaces is described by calculating centralities and topological measures for vertices in the connectivity graph that correspond to urban spaces (Hillier and Hanson 1984; Jiang and Claramunt 2004a, b). The connectivity graph where vertices represent named streets and edges intersections between streets is sometimes called the dual modelling approach for the street network (Crucitti et al. 2006) or an information city network (Rosvall et al. 2005). This is an appropriate method for the analysis of urban space, but taking the named streets or other larger spatial entities as basic spatial representation units distorts the topology of the original network and if used in our case, could potentially collapse several critical locations into only one graph element. Therefore our approach keeps the original topology of the street network and does not attempt to model it by adopting some other spatial representation.

In the complex network research, many studies have been undertaken in examining the vulnerability and survivability of critical network infrastructures. Some studies offer focus on global, continental or national networks, such as the Internet, electrical power systems and airport networks and either examine their scale-free or small-world properties or analyse their flow capabilities (Carlier et al. 1997, Latora and Marchiori 2004, Grubesic and Murray 2006, Guida and Maria 2007, Murray et al. 2007). The Internet in the USA is an example of a spatial information network that has received a particularly thorough investigation (Wheeler and O'Kelly 1999, Gorman and Malecki 2000, O'Kelly and Grubesic 2002, Grubesic et al. 2003, Gorman and Kulkarni 2004, Gorman et al. 2004). Less attention has been given to the vulnerability analysis of local spatial networks, such as the infrastructural networks within a city: the focus of this article.

2.3 Identification of Critical Locations in a Spatial Network

The approach presented in this article attempts to identify critical locations in a spatial network by using its line graph. The network is represented as an undirected and unweighted graph G from which its line graph $L(G)$ is derived. Critical locations correspond to edges in the original network, but since the centrality and topological measures can only be calculated for vertices and not for edges, we initially translate edges into vertices by generating a line graph. Once the line graph has been produced, three measures are then used to identify critical locations: cut vertices, the betweenness and the clustering coefficient. The framework for the identification of critical locations is presented in Figure 3 and explained in the remainder of this section.

The procedure is based on the assumption that the vertices of the line graph that correspond to critical locations have one or more of the following three properties:

- They are cut vertices of the line graph;
- They have a high betweenness; or
- They have a low clustering coefficient.

© 2008 The Authors. Journal compilation © 2008 Blackwell Publishing Ltd

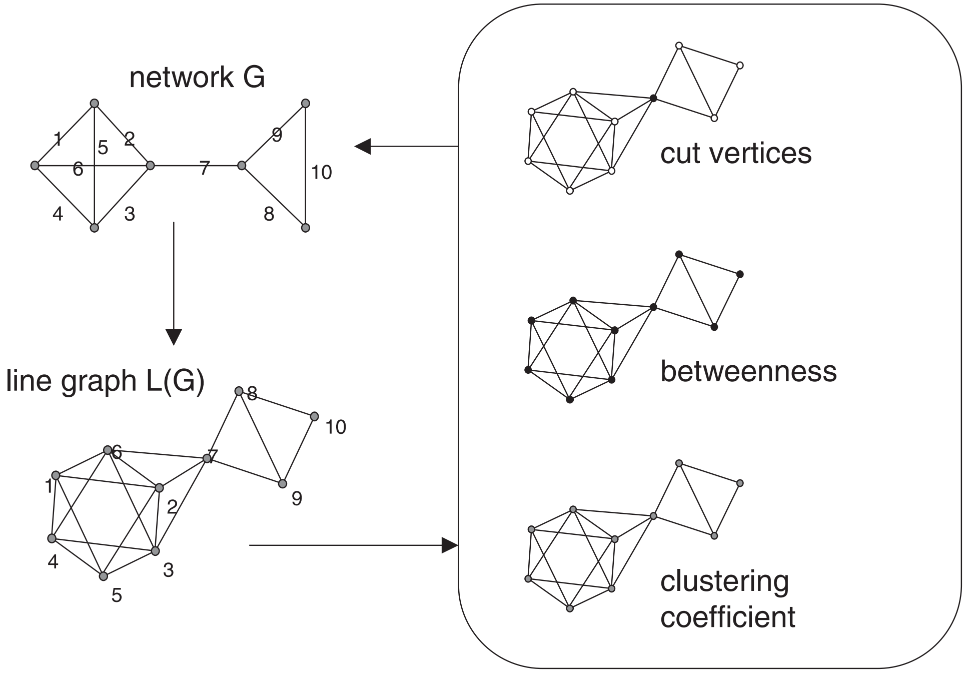

Figure 3 The framework for identification of critical locations

All cut vertices are clearly critical locations. If such a vertex is removed from the line graph (or its corresponding edge is removed from the original network), the graph falls apart into two or more disconnected components. However, the requirement that critical locations are cut vertices is by itself not sufficient for their identification. An example where this is not enough: consider the street network on an island which is connected to the mainland by two or more bridges. None of the vertices that correspond to the bridges in the respective line graph is in itself a cut vertex (as the removal of one of them is not enough to completely disconnect the island – there are still other bridges that connect it to the mainland), but removing or damaging each of the bridges will disrupt the flow through the network. Each of the bridges is therefore a critical location, but none of them is a cut vertex.

Of all three centrality measures, betweenness is the only one that is linked to the flow in the network. Degree is a local measure that only considers the immediate neighbourhood of a vertex and is therefore not useful for identification of critical locations. Closeness describes how well a vertex is integrated in the network: a vertex with high closeness is more connected to all other vertices in the sense that the shortest paths from this vertex to all other vertices are short compared to vertices with lower closeness. This property is not relevant for critical locations either. The final measure is betweenness. Removing a vertex with high betweenness from the line graph is equivalent to removing an edge with high edge-betweenness from the original graph. This breaks off many shortest paths between vertices in the original graph, which disrupts the flow or redirects it through a longer detour. This is one of the essential

© 2008 The Authors. Journal compilation © 2008 Blackwell Publishing Ltd

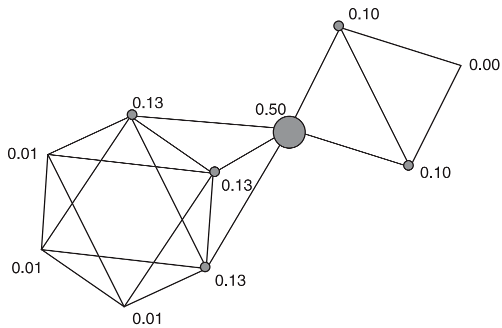

Figure 4 Betweenness of vertices in the line graph. The size and the labels of the vertices indicate the value of betweenness for each vertex

properties of a critical location. It is therefore reasonable to use high betweenness in the line graph as one of the criteria for the identification of critical locations. This choice has been experimentally confirmed in a preliminary analysis of these three centrality measures (Demšar et al. 2007). An example of a vertex with high betweenness is shown in Figure 4.

The clustering coefficient gives an estimate of how well connected is the immediate neighbourhood of a given vertex. Suppose that vertex v in the line graph $L(G)$ corresponds to the edge e in the original graph G. If v has a high clustering coefficient, this means that there are many edges between the neighbours of this vertex in the line graph, which in turn means that many edges adjacent to e in G share a common end vertex (Figure 5). These edges are represented as neighbouring vertices to vertex v in the line graph and are connected with each other in $L(G)$. The connections between them contribute to the clustering coefficient of the vertex v in $L(G)$. The common end vertex of these edges in G can either belong to the edge e or not. If it does (consider for example the common end vertex of edges 7 and 8 in Figure 5a, which also belongs to edge 10), this means that there is no parallel option to the edge e formed by these two edges and if edge e is removed, a possibly important connection between two parts of the graph is lost. However, if it does not (such as the common end vertex of edges 1 and 2 in Figure 5a, which does not belong to edge 10), then the path through these edges forms a parallel path to edge e. Even if e is removed, there still exists an alternative path connecting two parts of the network. It is more likely that such parallel paths exist for edges whose respective vertices in the line graph have a high clustering coefficient. If, however, the clustering coefficient of a vertex in the line graph is low, there is less chance for parallel paths through the nearest neighbourhood of the edge e. This implies that removing this edge would cause more disruption in the flow in the original graph than removing an edge whose respective vertex in the line graph has a high clustering coefficient. Therefore a low clustering coefficient could be used to identify critical locations.

After the calculation in the line graph the three topological measures are transferred back into the original network, by assigning the appropriate values to the edges of the network for analysis and visualisation purposes.

© 2008 The Authors. Journal compilation © 2008 Blackwell Publishing Ltd

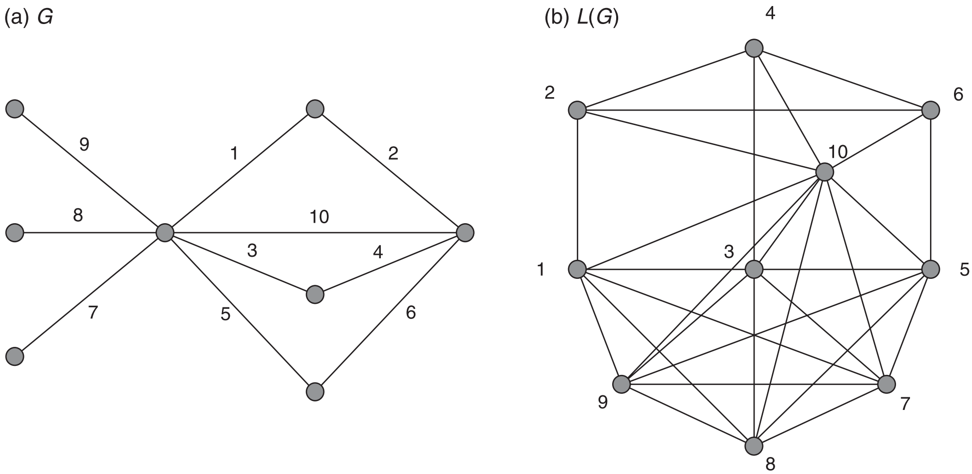

Figure 5 Original graph *G* in (a) and its line graph *L*(*G*) in (b). Edges 1 and 2 in *G* share a common end vertex with each other, which is not an end vertex of edge 10. Vertices 1 and 2 in *L*(*G*) are therefore connected and this connection contributes to the clustering coefficient of vertex 10 in *L*(*G*)

3 Experiment and Results

This section presents how the proposed method was tested on the street network of the Helsinki Metropolitan Area and how a measure for the vulnerability risk of network elements was experimentally defined based on the results of the test.

3.1 Testing the Procedure on the Street Network of the Helsinki Metropolitan Area

The method was tested on the street network of the Helsinki Metropolitan Area, which covers the municipalities of Helsinki, Espoo and Vantaa located in southern Finland. The network consists of 68,761 road segments and is available as a shapefile from Digiroad (Finnish Road Administration 2007). This file was first converted into an ASCII file where the only information retained was topology in the form of the numbers of from- and to-nodes for each road segment, i.e. the description of the graph representing the network. No geometrical or attribute information was used. The graph representing the network was then transformed into a line graph using a specially written programme. The line graph had 68,761 vertices, corresponding to the original road segments.

Cut vertices, betweenness and the clustering coefficient were calculated for the line graph using the network analysis package Pajek (Batagelj and Mrvar 2006). The size of the network prohibited a graphical display of these measures in Pajek, even though the package provides several graph-drawing algorithms (and was used for the creation of the figures in the previous sections). The values of all three measures were transferred back into the original shapefile by assigning appropriate values to the edges of the street network and then visualised as maps in GIS software. Figure 6 shows the distribution

© 2008 The Authors. Journal compilation © 2008 Blackwell Publishing Ltd

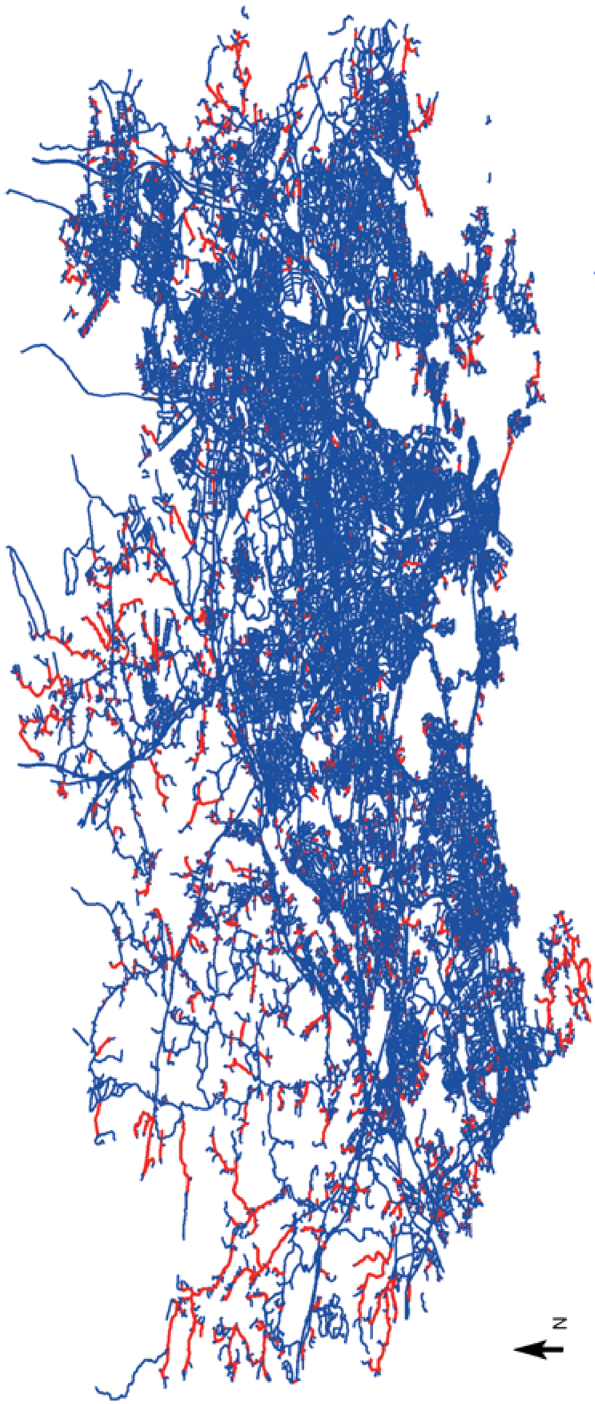

Figure 6 Edges in the street network of Helsinki Metropolitan Area that correspond to cut vertices in the line graph (marked in red)

© 2008 The Authors. Journal compilation © 2008 Blackwell Publishing Ltd

of the road segments in the street network which correspond to cut vertices in the line graph, Figure 7 the values of betweenness for each road segment and Figure 8 the clustering coefficient for each road segment in the street network.

3.2 Results and Definition of the Risk Measure

Each of the selected measures represents a different reason why a particular location should be considered as critical. To get an overall picture of these locations, the measures need to be combined into one single measure, the vulnerability risk, which could be added as an attribute to the edges of the original network. The combination should consider the nature of the particular measures and focus on their relevance. This section discusses the results of the experiment on the Helsinki street network and describes how the vulnerability risk was defined based on the results.

All identified cut vertices should be fully included in the resulting risk measure, as their removal disconnects parts of the network. These parts, however, often consist of one or two road segments only, and are as such too small to be really important. The simple cut vertices calculation resulted in a high number of identified critical locations, but in most cases, they connected areas that were really small. In an attempt to solve this, cut vertices were re-calculated in the line graph for a minimum size k of the biconnected components (i.e. each biconnected component in the line graph needed to have at least k vertices). Several values for this parameter were tried in the experiment.

Table 1 shows the number of identified cut vertices, the number of vertices in the biconnected components satisfying the minimum size criterion, and the number of vertices in the biconnected components that were too small for each value of the parameter k.

When the minimum size threshold of biconnected components was increased, the number of irrelevant cases decreased. However, some of the important cut vertices also disappeared for larger values of k. One such example was the cut vertices that represented the bridges composed of several road segments in the original network: each of the bridge segments was a cut vertex when no limitation was imposed on the size of the biconnected components. Once the size was set to more than 2, these edges did not count as cut vertices in the line graph anymore and disappeared from the result. In this way some of the important bridges were lost, but on the other hand, some not too important dead-ends were still identified. Figure 9 compares the cut vertices in the centre

Table 1 The number of the identified cut vertices, the vertices in the biconnected components of sufficiently large size (those that qualified) and the vertices in the too-small biconnected components (disqualified vertices) for selected values of the minimum size parameter

	Minimum size of biconnected components (k)					
Number of	1	3	4	5	7	10
cut vertices	2,752	2,233	122	63	37	21
qualified vertices	65,961	66,095	64,059	63,725	63,544	63,287
disqualified vertices	48	433	4,580	4,973	5,180	5,453

© 2008 The Authors. Journal compilation © 2008 Blackwell Publishing Ltd

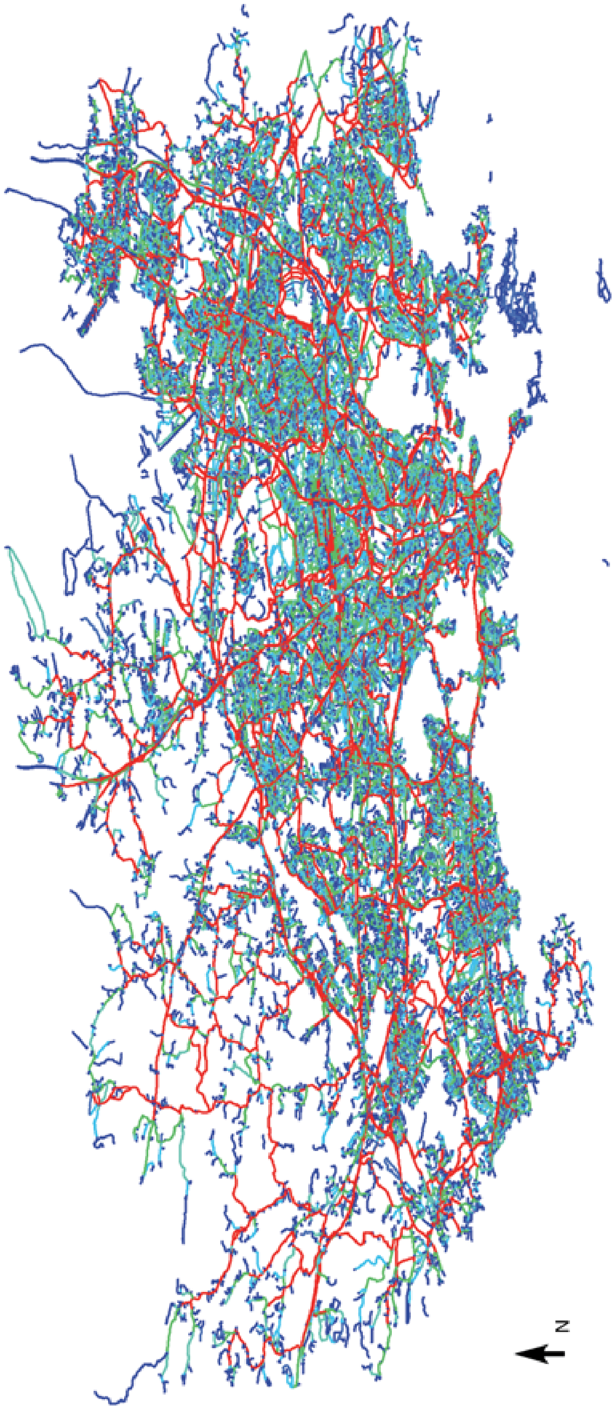

Figure 7 Betweenness. Red colour indicates high values, light blue medium values and dark blue low values of this measure

© 2008 The Authors. Journal compilation © 2008 Blackwell Publishing Ltd

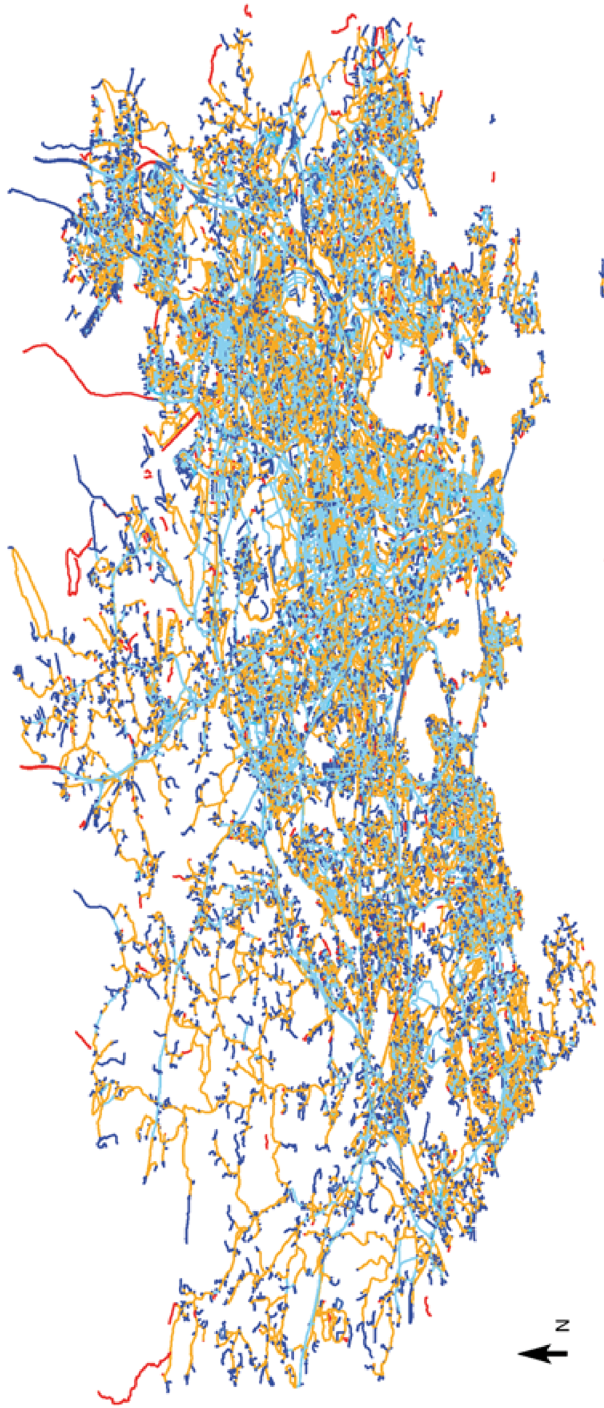

Figure 8 Clustering coefficient. Red and orange colours indicate low values, light blue medium values and dark blue high values of this measure

© 2008 The Authors. Journal compilation © 2008 Blackwell Publishing Ltd

of Helsinki for minimum sizes of biconnected components 1 and 5, respectively: the number of cut vertices for $k = 5$ in Figure 9b is significantly smaller than in Figure 9a, in fact, there are just four left. However, while in Figure 9a many bridges to the islands are clearly identified, there are also other thick black edges in the diagram that lead to parts of the network that only have one or two segments.

Figure 9 Cut vertices in the centre of Helsinki for (a) $k = 1$ and (b) $k = 5$. Road segments corresponding to cut vertices are shown as wide black lines, those that correspond to qualified vertices as light grey lines and edges that do not belong to biconnected components of the minimum size in dark grey

© 2008 The Authors. Journal compilation © 2008 Blackwell Publishing Ltd

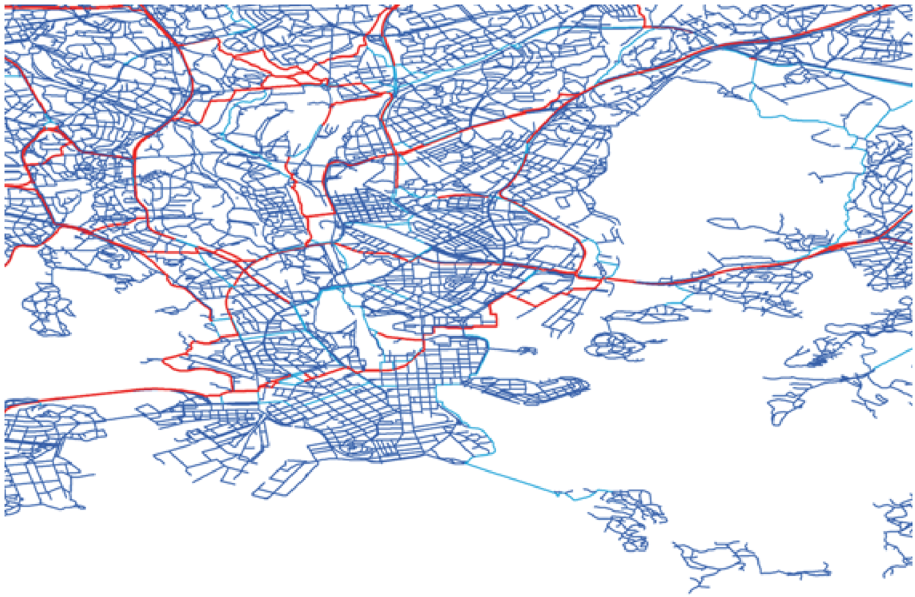

Figure 10 Betweenness in the centre of Helsinki – with an appropriate classification high betweenness (in red) corresponds to the network of real major roads and highways

Betweenness is the centrality measure that is linked to the flow in the network and can identify its functional characteristics. An interesting result that can be observed in Figure 7 is that road segments with high betweenness correspond very well to the network of main roads and highways in the Helsinki Metropolitan Area. The interesting fact is that they have been identified solely from topological information. No attribute information such as whether a particular road segment belongs to a highway or a small street has been used and yet a high betweenness identifies the main traffic routes in the network. Using an appropriate classification for the values of betweenness, it is possible to create a map of the real major roads as shown in Figure 10, which shows the betweenness values for the centre of Helsinki. Since edges with high betweenness indicate main traffic routes in the network, they are the ones that require protection against interdiction. Betweenness should therefore be included in the vulnerability risk measure.

The clustering coefficient gave the least convincing result. Even though the initial idea was that it could be used to count parallel paths through the immediate neighbourhood of an edge, this did not work as expected. Here we faced several problems. One of them is the non-uniqueness of line graphs: a line graph with two-edge parallel paths to a given edge *e* is equal to a line graph where there are no parallel paths to this edge. This is illustrated in Figure 11: graphs in Figures 11a and 11b are different, yet have the same line graph, shown in Figure 11c. Vertex 5 in Figure 11c represents an edge with two parallel paths in the first original graph (Figure 11a) and an edge without any parallel paths in the second one (Figure 11b). But the clustering coefficient of vertex 5 is the same, regardless of the original graph and can therefore not be used for separating the edges with parallel paths from those without such paths.

© 2008 The Authors. Journal compilation © 2008 Blackwell Publishing Ltd

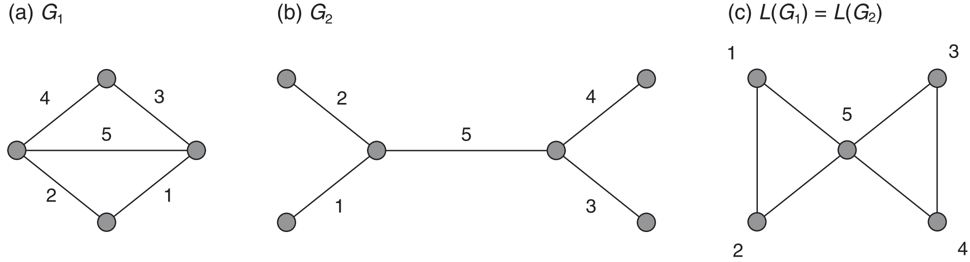

Figure 11 Non-uniqueness of line graph construction. Graphs G_1 in (a) and G_2 in (b) have the same line graph $L(G_1) = L(G_2)$ in (c). Two-edge paths 1–2 and 3–4 are parallel paths to edge 5 in G_1. The edge 5 in G_2 has no parallel paths

Topological errors and improper representation of the network are another cause for the failure of the clustering coefficient: disconnected edges (often dead ends) are assigned zeros. On the other hand, the main roads represented as several parallel line features have high values of the clustering coefficient because of their representation by several interconnected links. Only properly represented dead ends result in high values of the clustering coefficient as expected, but that does not help with identification of critical locations in the network.

Based on the results of this experiment, cut vertices for the minimum size of biconnected components equal to 5 were combined with betweenness in order to define the vulnerability risk for each road segment of the network. The clustering coefficient was not included in the vulnerability risk measure because of the above issues. The risk was calculated in the following way: whenever the cut vertex was identified, the edge was assigned a risk value 1 (high risk). In all other cases the risk was equal to normalised betweenness, so that it covered the range from 0 to 1. Figure 12 shows the vulnerability risk for the whole Helsinki Metropolitan Area.

The definition of the risk measure leaves some room for improvement and further development. If in the future some other graph measure is found to be relevant for the identification of critical locations, the vulnerability risk can be expanded as a weighted average of betweenness, cut vertices and this other measure.

In this experiment the size of the biconnected components was cut at 5, but since the suitable size is application-dependent, the decision of which size is relevant should be open for discussion with the crisis management experts as well as subject to sensitivity analysis, which however is beyond the scope of this paper. An attempt could also be made to disregard the minimum size of the components and eliminate the irrelevant cut vertices by some other method. One option would be to continue with spatial analysis of the result, try to investigate the structure of the nearest neighbourhood of each cut vertex and eliminate only those vertices that really lead into short dead-ends. Another option would calculate the biconnected components of a weighted line graph with weights derived from the attribute information of the street network and indicating which of the identified road segments are important enough to be included in the risk measure. The third option would be to first eliminate the edges corresponding to cut vertices of the components of a certain size and then run the procedure on the reduced network to estimate the importance of the remaining road segments. This could also

© 2008 The Authors. Journal compilation © 2008 Blackwell Publishing Ltd

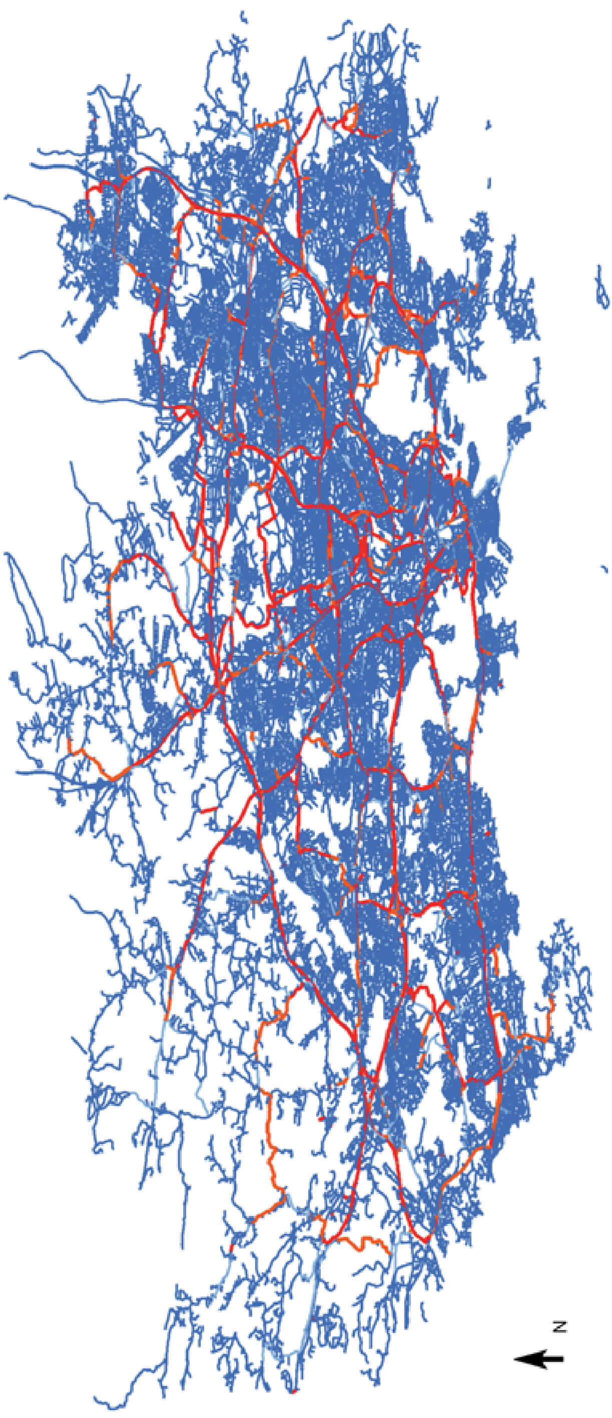

Figure 12 Vulnerability risk for the street network of the whole Helsinki Metropolitan Area. High risk is indicated in red and corresponds well to the highways and major roads of the area

© 2008 The Authors. Journal compilation © 2008 Blackwell Publishing Ltd

help with the identification of the second most important segments, if the network was damaged at a particular critical location and the goal was to identify the shortest detour path.

4 Conclusions and Further Research

This article presented an approach for the identification of critical locations in a spatial network by dual graph modelling, connectivity analysis and the calculation of topological measures for the line graph of a given spatial network. Of the presented graph theoretic measures, betweenness and cut vertices proved to be the two useful ones for the identification of critical locations. In the experiment where the method was tested on the street network of the Helsinki Metropolitan Area, betweenness identified the major roads using solely topological information as input. This shows that the method reveals some functional knowledge about the infrastructure, even though geometry and attribute information have not been used for analysis.

This knowledge could be enriched by including the available attribute information. For example, critical locations might more easily be identified if information such as known locations of vital points or vulnerable points or areas of the society would be taken into consideration. Examples of vulnerable points or areas are (Kohvakka and Valtonen 2004):

- All places where many people are located simultaneously: shopping centres, traffic nodes (railway and bus stations), but also mass events (sports competitions and rock concerts);
- All places where elderly people, handicapped or ill people or children are located simultaneously: homes for the elderly, hospitals, schools and kindergartens; and
- Protected objects: cultural heritage objects, churches, hospitals or bomb shelters.

The same authors also give examples of vital areas and points:

- Potential targets for strategic attacks: military locations, information networks and centres, electricity stations with the major transfer lines, central governmental buildings and major transportation lines; and
- Objects that pose a threat for the environment: industrial buildings, stores and transportation routes of dangerous materials.

Modelling the vulnerability risk for crisis management should include topological information from several co-located spatial networks (for example, transportation, energy and information networks, such as listed among the vital points and areas above), not just one. This poses two challenges for future research. One is how to model co-location: the problem is how to identify critical locations in several networks that are situated in the same area. For example, if a railway track and the water supply pipes are drawn through the same tunnel or over the same bridge, they are co-located in a critical location (bridge or tunnel) and this location is more critical than some other bridge or tunnel with only one spatial network drawn over/through it. The second challenge is how to analyse the risk for cascading events in interdependent co-located networks. A cascading failure occurs when an interdiction of one network produces a series of secondary failures in the networks that depend on the first one (Grubesic and

© 2008 The Authors. Journal compilation © 2008 Blackwell Publishing Ltd

Murray 2006). For example, a small localised electricity failure is usually not a major problem, but if it causes the breakdown of large telecommunication networks, such as the Internet or the mobile phone network or disrupts the regional gas supply system, the damage could be extensive.

From the mathematical point of view, graph theory encompasses much more than just the measures presented in this article. Other graph characteristics, such as transitivity, information and eigenvector centralities, colourings, cliquishness, flows, matchings, etc. (Jungnickel 2005, Newman 2007) could be considered for an improved risk modelling for spatial networks.

Since our method uses only topological information to reveal functional properties of a network, it could be applied more widely than just for analysis of local spatial networks. The method reveals knowledge from a structured but unlabeled dataset and could be as such considered a spatial data mining method and used for other purposes, such as cluster analysis and identification of link types. One potential idea is to use it for automatic metadata creation for unlabeled network data. For example, in cases when the street network has been automatically recognised from aerial or satellite images, no categorical information is attached to the roads. By using our method, the main roads could be at least partially identified from such data, which would provide valuable information about the infrastructure.

Another advantage is that the geometry of the network (i.e. the spatial location of its vertices and edges) does not play a role in the method – it is only used for visualisation purposes. Therefore the method could be applied to networks which do not have inherent geographic location, but which are viewed as spatial metaphors of non-geographic data. While geographic location is the core concept and research focus of GI Science, it has become apparent in recent years that models and methods that geographers have been using for a long time could successfully be applied to the representation of objects, phenomena or processes with spatial characteristics and behaviour in abstract spaces. The use of spatial metaphors to represent data which are not inherently spatial is called spatialisation (Skupin and Fabrikant 2007). Our method could therefore be used for knowledge discovery in any application where a network is seen as a spatial metaphor for the (not necessarily spatial) phenomenon under investigation.

Acknowledgements

The authors would like to thank Jukka M. Krisp from Helsinki University of Technology for fruitful discussions during the development stage of this method and for all their help with the preparation of the spatial network datasets used in this study.

References

Batagelj V and Mrvar A 2003 Pajek: Analysis and visualization of large networks. In Jünger M and Mutzel P (eds) *Graph Drawing Software*. Berlin, Springer Verlag: 77–103

Batagelj V and Mrvar A 2006 *Pajek: Program for Analysis and Visualization of Large Networks*, version 1.17 (available for download at: http://vlado.fmf.uni-lj.si/pub/networks/pajek/

Batty M and Rana S 2002 Reformulating Space Syntax: The Automatic Definition and Generation of Axial Lines and Axial Maps. WWW document, http://www.casa.ucl.ac.uk/working_papers/paper58.pdf

© 2008 The Authors. Journal compilation © 2008 Blackwell Publishing Ltd

Bessusi E 2006 Mapping European Research Networks. WWW document, http://www.casa.ucl.ac.uk/working_papers/paper103.pdf

Borgatti S P and Everett M G 2006 A Graph-theoretic perspective on centrality. *Social Networks* 28: 466–84

Carlier J, Li Y, and Lutton J-L 1997 Reliability evaluation of large telecommunication networks. *Discrete Applied Mathematics* 76: 61–80

Crucitti P, Latora V, and Porta S 2006 Centrality measures in spatial networks of urban streets. *Physical Review E* 73: 036125

Curtin K M 2007 Network analysis in geographic information science: Review, assessment and projections. *Cartography and Geographic Information Science* 34: 103–11

Demšar U, Špatenková O and Virrantaus K 2007 Centrality measures and vulnerability of spatial networks. In *Proceedings of the Fourth International ISCRAM Conference*, Delft, The Netherlands

Finnish Road Administration 2007 Digiroad: A National Road and Street Database. WWW document, http://www.digiroad.fi/en_GB/

Freeman L C 1979 Centrality in social networks: Conceptual clarification. *Social Networks* 1: 215–39

Girvan M and Newman M E J 2003 Community structure in social and biological networks. *Proceedings of the National Academy of Sciences of the USA* 99: 7821–26

Gorman S P and Kulkarni R 2004 Spatial small worlds: New geographic patterns for an information economy. *Environment and Planning B* 31: 273–96

Gorman S P and Malecki E J 2000 The networks of the Internet: An analysis of provider networks in the USA. *Telecommunications Policy* 24(2): 113–34

Gorman S P, Schintler L, Kulkarni R, and Stough R 2004 The revenge of distance: Vulnerability analysis of critical information infrastructure. *Journal of Contingencies and Crisis Management* 12(2): 48–63

Grubesic T H, O'Kelly M E, and Murray A T 2003 A geographic perspective on commercial Internet survivability. *Telematics and Informatics* 20: 51–69

Grubesic T H and Murray A T 2006 Vital nodes, interconnected infrastructures and the geographies of network survivability. *Annals of the Association of American Geographers* 96: 64–83

Guida M and Maria F 2007 Topology of the Italian airport network: A scale-free small-world network with a fractal structure? *Chaos, Solitons and Fractals* 31: 527–36

Hillier B and Hanson J 1984 *The Social Logic of Space*. Cambridge, Cambridge University Press

Jiang B and Claramunt C 2004a A structural approach to the model generalisation of an urban street network. *GeoInformatica* 8: 157–71

Jiang B and Claramunt C 2004b Topological analysis of urban street networks. *Environment and Planning B* 31: 151–62

Jungnickel D 2005 *Graphs, Networks and Algorithms* (Second Edition). Berlin, Springer-Verlag

Kohvakka K and Valtonen V 2004 *Fire Attack 2020: Views on Damage Effects and Preparedness in the Society* (in Finnish). Helsinki, The Defence College, Department of Tactics, Publication Series 2, 1(2004): 113

Latora V and Marchiori M 2004 How the science of complex networks can help develop strategies against terrorism. *Chaos, Solitons and Fractals* 20: 69–75

Miller H J and Shaw S L 2001 *Geographic Information Systems for Transportation*. New York, Oxford University Press

Murray A T, Matisziw T C, and Grubesic T H 2007 Critical network infrastructure analysis: Interdiction and system flow. *Journal of Geographical Systems* 9: 103–17

Newman M E J 2007 The mathematics of networks. In Blume L E and Durlauf S N (eds) *The New Palgrave Encyclopaedia of Economics* (Second Edition). London, Palgrave Macmillan

O'Kelly M E and Grubesic T H 2002 Backbone topology, access, and the commercial internet. *Environment and Planning B* 29: 533–52

de Nooy W, Mrvar A, and Batagelj V 2005 *Exploratory Social Network Analysis with Pajek*. Cambridge, Cambridge University Press.

Rosvall M, Trusina A, Minnhagen P, and Sneppen K 2005 Networks and cities: An information perspective. *Physical Review Letters* 94: 028701.

© 2008 The Authors. Journal compilation © 2008 Blackwell Publishing Ltd

Skupin A and Fabrikant S I 2007 Spatialization. In Wilson J P and Fotheringham A S (eds) *The Handbook of Geographic Information Science*. Oxford, Blackwell: 61–79

Thill J-C 2000 *Geographic Information Systems in Transportation Research*. Amsterdam, Pergamon

Watts D J and Strogatz S H 1998 Collective dynamics of 'small-world' networks. *Nature* 393: 440–2

Wheeler D and O'Kelly M E 1999 Network topology and city accessibility of the commercial internet. *The Professional Geographer* 51: 327–39

Worboys M and Duckham M 2004 *GIS: A Computing Perspective*. London, CRC Press

YETT 2006 The Strategy for Securing the Functions Vital to Society 2006, Governmental Resolution. WWW document, http://www.defmin.fi/index.phtml?l=en&s=335

© 2008 The Authors. Journal compilation © 2008 Blackwell Publishing Ltd