Res. Lett. Inf. Math. Sci., 2007, Vol.**11**, pp23-32
Available online at http://iims.massey.ac.nz/research/letters/

Node Importance Ranking and Scaling Properties of some Complex Road Networks

K.A. Hawick and H.A.James
Institute of Information and Mathematical Sciences
Massey University – Albany, North Shore 102-904, Auckland, New Zealand
Email: {k.a.hawick,h.a.james}@massey.ac.nz
Tel: +64 9 414 0800 Fax: +64 9 441 8181

The scaling and other quantifiable properties of a network have recently been proven valuable in understanding the robustness and vulnerability properties of various societal and infrastructural networks. In this paper we revisit the algorithms for computing various quantifiable properties of a planar road network and consider the algorithmic complexity and scalability in the light of recent technological advances. We compute properties for a sample of interesting trunk road networks and discuss their applicability in determining the relative importance or criticality to the whole network of a particular node. We discuss the implications of present and anticipated technological capabilities in calculating properties for anticipated network sizes in the light of 64-bit computer architectures and commodity parallel computing.

Keywords: circuits; all-pairs distance; scaling; network; complex system.

1 Introduction

Increasing traffic congestion in almost all parts of the world are driving greater levels of interest in road network problems [2]. A particular issue is that of the relative importance or criticality of various nodes in a road network. Traffic resource managers are interested in the effect of a particular node (place) being obstructed or temporarily removed from a network and planners may be interested in asking where new nodes or roads can be added to greatest effect for a limited extra resource. Road networks can be modelled as connected graphs [1] with various statistical vertex and linking properties.

In this paper we consider some planar graphs [5] representing main trunk routes in road networks. We are interested in what quantifiable metrics we can make on such graph, and the computational costs of doing so. In particular we consider the tradeoffs between obtaining accurate topological information using graph algorithms versus obtaining cheaper but less robust information from local measurements using the spatial or Euclidean geometry of a road-map.

We consider approximate trunk route maps for Scotland, and for the North and South islands of New Zealand. We show how these networks have distinct properties and how some nodes are significantly more important than others.

We explore graph theoretic metrics such as the Dijkstra all-pairs average distance and the number of elementary circuits in the graphs and review these as indicators of the relative importance of different nodes or places. In this paper we focus on the vertices or nodes in the graphs but do not consider detailed traffic flow and are not concerned with weighted graphs. We are interested in what happens to the network as a whole when a particular location is effectively removed. This question of overall network robustness and vulnerability has some relevance to current defence-related world issues.

In section 2 we consider some of the graph and geometric properties of vertices in a road network and in section 3 we discuss implications for their calculation given typical road network data and sizes.

In section 4 we present results for our analysis of trunk road networks for Scotland and the North and South New Zealand islands. We also note comparisons between topological connectedness and Euclidean geometrical proximity and how this can be used to define under- or over-connected nodes, and hence used to contribute towards a judgement of their relative importance in our conclusions in section 5.

2 Measures

There are various graph-theoretical measures or metrics that we can compute for particular networks. We discuss these in the context of planar graphs formed by some road network data sets.

Figure 1 shows a rendering of the Scottish road network produced using our *GraViz* Graph Visualisation Tool [6]. The network of 121 vertices located on the Scottish mainland are connected by 175 bi-directional arcs forming the main trunk road network. The rendering shows the nodes colour coded by degree – this being the simplest localised measure of connectivity we can compute. This network is fully connected so the minimum degree of any node is 1. "Perth" has the highest degree of any node (7) and the mean degree over the whole network is 5.785. Two nodes have been distinguished - "Hawick" and "John O' Groats" and the shortest (in terms of number of hops) path between them is shown (21 hops).

The Dijkstra all-pairs distance gives a measure of how close nodes on the whole network are to one another and is computed by average the pair-wise shortest path lengths over all pairs of nodes. The Scottish mainland trunk network as shown has an all-pairs average Dijkstra distance of approximately 8 hops. Figure 4 shows the histogrammed distribution of path lengths (in terms of hops) for the Scottish road network.

In this paper we consider only fully connected networks. An obvious metric of criticality is whether a particular node bisects the graph. Some nodes when removed will break the network into two or more clusters. We can identify and label the clusters or components of such a graph[4]. In the Scottish road network removal of "Kyle of Lochalsh" and hence access to the Skye road bridge will bisect the network in two, leaving the Isle of Skye separate from the mainland. Bisection has an anomalous effect on the interpretation of the Dijkstra all-pairs metric. This can lower the all-pairs distance due to the weighting effect of short distances in a small cluster component. For this paper we restrict attention to the largest component and report all-pairs averages over that.

Most real road and other transportation networks evolve for all sorts of social and geographical reasons. There is therefore often little correlation between the density of vertices in the sense of the embedding 2-dimensional Euclidean geometry and that of the graph topology. There may however be some positive correlation between the two measures – on average. We return to this issue below. Scotland in fact shows delineated geographic regions that are illustrated quite well in the rendering. The lowlands have a relatively flat landscapes and high densities of well connected nodes – whereas the Highlands have their inter node connectivities controlled more by mountain regions and other obstacles and general have a lower connectivity degree and node density.

Another metric for the network as a whole is the number of elementary circuits. A circuit or loop is defined as a path starting and finishing on the same vertex, and an elementary circuit has no crossovers and no repeat traversals of any arc (so figure of eight patterns are not elementary). The number of elementary circuits that pass through a particular vertex gives another quantifiable measure of how the vertex contributes to the overall connectivity of the graph. Although we restrict attention to fully connected graphs, so that all vertices contribute to some circuits – even

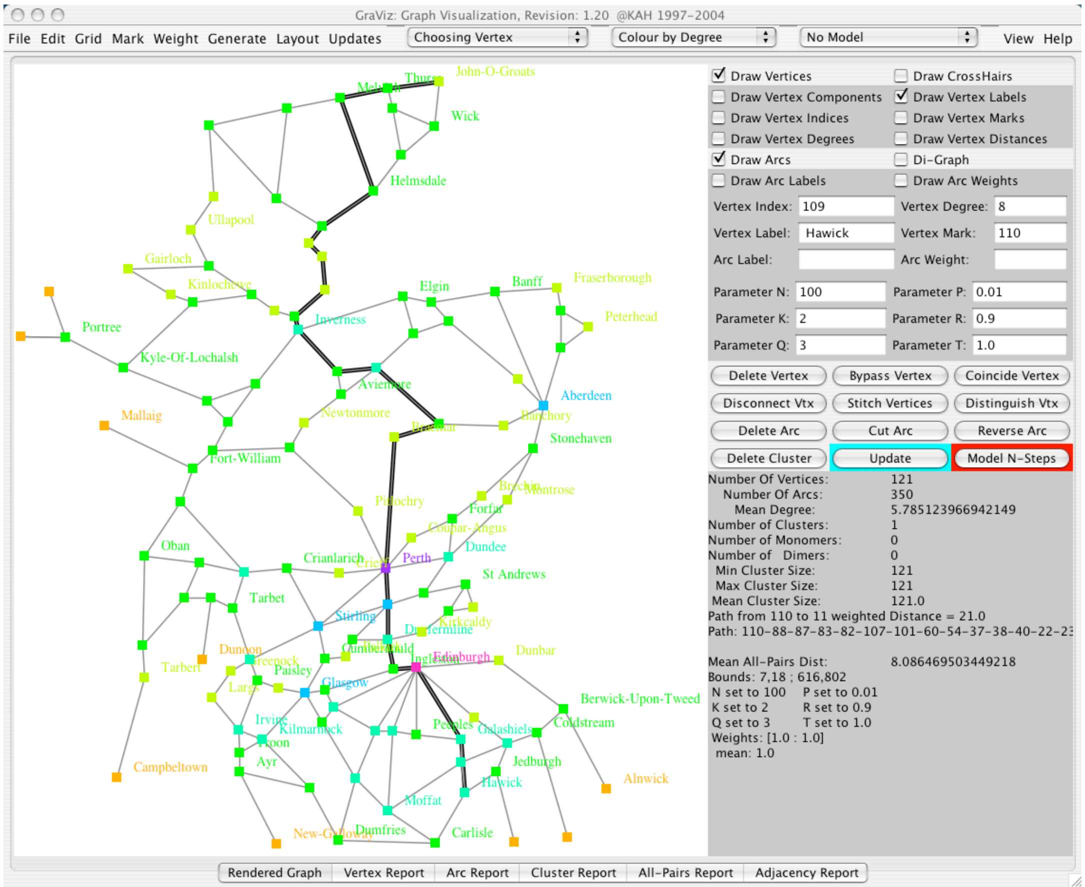

Figure 1: The Shortest Path (in terms of number of arcs traversed) from Hawick to John O'Groats

those trivial ones of length 2, by no means all graphs have all possible circuit lengths. In general the road networks we have studied have a very restricted distribution of possible circuit lengths, and in particular none had "Hamiltonian circuits" incorporating all vertices. Figure 5 shows the histogrammed distribution of circuit lengths in the Scottish road network.

A final measure can be constructed using the reachability of the vertices from all other vertices in the network. This is useful when considering **uni-directed arcs** as might be found in a road map with "one way streets" and can also indicate the level of a network's fragility or vulnerability to losing particular vertices. A robust network might display a phase transition to a regime of having unreachable regions. We do not consider this reachability index for our **bi-directional** trunk road networks however.

Table 2 summarises these static graph metrics for other road network data sets and some other graphs.

It is also possible to do a sensitivity or perturbation analysis using these metrics. The graph can have a vertex removed or disconnected and the effect of this on the path-lengths distance and number of possible circuits can be computed. We present this analysis on the Scottish and New Zealand road networks in section 4. These give us a number of measures that can be used as classification discriminators to rank vertices in terms of their importance.

3 Quantifiable Algorithms

It is important to consider precise definitions of what we mean by the paths and circuits properties of road networks. We describe algorithms for these in this section along with a review of the computational complexity for the algorithms needed to compute them. In analysis of an entire graph a metric some savings of order N can be made using appropriate book-keeping to record properties as we traverse all the nodes. This can alleviate the cost of some metrics but needless to add has little impact on NP algorithms like that for enumerating the elementary circuits.

Algorithms to compute the number of components in a graph are given in [4]. Dijkstra's algorithm for computing the shortest path between any two nodes is well-known and given in [3]. It is relatively trivial to adapt this algorithm to efficiently compute the all-pairs distance metric. Johnson's paper of 1975 [8] gives a stack-based algorithm for enumerating all the circuit in a graph. We have used a modification of this algorithm to construct node importance histograms rather than storing the enumerated circuits.

All these algorithms require an efficient means of traversing the vertex sets of the graphs and will typically also gain performance (prefactor if not complexity order) by having access to rapid lookup of arc properties such as arc-ij existence. For the size of networks involved in our study it is entirely feasible to store adjacency matrix tables rather than having to traverse neighbour lists by indirect addressing.

It is worth observing that although we have been able to enumerate and histogram the circuits in the 121-node Scottish road network – this took approximately 5 days' work on a 2GHz G5 processor and note also that the numbers involved would overflow 32-bit integers. We required the hardware support of 64-bit integers to manage this computation efficiently.

Metric / Algorithm	Order
Mean degree	N
Euclidean Centroid	N
Euclidean Radius of Gyration	N^2
Number of Component Clusters	N^2
All-Pairs Dijkstra Distance	N^3
Number of Elementary Circuits	NP

Table 1: Computational complexities for algorithms to compute the graph network measures.

Table 1 summarises the computational complexity of the algorithms we need consider here in terms of the number of vertices N.

In general we are severely limited in use of the algorithm for counting or enumerating the elementary circuits due to its intractability for anything but small networks of around 60 vertices and mean degree of around 6.

4 Model Network Results

Table 2 summarises the static properties of the road networks we analyse in this paper along with some other simple graphs of similar size for comparison. It is worth drawing attention to the simple lattice networks generated for comparisons. It was not feasible to compute the number of elementary circuits in a simple lattice of squares, triangles or hexagons even with numbers of nodes less than that of the Scottish road network data set. This emphasises the limited applicability of the circuits metric.

Nodes	Arcs	<Degree>	<All-Pairs>	Circuits	Description
60	160	5.333	6.6172	158,176	NZ South Island Network
63	186	5.905	6.2172	6,777,455	NZ North Island Network
121	350	5.785	8.0865	302,117,035,367	Scottish Road Network
49	168	6.857	4.5714	?	7x7 non-periodic squares
124	342	5.516	8.3385	?	6x8 non-periodic Hexagons
100	522	10.44	5.3374	?	5x5 non-periodic Triangles

Table 2: Summary of key properties of the test networks. Some other (non-road) graphs of similar sizes are shown for comparison. The number of circuits is an interesting measure of graph complexity for instance.

We present an analysis of the all-pairs and circuits perturbation analysis of the road networks.

Figure 2 shows the effect on the mean all-pairs Dijkstra distance of the Scottish road network when a single vertex is removed. The analysis is only for the largest remaining component, so the effect of the perturbation analysis is always to increase the all-pairs distance metric. It is interesting to see how this analysis automatically ranks the importance of individual vertices in terms of their effect on the rest of the network. This analysis is more discriminating than a simple inspection of the vertex degrees. It reveals that while the highest degree node is also significant in terms of all-pairs contribution, in fact Inverness is more critical. Removal of Inverness has the highest disruptive effect on the all-pairs connectivity of the network. Another interesting insight is that despite it suggestively high degree, Glasgow is not so important since it has many nearby nodes that would provide "bypass" connectivity were it removed from the network. In this figure the vertex indices are the same as the raw data in the graph network data set and we can identify individual nodes.

Figure 3 shows the all-pairs distances ranked in ascending order. The three curves are for the Scottish and two New Zealand island road networks. Note that each has a distinctly different mean all-pairs value and indeed profile. In particular the distributions are seen to be asymmetric about their means and the Scottish set is particularly so. Lowly ranked vertices can be identified as edge and other poorly connected vertices which normally have the effect of raising the all-pairs metric. Their removal thus has the beneficial effect of lowering the average. Highly ranked vertices are those which are important to the overall connectivity. Their removal raises the average all-pairs metric. The mid-ranked nodes are the medium importance nodes which are not critical and can be bypassed or removed without significant effects on the overall network connectivity.

Figure 4 shows the asymmetrical histogrammed distribution of all-pairs path lengths present in the Scottish road network. This illustrates the regional Lowland/Highland effect discussed above but also emphasises that a mean all-pairs measure needs to be interpreted with some caution. The distribution is not normal and hence the mean is not conveying information about all possible paths equally.

Figure 5 shows the histogrammed distribution of circuit lengths present in the Scottish road network. The small excursion at the start of the curve is just due to the definition of an elementary circuit of length 2. It is significant that again this distribution is not symmetrical. The large numbers of circuits – which grows exponentially in the number of vertices, makes the curve smooth and the hump at length around 50 is therefore significant. The curve also shows a sharp cutoff in the maximum circuit lengths supported by the network. It is unfortunately computationally infeasible for us to present a perturbation analysis of the circuits due to individual vertices.

Figure 6 shows the perturbation analysis applied to the two New Zealand island road networks.

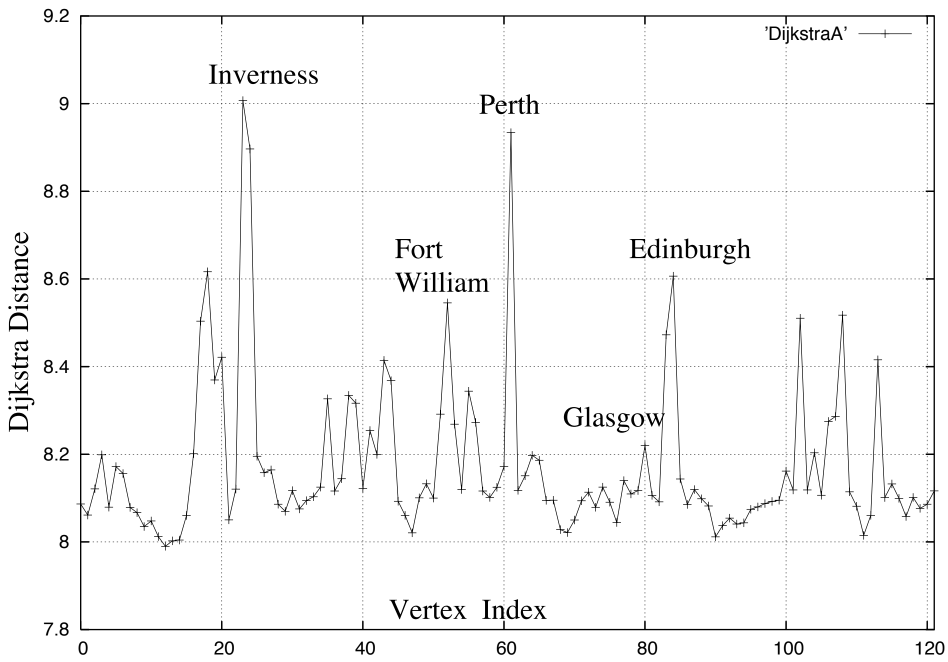

Figure 2: Dijkstra Distances for labelled vertices showing the most important nodes in the Scottish road network. The Dijkstra all-pairs distance has been measured for the entire network and the graph shows how this metric varies when a given node is "removed" from the network. Low values indicate better "connectedness" whereas high values show the impact of taking out particularly important nodes. Node the critical roles played by hub-cites like Perth, Inverness and Edinburgh. Glasgow registers (perhaps unfairly) as less critical just because of alternative nearby nodes.

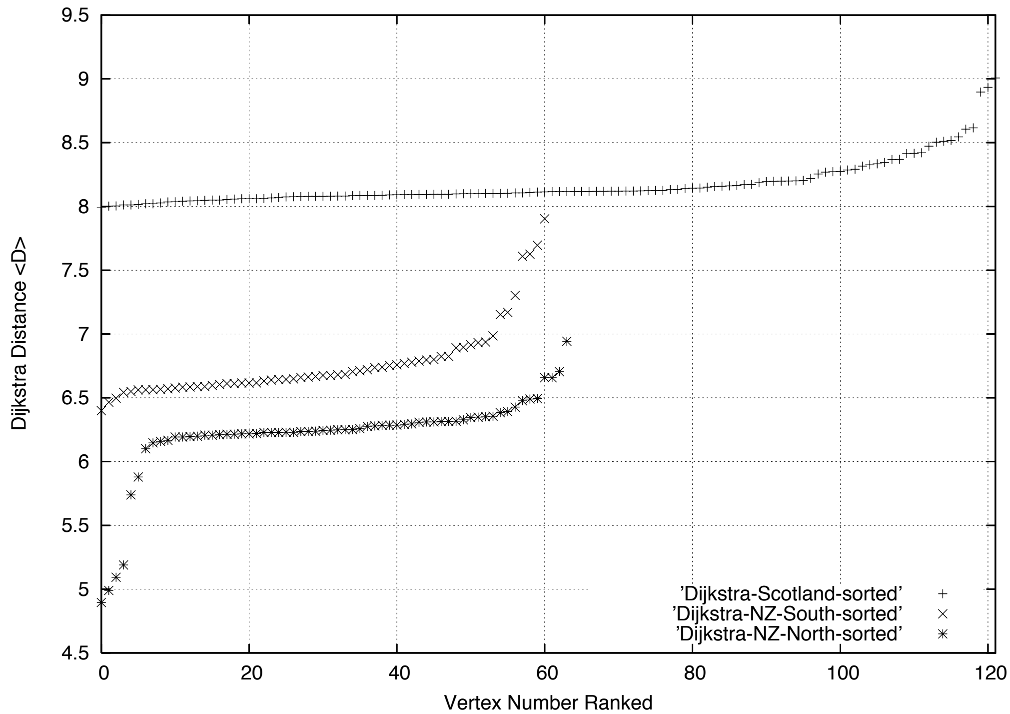

Figure 3: Ranked Dijkstra all-pairs distance for the whole (remaining) network with a particular vertex omitted. Removing some vertices will have a slight reduction of this metric from the average (an indication of their remoteness and unimportance), and removal of a select few vertices will significantly increase this distance metric – indicating their importance. The Scottish road network curve is noticeably more asymmetric than the NZ island curves.

The North and South islands have 63 and 60 nodes respectively but have 6,777,455 and 158,176 circuits respectively. The plots show the number of circuits left when the vertex labelled on the x-axis is removed from the network. The contribution curves are drawn on a logarithmic scale – note therefore the different scales for the two islands but also not the different characteristic distributions. The North island network is highly connected and quite robust against loss of particular nodes. The South island network is more sensitive to lost nodes – relatively. The circuits analysis identifies nodes like Kumara Junction (25), Hokitika (26) and Haast (27) as contributing a particularly high number of circuits towards the overall network. This is not obvious from simple degree considerations.

We can compute the average degree for the networks and compare this against the number of nodes that are in the (Euclidean) geometrical-proximity of a particular vertex. This is essentially a measure of the lost opportunity for local planners to build a connecting road between two nearby nodes. It has the particular advantage that it can be computed using computationally inexpensive metrics such as the degree. Figure 7 shows this measure used to automatically identify under-connected (compared to their local opportunities) nodes. The five nodes identified have more

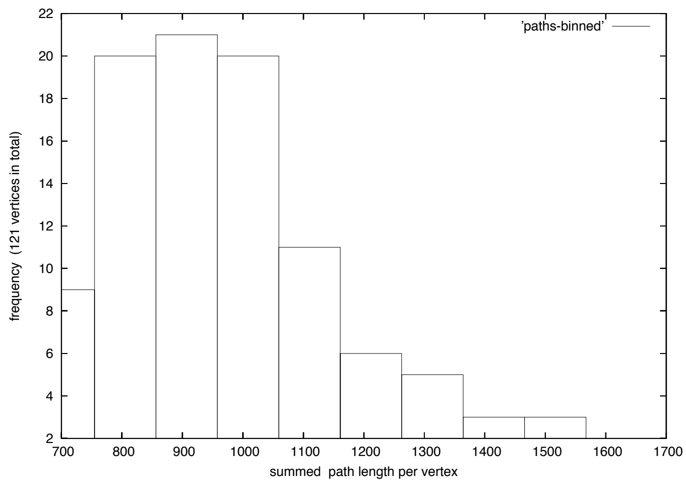

Figure 4: Asymmetrical distribution of the histogrammed path lengths in terms of numbers of "hops", for the Scottish road data set.

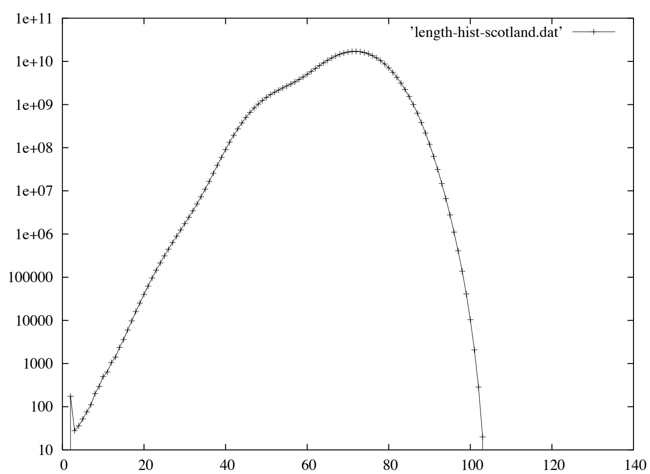

Figure 5: Asymmetrical distribution of the histogrammed circuits populations vs (topological) circuit-length (in number of "hops") for the Scottish road data set.

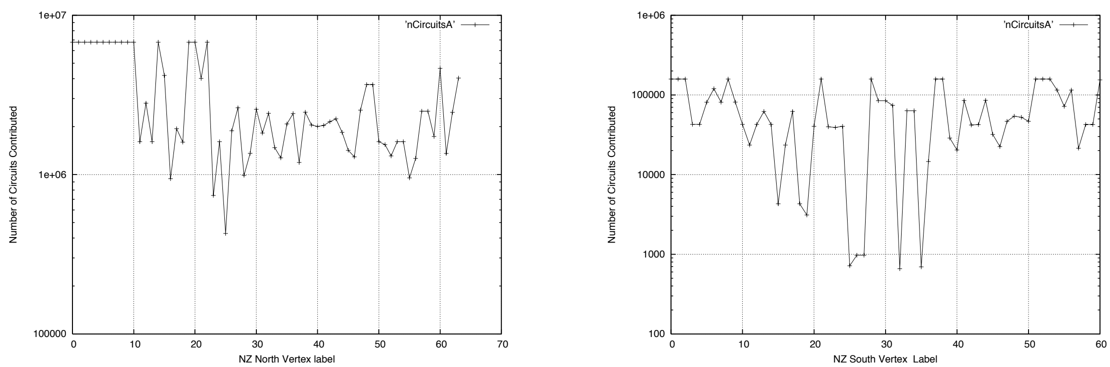

Figure 6: The contribution of each vertex (shown when it is removed) towards the number of elementary circuits in the **North** Island (top) and **South** (bottom) NZ Road network data.

nearby (geometrically) nodes than they actually connect to (topologically) and therefore could perhaps could be connected relatively cheaply.

5 Summary and Conclusions

We have presented a number of quantifiable graph metrics and showed how they can be used to analyse the relative importance of nodes in a road network with respect to the contributions nodes make to the overall network connectivity. As we have discussed some metrics can be computed cheaply as they have simple linear computational complexity. Unfortunately some of the most useful metrics have high order or worse still NP computational complexity.

We believe that analysis of the number of circuits in a network for instance is very powerful as an automatic discriminator to detect certain important nodes. It is however also possible to use heuristics derived from "computationally cheap" metrics based on the embedding Euclidean geometry to derive certain decision support information from a road network.

We believe there is further analysis to be done considering the full role of the asymmetries that typically occur in the distributions of circuit length or all-pairs path distances. It would also be worthwhile to extend the analyses we describe here to make use of Euclidean-distance weighted arcs.

Acknowledgements

The road networks described here are our approximations of trunk routes published by the UK and New Zealand Automobile associations. Transcription errors are however our own responsibility and the networks presented here are only indicative rather than definitive. This work has benefited from the use of Massey University's "Monte" compute cluster and thanks also to the Allan Wilson centre for the use of the "Helix" Supercomputer.

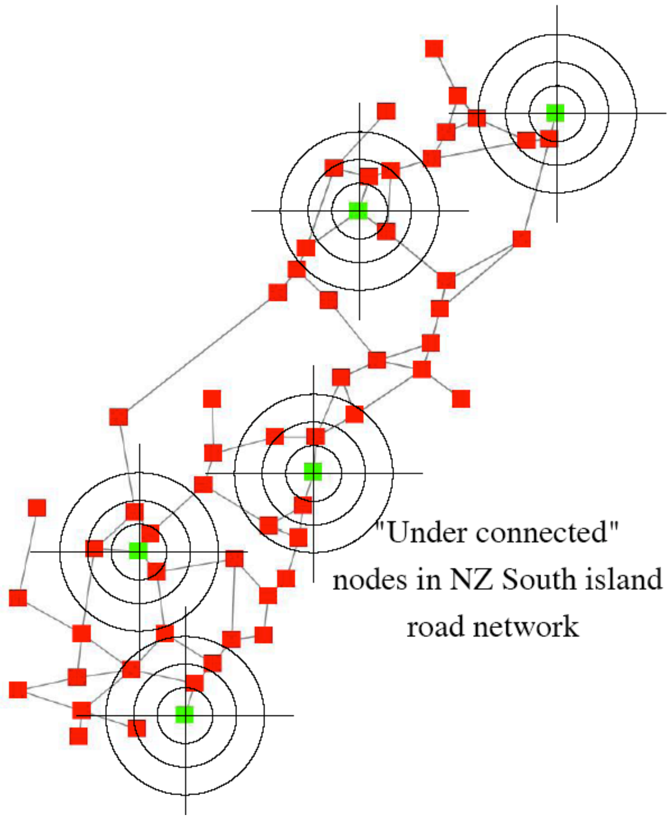

Figure 7: Under-connected vertices are identifiable automatically as those with an excess of Euclidean geometrically-near neighbours compared to actual graph-topological neighbours.

References

[1] Barabási, A.-L. *Linked*, Perseus Publishing, 2002. ISBN 0-7382-0667-9.

[2] Bovy, P.H.L. and Thijs, R. (Eds.) *Modelling for transportation systems planning*, DUP Satellite, 2002. ISBN 90-407-2274-9.

[3] Dijkstra, E.W. A Note on Two Problems in Connection with Graphs, *Numerische Math.*, (1959), 269-271.

[4] Gebremedhin, A.H. and Manne, F. Scalable Parallel Graph Coloring Algorithms *Concurrency: Practice and Experience*, Vol 12, 2000, pp 1131-1146

[5] Gould, R. *Graph Theory*, Benjamin/Cummings Publishing Company, Inc., 1988. ISBN 3-8053-6030-1.

[6] Hawick, K.A. *GraViz Graph Visualization and Analysis Program*, unpublished, 2004.

[7] Hawick, K.A. and James, H.A. Small-World Effects in Wireless Agent Sensor Networks, to appear in *Int. J. Wireless and Mobile Computing, Special Issue on Mobile Systems, E-Commerce and Agent Technology*, 2005

[8] Johnson, D.B. Finding all the Elementary Circuits of a Directed Graph, *SIAM J. Comput.* 4(1): 77-84 (1975)