Abstract for the FOURTH OPEN SOURCE GIS CONFERENCE - OSGIS 2012
Nottingham Geospatial Institute University of Nottingham
4th - 5th Sep 2012

An open source relational database schema and system for the analysis of large scale spatially interdependent infrastructure networks

Stuart Barr, Tomas Holderness, David Alderson, Craig Robson and Alistair Ford

School Civil Engineering & Geosciences, Newcastle University, Newcastle upon Tyne, U.K.
Corresponding author: Stuart Barr (s.l.barr@ncl.ac.uk).

Abstract

There is an increasing interest both academically, commercially and at a governmental level to spatially understand and model the performance of infrastructure networks at local, regional and national scales. Particular interest lies in understanding the vulnerability and probabilistic risk faced by large scale physical infrastructure networks to spatially localised extreme events such as natural hazards. In order to perform such analysis and modelling in a holistic manner one requires the ability to represent and model the dependencies and interdependencies that exist between different infrastructure networks, which in turn requires the ability to derive a ‘network-of-networks’ spatial representation and expression of physical infrastructure. Traditionally, spatial networks and their subsequent analysis have been performed within a GIS environment. However, the ability of GIS packages to explicitly represent large interdependent networks comprising of potentially many hundreds of thousands of nodes and edges has been limited. Moreover, the analytical network functionality provided by GIS packages often falls some way short of the complex analysis required to understand how failures propagate through individual networks, between dependent networks and feedback when interdependent relationships exist.

In order to address these issues we have developed a fully relationally compliant open source spatial database schema and a python coupled-interface to the NetworkX graph analysis package for the representation, encoding and analysis of networks and their dependencies/interdependencies. In particular we have used collaborative rapid-prototyping to develop our database schema for network representation and dependence/interdependence encoding using PostgreSQL and the spatial extension PostGIS. The schema is exposed through plpgsql functions, which are provided to the user wrapped in python class methods, forming the software API.

Figure 1 shows the entity relationship model of the PostgreSQL/PostGIS database scheme developed for the representation of geographical network dependencies and interdependence. In the database schema each separate unique infrastructure network comprises of three tables: a node table, an edge table, and an edge table that specifically contains only the edge geometry. Node tables describe infrastructure network hubs while the edges describe the linear links that exist between them. For an infrastructure network the geometry of the edges is stored separately from the edge table as multiple edges may exist between the same set of nodes which have distinct geometry; for example several electricity cables within an electrical transmission network may be connected to the same sub-stations forming a multi-graph representation. Interdependencies are represented within an interdependency edge table that links the nodes within two separate infrastructure networks and again have a separate geometry table to allow multiple edges between distinctive pairwise combinations of nodes within disparate infrastructure networks. Finally, recognition of a distinctive network is managed by a graphs table which encodes information on the name, node table, edge

table and the type of graph model (undirected, directed, multigraph etc.) of the networks within the database.

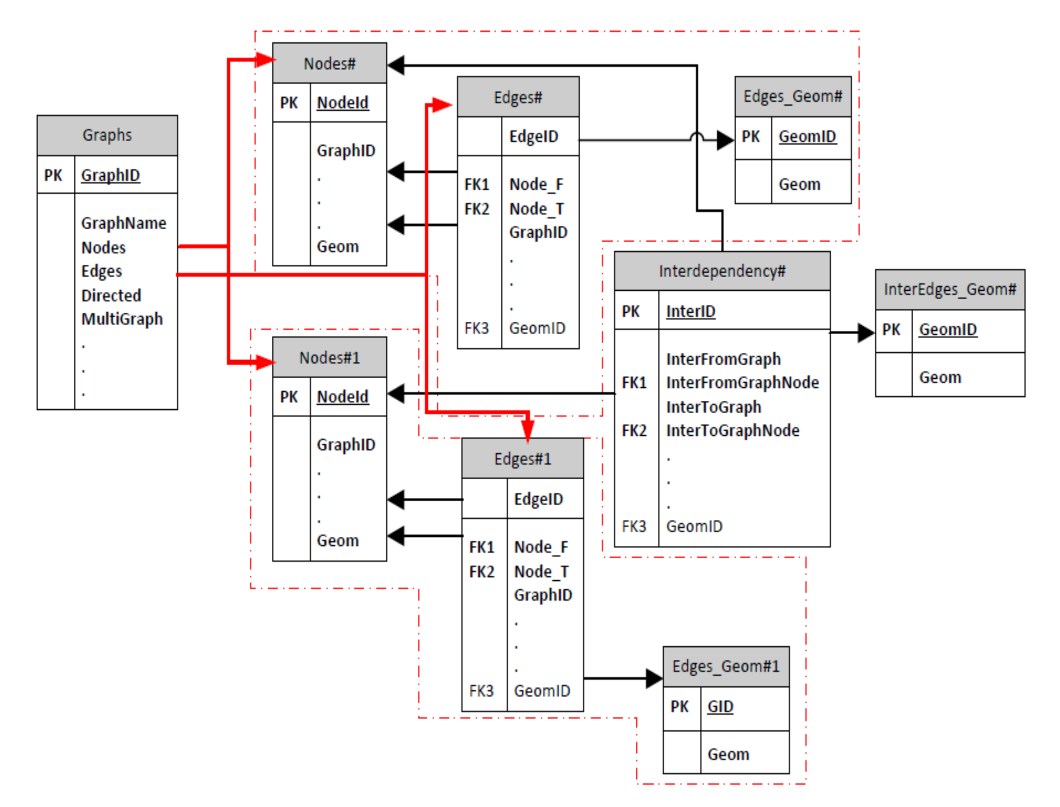

Figure 1: A generalised entity relationship model showing the tables and relationships developed in PostGIS for the representation of spatial infrastructure networks and their interdependencies.

In order to construct an infrastructure network model using the schema shown in Figure 1 a standard GIS layer representing a network but without network topology is required. This is used to create a table of Linestring geometry within the PostGIS database. In order to construct the network topology whilst retaining the geography of the edges, we have developed a series of python modules that utilise the open source graph/network package NetworkX and the geospatial data abstraction library (GDAL). NetworkX is a graph and network analysis package written in python but which does not as yet have an explicit representation of spatial networks. In our work we have developed a python module for NetworkX called nx_pgnet to construct spatial infrastructure models, with valid network topology. As NetworkX constructs network topology using only lists of edges and nodes, the nx_pgnet module ensures that a topologically and geographically valid network is created by encoding node and edge geometry as attributes within the NetworkX edges and nodes. Once NetworkX returns the topologically constructed network nx_pgnet then creates a new network model using the schema defined in Figure 1 and encodes the spatial representation of the network from node and edge attributes as PostGIS geometry. Thus, by using attributes of edges and nodes within NetworkX we can create the three tables in PostGIS required to represent an infrastructure network.

The database schema handles and represents the existence of interdependency between two nodes within separate networks in a similar manner to its handling of an edge between two nodes in a single network. Interdependencies are stored within instances of the Interdependency and Interdependency_Edge tables, which are analogous to the Edge and Edge_Geometry tables. This approach effectively enables both the user of the schema and the connected NetworkX python package to consider interdependencies as a network themselves. The interdependencies can be created and stored within the database, for example, via the use of one of the heuristic pre-processing functions, such as *Find_Nearest_NodeB_to_NodeA_by_NN*. A physical interdependency would be created as a result of executing this function, inferring some connection between two nodes of different networks based on proximity.

In addition to the python module nx_pgnet we have also developed a number of PostGIS functions within our database schema for data pre-processing purposes which utilise spatial heuristics in order to ensure that edges snap to nodes within disparate GIS layers that represent point and line infrastructure assets respectively. Using nx_pgnet an initial network model is encoded which can then be checked for topological completeness. The heuristic functions provided within the database schema can be executed to allow users to try and produce a topologically and geographically-valid network. The spatial heuristics currently employed in this infrastructure construction network toolbox comprise of:

I. A function that connects edge “ends” to nodes that are related by some relationship between two attributes; one attribute in the node table, and one attribute in the edge table - Connect_Edges_to_Node_ByRelationship;

II. A function that connects edge “ends” to each node that is within a specified search distance - Connect_Edges_To_Node_BySearch;

III. A function that connects the nearest point on an edge (which may or may not be an edge end) to each node that is related by some relationship between two attributes; one attribute in the node table, and one attribute in the edge table - Connect_Nodes_TO_Point_On_NearestEdge_Like;

IV. A function that combines all the geometries of duplicate edges that may result from running one of the above functions and merges them into a single edge geometry record - Detect_and_Combine_Duplicate_Edges; and,

V. A function that finds the closest node of nodeset B to each node in nodeset A - Find_Nearest_NodeB_to_NodeA_by_NN.

Figure 2 shows the result of running a number of the heuristic network construction tools outlined above. A series of gas transmission pipes that do not connect within the original GIS layers have been corrected to form a new topologically and geographically-valid representation of the gas transmission network. In the case of Figure 3 the nearest neighbour search between the node tables of two different infrastructure networks (electricity transmission substations (black lines) and high pressure gas transmission network (brown lines)) has been performed and a series of dependency/interdependency edges constructed.

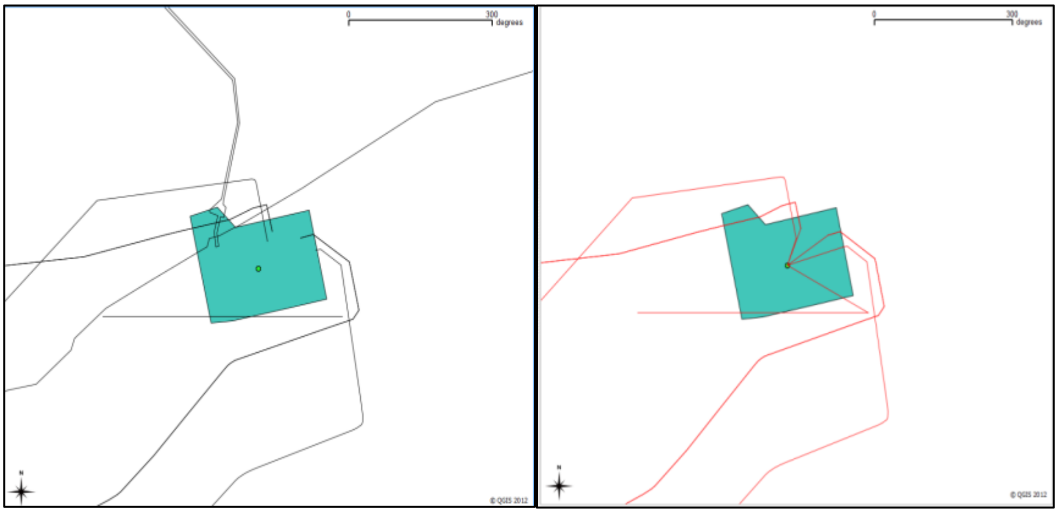

Figure 2: An example of using Connect_Edges_To_Node_BySearch to construct a valid connected spatial network model for gas pipe transmission from a pressured pumping station.

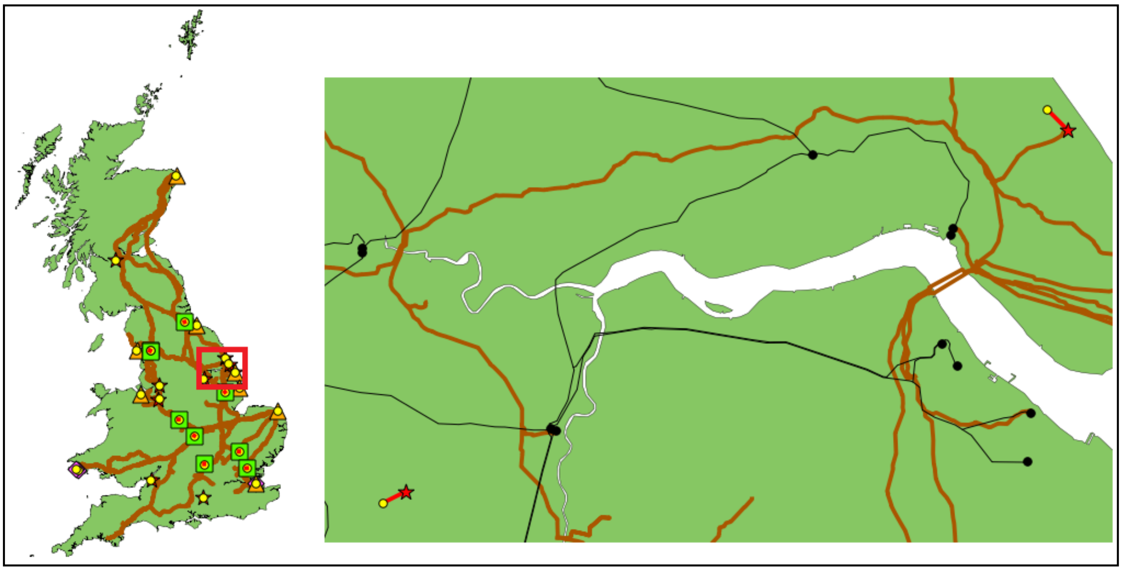

Figure 3: An example of using Find_Nearest_NodeB_to_NodeA_by_NN to construct spatial heuristically estimated dependency/interdependency edges between electricity substations and components of the high pressure gas transmission network.

To demonstrate the utility of the database schema and the related analysis tools developed using NetworkX and PostGIS we have undertaken a series of network and interdependent network analysis tasks. The first involves the construction of a network model of transport networks (Road, Rail and Light Rail) for the greater London area. To analyse this we have developed a further python module for transport network analysis, which includes generalised cost network analysis functionality. This has been used to derive an origin/destination (OD) matrix of generalised cost between the 633 wards of London on the basis of the road network encoded in the PostGIS database that contains 61,944 edges and 53,198 nodes. In the second we implement a simple failure model between the interdependencies of the electricity substation transmission network for the south-east and the London underground. To study the resilience of the London underground with respect to the supply of electricity we employ NetworkX graph-metrics to characterise how topological connectivity changes as a function of an increasing number of randomly assigned failures in the electricity substation network.

Acknowledgements

This research is sponsored under the UK EPSRC programme grant “Infrastructure Transitions Research Consortium“ (no: EP/I01344X/1) and EPSRC platform grant in “Earth Systems Engineering” (no: EP/G013403/1)