

A star schema for utility network analysis and visualisation in a Geo-business intelligence environment

Sasha Sadegholvaad^{a*}
Rohan Wickramasuriya^a
Jun Ma^a
Pascal Perez^a

Abstract: Utility network analysis is an established area of research in Geographic Information Science (GIS), but it is yet to feature in a Business Intelligence (BI) environment. Inclusion of this capability in BI can be achieved by modelling a utility network as a star schema. Star schema, which is a well-known data model used in data warehousing for BI solutions, organizes data into several fact tables referencing one or more dimension tables. This simplifies joins to provide fast execution of queries. Furthermore, star schema enables the analysis of data from multiple angles (slicing and dicing). However, modelling spatial data as star schema is still in its infancy for two main reasons, (a) only very recently have researchers started appreciating the importance of GIS capabilities in BI, (b) specific challenges associated with introducing geometry data with complex topological relationships into star schema. In this paper, we present a star schema to model geometric utility networks such as electricity, water and sewer systems. Our schema provides two-way benefits; it brings in an important new capability to BI in terms of spatial data analysis, and it gives non-technical users an opportunity carry out complex utility network analysis in an easy-to-use BI environment. Finally, we present an application of the star schema in service vulnerability assessment for electricity networks.

Key words – Infrastructure networks; Business Intelligence; Star schema.

I. Introduction

Geographic Information Systems (GIS) provides rich capabilities for network analysis resting strongly upon the theoretical basis of the mathematical subcategories of graph (network) theory and topology.¹ Several different methods and techniques are available within graph theory that enables one to describe measure and compare both individual graphs and groups of graphs in order to demonstrate their properties. Most of the mainstream GIS software tends to incorporate various subsets of these techniques in dedicated, domain specific network analysis modules.

^a SMART Infrastructure Facility, University of Wollongong, Northfields Ave, Wollongong NSW 2522, Australia

^{*}Corresponding author: E-mail: Sasha.Sadegholvaad@uowmail.edu.au
<http://dx.doi.org/10.14453/isngi2013.proc.40>

Network-based applications in GIS are usually geared towards modelling infrastructure networks with which the general public interacts on a regular basis. As long as networks of interest contain geographic elements, GIS tools have the ability to model and map them, for instance transportation, electricity and water networks which are in fact the main interest of this particular study.¹ Transportation network applications of GIS have experienced a significant interest over the last decade. Google Maps and Mapquest internet services are two good examples for such globally used facilities that provide route planning services and location services.

The other area of interest in this study is Business Intelligence (BI) which has traditionally been used by firms to analyse business data with the aim of obtaining Key Performance Indicators of interest.⁴ BI possesses powerful tools that can (a) harness disparate data sets in optimized data warehouses and (b) build highly interactive, online accessible reports and dashboard utilizing data stored in warehouses. BI reports and dashboards are easy to understand and use, hence ideal for use as public domain policy support tools. However, BI has traditionally lacked the capabilities to analyse and visualize geospatial data. Thankfully, BI community is realising the importance of integrating geospatial capabilities into BI software, particularly in the face of mounting geospatial data made available by new technologies like satellite remote sensing and Global Positioning Systems.

One of the primary structural components of any BI project is a data warehouse. Star schema is the most common schema design used in data warehouses that could be used to generate multidimensional cubes for BI analytics.⁵ The “must use” of star schema to build Online Analytical Processing (OLAP) cubes is one of many reasons that such schema design is preferred over the conventional one. Star schema organises data into one or more fact tables referencing any number of dimension tables. A star schema can simplify join queries and provide capability to analyse data from multiple dimensions, thus enabling a user to perform various drill down, roll up, slice and dice operations on data. Short responding time and processing in the same way that human process data are other advantages of star schema design.

Although star schema has several advantages over conventional normalised schema designs, it still faces challenges when it comes to introducing geometry data with complex topological relationships. Siqueira in his study aimed to tackle one of these challenges and proposed a benchmark (Spadawan) for spatial data warehouses with several appealing, yet imperfect, attributes.⁶ Spadawan (a) produces datasets containing points and polygons in spatial characteristics, (b) facilitates the assessment of spatial roll-up and drill-down processes, (c) enables the investigations of spatial data redundancy in spatial data warehouse by providing data schemas with spatial hierarchies and spatial dimensions, and (d) allows the modification of data volume and the spatial attribute selectivity.

Though there have been some attempts to exclusively model spatial data in a star schema, infrastructure networks have not been considered in any of these studies. Given the importance of infrastructure networks to governments at various levels, any BI tool built as a policy support tool needs to be able to store, analyse and visualize data related to infrastructure networks. Hence, the objective of this study is to design and demonstrate a star schema that can store

spatial topological networks in a data warehouse to support network-based analysis in a BI environment.

II. Design of network star schema

We designed the network star schema in a way that it can be connected to other schema available in the data warehouse using common dimension tables. Figure 1 illustrates the generic network star schema that we designed and its association with another schema that is designed to capture utility use (e.g. electricity, water) at geographic areas defined by the Australian Bureau of Statistics (ABS) in 2011.

The dimension table models the geographic hierarchy adopted by ABS in its 2011 census. This dimension table also supports postcode-based analysis. Note that this dimension table contains polygon geometry fields to store spatial extents of each census area. The dimension date models the date based on the ordinary calendar and the Australian season and financial calendar. This dimension enables multi temporal analysis of utility consumption. The fact table weather models temperature and rainfall derived from Bureau of Meteorology (BOM) daily records. The fact demography models demographic information extracted from and generated based on ABS census results.

In our data warehouse, each utility networks is modelled using a separate star schema. Figure 1 depicts a generic network star schema. The network schema consists of 10 fact tables and 7 dimension tables. DIM_Network_Main and DIM_Node tables contain the geometry data associate with links the nodes, respectively, in the modelled infrastructure network. This schema allows the user to perform several types of analysis: tracing upstream or downstream from a node or link, querying all geographic areas served by an asset, querying all assets linked to a given geographic area, and running what if scenarios such as finding all affected geographic areas given the failure of a particular asset.

To populate the network star schema with appropriate data, we relied on ArcGIS desktop GIS software. First, the directionality of the network flow was assigned to links. A python script was written to loop through all the elements (links and nodes) of the directional network in order to generate upstream/downstream links, nodes and geographic areas served.

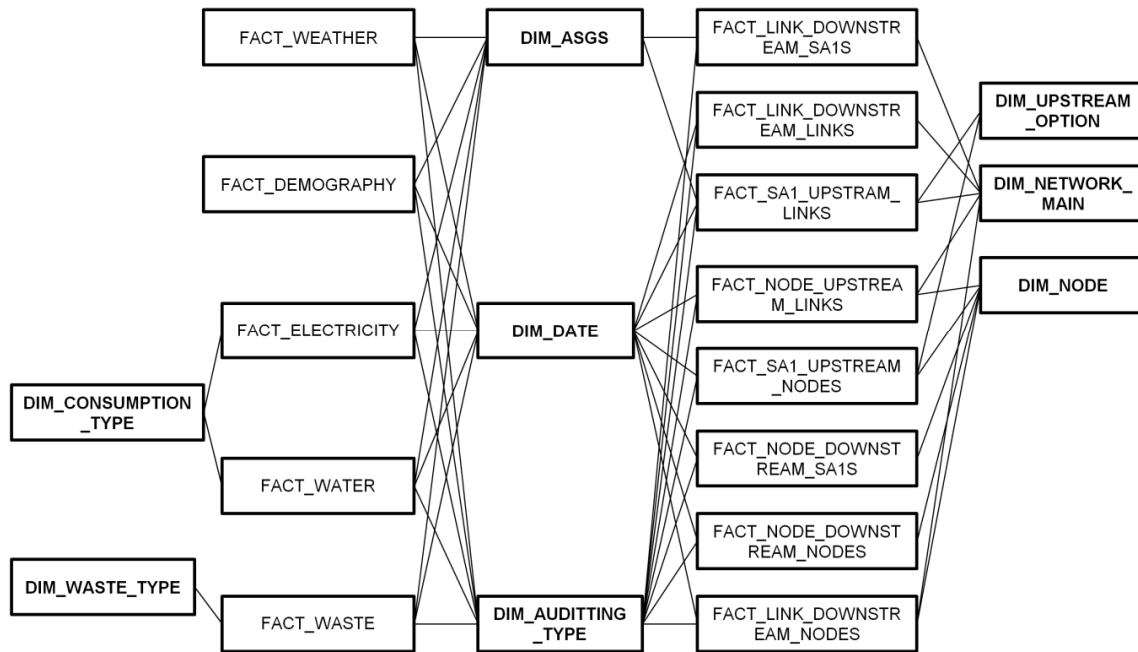


Figure 1: Overview of the star schema used in SID.

III. Star schema in action

In this section, we briefly explain one application of the network star schema that we designed. Service vulnerability is a critical aspect of infrastructure network management. Hence, tools that allow planners to perform service vulnerability assessments on infrastructure networks are invaluable. We built a map-based dashboard that enables planners to perform this kind of analysis on an electricity network modelled using our network star schema (Figure 2).

In Figure 2, blue bubbles represent electricity substations while conductors are represented as red lines. Hovering over a substation reveals ancillary information (such as asset ID and operating voltage) about the substation. Clicking on a substation brings up a drill through report, which again is map-based, that illustrates potentially affected postcodes given a failure of the selected substation. By hovering over any postcode in the drill through report, the user can reveal information such as total annual residential and commercial electricity usage and total population for that postcode (Figure 3).

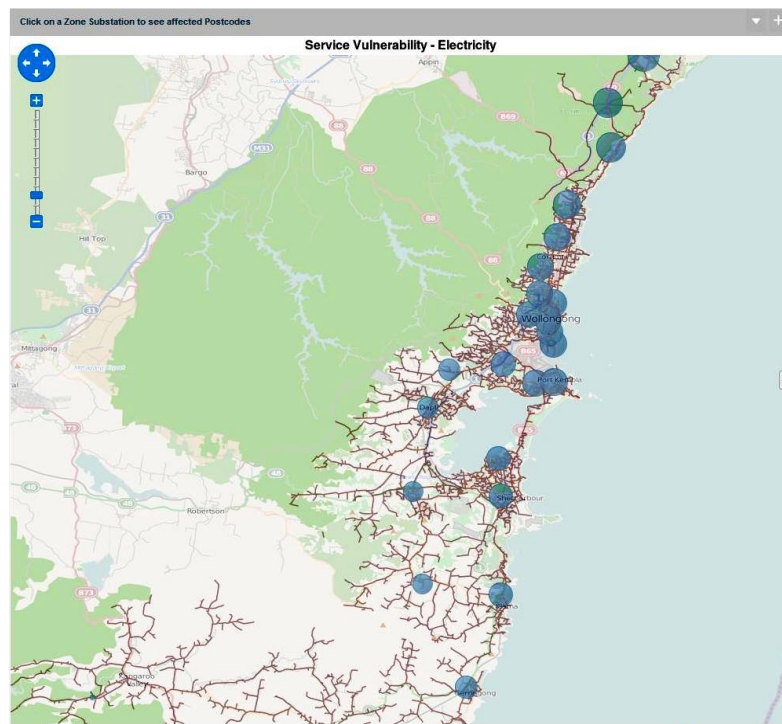


Figure 2. Map-based dashboard for service vulnerability assessment.

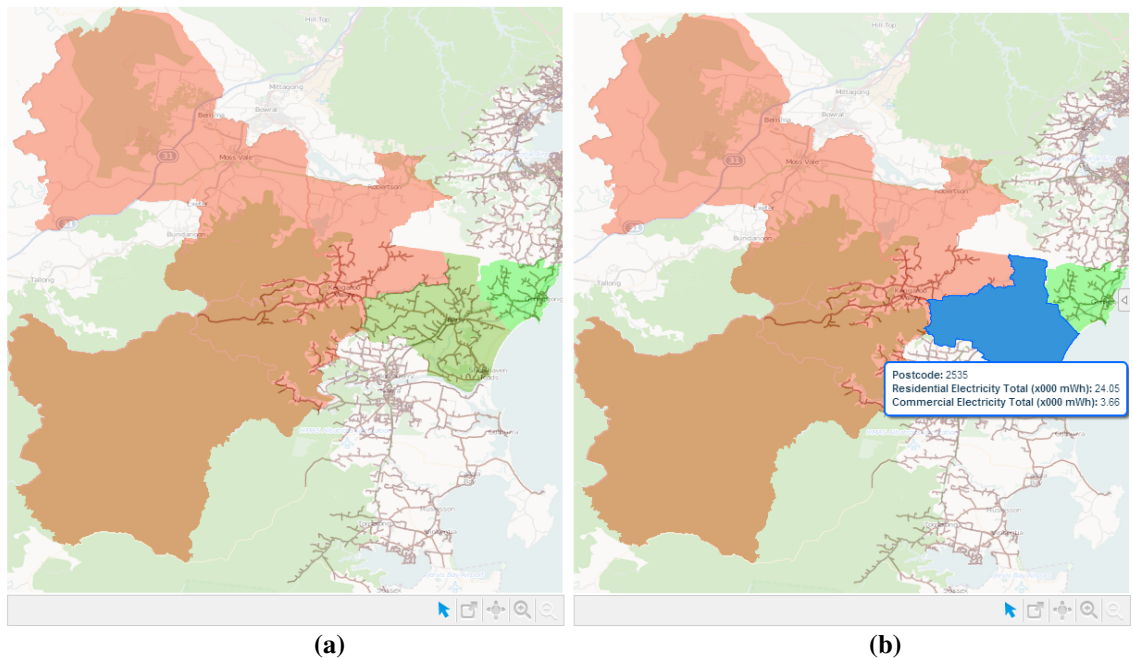


Figure 3. Postcode level affected area maps (a) Postcode map only; (b) Postcode map with demographic and electricity consumption usage details.

IV. Conclusion

Network analysis is a useful tool for infrastructure network management. Geographic Information Systems (GIS) traditionally incorporated capabilities to perform advanced network analysis. To use network analysis in GIS software, a user must possess advanced knowledge and skills. Even for advanced users, these capabilities are available in desktop GIS packages, limiting the usefulness of such analysis.

Business Intelligence (BI) software, on the other hand, consists of tools that are easy to use and understand, accessible through internet, and able to build visually rich, highly interactive content. BI is originally developed as a concepts and tools to generate Key Performance Indicators (KPI) for businesses striving to maximize their profits. BI, therefore, is a compelling alternative to GIS for network analysis, especially in a policy making environment. One critical issue though is the lack of proper geospatial capabilities in contemporary BI software.

At the heart of a BI project is an optimized data warehouse designed mainly using a star schema. The trick to unlock the power of BI for network analysis lies in successfully modelling a spatial topological infrastructure network using a star schema. In this study, we propose a star schema that can incorporate geometry data along with other network attributes to enable network analysis in BI. We demonstrate this star schema in action using a simple service vulnerability assessment on electricity networks.

References

- ¹Curtin, K.M., "Network analysis in geographic information science: review, assessment, and projections," *Cartography and Geographic Information Science*, Vol. 34, No. 2, 2007, pp. 103-111. <http://dx.doi.org/10.1559/152304007781002163>
- ²Scarponcini, P., "Generalised model for linear referencing in transportation," *Geoinformatica*, Vol. 6, No. 1, 2002, pp. 35-55. <http://dx.doi.org/10.1023/A:1013716130838>
- ³Franklin, C., "An introduction to geographic information systems: Linking maps to databases," *Database*, Vol. 15, No. 2, 1992, pp., 12-15.
- ⁴Angelaccio, M., Buttarazzi, B., Basili, A., & Ligouri, W., "Using geo-business intelligence to improve quality of life," *IEEE First AESS European Conference on Satellite Telecommunications (ESTEL)*, 2012, pp., 1-6.
- ⁵Poolet, A.M., "Discover the star schema," *SQL Server Pro Magazine*, Vol. 9, No. 7, 2007, pp., 29.
- ⁶Siqueira, T.L.L., Ciferri, R.R., Times, V.C., & Ciferri, C.D.A., "Benchmarking spatial data warehouses," *Data Warehousing and Knowledge Discovery, Lecture Notes in Computer Science*, No. 6263, 2010, pp., 40-51.