Vertical Topology – the last big challenge

Decades after GIS systems were created the creation and maintenance of vertical topology is still not well serviced.

Vertical topology is the correct alignment of geographic features across multiple themes, and is important because it makes spatial analysis more accurate and improves cartographic presentation.

The fact that vertical topology is not well serviced is because topological processing operations require very complex programming algorithms, with a great many tolerances put into place to test a wide variety of geometry.

Whilst each topological algorithm implements a succinct operation, it quickly becomes a spatial statistical game of immense proportions as for any given node or segment there are several routines that can come into play.

This article examines various aspects of software solutions that could be used to (re-)gain vertical topology.

State of play with Vertical Topology in GIS

Administration, statistical and geographic themes are derived from primary spatial themes like the digital cadastre, road, river and railway centerlines. The spatial accuracy of these primary themes are constantly being improved leaving the derived themes behind - causing the vertical topology problem, e.g. selecting parcels in a zoning category may select parcels outside the zooming polygon, simply because the zoning polygon's edge over hangs the extra parcels.

Primary spatial custodians, their clients and stakeholders have the following vertical topology problem categories:

- **Cost effective data capture** – ignoring vertical topology issues speeds data capture, but this is only effective if vertical topology can be restored at reasonable financial cost at some later time.
- **System performance** - GIS data structures designed for performance are at odds with those designed to assist maintain vertical topology, and visa versa.
- **Compatibility** – the custodians and their clients/stakeholders have different GIS systems, versions and local configurations, projections, spatial precisions and business goals.
- **Integration** – the custodians and their clients/stakeholders are looking for a seamless and integrated solution to suit the organisation's operational goals.

These translate into the following specific challenges for GIS managers.

Where to validate and correct Vertical Topology

There are several logical locations to validate and fix vertical topology.

One would be on the desktop at the time of capture. Creating themes from digitizing processes would benefit from a topology cleaning process, removing dangles, snapping nodes, giving the operator instant feedback as to what can and cannot be fixed automatically, leading to a reduced 'on-cost' of capturing topologically correct data.

Another would be on the desktop as part of some greater clean-up project.

Another logical location would be on the servers at the time the data is posted to the database, ensuring that the data in the database is at least topologically valid.

Implementation across a file based and database based GIS

The volume of data held in many organizations means that a file-based GIS is more appropriate. In such situations it is usually possible to run complex spatial operations over the whole dataset.

With greater data volumes comes greater economies of scale and the advantages of relational databases exceeds other factors, not the least of which is the ability to deal well with the enormous quantities of data that can be needed.

Another strategy is to process each change or added feature as it is loaded into the database. This is a good solution for incremental updates and small data loading exercises such as adding a subdivision into the digital cadastral theme, only causing the server load to peak for short periods.

This can be detrimental for large data loads as the production server can be kept very busy for an extended period thus degrading performance for other business requirements.

A 'two tier' architecture strategy with the topology processing functionality in the client means the load is taken off the server maintaining server performance.

Perhaps the best overall strategy is to the have the client based solutions do the initial topological processing and have the server based solutions ensure vertical topology is maintained. This gives the best performance and highest topological accuracy.

Keeping the business going

Correcting large datasets for vertical topology can take time and attention must be given to the day-to-day servicing of business needs whilst the processing is taking place. Updating the main spatial column with the corrected features should be as quick as possible because the business cannot stop while this takes place.

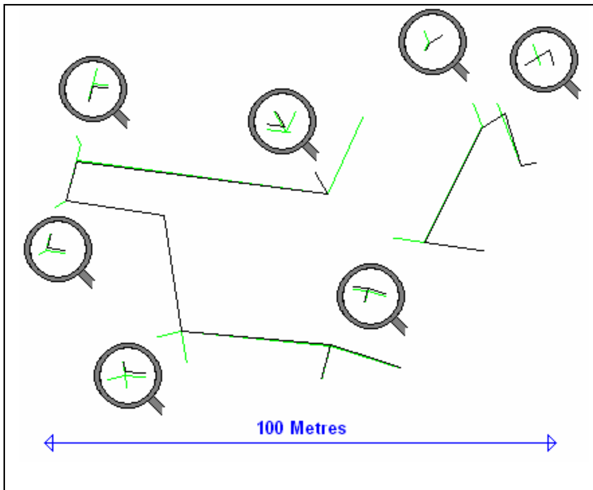
Relational databases that implement a spatial data-type often make it possible to store more than one such spatial column per table, making it possible for a topologically processed version of the original geometry to be stored alongside the original feature.

This has benefits for validation and for posting a completed set of topologically processed features to the main spatial column.

What constitutes a reasonable tool-set for Vertical Topology?

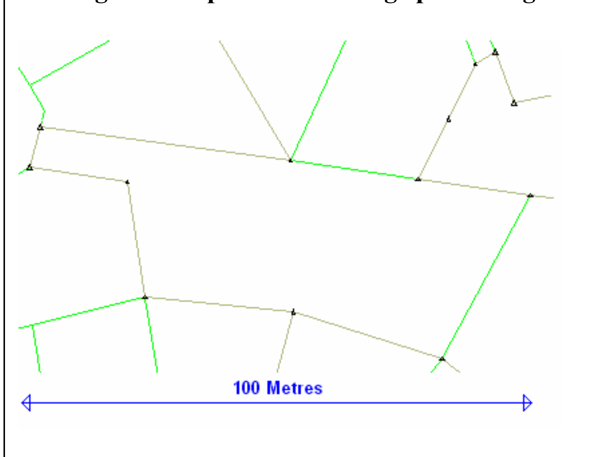
Vertical topology problems involving small shifts can be fixed with some fairly simple algorithms. A quick analysis (e.g. from Diagram **1**/) would likely arrive at:

- Node to ‘Snap to Theme’ Node
- Node to ‘Snap to Theme’ Edge
- Edge to ‘Snap to Theme’ Node
- Edge to ‘Snap to Theme’ Edge



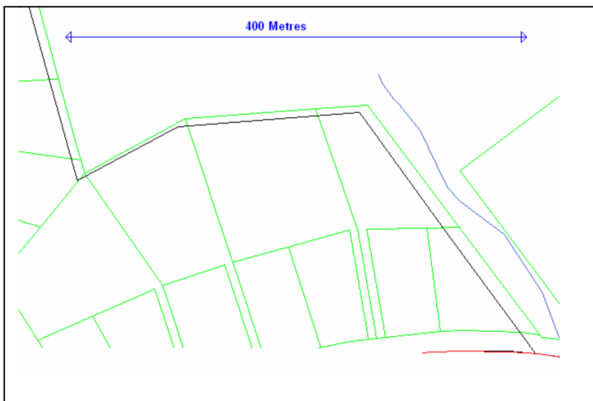
1/ Small shifts relative to line lengths

Edge to ‘Snap to Theme’ Edge processing.

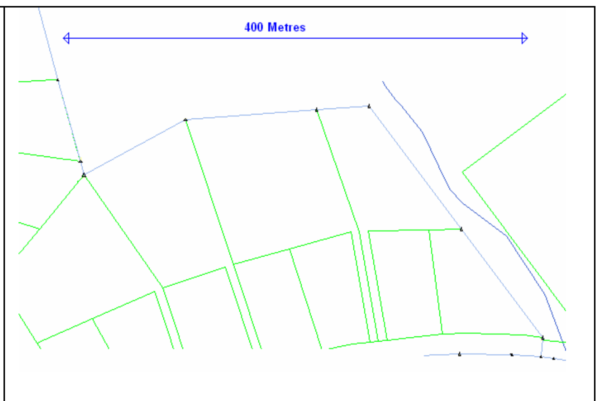


2/ Resulting topology from one algorithm

These routines do an excellent job where the line lengths are somewhat larger than the shifts involved. This extends to larger shifts if the lengths increase proportionally.



3/ Side boundaries provide additional nodes



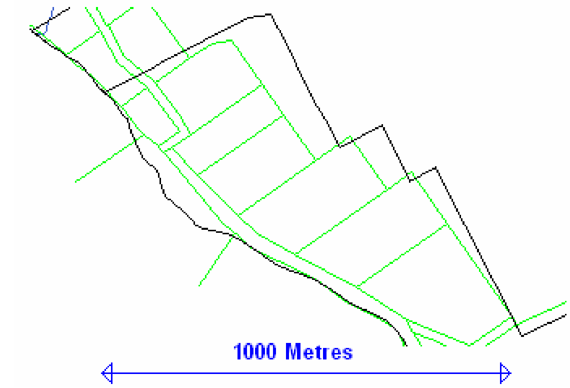
4/ Edge to ‘Snap to Theme’ Node is now needed

Diagram **3/** shows a slightly more complex scenario requiring snapping of edges to the side boundary nodes if the result in Diagram **4/** is to be achieved.

Fortunately, as these scenarios get increasingly complex they become fewer, with perhaps 97% of the effort being spent on just 3% of the data.

However, where the shifts are large and the lengths short the task of snapping these correct segments takes on a whole other challenge.

Diagram **5/** shows a statistical boundary in a non-urban area which plausibly mimics the digital cadastre, but is fortunately in the last 1% of difficulty.



5/ Large shifts relative to line lengths

When there are multiple primary themes involved in the definition of a composite theme the following items require careful consideration to stop the composite theme's features snapping to the wrong primary theme.

- The original scale of the data being digitized
- The current scale to which it is being compared
- The variance in the number of nodes between the various themes
- The variance in the link length
- The cleanness of the older dataset
- The number of themes to which the old theme must now fit
- The degree of detail between the old and new themes
- The degree of evolution of the datasets involved such as that caused by new subdivisions

Repeatability

At the top end of the GIS spectrum, complex technology and economies of scale justify the use of consultants to build systems that automate the maintenance of vertical topology.

Smaller sites require a simpler, more interactive solution.

In either case, when there are multiple primary themes there needs to be a way of ordering the topological processing functions and passing in parameters such as snap distances. Since such a sequence will want to be repeated, there must be at least a way of repeating, recording, editing and running such a sequence.

Integration with Relative Positioning

Many utilities and local governments place their field assets relative to the digital cadastre e.g. 2.2m from the front boundaries. Their position becomes invalid when the digital cadastre is moved.

The shifts needed to re-create vertical topology are the same shifts needed to maintain relative positioning to the digital cadastre, so it becomes an advantage for the topological processing tool to be integrated with the relative positioning tool.

Error analysis

When an automated topological processing or a relative positioning operation has been run there needs to be a mechanism to detect any topological matching errors.

Ideally these tests should be outside the correction tools and should include the overall shift of a feature, its change in length or area, changes in angles and other geometrical properties can be compared in the 'before' and 'after' features.

Where to now?

Given that so much of the spatial data held in so many formats around the world still does not approach GPS accuracy, the maintenance of Vertical Topology and Relative Position to a changing set of primary themes will undoubtedly be a topic for decades to come.

The complexity of the programming and exhaustive testing mean that automated solutions have long lead times to market. With few proven products available, it is critical for an organization to benchmark any solution thoroughly against its own datasets, with clearly defined accuracy and performance criteria.

A solution that fulfils these will solve one of the biggest and most expensive headaches to still plague GIS managers and spatial data custodians.

.Dr Roger Merritt is the Managing Director of Spatial Tapestry and has worked in the Spatial Industry for more than 25 years. Contact: Roger_Merritt@SpatialTapestry.com