Implementing the GMNS network specification for MATSim

Billy Charlton, Technische Universität Berlin <charlton@vsp.tu-berlin.de>

The General Modeling Network Specification (GMNS) [1] defines a new data format for routable roadway network files. This research describes findings while adapting MATSim networks to this specification.

The GMNS spec is thoughtful and mostly complete. A preliminary MATSim-to-GMNS converter was written along with a web-based GMNS viewer. Issues described below are the result of applying this theoretical research to an early, practical use case. The abstract ends with a call for finding a modern container format to tame the disadvantages of CSV tables.

CSV files

A GMNS network is a collection of comma-separated value (CSV) files. CSV is ubiquitous but has well known limitations, which GMNS tries to temper with schemas and a careful specification.

At least three files are needed to define GMNS: config.csv, node.csv, and link.csv, and often many more. The filenames are explicit, which precludes having multiple networks in one folder. The spec states that (unspecified) containers may be used; .zip, SQLite, etc. This leaves a bewildering number of options for how to actually implement a system based on GMNS.

All ID fields in GMNS are defined in the spec as type "any". Excel and most CSV parsing libraries convert numeric strings to numbers, which breaks IDs with leading zeroes. This is not hypothetical: upon creating our first network, links with zero-prefixed IDs failed to parse. IDs should be string.

The largest example network in the GMNS repository has 3000 links. For comparison, a MATSim network used for project studies in our group has 1.2 million links. Parsing a CSV of this size requires a huge amount of memory, and the CSV format precludes skipping unneeded columns. This crashes our web-based viewer! Workarounds such as streaming the file row-wise and discarding unneeded data exist, but then how to display dropped data? The entire file must be reloaded.

Geometry

GMNS uses the "Well-known text" (WKT) format for geometry.. WKT is verbose and does not compress well. WKB (almost identical, but binary) compresses far better. The geometry table could also reference link IDs instead of a duplicated geometry ID: in all examples they are identical.

Bounding boxes per element are not defined; these would allow filtering by zoom/location. This

slows down network display.

Complex networks

MATSim employs a table of time-dependent network changes to handle situations such as reversible lanes. GMNS has "time_set_definitions" that might be useful in the MATSim context. Similar more advanced issues are currently being investigated.

Advocating for a container

The success of the "Open Matrix" (OMX) standard was predicated on developing something that was complete, technically sound, and practical for daily use. For GMNS to achieve similar success, practical aspects of real world usage must be considered. Stopping at defining what goes into GMNS tables -- and leaving individuals to decide how to manage those tables -- is insufficient. OMX never would have been successful if it had left the matrices in CSV format. CSV is too slow, too big, and results in very many individual files.

This isn't failure: it is partial success. The next step is to select a modern container format for GMNS. There are many options and we will report on some possibilities.

This work is 100% open source:

Source code: github.com/simwrapper/gmns
NPM package: npmjs.com/package/@simwrapper/gmns
GMNS viewer: simwrapper.app/map

Example GMNS network for Berlin, Germany. Image shows a portion of a large MATSim network converted to GMNS using the scripts developed in this research; and displayed using the GMNS viewer that was added to the SimWrapper web-based tool. Network includes nodes, links, and geometry (curves)