Anonymising Water Networks for Open Access Publication

Morphing Networks

Richard Collins

Introduction

As part of the managing background leakage project there is a requirement to publish all the collected data and scripts. For this to be useful for future researchers this needs to be accompanied by network hydraulic data from the associated real systems. There are concerns with publishing network data as it may contain individuals data and that there may be issues if if is identifiable. As a result any published network data needs to be anonymised and not identifiable to a specific real world location, but retain the hydraulic properties to make it useful. Here I am proposing that the EPANET .inp file format will be the most ideal format to use for this process.

EPANET .inp File Format

The original EPANET was created in the 1990's as one of the first widely available network hydraulic solvers. Since then it has been updated relatively frequently with the latest version being released in 2020 (EPANET 2.2). EPANET was created by the US EPA and since 2014 been maintained by the Open Water Analytics group (https://github.com/Open-Web-Analytics), EPANET has always been freely available and for most of its history has been open source. EPANET is widely used around the world by water utilities as their primary network hydraulic tool, in addition most other commercial network hydraulic software either directly uses the EPANET hydraulic engine or have produced reworked versions of the code. All commercial network software is able to open or import EPANET .inp files, as a result the .inp file has become the de-facto method of transferring data between different software solvers.

The .inp file format is a human readable text file that contains information about the network nodes, links and hydraulic objects (pumps, valves, controls, demand patterns etc.) and the configuration of the hydraulic (and water quality) solver.

The EPANET format has a couple of key features that will allow good anonymisation of the network data. The primary one is that the network hydraulics are computed using the link (pipe, valve, pump etc.) properties which are stored directly in the link elements, for example pipes objects have a recorded specific length which is used for calculation of headlosses. Nodes have recorded positions, however this is only used for visualisation of the network. A pipe lengths is independent the distance between the nodes to which it is attached. As a result the network hydraulics can be preserved by

maintaining the link properties, but we can move the node positions arbitrarily to “disguise” the network and ensure that it cannot be mapped back to the “real world location”.

In addition the .inp file allows for the storing of additional link and node information in a way that will ensure that it can still be opened by original software, but can the additional information can be accessed by any scripts that are written as part of the project.

It should be noted that exporting networks from modern commercial network software to EPANET format does sometimes loose some information (usually about complex control options not available in EPANET solver), and we will need to verify that the EPANET export representation of the network is appropriate for the 25 DMAs used in this project.

Example

Original Network

As an example I will demonstrate the network topology anonymisation on a real world network. The network was provided to the University of Sheffield from Yorkshire Water as part of a previous research project.

Network Layout

In Figure 1 we can see the original layout of the network, with the pipes closely following the real world road layout, even without the nodal coordinates directly relating to the real world location (Figure 2) it would be possible with suitable effort to identify this network.

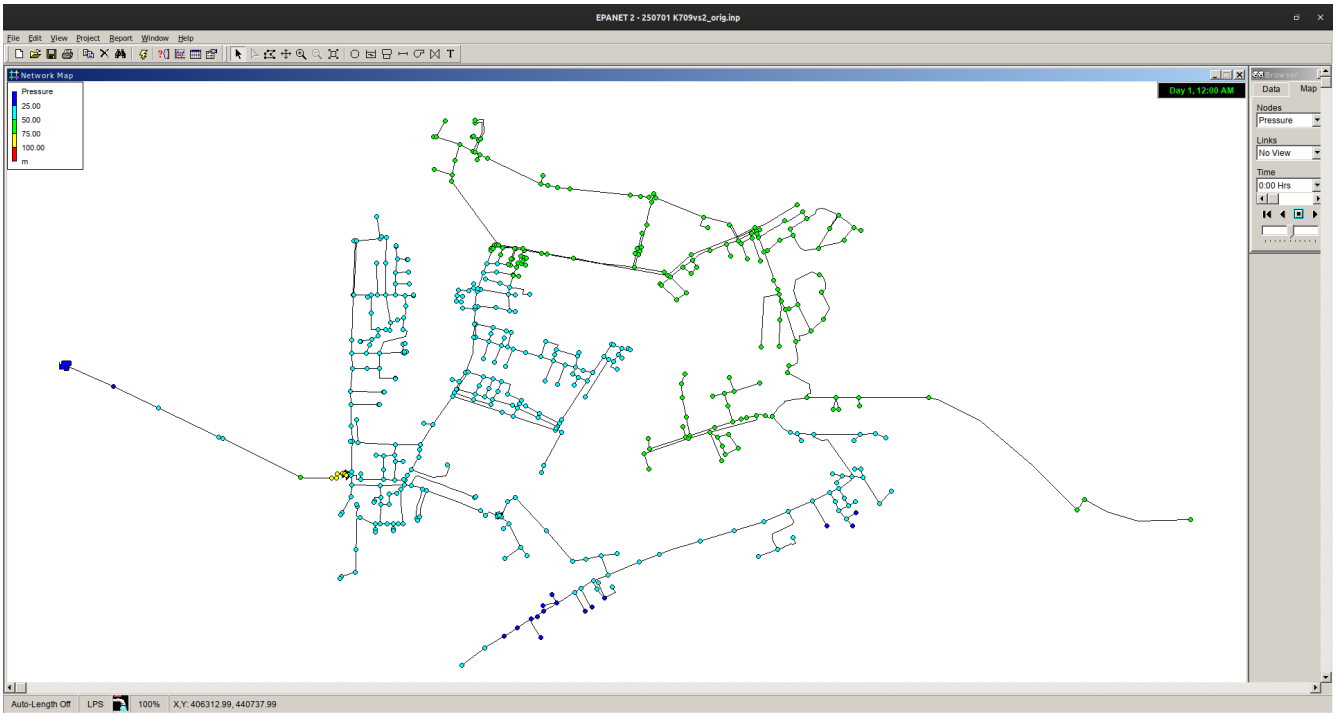

Figure 1: Original Network Layout

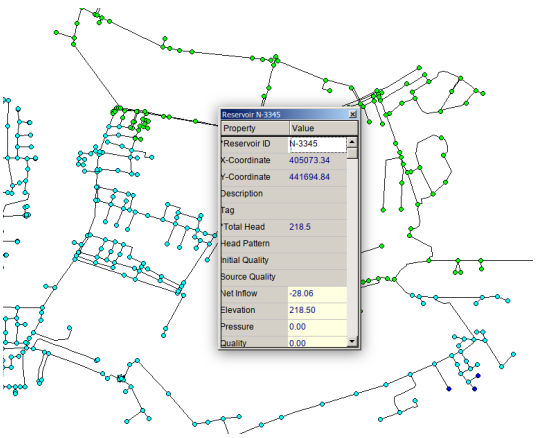

Figure 2: Coordinates of source reservoir in original network

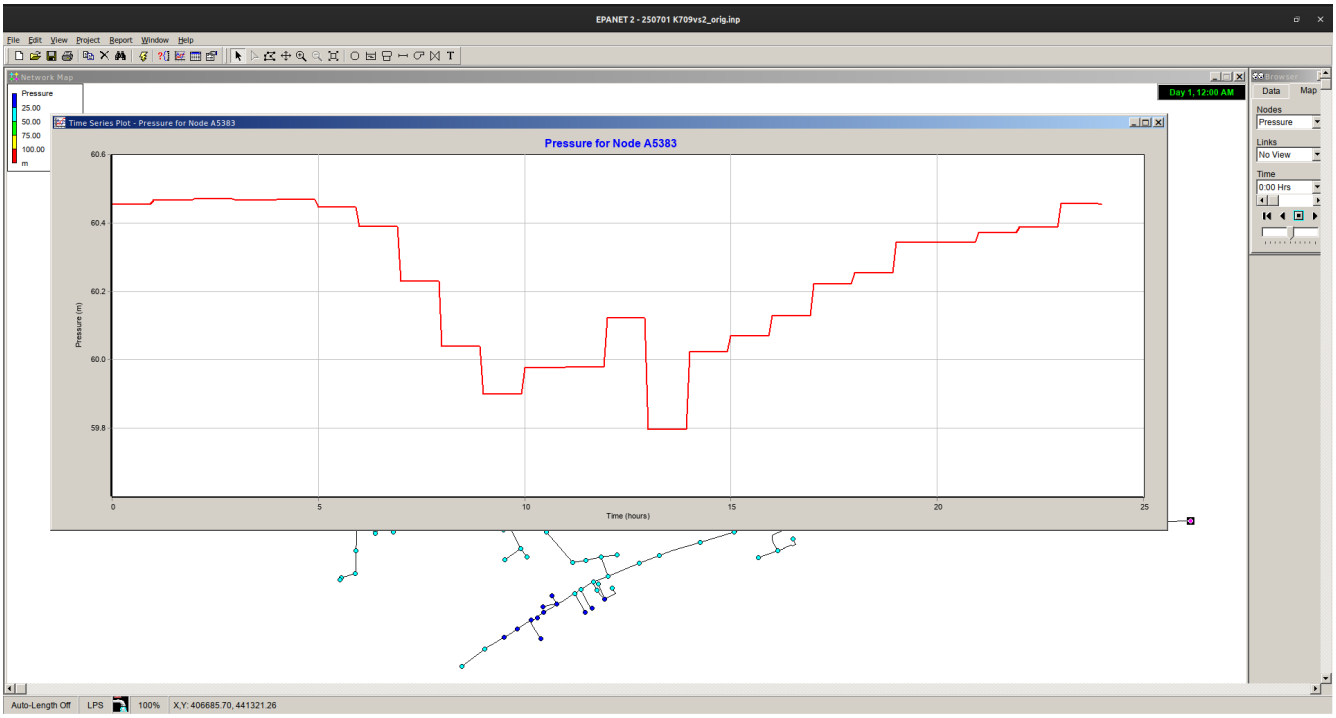

Figure 3: Pressure profile from Node A5383 in original network

Network Results

To ensure that hydraulic solution is unaffected by the network anonymisation in Figure 3 we can see the pressure profile of the node on the far right of the system A5383

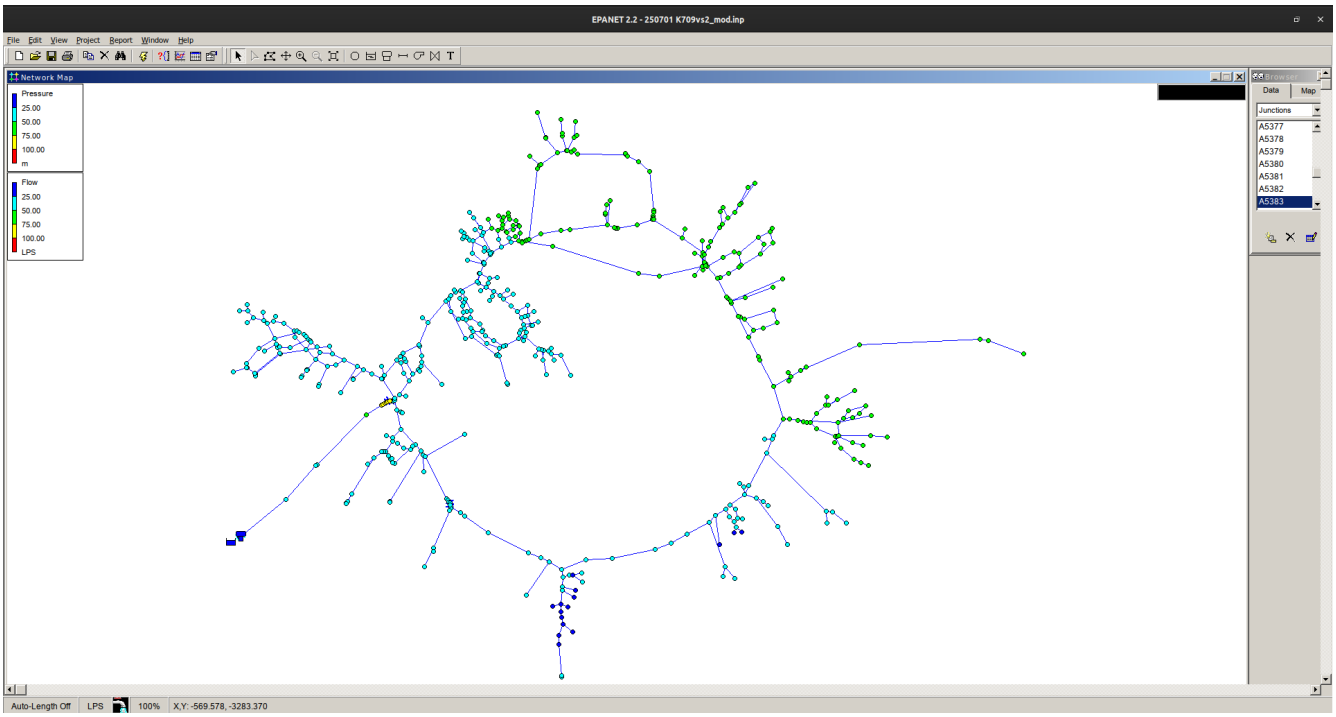

Figure 4: Anonymised Network Layout

Anonymised Network

A graph “morphing” algorithm was applied to the nodal locations to modify them in a way that, as close as possible, maintains the visualisation of the pipe lengths and any crossing points, but puts the nodes in new positions that no longer have any relation to their real world locations, Figure 4. In addition the network coordinates are now centred around an arbitrary point (here chosen to be 0,0), Figure 5.

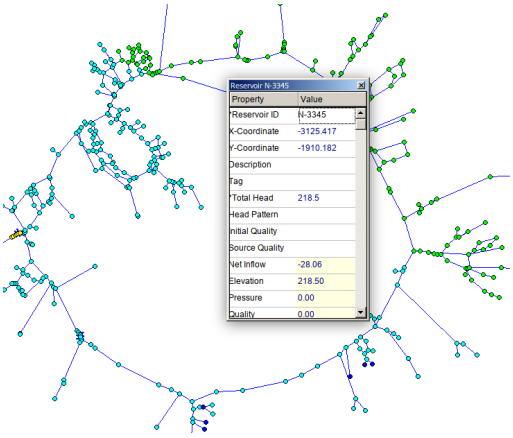

Figure 5: Morphed nodal position of source reservoir

Network Results

To confirm that the network hydraulics are unaffected by the anonymisation compare figure 6 with Figure 3 for the node at the right hand side of the network A5383.

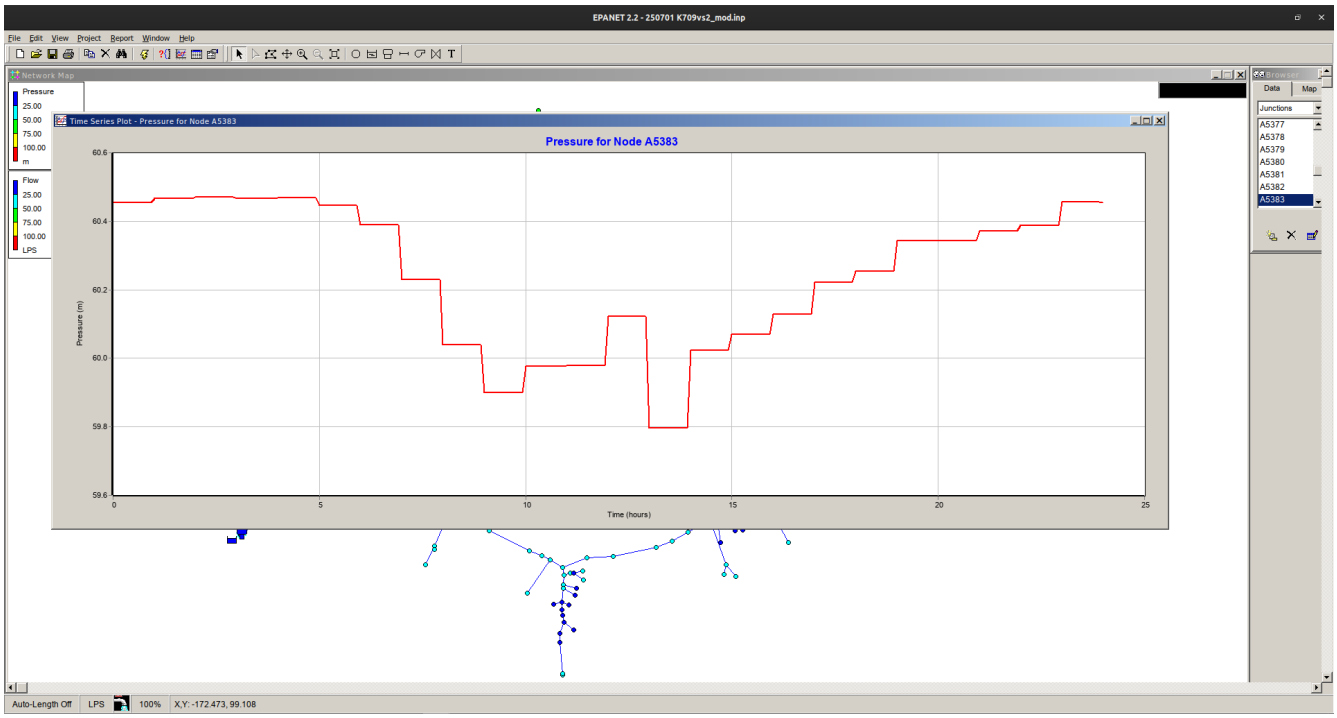

Figure 6: Anonymised network pressure profile for Node A5383

Mapping

In addition to the output anonymised network file this process also produces a “mapping” file that relates the original to anonymised locations. This file will be kept internal for the project but will allow us to add other data (customer meter data, pressure logging locations etc.) in a way that will ensure that it can be related to the correct points of the anonymised network.

Conclusions

This short document has demonstrated that it is possible to suitably anonymise a network topology to ensure that any open access published data is not able to be related back to the real world location. It should be noted that the above has not converted node or link names to allow for easy comparison between original and anonymised networks, but this process is easy undertake.

This document also proposes that the EPANET .inp file format should be used for open access publication as it is an open access format that is widely used by the water industry and is able to be extend depending on the requirements of this project.