ACM SIGSPATIAL Cup 2018 - Identifying Upstream Features in Large Spatial Networks

Dev Oliver[1], Bo Xu[2], Yuanyuan Pao[3]
[1]Environmental Systems Research Institute (Esri)
[2]HERE
[3]Lyft
doliver@esri.com, bo.5.xu@here.com, ypao@lyft.com

Abstract

ACM SIGSPATIAL Cup 2018 was the 7th GIS-focused algorithm contest hosted by the 26th ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems (ACM SIGSPATIAL 2018). The contest explored the problem of analyzing large spatial networks (e.g., utility networks) in order to find upstream *features from a given set of starting points (a feature is considered to be upstream if it is on a simple path between a starting point and a controller).*

1 Introduction

For over 26 years the ACM SIGSPATIAL conference has addressed a variety of topics in the field of geographic information systems. In addition to presenting scientific discoveries in the form of papers, short papers, posters, and demos, the conference began acknowledging the art of algorithm design and implementation from a practical perspective through the ACM SIGSPATIAL Cup. Each year, a challenging yet well-researched computational topic is chosen and participants are required to solve the given problem with high accuracy, quality, and performance.

This year's contest focused on analyzing large spatial networks in order to find upstream features from a given set of starting points. The identification of upstream features is a key problem in critical infrastructure analysis for locating assets that are crucial to a nation's economy, security, and health. An example is identifying transformers that supply electricity to a water pump that is needed by the sole hospital in a region. Given a spatial network, a set of starting points (which may be junctions or edges in the spatial network), and a set of controllers (which are also junctions in the spatial network), the goal is to find all features originating at starting points and terminating at controllers where each "upstream" feature returned is on a simple path between a starting point and a controller.

The type of spatial networks that the contest focused on were utility networks [5, 3, 1, 6], which are used to model utility systems such as electric, gas, water, and telecommunications. Utility networks define and manage a collection of database tables, metadata, business rules, and a network graph, used to store and model the connectivity information. They are used to represent the relations of the real-world objects that utility companies use to deliver resources to their customers. A utility network is subdivided into a set of zones or circuits. Each zone or circuit has a special set of junctions called controllers from which resources flow. The contest operated under the assumption that the zones or circuits are source-based (e.g., electric, water, and gas networks), meaning that resources will flow from controllers or sources to the rest of a zone or circuit (note that zones can also be sink-based such as in gravity-fed networks, e.g., sewer networks, where resources flow towards the controllers or sinks). The main tasks in this contest were to build the network topology and to infer flow direction in a source-based network such that features that are upstream to starting points can be identified. The term "starting point" defines the location where analysis begins, i.e., the feature for which we are trying to identify upstream features. In looped topologies, flow direction is assumed to be bidirectional.

2 Sponsors and Supporters

ACM SIGSPATIAL Cup 2018 was sponsored by IBM and Esri, which was a great help for the SIGSPATIAL community. The sponsorship resulted in cash awards of $500 for first place, $300 for second place, and $200 for third place. Additionally, the authors of the best three submissions were invited to write a short paper about the key ideas of their approaches and to give an oral presentation of their work in a conference session dedicated to the cup.

3 Problem Description

The problem that was explored in this contest is defined as follows:

Given:

- A collection of point features representing devices and junctions
- A collection of line features
- A set of controllers, which are a subset of the device features
- A set of starting points, S, where S is a subset of the device, junction, and line features

Find:

- All upstream features (a feature is considered to be upstream if it is on a simple path between a starting point and a controller).

Objective:

- Computational efficiency

3.1 Example

Figure 1 shows an example input and output of the problem. The input consists of a collection of point features, represented by junctions $J1$, $J2$, $J3$, $J4$, $J5$, $J6$, $J7$, $J8$, $J9$, $J10$, $J11$, $J12$, and $J13$, a collection of line features represented by edges $E1$, $E2$, $E3$, $E4$, $E5$, $E6$, $E7$, $E8$, $E9$, $E10$, $E11$, $E12$, $E13$, and $E14$, two controllers $J1$ and $J5$, and one starting point $J9$. The output is all upstream features which are highlighted (i.e., junctions $J1$, $J2$, $J3$, $J5$, $J6$, $J7$, $J8$, and $J9$ and edges $E1$, $E2$, $E4$, $E5$, $E6$, $E7$, $E8$, $E9$). The edges and junctions that were not included in the results were not on a simple path between a starting point and a controller.

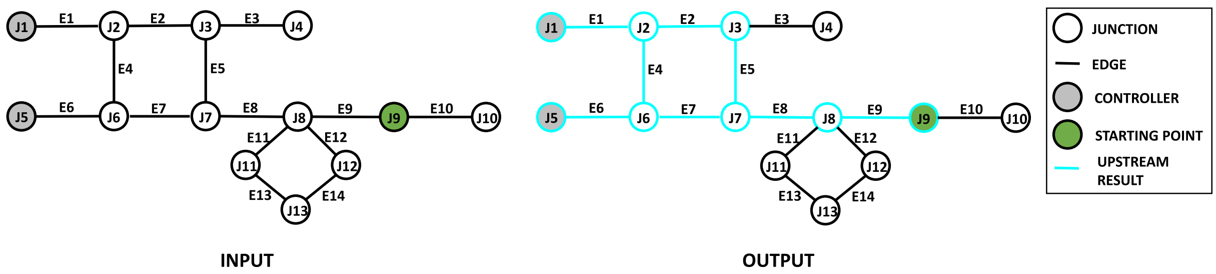

Figure 1: Example Input and Output of Identifying Upstream Features

3.2 Data

Two datasets were provided for this challenge. The first was a sample dataset comprised of 30 features, which was useful for participants to verify the correctness of their approaches. The second was Esri's Naperville Electric Network dataset (comprised of approximately 15,000 features). Figure 2 shows an example of identifying upstream features in Esri's Naperville electric dataset. The starting point is represented by the green circle towards the north western portion of the image and the controller is located towards the south western portion of the image. The features that are upstream are highlighted in blue.

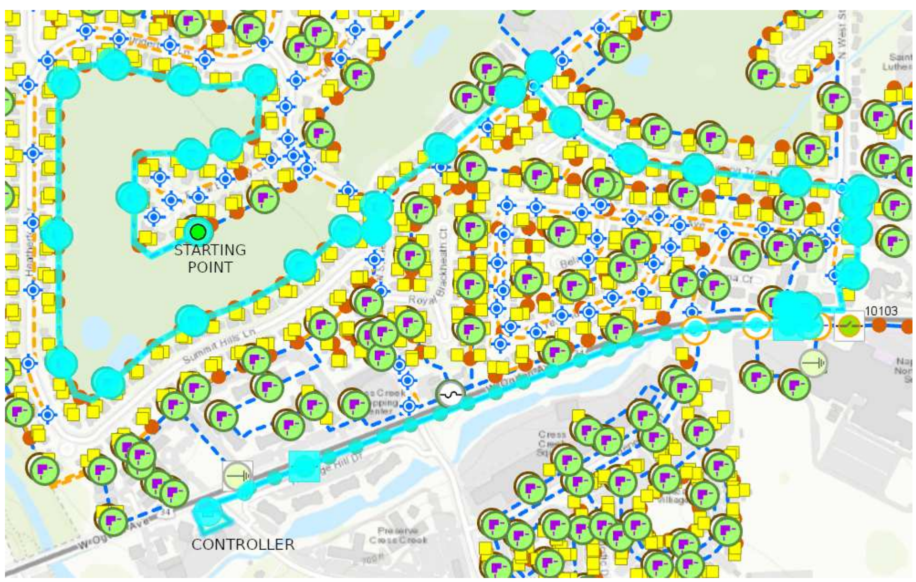

Figure 2: Example of identifying upstream features in Esri's Naperville Electric Dataset

Each dataset contained a json file for constructing the network topology, a set of shapefiles for visualization, a sample starting points file, and a sample output file. Submissions were required to consume the json file to construct the network topology and extract the controller information.

4 Submissions and Winners

The challenge received 15 very good submissions and the teams totaled 42 members from all over the world. Three entries were selected as winners. The best three submissions were written in C++. They all leveraged a block-cut tree for identifying upstream features from a set of starting points in the network.

The winning submission was by Salles Viana Gomes Magalhães, W. Randolph Franklin, and Ricardo dos Santos Ferreira [4]. Their submission computed articulation points and biconnected components using Tarjan's algorithm [7]. However, instead of reusing existing implementations of Tarjan's algorithm (such as the one provided by the Boost Graph Library), they implemented it from the ground up for performance purposes. Their implementation that was parallelized using OpenMP also used two 64-bit integers to represent the 32-digit (where each digit is in hexadecimal) global identifier for features (to reduce memory footprint and accelerate operations like comparing the ids of two features) and a custom parser to read the JSON files representing the network. The second place submission was by Thomas C. van Dijk, Tobias Greiner, Bas den Heijer, Nadja Henning, Felix Klesen, and Andre Löffler [8]. Their approach used a linear-time algorithm based on an annotated block-cut tree. The third place submission was by Zach Goldthorpe, Jason Cannon, Jesse Farebrother, Zachary Friggstad, and Mario A. Nascimento [2]. They implemented a linear time algorithm based primarily on a two-sweep depth-first search which decomposes a graph into its biconnected components prior to collecting the upstream features.

Acknowledgments

We wish to acknowledge the support from our sponsors IBM and Esri. Furthermore, we want to express our thanks to the organizing committee of the ACM SIGSPATIAL Conference on Advances in Geographic Information Systems for hosting this competition.

References

[1] P. Bakalov, E. G. Hoel, and S. Kim. A network model for the utility domain. In *Proceedings of the 25th ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems*, page 32. ACM, 2017.

[2] Z. Goldthorpe, J. Cannon, J. Farebrother, Z. Friggstad, and M. A. Nascimento. Using biconnected components for efficient identification of upstream features in large spatial networks. In *Proceedings of the 26th SIGSPATIAL International Conference on Advances in Geographic Information Systems*. ACM, 2018.

[3] E. Hoel, P. Bakalov, S. Kim, and T. Brown. Moving beyond transportation: utility network management. In *Proceedings of the 23rd SIGSPATIAL International Conference on Advances in Geographic Information Systems*, page 8. ACM, 2015.

[4] S. V. G. Magalhães, R. W. Franklin, and R. d. S. Ferreira. Fast analysis of upstream features on spatial networks. In *Proceedings of the 26th SIGSPATIAL International Conference on Advances in Geographic Information Systems*. ACM, 2018.

[5] B. Meehan. *Modeling electric distribution with GIS*. Esri Press Redlands, 2013.

[6] D. Oliver and E. G. Hoel. A trace framework for analyzing utility networks: A summary of results. In *Proceedings of the 26th SIGSPATIAL International Conference on Advances in Geographic Information Systems*. ACM, 2018.

[7] R. Tarjan. Depth-first search and linear graph algorithms. *SIAM journal on computing*, 1(2):146–160, 1972.

[8] T. C. van Dijk, T. Greiner, B. den Heijer, N. Henning, F. Klesen, and A. Löffler. Wüpstream: Efficient enumeration of upstream features. In *Proceedings of the 26th SIGSPATIAL International Conference on Advances in Geographic Information Systems*. ACM, 2018.