Estimation of Accumulated Upstream Drainage Values in Braided Streams Using Augmented Directed Graphs

National Geospatial Technical Operations Center
1400 Independence Road, Rolla Missouri, 65401

Lawrence V. Stanislawski
Science Applications International Corporation
Phone: (573) 308-3914
Fax: (573) 308-3652
E-mail: lstan@usgs.gov

Michael P. Finn
United States Geological Survey
Phone: (573) 308-3931
Fax: (573) 308-3652
E-mail: mfinn@usgs.gov

Michael J. Starbuck
United States Geological Survey
Phone: (573) 308-3687
Fax: (573) 308-3794
E-mail: mstarbuck@usgs.gov

E. Lynn Usery
United States Geological Survey
Phone: (573) 308-38
Fax: (573) 308-3652
E-mail: usery@uga.edu

Patrick Turley
United States Geological Survey
Phone: (573) 308-3890
Fax: (573) 308-3652
E-mail: pturley@usgs.gov

Abstract

Divergent-convergent tributaries that regularly occur in braided stream areas complicate the problem of traversing a flow-oriented surface drainage network and accumulating upstream values that describe geometric properties of the network and associated hydrologic catchments. Common traversal approaches associated with directed-graph data structures may implement a division rule at divergent nodes, but this approach can generate inappropriate estimates of accumulated upstream values when the objective is to have monotonically increasing values with downstream location on the network. This paper describes a traversal method that circumvents this problem by using an augmented directed graph that associates values to closed areas on the graph. The method is demonstrated on drainage network features for several watersheds of the US National Hydrography Dataset (NHD), which the US Geological Survey (USGS) and US Environmental Protection Agency have been developing over the last several years. The USGS is supporting the needs of the US for topographic mapping in the 21st century through a program known as *The National Map* and is developing and maintaining eight geospatial data layers: transportation, hydrography, boundaries, structures, elevation, land cover, orthographic images, and geographic names. The NHD is the hydrography layer of *The National Map*. We are applying this network traversal method in our research on automated generalization of the NHD.

Keywords: Drainage network, braided streams, upstream drainage area, directed graph, geospatial data structure

1. Introduction

The National Hydrography Dataset (NHD) is a vector data layer of *The National Map* representing the surface waters of the United States (USGS, 2000). It is stored in an ArcGIS geographic database (geodatabase) model at three levels of detail: medium (1:100,000-scale source), high (1:24,000-scale source), and local (1:12,000 or larger source) resolutions. Currently, only the medium resolution layer is complete for the conterminous U.S, with the high resolution layer nearing completion. Although the available resolutions are suitable for many applications, various uses of the NHD may require different levels of detail than what is available in the national database. While only additional data gathering can furnish more detailed data, generalization of existing data can furnish less detailed NHD data that is suitable for large-area regional or national studies. Furthermore, a robust generalization process could make it feasible to store and maintain only the highest resolution, most accurate layer of the NHD, and eliminate the need for storing and maintaining lower resolution, less accurate layers that were derived from smaller scale source maps.

A goal of this research is to develop a generalization strategy that will produce a subset of NHD features that maintains the NHD model format. Nearly all generalization processes include an initial step of selecting objects and attributes from the source database that are to be represented in the generalized dataset (McMaster and Shea, 1992). We refer to this object selection process as feature pruning.

Our feature pruning strategy extracts the most prominent network features based upon the relative amount of the watershed surface that runs off into the network features. Thus, a unique upstream drainage area must be assigned to each network feature in the watershed of interest.

This paper describes an approach we are developing to pre-process NHD network features and enable automated pruning in subsequent generalization procedures. The pre-processing approach should provide estimates of upstream drainage area for each feature in a network dataset. Estimates generated by the approach should be monotonically increasing with downstream location in the network, and should not incorporate duplicate values from diverging features into downstream features where the divergent tributaries converge.

Earlier attempts to generate upstream drainage area estimates included both problems, which are related to handling of divergent sections of a network. If upstream values are proportioned into downstream tributaries at divergences, then subsequent pruning may produce gaps where divergences exist. If entire upstream values are passed downstream to each divergent tributaries, then duplication of values at divergences are included downstream where divergent tributaries converge. This problem generates inflated upstream drainage area estimates, and subsequent pruning may improperly maintain network tributaries due to upstream braiding.

2. Methods

2.1 Augmented Directed Graph

The approach described in this paper implements a directed graph data structure. We refer to a graph as a set of points or "nodes" that are connected by a set of lines or "edges" (McCracken and Salmon 1987). Edge connections represent topological relationships between features in the graph. In a directed graph (Figure 1), edges are oriented in the direction of a process, such as the flow direction of a stream.

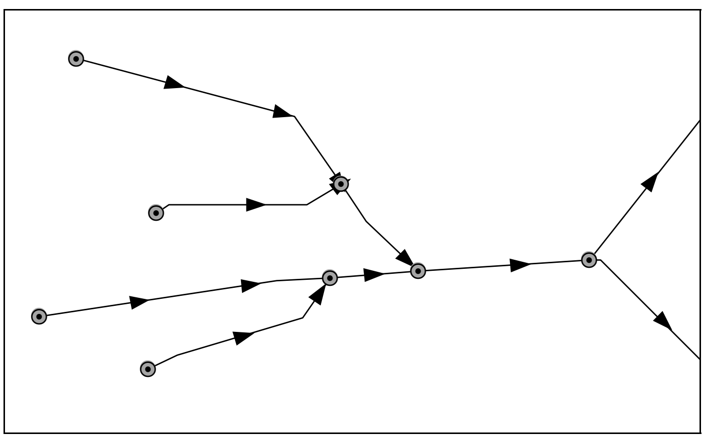

Figure 1. Section of a directed graph.

We are using the directed graph to represent the network of surface water features that flow over a watershed. The nodes of the graph represent confluences of surface water features and the edges portray features such as canals, ditches, streams, or pipes, which are oriented in the primary direction of water flow; that is, edges are directed downstream. Each edge of the surface drainage network is assigned a catchment area (km^2). The catchment associated with an edge is that portion of the watershed where surface runoff flows into the associated network edge feature.

Accumulated upstream drainage area for any edge is the sum of all upstream catchment area values including the catchment area associated with the edge of interest. The approach we are using to compute accumulated upstream drainage area for each edge in a network is to select the most upstream edges, or headwater tributaries, in the watershed and push the catchment areas into a sum field at each associated downstream node. Simultaneously, an in-counter field at each node is incremented each time a value is pushed into the sum for a node. Given that the number of inflows (inflowing edges) and outflows (out-flowing edges) are known for each node in the graph, the next step is to select all nodes that have the in-counter value equal to the number of incoming edges and push the sum at each of these nodes to the sum field on each out-flowing edge, where the edge sum field includes the catchment area associated with the edge. This process is illustrated in Figure 2.

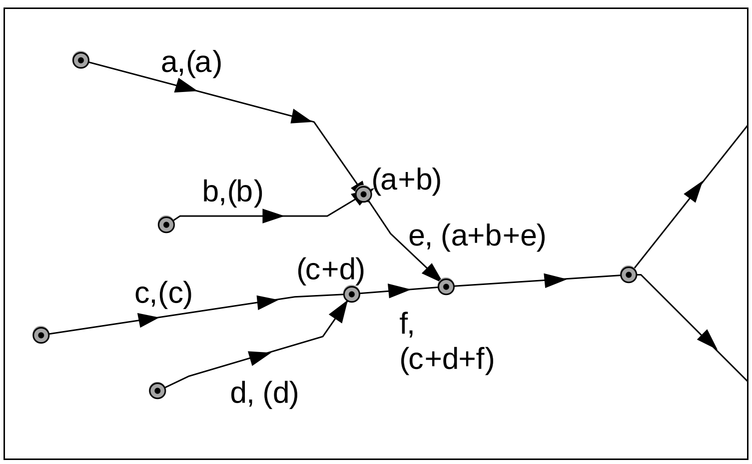

Figure 2. Directed graph with accumulated upstream values. Sums are shown in parentheses. Two values are associated with each edge: the value for the edge, and the sum. Only a sum is shown at nodes.

The process iteratively continues until all catchment areas are pushed into downstream sums. This accumulation process would work properly if no divergent sections of the graph converged at some node further downstream. However, when divergent sections of the graph converge, the sum existing at the upstream divergent node is duplicated in the sum at the downstream convergent node, resulting in improperly inflated sums. Inevitably, divergent-convergent sections, commonly referred to as braided streams, occur fairly often in drainage networks. In addition, swampy or coastal areas include complex divergent-convergent network systems. Consequently, a method to handle these cases must be included in the upstream accumulation algorithm.

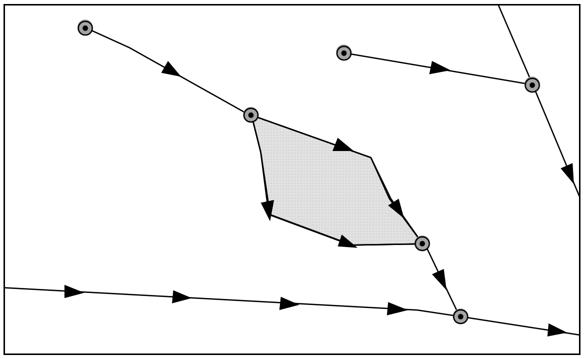

Figure 3. Divergent-convergent polygon (shaded area) with one incoming divergent node and one outgoing convergent node (one-in-one-out polygon).

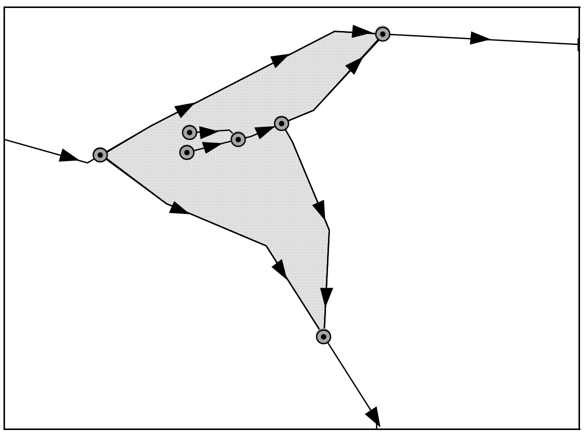

Figure 4. Divergent-convergent polygon (shaded area) with one incoming divergent node and two outgoing convergent nodes (many-out polygon).

To circumvent the problem, the directed graph is augmented to include the polygons created by divergent-convergent edges. These polygons are classified by the number of incoming divergent nodes and the number of outgoing convergent nodes.   Two different methods are applied at divergent nodes to handle divergent-convergent polygons. One method is applied to polygons with one incoming divergent node and one outgoing convergent node; referred to as the one-in-one-out polygon case (Figure 3). And the other method is applied to polygons with many outgoing convergent nodes, which may have one or more incoming divergent nodes. This second case is referred to as the many-out polygon case (Figure 4).

Handling of divergent nodes on many-out polygons consists of starting a list containing the divergent node number and its associated sum. The list is transferred to each of the edges flowing out of the divergent node. As the graph is being processed, these lists are passed to downstream edges and nodes and subsequently handled at convergent nodes. If a divergent node is passed a list from one or more upstream edges, then the new list built for the divergent node includes the upstream list appended with the new divergent node number and its associated sum. However, in this case, the sum associated with the new divergent node in the list is reduced by all previous sums in the list. The new list is passed to each outgoing edge.

Handling of divergent nodes on one-in-one-out polygons consists of passing the sum at the node to the associated polygon. However, if the divergent node has been passed a list of upstream divergent node sums, then the value passed to the polygon is the sum at the node reduced by each of the sums in the list. The list is not modified, but it is passed to downstream edges.

Convergences are handled the same way in all cases. The sum pushed to a convergent node may be reduced in up to two ways. First, it may be reduced by the sum pushed to the associated one-in-one-out polygon. Secondly, it may be reduced by the sums in the converging lists in the following manner: the sum associated with a listed node is multiplied by one less than the number of times a node occurs in all incoming lists, and the resulting product is subtracted from the convergent node sum. Thus, a node (referenced by its node number) must exist in at least two converging edge lists in order for it to affect the convergent node sum. The new list associated with a convergent node includes node numbers and associated sums from all incoming lists, but each is listed only once. The new list is passed to each outgoing edge.

2.2 Comparison to Brute-Force Tracing

To verify that the augmented directed graph (ADG) algorithm was working properly, it was compared to a brute-force tracing (BRT) algorithm that loops through all nodes in a network dataset. At each node, the BRT program traces all upstream paths in the network and sums the values associated with all traced features. The resultant sum is pushed to each downstream edge, with the value on the downstream edge being added to its sum.

The ADG and BRT algorithms were developed in Arc Macro Language programs and tested on several network datasets. Both programs computed the sum of all upstream catchment area values for each network feature. Values resulting from each program were compared for all features in each test dataset. Some of the test datasets were fabricated to include some special situations rarely found in real world data. Other test datasets consisted of single-subbasin networks extracted from the high resolution layer of the NHD.

3. Results

For all test datasets, the ADG method produced identical upstream drainage area values on each network feature as the BRT method. This outcome suggests that, at convergent nodes, the ADG program is properly removing the duplicate portion of sums that had been pushed to divergent edges in the network. In test datasets having only one out-flowing pour point, the sum of all catchment areas in the network equals the maximum upstream drainage area computed by the ADG program; further suggesting that the programs are functioning properly. In addition, it was verified that each sum was larger than the sum associated with the next upstream feature in the network. In more complex terms, the computed upstream drainage values are monotonically increasing with downstream location in the network for all tested datasets. A visual portrayal of program results are displayed in Figures 5 and 6.

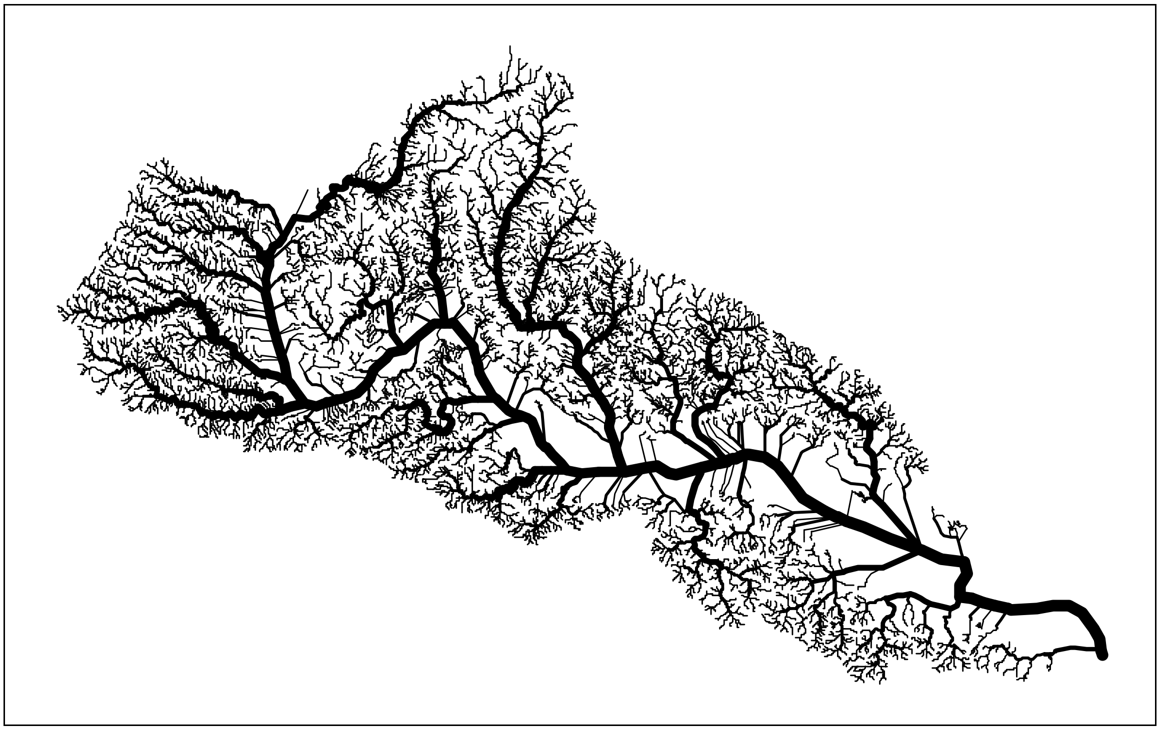

Figure 5. Graduated symbol display of accumulated upstream drainage area resulting from augmented directed graph algorithm for a high-resolution NHD subbasin.

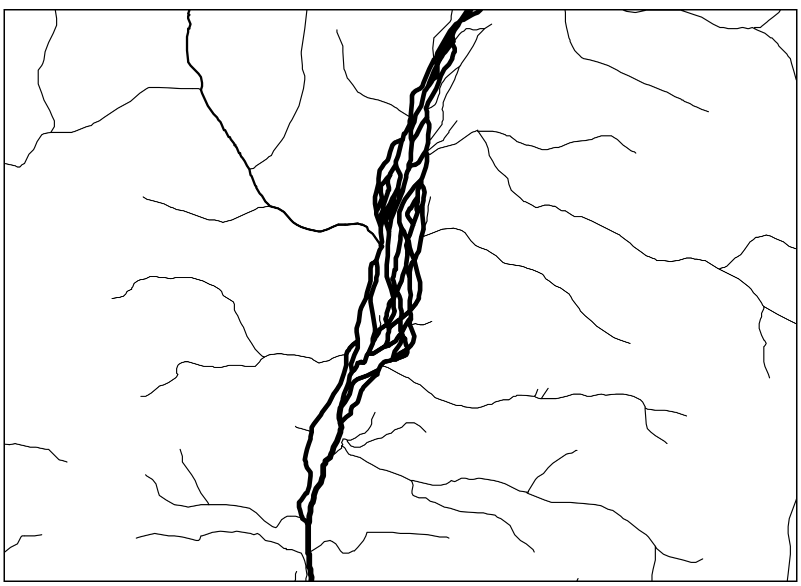

Figure 6. Graduated symbol display of accumulated upstream drainage area resulting from the augmented directed graph algorithm for a braided stream section of the high-resolution NHD.

The ADG program processed each dataset about 50 times faster than the BRT program. The ADG program took about 45 seconds to process each 1000 edges in the test data. After further testing of the ADG program, we plan to write the algorithm in a compiled language, such as Visual Basic or C, which should improve processing speed.

Currently, no method has been determined to reduce the size of lists associated with divergent sections of a network graph. For a 30,000 edge dataset of a complex coastal network, the ADG program generated about 1500 edge lists with the longest list including about 45 entries. Further testing and analysis is required to determine processing limitations of this algorithm and the associated data structure. However, it is anticipated that the ADG processing approach is well-suited for most inland drainage systems because they exhibit dendritic drainage patterns that include few many-out polygon sections.

4. Conclusions

This paper describes a processing algorithm that implements an augmented directed graph data structure. Results demonstrate that the algorithm can accurately accumulate values associated with the edges of the graph by accumulating them with the direction of flow. The algorithm has the following properties:

- iteratively pushes values downstream from multiple starting points,
- traverses each graph feature once,
- tracks duplicate values on divergent edges and removes the duplicates at convergent nodes, and
- provides monotonically increasing values with downstream location on the graph.

From a data processing perspective, the ADG approach appears to be a substantial improvement over existing graph traversal techniques that provide less precise proportioning alternatives for handling duplication at divergences.

The ADG approach is tailored for data that is properly oriented in the directed graph data structure. Inaccurate results or failed processing are likely when edge features are oriented improperly. Further research is required to validate this procedure and identify its limitations. It is expected that enhancements will be required to ensure proper handling of cycles that may exist in a graph.

In the future, we will use this approach to generate values for the subsequent pruning process in our generalization strategy for the NHD layer of *The National Map*. Alternative uses for ADG processing of network geospatial data are likely to be identified.

References

McCracken, D., and W. Salmon. 1987. *A second course in computer science with Modula-2*. New York, New York: John Wiley and Sons.

McMaster, R., and K. Shea. 1992. *Generalization in digital cartography*. Association of American Geographers. Washington, D.C.

U.S. Geological Survey. 2000. *The National Hydrography Dataset: Concepts and Contents*. Accessed May 5, 2006, at URL http://nhd.usgs.gov/chapter1/index.html.