Towards the topologically correct drainage network: integration between TAS GIS and GRASS GIS software

Jarosław Jasiewicz

Adam Mickiewicz University, Department of Paleogeography and Quaterary Geology,
Dzięgielowa 27, 60-680 Poznań, Poland
jarekj@amu.edu.pl

abstract: In the segment of freely available software (Freeware and FLOSS) there are no tools which individually allow to build topologically corrected drainage network and use it to extended spatial analysis.Some of the useful functions are diffused among different tools; two programs distinguish among other: GRASS GIS: the open source fully featured universal GIS suite and TAS GIS: (free software developed by J. Lindsay) specialized in hydro-geomorphological terrain analysis with lot of procedures unavailable in other programs. The main advantage of GRASS is its interoperability with Linux system, PostgreeSQL database and R statistical environment, which allow easily process data in batch mode in any manner. Unfortunately GRASS GIS hydrological modeling and raster-vector conversion tools do not create correct topological structure. On the other hand TAS GIS has very limited batch procedures and no tools for network analysis but has tools to generate topologically corrected vector data structure of drainage network. Additional information, like Starhler stream order, channel type and many others can be acquired in raster format.

It is important to take into account that drainage network is a tree data structure called rooted tree where one node (outlet of the catchment) is designated the root. Segments (channels) of the network have a natural orientation towards the root. Most of the network has internal topology based on topology of the GIS data structure, however the lack of data describing the spatial relationship within the catchment often limits the possibility of advanced spatial analysis. This approach must deal with the specificity of the network considered as a directed tree and allow to model characteristics not only on the network nodes (junctions) but also along the sections between nodes.

The procedure of data interoperability between TAS and GRASS GIS use vector-format drainage network and raster data imported to GRASS GIS from TAS. Procedure uses internal GRASS vector topology and SQL statements. Additional information about relationship: ancestor-successor, Strahler stream order of particular channel, Shreve stream magnitude are derived with sampling raster data and build into SQL relationship. Such network, connected with additional SQL tables can store bundle of geomorphological data both for nodes (junctions) and particular vertex of any channel. Storing data in GRASS GIS topological format allow to transfer geometrical data to R environment where geometrical properties of network components like channel directions can be easily calculated without particular skills in programing.

keywords: drainage networks, geomorphology, OpenSource GIS, PostgreSQL, hierarchical structures,

Preface

The fluvio-denudational geomorphological processes are actively involved in landscape formation processes mostly at the upper reaches of valleys. Therefore the identification, mapping and management of stream channels is important in cartography, hydrology, water resource management, and last but not least in geomorphology applications. Drainage channels are distinct geomorphological features which can be represented both as areal and linear features. Thus, there is no doubt that the drainage network is a typical hierarchical tree data structure. The linear representation of a channel allows to analyze drainage networks as a topological tree structure with a build-in hierarchy. The main problem of this ascertainment is a lack of tools to create and analyze such structures in a popular GIS tools. For example the network analysis tools present in the GRASS GIS software poorly support hierarchical structures.

The lack of data describing the spatial variety within the drainage network often limits its application in spatio-geomorphological analysis. This is particularly true for the variability of local drainage network characteristics, such as length of channel, cross- and long-section geometry or drainage direction. Remote sensing technology usually does not provide accurate estimates of these characteristics.

The aim of this work is to provide a method of building a hierarchical tree structures with a widely available free and open source tools: TAS GIS (Lindsay 2005), GRASS GIS (GRASS 2008) and PostgreSQL database (PostgreSQL 2008). Since the information about the hierarchy is stored in an external database, the advantage of this method is its portability to other GIS systems and the ability to manage the data structure with SQL commands without additional programing effort.

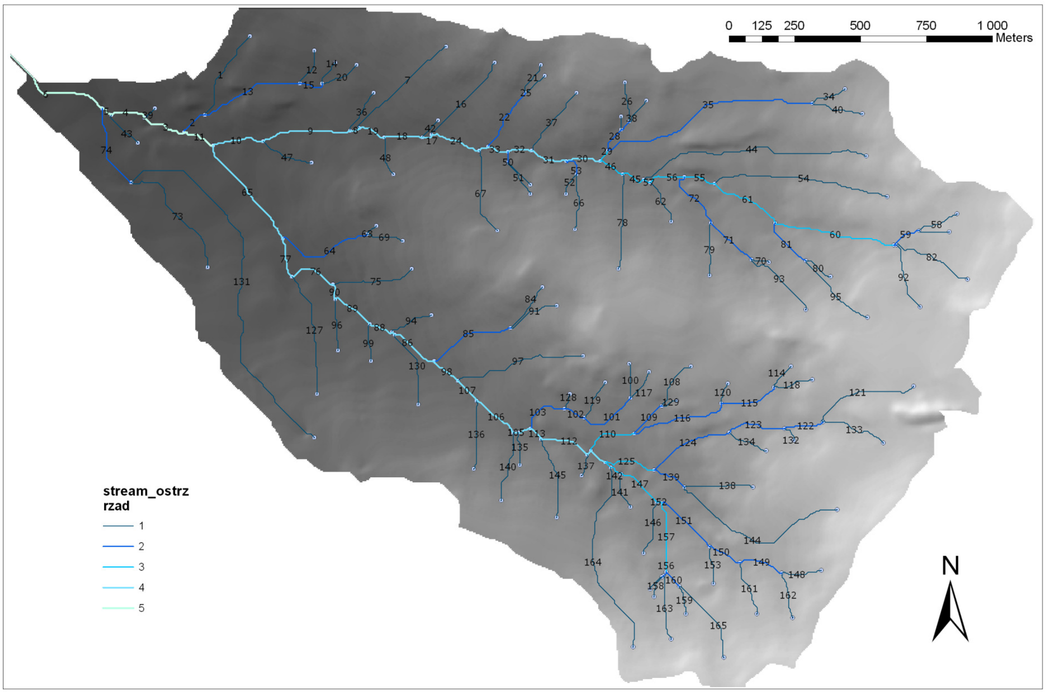

Fig. 1. Drainage network with stream identifiers ans the stream order according to Strhaler (1952)

An example drainage network used in this paper (FIG 1) consist of a northern part of the area of glacially thrusted end moraine of Ostrzeszów Hills (Rotnicki 1967). Digital terrain model has been developed by Marcin Wierzbiński under the supervision of Karol Rotnicki and the author. It must be pointed out that the "drainage network" consist both of stream channel with active flow and "dry valleys" - a remnant of vistulian fluvio-denudational system.

Topological definition of drainage network (according to Bailly et al 2006)

The drainage network (FIG. 2) can be modeled as a directed tree *T*. *T=[N,A]* is described as a connected graph with no circuits where its edges represents drainage channels. Let *N* denote the set of nodes *n* and *A* set of arc *a*. Every edge corresponds to channels segment linking to two nodes: *a=(i,j)* and is naturally directed by water flow from upstream node *i* to downstream one *j*. Both nodes are extremities of an arc/channel. A node: an extremity of directed tree is called source when it is initial extremity of an tree and outlet if it is a final one. From these definitions it follows that we can designate only one node of the network as an outlet. This node is chosen to be the root of the directed tree, but there can be more than one source. Nodes which are not extremities are called junction.

We can consider *T* as the valued graph by associating a value to each arcs and nodes. A directed path between two nodes (*k,l*) corresponds to the water course of drainage network from *k* - the upstream node to *l* - the downstream node.

Drainage network extraction

The procedure of the drainage network extraction has been carried out in the TAS GIS software. In the first step the DEM correction was made using procedure proposed by Lindsay and Creed (2006). Then terrain parameters: D8 specific catchment area (SCA) and slope inclination (SLOPE) were calculated. These parameters became the basis for determining of the drainage network mask using modified method proposed by ~~the~~ Montgomery and Dietrich (1989) according to the formula: *SCA x SLOPE < threshold w*here the threshold was qualified experimentally. Montgomery and Georgiou (1993) suggest raising SLOPE parameter to the power of 2, but this does not appear to result on the gently inclined areas. The drainage network mask was used to designate the topological parameters of the network:

- individual channel identifier (segments) in raster and vector format (FIG 1). Even though GRASS GIS allow to create an individual stream identifier with r.watershed module, limitations of GRASS raster-to-vector conversion tools make their usability for this operation very limited (FIG 3);

- class of channel and junction: channels, initial channels, junction points, source points, and outlets. This structure of data is defined as the leaf-branch directed tree graph, and deal with definition of drainage network proposed by Bailly et al. (2006);
- additionally the channel order according to Strahler classification is created (See FIG 1).

In the next step, all the data was imported to GRASS software, for further analysis.

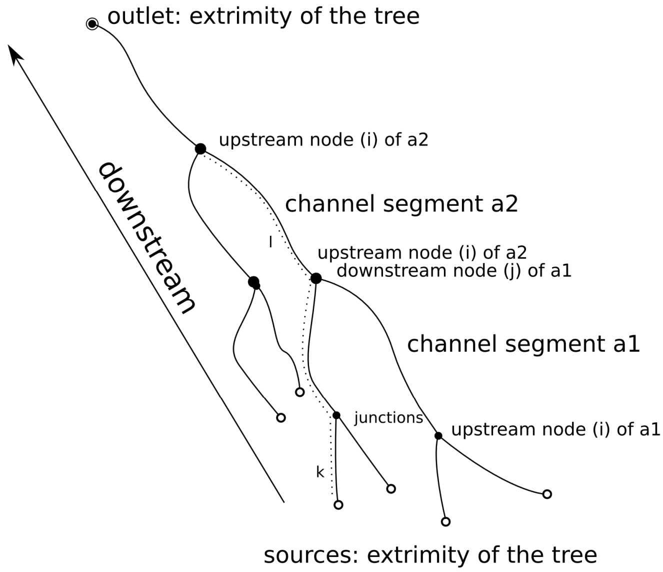

Fig. 2 Definition of network topology

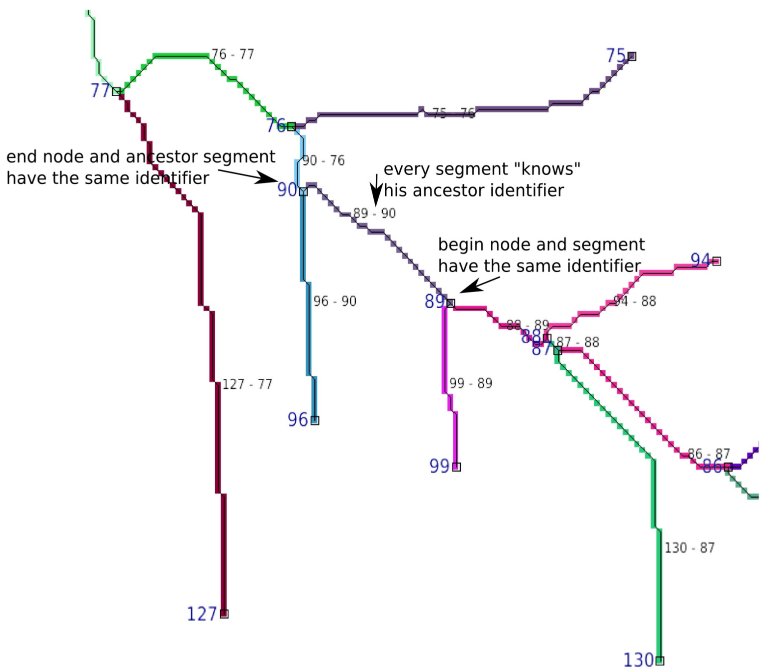

Fig. 3 A comparation between TAS GIS vector network topology and GRASS GIS conversion problems

Building network structure

Internal topology of GRASS GIS (similar to other GIS systems) include information about relation arcs-nodes

and can be used, for example, to find paths between particular arc/nodes using the network tools analysis. Unfortunately, both internal topology and information ~~of~~ on paths tracts are not portable between different GIS systems. It is mostly caused by the fact that GIS data sets are transfered using OGC Simple Feature data model (or even only primitive geometry information) which can transfer attribute table (If we use the ESRI shapefile format) but cannot transfer topology information.
Therefore certain information must be stored in relational data base management system as an attribute table, and then it must follow rules of SQL limitation. The SQL-based data model use relational algebra that is not suitable for hierarchical structures. In advanced RDBMS there are extensions or build-in clauses like CONNECT BY of Oracle or DB2 or WITH RECURSIVE of MS SQL Server. Those clauses are very helpful in managing hierarchies but not belongs to SQL standard and are not parsed by internal engines of popular GIS systems.
GIS data structures can be extended to hierarchical data tree with two data models: the neighborhood model and the materialized path model. The first one uses current arc identifier and arc ancestor identifier and follow two rules:

- every channel segment (segment) has to "know" about his ancestor. The attribute link is the identifier that allow each stream links to be related forward to the ancestor;
- every channel segment (segment) has to "know" about his initial node and every node has to "know" his ancestor channel;

The data model described above is almost identical with the one proposed by Tarboton (1997, 2008) in TauDEM package. Tarboton's (1997, 2008) proposal allows to build ordered trees using recursive SQL queries. Unfortunately, such queries are not available in most databases systems, and even if they are, parsing modules of most GIS applications do not accept such clauses as correct.
The second model uses the series of IDs delimited by separator: ....56:77:122..., where numbers are consecutive identifiers of arcs and can be easily managed with simple SQL queries.

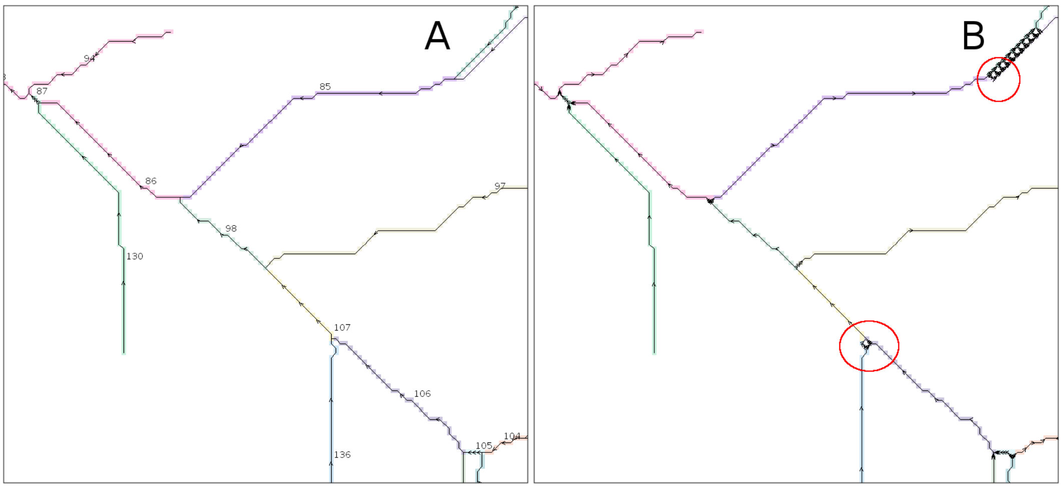

Fig. 4 The procedure of sampling the data to bulid tree topology

To meet situation where each network segment explicitly knows the segment it is associated to, it requires some processing to find segment ancestor. In GRASS GIS software each segment can store its begin and end nodes identifier in another vector layer created on basis of the segment layer (FIG. 4). Simple procedure of sampling data allows to specify particular segment identifier (begin node) and ancestor identifier (end node). This can then be used to establish a relationship between the current segment (link) and the attached segment- (link_to) with simple SQL query where virtual tables of begin nodes and end nodes are joined one another. Next, table received as a result of that query, is attached to the segment layer.

Building the extended network topology as a hierarchical tree

PostgreSQL RDBMS offers extension *tabletools* (Convay 2008) with connectby() function (available form version 8.3), based on CONNECT BY clause of Oracle and DB2. In PostgreSQL a function connectby() applied to the neighborhood model can only perform descending queries (upstream: from any *k* outlet node to all source nodes: FIG. 5a) and choose the subnets belonging to the designated outlet. Fortunately, for each stream function connectby() also generates the path containing the IDs of all streams delimited by separators, from the current one to the one designated as the root of a given subnet. This path is simply a next structure of the data - the materialized path.

The function connectby() is limited only to the PostgreSQL database and GIS system which allows for full processing of PostgreSQL queries. There is a need to make also possible performing hierarchical descending queries, with methods available in a standard SQL. For this reason we can used materialized path. Because it is a text string, it can be easily used to select any subnet for a given stream/outlet with a simple query, using LIKE operator: *LIKE '%:57:%';* where the number used in the pattern means the root stream/node of the subnet. LIKE operator is available in all database systems.
Ascending queries (downstream from any source node *k* to any outlet node *l*: FIG. 4b), selecting all streams along the path from the current one to the outlet are yet difficult to implement. If both the database system and GIS software support the type *explode/split* command or function, it can be used for the field containing the path definition. In the PostgreSQL for this purpose regexpr_split_to_table() function can be used, and the result of its action are nested in the IN expression.

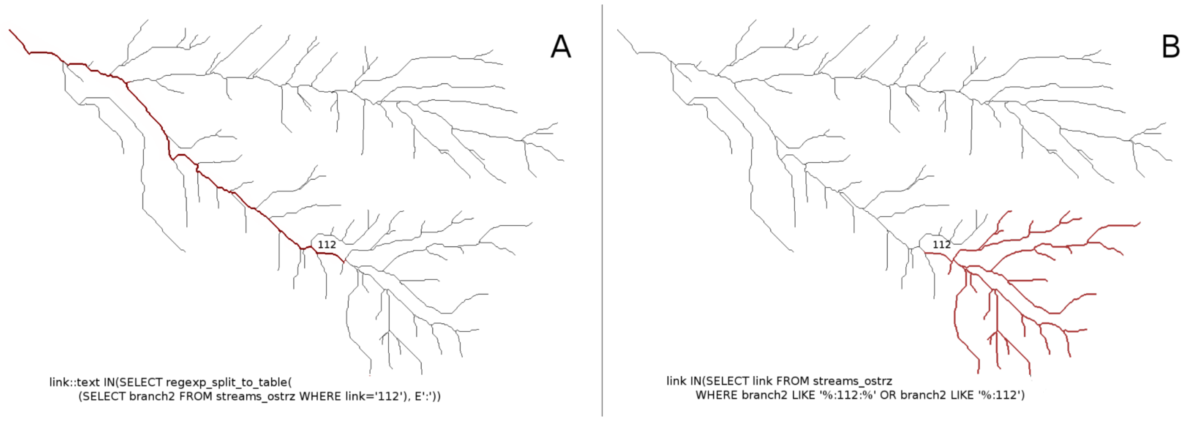

Fig. 5 An example of descending (A) and ascending (B) queries

Managing the tree hierarchy

More advanced operations on drainage networks or generally speaking hierarchical trees with OPEN SOURCE GRASS GIS PostgreSQL suite can be implemented using PostgreSQL extension *ltree* (Sigayev, Bartunov 2008) (fully available form 8.0 version of PostgreSQL). The ltree data type was designed to support leaf-and-branch data structures and represents a path from the root of a hierarchical tree to a particular node/branch. This type of data is nothing more than the materialized path described above and consists of identifiers delimited by separator '.'. Materialized path, as a result of actions connectby() function can be cast to type ltree, using text2ltree() function without additional effort.
The data processing can be easily performed using operators based on the POSIX regular expression language. The data processing covers a wide range of operations including searching the tree, comparing and combining subnets, analyzing individual paths. The combination of ltree capabilities with operations on the numerical attributes of individual channels gives new opportunities in geomorphometry, but goes beyond the scope of this article.

Conclusion

The potential of the terrain analysis offered by TAS GIS in conjunction with the wide possibilities of data processing offered by GRASS GIS / PostgreSQL allow the use of hierarchical data structures using the attributes stored in the table attached to vector data layer. Commonly known neighborhood model can be obtained by means of simple GIS operations and transformed into a materialized path, supported by most GIS applications in order to carry out simple selection operations. Such network, connected with additional SQL tables can store bundle of geomorphological data both for nodes (junctions) and particular vertex of any channel. Storing data in GRASS GIS topological format allow to transfer geometrical data to R environment where geometrical properties of network components like channel directions can be easily calculated without particular skills in programing.
The nature of cooperation between PostgreSQL and GRASS GIS allows the use of ltree data type -for operations that go beyond the means of topological network analysis.

Reference:

Bailly, J-S., Monestiez, P., Lagacherie, P., Modelling Spatial Variability Along Drainage Networks with Geostatistics, Mathematical Geology, *2006, 38(5),* ISSN 0882-8121.

Convay, J., tablefunc. PostgreSQL PostgreSQL 8.3.4 Documentation, *2008.*

Lindsay J.B., The Terrain Analysis System: A tool for hydro-geomorphic applications. *Hydrological Processes*, *2005 19(5),* ISSN 1099-1085.

Lindsay, J. B, Creed, I. F., Distinguishing between artefact and real depressions in digital elevation data. Computers and Geosciences, *2006, 32(8),* ISSN 0098-3004.

Rotnicki, K., 1967: Geneza Wzgórz Ostrzeszowskich. Badania Fizjograficzne nad Polską Zachodnią 19, ISBN 978-83-7063-538-1

Sigayev, T., Bartunow, O., Ltree module for PostgreSQL: http://www.sai.msu.su/~megera/postgres/gist/ltree/, *2008.*

Tarboton, D. G., A New Method for the Determination of Flow Directions and Contributing Areas in Grid Digital Elevation Models, Water Resources Research, *1997, 33(2),* ISSN 0043-1397.

Tarboton D. G., Terrain Analysis Using Digital Elevation Models (TauDEM),
http://www.engineering.usu.edu/dtarb/, *2008.*

Software:

TAS GIS: Terrain Analysis System, Lindsay (2005): http://www.uoguelph.ca/~hydrogeo/TAS

GRASS GIS: *G*RASS Development Team, 2008. Geographic Resources Analysis Support System (GRASS) Software, Version 6.3.0. http://grass.osgeo.org

PostgreSQL: PostgreSQL Database Management System, 1999-2009. PostgreSQL Core Development Team, http://www.postgreSQL.org