Using Data Analytics to Explore "J" (Muddy) Channel Sandstones in the Denver-Julesburg Basin and Test a ML Exploration Technique

R. Mark Maslyn[1]

[1]Consulting Geologist, 10268 Dusk Way, Littleton, CO 80125

Abstract

Some geologic features including river systems can form branching hierarchical network patterns. A trend in storing hierarchical data is to use a NoSQL hierarchical database. Graph databases are an example of a NoSQL database that can model this data category. Graph databases consist of nodes representing entities, with connecting edges representing relationships. Channel sandstones can be modeled as edges with the channel segment starting / ending points and direction. Locations where two channels intersect are modeled as nodes. Once the data is in a graph database data analytics can derive additional information. Different graph specific data analytics can be applied to derive insights and potentially aid exploration. The example described in this presentation uses published data from the Cretaceous "J" (Muddy) Sandstone channels of the Dakota Group of the Denver-Julesburg Basin stored in a graph database, and after applying data analytics, illustrates the results. The "J" Sandstone is the uppermost member of the Cretaceous-age Dakota Group in the Denver-Julesburg Basin of Colorado, Nebraska and Kansas. This unit was deposited in channels eroded into underlying Cretaceous sediments during a regressive shoreline phase. It consists of two members: the basal Fort Collins Member, where present, includes deltaic and marine sediments and overlying Horsetooth Member which includes fluvial channel and estuarine sandstones. A published "J" Sandstone channel sandstone map indicating the directions of water flow and the channel network pattern in the Denver-Julesburg Basin was digitized and stored in a graph database. Individual channel sandstone segments identified in the published mapping were digitized and stored as edges in the graph database, while channel intersections were identified and stored as nodes connected by channel segment edges. Although there are several categories of data analytics that could be applied to this type of data, Google's PageRank was selected because its use for analyzing link flow between websites can be compared to analyzing intersecting flows in channel networks. The PageRank algorithm applied to incoming website links was originally developed by Google to numerically score search results. In simplified terms, the importance of a website is related to the number of external websites that have links to the site being ranked. Sites with greater numbers of incoming links are assumed to be more important and receive higher scores. The numeric ranking of each site is propagated through the nodes in the network and combined by the algorithm with other node's numeric scores to produce a PageRank value. The "J" Sand example uses a similar approach to search result ranking, but instead uses counts of upstream tributary channels as inputs to the algorithm to compute PageRank values. Higher PageRank scores may indicate areas with greater channel flow and exploration potential.

Using Machine Learning to Supplement Geology Tuesday, July 26 3:40 PM