An improved community detection algorithm for classification of catchments in a large region

*Siti Aisyah Tumiran[1,2], Sivakumar Bellie[1,3]

1. School of Civil and Environmental Engineering, The University of New South Wales, Sydney, NSW 2052, Australia, 2. Faculty of Science and Natural Resources, Universiti Malaysia Sabah, Kota Kinabalu, Malaysia, 3. Department of Land, Air and Water Resources, University of California, Davis, CA 95616, USA

In recent decades, there has been significant interest in the development of a catchment classification framework, for identification of appropriate catchment model complexity and predictions in ungaged basins, among other purposes. There exist numerous approaches for classification, with different bases and assumptions, which have been applied for catchment classification. The concepts of complex networks, and particularly community structure, have emerged as important tools for classification, and are currently gaining attention in catchment classification. Among the many community structure-based methods, the edge betweenness (EB) algorithm, which applies a hierarchical clustering concept, is one of the most basic methods for identification of communities (groups) in large dynamically-evolving networks, such as catchment systems. The method’ s signature steps include: (1) an iterative removal of edges by calculation of edge betweenness values that pass through the shortest paths between vertices (i.e. nodes); (2) recalculation of the betweenness values after each iterative removal of edges; and (3) formation of communities using a modularity measure, as the maximum value of modularity representing the best partition of the network. Although the EB method has been effectively applied for classification in many different fields, including in hydrology, the modularity measure that is used to form the best partition of community structure is susceptible to network (or data) resolution or scale problem. As a consequence, communities may change when the size of the network changes. To overcome this resolution or scale problem for catchment classification, we propose an improved EB algorithm, by replacing the modularity measure with the modularity density function, so that the best formation of community structure in the network is represented by the maximum value of the modularity density. We apply this improved algorithm to monthly streamflow data for classification of catchments in the United States. To demonstrate its effectiveness, we study three different scenarios of network sizes: (1) 639 streamflow stations; (2) 300 randomly selected stations (with 100 different realizations) from these 639 streamflow stations –purely to address the network size; and (3) stations in each of 14 different hydrologic units –to address the network size and regional similarity and influence. The results are interpreted in terms of the number of communities that are formed and the number of stations that change from their communities when the network size changes.

Keywords: Catchment classification, complex networks, edge betweenness, modularity density

