IDS 564: Dr. Ali Tafti, Spring 2016

Advanced Lab 3: Community Structure of the SAP Online Knowledge Community Platform

In this optional advanced lab, you will examine a network drawn from the SAP Online Knowledge Community platform (http://go.sap.com/community.html). This is an induced sub-network based on 10% of nodes drawn from the dataset generously provided by Prof. Peng Huang, of the University of Maryland. See here for more on his research studies: http://www.rhsmith.umd.edu/directory/peng-huang.

You are provided with two CSV files: 1) The file "SAPSUB_edges.csv" has the edge list. The right-hand column represents the ID of a user who posts a question that starts a new thread in a SAP community user forum (*thread_author_id*). The left-hand column represents the ID of a user who provides an answer to the posted question (*message_author_id*). So, a directed edge from the left-hand node to a right-hand node represents an answer provided to a question. Since a user can answer multiple questions posted by another user on one or more threads, the file contains duplicate edge-pairs, which can be combined to form directed edges of varying weights (using the simplify function in R). 2) The file "SAPSUB_nodes.csv" has the list of nodes and their attributes. User ids are identified by "*author_id*", which correspond to the thread and message authors in the edges file.

You may also refer to the description and specifications for Advanced Lab 2.

Please do the following, and submit your report in a PDF file of 5 pages or less. Also, please submit a completed R script that generates any plots or statistics discussed in your report.

Community structure

Please visualize, analyze and describe the community structure of the SAP online knowledge community; and evaluate the results of the faster community detection algorithms available in igraph (i.e. try the fast-greedy and walktrap algorithms, then see if there are any other reasonably fast algorithms).

Community structure, Nationality and other observable attributes

Please explain to what extent the nationality of individuals places them into distinct communities. What other observable attributes given for each node might explain the community structure?

Community structure and points generated

The attribute *ln_points* gives the logarithm of award points accumulated by each user in the network. When aggregated to the community level, this measure could represent the productivity level of each community in producing new knowledge in the system. Please provide an analysis for how, and to what extent, certain network attributes measured at the community level might explain knowledge contributions (points awarded to individuals in the community).