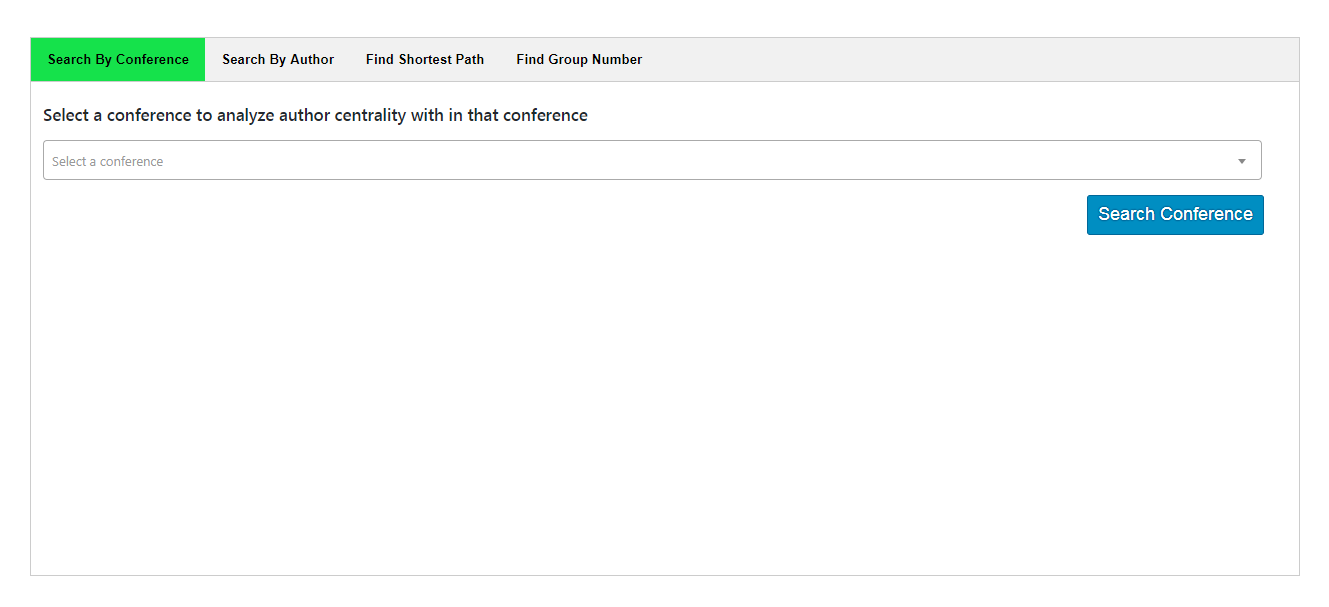
Algorithmic Methods of Data Mining

Homework 4

Group 3 – Vamsi Krishna Varma Gunturi, Matteo Manzari, Andrea Ferrara

To better visualize our hard work ☺ done by our group, we decided to create a small web application (an idea of Vamsi). Below a screenshot:

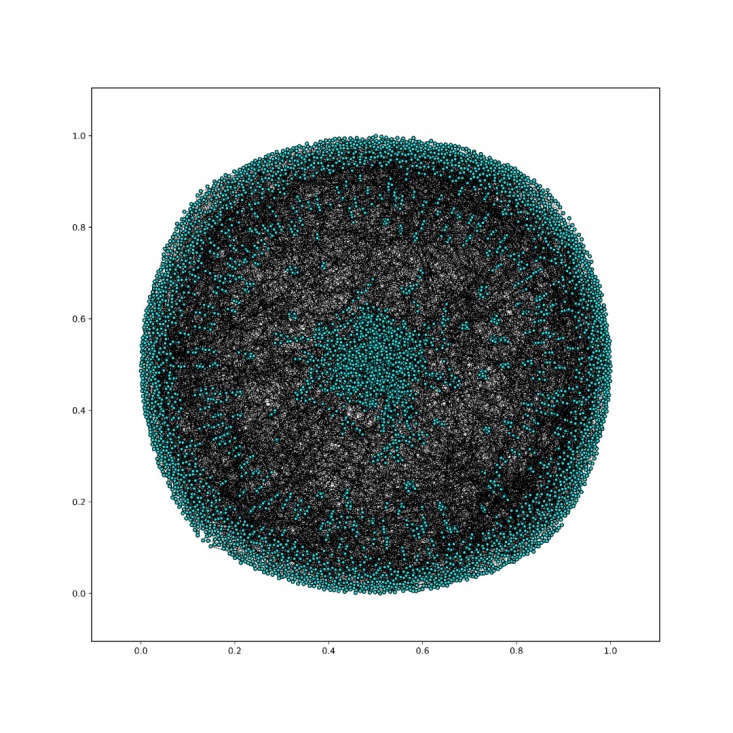


There are 4 sections that refer (in order) to points 2.a, 2.b, 3.a and 3.b and that points will be explained below.

Part 1 – By processing Json file create a graph G

A graph G= (V,E) consist of a set of nodes V, and a set of edges E. In our case we define as nodes the authors and the edges as weighted that is compute by 1- J(p1,p2) or 1- Jaccard similarity between two sets of pubblications thats “are contained” in the authors.

In our case we considered that graph G undirected and we assume that (u,v) and (v,u) are the same edge.The representation of the graph (reduced file json in input) is shown below:



Part 2 – Compute some statistics and visualizations

2.a

Centrality measures how central is the node in the graph. Depending on what we mean by “central”, there are versions of centrality measure.

1. Degree centrality:

is a simple centrality measure that counts how many neighbors a node has.

1. Closeness centrality:

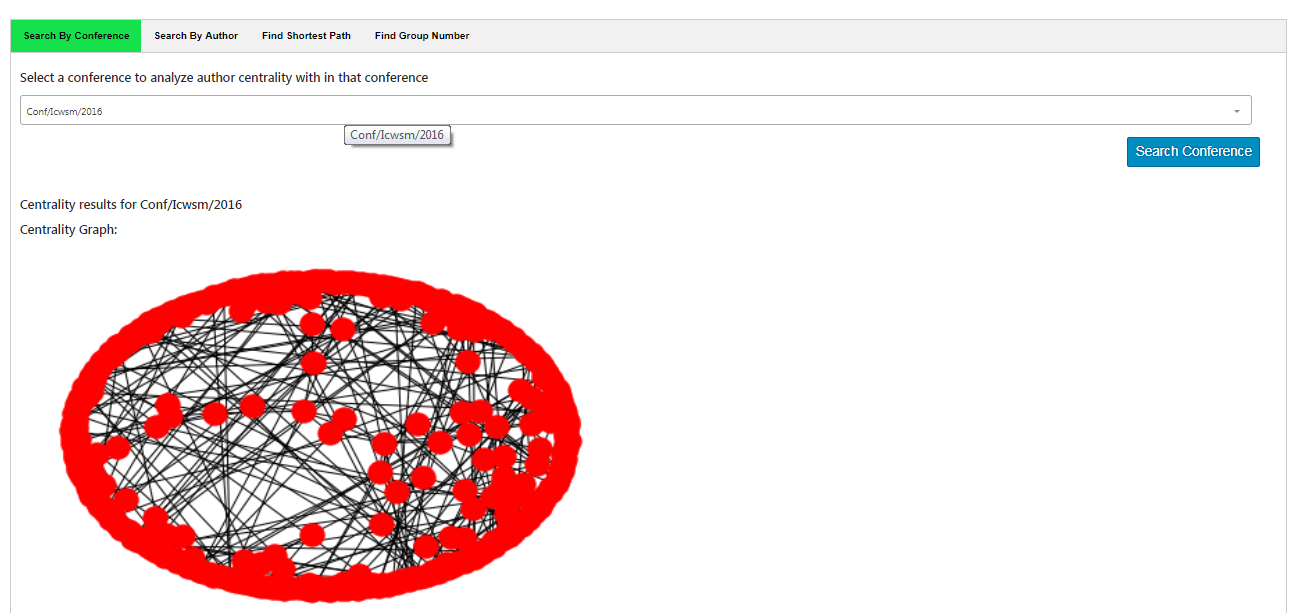
measures the mean distance from a vertex to other vertices. Intuitively a node is

central if its distance from the other nodes is small.

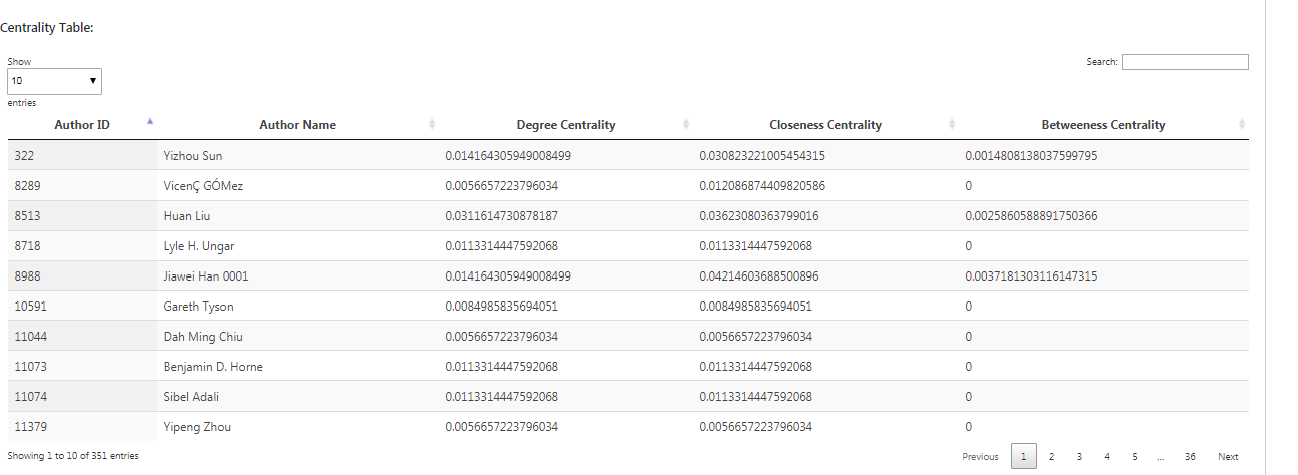
1. Betweeness centrality:

measures the extent to which a vertex lies on paths between other vertices.

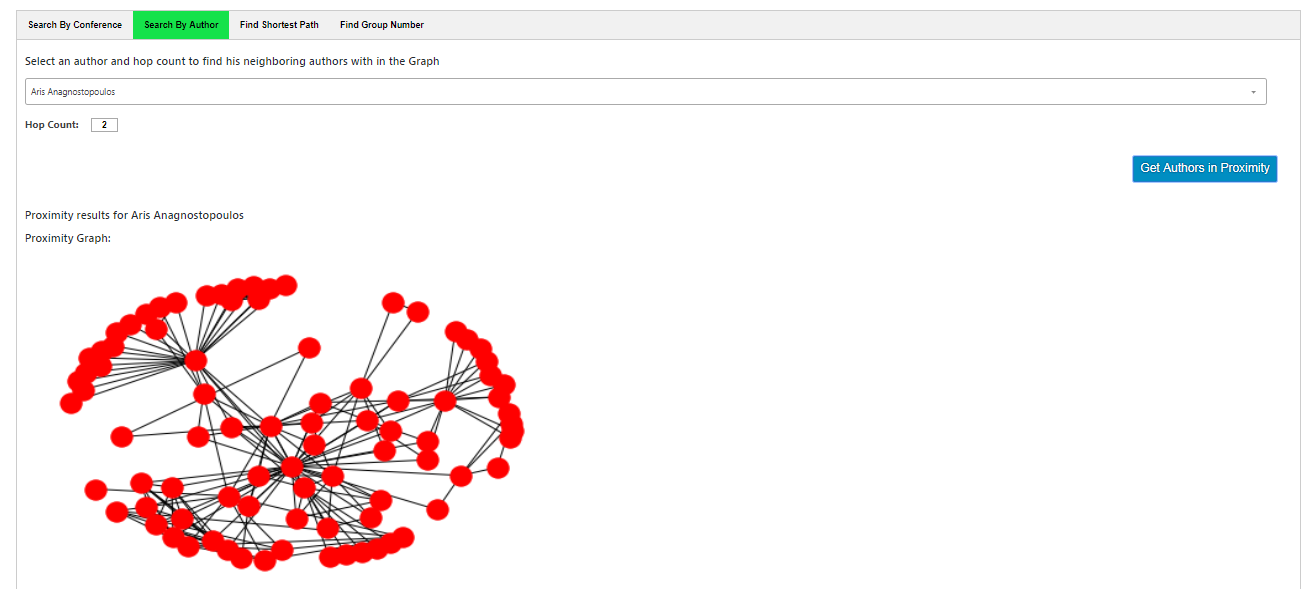
In the web application “Search by Conference”, below is an example for the conference “Conf/Icwsm/2016”



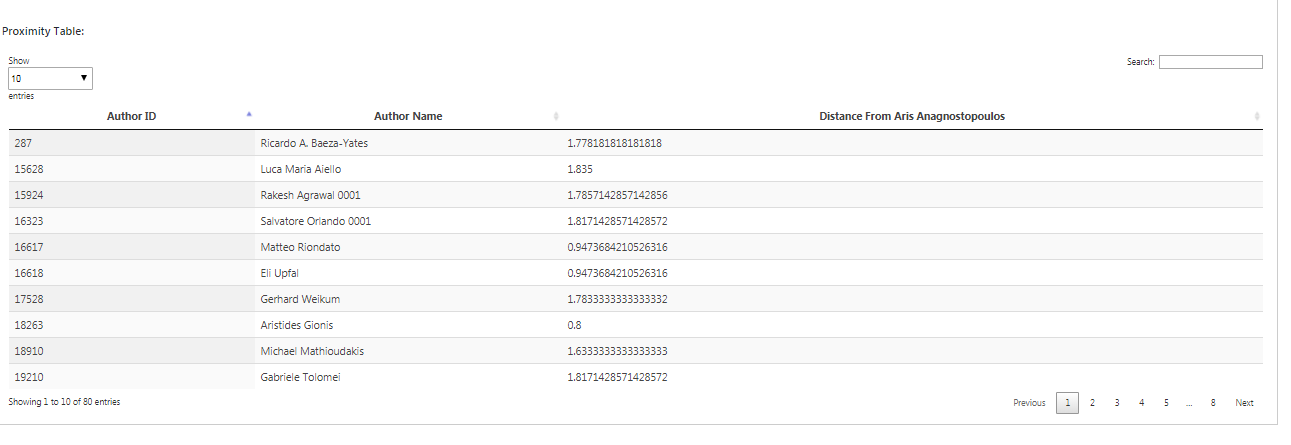
and its table with the results of centrality measures:



2.b In the web application if you select an author and the number of the hop count you can visualize this (in this example we have selected our fantastic professor ☺ ):



And its table whith the measure of the distances between Aris and another authors:



Part 3 – Compute some generalized version of the Erdos number.

# To compute “Aris number” we have implemented from scratch the algorithm to compute the shortest path. So we have following the logic of the Dijkstra’s Algorithm.

# In the web application you can see this in the “Find Shortest Path”.

For instance we compute the weight of the shortest path that connects “george brova” and Aris and the result is: