

An analysis of co-authorship network

Team 19: Fengyu Cai, Wanhao Zhou, Junze Li, Liangwei Chen

Purpose

The project we propose is to discover both static and dynamic patterns of the co-authorship network.

Dataset

We will use a [co-authorship graph](#) built from the DBLP library, collected from the main conferences or journals in the field of Data Mining and Database communities between January 1990 and February 2011. Each node will represent one author, and each link between nodes will represent the co-authorship at least one paper (whether the conference or journals) between them. There are 38 properties for each node, including the number of publications in 29 selected conferences or journals, and 9 more topology properties such as PageRank.

Meanwhile, the data provider also offers a dynamic co-authorship dataset, chunked by a time period of five years. This subset of data will focus on fewer authors who had at least 10 publications, and more properties such as timestamps.

Goals

We would like to analyze from both micro and macro perspectives.

From the macro perspective, we will focus on:

- the detection of existing communities,
- finding patterns of such communities (size, sub-communities, and relations with adjacent communities),
- discovering the evolutionary characteristics of these communities (e.g. the emergence/diminishing of a community)

From the micro perspective, we will focus on:

- discriminating whether a certain individual author belongs to a given community or not,
- predicting whether there exists co-authorship between two newly introduced vertices with the graph features we learned,
- for a certain individual author, how his/her membership to communities changes with time evolves