Network Analysis

Method background

Main idea of Network Analysis approach is to to find, represent, and analyze the impact of relationship patterns. Network analysis looks at the relationship between two basic components - nodes (people, documents) and ties (relationships, associations). The goal of network analysis is to gather information about the structure of the network and uncover patterns of connections and information flow between them.

Some common network analysis applications include data aggregation and mining, network propagation modeling, network modeling and sampling, user attribute and behavior analysis, community-maintained resource support, location-based interaction analysis, social sharing and filtering, recommender systems development, and link prediction and entity resolution. In the private sector, businesses use social network analysis to support activities such as customer interaction and analysis, information system development analysis, marketing, and business intelligence needs.

The nature of method can be both inductive and deductive. Network analysis can both help to explore new phenomena and proof or reject already existing hypothesis. Network Analysis method collocates with pragmatical philosophical worldview sharing the idea that knowledge is generated through deductive and inductive tools. From the point of Jarvinen's taxonomy NA has a certain attitude to Mathematical approaches.

Although there are no unified grandfathers of this method, early sociologists Georg Simmel and Émile Durkheim, who wrote about the importance of studying patterns of relationships that connect social actors in 1890-1900s, can be called as founders of Network Analysis method.

Example of applying this research method can be an article called "A Clustering Approach for Collaborative Filtering Recommendation Using Social Network Analysis".

Data Collection

For this method researcher can rely on either secondary data (like publicly available records) or primary data (where own information is collected) to conduct analysis.

The datasets usually represent structured data with predefined number of attributes. All attribute types including discrete attributes (nominal and ordinal) and continuous attribute (interval and ratio) can be used for analysis.

The most common sources of data are questionnaires, direct observations, written records (archival or diary), experiments and derivation.

Method Implementation

Let's assume the case where student/researchers of the university (or any other organization) want to analyze the incoming traffic to their official website.

In this case the dataset represents a log file from HTTP server that hosts the main website of the university (given organization). Every line of the dataset represents a request from particular user and contains the following attributes: timestamp, target URL path, session identifier, IP address of the user that sent the request, returned HTTP code, amount of data returned (in KB).

Applying network analysis method to this dataset we will be able to build and visualize a network where every node represents a particular HTML page (endpoint) and connections show history of visiting different pages of particular user. The user can be identified by unique IP address and one history of redirects can be defined as one unique Session ID.

Having built such network it will be possible to find and analyze common behaviours of users (list of pages they usually attend and in which order), identify which pages are more popular than others and from which endpoints the visit to the university (organization) web-resource starts and where it ends.

It's worth noting that similar characteristics of the dataset can be found using other research methods as well, but using network analysis we will benefit from building a graph that can be easily visualized and used to identify some patterns and properties manually without hard computational calculations. So first hypothesis can be made by visually exploring produced graph, and then the proposed theories can be proved/rejected by using other research methods.

Taking into account large amount of algorithms and platforms for network analysis we can be confident in reliability and validity of this method. Large amount of open-source projects gives opportunity to check that the algorithm does what we exactly want and vast amount of platform allows us to process and make research with huge amount of data measured in gigabytes and terabytes.