Stanford

World Wide Web, blogging platforms, instant messaging and Facebook can be characterized by the interplay between rich information content, the millions of individuals and organizations who create and use it, and the technology that supports it. This thesis will cover recent research on the structure and analysis of large social and transaction networks and on models and algorithms that abstract their basic properties. Unusual ways have been explored how to practically analyze large scale network data and how to reason about it through models for network structure. Topics include methods for network community detection, their connection with transactional graphs.

Article: 0709.2938v1

Community detection and analysis is an important methodology for understanding the organization of various real-world networks and has applications in problems as diverse as consensus formation in social communities. Currently used algorithms that identify the community structures in large-scale real-world networks require a priori information such as the number and sizes of communities or are computationally expensive. I intend to rely more on algorithms, which use the network structure as their guide instead of this priori information. Finding community structures in networks is another step towards understanding the complex systems they represent. Social networks are represented by people as nodes and their relationships by edges.

Girvan-Newman algorithm

The Girvan-Newman algorithm for the detection and analysis of community structure relies on the iterative elimination of edges that have the highest number of shortest paths between nodes passing through them. By removing edges from the graph one-by-one, the network breaks down into smaller pieces, so-called communities. The algorithm was introduced by Michelle Girvan and Mark Newman. The idea was to find which edges in a network occur most frequently between other pairs of nodes by finding edges betweenness centrality. The edges joining communities are then expected to have a high edge betweenness. The underlying community structure of the network will be much more fine-grained once the edges with the highest betweenness are eliminated which means that communities will be much easier to spot.

The Girvan-Newman algorithm can be divided into four main steps:

1. For every edge in a graph, calculate the edge betweenness centrality.
2. Remove the edge with the highest betweenness centrality.
3. Calculate the betweenness centrality for every remaining edge.
4. Repeat steps 2-4 until there are no more edges left.

NetworkX

NetworkX is a Python package for the creation, manipulation, and study of the structure, dynamics, and functions of complex networks. It supports a variety of features for complex networks.

* Data structures for graphs, digraphs, and multigraphs
* Many standard graph algorithms
* Network structure and analysis measures
* Generators for classic graphs, random graphs, and synthetic networks
* Nodes can be "anything" (e.g., text, images, XML records)
* Edges can hold arbitrary data (e.g., weights, time-series)
* Open source 3-clause BSD license
* Well tested with over 90% code coverage
* Additional benefits from Python include fast prototyping, easy to teach, and multi-platform