Explorative analysis of real world dynamic networks

Benedikt Meuthrath, AG NBI, Freie Universität Berlin

Abstract

Network analysis is an established method in mathematics, physics, biology, computer sciences, sociology and management studies to analyze significantly unlike relationships between entities such as paths between cities, interaction between living cells or the flow of information between human beings. In this context, the interpretation of different network metrics helps to gain an understanding of travel routes, how complex organisms work or how knowledge is spread in a company.

A network consists of a set of nodes and a set of edges, which connect nodes. Often the analyzed data set, what from the network is extracted, includes a time factor, e.g. timely shifted cell growth or the evolution of a wiki's content. The dynamics, responsible for this time factor, result from multiple change processes such as natural evolutionary processes including learning, birth and aging as well as intervention processes such as altering the set of individuals who lead a system. Therefore this time factor has to be considered in analyzing networks: dynamic network analysis (DNA). There are two aspects when analyzing dynamic network data: statistical analysis and a utilization of simulation to address issues of network dynamics. One challenge in DNA in contrast to traditional social network analysis is that in order to understand the dynamics of real world networks, a lot more of different data needs to be examined, that means dynamic networks are larger, multi-modal, dynamic, multi-plex networks and they may contain varying levels of uncertainty.

Explorative analysis means there is no hypothesis given that should be proven with a given methodology. In contrast by visualizing and analyzing the data and measures on it the analyst explores the information space spanned by the data for the purpose of formulating hypotheses worth analyzing. Because exploration is an interactive process, calculation and visualization should be done as fast as possible. For that parallel computing is one approach that is lay out in diverse publications. Visualization and analysis of *dynamic* networks is a challenge because from a network's point of view dynamics mean changes in structure and positions. Therefore network measures and layout have to be recalculated for every point in time that is analyzed. Because of the dimension of real world dynamic networks a complete recalculation would consume much time.

The hypothesis is that semantical information about

- how calculations for measures and visualization are accomplished,
- how they can be parallelized,
- how they depend on each other and
- on which precalculations they depend (like e.g. shortest paths, minimal spanning trees)

can be used at runtime to avoid recalculation and to identify parts of the network or (parts of) measures that don’t need to be recalculated. Results in form of calculation times are quantitatively compared to other approaches in dynamic network analysis.