Modern tools for complex networks analysis

In today's world, both scientists and employees of various industries have to deal with processing enormous volumes of data. These are often data that naturally form networks (so-called real-world data networks). Examples include large data sets derived from social networks (like Facebook, Twitter, or Pinterest), biological networks (e.g., protein interaction networks or neuron connections in the brain), telecommunications networks, power grid networks, transportation networks (e.g., airline networks), etc.

Mathematical models that represent this kind of data are large graphs, known as complex networks. These are structures consisting of vertices and edges. Depending on what the graph represents, vertices and edges can have different meanings. For instance, (i) in social networks vertices correspond to people/entities registered, and an edge between two vertices reflects a friendship relation or the act of following; (ii) in biological networks vertices correspond to protein molecules, and edges of various kinds mimic different types of interactions between molecules. The dynamic nature of data in real-world networks (e.g., the inflow and outflow of users in social networks or appearing and disappearing connections between neurons) is reflected in graph models by adding or removing vertices and edges.

The large size of real-world data and their dynamic nature pose a considerable challenge for both analysts and algorithm developers. Gathering such data in the memory of a single machine and subjecting it to static analysis is often impossible. Similarly, applying standard algorithms to such data is unfeasible as they would require the use of complete knowledge about the studied network. Due to memory constraints, 'online' methodologies that assume we learn the network structure over time also prove ineffective. Hence, the need arises for applying unconventional data processing and analyzing techniques, e.g., methods derived from, so-called, Big Data analytics. An example might be generating a data sketch - a small dataset that allows recreating the most crucial information about a real-world network.

The project's aim is to construct new and improve existing tools for working with large real-world data networks. The research problems have been divided into three groups:

1. Designing mathematical models that match well contemporary real-world networks.

2. Identification and analysis of parameters of complex networks which are significant from the algorithmic point of view.

3. Construction of efficient algorithms that allow estimating selected parameters of real-world networks using Big Data analytics techniques.

Having accurate mathematical models will allow to understand better phenomena occurring in the systems surrounding us and reliably predict their future behaviors. Possessing methods that precisely estimate selected parameters of real-world networks is necessary to construct algorithms appropriately tailored to the data. This will result in, for instance, improving recommendation systems, more effective spam filtering, identifying bothersome bots operating in networks or effective fraud detection .