Network Science: algorithms to analyze and optimize networks structures and solve business problems

Network science is the study of connected things. Things can be represented by people, devices, companies, governments, etc. Connected can be represented by calls, messages, likes, money transferring, references, geo-positioning, contracts, etc. The study of networks has emerged in diverse disciplines as a means of analyzing complex relational data. Every industry, in every domain, has information that can be analyzed in terms of linked data. Many practical applications of network analysis and network optimization depend on an underlying network. Networks appear explicitly and implicitly in many application contexts. Networks are often constructed from natural co-occurrence types of relationships, such as among researchers who coauthor articles, actors who appear in the same movie, words that occur in the same document, items together in a shopping basket, terrorism suspects who travel together or are seen in the same location, etc. In these types of relationships, the strength of interaction is modeled as weights on the links of the resulting network. Network analysis can be employed to avoid churn, diffuse products, and services, detect fraud and abuse, identify anomalies, and many other applications, in a wide range of industries such as communications and media, banking, insurance, retail, utilities, and travel and transportation. Network optimization includes algorithms that can augment more generic mathematical optimization approaches. In network analysis and network optimization, nodes can represent people, mobile phones, computers, stores, locations, cell towers, bank accounts, vehicles, trains, airplanes, or even a walking person. The occurrence of a transaction involving two nodes can represent a link between them. For example, trucks delivering goods to multiple stores. The stores represent the nodes. The truck going through these stores represent links between the stores. The weight of the links can be defined by the frequency of the trips between any two stores, or the number of goods delivered, or a combination of both. Network methods can represent nodes and link in the same way but use them differently when distinct analysis and optimization algorithms are applied. For example, in the finance industry, bank accounts, individual people and organizations can represent nodes. All transactions between bank accounts, persons and organizations can represent links. Money transfers, withdraws, payments, wires, etc., they all represent different types of connections between these nodes, which may have different importance, or weights. In traditional network analysis, components, cores, or communities of nodes can be detected, revealing groups of accounts, persons or organizations that are strongly related. Highly weighted links can highlight connection between accounts, persons and organizations that are not usual. In network optimization, the same network, based on the same graph representation, can be used differently under distinct optimization algorithms. Cycles, cliques, paths, and shortest paths can reveal patterns on the connections that cannot be highlighted by traditional network analysis. Pattern matching can search for specific groups of nodes with particular types of links among them. A network where stores are nodes and trucks traveling are links; optimization algorithms can search for optimal routes to minimize the time or the distance when trucks are delivering the goods. By adding a depot as a node and a vehicle capacity for the truck, network optimization algorithms can also search for the optimal trips that truck needs to perform when delivering the goods from the depot throughout all the stores. This presentation covers several algorithms in network analysis and network optimization, in a step-by-step coding method, exploring changes in variables, resources and constraints, and the final outcomes through interactive graphs for network analysis and interactive geo maps for network optimization.

<u>Outline:</u>

Sub-Networks

- Connected components
- Bi-connected components
- Community detection
- K-Core decomposition
- Reach network
- Projection
- Node similarity
- Pattern match

Network Centrality Measures

- Degree
- Influence
- Clustering coefficient
- Closeness
- Betweenness
- Eigenvector
- PageRank
- Hub and Authority

Network Optimization

- Clique
- Cycle
- Linear assignment
- Minimum-cost network flow
- Maximum network flow problem
- Minimum cut
- Minimum spanning tree
- Path
- Shortest path
- Transitive closure
- Traveling salesman problem
- Vehicle routing problem
- Topological sort

An introduction to network science, including basic concepts, the type of data, and network structures. This presentation is divided into three sections. The first one explores the concepts of subnetworks, including algorithms to search for connected components, bi-connected components, to community detection, k-core decomposition, reach network, projection, nodes similarity and pattern match. Section two covers network centralities, including degree, influence, clustering coefficient, closeness, betweenness, eigenvector, PageRank, hub, and authority. The final section covers network optimization algorithms, including clique, cycle, linear assignment, minimum-cost network flow, maximum network flow problem, minimum cut, minimum spanning tree, path, shortest path, transitive closure, traveling salesman problem, vehicle routing problem, and topological sort.