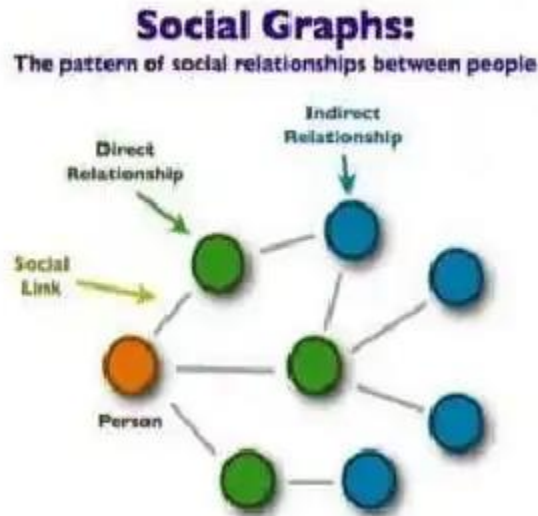


A survey of link prediction in social networks


Link prediction in social networks is a critical task that involves predicting future or missing connections (edges) between nodes (users, entities, etc.) in a social network graph. It has numerous applications, including friend recommendation, identifying potential collaborations, and understanding network dynamics.

A single person is the node of the network while edges, that link nodes and are called also "connections", "links", correspond to relationships between people as represented.



Social networks are graph structures whose nodes or vertices represent people or other entities embedded in a social context, and whose edges represent interaction or collaboration between these entities. Social networks are highly dynamic, evolving relationships among people or other entities. This dynamic property of social networks makes studying these graphs a challenging task. A lot of research has been done recently to study different properties of these networks. Such complex analysis of large, heterogeneous, multi-relational social networks has led to an interesting field of study known as Social Network Analysis (SNA) Link prediction, the focus of my survey, is a sub-field of social network analysis.

Link prediction is concerned with the problem of predicting the (future) existence of links among nodes in a social network. Link prediction is the only sub-field of SNA which has focus on links between objects rather than objects themselves. This makes link prediction interesting and different from traditional data mining areas which focus on objects.

	On Objects	On Links
Predictive Task	X	
Descriptive Task	X	

Link prediction is applicable to a wide variety of areas like bibliographic domain, molecular biology, criminal investigations and recommender systems. Most of the surveyed papers use bibliographic domain data as it proves to be good representative data for link prediction. Also, such data is readily available on the Internet sites like DBLP, BIOBASE, CiteSeer, etc. Thus, most of the papers use co-authorship as a running example. Link prediction problem can be posed as a binary classification problem that can be solved by employing effective features in a supervised learning framework. Thus, in every paper that I surveyed, link prediction is done using topological features of the social network i.e. existing links, and various classification algorithms are applied to classify future link as Yes/No.