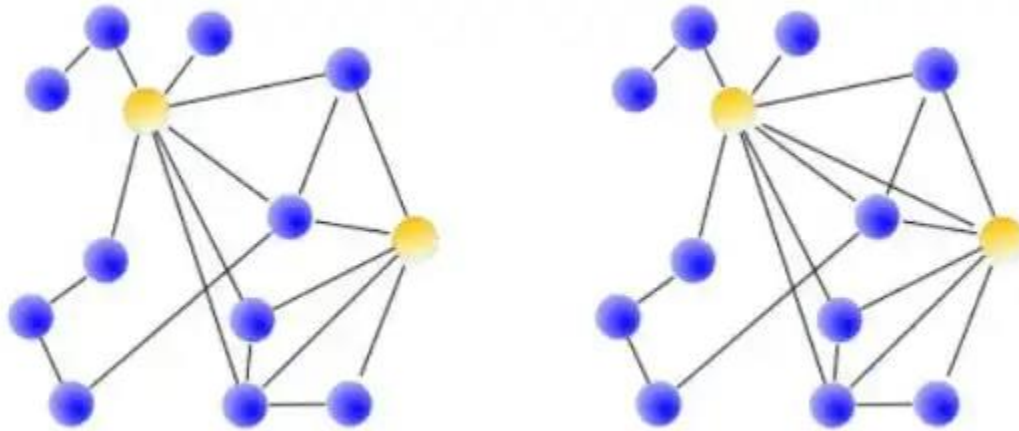# Link Prediction

The nodes in a sociogram are linked in a complex web of relationships that change over time. These relationships emerge, strengthen and decay as a result of individuals' positions in the network, their behaviour and the influence of the environment. Predicting changes to a social network is called the link prediction problem.

Liben-Nowell and Kleinberg explain it as:

Given a snapshot of a social network at time t, we seek to accurately predict the edges that will be added to the network during the interval from time t to a given future time t'. The problem is illustrated in the graphs below, where a link is forming between the orange nodes.

Link prediction is generally approached in two separate (but complementary) ways: relational analysis (or topological analysis) and feature analysis. The first way examines a sociogram for unbalanced social structures that should tend towards a state of equilibrium (e.g. two people who have a lot of mutual friends should eventually meet). The second way does not involve graph theory at all, but rather looks at the content of communications between individuals to search for common interests (e.g. two people who discuss both fly-fishing and abstract algebra in emails should eventually meet).

In this it concentrates only on the former method – predicting links using graph theory – and disregards the content of communications. This method was chosen so as to concentrate on only one part of the problem and perform detailed research on it alone. The purely graph theoretical technique will almost certainly not be as accurate as one that uses both graph theory and content analysis combined, but it means that it is suitable for analysing networks where no content is available (such as email logs, which are numerous and easily accessible to intelligence and business analysts).

Additionally, a purely graph theoretical approach to link prediction can later be augmented with content analysis.