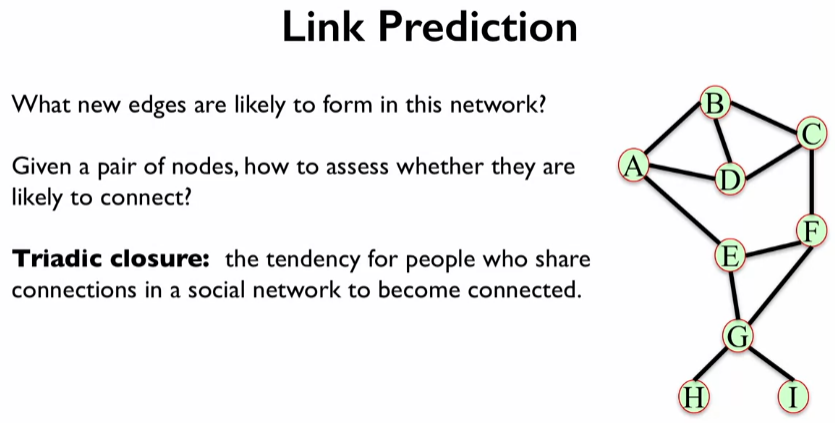
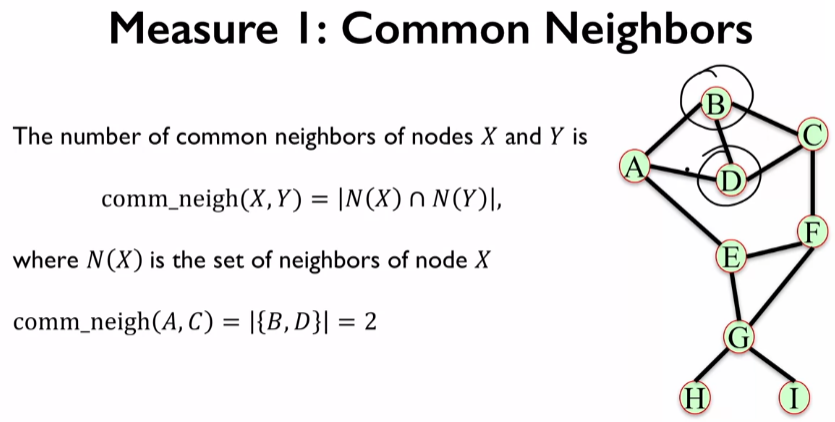
**Link Predictions:**

This document will look at networks as dynamic structures that can change over time. We’re going to look at a fixed network and then try to predict how its going to grow over time.

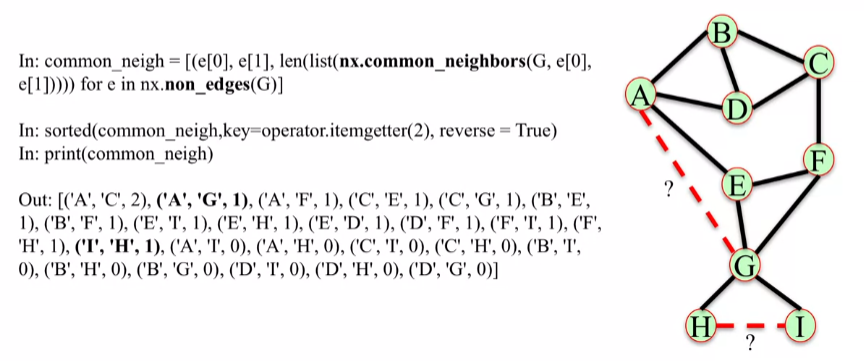
**Link predictions** looks at what new edges are likely to form. This is very important to companies like Facebook, who’s platforms success is partly based on their friend recommendation system, which is essentially new edges in a network.



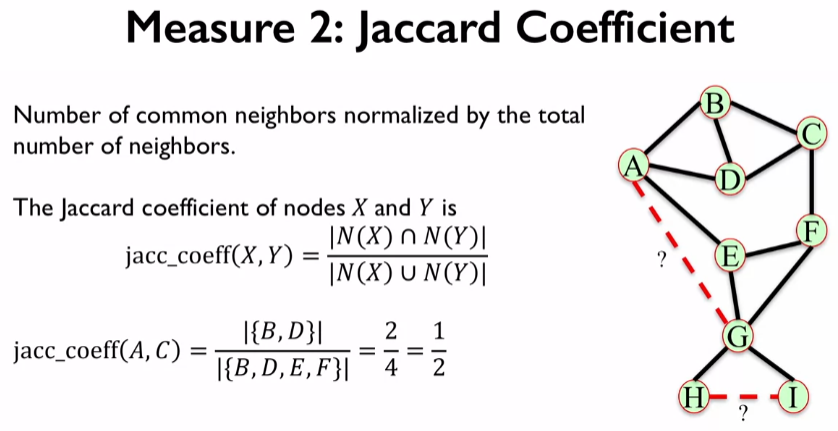
The first measure we’re going to look at is the number of common neighbours a nodes has; **Triadic closure** give us a hint at this measure.

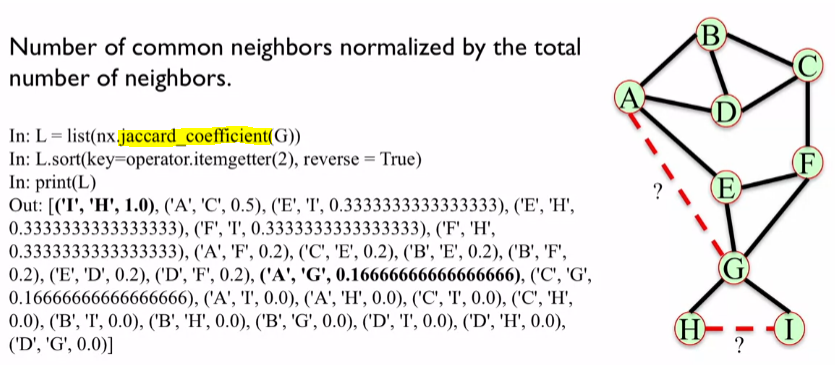


Its important to note that the common neighbours X, Y cannot be directly related, they need more than an edge between them.

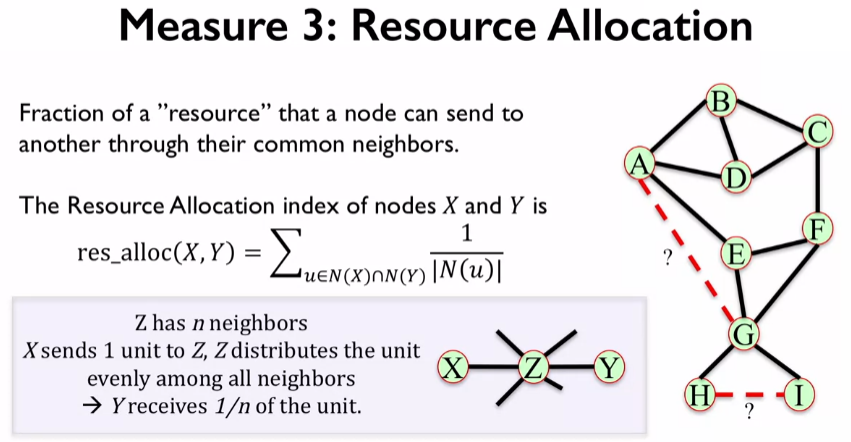


The output of the above code is a tuple of the two nodes and the number of common neighbours they have. If we wanted to ask the question of which nodes would be more likely to connect, we could consult the number returned from the about code. In this case is H-I or A-G more likely to form a connection, they are both equally likely as they both have 1 mutual connection. We can get more information to help us decide by using another measure.

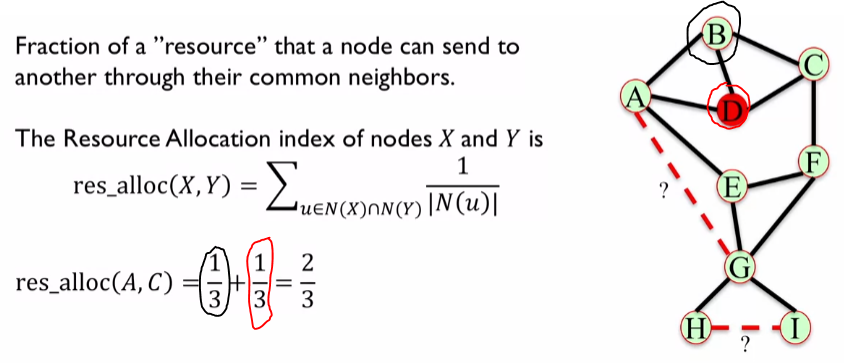


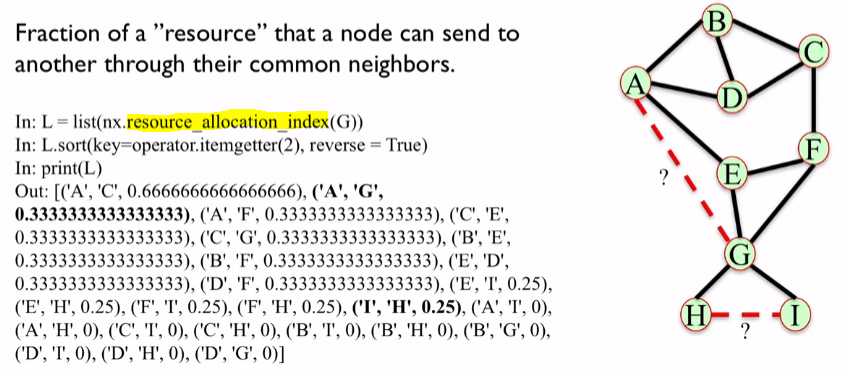


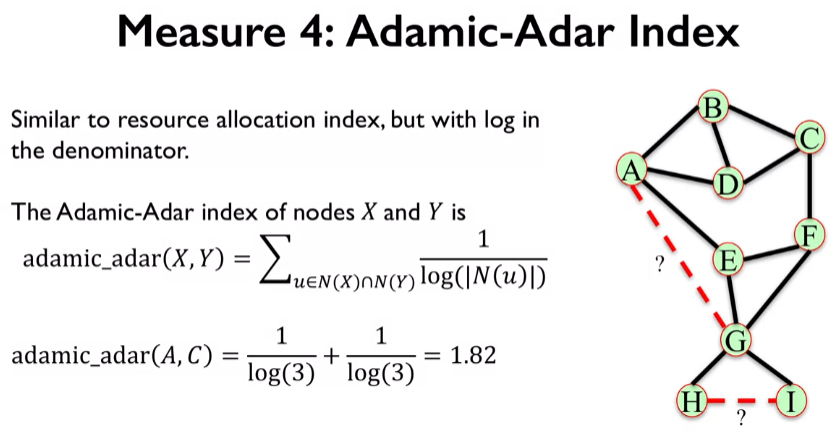
Going back to the question we can see that H-I is more likely to form a connection.

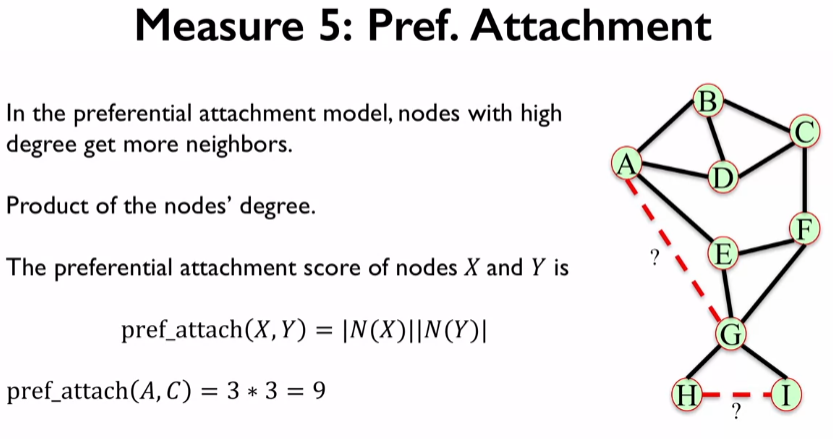


Consider the nodes X and Y, and that X wants to share some chocolate with Y. X passes this to Z first, who then divides the chocolate evenly between all its connections and only ends up giving Y 1/5th of the chocolate. In this case the fraction of chocolate received by Y depends on the degree of Z.

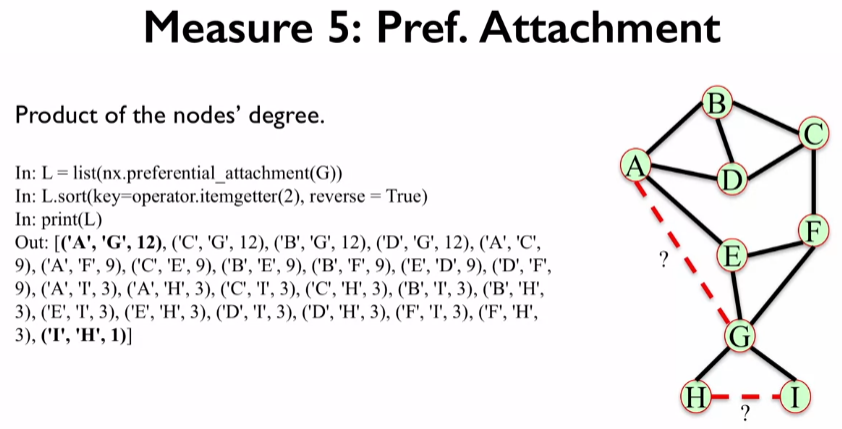




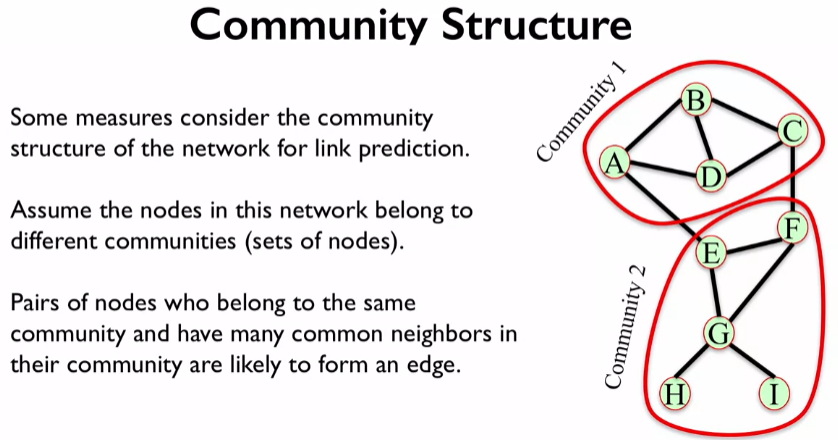


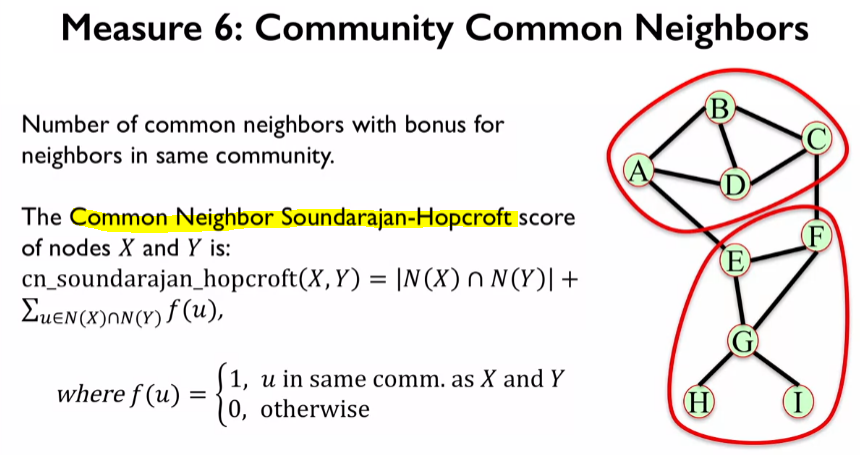


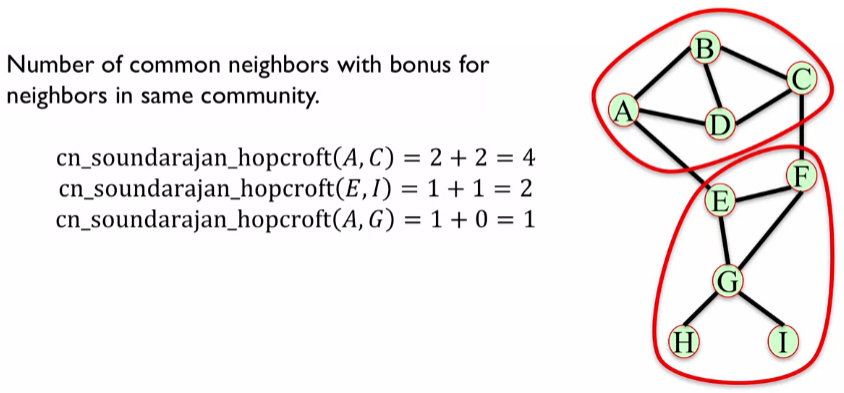
Where N(X) is the number of connections that X has.



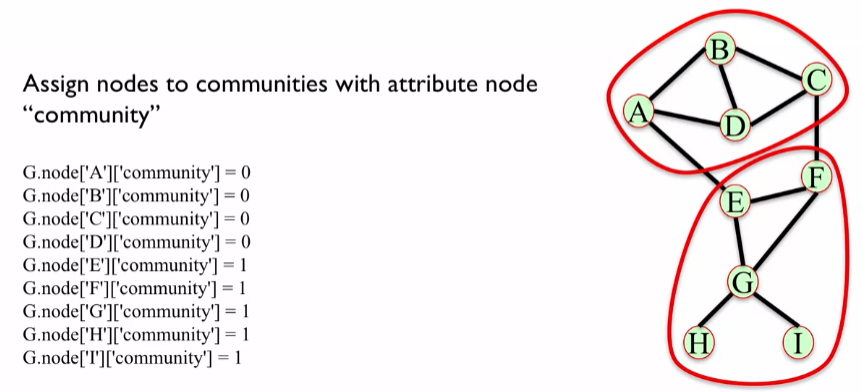
Now we’re going to look at two different measure that consider the community structure of the network. E.g. this could be different departments in a business. This allows us to compute new measures.

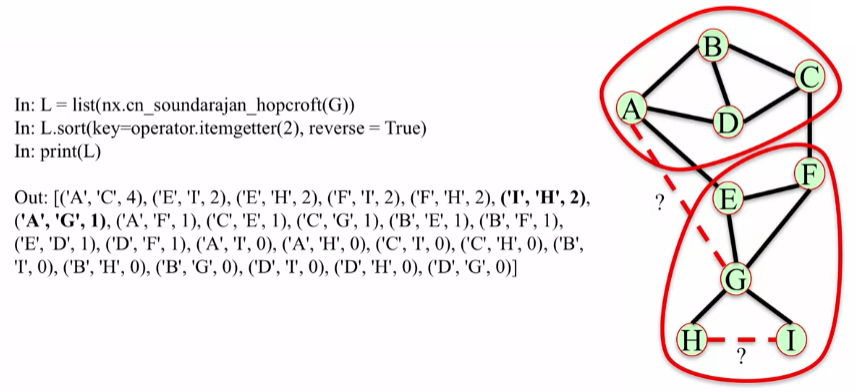




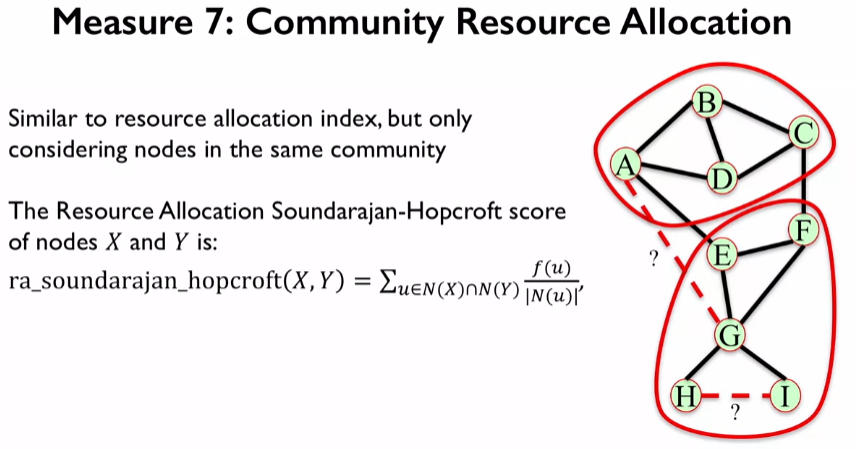


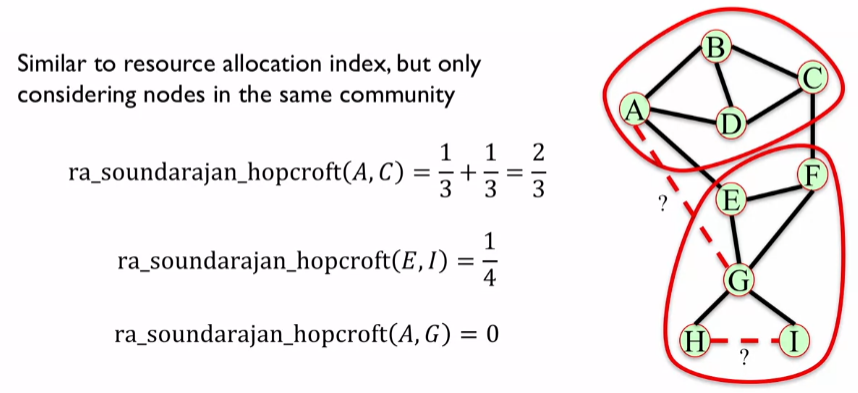
The sum of the intersection of X and Y (f(u)) is a count of the number of common neighbours in the same community, returning 1 if they are, and 0 if they aren’t.

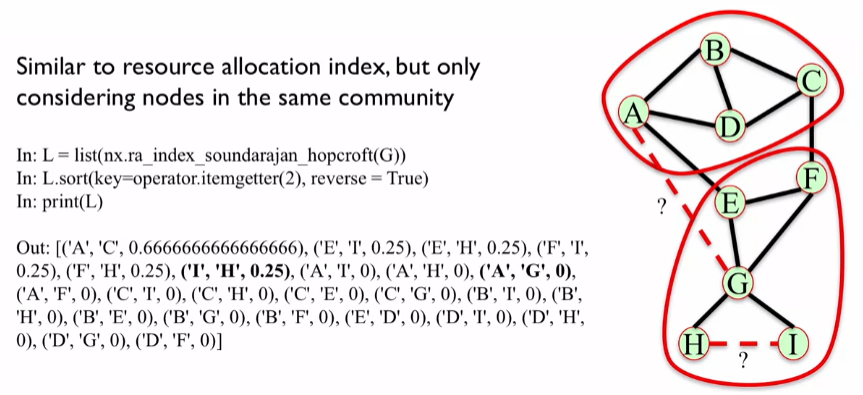




The final measure is similar to the resource allocation measure, but it only considers those who are in the same community. Where f(u) is the same as above.







All of these measures are used to try to predict if a new connection will form between nodes. Different measure can give different scores, so they aren’t consistent at predicting new connections. What you would do, is you would build a dataset with features as the above measures and then use a machine learning algorithm to predict where the connection would form.