

Basic Network Metrics

Notes by Daniel Whitney
January 5, 2008

Introduction

Researchers are constantly seeking metrics to characterize networks (and other complex things) as a way of, or in the hope of, summarizing complex situations and behaviors in a convenient and compact way. As indicated below, this does not always work, and we should not be surprised at this. Interest in structural metrics has been strong in the network research community in part because networks tend to be very large and in need of some kind of summary characterization. Another reason is the hope that structure, captured conveniently in metrics, can be linked to observable or even hidden behaviors.

Interest has been very strong in understanding the internet, by far the largest and potentially most interesting network in existence. The structure of this network has been studied for at least 10 years and different researchers have come up with different proposed structures. There is at present no way to validate these different proposals. Several methods have been used to estimate its degree sequence and most researchers agree that it follows a power law: $p(k) \sim k^{-\gamma}$ where $\gamma \sim 2.5 - 3$. [Barabasi and Albert] Since such a degree distribution implies a few nodes of very high degree, some researchers concluded that these nodes linked preferentially to each other, creating a network of centrally located hubs, which presumably were major servers. However, it has emerged that while structure drives structural metrics, the metrics themselves imply little or nothing about the actual structure, and networks with highly different structure can have the same degree distribution, average nodal degree, and power law coefficient γ . [Li, Alderson, et al]

In these notes, we describe a number of metrics that seek to characterize networks in one way or another. Some of these provide metrics about individual nodes whereas others return a value that applies only to the network as a whole. Some of the simplest metrics are given in the notes called Network Models and Basic Operations. These include the number of nodes, number of edges, and average nodal degree. In the present notes, the metrics relate to structural topics such as how nodes of different degrees tend to link to each other or tend to cluster into relatively tightly connected subregions of the network even while the network as a whole remains connected. Also included in these notes are methods for generating graphs that have specified degree sequences or have degree sequences that are characterized statistically as being drawn from an ensemble with a given probabilistic distribution.

Degree Correlation

Degree correlation is a basic structural metric that calculates the likelihood that nodes link to nodes of similar or dissimilar nodal degree. The former case is called positive degree correlation while the latter is called negative degree correlation. In the social sciences, a network with positive degree correlation is called assortative while one with negative degree correlation is called dis-assortative. Three ways of characterizing the amount of degree correlation are used, each containing less detail and expressing the result in more compact terms. These are the joint degree distribution, the k -nearest neighbors metric, and the Pearson Degree Correlation. These are discussed below in turn.

Joint Degree Distribution

The joint degree distribution examines each pair of connected nodes and notes their respective nodal degrees. Each time a pair has degrees (k_1, k_2) , one is added to the (k_1, k_2) entry of the matrix JDD. If 12 nodes of degree 3, for example, are linked, then there will be a 12 in the $(3,3)$ entry in JDD. If many nodes of similar degree are linked, then JDD will have large entries along and near its diagonal. Conversely, if many nodes of dissimilar degree are linked, then JDD will have large entries on the anti-diagonal.

The text of the JDD routine *JDD1* is below:

```
function [JDD]=JDD1(A,name)
%computes joint degree dist of A
kvA=kvec(A);
kmax=max(kvA);
JDD=zeros(kmax);
N=size(A,1);
M=sum(kvA);
for i = 1:N
    for j=1:N
        if A(i,j)==1
            k1=kvA(i);
            k2=kvA(j);
            JDD(k1,k2)=JDD(k1,k2)+1;
        end
    end
end
JDD=JDD/M;
figure;pcolor(JDD);colorbar
shading interp
xlabel('k1')
ylabel('k2')
title(['Joint Degree Distribution for ',name])
```

Here is code that creates a simple undirected network from a nodelist file, plots it (see result in Figure 1), and calculates its JDD (see result in Figure 2):

```
>> test3=dlmread('testmatrix3.csv')
```

```

test3 =
  1  2  3  5
  2  1  4  5
  3  1  0  0
  4  2  5  0
  5  1  2  4
>> test3adj=adjbuildn(test3)
test3adj =
  0  1  1  0  1
  1  0  0  1  1
  1  0  0  0  0
  0  1  0  0  1
  1  1  0  1  0
>> CNet(test3adj) %CNet plots a network with node 1 at 3 o'clock, nodes numbered
counter-clockwise
>> JDD1(test3adj,'test3adj')
ans =
  0  0  0.0833
  0  0  0.1667
  0.0833  0.1667  0.5000
>>

```

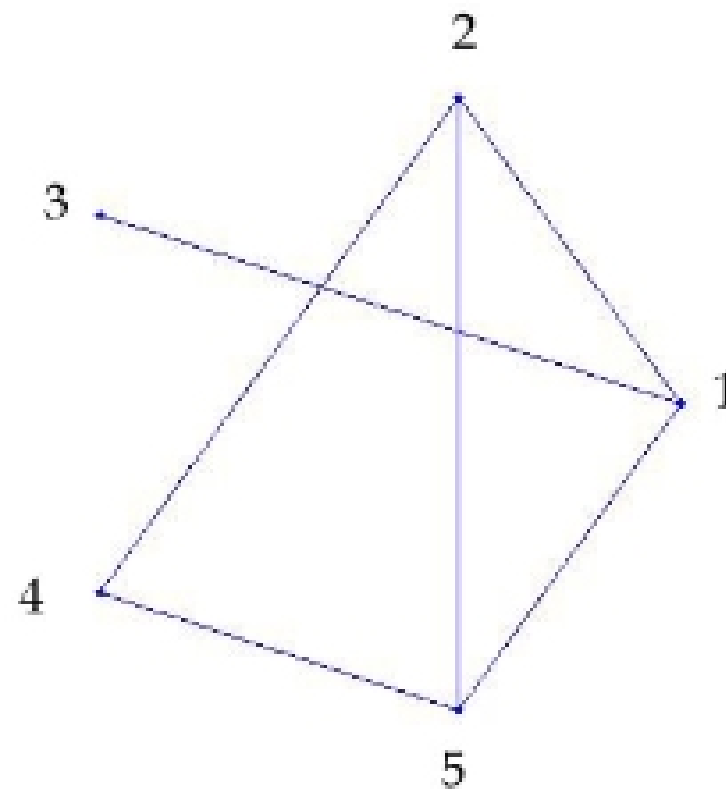


Figure 1. A Simple Undirected Network (test3adj)