COMPLEX NETWORKS, CSYS303

HW01 WRITE-UP

David W. Landay

University of Vermont

Graduate Student, Comp. Systems and Data Sci.

0.1 Problem 1

Record in a table the following basic characteristics:

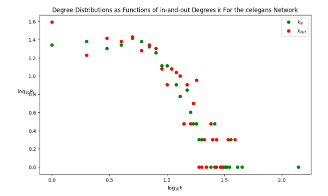
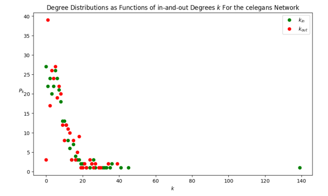
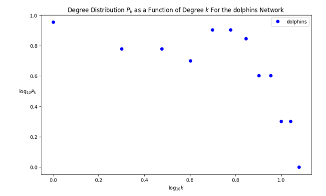
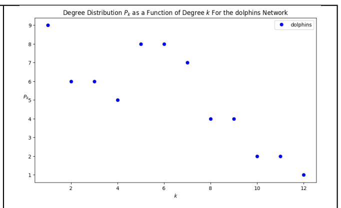
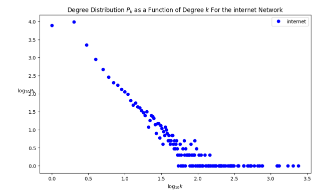
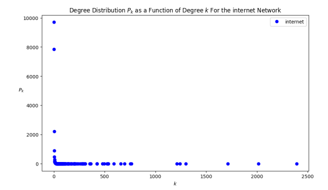
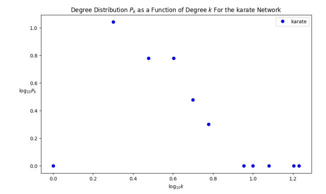
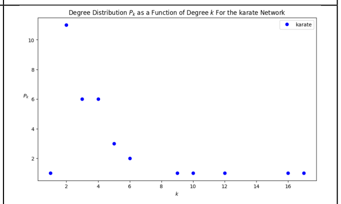
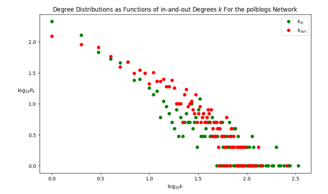
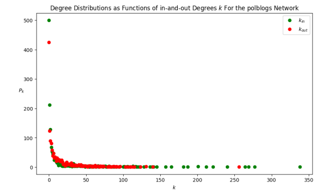
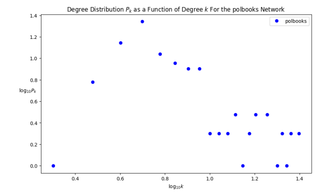
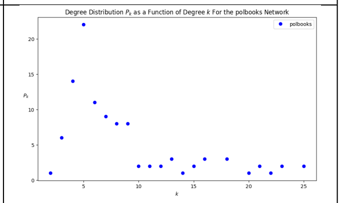
1. N, the number of nodes;
2. m, the total number of links;
3. Whether the network is undirected or directed based on the symmetry of the adjacency matrix;
4. k, the average degree ($k_{in} k_{out}$ if the network is directed);

Name	Directed	N	m	$<k>$	$<k_{in}>$	$<k_{out}>$	k^{max}	k^{min}	k_{in}^{max}	k_{out}^{max}	k_{in}^{min}	k_{out}^{min}
celegans	True	297	2359	--	7.9428	7.9428	--	--	139	39	0	0
dolphins	False	62	159	5.1290	--	--	12	1	--	--	--	--
internet	False	22963	48436	4.2186	--	--	2390	1	--	--	--	--
karate	False	34	78	4.5882	--	--	17	1	--	--	--	--
polblogs	True	1490	19090	--	12.8121	12.8121	--	--	338	256	0	0
polbooks	False	105	441	8.4000	--	--	25	2	--	--	--	--

Figure 1: Summary information of six networks

0.2 Problem 2

(a) Plot the degree distribution P_k as a function of k. In the case that P_k versus k is uninformative, also produce plots that are clarifying. For example, $\log_{10} P_k$ versus $\log_{10} k$. (Note: Always use base 10.)

			
“C. Elegans log-log”	“C. Elegans”	“Dolphins log-log”	“Dolphins”
			
“Internet log-log”	“Internet”	“Karate log-log”	“Karate”
			
“Polblogs log-log”	“Polblogs”	“Polbooks log-log”	“Polbooks”

There are various options for "goodness of fit" tests, which involve hypothesis testing to determine/estimate whether a set of observations was drawn from a particular distribution, or to compare a sample and a reference probability distribution. Such tests may include the Anderson-Darling test, Kolmogorov-Smirnov test, and various others. Performing these tests are likely outside of the scope of this assignment, so we will give our best estimate.

It is difficult to infer what probability distributions the degree distributions of each network compares most to. For example, the "Karate" network is very small; containing only 34 nodes and 78 edges. We cannot confidently rely on the shape of the data to determine what type of distribution best represents our observations. However, by my best estimation and from what I have seen in lectures, side-by-side comparisons of the degree distributions in log space and non-log space of the networks seem to exhibit the following characteristics:

C.Elegans:	Poisson Distribution (possibly log-normal)
Dolphins:	Difficult to determine. Possibly a Log Normal Distribution and another scaling regime
Internet:	Heavy tail indicates a Power-law Distribution
Karate:	Difficult to determine. With more interactions, probably Power-law Distribution
Polblogs:	Heavy tail indicates a Power-law Distribution
Polbooks:	Poisson Distribution

0.3 Problem 3

Measure the clustering C_2 where

$$C_2 = \frac{3 \times \text{ triangles}}{\text{triples}}$$

For directed networks, transform them into undirected ones first. One approach is to compute C_2 as

$$C_2 = \frac{3 \times \frac{1}{6}\mathbf{Tr}(A)^3}{\frac{1}{2}\left(\sum_{ij}\left[A^2\right]_{ij} - \mathbf{Tr}A^2\right)}$$

*Note: avoiding computing A^3 is important and can be done.

We calculate the global C_2 values for each of the six networks. C_2 represents the transitivity of the network; i.e. the proportion of triads that are transitive: if node A directs to node B, and node B directs to node C, then node A directs to node C.

Name	C_2	Time to calculate (s)
C.Elegans	0.18	0.0685
Dolphins	0.31	0.002
Internet	0.011	5.97
Karate	0.255	0.0015
Polblogs	0.226	0.83
Polbooks	0.348	0.007