ESD.342 Spring 2006

Assignment # 3

Due to all three instructors electronically on May 2 at noon

The purpose of this assignment is to become more familiar with a small slice of the network literature and to become more deeply knowledgeable about one other network that has been quantitatively assessed (to variable levels) in the literature. Each of you is assigned **one network from Table II** in Newman's Complex Networks Review Paper. The Table below gives for each student an assigned network based on the terminology Newman uses in his Table II. A modified version of this is given at the end of this assignment and this will also be distributed in class in an upcoming lecture.

Student Name	Network	Student Name	Network
Avnet	Film Actors	Bonnefoy	Company Directors
Bounova	Math co-authorship	Castro	Physics co-authorship
Frank	Biology co-authorship	Hanowsky	Telephone call graph
Lin	Email messages	Lindsey	Email address books
Livengood	Student relationships	Long	WWW nd.edu
Martin	www altavista	Mozdzandowska	Citation network
Nicol	Roget's Thesaurus	Noor	Word co-occurrence
Rayside	Internet	Shah	Power grid
Song	Software packages	Steel	Software classes
Sudarsanam	Electronic circuits	Tapia-Ahumada	Peer-to-peer network
Underwood	Metabolic network	Vaishnav	Protein interactions
Weibel	Marine food web	Wirthlin	Freshwater food web
Yang	Neural network		

I. *Required*

Become familiar with your assigned network. You will need to have access to the referenced sources (as given in Newman's Table II-see below) for each of these networks to answer most parts of this question. Thus, you should first obtain and read the key references (for many systems there are two references in Table II but there have been others published in the interim for most of these systems). **For your assigned network, answer the following:**

- Do a brief literature survey describing the work in the original references and other papers you find through your searching.
- Check the data in the paper(s) to see if it agrees with the data quoted in Table II. This may require you to use some of the Matlab routines provided or to write your own (or to use or UCINET).

- How are the nodes and links actually described for the network? Are precise and reproducible definitions of "node" and "arc" or "edge" given? Are there alternative nodes and link definitions for the same or a closely related network?
- Comment on the categorization in Table II ("technological," "biological," etc.). Is it suitable? Is it sufficient?
- Comment on whether the author(s) of the paper is (are) experts in the domain of the data and underlying system, and whether they explain convincingly that their data or model capture the important aspects of the underlying system or context.
- What system (or network) properties –if any- were the authors most interested in? What (other) properties could be important in the system?
- Is the network a bipartite network? Might the network actually reflect affiliation structure not known or measured by the authors?
- Is there hierarchy in the system the network is attempting to represent? Are there possible layers in the actual system that the network description is missing?
- Is the data used to derive the metrics actually available in the paper or would it have to be taken from plots to develop or test another metric? Does the author have a web site where more data are available or do they answer queries for more detailed data?
- How reliable do you feel the data is? How reliable do you feel the analysis of the data to yield metrics or model tests is? What is your estimate of the "noise" in the estimates?
- If the authors applied a network model to the data, comment on its appropriateness. Suggest a model (not tried by the authors) that might be interesting to test against the network data.
- Does the network of interest exhibit any "anomalies" relative to its metrics as compared to the other networks in Table II? What might be an explanation of this anomaly?
- For what you think is interesting relative to the network you were assigned, justify its potential importance and apply it to the data you have and report what is learned.

II. *Optional*

- Develop a model for each network different than that used by the authors and explore what is learned by exercising that model.

The assignment is not due until May 2 but you are encouraged to begin now to collect the references (starting with Newman's paper) and the data sources. The answer to the required questions above is expected to be brief but thoughtful.

The table below gives the information from Newman's review paper and includes my calculations for random network approximation for path length and clustering. It is shown in landscape mode and includes the references in the last column (from numbering in Newman's paper) where the networks that have been assigned are further described.

	Network	Type	n	m	$\langle k \rangle$	l	$\frac{\log n}{\log \langle k \rangle}$	α	$C^{(1)}$	$C^{2)}$	$\langle k \rangle / n$	r	Ref (s)
Social	film actors	undirected	449 913	25 516 482	113.4	3.48	2.75	2.3	0.20	0.78	.00025	0.208	20, 416
	company directors	undirected	7 673	55 392	14.44	4.60	3.35	-	0.59	0.88	.002	0.276	105, 323
	math coauthorship	undirected	253 339	496 489	3.92	7.57	9.1	-	0.15	0.34	.000018	0.120	107, 182
	physics coauthorship	undirected	52 909	245 300	9.27	6.19	4.9	-	0.45	0.56	.00017	0.363	311, 313
	biology coauthorship	undirected	1 520 251	11 803 064	15.53	4.92	5.2	-	0.088	0.60	.00001	0.127	311, 313
	telephone call graph	undirected	47 000 000	80 000 000	3.16	-	15.4	2.1	-		7×10^{-8}	-	8, 9
	email messages	directed	59 912	86 300	1.44	4.95	30.5	1.5/2.0	-	0.16	.000025	-	136
	email address books	directed	16 881	57 029	3.38	5.22	8.0	-	0.17	0.13	.0002	0.092	321
	student relationships	undirected	573	477	1.66	16.01	12.2	-	0.005	0.001	.003	−0.029	45
	sexual contacts	undirected	2 810	-	-	-	-	3.2	-	-	-	-	265, 266
Information	WWW `nd.edu`	directed	269 504	1 497 135	5.55	11.27	7.29	2.1/2.4	0.11	0.29	.00002	−0.067	14, 34
	WWW Altavisa	directed	203 549 046	2 130 000 000	10.46	16.18	4.32	2.1/2.7	-	-	5×10^{-8}	-	74
	Citation network	directed	783 339	6 716 198	8.57	-	7.86	3.0/-	-	-	.00001	-	351
	Roget's Thesaurus	directed	1 022	5 103	4.99	4.87	4.31	-	0.13	0.15	.005	0.157	244
	Word co-occurrence	undirected	460 902	17 000 000	70.13	-	3.07	2.7	-	0.44	.00014	-	119, 157
Technological	Internet	undirected	10 697	31 992	5.98	3.31	5.19	2.5	0.035	0.39	.0005	−0.189	86, 148
	power grid	undirected	4 941	6 594	2.67	18.99	8.67	-	0.10	0.080	.0004	−0.003	416
	Train routes	undirected	587	19 603	66.79	2.16	1.52	-		0.69	.12	−0.033	366
	software packages	directed	1 439	1 723	1.20	2.42	39.81	1.6/1.4	0.070	0.082	.0008	−0.016	318
	software classes	directed	1 377	2 213	1.61	1.51	15.18	-	0.033	0.012	.0012	−0.119	395
	electronic circuits	undirected	24 097	53 248	4.34	11.05	6.87	3.0	0.010	0.030	.0002	−0.154	155
	peer-to-peer network	undirected	880	1 296	1.47	4.28	17.6	2.1	0.012	0.011	.002	−0.366	6, 354
Biological	metabolic network	undirected	765	3 686	9.64	2.56	2.9	2.2	0.090	0.67	.012	−0.240	214
	protein interactions	undirected	2 115	2 240	2.12	6.80	10.19	2.4	0.072	0.071	.001	−0.156	212
	marine food web	directed	135	598	4.43	2.05	3.30	-	0.16	0.23	.03	−0.263	204
	freshwater food web	directed	92	997	10.84	1.90	1.40	-	0.20	0.087	.12	−0.326	272
	neural network	directed	307	2 359	7.68	3.97	2.81	-	0.18	0.28	.025	−0.226	416, 421