Fall 2017

# ECE 18-755: Networks in the Real World

## HW1 Assigned 09/19/17 DUE 10/03/17 (midnight - 11:59:59pm ET)

## Problem 1 (15 pts)

a) Consider 10 isolated nodes. For each pair of nodes, throw a (fair) die and connect them if the number on the die is 1. Describe the graph you obtain. Is it connected or not? What is the average degree and the degree distribution? Are there any cycles?

Now repeat the experiment and connect nodes if the number on the die is 1 or 2. How is the graph different from the previous case?

b) Install the following software tool: Gephi: http://gephi.org/. Use it to draw and analyze the graphs you obtained in part a) (*i.e.*, report the number of unreachable pairs of nodes, average distance among reachable pairs of nodes, network diameter, clustering coefficient, etc). Provide a screenshot of your work.

c) Now build and analyze a few real world networks. More specifically, we look into how microstructures of different materials can affect the network features of crack lines and compare how these propagate through different materials.

Using the images[1] below, model the fractures of the wood, asphalt, and glass as networks in Gephi using the Edge and Node Pencil tools. To this end, you should model material fractures as a network, *i.e.*, a node represents the intersection of two or more cracks, while a link is represents the crack itself.

After your build the network based on these images, you need to determine various network properties (*e.g.*, average degree, path length, connected components, and clustering coefficient), compare them across materials, and provide possible explanations as to how these network properties may or may not reflect the material mechanical properties (*e.g.*, material plasticity, brittleness, stiffness, *etc.*).

wood

asphalt

glass

1. To build the Gephi network, you should use the higher quality JPG images available under the HW1 folder.

## Problem 2 (10 pts)

Consider the network of loans (as directed arcs) between 16 Renaissance Florentine families given in the text file *Renaissance.net*[2].

a) Using Gephi plot the corresponding graph of the file *Renaissance.net*. What is the *IN* and *OUT* degree distribution of the network?

b) What is the average distance between the nodes of the largest strongly connected component? How does this compare to the network diameter?

c) What are the most important nodes in terms of total degree and betweenness? Illustrate one case where these two statistics are different and explain why.

## Problem 3 (20 pts)

The file "arXiv_lcc.gml" is an undirected network of scientific collaborations between authors who submitted to General Relativity and Quantum Cosmology. Use the Gephi tool to:

a) Visualize and calculate the average degree, average path length, diameter, and average clustering coefficient of this network.

b) Plot the degree distribution using both probability density function (PDF) and rank frequency plot (both in log-log scale).

c) Generate 5000 exponentially-distributed numbers with a mean identical to the average degree. Plot the rank frequency plot in log-log scale and contrast it to the result in b). What kind of network is this collaboration network? Justify your answer.

## Problem 4 (25 pts)

Based on the concepts of characteristic path length and clustering coefficient learned in class, study the impact of increased randomness in the network structure using a regular lattice with $N = 100$ nodes and connectivity $k = 4$ (*i.e.*, each node has 4 directed links to its neighbors as described below). More precisely, to introduce some randomness into the regular lattice proceed as follows:

- Consider first a regular lattice where nodes are distributed uniformly over a circle; each node is directly connected to its first and second immediate neighbors (that is, $k = 4$ as discussed in class). Note, that the circle is not divided into 16 parts as in the lecture notes, but instead in 100 parts.
- For each node and the edge connecting it to its first and second order neighbors (clockwise), re-wire this edge with a probability $p$ to a new node chosen uniformly at random over the entire circle. If the two nodes are already connected, then leave the edge as it was. Stop iterating when each node was considered exactly once.

---

2. The *Renaissance.net* file contains a network of 16 nodes and 38 directed arcs.

Using a logarithmic scale, vary the probabilities $p$ between 0.0001 and 1 and plot on the same graph the normalized values of the characteristic path length and clustering coefficient (that is $L(p)/L(0)$ and $C(p)/C(0)$ similar to what we discussed in class). Comment on the results you obtain.

## Problem 5 (30 pts)

Based on what we have learned in class, study the characteristics of an Erdos-Renyi (ER) graph and a Scale-Free (SF) network generated as described below.

**Scale-free:** Given $N$ = 1000 nodes and a preferential attachment mechanism for linking nodes construct a SF graph as follows:

- Initially, choose $m$ = 5 nodes and randomly connect them. Make sure that you do not add self-loops and multiple edges between two nodes.
- At each iteration, pick a node from the remaining ($N$ - $m$) nodes and connect the new node to at most $m$ nodes with a certain probability. More precisely, the probability that the new node $i$ will be connected to a node $j$ is proportional with its degree $k_j$; this probability can be written as $P(i \to j) \sim k_j \;/\; \sum_j k_j$. Note that we consider only undirected links so if $i$ and $j$ are connected, then both entries $A(i,j)$ and $A(j,i)$ in the adjacency[3] matrix are set to 1. Make sure you do not add self-loops to the generated graph.
- Continue this process until all the initial nodes are linked at least once to the growing network.

**Erdos-Renyi:** Given $N$ = 1000 isolated nodes, a probability $p$ = 0.05 of connecting two randomly selected nodes, and a maximum number of edges equal to the number of edges in the SF graph, generate the ER graph as follows:

- At each iteration pick two random nodes and if they are not already connected, then link them with probability $p$.
- Continue this process until the maximum number of iterations is reached.

Save the two adjacency matrices into "ER.txt" and "SF.txt" files, respectively. Answer the following questions:

a) Compute the degree distribution of the two networks and explain their properties (*e.g.*, maximum and minimum degree, existence of hubs).

b) Estimate and plot the empirical PDF of the degree distributions for the two networks. Use any of the software tools to display the two networks.

c) Compute the clustering coefficient for the two networks.

3. The adjacency matrix $A$ of a finite undirected graph of $N$ vertices and $E$ edges is a matrix of size $N$x$N$, where each row corresponds to a distinct node and the non-zero entries in a row denote an edge between two nodes. For instance, if $A(i,j)$=1 then it means that there is an undirected edge between nodes $i$ and $j$. Note that we do not allow self-loops in these graphs so the diagonal elements of $A(i,i)$ = 0. Moreover, the sum of all the elements in the upper triangular part of the adjacency matrix $A$ should equal the total number of edges $E$ in the graph.

d) For both graphs remove now 2% of the nodes with the highest degree. Recompute the clustering coefficient for both graphs and compare the results with the previously computed coefficients. Explain what do you observe.

## Note on submission:

***** For all problems in this homework, everything is handled electronically. Prepare the answers using either Word or Latex and ***create a PDF file for submission***. Also, put the source code and all related files for each problem (as needed) in a separate folder. Finally, compress everything and name it as "yourAndrewId_hw1.zip", and deposit it on the Canvas under Assignments > Homeworks Fall 2017 > HW1 Submission.

***** Work *individually* on all problems.

***Good luck!***