1. (20 pts) In this problem we consider a personalized version of HITS for product recommendation. Given a set of products and a set of consumers, we set $L_{uv} = 1$ if consumer $u$ has bought product $v$ and $L_{uv} = 0$ otherwise. Let $h_u$ be the hub score for consumer $u$ and $a_v$ be the authority score for product $v$. Note that we do not associate hub scores with products or authority scores with consumers. For $\epsilon \in [0, 1]$, the personalized HITS equations for consumer $i$ are:

$$h_u = (1 - \epsilon) \sum_v L_{uv} a_v + b_u,$$

where $b_u = \epsilon$ if $u = i$ and 0 otherwise; and

$$a_v = \sum_u L_{uv} h_u.$$

We have the following data for consumers $\{c_1, c_2, c_3, c_4, c_5, c_6\}$ and products $\{p_1, p_2, p_3, p_4, p_5, p_6\}$:

- Consumer $c_1$ has bought products $p_1$ and $p_2$.
- Consumer $c_2$ has bought products $p_1$, $p_2$ and $p_3$.
- Consumer $c_3$ has bought products $p_5$ and $p_6$.
- Consumer $c_4$ has bought products $p_4$ and $p_6$.
- Consumer $c_5$ has bought products $p_3$, $p_5$ and $p_6$.
- Consumer $c_6$ has bought product $p_6$.

Compute product recommendations for consumer $c_1$ for $\epsilon = 0.1$ and $\epsilon = 0.6$. In particular, compute the hub scores of consumers and the authority scores of products. Which product would be recommended to $c_1$ in each case? Explain.

Hint: Use matlab, Excel or any other program to compute the solution iteratively. Remember to normalize after each step.

2. (10 pts) We have seen in class how PageRank can be gamed and now consider how HITS can be gamed. Suppose HITS is being used by a collaborative recommendation web-site. Products play the role of authorities, and users of the system play the role of hubs. There is a link from a user to a product if the user recommends that product. One of the users is actually a pseudonym for the manufacturer of one of the products. If this user is allowed to place exactly $k$ product recommendations, intuitively explain which products he should recommend in order to increase the authority score of his product. Assume that he can not post more than one recommendation for a particular product.

3. One increasingly important use of network models is to determine how a seed investment in advertising will help brand awareness propagate through a network.
Consider the following simple scenario: You are given a social network with $N$ individuals. A company invests $X$ in educating one person, say $v_1$, about its product. This person then informs exactly one more person, $v_2$, chosen uniformly at random from the remaining $N - 1$ individuals, about the product. The newly educated person $v_2$ then informs one more person $v_3$, chosen uniformly at random from the remaining $N - 1$ individuals (i.e. other than $v_2$), about the product, and so on, forming a chain. The process terminates when the chain revisits an already informed individual. For example, if $v_3 = v_1$, the process terminates and only two individuals, $v_1$ and $v_2$ get informed; in this case we will say that the length of the chain is 2.

(a) With what probability is the length of the chain at least $k$? [5 pts]

(b) Assume that the expected length of the chain, $l$, can be modeled by the equation: $l(N) = aN^e$, for some constants $a$ and $e$. Estimate the exponent $e$. Use a software program, Excel, or your own code to estimate the expected length of a chain for several different values of $N$; or download off the website data that has already been generated. Then fit this data to the model equation to determine the constant $e$. [10 pts]

(c) BONUS [10 pts] Find or approximate $e$ mathematically.

(d) Comment on the power of viral marketing. [5 pts]

4. Concisely explain the winners’ curse and the optimizers’ curse. Make the connection between them and uncertainty. [10 pts]

5. Explain in an essay the business model of Skype. Relate the business model to the technical design. Use no more than 500 words total. [20 pts]