Question 1 (Chapter 13: Structure of the Web)

a) Is the graph in Figure 1 strongly connected?

b) Identify two edges which if deleted make the size of the SCC five. Calculate the IN, OUT and tendrils in the resulting graph.

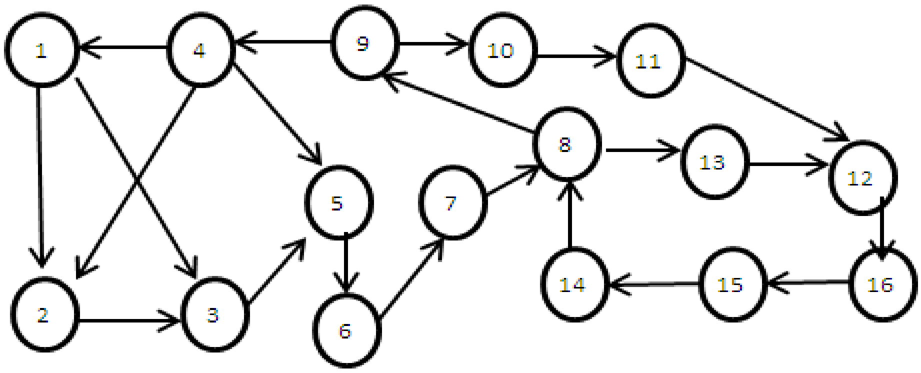

Figure 1

Question 2 (Chapter 14: PageRank)

a) Which of the following is most effective for increasing the PageRank score of a page:
   1. adding an inlink;
   2. adding an outlink;
   3. deleting an inlink; and
   4. deleting an outlink?

b) Which is most likely going to decrease the PageRank score of the page?

c) Compute the pagerank for the graph in Figure 2.

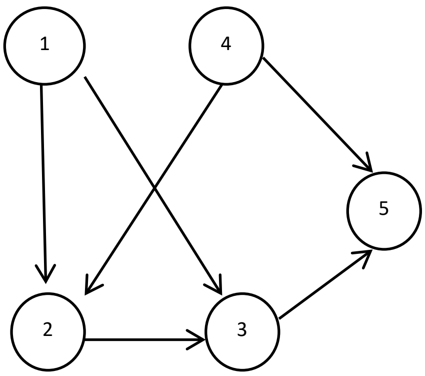

Figure 2

Question 3 (Chapter 16, Information cascades)

Bayes rules states:

$$P(G|S) = \frac{P(G)P(S|G)}{P(S)}$$

where we often interpret S as a *signal* and G as an actual event in the real world

(a) What is the importance of Bayes' rule in terms of translating experimental results (lab-experiments) to the real world situation?

(b) Let $\bar{G}$ denotes the complement of G. Show, using the definition of conditional probability that Bayes' rule can also be written as:

$$P(G|S) = \frac{P(G)P(S|G)}{P(S|G)P(G) + P(S|\bar{G})P(\bar{G})}$$

(c) Show that the above formula implies that if $P(S \mid \overline{G}) = P(S \mid G)$ then $P(G|S) = P(G)$, and give an interpretation.

(d) Assume there are two bowls of chips. One with 8 yellow and 4 red chips (called Majority Red) and one with 8 yellow and 4 red chips (Majority Yellow). Suppose two students have drawn a chip and announced 'Majority Yellow'. Now the third student comes and draws a red chip. What will he or she announce? Support your answer with a calculation using Bayes' rule.

Answer

(a) Bayes rule translates $P(S|G)$ to $P(G|S)$. S can be interpreted as a signal. G is the 'real world'. The meaning is that $P(S|G)$ can rather easily be measured in a lab situation/controlled environment, and that $P(G|S)$ is the information on the real world that you really want to have.

(b) We have to show $P(S) = P(S|G)P(G) + P(S|\bar{G})P(\bar{G})$. Conditional probability:

$$P(G|S) = \frac{P(G \cap S))}{P(S)}$$

$S = (S \cap G) \cup (S \cap \bar{G})$ (0.5 point). These sets are disjoint. So $P(S) = P(S \cap G) + P(S \cap \bar{G})$. By conditional probability $P(S \cap G) = P(S)P(G|S)$ (0.5). Likewise $P(S \cap \bar{G}) = P(S)P(\bar{G}|S)$. This finishes the proof.

(c) Suppose $P(S|\bar{G}) = P(S|G)$. The denominator can then be written as $P(S|G)(P(G) + P(\bar{G})) = P(S|G)\,.\,1$. Now, we can divide out the $P(S|G)$. This leads to $P(G|S) = P(G)$. Interpretation: if a signal occurs equally likely under G or under $\bar{G}$, then the fact that a signal occurs does not have any influence on G. It does not add/take-away information.

(d) Suppose $P(S|\bar{G}) = 0$, then the denominator becomes $P(S|G)P(G)$ which is equal to the numerator. This shows $P(G|S) = 1$. Interpretation: if a signal S never occurs when $\bar{G}$ holds, then when it occurs it must mean that G holds.

(e) We have situation of the Herding Experiment. Let G = 'reality is majority yellow', H = 'drawing of yellow chip', and L = 'drawing a red chip'. Then $P(G) = ½$, $P(H|G) = P(L|\bar{G}) = 2/3$, and $P(L|G) = P(H|\bar{G}) = 1/3$. From the announcements of the first two students, the third

student can derive that both of them drew a yellow chip. Hence, the information that the third student has is (H,H,L). Let $S = (H,H,L)$. If $P(G|S) > ½$, then the third student will announce 'majority yellow' (despite that he observes a red chip), otherwise he will follow his own private signal (L) and announce 'majority red'. $P(G|S)$ can be calculated by applying Bayes rule. Numerator: $P(G).P(S|G) = 2/27$. Denominator: $P(S) = P(S|G).P(G) + P(S|\bar{G}).P(\bar{G}) = 2/27 + 1/27 = 3/27$. Therefore $P(G|S) = 2/3 > ½$ and the third student will announce 'majority yellow'.
Note: $P(S|G) = P((H,H,L)|G) = P(H|G).P(H|G).P(L|G)$. Similar $P(S|\bar{G})$.

Question 4 (Chapter 17, Network effects)

Consider the economic model in Chapter 17 where consumers occupy the interval [0,1] and their reservation price without network effects is given by r(x) . The function f(z) measures the benefit to each customer from having a fraction z of the population using the good. When proportion z of the population buys the product then the new reservation price of customer x with network effects is

$p(x) = r(x)\, f(z).$

Assume that the price p* is strictly between 0 and 1. When everyone expects an audience size of z, the fraction of people who actually use the product is:

$\hat{z} = g(z)$, when the condition $p^*/f(z) \le r(0)$ holds.

(a) Explain why the condition $p^*/f(z) \le r(0)$ must hold.
(b) Let $r(x)=1-x$, and $f(z) = 1 + 4z^2$, assume p = 0.95. Suppose initially nobody buys the product. What will be the fraction of the population that buys the product after the first time step?
(c) The graph is shown in Figure 3 for $0 \le z \le 2$. What is the nature of the first equilibrium at about z=0.065 and why?
(d) Read from the graph how the price should be adapted (higher, lower, same) to remove this equilibrium, and approximately how much.

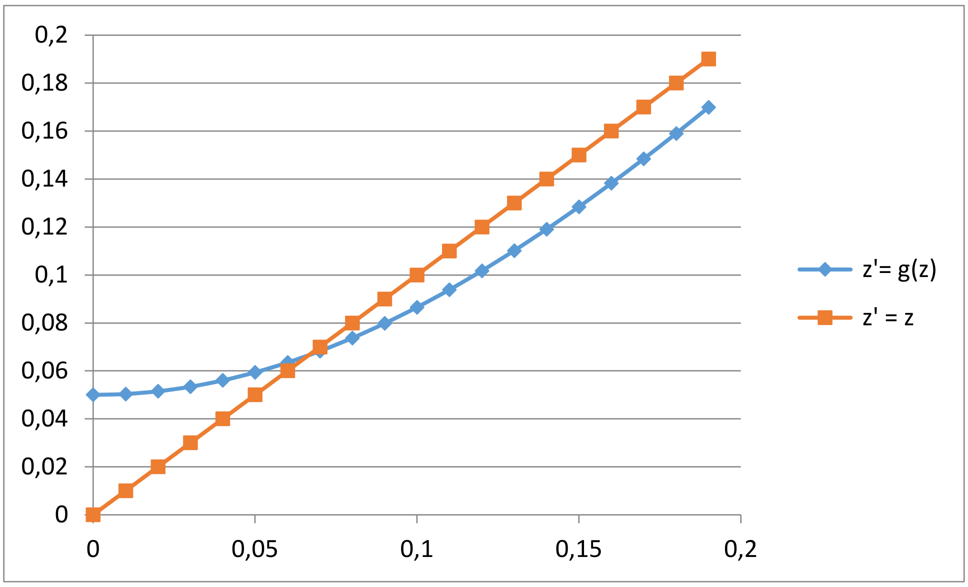

Figure 3

Answer

(a) Suppose p* / f(z) > r(0), combined with the fact that r(x) is decreasing this means that p* / f(z) > r(x) for all x ϵ [0, 1]. This means that there are no customers that have a reservation price that makes them want to buy the product, hence there is nobody that will buy the product.

(b) $g(z) = 1 – (p/(1+4z^2))$. So with p = 0.95 it follows that g(z) = 0.05. This means that after the first timestep 0.05 of the population uses the product. Other way: g(z) can be calculated from $p(x) = (1 - x)(1 + 4z^2)$, where p(x) denotes the reservation price of customer x. Starting with z = 0 it follows we get an equilibrium when 1 - x = p* = 0.95, so x = 0.05.

(c) It is a stable equilibrium. If we start below it we will converge to the equilibrium value. If we start above it, we will also converge to the equilibrium value. This can be seen from the figure. E.g., start at z =0.15 gives a value of around 0.12. Next step we start at 0.12, and will move closer the equilibrium value.

(d) If we could move the graph approximately 0.03 upwards then this equilibrium would no longer exist. This means the price should be made lower with about 0.03.

Question 5 (Chapter 18, Preferential attachment)

(a) Describe the preferential attachment model of Barabasi-Albert.

(b) Figure 4 shows a plot of the logarithm of the degrees against the logarithm for the number of nodes (with a given degree) for a simulation of the Barabasi-Albert model. We see the degree distribution is close to a straight line. This means there is a power law. Why ?

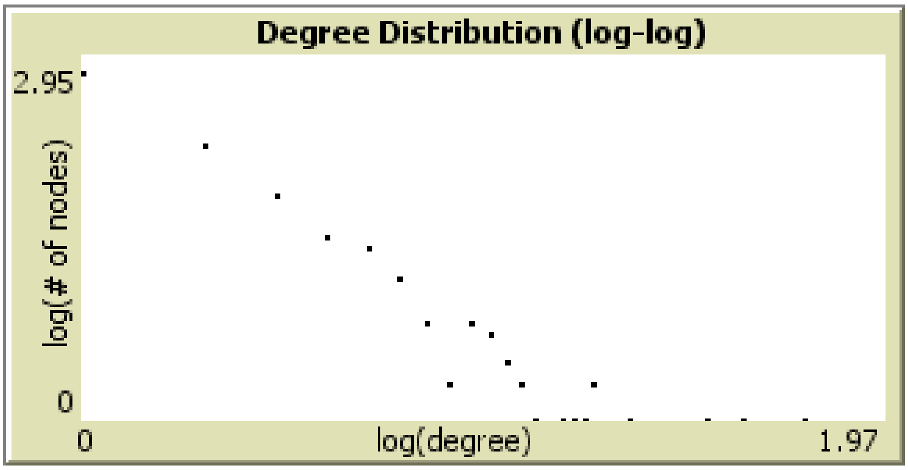

Figure 4

(c) The BA model leads to the following equation for k>1:

$$N(k, t+1) = \frac{k-1}{2t} N(k-1, t) + (1 - \frac{k}{2t}) N(k, t)$$

Explain the factors before the terms N(k-1,t) and N(k,t), where k denotes the degree and N(k,t) denotes the long-term average number of nodes with degree k at time t.

(d) How does the model look for k=1 ?

(e) Let p_k denote the long-term fraction of the nodes with degree k. Show that the following recursive relation can be derived:

$$p_k = \frac{k-1}{k+2} p_{k-1}$$

Show that the equation below is a solution to this recursive relation, and discuss the relation with power laws:

$$p_k = \frac{4}{k(k+1)(k+2)}$$

Answer

(a) The only parameter of the Barabarasi-Albert model is the *m* (Answer formulated for only m = 1 is correct). According to the description in their paper, the initial graph is arbitrary, but for definiteness we assume $V_1 = \{v_1, v_2\}$ linked with m parallel edges. At time step t we add v_{t+1} to $\{v_1, \ldots, v_t\}$ by adding m edges independently, where the probability of adding a single edge to $u \in \{ v_1, \ldots, v_t \}$ is given by:

$p(v_{t+1}$ connects to $u) = p_u = d(u) / \sum_{v \in V(G_t)} d(v) = d(u)/2mt$,

where $d(u)$ is the degree of node u. So at any time t we have that $G_t(m)$ has a total degree of $2mt$ (so mt edges), and $t + 1$ vertices.

(b) We have a straight line on a log-log plot. This means ln('nr nodes') = $-c$. ln('degree') + b, for some appropriate b and c. Let y = 'nr nodes' and x = 'degree'.
Then $\ln(x) = -c.\ln(x) + b = \log(x^{-c}) + \log(e^b) = \ln(x^{-c}.e^b)$. Exponentiating, and writing $B = e^b$, we get $x = B.x^{-c}$.

(c) Here, the first term represents the edges of degree k-1 getting an extra connection, with probability $(k-1)/2t$. Further, the second term represents the edges of degree k that are not getting an extra connection, with probability $1 - (k/2t)$.

(d) Consider the case k = 1. We find: $N(1, t+1) = 0 + (1 - 1/2t)N(1,t) + 1$. This is the same as the model above, except for the fact that at each time step a node of degree 1 is added.
Take the equation for k-1:

$$p_{k-1} = \frac{4}{(k-1)(k)(k+1)}$$

Now, if we p_k by p_{k-1} using these equations, the k and k+1 cancel out, and we find

$$p_k/p_{k-1} = \frac{k-1}{k+2}$$

as required. Relation with power laws:

$$p_k = \frac{4}{k(k+1)(k+2)}$$

$$\approx 4k^{-3}$$

Question 6 (Chapter 19, Cascading)

Consider a networked coordination game on the infinite grid as depicted in Figure 5, in which each node as choice between two possible behaviours, A and B.

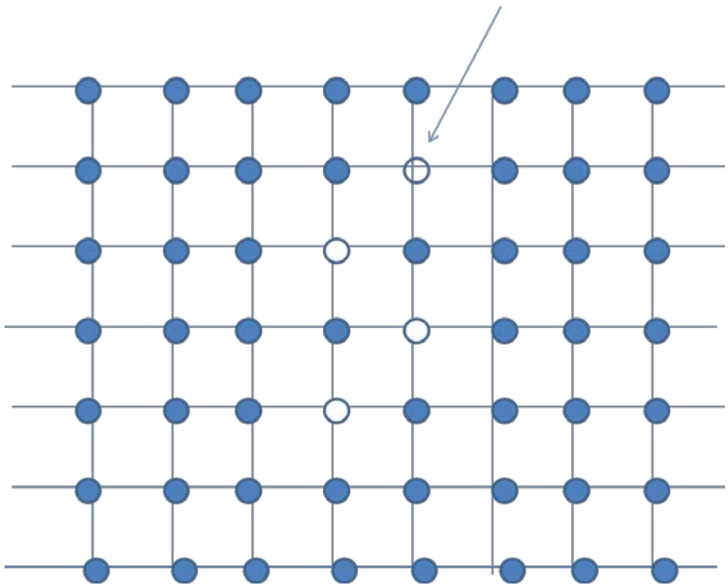

Figure 5

Both two nodes connected by an edge both receive a payoff of a if they both select A, a payoff of 1 if they both select B, and a payoff of 0 otherwise:

	A	**B**
A	*a,a*	0,0
B	0,0	1,1

The set of initial adopters of *A* is indicated by the open circles in Figure 1. We are interested in the question whether each node will adopt A at some point in time.

(a) Let a=5. Calculate the threshold q so that a node v will adopt *A* if at least of fraction q of v's neighbors has adopted a. Motivate your answer.

(b) Let a_0 be so that for $a > a_0$ everyone in the network adopts *A* at some point in time. What is the smallest value of a_0 given this set of initial adopters. Motivate your answer.

(c) Show that the cascade capacity of this grid is ¼. The cascade capacity is the largest value for the threshold q for which some finite set of early adopters can cause a complete cascade.

(d) The book states "the fact that the cascade capacity of the grid is ¼ means that when q is between ¼ and ½, *A* is the better technology, but the structure of the network makes *B* so heavily entrenched that no finite set of initial adopters can cause *A* to win. Explain this quote.

(e) Now a is lowered to 1/3 so that q =3/4. Let *S* be an *arbitrary* set of 4 initial adopters of *A*. After how many steps will the process come to an end.

Answer

(a) $q = b/(a+b) = 1/(5+1) = 1/6$.

(b) $a_0 = 3$ is the smallest value. If we would take a_0 smaller, then the cascade would not spread when a node has only a single edge into the set of adopters of *A*.

(c) Assume a finite set of initial adopters. Without loss of generality we may assume these are contained in some square *X* of *x* by *x*. This means that there are nodes incident to only one node of the square, and their remaining neighbors outside the square. Now if the threshold would be higher than ¼, then this node would not adopt the behavior. On the other hand, if threshold would be lower, then it will accept the behavior.

(d) A threshold of ¼ means that $b/(a+b)$ must be ¼. This means that a has three times more value (payoff) than b in order for *A* to spread. This can be interpreted that *A* has to be three times 'better' than *B* to get accepted. If a and b would be equal then the threshold would be ½. This discrepancy shows that the network hinders innovation, in order to get accepted A should be three times better than *B*.

(e) For $q > 1/2$ the size of the interface decreases at least 1 with every time step. We start with 4 nodes, and an interface size of at most 16. This means that at most 16 more nodes can adopt *A* (in total 20 nodes) . Most people looked at configurations where the cascade can continue for at least a few steps, concluding that the configuration depicted in Figure 5 works best. This gives a much better bound of 2 or 3. Such answers have been judged as correct.