Web Science (201500025) Exam Part 1

December 10, 2018, 13:45-16:45 hrs, Therm

Please pay attention to the following:

- This is a closed book exam: it is not allowed to use the book or any other material.
- Use of calculators, mobile phones etc. is also not allowed.
- Please write your name and student ID on your solution pages.
- Each question mentions the related chapter/topic from the book.
- There are 7 regular questions and one bonus question. The calculation of the partial exam grade G is as follows:
  if $R + B < 51$: $G = 1 + 9(R + B)/51$
  otherwise: $G = 10$
  where
  - R is the number of points you get from answering the regular questions;
  - B is the number of points you get from answering the bonus question.

Success!

Question 1 (Chapters 3+4: Strong and Weak Ties) – 5 x 2 = 10 points

a) In the social network depicted in Figure 1, in which edge is labeled as strong (s) or a weak (w), indicate for nodes {2, 4, 7} whether it satisfies the STC (Strong Triadic Closure) property. Explain your answer.

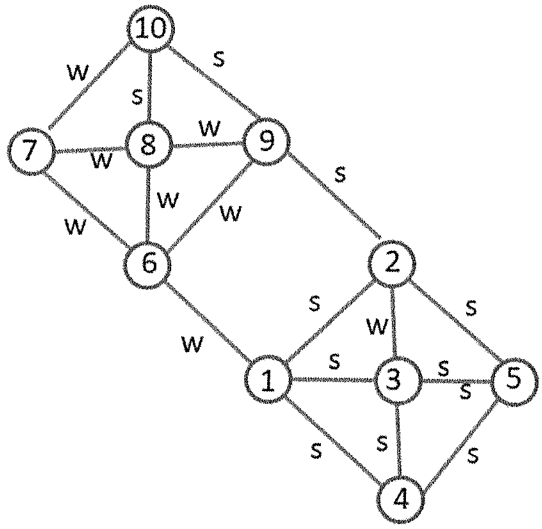

Figure 1

b) Give the definition of *betweenness* of an edge *e*. Using this definition, calculate the betweenness of edge {2, 9}. (Hint: you do not have to use the Girvan-Newman method.)

c) Draw the breadth-first decomposition of the network, starting at node 4.

d) Determine the number of shortest paths from node 4 to each of the other nodes.

e) The Girvan-Newman method assumes one unit of flow between each pair of nodes. Consider the flow *F* between node 4 and node 10. Determine how much of flow F is carried by the edges {2, 9}, {6, 9}, {4, 10}, {8, 10}, {7, 10}.

Question 2 (Chapter 13: Structure of the Web) – 2 x 2 = 4 points

a) Which of the following actions are most effective for increasing the PageRank score of a page and why?

   i. adding an inlink;
   ii. adding an outlink;
   iii. deleting an inlink;
   iv. deleting an outlink.

b) Which of the 4 actions above is most likely going to decrease the PageRank score of the page and why?

Question 3 (Chapter 14: Link Analysis and Web Search) – 3 points

Compute the page-rank for the web-graph depicted in Figure 2 using the PageRank algorithm with spider-trap avoidance and β=0.45. Indicate in a few lines if you think your answer is intuitive or not?

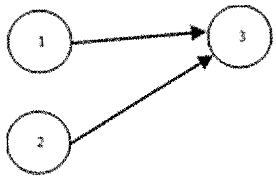

Figure 2

Question 4 (Chapter 14: Link Analysis and Web Search) – 4 points

Apply the HITS page-ranking algorithm to the Web-graph in Figure 3 . Compute manually the first 3 iterations of the algorithm. Can you predict what the page ranks will be, in the limit? What are the ordinal ranks (in other words, the order in which the user is shown these pages) at the end of your computation?

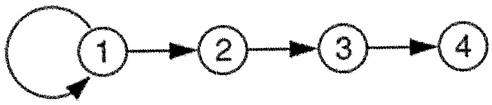

Figure 3

Question 5 (Chapter 16: Information Cascades) – 5 x 2 = 10 points

In the setting of Section 16, consider two urns, each containing 10 marbles: urn A contains k red marbles ($k \in \{1, \ldots, 4\}$) and 10 - k blue marbles, and urn B containing k blue marbles and 10 - k red marbles. Urn A is called 'Majority Blue' (MB), Urn B is called 'Majority Red' (MR). With probability 1/2 either Urn A or Urn B is selected (it is unknown which one). Once the urn is fixed, three rational people draw a marble and announce whether they think the urn is MB or MR.

a) Formulate Bayes' rule, and provide a short discussion relating it to observed signals and events in the real world.
b) Suppose you're the first person to draw, and you draw a red marble. Given this fact, what is the probability that the urn is MR?
c) Suppose you're the third person to draw a marble. Your predecessors have announced MB, MB. Then you draw a red marble. What will you announce and why?
d) Express the probability that an *incorrect* cascade emerges, for $k \in \{1, \ldots, 4\}$ as function of k. For which value of k is it maximized?
e) Express the probability that a cascade emerges (correct or incorrect), for $k \in \{1, \ldots, 4\}$ as function of k. For which value of k is it maximized, respectively minimized? Briefly discuss the difference or similarity in the answers of (d) and (e).

Question 6 (Chapter 17: Network Effects) – 4 x 2 = 8 points

Consider the economic model in Chapter 17 where consumers occupy the interval [0, 1] and their reservation price without network effects is given by $r(x) = 1 - x$.

a) First consider the situation without network effects. Let $1 \geq p^* > 0$. Give the function indicating the fraction of customers that buys the product for given p^*. Explain why this point is a stable market equilibrium.

b) Now consider the situation *with* network effects. The function $f(z)$ measures the benefit to each customer from having a fraction z of the population buying the good. So when proportion z of the population buys the product, then the new reservation price of customer x with network effects is $r(x)f(z)$. Given a price p^*, when the shared expectation is $z \geq 0$, the fraction of people $\hat{z}$ that will actually buy the product can be denoted as : $\hat{z} = g(z)$. Using $r(x) = 1 - x$, provide an expression for $g(z)$, based on z, $f(z)$ and p^* for all values of z in [0,1].

c) Now let $f(z) = z^2$ and p=1/8. The functions $g(z)$ and the identity function are sketched in the graph in Figure 4. From the figure we see that $g(z) = 0$ when $z \leq c$, and $g(z) > 0$ when $z > c$, with $c \approx 0.35$. Calculate the exact value of c (for $p = 1/8$).

d) Give the exact or approximate values of the equilibria for $p = 1/8$, show that these are indeed (approximate) equilibria, and indicate if they are stable or not, and why.

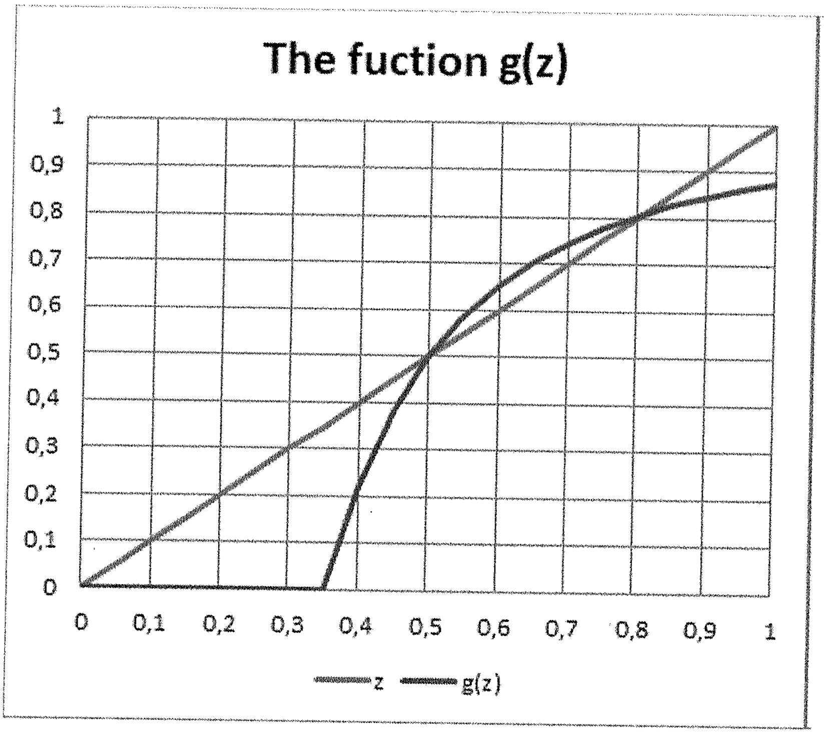

Figure 4

Question 7 (Chapter 19: Extensions of the Cascade Model) – 6 x 2 = 12 points

Consider the model from Chapter 19. There is a set of initial adopters of A. The remaining nodes with behavior *B* and any node will switch to *A* if at least a fraction *q* of its neighbors has adopted *A*. Such a model can be derived from a networked coordination game in which each node has a choice between two possible behaviors, *A* and *B*. Two nodes connected by an edge both receive a payoff *a* if they both select *A*, a payoff *b* if they both select *B*, and a payoff 0 otherwise. In case of equal payoffs a node prefers *A*.

a) Derive an expression for the threshold q in terms of *a* and *b*.

b) Suppose $a = 2$ and $b = 1$. Consider the network depicted in Figure 5. Assume node 2 and 9 start with behavior *A*. Indicate how the process of adopting *A* will proceed. Which nodes will eventually adopt *A*?

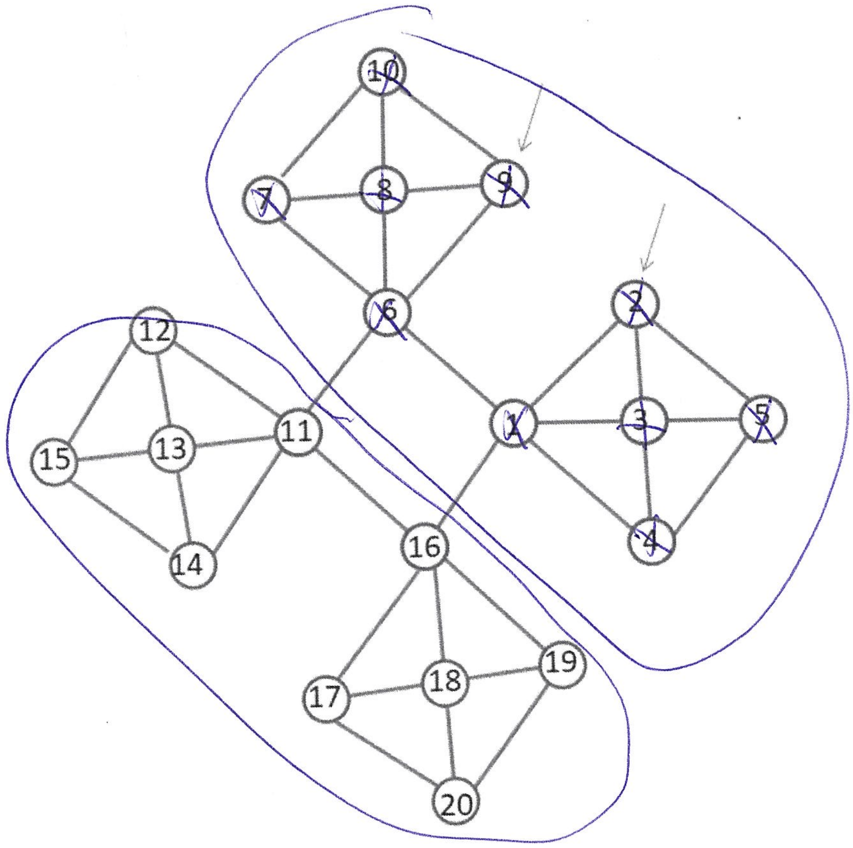

Figure 5

c) In case a complete cascade does not occur: describe in detail which two measures could be taken to ensure all nodes adopt *A*.

d) Find two disjoint (=non-overlapping) clusters in the network of density greater than 2/3.

e) Let *q=1/3*. Find a set of two initial adopters from *A* that could cause the entire network to adopt *A*. Explain your answer.

f) Let *q=1/3*. Can there be a single node in the network that could cause a complete cascade? Either indicate such a set or explain why this is not possible.

Question 8 – BONUS (Chapter 18: Power Laws) – 4 points

Consider the following model for the creation of links between Web pages:

(1) Pages are created in order, and named 1, 2, 3; ..., *N*. Each page creates a single link to another page as follows:

(2) When page *j* is created, it produces a link to an earlier web page by choosing between actions (a) and (b) below, according to the following probabilistic rule

- With probability *1/2* page *j* chooses a page *j* uniformly at random from among all earlier pages and creates a link to this page *i*.
- With probability *1/2*, page *j* instead chooses a page *i* uniformly at random from among all earlier pages in creates a link to the page that *i* points to.

Show that this model leads to a power law for the fraction of pages of in-degree *k* when *N* tends to infinity.