Web and Cloud Computing
academic year 2014-15

Rules

- This is a closed books exam.
- The operation of any electronic device is prohibited (e.g, no calculator, phone or PDA).
- Answer the questions being *precise, complete*, and *formal*.
- Write as *clearly* as possible, both in terms of handwriting and wording.

1. Consider the graph in Figure 1 and give:
   (a) The degree of each node and the average degree of the graph
   (b) The diameter
   (c) Now add directions to the edges in such a way as to obtain the highest possible PageRank for node c
   (d) Provide the $\mathbf{H}$ matrix representation of the directed graph
   (e) The PageRank at iteration two π_2^T without adjustments, i.e., $\pi_2^T = \pi_0^T \mathbf{HH}$
   (f) The stochastically adjusted matrix $\mathbf{S}$, and the Google matrix $\mathbf{G}$ (supposing $\alpha = 0.5$)

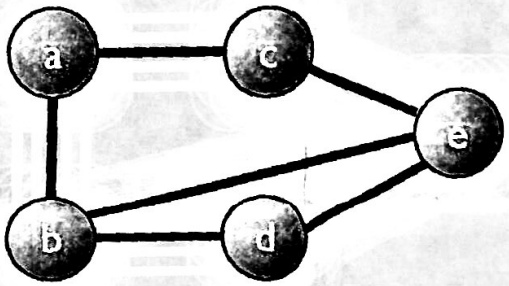

Figure 1: A graph.

2. What is index-free adjacency property? Why is it so important for graph databases?
3. Amazon contacted you with a request to improve their recommendation engine that determines what products should be shown to the user on the main page. You would like to base your recommendation on other products that have already been bought by the user. What other facts could you consider? How would you model the data? What kind of query do you need to write to answer the question for a particular user (based on your suggested model)? *Hint: use graph databases.*
4. How would you implement a leader election algorithm using Zookeeper?
5. Twitter is interested to calculate the amount of unique users who have been exposed to a certain topic (e.g., some URL, or a term/word, or a hashtag). We say that a person is "exposed" to a topic if he either tweets about it himself or he follows someone who did it. You are given full access to the internal Twitter database with all relevant data available. How would you compute the amount of unique people exposed to a given topic using Twitter Storm?
6. Answer the previous question by using Map/Reduce or Spark (use what you are more comfortable with). Compare your solution with the Twitter Storm solution, provided in the previous question.
7. What are R, W, and N parameters for data replication? What can you tell about the database consistency level looking at these parameters? What are the values of R, W, and N for a MongoDB replication based on replica sets?