Assignment 3 (Optional Assignment)

COMP 5703 - Fall 2013

December 2, 2013

1 Instructions

This is due in SCS office (or my office) by 9 AM on Monday December 16, 2013. This Assignment is optional, as I will take best two Assignments out of three. Please write clearly and answer questions precisely. As a thumb rule, the answer should be limited to ≤ 2 written pages, with ample spacing between lines and in margins, per question. Always start a new question on a new page, starting with Question 1, followed by Question 2, ..., Question n. Please cite all the references (including web-sites, names of friends, etc.) which you have used/consulted as the source of information for each of the questions. BTW, when a question asks you to design an algorithm - it requires you to (1) Clearly spell out the steps of your algorithm in pseudocode (2) Prove that your algorithm is correct and (3) Analyze the running time. By default a graph \( G = (V, E) \) is simple, undirected and connected.

2 Problems

1. Show that the Jaccard Distance which is defined as \( 1 - \) the Jaccard Similarity between the two sets is a metric.

2. Look at Algorithm 1.6 and Lemma 1.5 in the report on MST Verification. Take a connected weighted graph (with distinct weights) consisting of at least 10 nodes and 16 edges, and construct some spanning tree \( T \), its corresponding Full Branching Tree \( B \) and present some reasoning in your own words why Lemma 1.5 is true.

3. Prove that a matching is maximum if and only if there are no augmenting paths with respect to that matching. You may refer to the talk on Perfect Matching.

4. Prove that a bipartite graph \( G = (V = A \cup B, E) \) has a perfect matching if and only if for any subset \( S \subseteq A \) the number of vertices adjacent to \( S \) in \( B \) (denote it by \( N(S) \)) must be as large as \( |S| \) (i.e. \( |N(S)| \geq |S|, \forall S \in A \)).

5. Construct a non-trivial instance of a flow network, say with at least 6 nodes and at least 10 edges with distinct integral capacities. Execute the preflow algorithm and give some reasoning on the correctness of Lemma 1.9 in the report on PreFlow-Push Network Flow algorithm.

6. Present a proof, in your own words, of the Isolation Lemma, i.e. Lemma 1 in the report on “Parallel Algorithm for Perfect Matching”. What is it connection with the parallel algorithm for maximum matching?
7. Refer to Theorem 3.3 in the Bottleneck Spanning Tree report. Provide a formal reasoning on why this algorithm runs in $O(m \log^* n)$ time. Basically need to show understanding of Step 2 and where is this $\log^* n$ showing up in Step 2.

8. When applying amplification constructions to a locality-sensitive family of functions, which order of composition is ‘better’, and why? Explain when you would want to use different orders of construction.

9. Take two graphs, one which is 4-connected, and one which is almost 4-connected (i.e. for example by adding one or two extra edges, it will become 4-connected). Execute the Scan First Search based 4-connectivity certificate algorithm. Show that for the first graph you have a certificate (a subgraph) which is 4-connected and for the second graph you do not have such a certificate. Try to give some reasoning in terms of which Lemma in the proof was able to differentiate the two graphs in terms of 4-connectivity. You need to refer to the report on Sparse Certificates.

10. This problem requires you to understand Cauchy-Schwarz inequality. Present some ideas on the following problem. Suppose you have a good estimate of interpoint distances in the $L_1$ metric for a set of points. Using the Cauchy-Schwarz inequality, how can you obtain very good estimates for interpoint distances in any $L_p$ norm.