Seoul National University

M1522.001400 Introduction to Data Mining

Spring 2016, Kang

Homework 7: Advertising on the Web (Chapter 8)

Due: May 23, 09:30 AM

Reminders

- The points of this homework add up to 100.
- Like all homeworks, this has to be done individually.
- Lead T.A.: Jinhong Jung (montecast9@gmail.com)
- Please type your answers in English. Illegible handwriting may get no points, at the discretion of the graders.
- If you have a question about assignments, please upload your question in eTL.
- If you want to use slipdays or consider late submission with penalties, please note that you are allowed one week to submit your assignment after the due date.

Remember that:

- Whenever you are making an assumption, please state it clearly

Question 1

Define the graph G_n to have the $2n$ node

$$a_0, a_1, \dots, a_{n-1}, b_0, b_1, \dots, b_{n-1}$$

and the following edges. Each node a_i, for $i = 0, 1, \dots, n-1$, is connected to the nodes b_j and b_k, where

$$j = 2i \; mod \; n \text{ and } k = (2i+1) \; mod \; n$$

For instance, the graph G_4 has the following edges: (a_0, b_0), (a_0, b_1), (a_1, b_2), (a_1, b_3), (a_2, b_0), (a_2, b_1), (a_3, b_2), and (a_3, b_3). Answer the following questions. [30 points]

(a) Find one of perfect matchings for G_4. [10 points]

One of the following perfect matchings:

1. $(a_0, b_0), (a_1, b_2), (a_2, b_1), (a_3, b_3)$
2. $(a_0, b_0), (a_1, b_3), (a_2, b_1), (a_3, b_2)$
3. $(a_0, b_1), (a_1, b_2), (a_2, b_0), (a_3, b_3)$
4. $(a_0, b_1), (a_1, b_3), (a_2, b_0), (a_3, b_2)$

(b) Find one of perfect matchings for G_5. [10 points]

One of the following perfect matchings:

1. $(a_0, b_0), (a_1, b_2), (a_2, b_4), (a_3, b_1), (a_4, b_3)$
2. $(a_0, b_1), (a_1, b_3), (a_2, b_0), (a_3, b_2), (a_4, b_4)$

(c) How many perfect matchings do the graphs G_4 and G_5 have? [10 points]

G_4 has 4 perfect matchings, and G_5 has 2 perfect matchings.

Question 2

Whether or not the greedy algorithm for matching on a bipartite graph gives us a perfect matching for the graph of Figure 1 depends on the order in which we consider the edges. Among the 6! possible orders of the six edges, how many orders do they give us a perfect matching? Give a simple test for distinguishing those orders that do give the perfect matching from those that do not. [30 points]

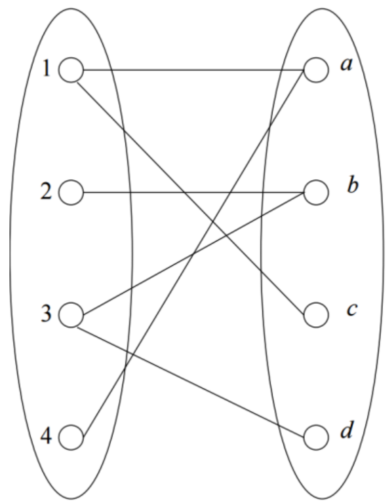

Figure 1. A bipartite graph

1. A simple test: when an order satisfies that $\big((\mathbf{1},\mathbf{c})\ \mathbf{or}\ (\mathbf{4},\mathbf{a})\ \prec (\mathbf{1},\boldsymbol{a})\big)$ **and** $\big((\mathbf{2},\mathbf{b})\ \mathbf{or}\ (\mathbf{3},\mathbf{d})\ \prec (\mathbf{3},\boldsymbol{b})\big)$**,** it gives us a perfect matching where $\mathrm{x} \prec \mathrm{y}$ indicates x precedes y in an order.
   - Note that $(1, \mathrm{c}) \prec (1, \mathrm{a}) \prec (4, \mathrm{a}) \prec (2, \mathrm{b}) \prec (3, \mathrm{b}) \prec (3, \mathrm{d})$ is one of possible orders.
2. There are 320 orders which give us a perfect matching.
   - Split the graph into two chains: $\{(1, c), (4, a), (1, a)\}$ and $\{(2, b), (3, \mathrm{d}), (3, \mathrm{b})\}$.
   - For each chain, there are 4 orders satisfying the above condition, respectively.
   - $\#$ of ways to mix the two chains $\times$
$\#$ of orders of the first chain with the condition $\times$
$\#$ of orders of the second chain with the condition $= \frac{6!}{3! \times 3!} \times 4 \times 4 = 320$

Question 3

Suppose that there are three advertisers, A, B, and C. There are three queries, x, y, and z. Each advertiser has a budget of 2. Advertiser A bids only on x; B bids on x and y, while C bids on x, y, and z. Note that on the query sequence xxyyzz, the optimum off-line algorithm would yield a revenue of 6, since all queries can be assigned. Answer the following questions. [40 points]

(a) Show that the greedy algorithm will assign at least 4 of these 6 queries. [20 points]

Greedy algorithm for this online matching problem can be defined as pairing a query with any advertiser who has budget on the query. If there are advertisers who have same budgets on the query, we can choose an advertiser as following pre-designated priority like (A > B > C). There are 6 possible priority orders, which are (A > B > C), (A > C > B), (B > A > C), (B > C > A), (C > A > B), (C > B > A).

Priority order	**xxyyzz**
A > B > C	AABBCC
A > C > B	AACC--
B > A > C	BBCC--
B > C > A	BBCC--
C > A > B	CCBB--
C > B >A	CCBB--

Table 1 shows that the greedy algorithm assigns at least 4 of the 6 queries.

(b) Find another sequence of queries such that the greedy algorithm can assign as few as half the queries that the optimum off-line algorithm assigns on that sequence. [20 points]

One possible sequence of queries is 'xxzz'. For this query, the worst-case matching of the greedy algorithm is 'CC--', but optimal matching of off-line algorithm is 'BBCC'. (Other queries sequences that satisfy above conditions are right answers)