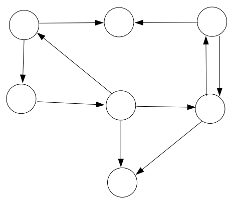
Maths 3467 quizzes
5: Thursday 5/12/13 due by noon Fri 13/12

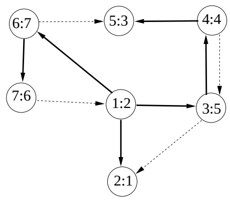
Please be sure you understand what is wanted. If in doubt, send e-mail. Collaboration is permissible.

Attempt three questions: two from questions 1–4, worth 12 marks each, and one from questions 5–7, worth 26 marks each.

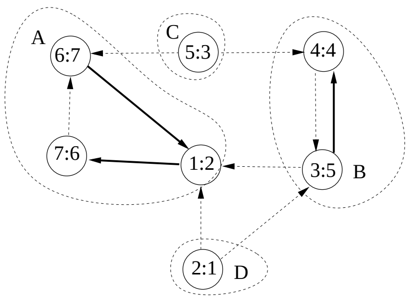
(1) Simulate both Sharir's algorithm and the Hopcroft-Tarjan algorithm (? the other one) on the following digraph.



Answer. Sharir's method: first dfs to get postorder rank.



Reverse the edges, start at highest postorder, dfs, etcetera: giving four components A,B,C,D, as illustrated.



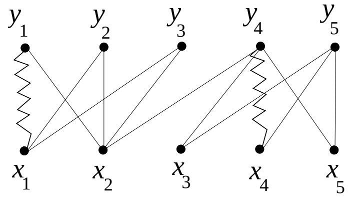
For Tarjan's method.

```
dfs(1)
  stack 1
  {1,2}
  dfs(2)
    stack 1 2
    end of dfs(2),
    link_rank = 2,
    scc {2}
    stack 1
```

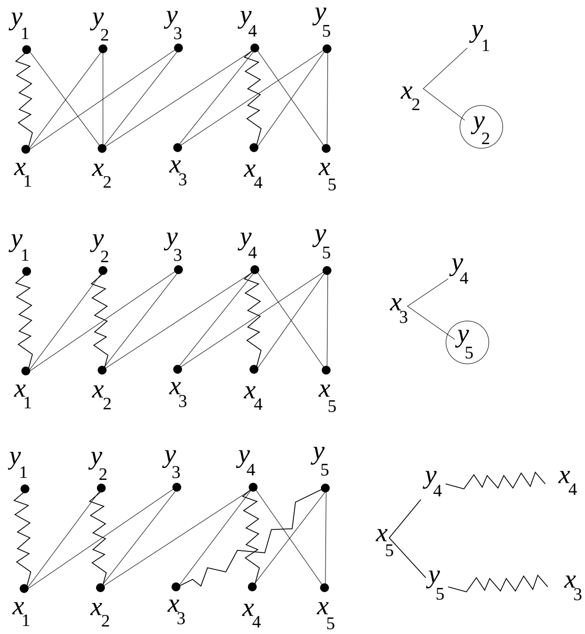
```
{1,3}
dfs(3)
  stack 1 3
  {3,2} 2 cleared no action.
  {3,4}
  dfs(4)
    stack 1 3 4
    {4,3} 3 uncleared link_rank(4) = 3
    {4,5}
    dfs(5)
      stack 1 3 4 5
      end of dfs(5), link_rank = 5
      scc {5}
      stack 1 3 4
    dfs(4) ends, link_rank(4) = 3
  in dfs(3), link_rank(3) unchanged
  dfs(3) ends, link_rank = 3
  scc {3,4}
{1,6}
dfs(6)
  stack  1 6
  {6,5} 5 cleared no action
  {6,7}
  dfs(7)
    stack 1 6 7
    {7,1} 1 not cleared link_rank(7) = 1
    dfs(7) ends
  in dfs(6), link_rank = 1
  dfs(6) ends
dfs(1) ends
  scc {1,6,7}
```

---

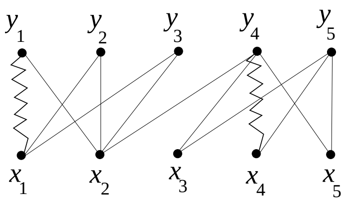
(2) By repeated use of augmenting paths, convert the matching illustrated below to a maximal matching. You should show that the matching is maximal. Begin the search for augmenting paths at the exposed vertex x_2.



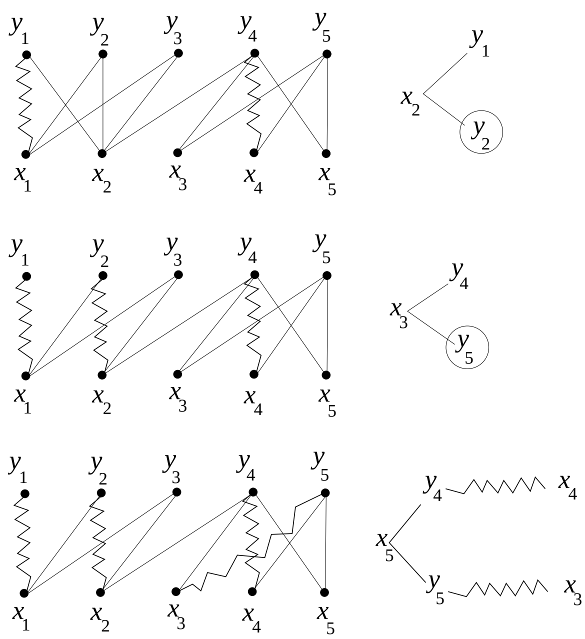
Answer. There are two augmentations. The last attempt at augmentation shows that x_3, x_4, x_5 are adjacent only to y_4 and y_5, so there is no perfect matching, so the 4-edge matching is maximal.



(2) By repeated use of augmenting paths, convert the matching illustrated below to a maximal matching. You should show that the matching is maximal. Begin the search for augmenting paths at the exposed vertex x_2.



Answer. There are two augmentations. The last attempt at augmentation shows that x_3, x_4, x_5 are adjacent only to y_4 and y_5, so there is no perfect matching, so the 4-edge matching is maximal.



(3) Let G be a graph (undirected). G is *connected* if for every two vertices u, v, there exist paths from u to v and v to u. You may assume (it's easy to prove) that every connected graph has a spanning tree.

If G is connected, an *articulation point* is a vertex v such that $G\backslash\{v\}$ is disconnected.

Given G is nonempty and connected, prove that there exists a vertex v which is *not* an articulation point, i.e., such that $G\backslash\{v\}$ is connected.

Hint: if you delete a leaf from a tree, what remains is still a tree — i.e., connected.

Answer. Let T be a spanning tree and v a leaf of T. Possibly it is the root as well, in which case G is trivial and $G\backslash\{v\}$ is empty and connected.

Otherwise, every vertex in T is reachable from the root, and if w is any vertex different from v, the path from the root to w does not go through v, so $T\backslash\{v\}$ is connected; so is $G\backslash\{v\}$. ∎

(4) Deduce from (3) that if G is a connected graph, then it is possible to find an expanding sequence

$$G_1 \subseteq G_2 \subseteq \ldots \subseteq G_n = G$$

of subgraphs, where each G_j has j vertices, and each G_j is connected.

Proof. If G is trivial the result is immediate, so we assume G contains at least two vertices. Start with $G_n = G$. Let v_n be an articulation point in G_n. Let $G_{n-1} = G\backslash\{v_n\}$. It is connected and nonempty, so by induction we can assume $G_1 \subseteq \ldots \subseteq G_{n-1}$ which can be extended to G_n. ∎

Optional extra. Show that if S is a collection of n (topological) closed discs in $\mathbb{R}^2$, whose union is connected. then S can be indexed in a sequence $D_1, \ldots, D_n$, such that for each k, $\bigcup_{j \leq k} D_j$ is connected.

(5) Let p be a (large) prime and let $W \subseteq \{0, \ldots, p-1\}$ be a set of cardinality n, where $n^2 < p$. Prove that there exists k, $1 \leq k \leq p-1$, such that the hash function

$$f_k : \quad x \mapsto (kx \mod p) \mod n^2$$

is injective **on** W.

Hints.

- Let $b_{k,j} = |\{x \in W : f_k(x) = j\}|$, where $1 \leq k \leq p-1$ and $0 \leq j \leq n^2 - 1$. Also let

$$s_k = \sum_j \binom{b_{k,j}}{2}$$

- For fixed k, how do you interpret s_k? Note that f_k is injective on W iff $s_k = 0$.

- It is enough, therefore, to show that

$$\sum_k s_k < p - 1.$$

- Let

$$T = \{(x, y, k) : \ x < y \in W \ \wedge \ f_k(x) = f_k(y)\}.$$

- Fix $x < y \in W$. Note that $y - x$ is invertible modulo p so the map $k \mapsto k(y - x)$ is injective. Give an upper bound on $|\{k : \ (x, y, k) \in T\}|$, and hence an upper bound on $|T|$.

To establish this, we evaluate the double summation differently. Fix $x > y \in W$. For $1 \leq k \leq p - 1$, we consider the situation where $f_k(x) = f_k(y)$.

$$(k(x - y) \mod p) \mod n^2 = 0$$

That is, $k(x - y) \mod p$ is in the set

$$-n^2, n^2, 2n^2, \ldots$$

Notice that since $k \neq 0$ and $x - y \neq 0$, $k(x - y) \mod p$ cannot be zero, hence zero is omitted from this list. That is, $k(x - y) \mod p$ belongs to a set of size

$$2\lfloor \frac{p-1}{n^2} \rfloor.$$

Since the map $k \mapsto k(x - y) \mod p$ is injective,

$$|\{k : \ (k(x - y) \mod p) \mod n^2\}| \leq 2\lfloor \frac{p-1}{n^2} \rfloor.$$

There $\binom{n}{2}$ such pairs x, y, whence

$$\sum_k s_k < \binom{n}{2} 2 \frac{p-1}{n^2} < p - 1$$

as required. ∎

(6) The *Hamiltonian cycle* problem is to determine of an undirected graph G whether its vertices belong to a single (simple) cycle. It is 'NP-complete,' which means it is unlikely to have efficient solutions. On the other hand, it is quite easy to determine whether the vertices belong to a union of disjoint cycles. Show this.

Unfortunately, this is stated incompletely (the cycles should be nontrivial), and a student discovered that it is *wrong*. The graph should be *directed*, and the Hamiltonian cycle should be the *directed Hamiltonian cycle* problem, which is also NP-complete.

Hint: Matching. Given $G = (V, E)$, where $V = \{v_1, \ldots, v_n\}$, choose disjoint sets $X = \{x_1, \ldots, x_n\}$ and $Y = \{y_1, \ldots, y_n\}$, and consider the bipartite graph with vertices $X \cup Y$ and edges

$$\{\{x_i, y_j\} : \ \{v_i, v_j\} \in E\}.$$

changed to

$$\{\{x_i, y_j\} : \ (v_i, v_j) \in E\}.$$

Answer. Build the bipartite graph as suggested. Call it H. Claim that H has a perfect matching if and only if G can be covered by a union of disjoint cycles.

Suppose M is a perfect matching. Define a permutation σ of the graph vertices:

$$\sigma(v_i) = v_j \qquad \iff \qquad \{x_i, y_j\} \in M.$$

Since M is a perfect matching, σ is a bijection. Since $(v_i, \sigma(v_i))$ is always an edge of G, σ has no fixed points.

Let $C_1, C_2, \ldots, C_k$ be its cycle decomposition. That is, $C_1 = v_{i(1)}, v_{i(2)} = \sigma(v_{i(1)}, \ldots,$ and since successive vertices in C_1 are adjacent in G (including the last, followed by $v_{i(1)}$), and the vertices on C_1 are distinct, C_1 is a cycle. Similarly for the other cycles. Therefore G is covered by nontrivial cycles.

Conversely, if G is covered by nontrivial cycles, there is associated a permutation σ.

Define

$$M = \{\{x_i, y_j\} : v_j = \sigma(v_i)\}$$

Then M is a perfect matching. ∎

(7) Show that quickselect has $O(n)$ average cost. Quickselect finds the k-th smallest in a set S of keys (assumed distinct for simplicity), as follows.

- Take a key f, for simplicity, always the first, in S and split S into three sets L, E, H: keys $< f$, $= f$, and $> f$.
- If $k \leq |L|$ then return quickselect (k, L).
- If $k = |L| + 1$ then return f.
- Otherwise return quickselect $(k - |L| - 1, H)$.

Taking the average cost, show that Quickselect is $O(n)$ on average.

Hint. I found it hard to estimate the average cost with quicksort-like recurrences. It is easier (I think) to consider, for each key sequence S, the average cost of finding the k-th smallest, for all k, $1 \leq k \leq n$. Also imagine the tree T which would be constructed by inserting the keys in the order of occurrence in S. Then the procedure is really to follow a path from the root to a given node in the tree T. The point is, however, that not all nodes in T are inspected: given k, indeed, only the branch, leading to the k-th node in inorder, is inspected. The cost of the search is the quantity Q-value which came up in an earlier quiz.

Answer. The total Q-value of an average tree (built by insertions) is $O(n^2)$ (Quiz 2, question 3). Hence the average Q-value of a tree with n nodes is $O(n)$, as required. ∎