Full Name: _____

Andrew ID: _____ **Section:** _____

15–210: Parallel and Sequential Data Structures and Algorithms

Practice Exam II

April 2016

- There are 17 pages in this examination, comprising 6 questions worth a total of 116 points. The last few pages are an appendix with costs of sequence, set and table operations.

- You have 80 minutes to complete this examination.

- Please answer all questions in the space provided with the question. Clearly indicate your answers.

- You may refer to your one double-sided $8\frac{1}{2} \times 11$in sheet of paper with notes, but to no other person or source, during the examination.

Circle the section YOU ATTEND

		Sections	
A	9:30am - 10:20am	Andra/Charlie
B	10:30am - 11:20am	Andres/Emma
C	12:30pm - 1:20pm	Anatol/John
D	12:30pm - 1:20pm	Aashir/Ashwin
E	1:30pm - 2:20pm	Nathaniel/Sonya
F	3:30pm - 4:20pm	Teddy/Vivek
G	4:30pm - 5:20pm	Alex/Patrick

Full Name: _____   **Andrew ID:** _____

Question	Points	Score
Short Answers	30	
(Sets and Tables) Bingled	16	
Dijkstra and A*	15	
(Shortest Paths) Wormholes	10	
Strongly Connected Components	20	
MST and Tree Contraction	25	
Total:	116	

Question 1: Short Answers (30 points)

Please answer the following questions each with a few sentences, or a short snippet of code (either pseudocode or SML).

(a) (4 points) Consider an undirected graph G with unique positive weights. Suppose it has a minimum spanning tree T. If we square all the edge weights and compute the MST again, do we still get the same tree structure? Explain briefly.

(b) (5 points) Lets say you are given a table that maps every student to the set of classes they take. Fill in the algorithm below that returns all classes, assuming there is at least one student in each class. Your algorithm must run in $O(m \log n)$ work and $O((\log m)(\log n))$ span, where n is the number of students and m is the sum of the number of classes taken across all students. Note, our solution is one line.

(c) (5 points) A new startup *FastRoute* wants to route information along a path in a communication network, represented as a graph. Each vertex represents a router and each edge a wire between routers. The wires are weighted by the maximum bandwidth they can support. *FastRoute* comes to you and asks you to develop an algorithm to find the path with maximum bandwidth from any source s to any destination t. As you would expect, the *bandwidth* of a path is the minimum of the bandwidths of the edges on that path; the minimum edge is the *bottleneck*.

Explain how to modify Dijkstra's algorithm to do this. In particular, how would you change the priority queue and the following relax step?

```
fun relax (Q, (u,v,w)) = PQ.insert (d(u) + w, v) Q
```

Justify your answer.

(d) (5 points) Given a graph with integer edge weights between 1 and 5 (inclusive), you want to find the shortest *weighted* path between a pair of vertices. How would you reduce this problem to the shortest *unweighted* path problem, which can be solved using BFS?

(e) (5 points) Recall the implementation of DFS shown in class using the `discover` and `finish` functions. Circle the correct answer for each of the following statements, assuming DFS starts at A:
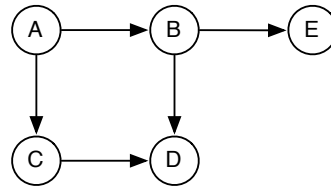
`discover D` could be called before `discover E`:	True	False
`discover E` could be called before `discover D`:	True	False
`discover D` could be called before `discover C`:	True	False
`finish A` could be called before `finish B`:	True	False
`finish D` could be called before `discover B`:	True	False

(f) (6 points) Circle *every* type of graph listed below for which star contraction will reduce the number of *edges* by a constant factor in expectation in every round until fully reduced (and hence imply $O(|E|)$ total work). You can assume redundant edges between vertices are removed.

(a) a graph in which all vertices have degree at most 2
(b) a graph in which all vertices have degree at most 3
(c) a graph in which all vertices have degree $\sqrt{|V|}$
(d) a graph containing a single cycle (i.e. a forest with one additional edge)
(e) the complete graph (i.e. an edge between every pair of vertices)
(f) any graph (still circle others if relevant)

Question 2: (Sets and Tables) Bingled   (16 points)

After forming your company Bingle to index the web allowing word searches based on logical combination of terms (e.g. "big" and "small"), you discover that there are already a couple companies out there that do it....and lo-and-behold, they even have similar names. You therefore decide to extend yours with additional features. In particular you want to support phrase queries: e.g. find all documents where "fun algorithms" appears.

You decide the right way to represent the index is as a table of sets where the keys of the table are strings (i.e. the words) and the elements of the sets are pairs of values consisting of a document identifier and an integer location in the document where the string appears. So, for example the following collection of three documents with integer document identifiers:

$$\langle\ (1,\ \textit{"the big dog"}),$$
$$(2,\ \textit{"a big dog ate a hat"}),$$
$$(3,\ \textit{"i read a big book"})\ \rangle$$

the document index would be represented as

$$\texttt{idx}\ =\ \{\ \ \textit{"a"} \mapsto \{(2,0),(2,4),(3,2)\}\,,$$
$$\textit{"big"} \mapsto \{(1,1),(2,1),(3,3)\}\,,$$
$$\textit{"dog"} \mapsto \{(1,2),(2,2)\}\,,$$
$$\dots\ \}$$

In particular you want to support the following interface

```
signature INDEX = sig
  type word = string
  type docId = int
  type index = docIdIntSet wordTable

  (* represents all documents and all locations where a phrase appears *)
  type docList

  val makeIndex : (docId * string) seq → index
  val find : index → word → docList
  val adj : docList * docList → docList
  val toSeq : docList → docId seq
end
```

where, given an index I, `toSeq (adj (find I "210", find I "rocks"))` would return a sequence of identifiers of documents where "210" appears immediately before "rocks", and

```
toSeq (adj (find I "Phil", adj (find I "loves", find I "cats")))
```

would return a sequence of identifiers of documents where the phrase "Phil love cats" appears.

(a) (8 points) Show SML code to generate the index from the sequence of documents. It should not be more than 8 lines of code and assuming all words have length less than some constant, must run in $O(n \log n)$ work and $O(\log^2 n)$ span, where n is the total number of words across all documents. A function to break a string into words has been given to you.

```
type index = docIdIntSet wordTable

fun makeIndex (docs : (docId * string) seq) : index =
  let
    fun toWords str = Seq.tokens (fn c ⇒ not (Char.isAlphaNum c)) str
```

(b) (8 points) Define the `docList` type and implement the function `adj` as defined above. You might find the function `setmap` useful. The solution should only be a few lines of code.
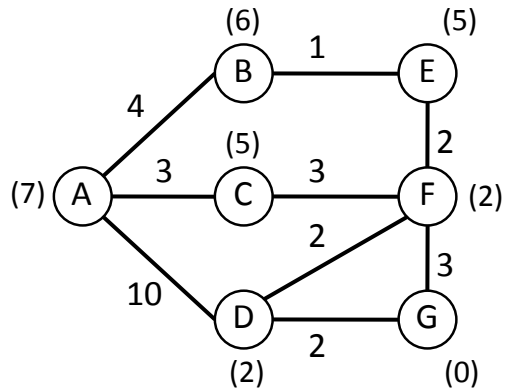
```
fun setmap f s = Set.fromSeq (Seq.map f (Set.toSeq s))


type docList =

fun adj (                    ,                    ) : docList =
```

Question 3: Dijkstra and A* (15 points)

(a) (6 points) Consider the graph shown below, where the edge weights appear next to the edges and the heuristic distances to vertex G are in parenthesis next to the vertices.

i. Show the order in which vertices are visited by Dijkstra when the source vertex is A.

ii. Show an order in which vertices are visited by A^* when the source vertex is A and the destination vertex is G.

(b) (4 points) What is the key reason you would choose to use A^* instead of Dijkstra's algorithm?

(c) (5 points) Show a 3-vertex example of a graph on which Dijkstra's algorithm always fails. Please clearly identify which vertex is the source.

Question 4: (Shortest Paths) Wormholes (10 points)

(a) (10 points) In your new job for a secret Government agency you have been told about the existence of wormholes (also known as Einstein-Rosen bridges) that connect various locations in the country. You have been tasked with designing an algorithm for finding the shortest path using a combination of roads and wormholes between a pair of locations. Traveling through a wormhole is instantaneous, for all practical purposes, but it turns out that on a given trip someone can only go through two wormholes otherwise they risk rearrangement of their atomic structure. The wormhole problem is therefore the weighted shortest path problem (assuming non-negative edge weights) with the additional constraint that

- Some edges are specially marked
- A path can take at most two of those edges

You still have your Dijkstra code from 210. You don't want to change your code after all you forgot how ML works so you just want to preprocess your graph so that a call to your code $SP(s, t)$ returns the correct solution to the wormhole problem. Explain how to do this. **At most 5 sentences**.

Question 5: Strongly Connected Components   (20 points)

In this question, you will write 2 functions on directed graphs. We assume that graphs are represented as:

```
type graph = vertexSet vertexTable
```

with key comparisons taking $O(1)$ work.

(a) (10 points) Given a directed graph $G = (V, E)$, its transpose G^T is another directed graph on the same vertices, with every edge flipped. More formally, $G^T = (V, E')$, where

$$E' = \{(b, a) \mid (a, b) \in E\}.$$

Here is a skeleton of an SML definition for `transpose` that computes the transpose of a graph. Fill in the blanks to complete the implementation. Your implementation must have $O(|E| \log |V|)$ work and $O(\log^2 |V|)$ span.

```
fun transpose (G :  graph) :  graph =
  let
    val S = vertexTable.toSeq(G)        (* returns (vertex*vertexSet) seq *)

    fun flip(u,nbrs) = Seq.map (_____) (vertexSet.toSeq nbrs)

    val ET = Seq.flatten(Seq.map flip S)

    val T = vertexTable._____ ET
  in
    vertexTable.map _____ T
  end
```

(b) (10 points) A *strongly connected component* of a directed graph $G = (V, E)$ is a subset S of V such that every vertex $u \in S$ can reach every other vertex $v \in S$ (i.e., there is a directed path from u to v), and such that no other vertex in V can be added to S without violating this condition. Every vertex belongs to exactly one strongly connected component in a graph.

Implement the function:

```
val scc : graph * vertex -> vertexSet
```

such that `scc(G,v)` returns the strongly connected component containing v. You may assume the existence of a function:

```
val reachable : graph * vertex -> vertexSet
```

such that `reachable(G,v)` returns all the vertices reachable from v in G. Not including the cost of `reachable`, your algorithm must have $O(|E| \log |V|)$ work and $O(\log^2 |V|)$ span. You might find `transpose` useful and can assume the given time bounds.

```
fun scc (G :  graph, v :  vertex) :  vertexSet =
```

_____

_____

Question 6: MST and Tree Contraction (25 points)

In *SegmentLab*, you implemented Borůvka's algorithm that interleaved star contractions and finding minimum weight edges. In this question you will analyze Borůvka's algorithm more carefully.

We'll assume throughout this problem that the edges are undirected, and each edge is labeled with a unique identifier (ℓ). The weights of the edges do not need to be unique, and $m = |E|$ and $n = |V|$.

```
1   % returns the set of edges in the minimum spanning tree of G
2   function MST(G = (V, E)) =
3       if |E| = 0 then {}
4       else let
5           val F = {min weight edge incident on v : v ∈ V}
6           val (V', P) = contract each tree in the forest (V, F) to a single vertex
7                       V' = remaining vertices
8                       P = mapping from each v ∈ V to its representative in V'
9           val E' = {(P_u, P_v, ℓ) : (u, v, ℓ) ∈ E | P_u ≠ P_v}
10      in
11          MST(G' = (V', E')) ∪ {ℓ : (u, v, ℓ) ∈ F}
12      end
```

(a) (4 points) Show an example graph with 4 vertices in which F will not include all the edges of the MST.

(b) (4 points) Prove that the set of edges F must be a forest (i.e. F has no cycle).

(c) (4 points) Suggest a technique to efficiently contract the forest in parallel. What is a tight asymptotic bound for the work and span of your contract, in terms of n? Explain briefly. Are these bounds worst case or expected case?

(d) (4 points) Argue that each recursive call to MST removes, in the worst case, at least *half* of the vertices; that is, $|V'| \leq \frac{|V|}{2}$.

(e) (4 points) What is the maximum number of edges that could remain after one step (i.e. what is $|E'|$)? Explain briefly.

(f) (5 points) What is the expected work and span of the overall algorithm in terms of m and n? Explain briefly. You can assume that calculating F takes $O(m)$ work and $O(\log n)$ span.

Appendix: Library Functions

```
signature SEQUENCE =
sig
  type 'a t
  type 'a seq = 'a t
  type 'a ord = 'a * 'a -> order
  datatype 'a listview = NIL | CONS of 'a * 'a seq
  datatype 'a treeview = EMPTY | ONE of 'a | PAIR of 'a seq * 'a seq

  exception Range
  exception Size

  val nth : 'a seq -> int -> 'a
  val length : 'a seq -> int
  val toList : 'a seq -> 'a list
  val toString : ('a -> string) -> 'a seq -> string
  val equal : ('a * 'a -> bool) -> 'a seq * 'a seq -> bool

  val empty : unit -> 'a seq
  val singleton : 'a -> 'a seq
  val tabulate : (int -> 'a) -> int -> 'a seq
  val fromList : 'a list -> 'a seq

  val rev : 'a seq -> 'a seq
  val append : 'a seq * 'a seq -> 'a seq
  val flatten : 'a seq seq -> 'a seq

  val filter : ('a -> bool) -> 'a seq -> 'a seq
  val map : ('a -> 'b) -> 'a seq -> 'b seq
  val zip : 'a seq * 'b seq -> ('a * 'b) seq
  val zipWith : ('a * 'b -> 'c) -> 'a seq * 'b seq -> 'c seq

  val enum : 'a seq -> (int * 'a) seq
  val filterIdx : (int * 'a -> bool) -> 'a seq -> 'a seq
  val mapIdx : (int * 'a -> 'b) -> 'a seq -> 'b seq
  val update : 'a seq * (int * 'a) -> 'a seq
  val inject : 'a seq * (int * 'a) seq -> 'a seq

  val subseq : 'a seq -> int * int -> 'a seq
  val take : 'a seq -> int -> 'a seq
  val drop : 'a seq -> int -> 'a seq
  val splitHead : 'a seq -> 'a listview
  val splitMid : 'a seq -> 'a treeview


  val iterate : ('b * 'a -> 'b) -> 'b -> 'a seq -> 'b
  val iteratePrefixes : ('b * 'a -> 'b) -> 'b -> 'a seq -> 'b seq * 'b
  val iteratePrefixesIncl : ('b * 'a -> 'b) -> 'b -> 'a seq -> 'b seq
  val reduce : ('a * 'a -> 'a) -> 'a -> 'a seq -> 'a
  val scan : ('a * 'a -> 'a) -> 'a -> 'a seq -> 'a seq * 'a
  val scanIncl : ('a * 'a -> 'a) -> 'a -> 'a seq -> 'a seq

  val sort : 'a ord -> 'a seq -> 'a seq
  val merge : 'a ord -> 'a seq * 'a seq -> 'a seq
  val collect : 'a ord -> ('a * 'b) seq -> ('a * 'b seq) seq
```

```
  val collate : 'a ord -> 'a seq ord
  val argmax : 'a ord -> 'a seq -> int

  val $ : 'a -> 'a seq
  val % : 'a list -> 'a seq
end
```

ArraySequence	Work *	Span *
empty ()		
singleton a		
length s	$O(1)$	$O(1)$
nth s i		
subseq s (i, len)	*	*

tabulate f n
   if $f(i)$ has W_i work and S_i span
 map f s
   if $f(s[i])$ has W_i work and S_i span, and $|s| = n$
 zipWith f (s, t)
   if $f(s[i], t[i])$ has W_i work and S_i span, and $\min(|s|, |t|) = n$
	$O\left(\sum_{i=0}^{n-1} W_i\right)$
	
	 *
		$O\left(\max_{i=0}^{n-1} S_i\right)$
		
		 *

reduce f b s
   if f does constant work and $|s| = n$
 scan f b s
   if f does constant work and $|s| = n$
 filter p s
   if p does constant work and $|s| = n$
	$O(n)$	$O(\lg n)$

flatten s	$O\left(\sum_{i=0}^{n-1} \left(1 + |s[i]|\right)\right)$	$O(\lg |s|)$
sort cmp s if cmp does constant work and $|s| = n$	$O(n \lg n)$	$O(\lg^2 n)$
merge cmp (s, t) if cmp does constant work, $|s| = n$, and $|t| = m$	$O(m + n)$	$O(\lg(m + n))$
append (s,t) if $|s| = n$, and $|t| = m$	$O(m + n)$	$O(1)$

```
signature TABLE =
sig
  structure Key : EQKEY
  structure Seq : SEQUENCE

  type 'a t
  type 'a table = 'a t

  structure Set : SET where Key = Key and Seq = Seq

  val size : 'a table -> int
  val domain : 'a table -> Set.t
  val range : 'a table -> 'a Seq.t
  val toString : ('a -> string) -> 'a table -> string
  val toSeq : 'a table -> (Key.t * 'a) Seq.t

  val find : 'a table -> Key.t -> 'a option
  val insert : 'a table * (Key.t * 'a) -> 'a table
  val insertWith : ('a * 'a -> 'a) -> 'a table * (Key.t * 'a) -> 'a table
  val delete : 'a table * Key.t -> 'a table

  val empty : unit -> 'a table
  val singleton : Key.t * 'a -> 'a table
  val tabulate : (Key.t -> 'a) -> Set.t -> 'a table
  val collect : (Key.t * 'a) Seq.t -> 'a Seq.t table
  val fromSeq : (Key.t * 'a) Seq.t -> 'a table

  val map : ('a -> 'b) -> 'a table -> 'b table
  val mapKey : (Key.t * 'a -> 'b) -> 'a table -> 'b table
  val filter : ('a -> bool) -> 'a table -> 'a table
  val filterKey : (Key.t * 'a -> bool) -> 'a table -> 'a table

  val reduce : ('a * 'a -> 'a) -> 'a -> 'a table -> 'a
  val iterate : ('b * 'a -> 'b) -> 'b -> 'a table -> 'b
  val iteratePrefixes : ('b * 'a -> 'b) -> 'b -> 'a table -> ('b table * 'b)

  val union : ('a * 'a -> 'a) -> ('a table * 'a table) -> 'a table
  val intersection : ('a * 'b -> 'c) -> ('a table * 'b table) -> 'c table
  val difference : 'a table * 'b table -> 'a table

  val restrict : 'a table * Set.t -> 'a table
  val subtract : 'a table * Set.t -> 'a table

  val $ : (Key.t * 'a) -> 'a table
end
```

```
signature SET =
sig
  structure Key : EQKEY
  structure Seq : SEQUENCE

  type t
  type set = t

  val size : set -> int
  val toString : set -> string
  val toSeq : set -> Key.t Seq.t

  val empty : unit -> set
  val singleton : Key.t -> set
  val fromSeq : Key.t Seq.t -> set

  val find : set -> Key.t -> bool
  val insert : set * Key.t -> set
  val delete : set * Key.t -> set

  val filter : (Key.t -> bool) -> set -> set

  val reduceKey : (Key.t * Key.t -> Key.t) -> Key.t -> set -> Key.t
  val iterateKey : ('a * Key.t -> 'a) -> 'a -> set -> 'a

  val union : set * set -> set
  val intersection : set * set -> set
  val difference : set * set -> set

  val $ : Key.t -> set
end
```

MkTreapTable	Work	Span
size T	$O(1)$ ${}_*$	$O(1)$ ${}_*$
filter f T map f T	$\displaystyle\sum_{(k\mapsto v)\in T} W(f(v))$	$\lg|T|+$ $\displaystyle\max_{(k\mapsto v)\in T} S(f(v))$
tabulate f X	$\displaystyle\sum_{k\in X} W(f(k))$	$\displaystyle\max_{k\in X} S(f(k))$
reduce f b T if f does constant work	$O(|T|)$ ${}_*$	$O(\lg|T|)$ ${}_*$
insertWith f $(T,(k,v))$ if f does constant work find T k delete (T,k)	$O(\lg|T|)$ ${}_*$	$O(\lg|T|)$ ${}_*$
domain T range T toSeq T	$O(|T|)$ ${}_*$	$O(\lg|T|)$ ${}_*$
collect S fromSeq S	$O(|S|\lg|S|)$	$O(\lg^2|S|)$

For each argument pair (A,B) below, let $n = \max(|A|,|B|)$ and $m = \min(|A|,|B|)$.

MkTreapTable	Work ${}_*$	Span ${}_*$
union f (X,Y) intersection f (X,Y) difference (X,Y) restrict (T,X) subtract (T,X)	$O\big(m\lg(\tfrac{n+m}{m})\big)$	$O\big(\lg(n+m)\big)$