Chapter 99

Review session

CS 473: Fundamental Algorithms, Fall 2011
September 22, 2011

99.0.0.1 Why Graphs?

(A) Graphs help model networks which are ubiquitous: transportation networks (rail, roads, airways), social networks (interpersonal relationships), information networks (web page links) etc etc.
(B) Fundamental objects in Computer Science, Optimization, Combinatorics
(C) Many important and useful optimization problems are graph problems
(D) Graph theory: elegant, fun and deep mathematics

99.0.0.2 Basic Graph Search

Given $G = (V, E)$ and vertex $u \in V$:

```
Explore(u):
    Initialize S = {u}
    while there is an edge (x, y) with x ∈ S and y ∉ S do
        add y to S
```

99.0.0.3 DFS in Directed Graphs

```
DFS(G)
    Mark all nodes u as unvisited
    T is set to ∅
    time = 0
    while there is an unvisited node u do
        DFS(u)

    Output T
```

```
DFS(u)
    Mark u as visited
    pre(u) = ++time
    for each edge (u, v) in Out(u) do
        if v is not marked
            add edge (u, v) to T
            DFS(v)
    post(u) = ++time
```

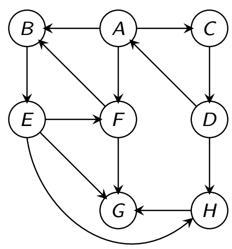
99.0.0.4 pre and post numbers

Node u is ***active*** in time interval $[\mathrm{pre}(u), \mathrm{post}(u)]$

Proposition 99.0.1 *For any two nodes u and v, the two intervals $[\mathrm{pre}(u), \mathrm{post}(u)]$ and $[\mathrm{pre}(v), \mathrm{post}(v)]$ are disjoint or one is contained in the other.*

99.0.0.5 Connectivity and Strong Connected Components

Definition 99.0.2 *Given a directed graph G, u is strongly connected to v if u can reach v* and *v can reach u. In other words $v \in rch(u)$ and $u \in rch(v)$.*



99.0.0.6 Directed Graph Connectivity Problems

(A) Given G and nodes u and v, can u reach v?
(B) Given G and u, compute $\mathrm{rch}(u)$.
(C) Given G and u, compute all v that can reach u, that is all v such that $u \in \mathrm{rch}(v)$.
(D) Find the strongly connected component containing node u, that is SCC(u).
(E) Is G strongly connected (a single strong component)?
(F) Compute *all* strongly connected components of G.

First four problems can be solve in $O(n+m)$ time by adapting **BFS**/**DFS** to directed graphs. The last one requires a clever **DFS** based algorithm.

99.0.0.7 DFS Properties

Generalizing ideas from undirected graphs:
(A) $DFS(u)$ outputs a directed out-tree T rooted at u
(B) A vertex v is in T if and only if $v \in \mathrm{rch}(u)$
(C) For any two vertices x, y the intervals $[\mathrm{pre}(x), \mathrm{post}(x)]$ and $[\mathrm{pre}(y), \mathrm{post}(y)]$ are either disjoint are one is contained in the other.
(D) The running time of $DFS(u)$ is $O(k)$ where $k = \sum_{v \in \mathrm{rch}(u)} |Adj(v)|$ plus the time to initialize the Mark array.
(E) **DFS**(G) takes $O(m+n)$ time. Edges in T form a disjoint collection of of out-trees. Output of $DFS(G)$ depends on the order in which vertices are considered.

99.0.0.8 DFS Tree

Edges of G can be classified with respect to the **DFS** tree T as:

(A) ***Tree edges*** that belong to T

(B) A ***forward edge*** is a non-tree edges (x, y) such that $\text{pre}(x) < \text{pre}(y) < \text{post}(y) < \text{post}(x)$.

(C) A ***backward edge*** is a non-tree edge (x, y) such that $\text{pre}(y) < \text{pre}(x) < \text{post}(x) < \text{post}(y)$.

(D) A ***cross edge*** is a non-tree edges (x, y) such that the intervals $[\text{pre}(x), \text{post}(x)]$ and $[\text{pre}(y), \text{post}(y)]$ are disjoint.

99.0.0.9 Algorithms via DFS

$SC(G, u) = \{v \mid u \text{ is strongly connected to } v\}$

(A) Find the strongly connected component containing node u. That is, compute $\text{SCC}(G, u)$.

$\text{SCC}(G, u) = \text{rch}(G, u) \cap \text{rch}(G^{rev}, u)$

Hence, $\text{SCC}(G, u)$ can be computed with two **DFS**es, one in G and the other in G^{rev}. Total $O(n + m)$ time.

99.0.1 Linear Time Algorithm

99.0.1.1 ...for computing the strong connected components in G

```
do DFS(G^rev) and sort vertices in decreasing post order.
Mark all nodes as unvisited
for each u in the computed order do
    if u is not visited then
        DFS(u)
        Let S_u be the nodes reached by u
        Output S_u as a strong connected component
        Remove S_u from G
```

Analysis

Running time is $O(n + m)$. (Exercise)

Example: Makefile

99.0.1.2 BFS with Distances

```
BFS(s)
    Mark all vertices as unvisited and for each v set dist(v) = ∞
    Initialize search tree T to be empty
    Mark vertex s as visited and set dist(s) = 0
    set Q to be the empty queue
    enq(s)
    while Q is nonempty do
        u = deq(Q)
        for each vertex v ∈ Adj(u) do
            if v is not visited do
                add edge (u, v) to T
                Mark v as visited, enq(v)
                and set dist(v) = dist(u) + 1
```

Proposition 99.0.3 **BFS**(s) *runs in* $O(n+m)$ *time.*

99.0.1.3 BFS with Layers

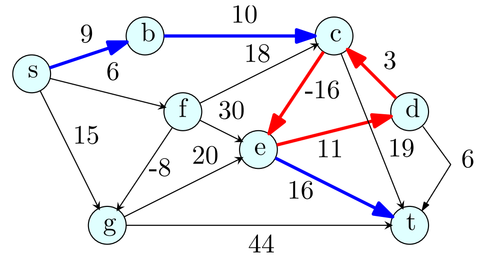
```
BFSLayers(s):
Mark all vertices as unvisited and initialize T to be empty
Mark s as visited and set L_0 = {s}
i = 0
while L_i is not empty do
        initialize L_{i+1} to be an empty list
        for each u in L_i do
            for each edge (u, v) ∈ Adj(u) do
            if v is not visited
                    mark v as visited
                    add (u, v) to tree T
                    add v to L_{i+1}
        i = i + 1
```

Running time: $O(n+m)$

99.0.2 Checking if a graph is bipartite...

99.0.2.1 Linear time algorithm

Corollary 99.0.4 *There is an* $O(n+m)$ *time algorithm to check if* G *is bipartite and output an odd cycle if it is not.*



99.0.2.2 Dijkstra's Algorithm

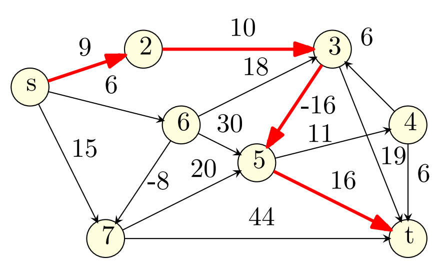
`Initialize for each node v,` $\mathrm{dist}(s,v) = \infty$
`Initialize` $S = \{s\}$, $\mathrm{dist}(s,s) = 0$
for $i = 1$ `to` $|V|$ **do**
  `Let v be such that` $\mathrm{dist}(s,v) = \min_{u \in V-S} \mathrm{dist}(s,u)$
  $S = S \cup \{v\}$
  for `each` u `in` $\mathrm{Adj}(v)$ **do**
    $\mathrm{dist}(s,u) = min\Big(\mathrm{dist}(s,u),\ \mathrm{dist}(s,v) + \ell(v,u)\Big)$

(A) Using Fibonacci heaps. Running time: $O(m + n \log n)$.
(B) Can compute shortest path tree.

99.0.2.3 Single-Source Shortest Paths with Negative Edge Lengths

Single-Source Shortest Path Problems
Input: A *directed* graph $G = (V, E)$ with arbitrary (including negative) edge lengths. For edge $e = (u, v)$, $\ell(e) = \ell(u, v)$ is its length.

- Given nodes s, t find shortest path from s to t.
- Given node s find shortest path from s to all other nodes.



99.0.2.4 Negative Length Cycles

Definition 99.0.5 *A cycle C is a negative length cycle if the sum of the edge lengths of C is negative.*

99.0.2.5 A Generic Shortest Path Algorithm

Dijkstra's algorithm does not work with negative edges.

```
Relax(e = (u, v))
    if (d(s, v) > d(s, u) + ℓ(u, v)) then
        d(s, v) = d(s, u) + ℓ(u, v)
```

```
d(s, s) = 0
for each node u ≠ s do
    d(s, u) = ∞

while there is a tense edge do
    Pick a tense edge e
    Relax(e)

Output d(s, u) values
```

99.0.2.6 Bellman-Ford to detect Negative Cycles

```
for each u ∈ V do
    d(s, u) = ∞
d(s, s) = 0

for i = 1 to |V| − 1 do
    for each edge e = (u, v) do
        Relax(e)

for each edge e = (u, v) do
    if e = (u, v) is tense then
        Stop and output that s can reach a negative length cycle

Output for each u ∈ V:  d(s, u)
```

(A) Total running time: $O(mn)$.
(B) Can detect negative cycle reachable from s.
(C) With appropriate construction - can detect any negative cycle in a graph.

99.0.3 Shortest paths in DAGs

99.0.3.1 Algorithm for DAGs

```
ShortestPathInDAG(G, s):
    s = v_1, v_2, v_{i+1}, ..., v_n be a topological sort of G
    for i = 1 to n do
        d(s, v_i) = ∞
    d(s, s) = 0

    for i = 1 to n − 1 do
        for each edge e in Adj(v_i) do
            Relax(e)

    return d(s, ·) values computed
```

Running time: $O(m+n)$ time algorithm! Works for negative edge lengths and hence can find *longest* paths in a DAG.

99.0.3.2 Reduction

Reducing problem A to problem B:

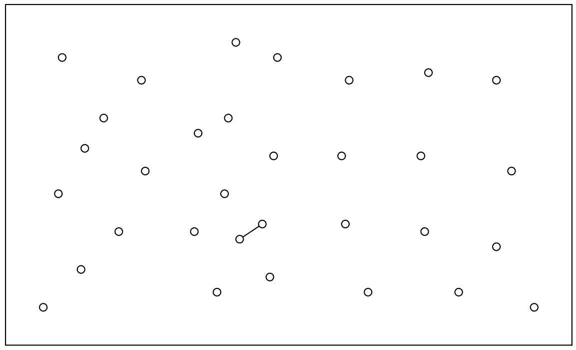
(A) Algorithm for A uses algorithm for B as a *black box*.
(B) Example: Uniqueness (or distinct element) to sorting.

99.0.3.3 Recursion

(A) Recursion is a very powerful and fundamental technique.
(B) Basis for several other methods.
    (A) Divide and conquer.
    (B) Dynamic programming.
    (C) Enumeration and branch and bound etc.
    (D) Some classes of greedy algorithms.
(C) Recurrences arise in analysis.

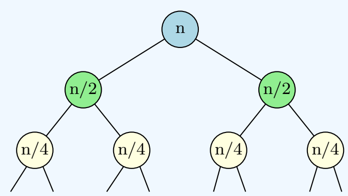
Examples seen:

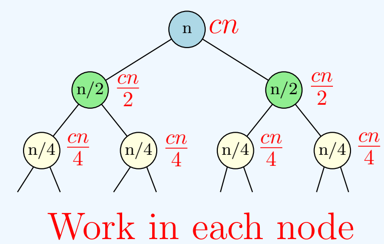
(A) Recursion: Tower of Hanoi, Selection sort, Quick Sort.
(B) Divide & Conquer:
    (A) Merge sort.
    (B) Multiplying large numbers.



99.0.4 Solving recurrences using recursion trees

99.0.4.1 An illustrated example...





99.0.5 Solving recurrences

99.0.5.1 The other "technique" - guess and verify

(A) Guess solution to recurrence.
(B) Verify it via induction.

Solved in class:

(A) $T(n) = 2T(n/2) + n/\log n$.
(B) $T(n) = T(\sqrt{n}) + 1$.
(C) $T(n) = \sqrt{n}T(\sqrt{n}) + n$.
(D) $T(n) = T(n/4) + T(3n/4) + n$

99.0.5.2 Closest Pair - the problem

Input Given a set S of n points on the plane

Goal Find $p, q \in S$ such that $d(p, q)$ is minimum

Algorithm:

One can compute closest pair points in the plane in $O(n \log n)$ time using divide and conquer.

99.0.5.3 Median selection

Problem

Given list L of n numbers, and a number k find kth smallest number in n.

(A) Quick Sort can be modified to solve it (but worst case running time is quadratic (if lucky linear time).

(B) Seen divide & conquer algorithm...
Involved, but linear running time.

99.0.6 Recursive algorithm for Selection

99.0.6.1 A feast for recursion

```
select(A, j):
    n = |A|
    if n ≤ 10 then
        Compute jth smallest element in A using brute force.
    Form lists L_1, L_2, ..., L_⌈n/5⌉ where L_i = {A[5i − 4], ..., A[5i]}
    Find median b_i of each L_i using brute-force
    B is the array of b_1, b_2, ..., b_⌈n/5⌉.
    b = select(B, ⌈n/10⌉)
    Partition A into A_less or equal and A_greater using b as pivot
    if |A_less or equal| = j then
        return b
    if |A_less or equal| > j) then
        return select(A_less or equal, j)
    else
        return select(A_greater, j − |A_less or equal|)
```

99.0.6.2 Back to Recursion

Seen some simple recursive algorithms:

(A) Binary search.

(B) Fast exponentiation.

(C) Fibonacci numbers.

(D) Maximum weight independent set.