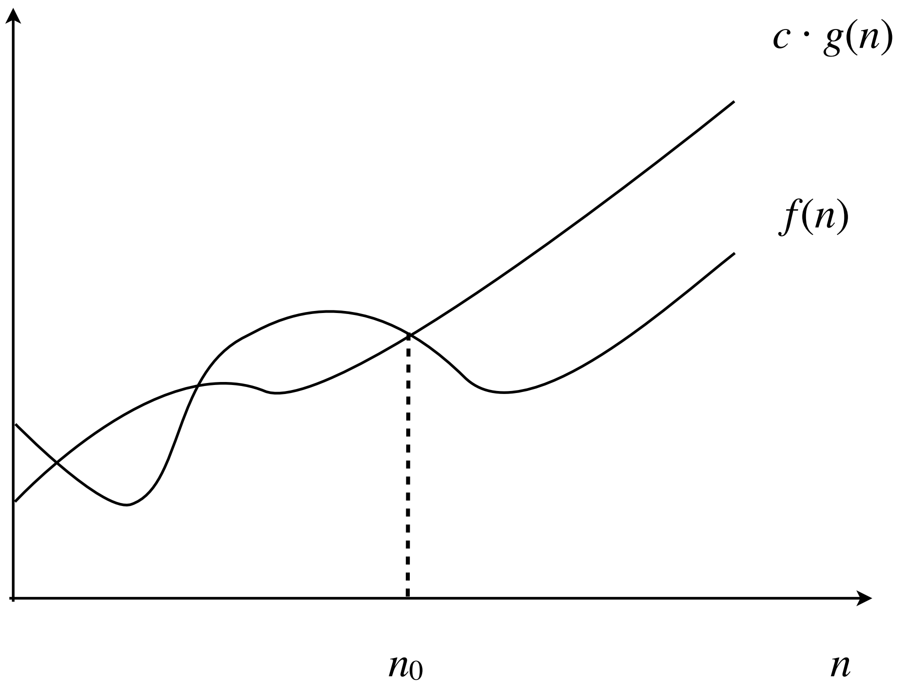
Graphs and Traversals

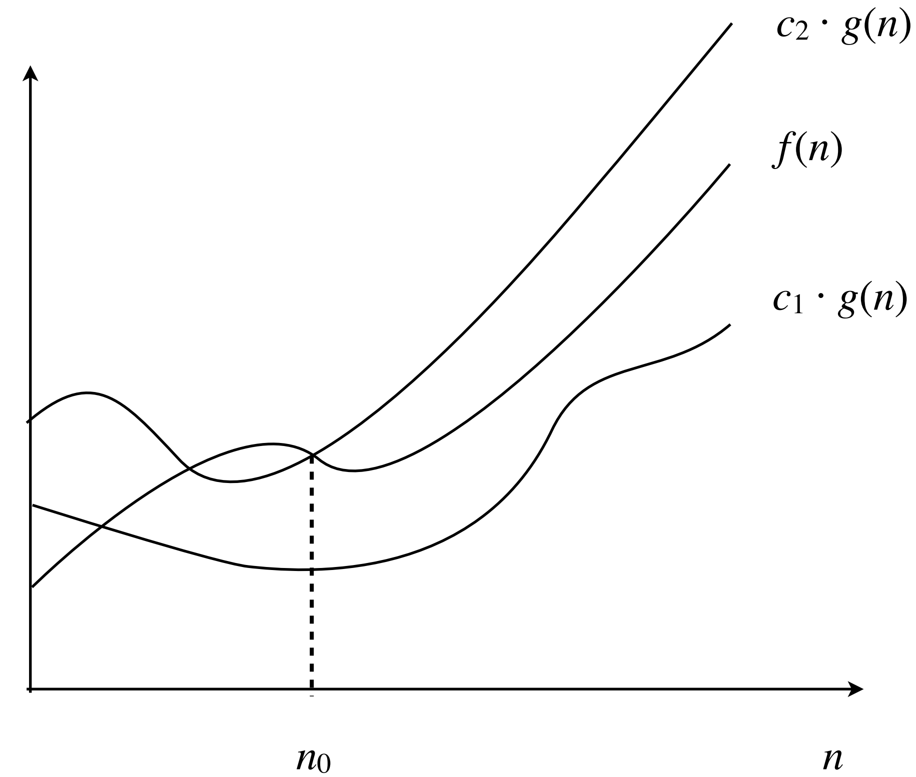
Reminders/ Check in

- **Assignment 0** due tonight at 11 pm
- Assignment 1 will be released later today
- If you have done so already, check out Problem Set Advice
- Take advantage of office hours:
  - Mine: 1.30-3 pm, TAs: 3-5pm, 7-11 pm

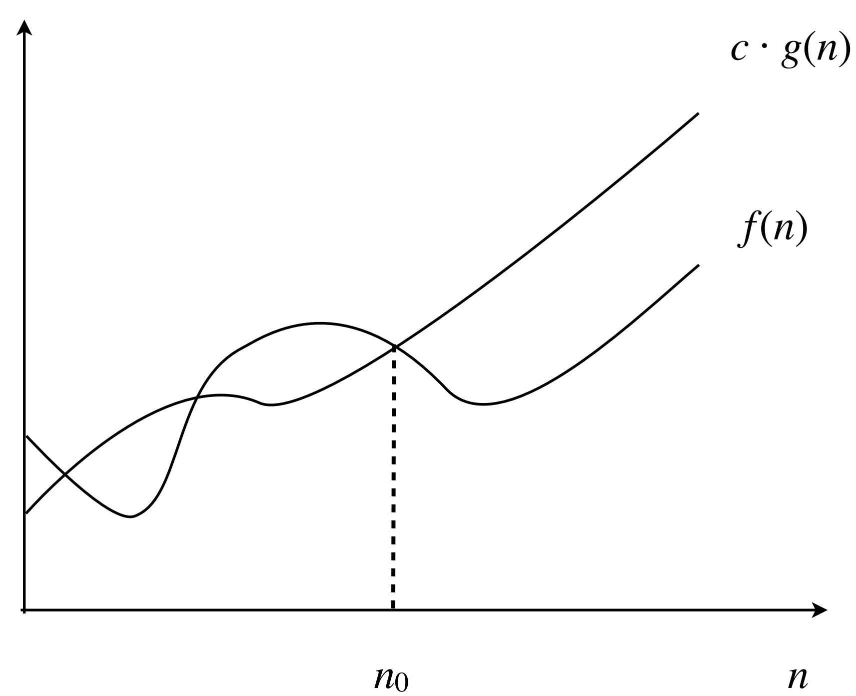
Asymptotics



$$f(n) = O(g(n))$$



$$f(n) = \Omega(g(n))$$



$$f(n) = \Theta(g(n))$$

Tools for Comparing Asymptotics

- Logs grow slowly than any polynomial:
  - $\log_a n = O(n^b)$ for every $a > 1,\ b > 0$
- Exponentials grow faster than any polynomial:
  - $n^d = O(r^n)$ for every $d > 1,\ r > 0$
- Taking logs
  - As $\log x$ is a strictly increasing function for $x > 0$, $\log(f(n)) < \log(g(n))$ implies $f(n) < g(n)$
  - E.g. Compare $3^{\log n}$ vs 2^n
    - Taking log of both, $\log n \log 3$ vs n
  - Beware: when comparing logs, constants matter!

Tools for Comparing Asymptotics

- Using limits
  - If $\lim_{n\to\infty} \frac{f(x)}{g(x)} = 0$, then $f(x) = O(g(x))$
  - If $\lim_{n\to\infty} \frac{f(x)}{g(x)} = c$ for some constant $0 < c < \infty$, then $f(x) = \Theta(g(x))$

Analyzing
Gale Shapley

Gale-Shapely Algorithm

GALE–SHAPLEY (*preference lists for hospitals and students*)

INITIALIZE M to empty matching.

WHILE (some hospital h is unmatched and hasn't proposed to every student)

  $s \leftarrow$ first student on h's list to whom h has not yet proposed.

  IF (s is unmatched)

    Add h–s to matching M.

  ELSE IF (s prefers h to current partner h')

    Replace h'–s with h–s in matching M.

  ELSE

    s rejects h.

RETURN stable matching M.

Gale-Shapely Algorithm

Number of iterations?

GALE–SHAPLEY (*preference lists for hospitals and students*)

INITIALIZE M to empty matching.

WHILE (some hospital h is unmatched and hasn't proposed to every student)

  $s \leftarrow$ first student on h's list to whom h has not yet proposed.

  IF (s is unmatched)

    Add h–s to matching M.

  ELSE IF (s prefers h to current partner h')

    Replace h'–s with h–s in matching M.

  ELSE

    s rejects h.

RETURN stable matching M.

How long does each iteration take?

Analyzing the Algorithm: Performance

- Each hospital makes an offer to each student at most once
- Upper bound on total number of offers that can ever be made?
    - At most n^2 iterations
- What do we do in each iteration?

Gale-Shapely Algorithm

GALE–SHAPLEY (*preference lists for hospitals and students*)

INITIALIZE M to empty matching.

WHILE (some hospital h is unmatched and hasn't proposed to every student)

  $s \leftarrow$ first student on h's list to whom h has not yet proposed.

  IF (s is unmatched)

    Add h–s to matching M.

  ELSE IF (s prefers h to current partner h')

    Replace h'–s with h–s in matching M.

  ELSE

    s rejects h.

RETURN stable matching M.

Find top student in h's list

Check matched status, update matching

check if s prefers h to h'?

Analyzing the Algorithm: Performance

- Each hospital makes an offer to each student at most once
- Upper bound on total number of offers that can ever be made?
    - At most n^2 iterations
- What do we do in each iteration?
    - Select a free hospital h
    - Find top ranked s not yet offered a post by h
    - Find s's ranking of a given hospital
    - Add to & delete from set of matched pairs
- How long does it take?
    - Depends on how we are store the input!

Analyzing Time Complexity

- **Input representation.** Index students and hospitals $1, \ldots, n$
  - Each student provides a sorted list of hospitals (most to least preferred) and each hospital provides a sorted list of students
- Of students not yet offered a post by h, find most preferred
  - Takes $O(1)$ *time*
- Does s prefer h to the current hospital h'?
  - How to do this operation efficiently? (In $O(1)$ time?)
  - If we scan the list, it will take $O(n)$ time
- Identify efficient data structures for operations:
  - For each s, create inverse of preference list of hospitals

Analyzing Time Complexity

- Does s prefer h to the current hospital h'?
    - For each s, create inverse of preference list of hospitals
- Once we have inverse list, this step takes $O(1)$ time
- How long does this preprocessing step take?
    - $O(n)$ time for each student, $O(n^2)$ overall

student prefers hospital 4 to 6 since rank[4] < rank[6]

Student preference list indexed by rank

	1st	2nd	3rd	4th	5th	6th	7th	8th
pref[]	8	3	7	1	4	5	6	2

Inverse pref-list indexed by hospital #

	1	2	3	4	5	6	7	8
rank[]	4th	8th	2nd	5th	6th	7th	3rd	1st

```
for i = 1 to n
    rank[pref[i]] = i
```

Analyzing Time Complexity

Analyzing overall running time:

- Preprocessing to create inverse list: $O(n^2)$
- Maintain free hospitals: Queue: $O(1)$ for get() and put()
- Add to & delete from set of matched pairs:
    - Array, Matched(*s*) = *h* currently matched to s (or 'free') : Creation time (preprocessing) $O(n)$; update time $O(1)$
- Each iteration thus takes $O(1)$ time
- Overall, $O(n^2)$ time preprocessing + $O(n^2)$ time in iterations
- Is this a linear time algorithm?
    - Yes! Here input size is $O(n^2)$ size, *linear in input size*

Stable Matching Fun Facts

- It matters which side of the market you are on!
- In the hospital-proposing algorithm, each hospital gets matched to its best-achievable partner, and each student gets matched to their worst-achievable partner!
- **Lemma**. $S^* = \{(h, \text{best}(h)) \mid h \in H\}$
  $= \{(\text{worst}(s), s) \mid s \in S\}$
  is the unique output of the Gale-Shapely algorithm (regardless of the order of execution).
- There are many stable matchings:
  - a whole lattice of stable matchings in between the hospital-optimal and student-optimal matchings

The Match

- **National Resident Matching Program**
    - Original 1952 implementation of the DA algorithm was the **hospital-optimal version**
- Students protested that the match was favoring hospitals

MONDAY, OCTOBER 22, 1951. The New York Times MONDAY, OCTOBER 22, 1951. 25 L+

MEDICAL SENIORS HIT INTERNE PLAN

Delegates of 44 Schools Meet Here to Protest Selection by 'Matching Machine'

TEACHERS PRAISE SYSTEM

They Argue That It Bars Unfair Recruiting—Students Insist on Choosing Their Hospitals

Delegates representing seniors in nearly all of the country's leading medical schools met here yesterday to express overwhelming opposition to a proposed mathematically contrived plan to place medical students in hospitals as internes.

They indicated that a great majority of their classmates preferred the present system whereby the country's hospitals, which have 10,000 internships, scramble for the best of a year's 6,000 medical graduates.

The meeting was held at Bard Hall of the College of Physicians and Surgeons of Columbia University, but it was made clear that the university was not its sponsor.

By Winston Churchill: The Second World War

Volume V—Closing the Ring

INSTALLMENT 15—TEHERAN: CONCLUSIONS

Book II—Teheran to Rome

Conversation at Luncheon, December 1—The Frontiers of Poland—The "Curzon Line", and the Line of the Oder — Finland — "No Annexations and No Indemnities"—The Question of Germany — Partition? — President Roosevelt's Suggestion—I Unfold a Personal View — Marshal Stalin's Standpoint — Broad Agreement on Military Policy — Political Aspects Remote and Speculative—Deep Fear of German Might at This War Climax—The Present Partition—"It Cannot Last".

SEVERAL of our gravest political issues stood out before and after the main decision on strategy had been reached [at the Teheran conference]. The Three lunched together again at the President's table in the Soviet Legation on December 1 [1943]. In addition on this occasion Molotov, Hopkins, Eden, Clark Kerr, and Harriman were present. The question of inducing Turkey to enter into the war was our first topic.

There was a very great measure of agreement on the limited steps for which I asked in order to win the great prize of bringing Turkey into the war.

* * *

Poland was the next important subject.

The New York Times.

ROOSEVELT, STALIN, CHURCHILL AGREE ON PLANS FOR WAR ON GERMANY IN TALKS AT TEHERAN; 1,500 MORE TONS OF BOMBS DROPPED ON BERLIN

AGREEMENT: The news on the morning of Dec. 4, 1943.

Anglo-American Discussions in Cairo —Andaman Islands Plan—No Agreement at Our First Plenary Meeting, December 4—The President Agrees to Abandon Andamans Plan, December 5—President Roosevelt Decides to Appoint General Eisenhower to Command "Overlord"—The President and I Visit the Sphinx.

of both parties by the evening of Sunday, December 5.

I said that I did not wish to leave the Conference in any doubt that the British delegation viewed our early dispersal with great apprehension. There were still many questions of first-class importance to be settled. Two decisive events had taken place in the last few days. In the first place, Marshal Stalin had voluntarily proclaimed that the Soviet would declare war on Japan the moment Germany was defeated. This would give us better bases than we could ever find in China, and made it all the more important that we should concentrate on making "Overlord" a success. It would be necessary for the Staffs to examine how this new fact would affect operations in the Pacific and South-East Asia.

The second event of first-class importance was the decision to cross the Channel during May. I myself would have preferred a July date, but I was determined nevertheless to do all in my power to make a May date a complete success.

NEED FOR TEACHERS EXPECTED TO GROW

1,200 More a Year Required in State, Board Officials Say —Triple Sessions Feared

LACK IS WORST IN GRADES

Special Subjects Also Suffer—Syracuse Parley Cites High Birth Rate, Low Salaries

By LEONARD BUDER

Special to THE NEW YORK TIMES.

SYRACUSE, N. Y., Oct. 21—A growing shortage of teachers in the elementary grades and in specialized subjects is complicating the problems caused by the postwar increase in school enrollments, officers of the New York State School Boards Association said today at the organization's annual meeting here.

The state's school systems, which have never fully recovered from the wartime teacher shortage, will need 1,200 new teachers each year for the next five or six years, they declared. This figure, which exceeds the total of students expected to be graduated by teacher training institutions, does not include

The Match

- A new algorithm was adopted in 1997
    - Primary motivated was to give couples the option to get placed in geographically nearby programs
    - But in addition was made student-proposing

MONDAY, OCTOBER 22, 1951. The New York Times MONDAY, OCTOBER 22, 1951. 25 L+

MEDICAL SENIORS HIT INTERNE PLAN

Delegates of 44 Schools Meet Here to Protest Selection by 'Matching Machine'

TEACHERS PRAISE SYSTEM

They Argue That It Bars Unfair Recruiting—Students Insist on Choosing Their Hospitals

Delegates representing seniors in nearly all of the country's leading medical schools met here yesterday to express overwhelming opposition to a proposed mathematically contrived plan to place medical students in hospitals as internes.

They indicated that a great majority of their classmates preferred the present system whereby the country's hospitals, which have 10,000 internships, scramble for the best of a year's 6,000 medical graduates.

The meeting was held at Bard Hall of the College of Physicians and Surgeons of Columbia University, but it was made clear that the university was not its sponsor.

By Winston Churchill: *The Second World War*

Volume V—Closing the Ring

INSTALLMENT 15—TEHERAN: CONCLUSIONS

Book II—Teheran to Rome

Conversation at Luncheon, December 1—The Frontiers of Poland—The "Curzon Line", and the Line of the Oder — Finland — "No Annexations and No Indemnities"—The Question of Germany—Partition? —President Roosevelt's Suggestion—I Unfold a Personal View — Marshal Stalin's Standpoint — Broad Agreement on Military Policy — Political Aspects Remote and Speculative—Deep Fear of German Might at This War Climax—The Present Partition—"It Cannot Last".

SEVERAL of our gravest political issues stood out before and after the main decision on strategy had been reached [at the Teheran conference]. The Three lunched together again at the President's table in the Soviet Legation on December 1 [1943]. In addition on this occasion Molotov, Hopkins, Eden, Clark Kerr, and Harriman were present. The question of inducing Turkey to enter into the war was our first topic.

There was a very great measure of agreement on the limited steps for which I asked in order to win the great prize of bringing Turkey into the war.

* * *

Poland was the next important subject.

The New York Times.

ROOSEVELT, STALIN, CHURCHILL AGREE ON PLANS FOR WAR ON GERMANY IN TALKS AT TEHERAN; 1,500 MORE TONS OF BOMBS DROPPED ON BERLIN

AGREEMENT: The news on the morning of Dec. 4, 1943.

Anglo-American Discussions in Cairo—Andaman Islands Plan—No Agreement at Our First Plenary Meeting, December 4—The President Agrees to Abandon Andamans Plan, December 5—President Roosevelt Decides to Appoint General Eisenhower to Command "Overlord"—The President and I Visit the Sphinx.

of both parties by the evening of Sunday, December 5.

I said that I did not wish to leave the Conference in any doubt that the British delegation viewed our early dispersal with great apprehension. There were still many questions of first-class importance to be settled. Two decisive events had taken place in the last few days. In the first place, Marshal Stalin had voluntarily proclaimed that the Soviet would declare war on Japan the moment Germany was defeated. This would give us better bases than we could ever find in China, and made it all the more important that we should concentrate on making "Overlord" a success. It would be necessary for the Staffs to examine how this new fact would affect operations in the Pacific and South-East Asia.

The second event of first-class importance was the decision to cross the Channel during May. I myself would have preferred a July date, but I was determined nevertheless to do all in my power to make a May date a complete success.

NEED FOR TEACHERS EXPECTED TO GROW

1,200 More a Year Required in State, Board Officials Say—Triple Sessions Feared

LACK IS WORST IN GRADES

Special Subjects Also Suffer—Syracuse Parley Cites High Birth Rate, Low Salaries

By LEONARD BUDER

Special to THE NEW YORK TIMES.

SYRACUSE, N. Y., Oct. 21—A growing shortage of teachers in the elementary grades and in specialized subjects is complicating the problems caused by the postwar increase in school enrollments, officers of the New York State School Boards Association said today at the organization's annual meeting here.

The state's school systems, which have never fully recovered from the wartime teacher shortage, will need 1,200 new teachers each year for the next five or six years, they declared. This figure, which exceeds the total of students expected to be graduated by teacher training institutions, does not include

Nobel Prize in Economics

- In 1952, the National Resident Matching Program (NRMP) adopted the “Boston Pool” algorithm named after regional clearinghouses in Boston
- In 1962, David Gale and Lloyd Shapley formally analyzed a generalization of the Boston Pool algorithm
- Shapley & Roth (who extended his work) were awarded **the 2012 Nobel Prize in Economics** (Gale did not share the prize, because he died in 2008.)
- Read https://www.nobelprize.org/uploads/2018/06/popular-economicsciences2012-1.pdf

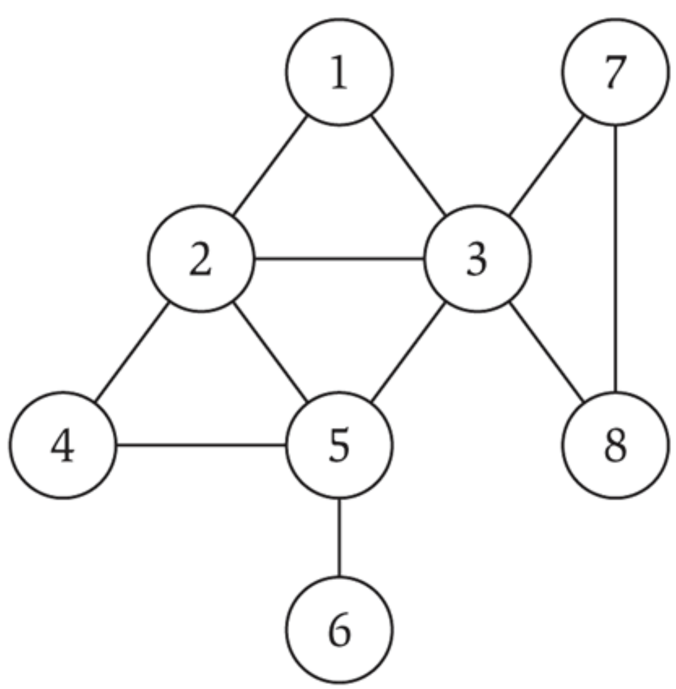
Graphs and Traversals

Review: Undirected Graphs

An undirected graph $G = (V, E)$

- V is the set of nodes, E is the set of edges
- Captures pairwise relations between objects
- Graph size parameters: $n = |V|$, $m = |E|$

Sometimes we consider weighted graphs, where each edge e has a weight $w(e)$



$V = \{ 1, 2, 3, 4, 5, 6, 7, 8 \}$

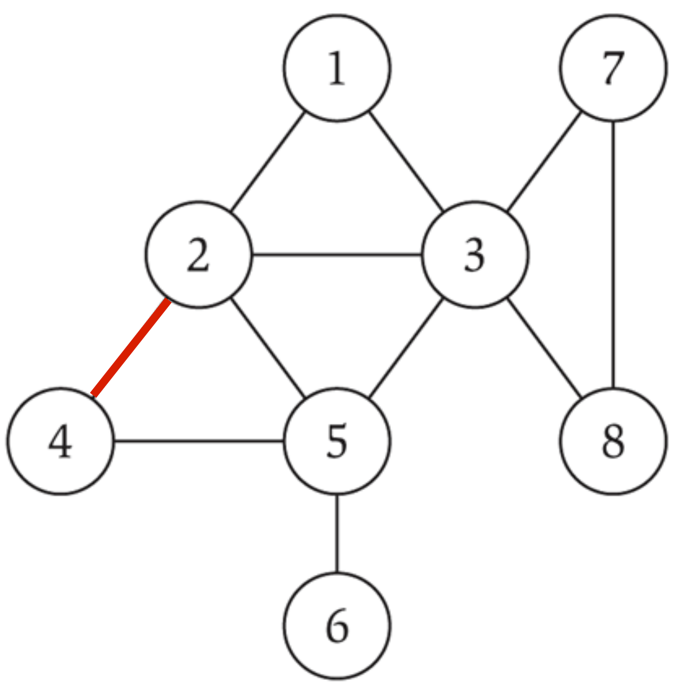
$E = \{ 1–2, 1–3, 2–3, 2–4, 2–5, 3–5, 3–7, 3–8, 4–5, 5–6, 7–8 \}$

$m = 11, n = 8$

Representing Graphs (Review)

Adjacency matrix.

- n-by-n matrix where $A[u][v] = 1$ if $(u, v) \in E$
- Space $O(n^2)$
- Checking if $(u, v) \in E$ takes _____ time?

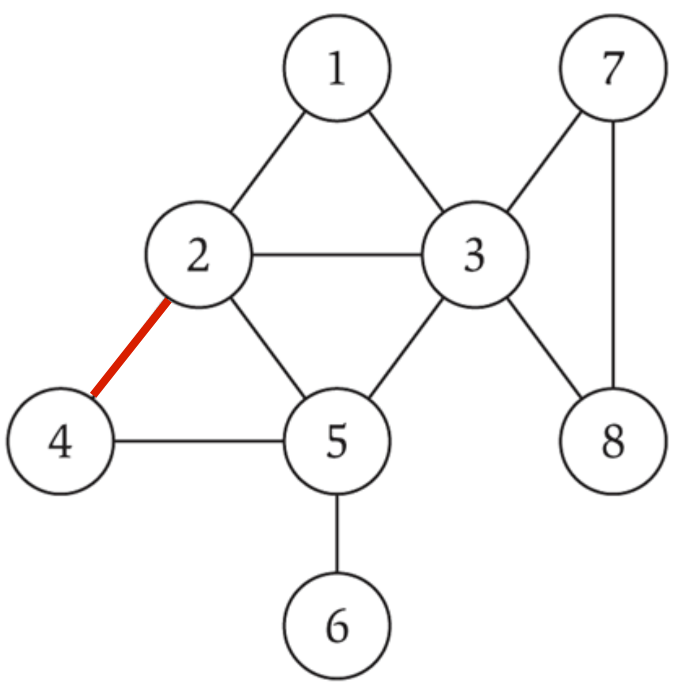


	1	2	3	4	5	6	7	8
1	0	1	1	0	0	0	0	0
2	1	0	1	1	1	0	0	0
3	1	1	0	0	1	0	1	1
4	0	1	0	0	1	0	0	0
5	0	1	1	1	0	1	0	0
6	0	0	0	0	1	0	0	0
7	0	0	1	0	0	0	0	1
8	0	0	1	0	0	0	1	0

Representing Graphs (Review)

Adjacency matrix.

- n-by-n matrix where $A[u][v] = 1$ if $(u, v) \in E$
- Space $O(n^2)$
- Checking if $(u, v) \in E$ takes $O(1)$ time

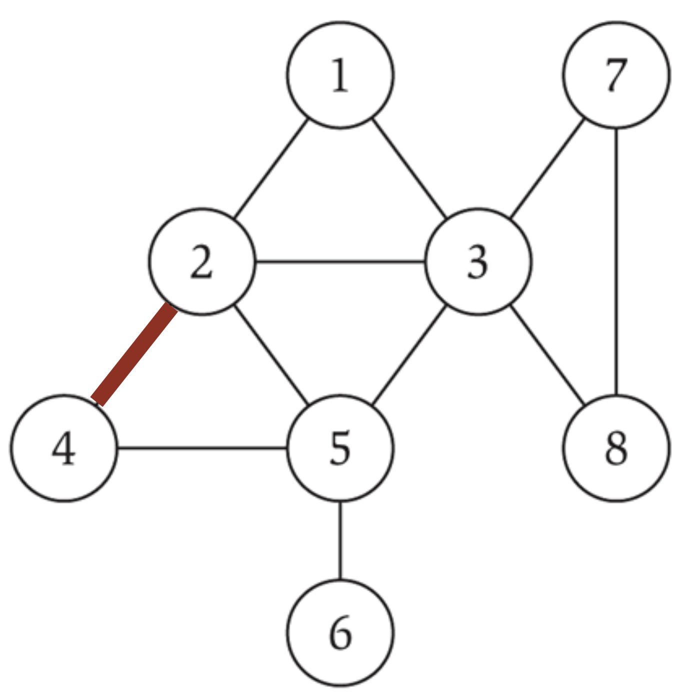


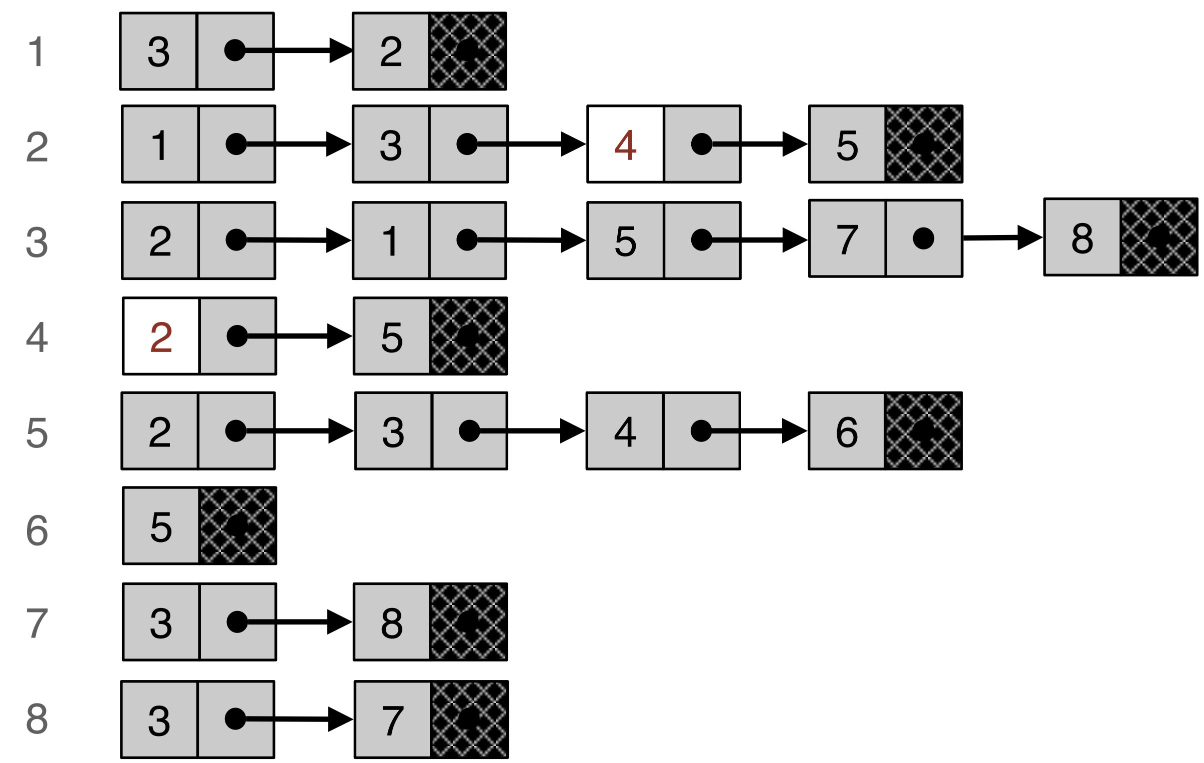
	1	2	3	4	5	6	7	8
1	0	1	1	0	0	0	0	0
2	1	0	1	1	1	0	0	0
3	1	1	0	0	1	0	1	1
4	0	1	0	0	1	0	0	0
5	0	1	1	1	0	1	0	0
6	0	0	0	0	1	0	0	0
7	0	0	1	0	0	0	0	1
8	0	0	1	0	0	0	1	0

Representing Graphs (Review)

Adjacency list.

- Array of lists, where each list represents the neighbors of a given node
- Space $O(n + m)$
- Checking if $(u, v) \in E$ takes _____ time?

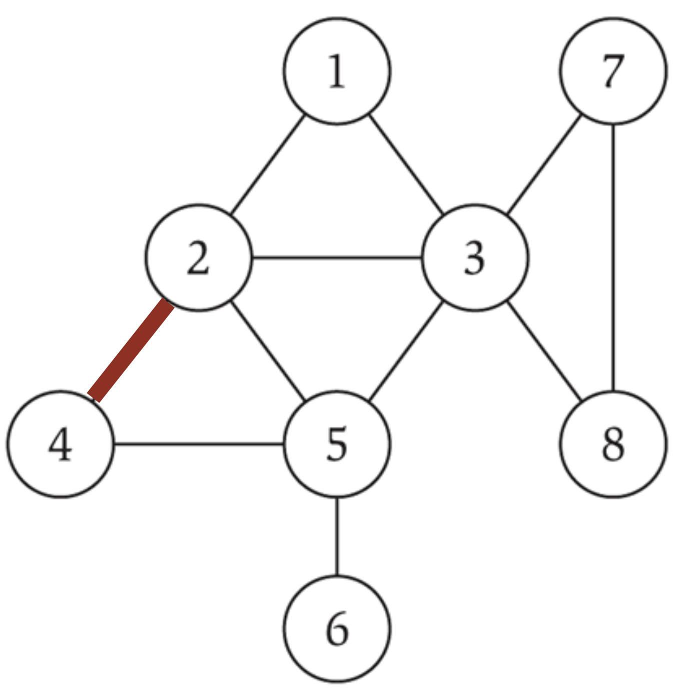


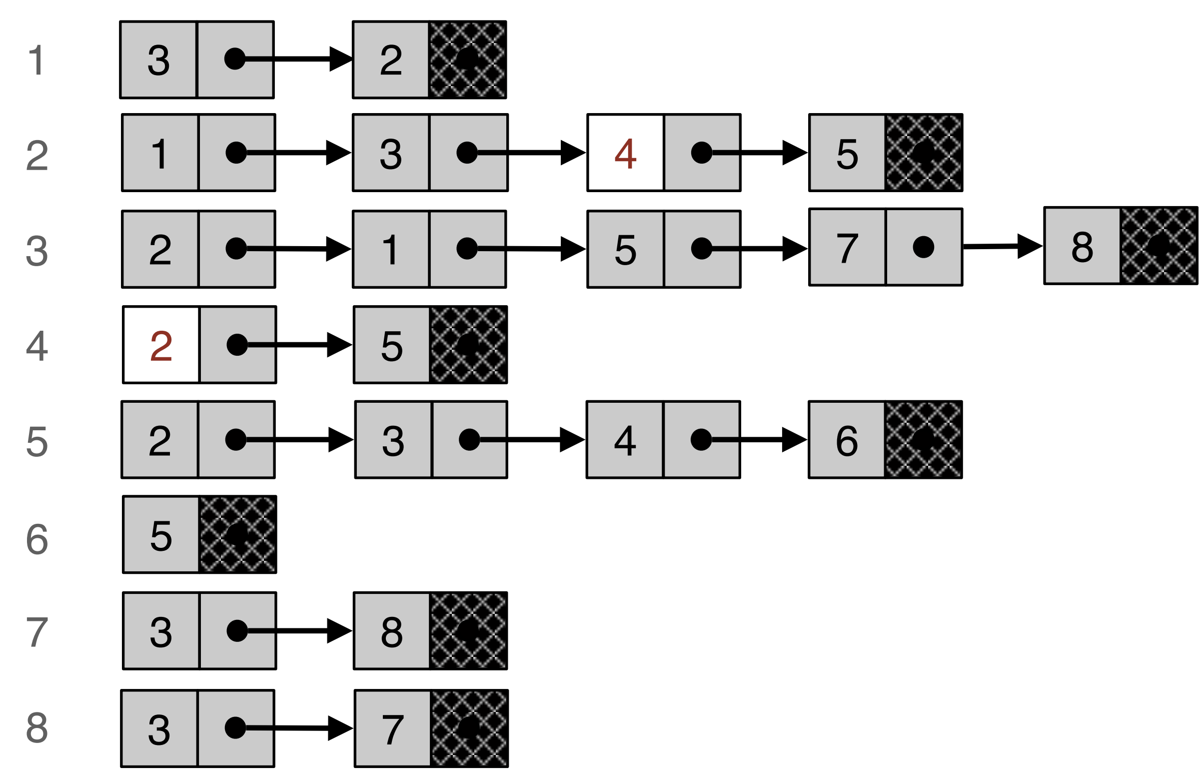


Representing Graphs (Review)

Adjacency list.

- Array of lists, where each list represents the neighbors of a given node
- Space $O(n + m)$
- Checking if $(u, v) \in E$ takes $O(\text{degree}(u))$ time





Graph Terminology

- A **path** in an undirected graph $G = (V, E)$ is a sequence of nodes $u_1, u_2, \ldots, u_k$ such that every pair $(u_{i-1}, u_i) \in E$.
- A path is **simple** if all nodes are distinct.
- The **length** of a path is **the number of edges on the** path

- An undirected graph is **connected** if for every pair of nodes u and v, there is a path between u and v

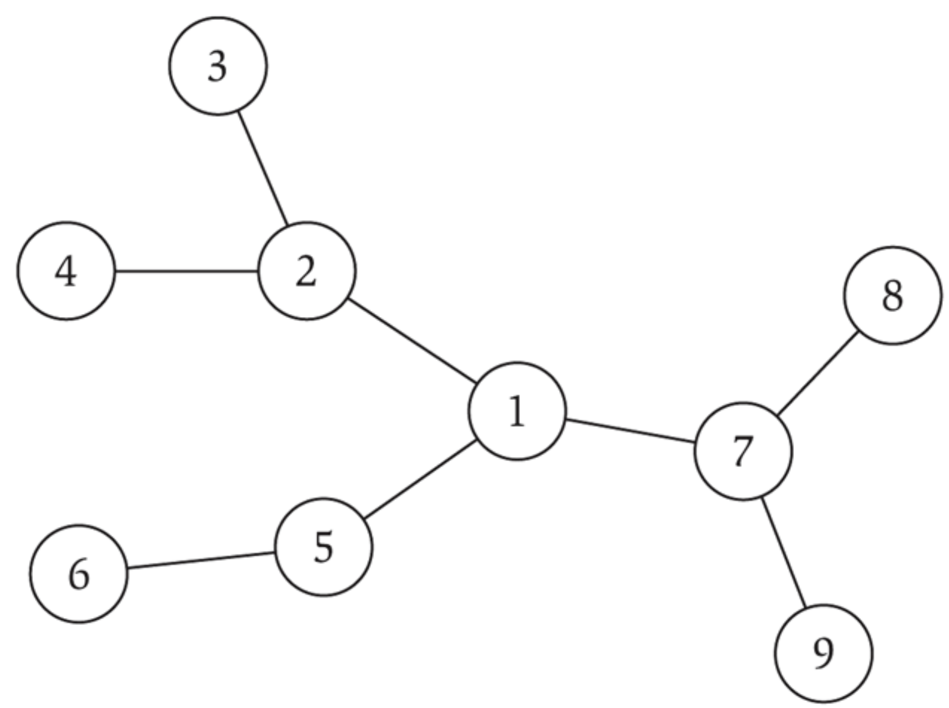
- A **cycle** is path $u_1, u_2, \ldots, u_k$ where $u_1 = u_k$ $(k \geq 2)$
- A cycle is **simple** if all internal nodes are distinct

Trees

- An undirected graph is a tree if it is connected and does not contain a cycle

Lemma. Let G be an undirected graph with n nodes. Then any two of these conditions imply the third

- G is connected
- G does not contain a cycle
- G has $n - 1$ edges

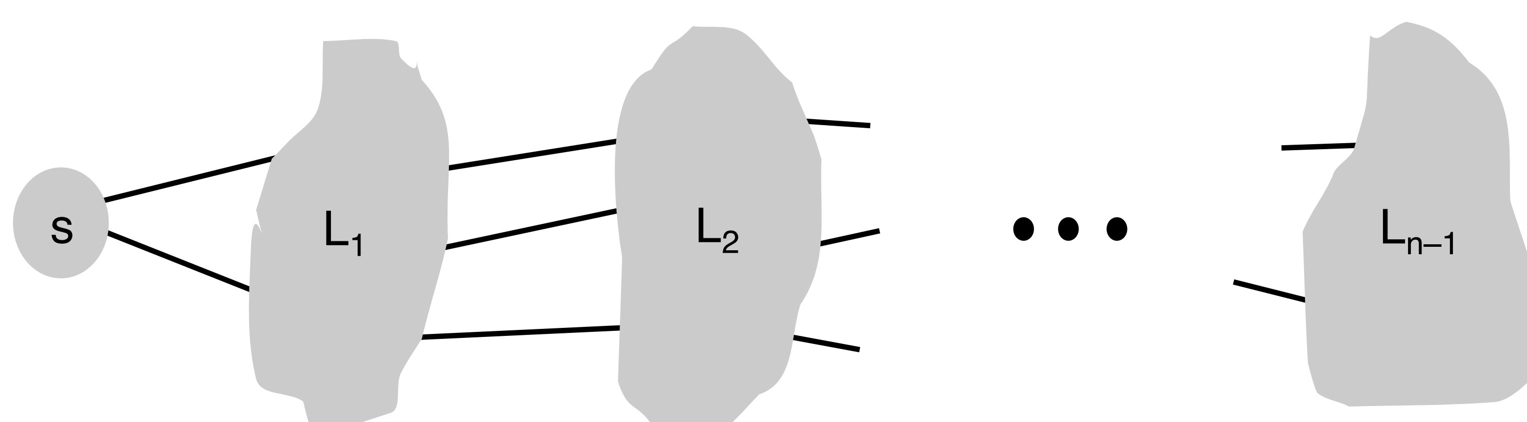


Graph Traversals

- **Connectivity.** How do we verify if a graph is connected?
- **Path.** Given $s, t \in V$, is there a path between them?
- Determined by "traversing the graph"
- Two classic graph traversal algorithms:
  - Breadth-first search (BFS)
  - Depth-first search (DFS)
  - Both have different applications
    - Bipartite testing (BFS)
    - Topological ordering (DFS), etc

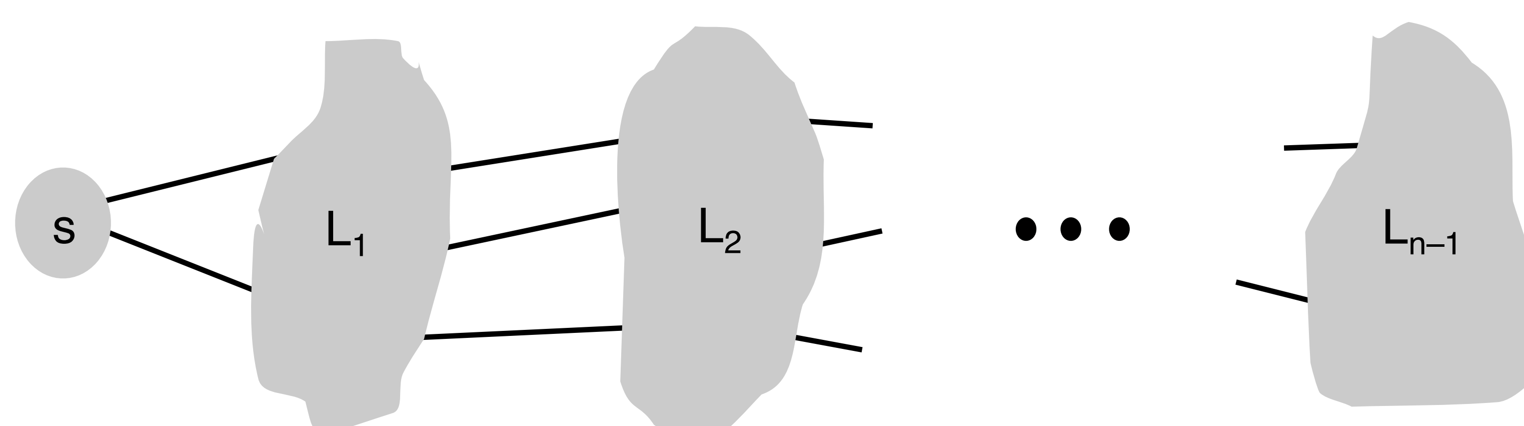
Breadth-first Search

- Explore outwards in all possible direction from starting point, peeling "one layer after another"
- BFS algorithm: Initialize $L_0 = \{s\}$
  - $L_1 =$ all neighbors of L_0
  - $L_2 =$ all nodes that do not belong to L_0 or L_1 that are adjacent to a node in L_1
  - …
  - $L_{i+1} =$ all nodes that do not belong an earlier layer that are adjacent to a node in L_i



BFS Implementation

- Nodes that we have not seen yet
- Nodes that we have visited
- Nodes that have been “explored” (visited all its neighbors as well)
    - Suppose we are currently exploring u
    - Its neighbors will be marked but when should they be explored compared to other marked unexplored nodes?
    - Want to explore all nodes at level i before moving on to level $i+1$ (first visited is first to be explored)
    - Which data structure?
        - Queue

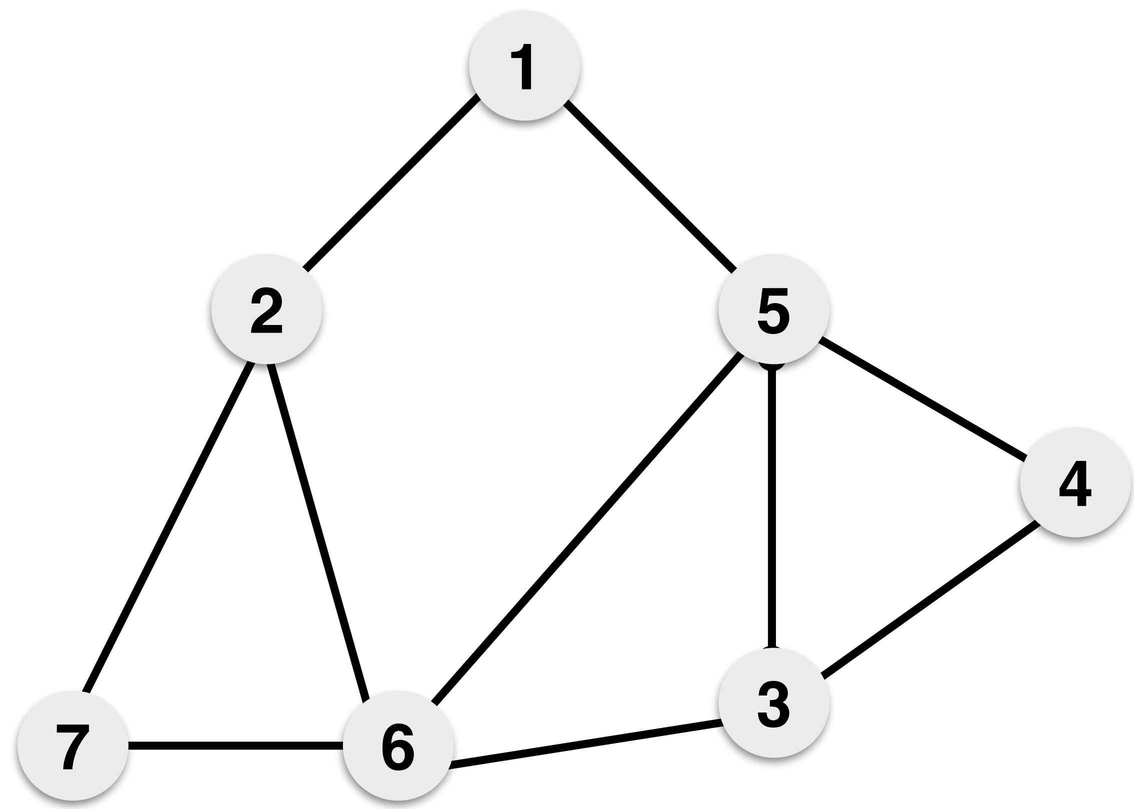


BFS Implementation: Queue

- Nodes that we have not seen yet (never been added to queue)
- Nodes that we have visited (added to queue but not marked)
- When a node is marked (after extraction from queue), all its neighbors are visited: next time we see it we can ignore it —-its been explored!

```
BFS (G, s):
 Put s in the queue Q
 While Q is not empty
   Extract v from Q
     If v is unmarked
       Mark v
       For each edge (v, w):
          Put w into the queue Q
```

BFS Example



The BFS Tree

- Can remember parent nodes (the node at level i that lead us to a given node at level $i+1$)

```
BFS-Tree(G, s):
 Put (∅, s) in the queue Q
 While Q is not empty
   Extract (p, v) from Q
     If v is unmarked
       Mark v
       parent(v) = p
       For each edge (v, w):
          Put (v, w) into the queue Q      (*)
```

BFS Analysis

- Inserting and extracting an edge from a queue
  - $O(1)$ time
- For each marked node v, we run the for loop for its edges: $O(n)$
- Overall running time?
  - Easy to prove $O(n^2)$ time
- Can improve the analysis to $O(n+m)$
  - Node u has degree(u) incident edges (u, v)
  - Total time processing edges: $\sum_{u \in V} \text{degree(u)} = 2m$

each edge (u, v) is counted exactly twice

in sum: once in degree(u) and once in degree(v)