Suppose $f$ is $O(g)$. Which of the following is true?

A. $g$ is $O(f)$
B. $g$ is not $O(f)$
C. $g$ may be $O(f)$, depending on the particular functions $f$ and $g$

Which is an equivalent definition of big Omega notation?

A. $T(n)$ is $\Omega(f(n))$ if $f(n)$ is $O(T(n))$
B. $T(n)$ is $\Omega(f(n))$ if there exists a constant $c > 0$ such that $T(n) \geq c \cdot f(n)$ for infinitely many $n$
C. Both A and B
D. Neither A nor B
Exercise: let $T(n)$ be the running time of \texttt{sum-product}. Show that $T(n)$ is $\Omega(n^2)$

Algorithm \texttt{sum-product}
\begin{verbatim}
sum = 0 
for i = 1 to n do 
  for j = i to n do 
    sum += A[i]*A[j] 
  end for 
end for
\end{verbatim}

Solution

Hard way
\begin{itemize}
  \item Count exactly how many times the loop executes
  \end{itemize}
\begin{equation}
1 + 2 + \cdots + n = \frac{n(n+1)}{2} = \Omega(n^2)
\end{equation}

Easy way
\begin{itemize}
  \item Ignore all loop executions where $i > n/2$ or $j < n/2$
  \item The inner statement executes at least $(n/2)^2 = \Omega(n^2)$ times
\end{itemize}

Definition: the function $T(n)$ is $\Theta(f(n))$ if it is both $O(f(n))$ and $\Omega(f(n))$.

Example. $T(n) = 32n^2 + 17n + 1$
\begin{itemize}
  \item $T(n)$ is $\Theta(n^2)$
  \item $T(n)$ is neither $\Theta(n)$ nor $\Theta(n^3)$
\end{itemize}

Additivity Revisited

Suppose $f$ and $g$ are two (non-negative) functions and $f$ is $O(g)$

Old version: Then $f + g$ is $O(g)$

New version: Then $f + g$ is $\Theta(g)$

$$\frac{n^2}{g} + 42n + n \log n \text{ is } \Theta(n^2)$$

Running Time Analysis

Mathematical analysis of \textit{worst-case} running time of an algorithm as function of input size. Why these choices?
\begin{itemize}
  \item \textbf{Mathematical}: describes the \textit{algorithm}. Avoids hard-to-control experimental factors (CPU, programming language, quality of implementation), while still being predictive.
  \item \textbf{Worst-case}: just works. (“average case” appealing, but hard to analyze)
  \item \textbf{Function of input size}: allows predictions. What will happen on a new input?
**Efficiency**

When is an algorithm efficient?

Stable Matching Brute force: \( \Omega(n!) \)
Propose-and-Reject?: \( O(n^2) \)
We must have done something clever

**Polynomial Time**

*Definition:* an algorithm runs in polynomial time if its running time is \( O(n^d) \) for some constant \( d \)

**Polynomial Time: Examples**

These are polynomial time:

- \( f_1(n) = n \)
- \( f_2(n) = 4n + 100 \)
- \( f_3(n) = n \log(n) + 2n + 20 \)
- \( f_4(n) = 0.01n^2 \)
- \( f_5(n) = n^2 \)
- \( f_6(n) = 20n^2 + 2n + 3 \)

Not polynomial time:

- \( f_7(n) = 2^n \)
- \( f_8(n) = 3^n \)
- \( f_9(n) = n! \)

**Why Polynomial Time?**

Why is this a good definition of efficiency?

- Matches practice: almost all practically efficient algorithms have this property.
- Usually distinguishes a clever algorithm from a “brute force” approach.
- Refutable: gives us a way of saying an algorithm is not efficient, or that no efficient algorithm exists.

**Graphs: Motivation**

- Shortest driving route from Amherst to Florida?
- Number of “degrees of separation” between you and Tom Brady (or Theresa May) in online social network?
- Find influencers and bots on twitter?
- Find reputable web pages?

  How do we build algorithms to answer these questions?

Graphs and graph algorithms.

**Networks**
2.2. PATHS AND CONNECTIVITY

Figure 2.2: A network depicting the sites on the Internet, then known as the Arpanet, in December 1970.

In fact, every edge in the 1970 Arpanet belongs to a cycle, and this was by design: it means that if any edge were to fail (e.g. a construction crew accidentally cut through the cable), the entire network would still be connected. This particular design of a network is known as a ring network.

Theoretic notions have been studied; the social scientist John Barnes once described graph theory as “a language the scientists can use.” Perhaps because graphs are so simple to define and work with, an enormous range of graph-theoretic notions have been studied; the social scientist John Barnes once described graph theory as “a language the scientists can use.”

We now turn to some of the fundamental concepts and definitions surrounding graphs. Per- haps because graphs are so simple to define and work with, an enormous range of graph-theoretic notions have been studied; the social scientist John Barnes once described graph theory as “a language the scientists can use.”

There are many cycles in Figure 2.3: a path with at least three edges, in which the first and last nodes are the same, but otherwise all nodes are distinct. A cycle is an example of a graph that is a “ring” structure such as the sequence of nodes.“liberal blogs link to another blog, while only 67% are linked to by another blog. In fact, 91% of the links originating within either the conservative or liberal blogospheres originate on the same blog. Conserva- tives also tend to cite other conservatives slightly more often than liberals tend to cite other liberals. The size of each blog reflects the number of other blogs it cites; red for conservative, blue for liberal. Orange links indicate a link between blogs that are of similar partisanship. More links than expected come from liberal to conservative blogs, while many others simply link in the opposite direction. The one blog that links to a large number of blogs is CNN.com, which informally is a “hub” in the graph. As we have defined it here, a path can repeat nodes: for example, the path that visits ::lincoln, mit, utah, sri, ucsb:: is as short an example as possible. In fact, every edge in the 1970 Arpanet belongs to a cycle, and this was by design: it means that if any edge were to fail (e.g. a construction crew accidentally cut through the cable), the entire network would still be connected. This particular design of a network is known as a ring network.

There are many cycles in Figure 2.3: a path with at least three edges, in which the first and last nodes are the same, but otherwise all nodes are distinct. A cycle is a path that visits a sequence of edges linking these nodes. For example, the sequence of nodes “::lincoln, mit, utah, sri, ucsb::” is as short an example as possible. In fact, every edge in the 1970 Arpanet belongs to a cycle, and this was by design: it means that if any edge were to fail (e.g. a construction crew accidentally cut through the cable), the entire network would still be connected. This particular design of a network is known as a ring network.

Figure 2.3: An alternate drawing of the 13-node Internet graph from December 1970.

Example: Internet in 1970

Since posts usually contain sparse references to other blogs, and blogrolls usually contain links to blogs that are different from the one we are currently on, we started with a balanced set of blogs, conservative blogs show a greater tendency to link to other conservative blogs than liberal blogs show to link to other liberal blogs. The size of each blog reflects the number of other blogs it cites; red for conservative, blue for liberal. Orange links indicate a link between blogs that are of similar partisanship. More links than expected come from liberal to conservative blogs, while many others simply link in the opposite direction. The one blog that links to a large number of blogs is CNN.com, which informally is a “hub” in the graph.

The size of each blog reflects the number of other blogs it cites; red for conservative, blue for liberal. Orange links indicate a link between blogs that are of similar partisanship. More links than expected come from liberal to conservative blogs, while many others simply link in the opposite direction. The one blog that links to a large number of blogs is CNN.com, which informally is a “hub” in the graph.

Graphs

A graph is a mathematical representation of a network.

- Set of nodes (vertices) $V$
- Set of pairs of nodes (edges) $E$

Graph $G = (V, E)$

**Notation:** $n = |V|$, $m = |E|$ (almost always)

Example: Internet in 1970

**Definitions:**

- **Edge** $e = \{u, v\}$ — but usually written $e = (u, v)$
- $u$ and $v$ are neighbors, adjacent, endpoints of $e$
- $e$ is incident to $u$ and $v$
Definitions:
A path is a sequence $P = v_1, v_2, \ldots, v_{k-1}, v_k$ such that each consecutive pair $v_i, v_{i+1}$ is joined by an edge in $G$

Path “from $v_1$ to $v_k$”. A $v_1$-$v_k$ path

Example: Internet in 1970

Definitions: Simple path, cycle, distance

Example: Internet in 1970

Definitions: Connected graph = graph with paths between every pair of vertices.

Definitions:
- **Simple path**: path where all vertices are distinct
- **(Simple) Cycle**: path $v_1, \ldots, v_{k-1}, v_k$ where
  - $v_1 = v_k$
  - First $k - 1$ nodes distinct
  - All edges distinct ($k > 3$)
- **Distance** from $u$ to $v$: minimum number of edges in a $u$-$v$ path

Definitions:
- **Connected component**: maximal subset of nodes such that a path exists between each pair in the set
  - **maximal** = if a new node is added to the set, there will no longer be a path between each pair

Example: Internet in 1970

Definitions:
Q: Which is not a path?
A. UCSB - SRI - UTAH
B. LINC - MIT - LINC - CASE
C. UCSB - SRI - STAN - UCLA - UCSB
D. None of the above
Definitions

**Tree**: a connected graph with no cycles

- Q: Is this equivalent to trees you saw in Data Structures?
- A: More or less.

**Rooted tree**: tree with parent-child relationship
- Pick root \( r \) and “orient” all edges away from root
- Parent of \( v \) = predecessor on path from \( r \) to \( v \)

Directed Graphs

Graphs can be *directed*, which means that edges point *from* one node to another, to encode an asymmetric relationship. We’ll talk more about directed graphs later.

Graphs are *undirected* if not otherwise specified.

Graph Traversal

Thought experiment. World social graph.
- Is it connected?
- If not, how big is largest connected component?
- Is there a path between you and Tom Brady? What about Theresa May?

How can you tell algorithmically?

Answer: graph traversal! (BFS/DFS)