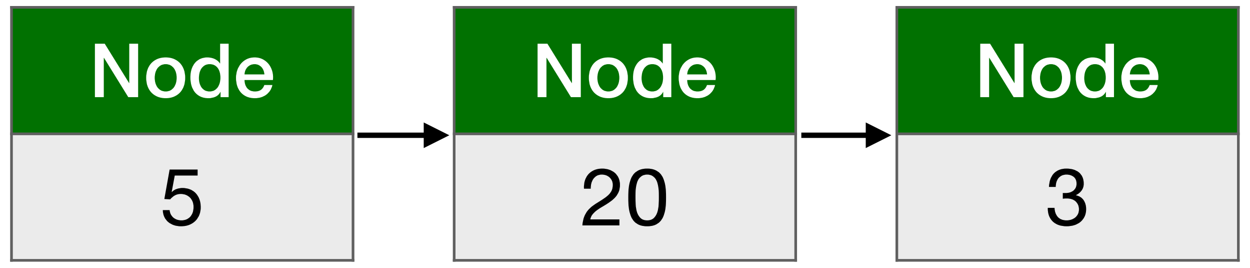
Graphs

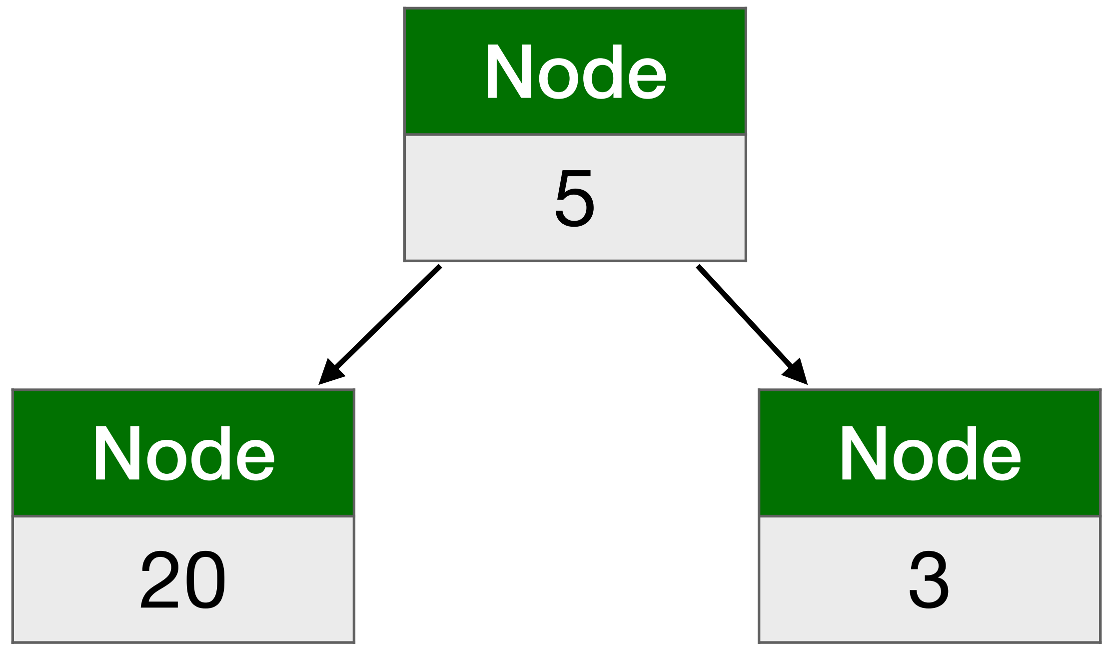
Data Structures: Review

- Sequential Data Structures
  - Elements stored in a specific order
  - Ex: Array, List
- Key-Value Store
  - Stores pairs of elements with no particular order
  - Each key is associated with one value
  - Ex. Map, Dictionary, Object
- Tree
  - Non-linear structure
  - Each element can be associated with multiple other elements

Index	0	1	2
Value	5	20	3

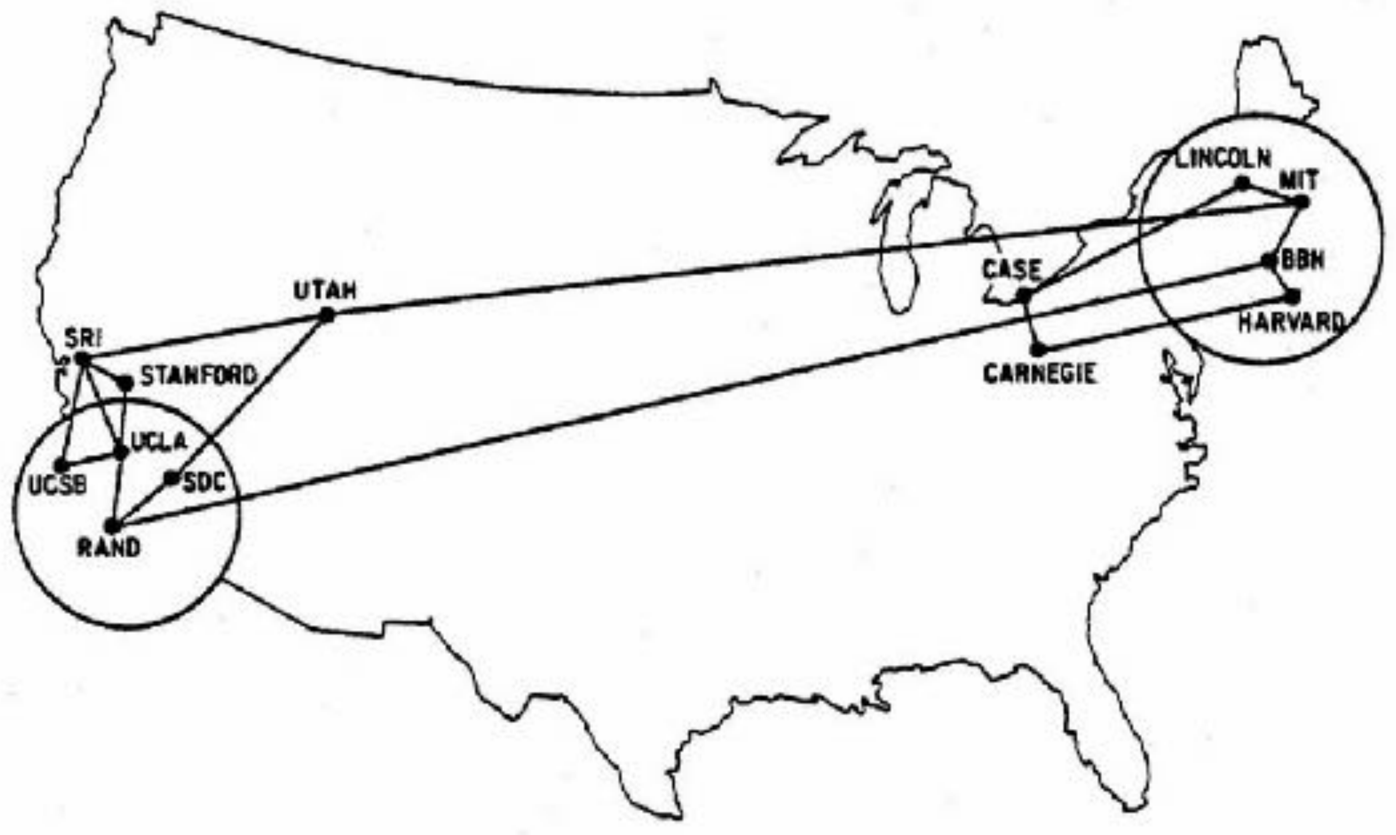


Key	"cse"	"mga"	"geo"
Value	20	3	5



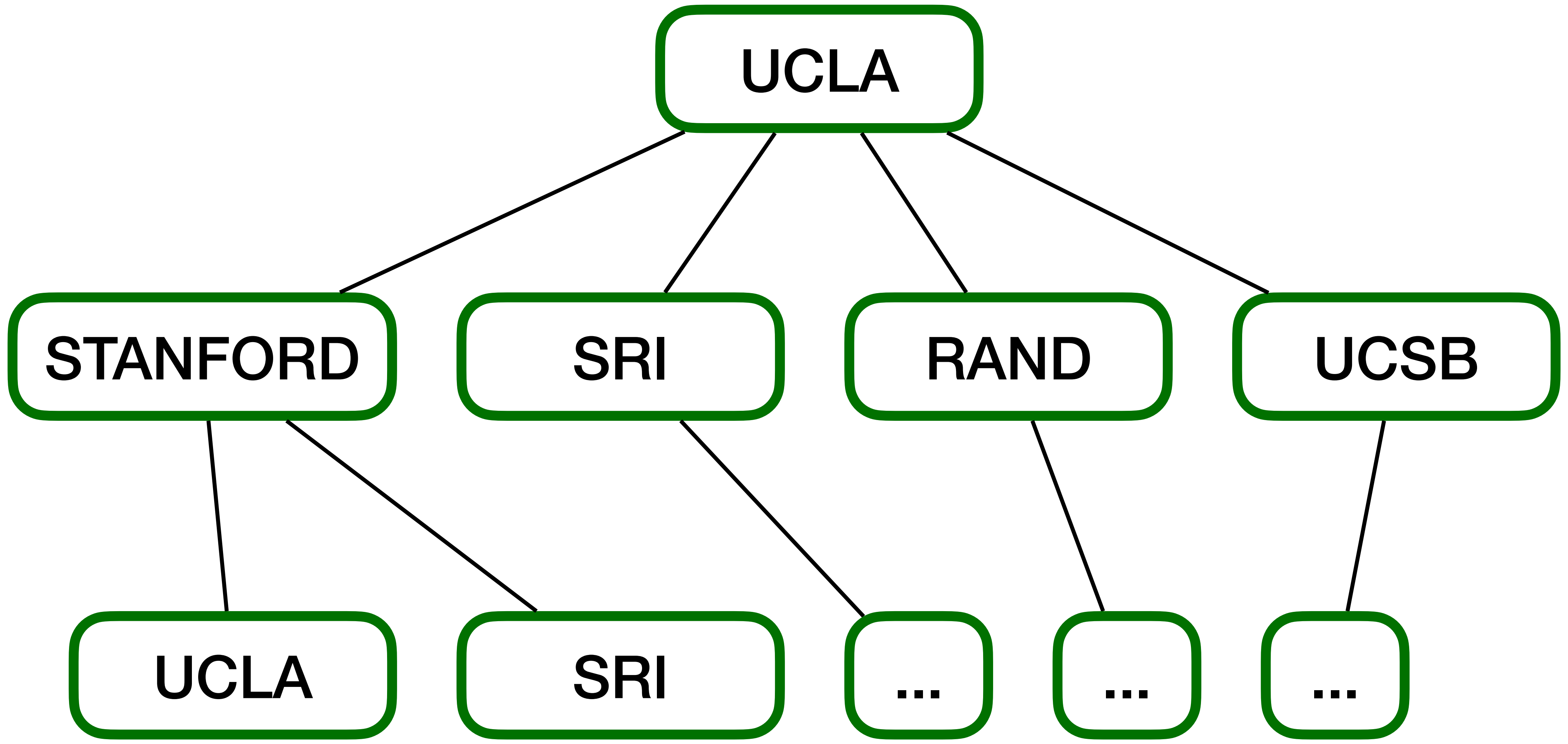
Data Structures

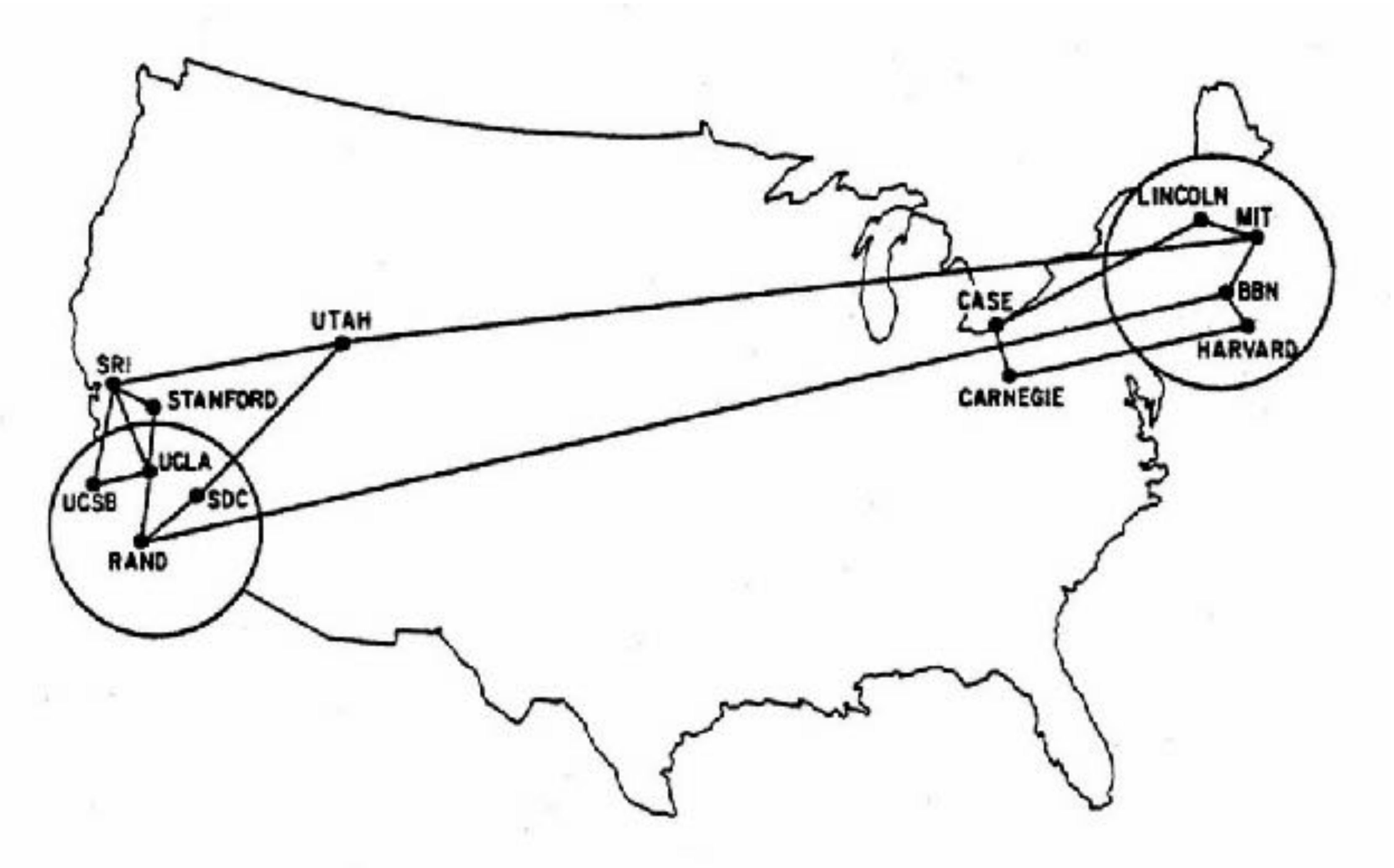
- How do we store data with multiple interconnected associations?
- A [station, intersection, city] can have multiple connections



Data Structures

- Let's use trees
- Start with UCLA as the root
- Recursively add all children

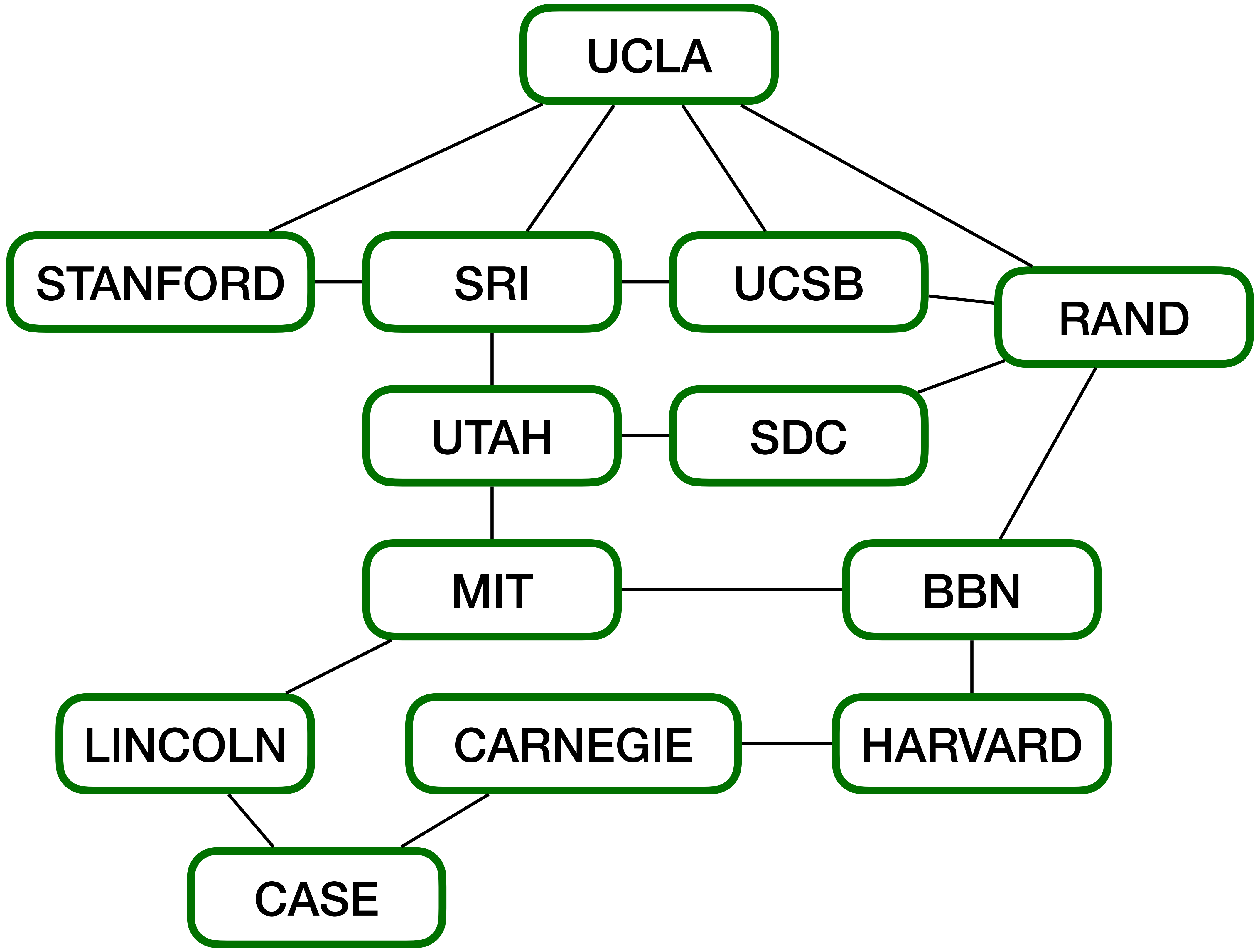


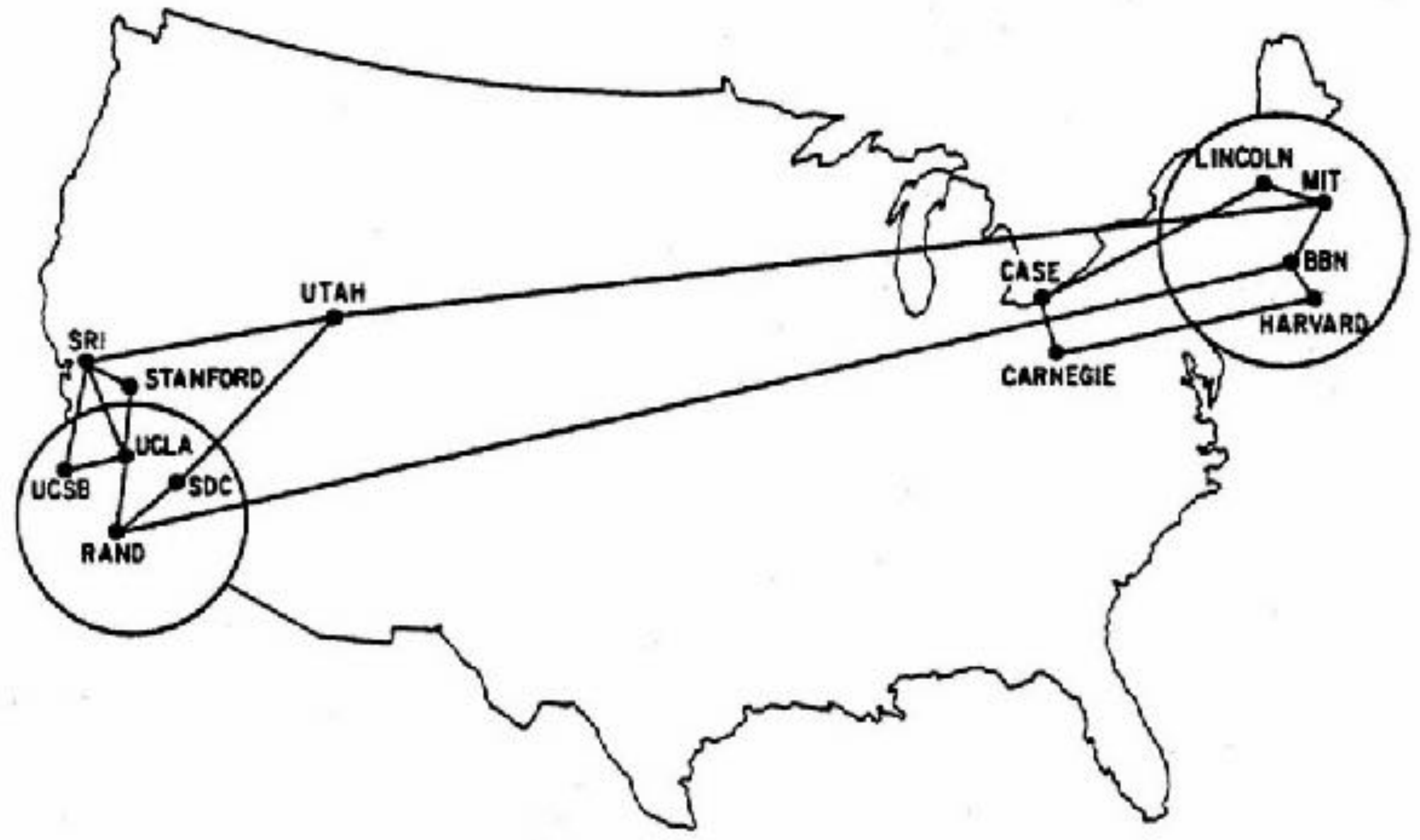


- Oops
    - We have duplicates in our data structure

Data Structures

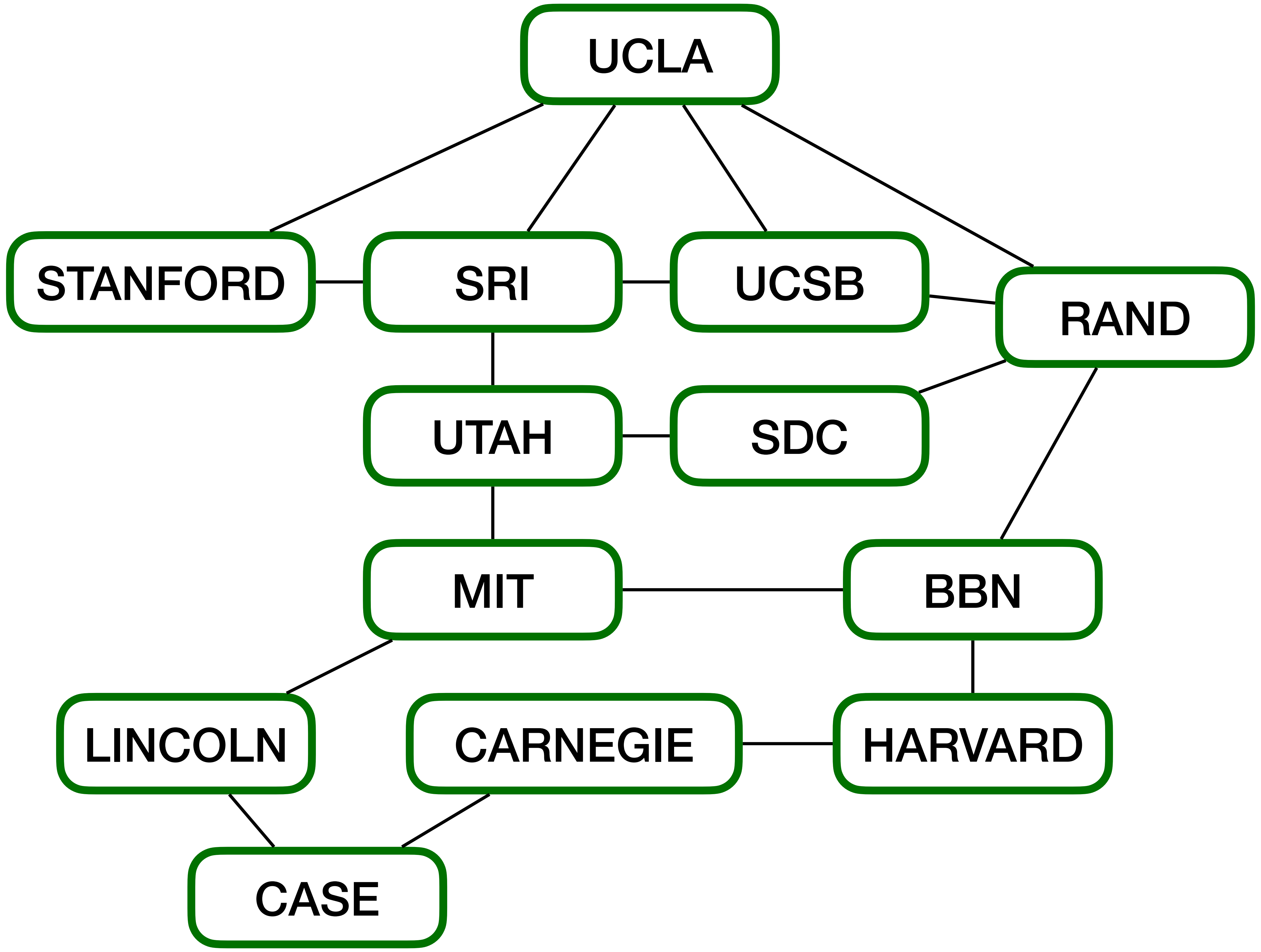
- Let's try again
  - When we try to add a duplicate, add a reference to the existing node

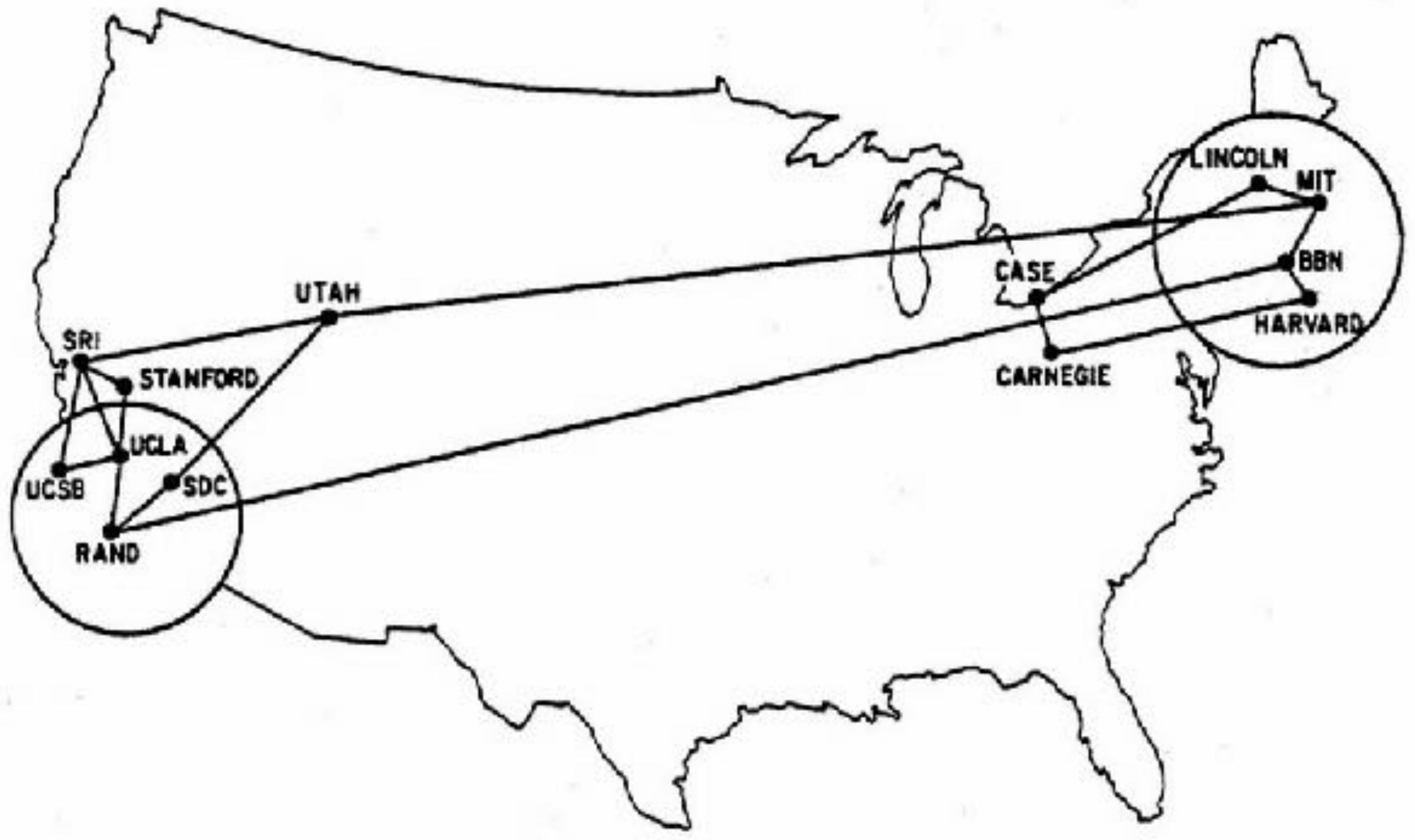




Graphs

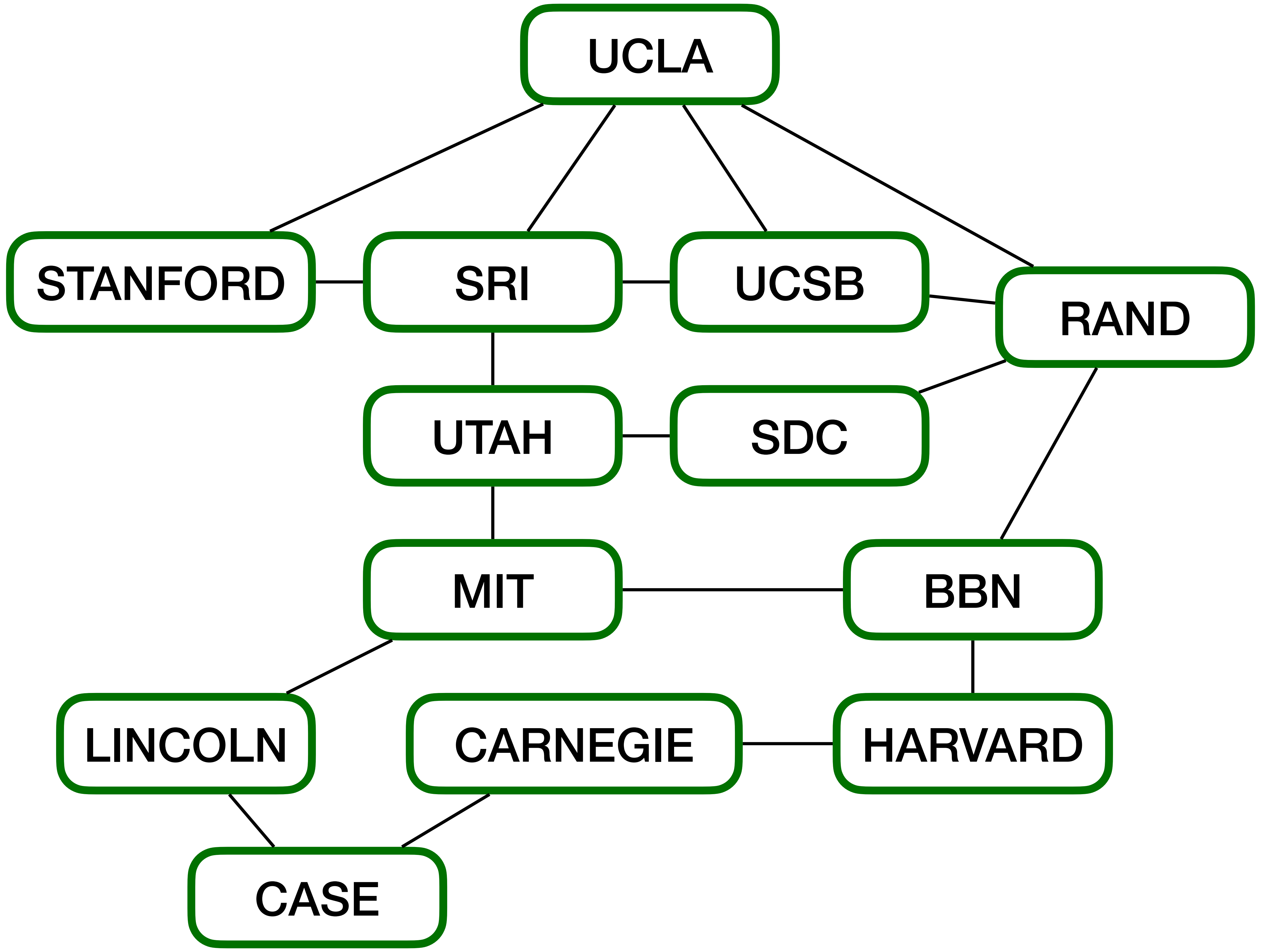
- This is a graph
- Similar to a tree, except cycles are allowed
  - Cycle: Can "travel" from a node back to itself without repeating a node

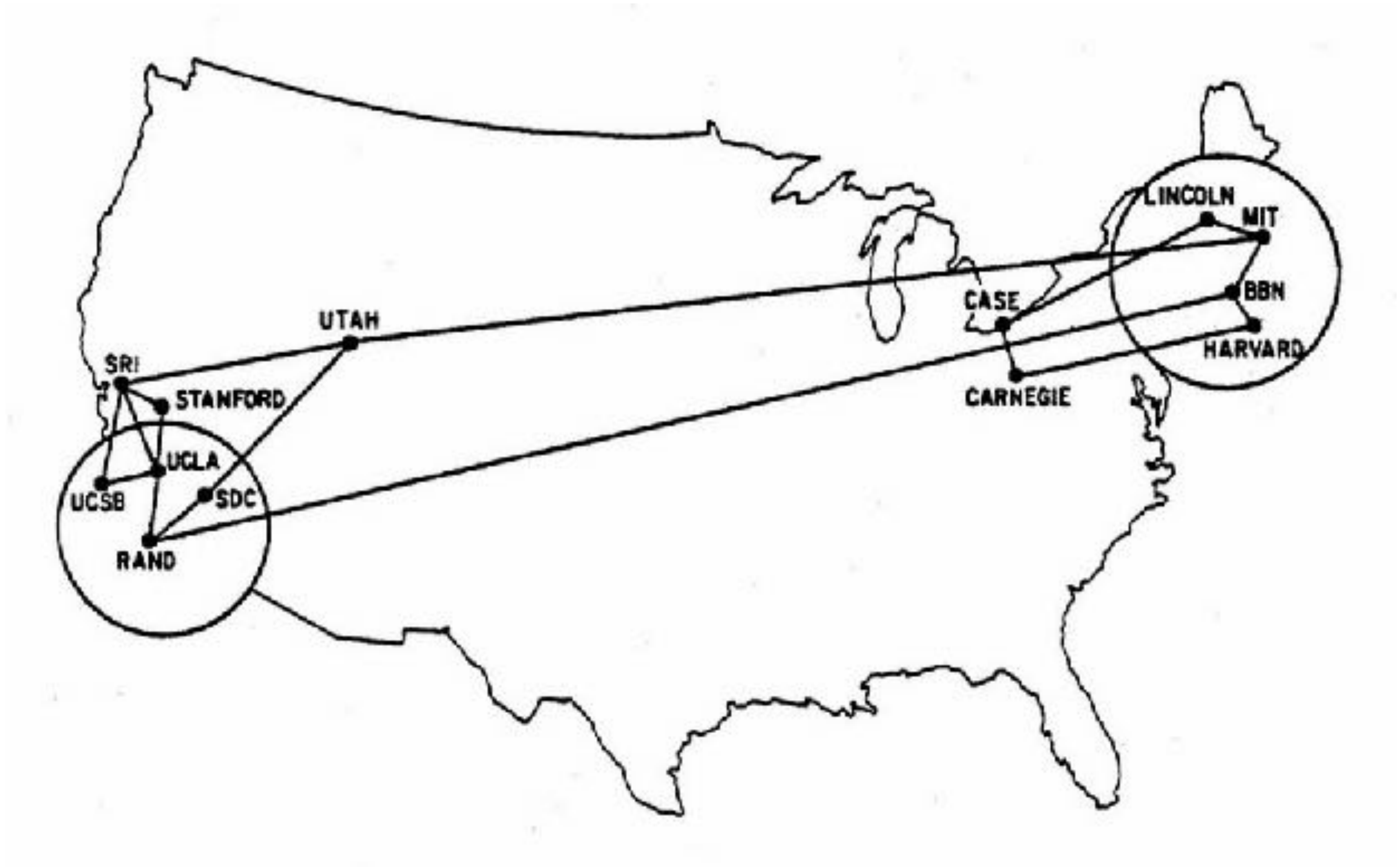




Graphs

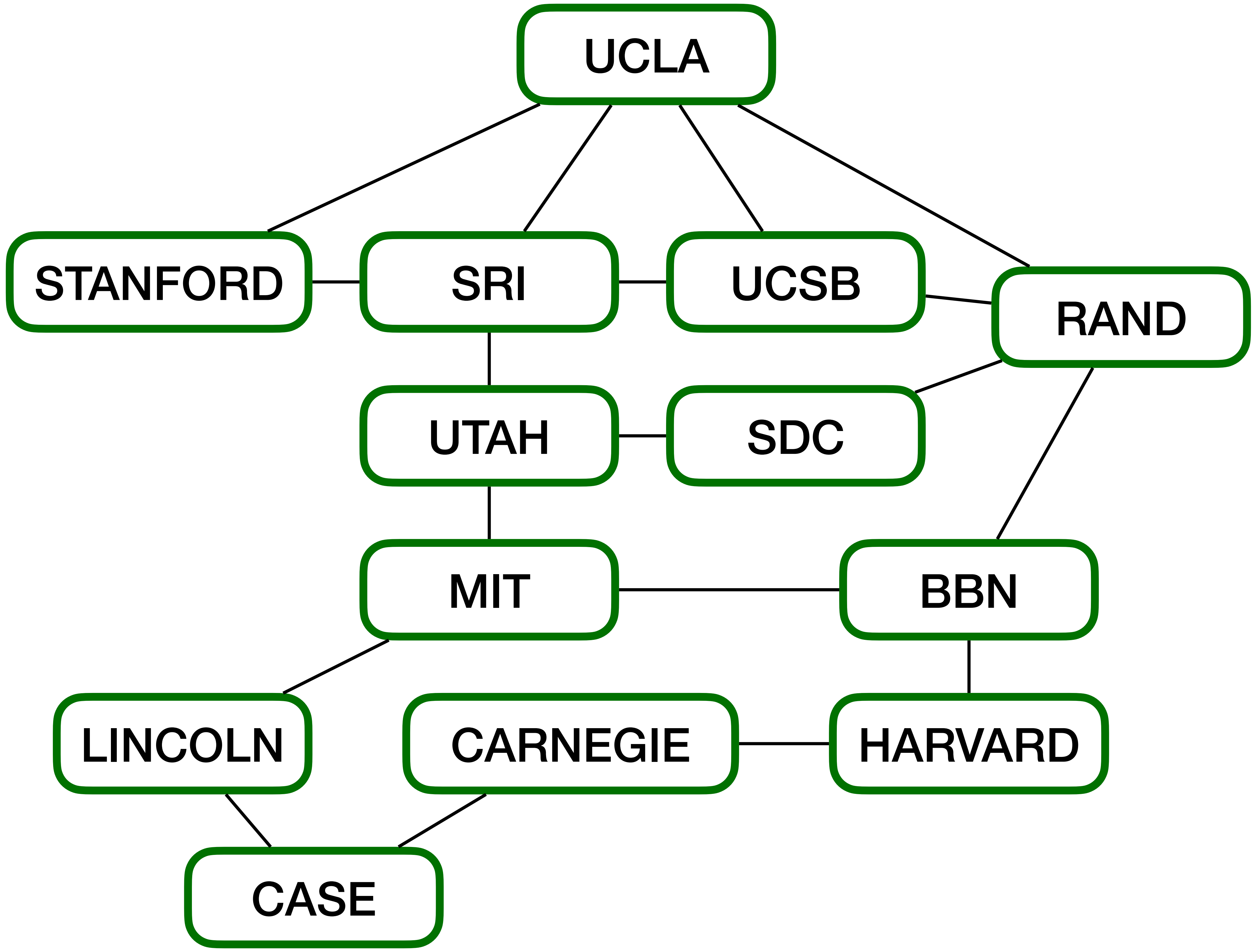
- Because of the cycles, our tree traversals will get stuck in infinite recursion
  - No leaves (node with no children) to terminate the recursion

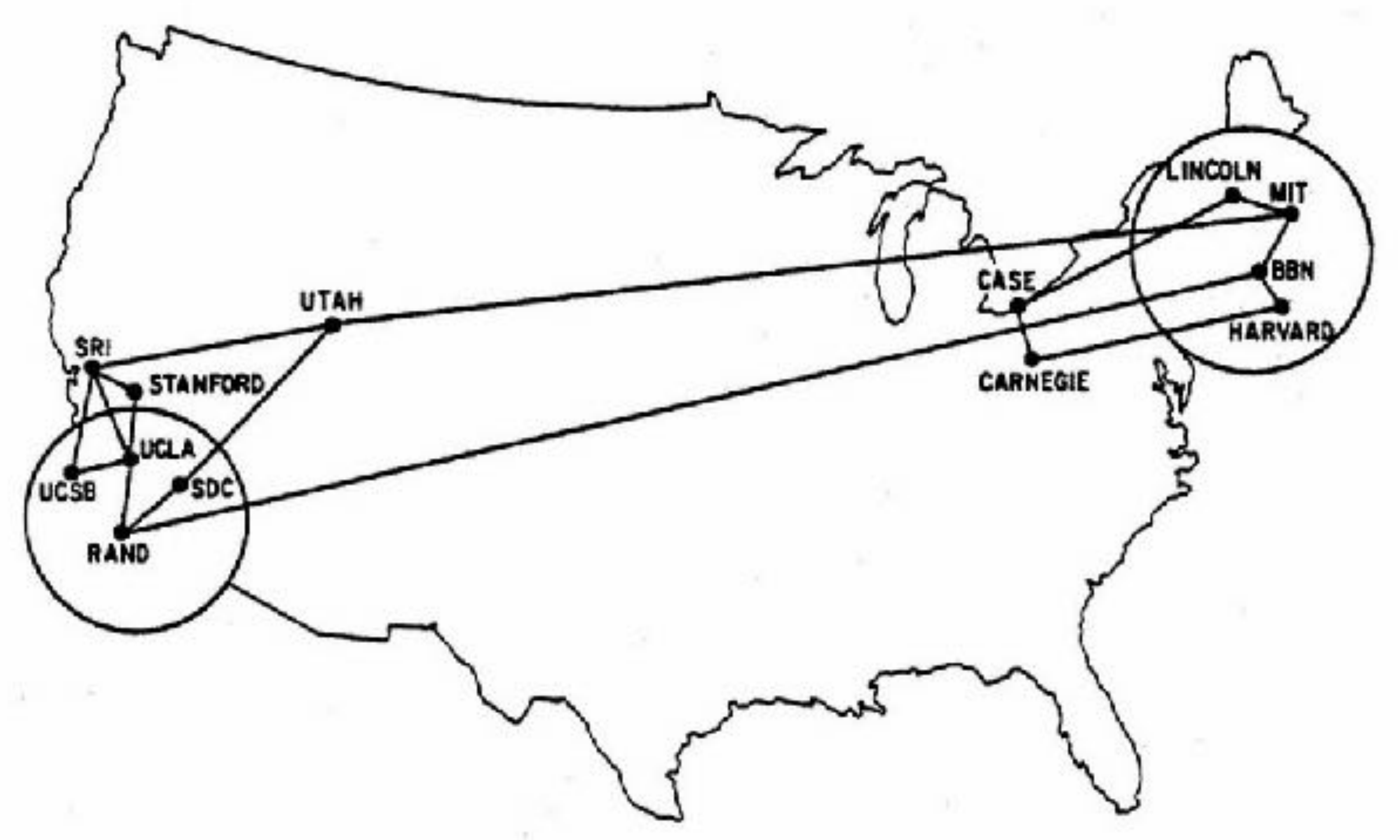




Graphs

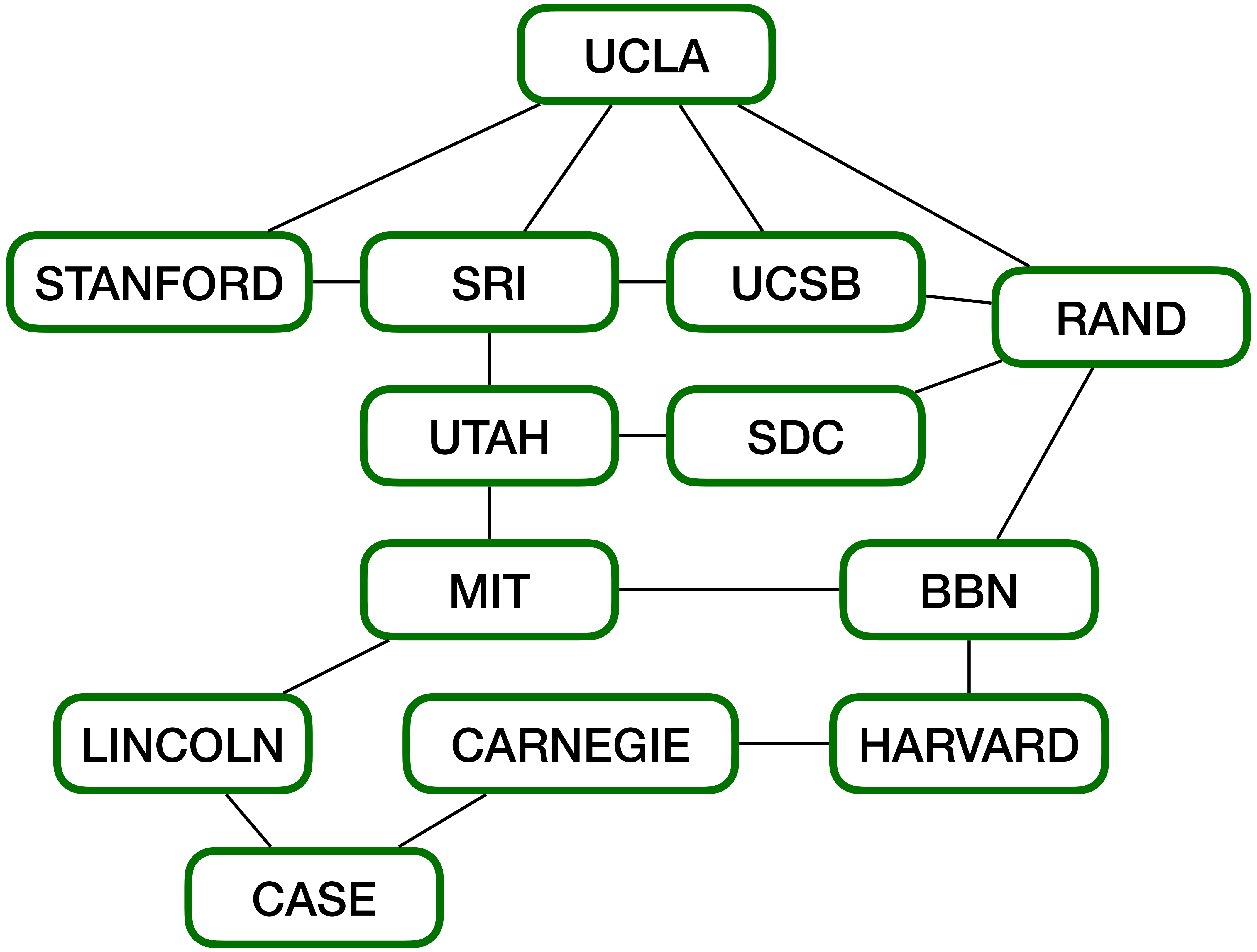
- We'll need a new way of representing this data structure and new algorithms to work with the data
- Store the nodes and edges

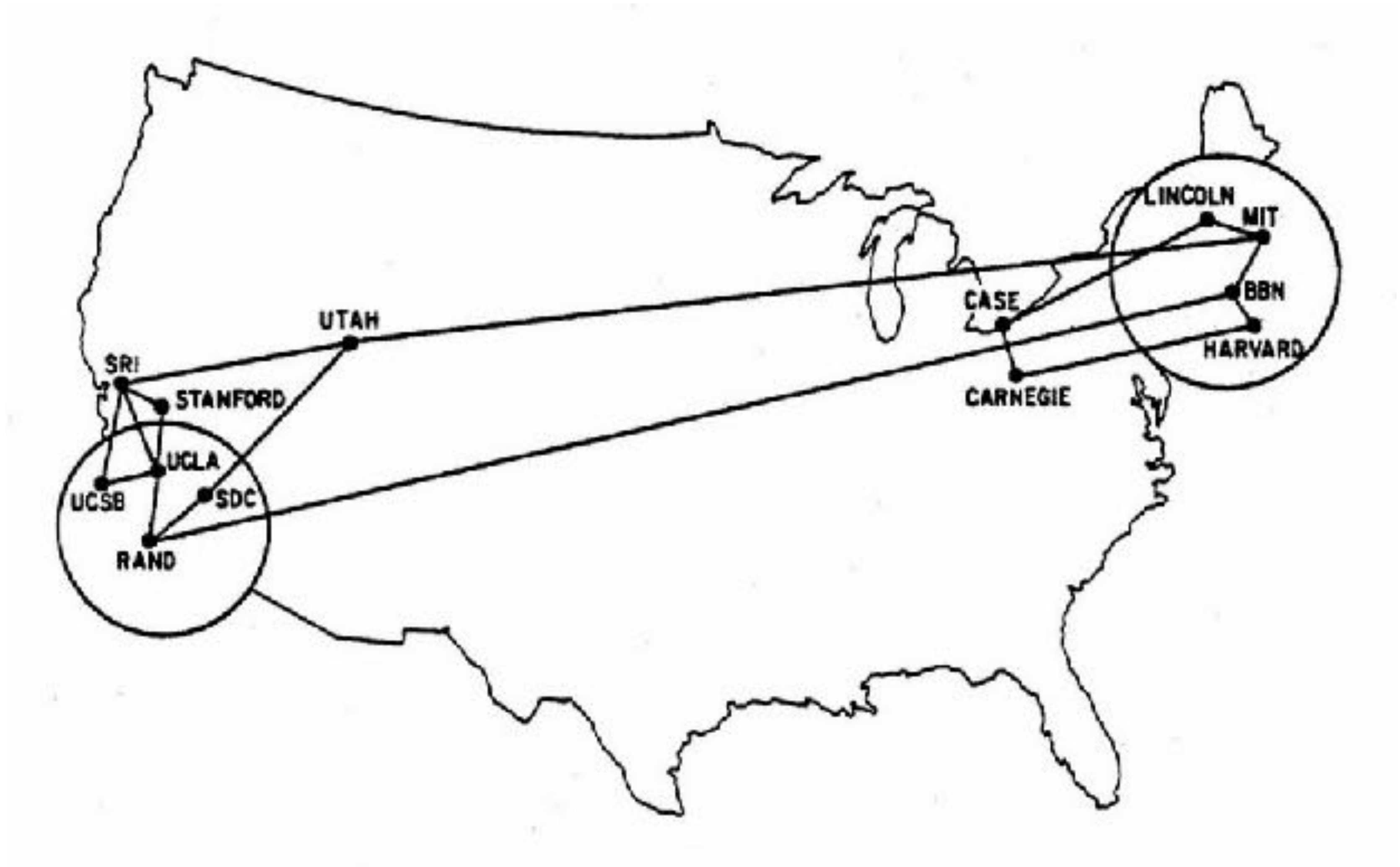




Graphs - Nodes and Edges

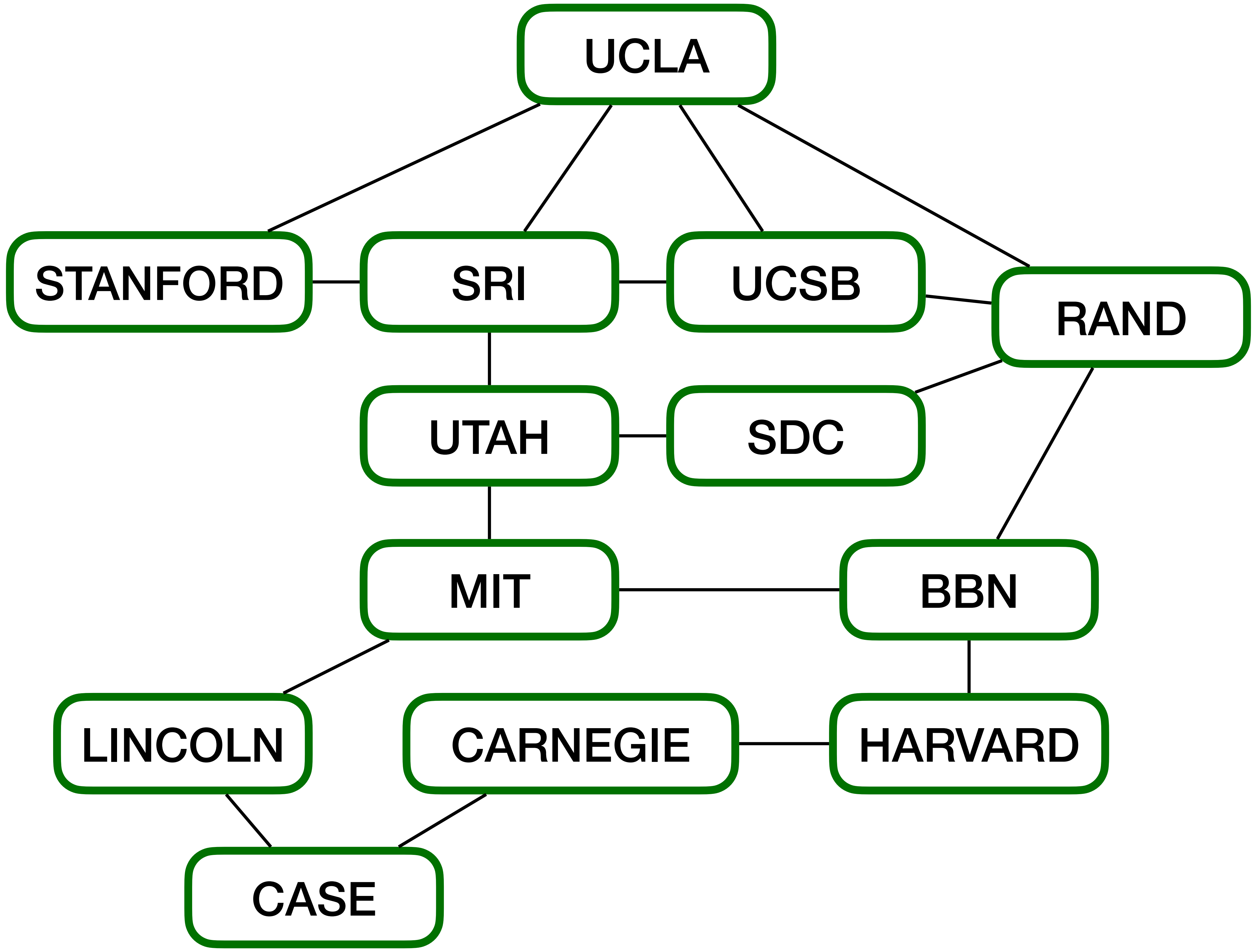
- Node: Each data element is stored in a node, similar to linked lists and trees
- Edge: A connection between two nodes





Graphs - Adjacency List

- A map of nodes to all nodes connected to it through an edge
- This is how we'll represent graphs



UCLA	STANFORD, SRI, UCSB, RAND
STANFORD	UCLA, SRI
SRI	UCLA, UCSB, UTAH, STANFORD
UCSB	UCLA, SRI, RAND
RAND	UCLA, UCSB, SDC, BBN
UTAH	SRI, SDC, MIT
SDC	UTAH, RAND
MIT	UTAH, BBN, LINCOLN
BBN	RAND, MIT, HAVARD
LINCOLN	MIT, CASE
CARNEGIE	CASE, HARVARD
HARVARD	BBN, CARNEGIE
CASE	LINCOLN, CARNEGIE

Graphs - Adjacency List

- When creating a graph, we'll assign each node a unique ID as an int
  - Allows nodes with identical values, but different IDs
  - Allows a generic Graph using Int to store any type parameter

```scala
class Graph[A] {

  var nodes: Map[Int, A] = Map()
  var adjacencyList: Map[Int, List[Int]] = Map()

  def addNode(index: Int, a: A): Unit = {
    nodes += index -> a
    adjacencyList += index -> List()
  }

  def addEdge(index1: Int, index2: Int): Unit = {
    adjacencyList += index1 -> (index2 :: adjacencyList(index1))
    adjacencyList += index2 -> (index1 :: adjacencyList(index2))
  }

}
```

UCLA	STANFORD, SRI, UCSB, RAND
STANFORD	UCLA, SRI
SRI	UCLA, UCSB, UTAH, STANFORD
UCSB	UCLA, SRI, RAND
RAND	UCLA, UCSB, SDC, BBN
UTAH	SRI, SDC, MIT
SDC	UTAH, RAND
MIT	UTAH, BBN, LINCOLN
BBN	RAND, MIT, HAVARD
LINCOLN	MIT, CASE
CARNEGIE	CASE, HARVARD
HARVARD	BBN, CARNEGIE
CASE	LINCOLN, CARNEGIE

Graphs - Adjacency List

- IDs for each node are arbitrary as long as they are unique
- Methods will work with IDs
- Values are only accessed when needed

```scala
class Graph[A] {

  var nodes: Map[Int, A] = Map()
  var adjacencyList: Map[Int, List[Int]] = Map()

  def addNode(index: Int, a: A): Unit = {
    nodes += index -> a
    adjacencyList += index -> List()
  }

  def addEdge(index1: Int, index2: Int): Unit = {
    adjacencyList += index1 -> (index2 :: adjacencyList(index1))
    adjacencyList += index2 -> (index1 :: adjacencyList(index2))
  }

}
```

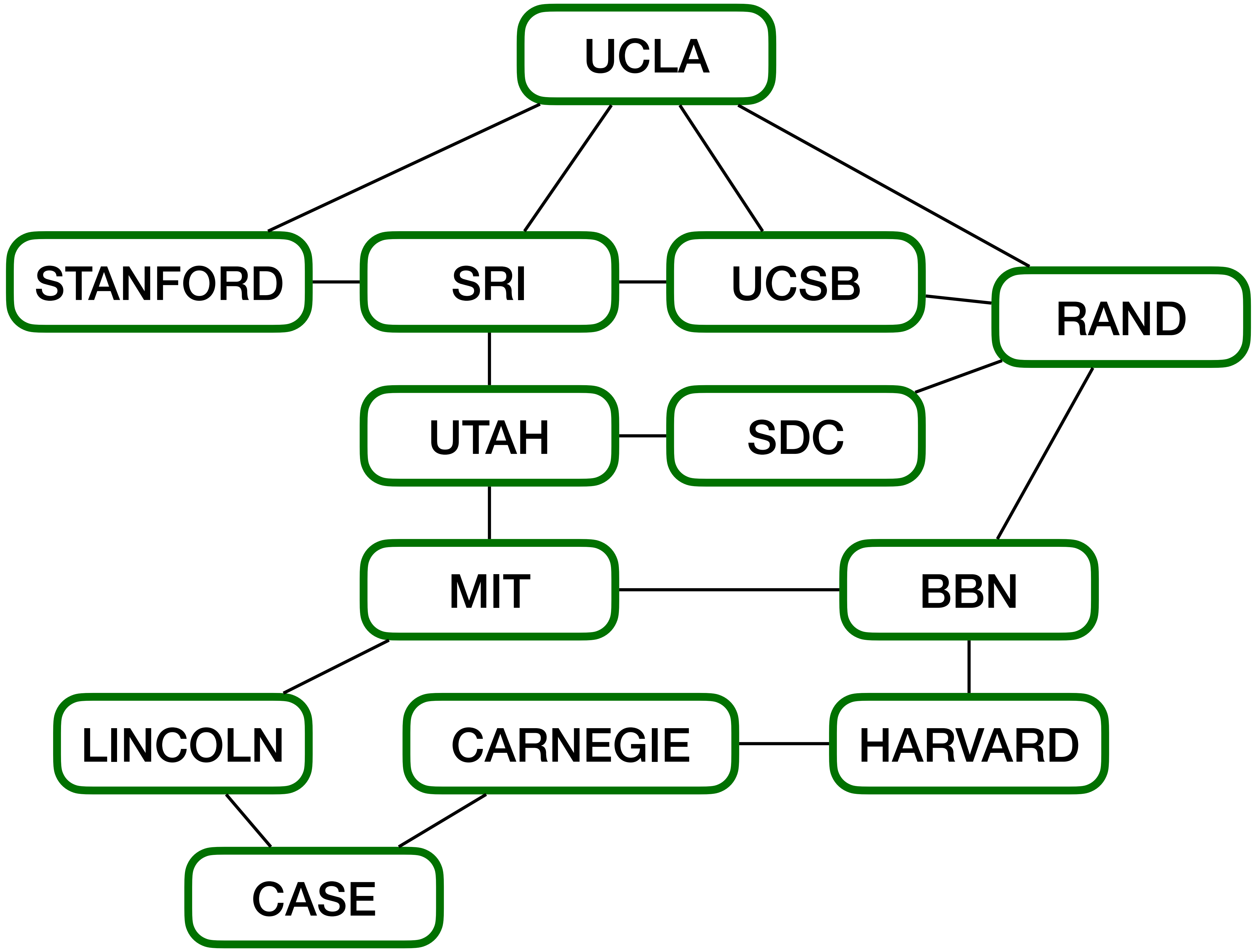
```scala
object GraphExample {

  def main(args: Array[String]): Unit = {
    val graph: Graph[String] = new Graph()
    graph.addNode(0, "UCLA")
    graph.addNode(1, "STANFORD")
    graph.addNode(2, "SRI")
    graph.addNode(3, "UCSB")
    graph.addNode(4, "RAND")
    graph.addNode(5, "UTAH")
    graph.addNode(6, "SDC")
    graph.addNode(7, "MIT")
    graph.addNode(8, "BBN")
    graph.addNode(9, "LINCOLN")
    graph.addNode(10, "CARNEGIE")
    graph.addNode(11, "HARVARD")
    graph.addNode(12, "CASE")

    graph.addEdge(0,1)
    graph.addEdge(0,2)
    graph.addEdge(0,3)
    graph.addEdge(0,4)
    graph.addEdge(1,2)
    graph.addEdge(2,3)
    graph.addEdge(3,4)
    graph.addEdge(2,5)
    graph.addEdge(4,6)
    graph.addEdge(5,6)
    graph.addEdge(5,7)
    graph.addEdge(4,8)
    graph.addEdge(7,8)
    graph.addEdge(7,9)
    graph.addEdge(9,12)
    graph.addEdge(12,10)
    graph.addEdge(10,11)
    graph.addEdge(11,8)
  }
}
```

Paths

- A path is a sequence of nodes where each pair of adjacent nodes are connected by an edge
- ["UCLA", "SRI", "UTAH", "MIT", "BBN", "RAND"] is a path in this graph
- ["SRI", "UTAH", "BBN"] is not a path since UTAH and BBN are not connected by an edge

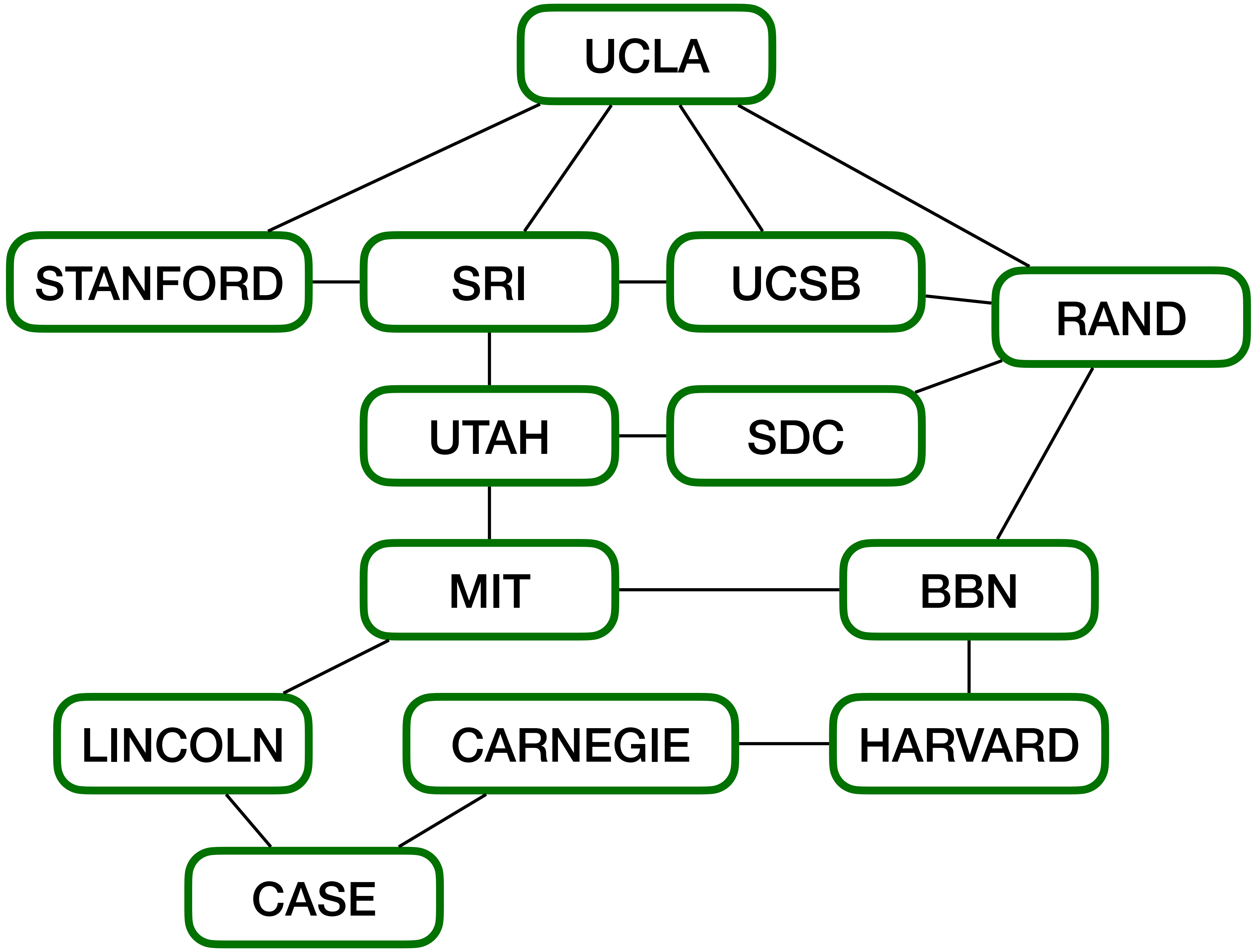


UCLA	STANFORD, SRI, UCSB, RAND
STANFORD	UCLA, SRI
SRI	UCLA, UCSB, UTAH, STANFORD
UCSB	UCLA, SRI, RAND
RAND	UCLA, UCSB, SDC, BBN
UTAH	SRI, SDC, MIT
SDC	UTAH, RAND
MIT	UTAH, BBN, LINCOLN
BBN	RAND, MIT, HAVARD
LINCOLN	MIT, CASE
CARNEGIE	CASE, HARVARD
HARVARD	BBN, CARNEGIE
CASE	LINCOLN, CARNEGIE

Breadth-First Search (BFS)

Connected Component

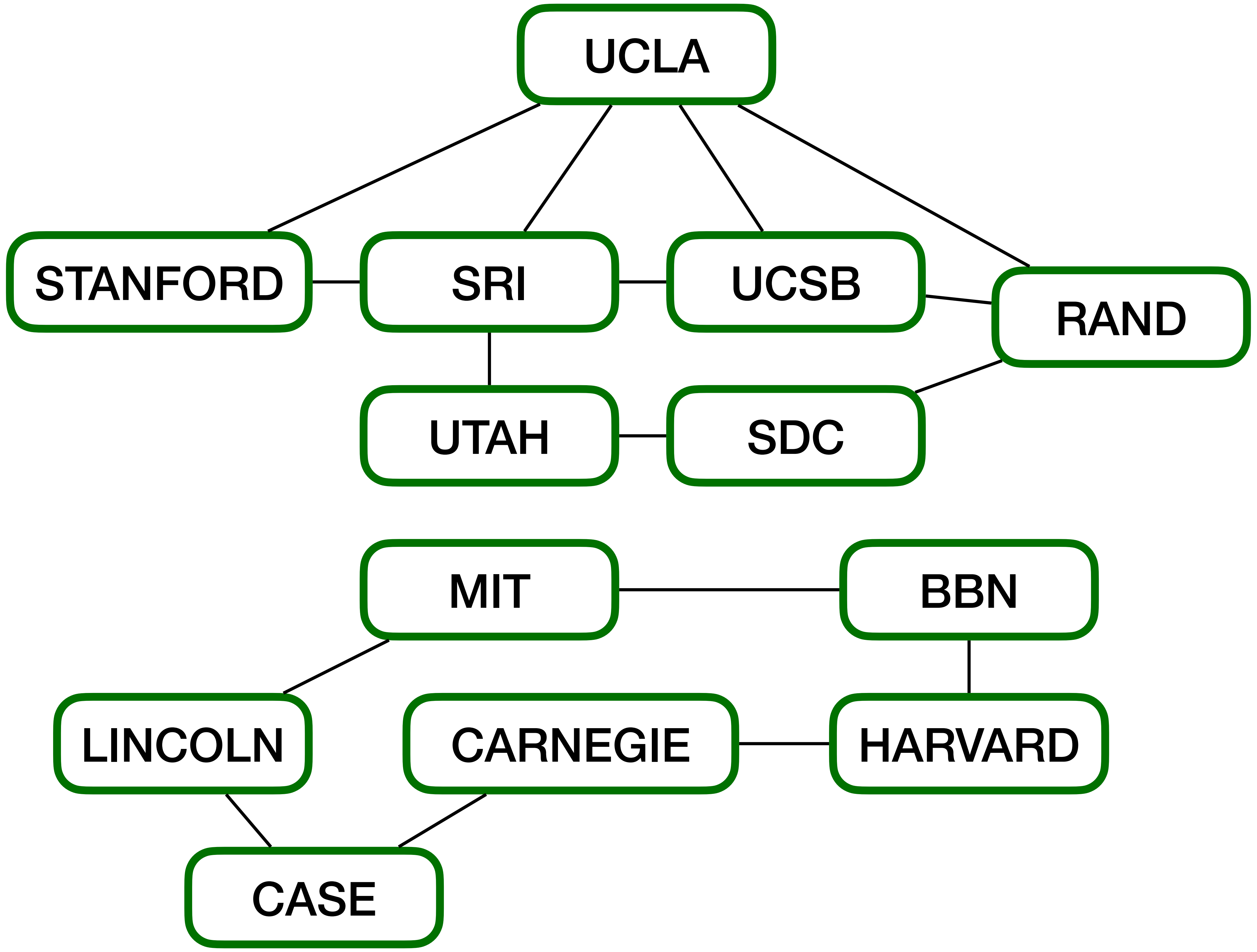
- This graph is connected
  - There exists a path between any 2 nodes in the graph



UCLA	STANFORD, SRI, UCSB, RAND
STANFORD	UCLA, SRI
SRI	UCLA, UCSB, UTAH, STANFORD
UCSB	UCLA, SRI, RAND
RAND	UCLA, UCSB, SDC, BBN
UTAH	SRI, SDC, MIT
SDC	UTAH, RAND
MIT	UTAH, BBN, LINCOLN
BBN	RAND, MIT, HAVARD
LINCOLN	MIT, CASE
CARNEGIE	CASE, HARVARD
HARVARD	BBN, CARNEGIE
CASE	LINCOLN, CARNEGIE

Connected Component

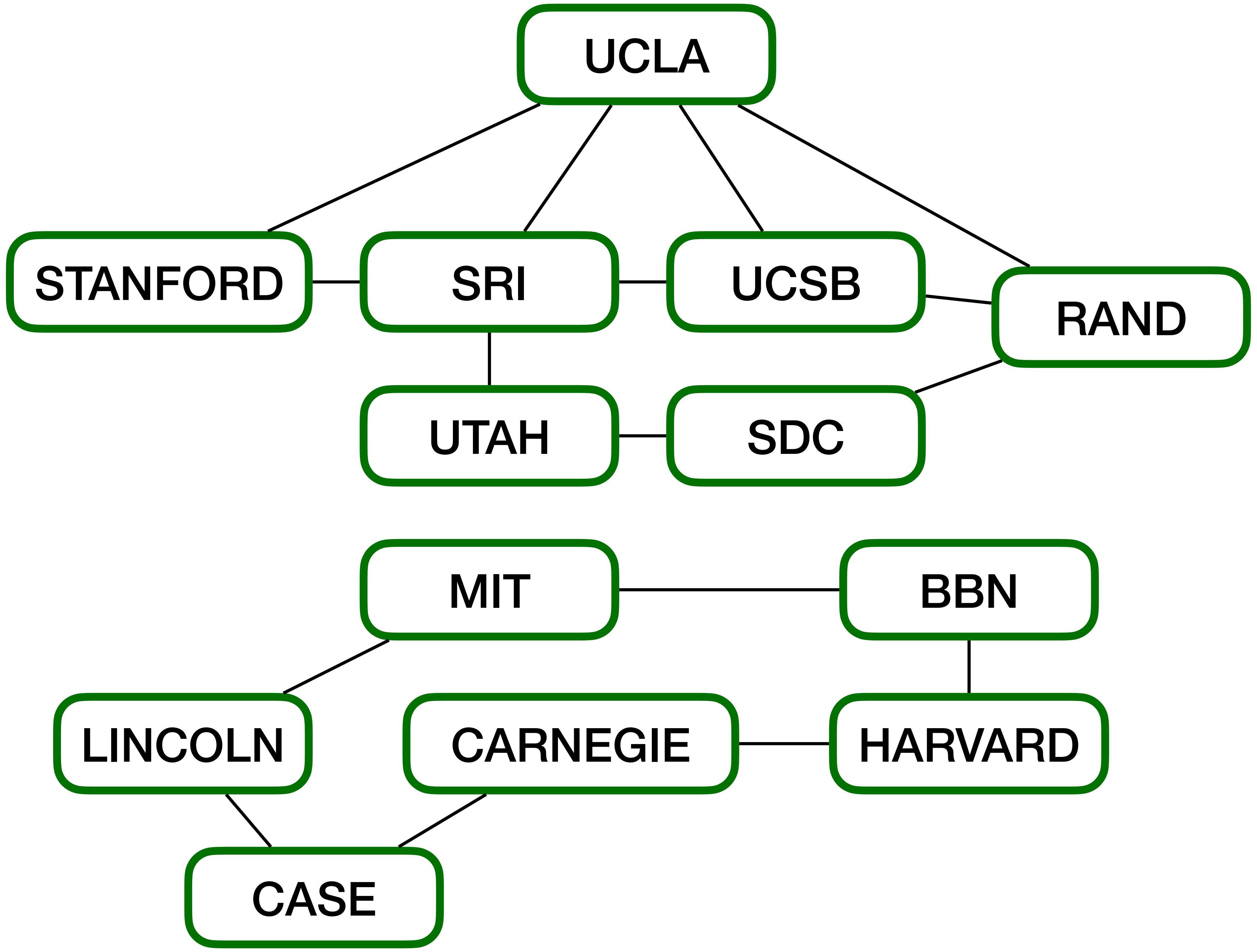
- What if a few connections are broken?
  - How can we tell if two nodes are connected?

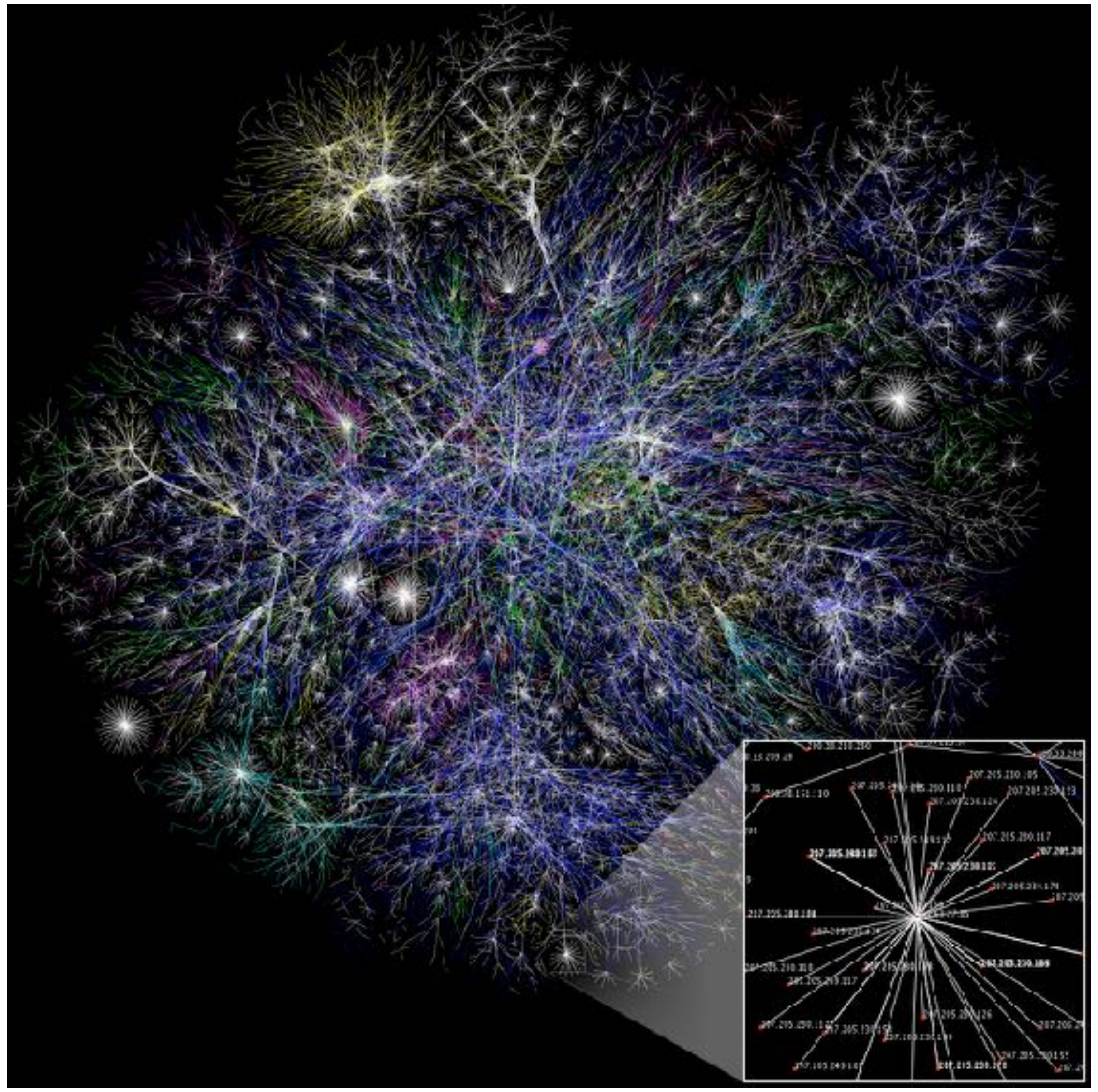


UCLA	STANFORD, SRI, UCSB, RAND
STANFORD	UCLA, SRI
SRI	UCLA, UCSB, UTAH, STANFORD
UCSB	UCLA, SRI, RAND
RAND	UCLA, UCSB, SDC
UTAH	SRI, SDC
SDC	UTAH, RAND
MIT	BBN, LINCOLN
BBN	MIT, HAVARD
LINCOLN	MIT, CASE
CARNEGIE	CASE, HARVARD
HARVARD	BBN, CARNEGIE
CASE	LINCOLN, CARNEGIE

Connected Component

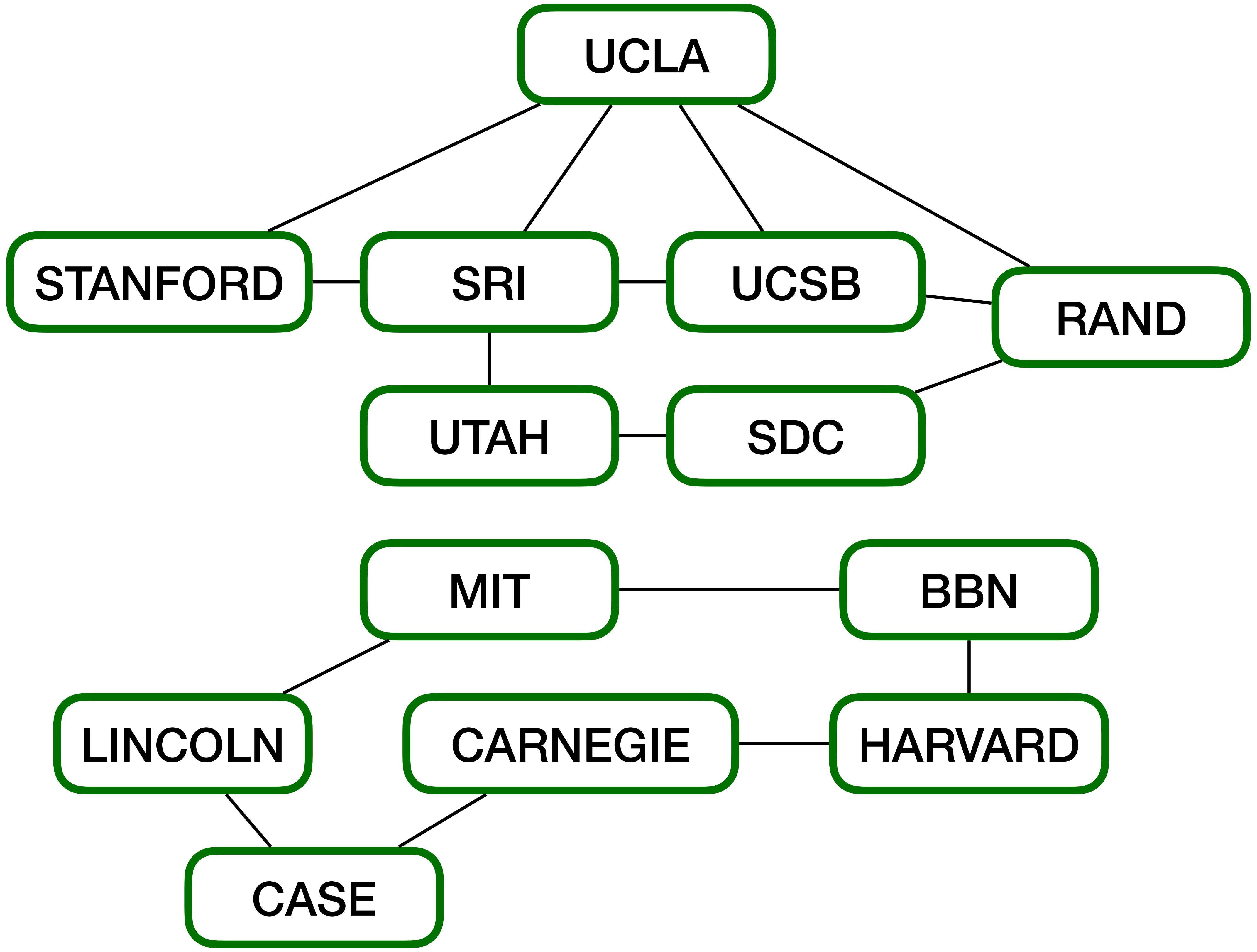
- We could verify manually for this graph
- But the Internet has gotten a *little* bigger over time
- Need to code an algorithm to solve this for us





BFS

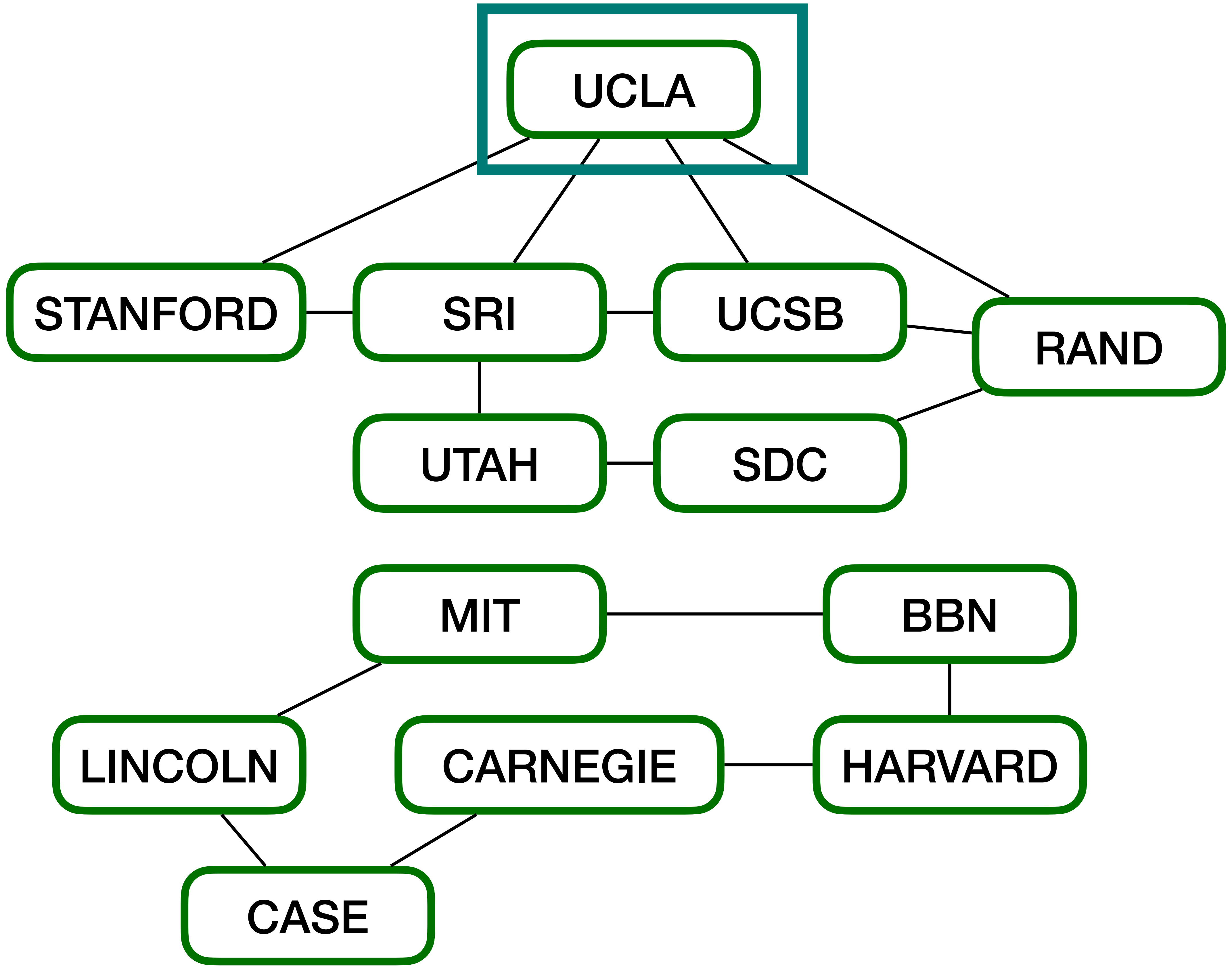
- The Algorithm: Breadth-First Search (BFS)
  - Choose a starting node
  - Continuously explore connected nodes



UCLA	STANFORD, SRI, UCSB, RAND
STANFORD	UCLA, SRI
SRI	UCLA, UCSB, UTAH, STANFORD
UCSB	UCLA, SRI, RAND
RAND	UCLA, UCSB, SDC
UTAH	SRI, SDC
SDC	UTAH, RAND
MIT	BBN, LINCOLN
BBN	MIT, HAVARD
LINCOLN	MIT, CASE
CARNEGIE	CASE, HARVARD
HARVARD	BBN, CARNEGIE
CASE	LINCOLN, CARNEGIE

BFS

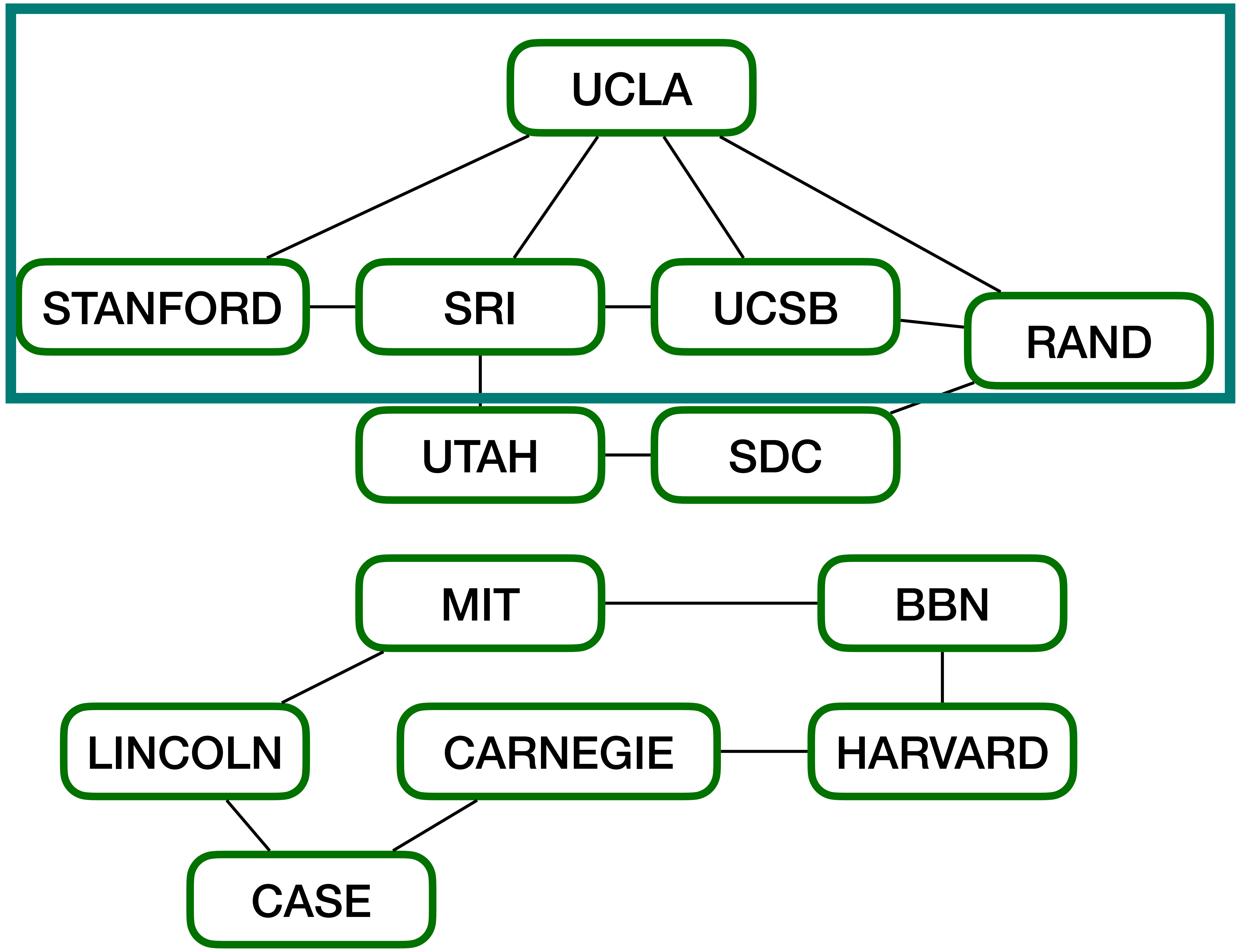
- Choose a starting node



UCLA	STANFORD, SRI, UCSB, RAND
STANFORD	UCLA, SRI
SRI	UCLA, UCSB, UTAH, STANFORD
UCSB	UCLA, SRI, RAND
RAND	UCLA, UCSB, SDC
UTAH	SRI, SDC
SDC	UTAH, RAND
MIT	BBN, LINCOLN
BBN	MIT, HAVARD
LINCOLN	MIT, CASE
CARNEGIE	CASE, HARVARD
HARVARD	BBN, CARNEGIE
CASE	LINCOLN, CARNEGIE

BFS

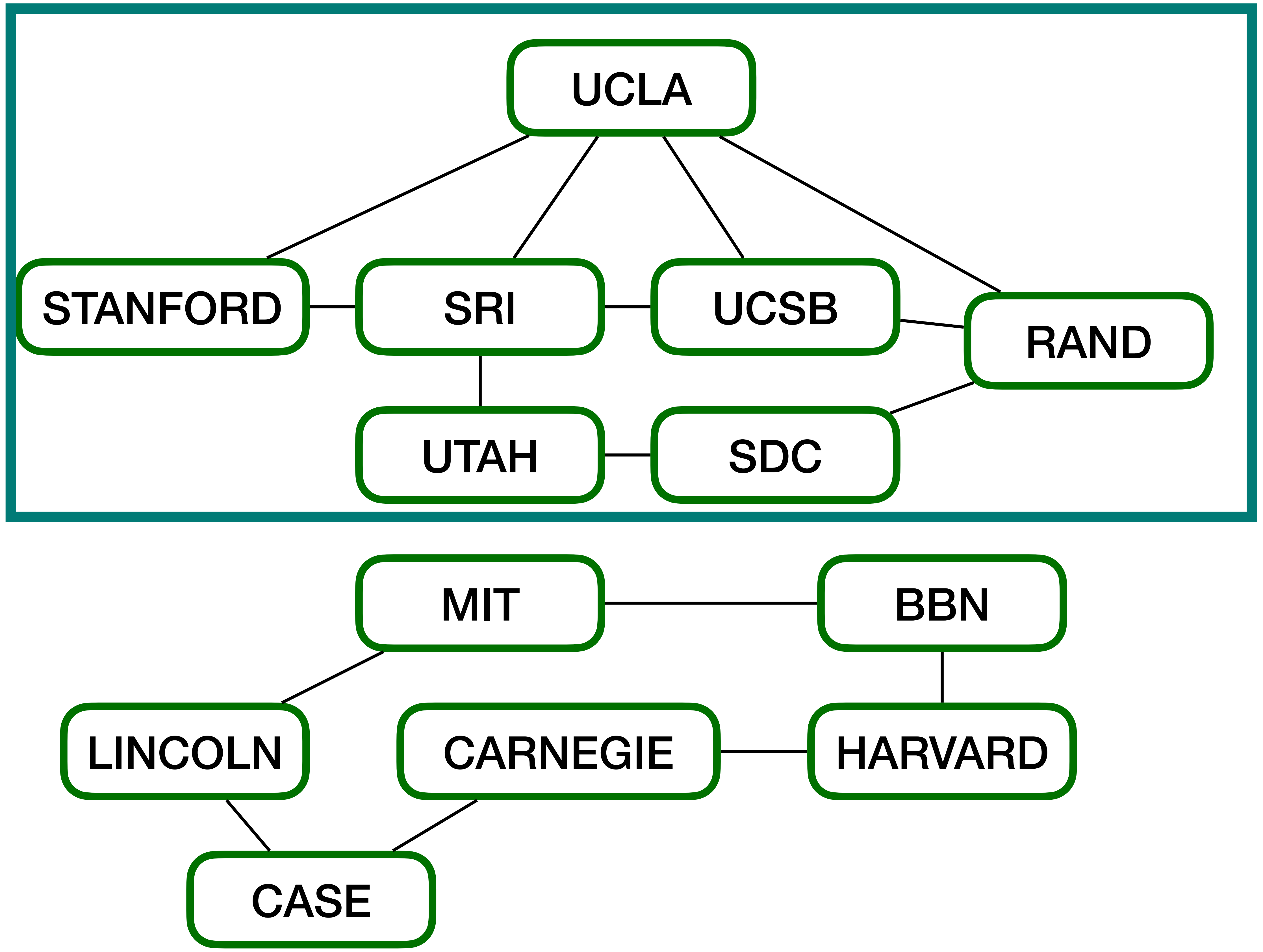
- Explore all nodes connected to the striating node



UCLA	STANFORD, SRI, UCSB, RAND
STANFORD	UCLA, SRI
SRI	UCLA, UCSB, UTAH, STANFORD
UCSB	UCLA, SRI, RAND
RAND	UCLA, UCSB, SDC
UTAH	SRI, SDC
SDC	UTAH, RAND
MIT	BBN, LINCOLN
BBN	MIT, HAVARD
LINCOLN	MIT, CASE
CARNEGIE	CASE, HARVARD
HARVARD	BBN, CARNEGIE
CASE	LINCOLN, CARNEGIE

BFS

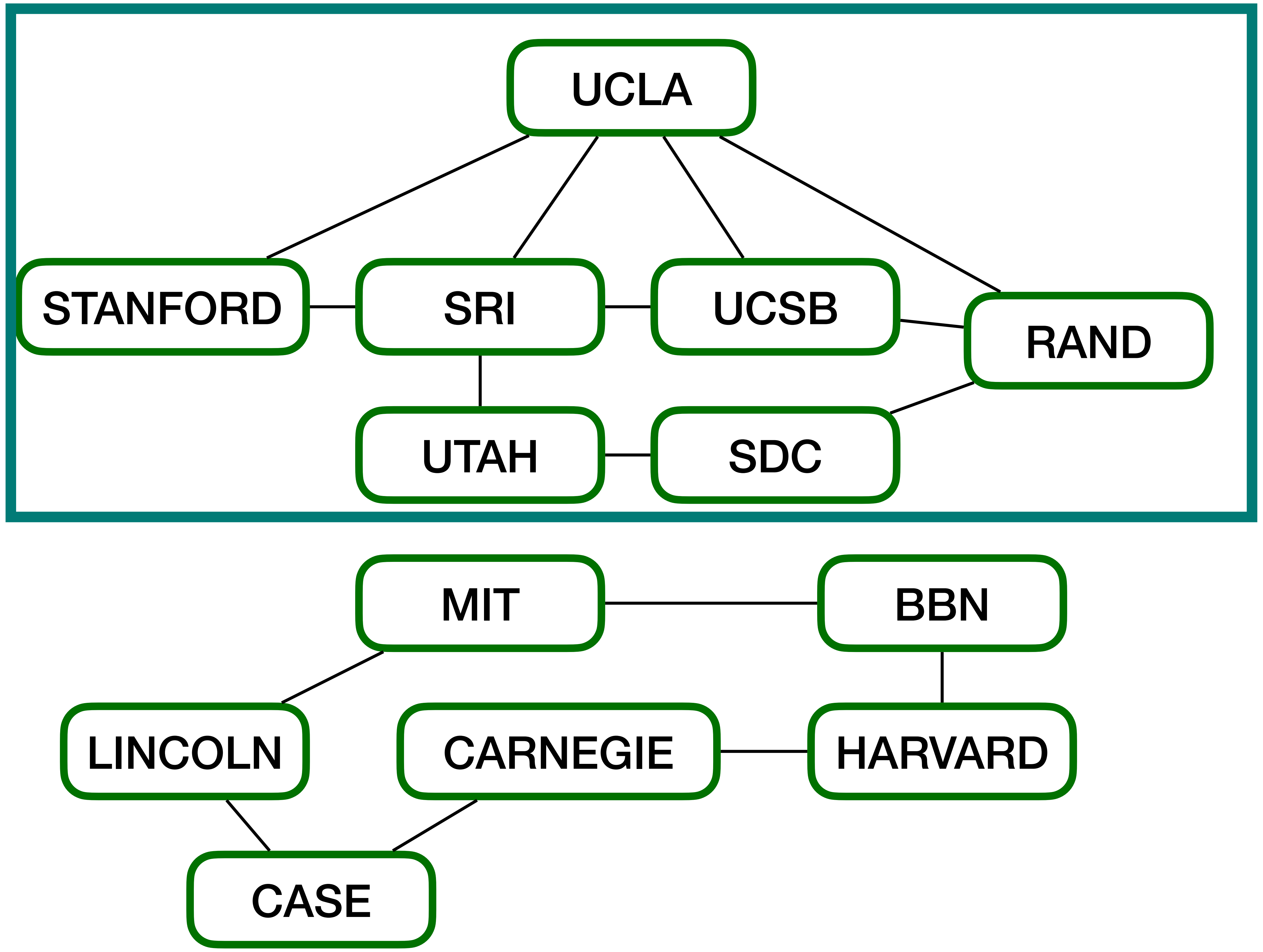
- Repeatedly explore nodes that were visited in the last round



UCLA	STANFORD, SRI, UCSB, RAND
STANFORD	UCLA, SRI
SRI	UCLA, UCSB, UTAH, STANFORD
UCSB	UCLA, SRI, RAND
RAND	UCLA, UCSB, SDC
UTAH	SRI, SDC
SDC	UTAH, RAND
MIT	BBN, LINCOLN
BBN	MIT, HAVARD
LINCOLN	MIT, CASE
CARNEGIE	CASE, HARVARD
HARVARD	BBN, CARNEGIE
CASE	LINCOLN, CARNEGIE

BFS

- Repeat until no new nodes are added
- Never visit a node twice



UCLA	STANFORD, SRI, UCSB, RAND
STANFORD	UCLA, SRI
SRI	UCLA, UCSB, UTAH, STANFORD
UCSB	UCLA, SRI, RAND
RAND	UCLA, UCSB, SDC
UTAH	SRI, SDC
SDC	UTAH, RAND
MIT	BBN, LINCOLN
BBN	MIT, HAVARD
LINCOLN	MIT, CASE
CARNEGIE	CASE, HARVARD
HARVARD	BBN, CARNEGIE
CASE	LINCOLN, CARNEGIE

BFS

- Use a queue to track the order of nodes to visit
- Start with starting node in the queue
- When visiting a node, add all unexplored neighbors to the queue
- Visit neighbors of the node at the front of the queue until the queue is empty

UCLA	STANFORD, SRI, UCSB, RAND
STANFORD	UCLA, SRI
SRI	UCLA, UCSB, UTAH, STANFORD
UCSB	UCLA, SRI, RAND
RAND	UCLA, UCSB, SDC
UTAH	SRI, SDC
SDC	UTAH, RAND
MIT	BBN, LINCOLN
BBN	MIT, HAVARD
LINCOLN	MIT, CASE
CARNEGIE	CASE, HARVARD
HARVARD	BBN, CARNEGIE
CASE	LINCOLN, CARNEGIE

BFS

- More BFS details to come

UCLA	STANFORD, SRI, UCSB, RAND
STANFORD	UCLA, SRI
SRI	UCLA, UCSB, UTAH, STANFORD
UCSB	UCLA, SRI, RAND
RAND	UCLA, UCSB, SDC
UTAH	SRI, SDC
SDC	UTAH, RAND
MIT	BBN, LINCOLN
BBN	MIT, HAVARD
LINCOLN	MIT, CASE
CARNEGIE	CASE, HARVARD
HARVARD	BBN, CARNEGIE
CASE	LINCOLN, CARNEGIE

Connectivity

- If you start at nodeA and explore nodeB during the algorithm
    - nodeA and nodeB are connected