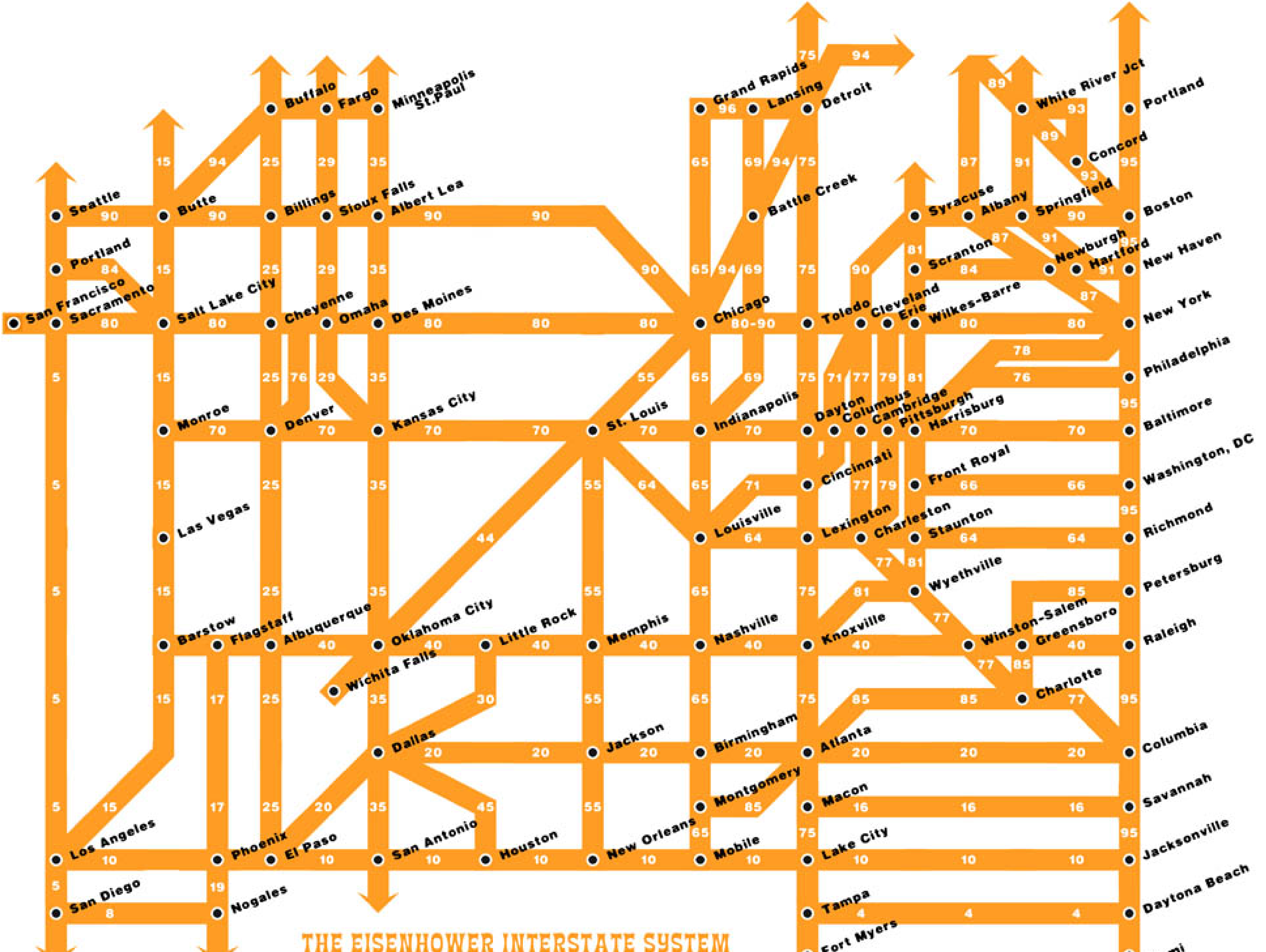
Graph Theory

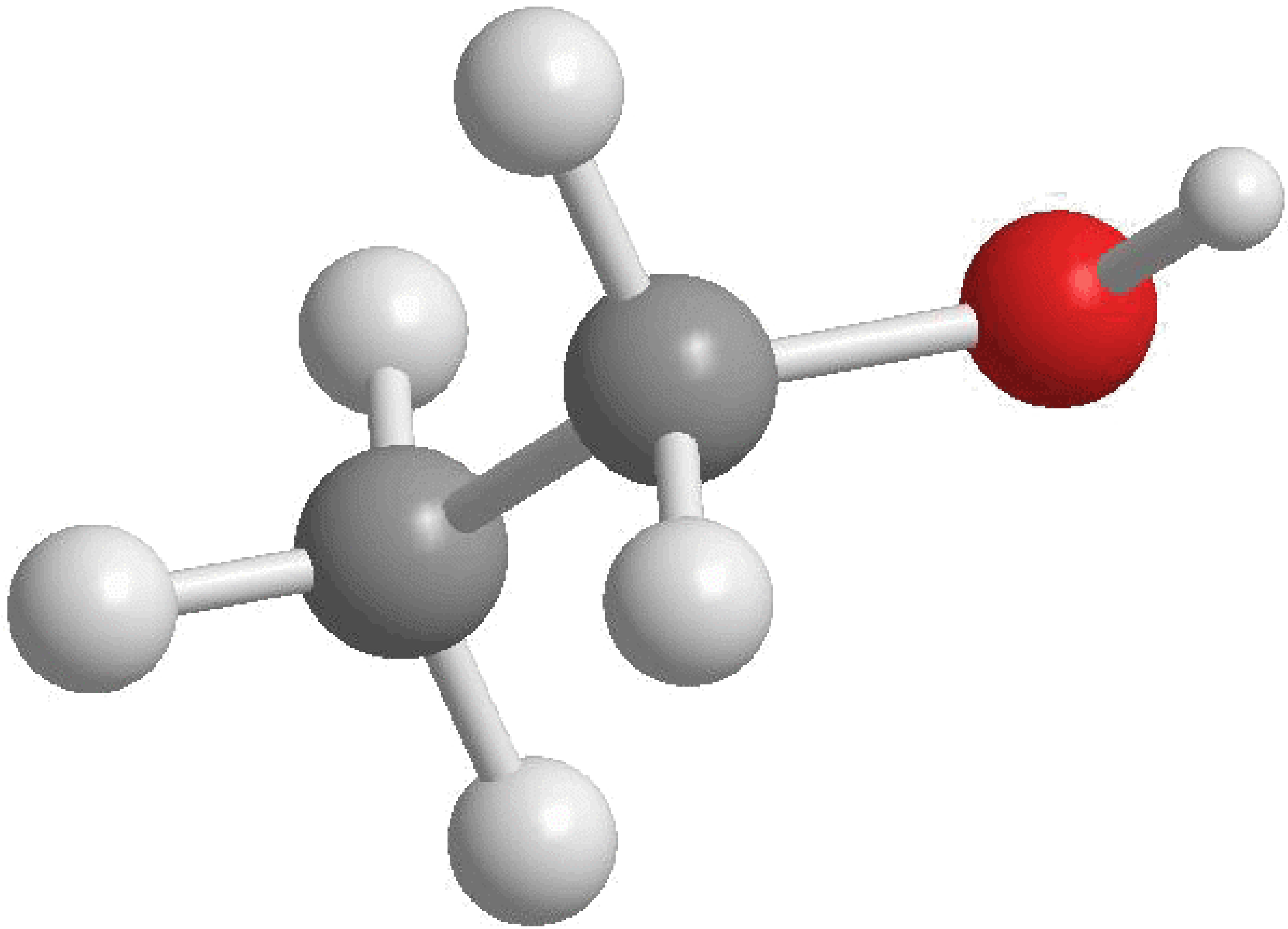


THE EISENHOWER INTERSTATE SYSTEM

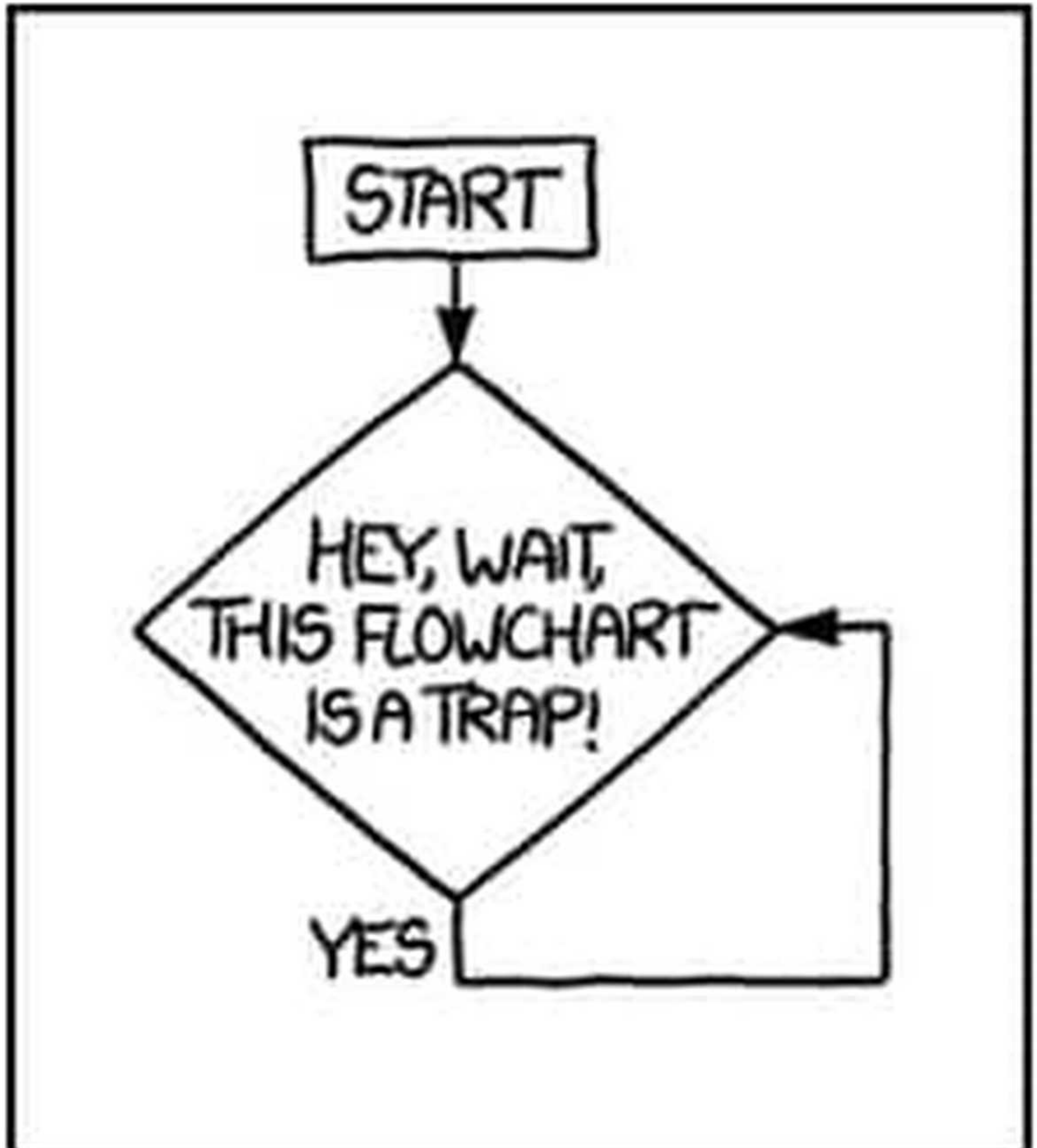
(simplified)

CHRIS YATES 2007

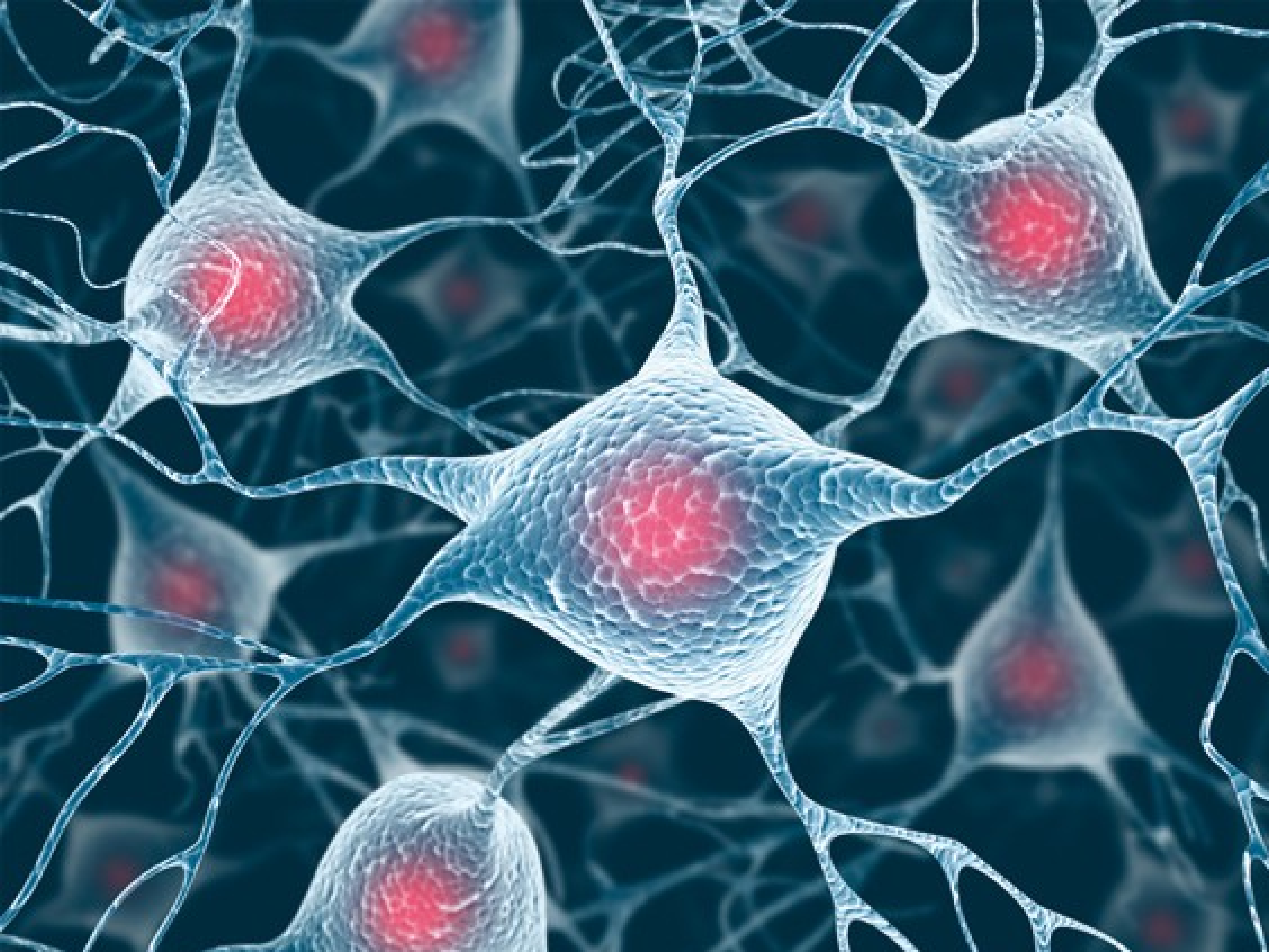
Chemical Bonds



http://4.bp.blogspot.com/-xCtBJ8lKHqA/Tjm0BONWBRI/AAAAAAAAAK4/-mHrbAUOHHg/s1600/



https://xkcd.com/1195/

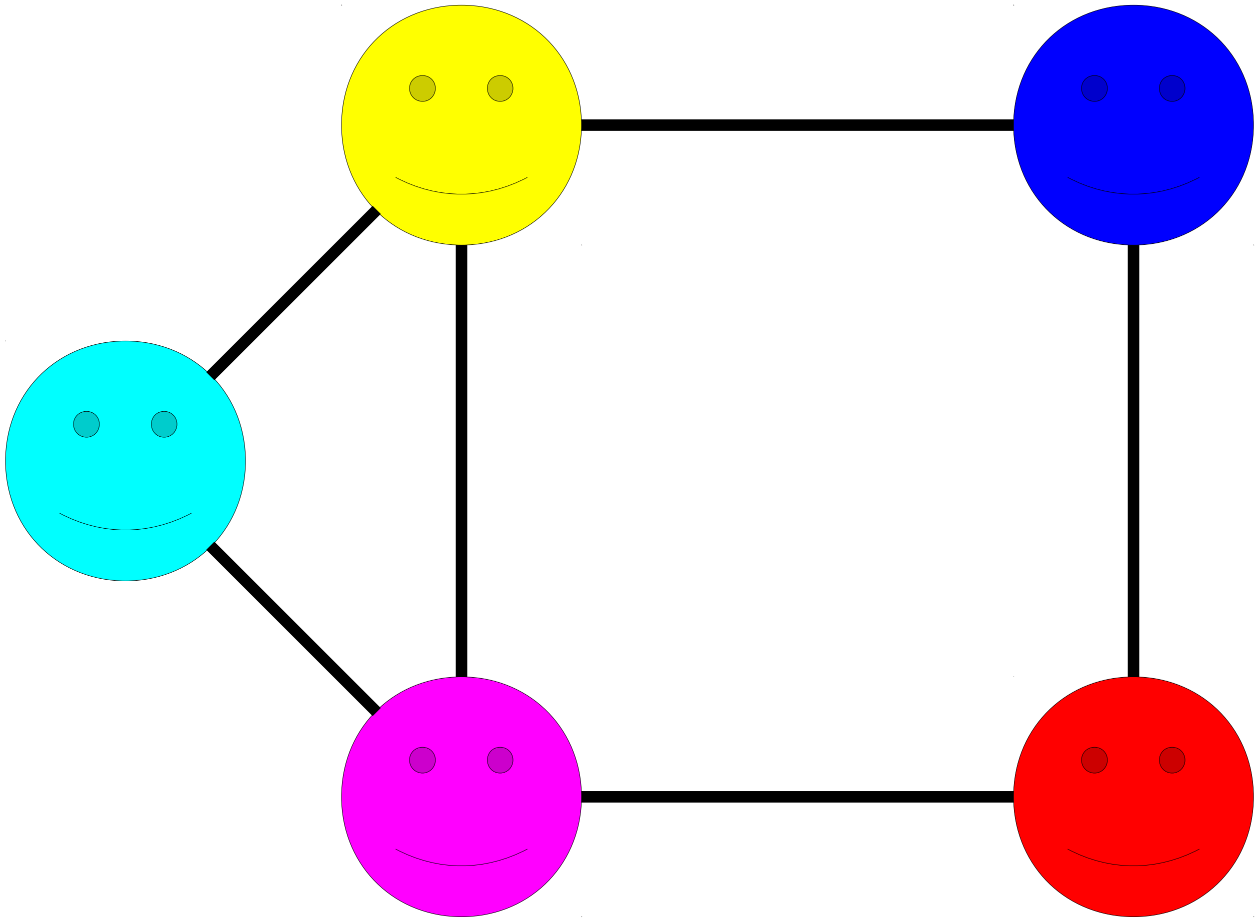


facebook®

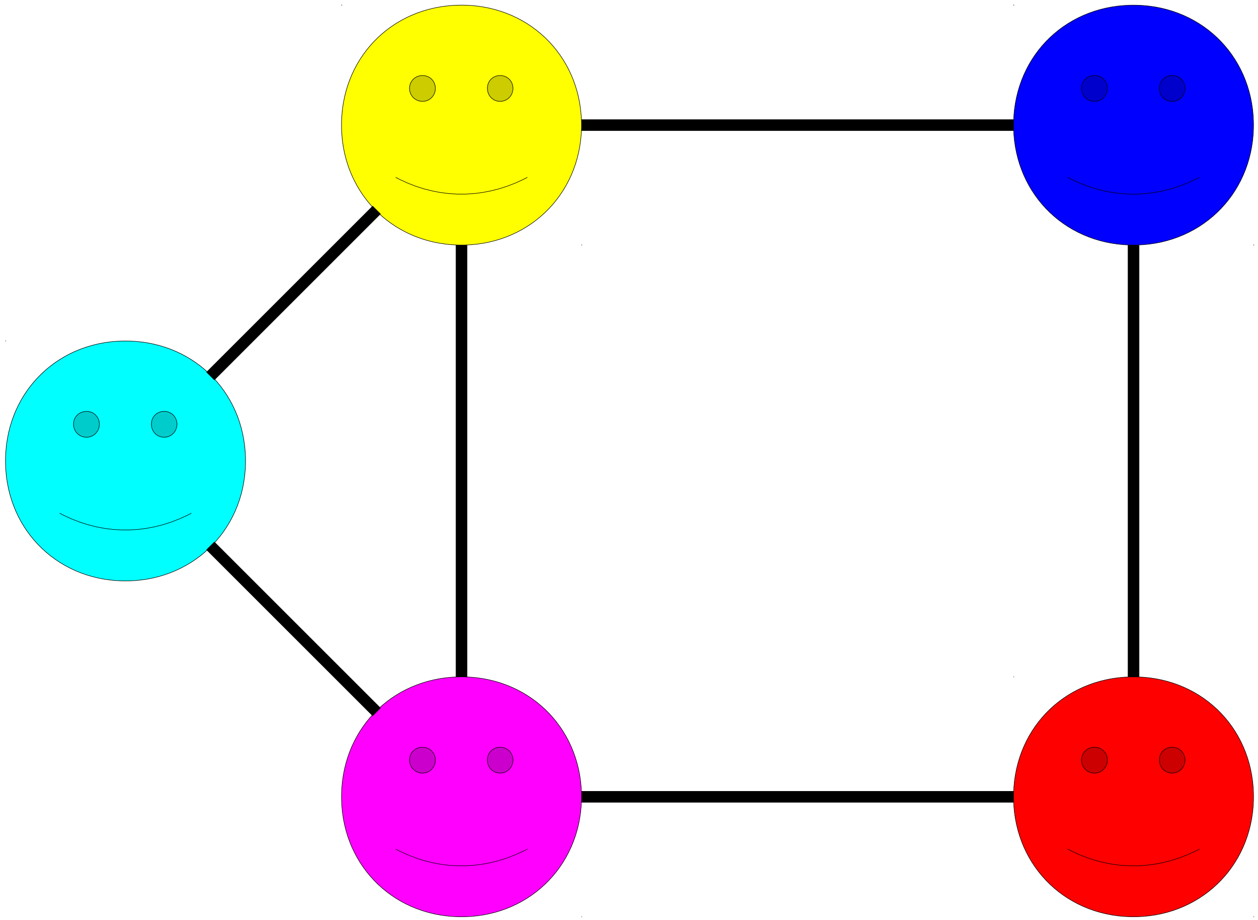
What's in Common

- Each of these structures consists of
  - a collection of objects and
  - links between those objects.
- ***Goal:*** find a general framework for describing these objects and their properties.

A ***graph*** is a mathematical structure for representing relationships.

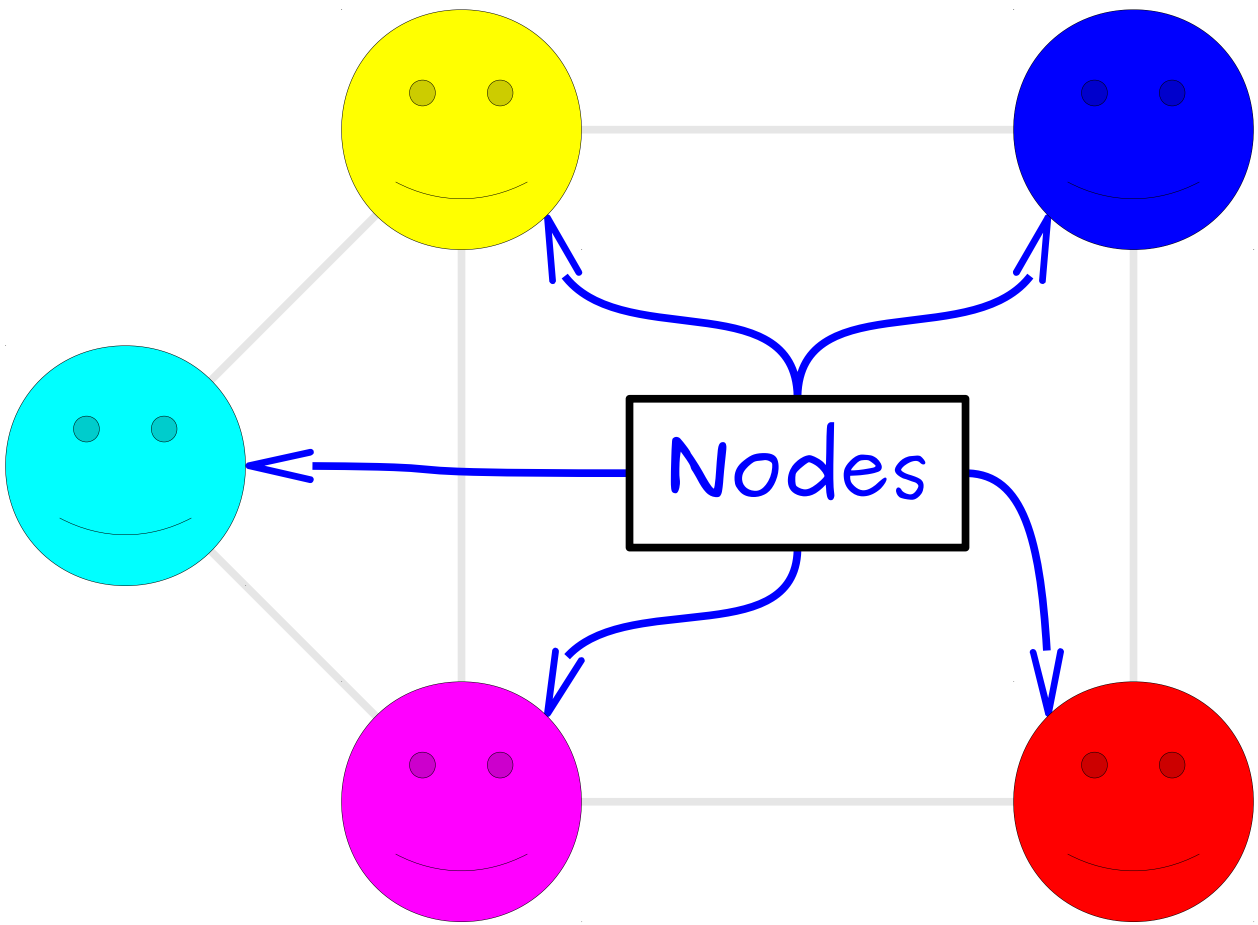


A ***graph*** is a mathematical structure for representing relationships.



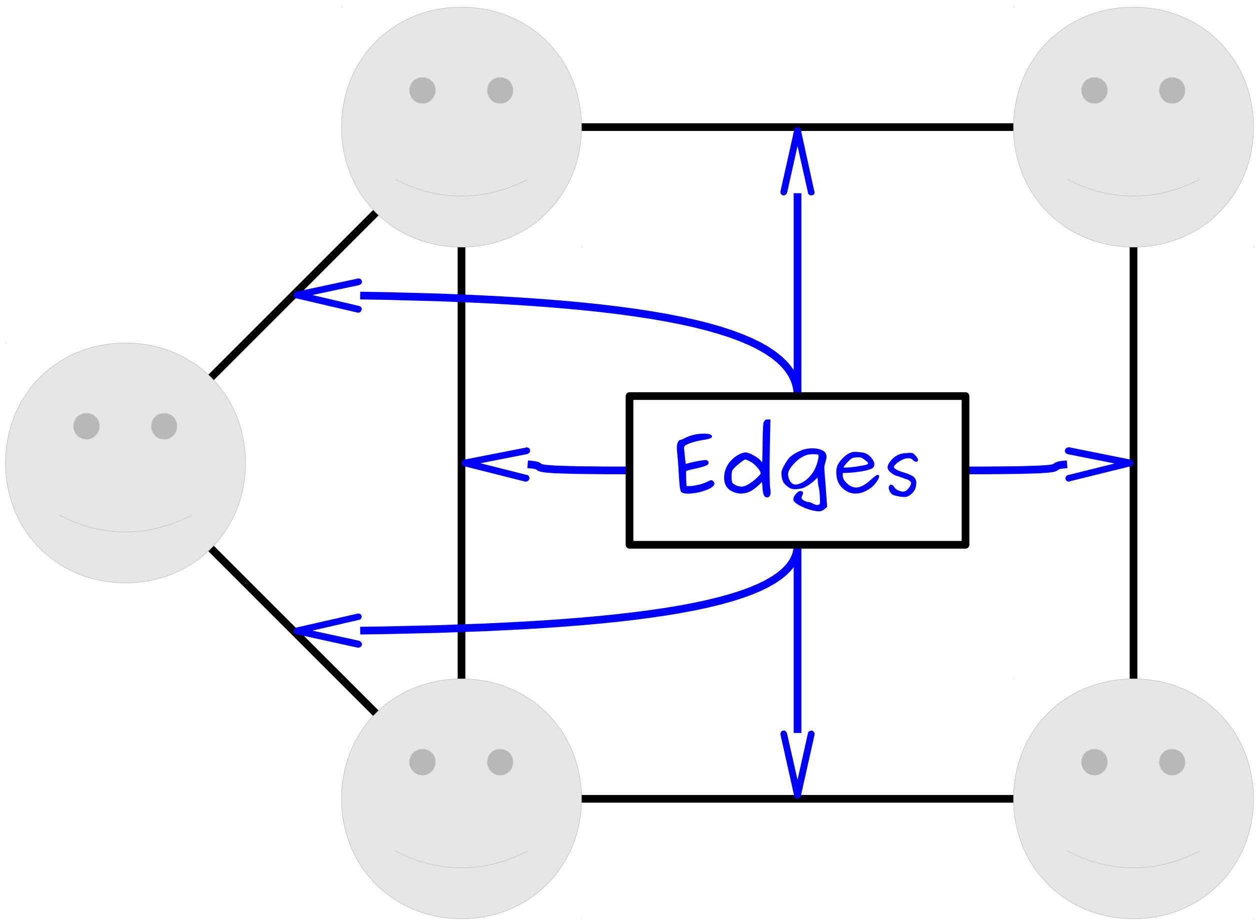
A graph consists of a set of ***nodes*** (or ***vertices***) connected by ***edges*** (or ***arcs***)

A ***graph*** is a mathematical structure for representing relationships.



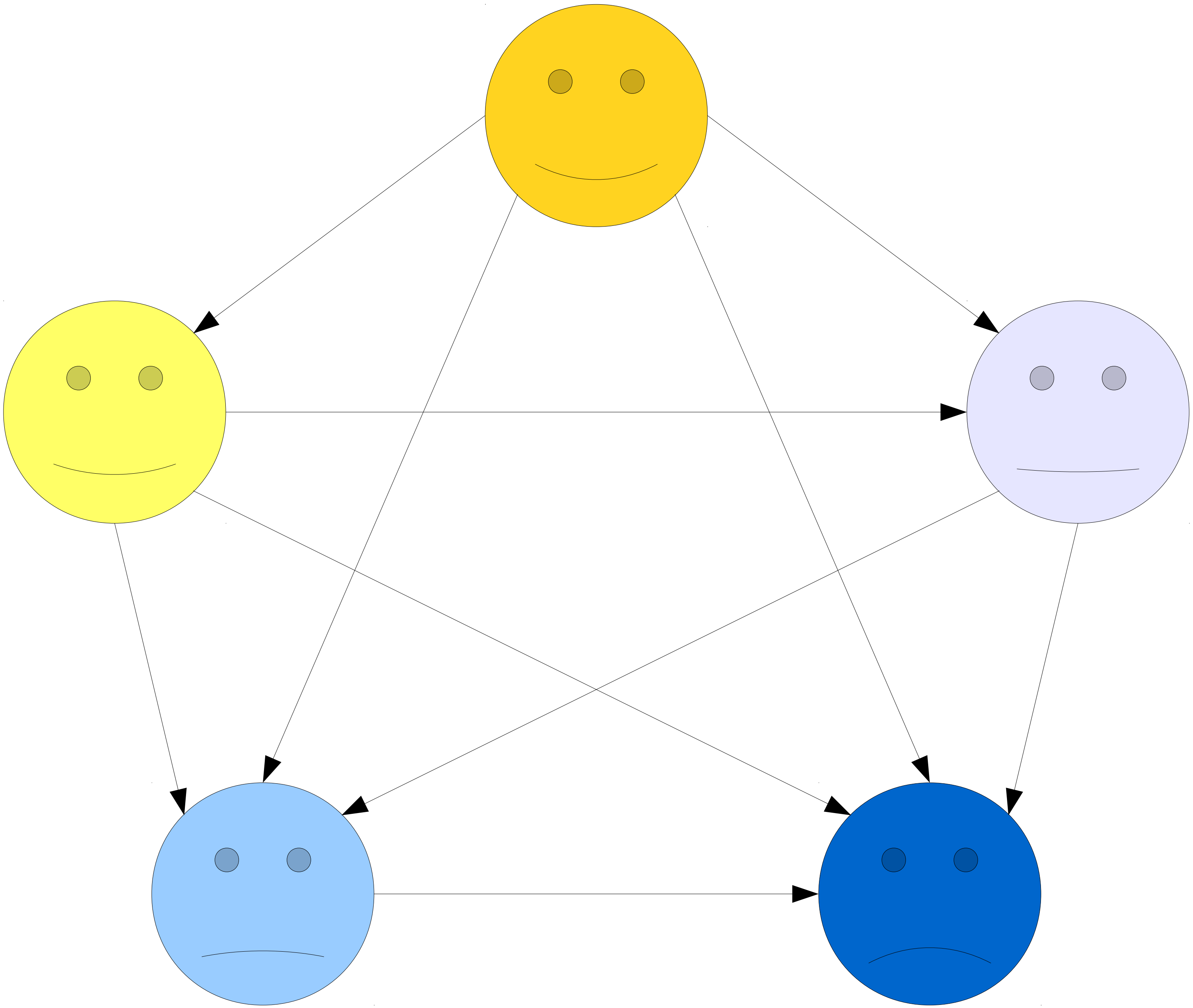
A graph consists of a set of ***nodes*** (or ***vertices***) connected by ***edges*** (or ***arcs***)

A ***graph*** is a mathematical structure for representing relationships.

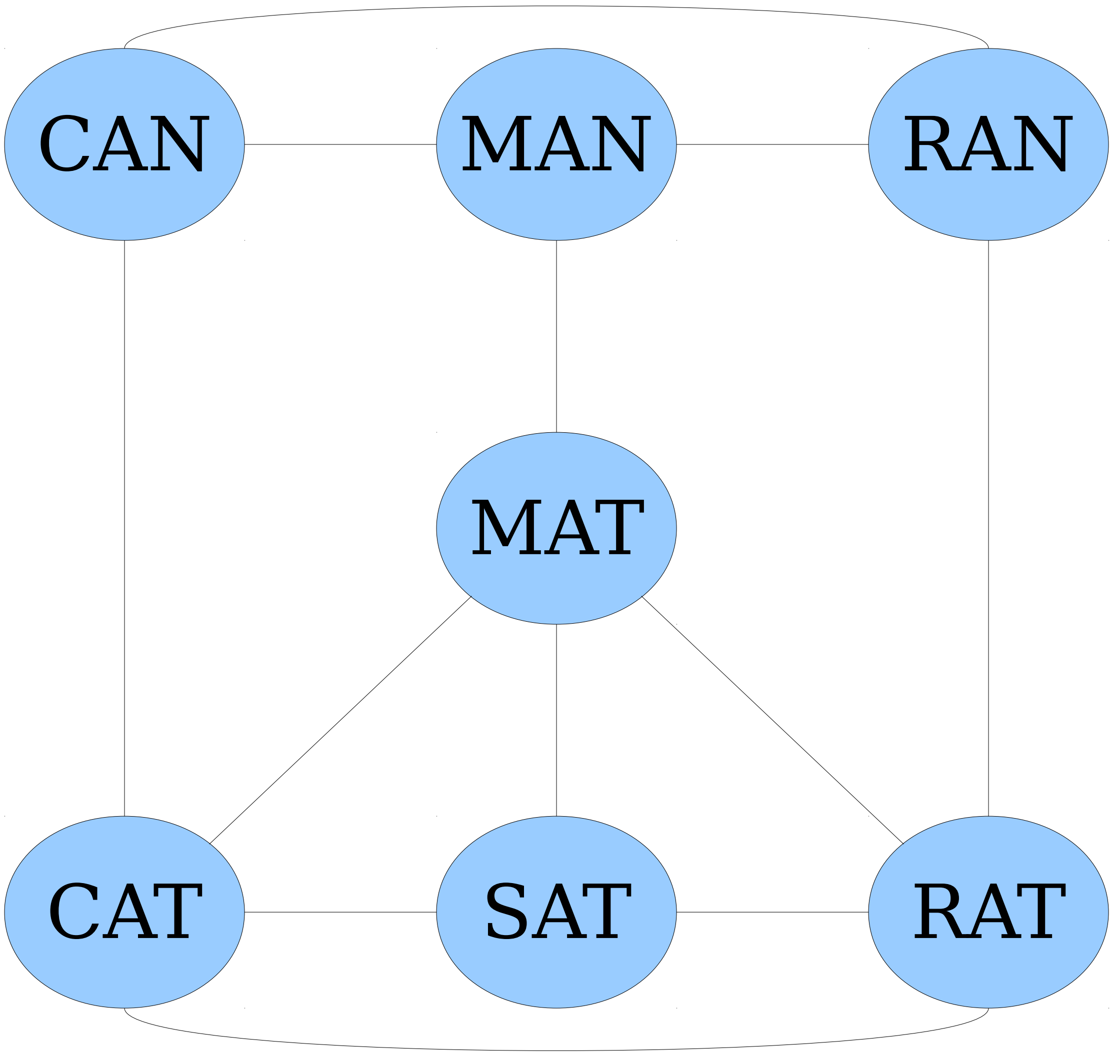


A graph consists of a set of ***nodes*** (or ***vertices***) connected by ***edges*** (or ***arcs***)

Some graphs are ***directed***.



Some graphs are ***undirected***.



Going forward, we're primarily going to focus on undirected graphs.

The term “graph” generally refers to undirected graphs with a finite number of nodes, unless specified otherwise.

Formalizing Graphs

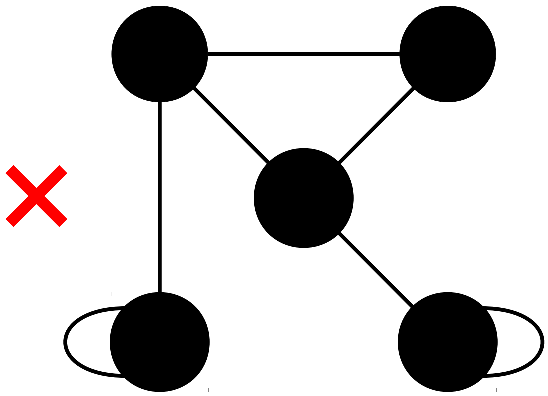
- How might we define a graph mathematically?
- We need to specify
  - what the nodes in the graph are, and
  - which edges are in the graph.
- The nodes can be pretty much anything.
- What about the edges?

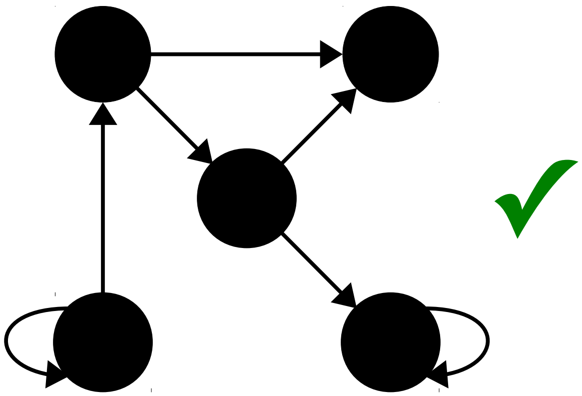
Formalizing Graphs

- An ***unordered pair*** is a set $\{a, b\}$ of two elements $a \neq b$. (Remember that sets are unordered).
  - $\{0, 1\} = \{1, 0\}$
- An ***undirected graph*** is an ordered pair $G = (V, E)$, where
  - V is a set of nodes, which can be anything, and
  - E is a set of edges, which are unordered pairs of nodes drawn from V.
- A ***directed graph*** is an ordered pair $G = (V, E)$, where
  - V is a set of nodes, which can be anything, and
  - E is a set of edges, which are *ordered* pairs of nodes drawn from V.

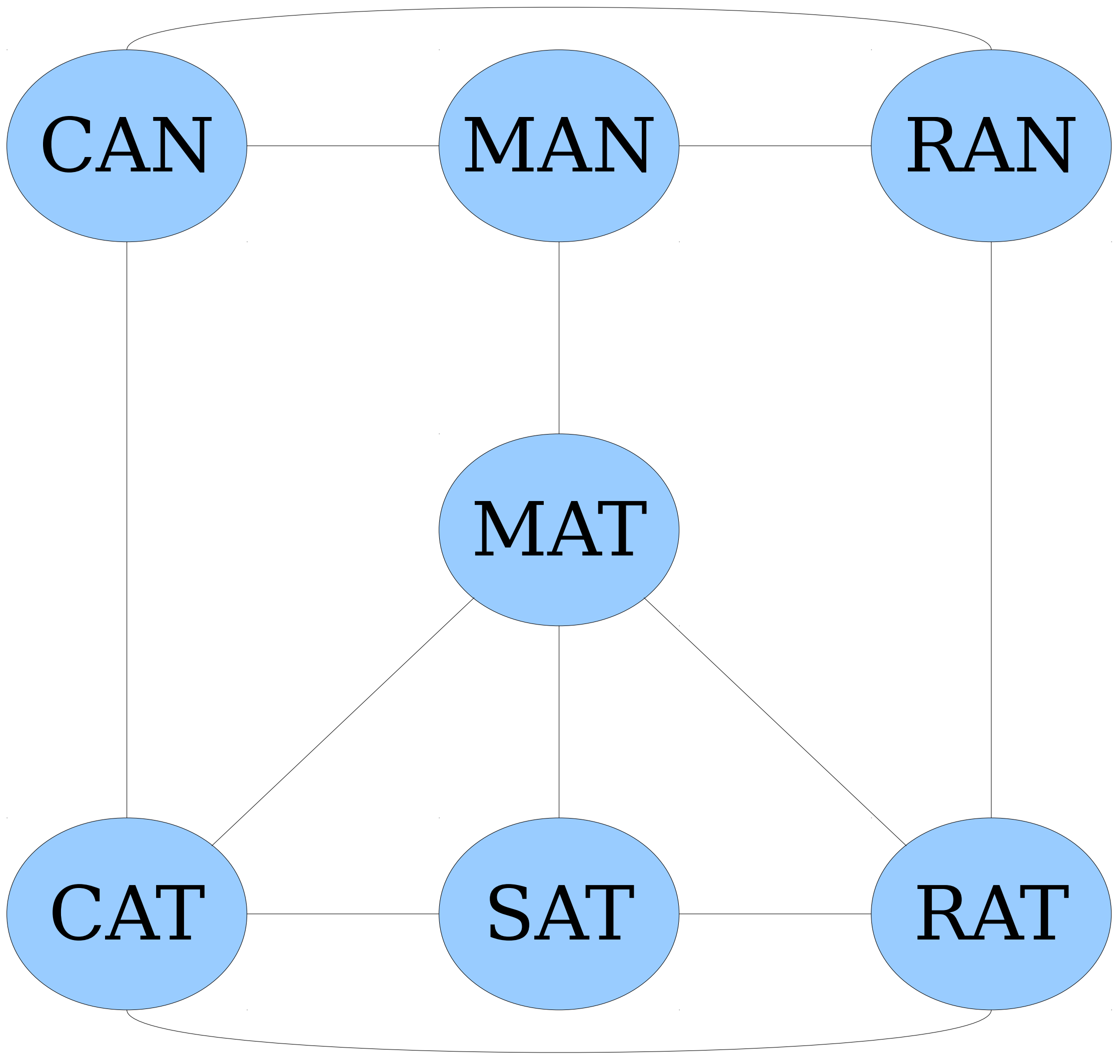
Self-Loops

- An edge from a node to itself is called a ***self-loop***.
- In undirected graphs, self-loops are generally not allowed.
  - Can you see how this follows from the definition?
- In directed graphs, self-loops are generally allowed unless specified otherwise.

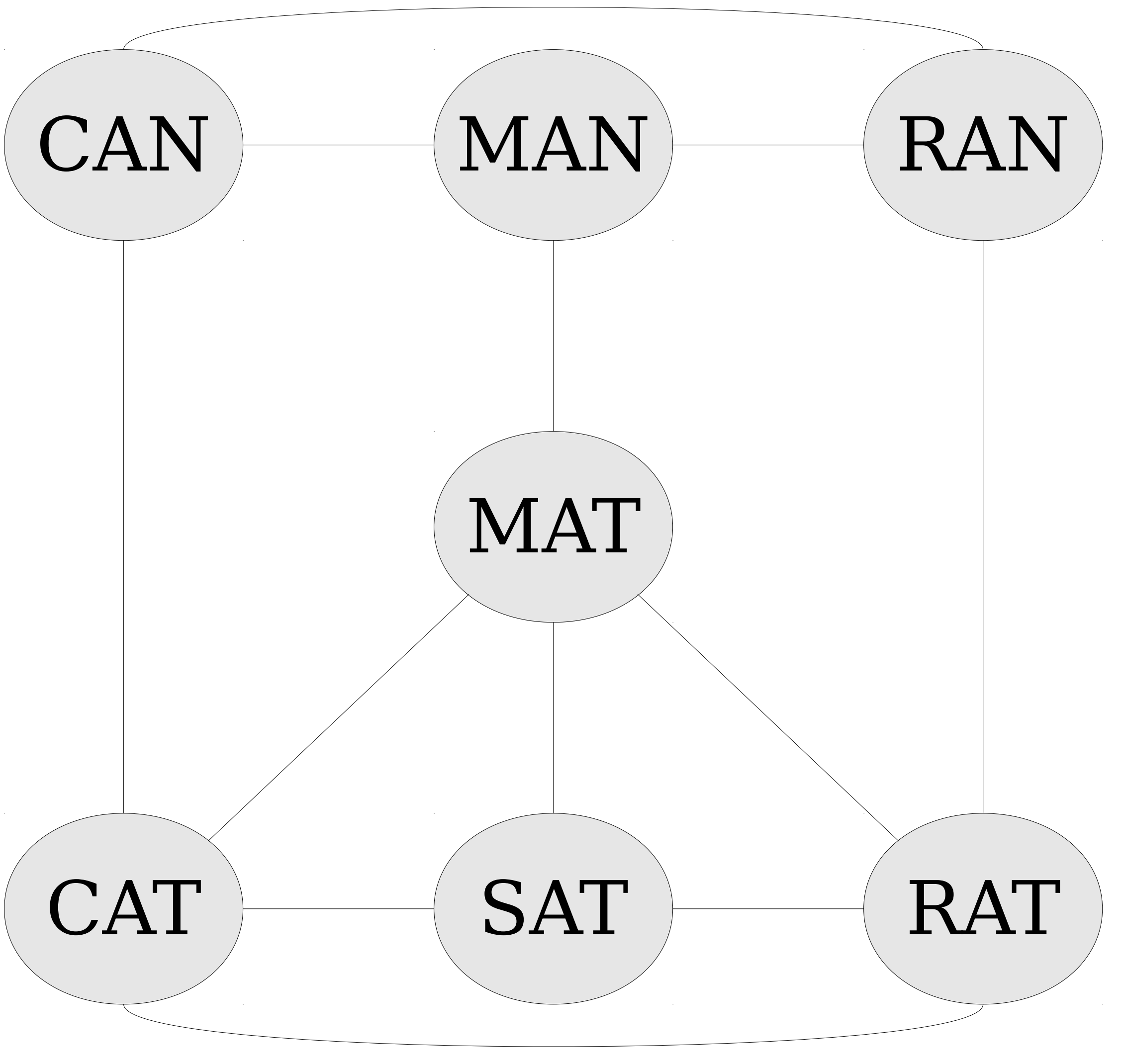




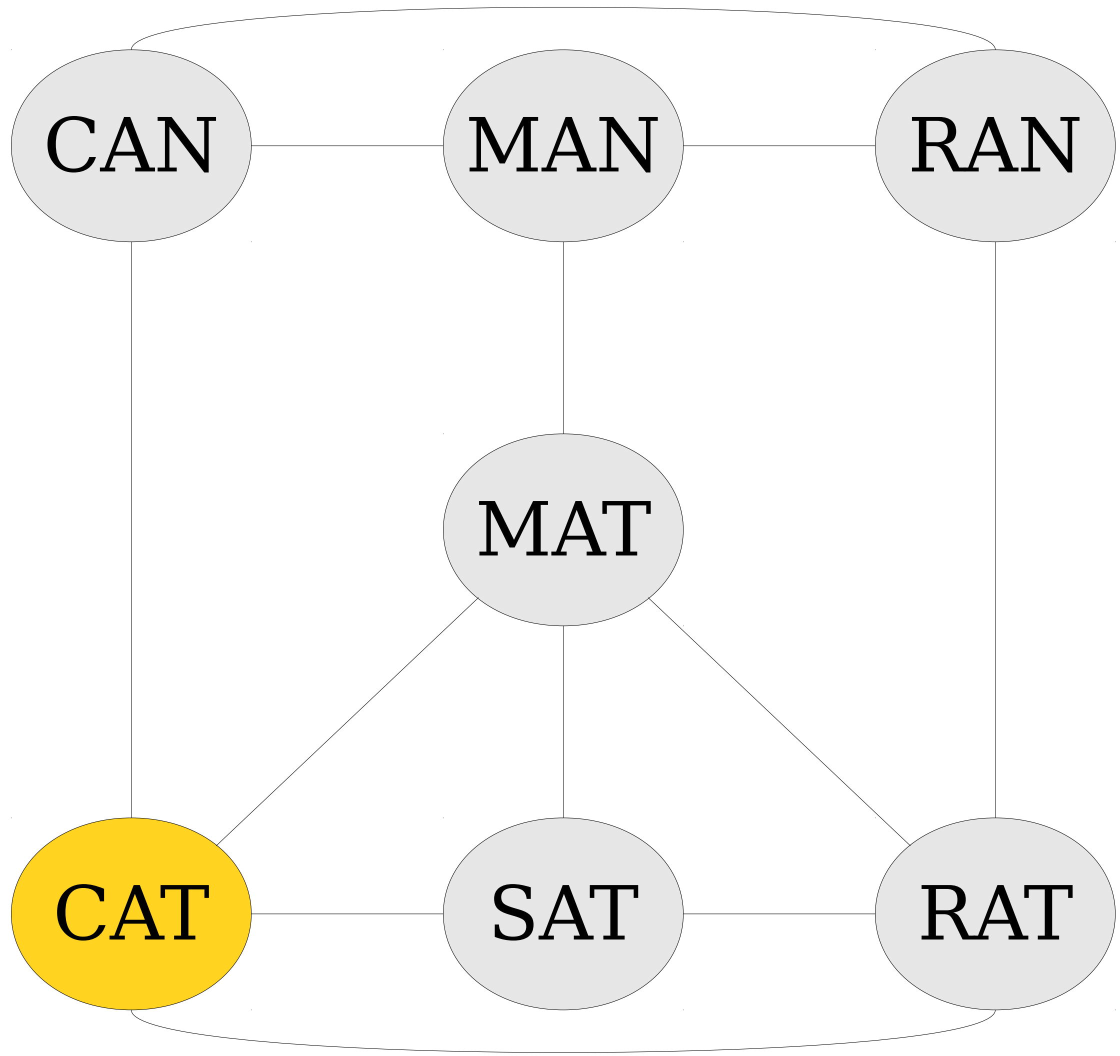
Standard Graph Terminology



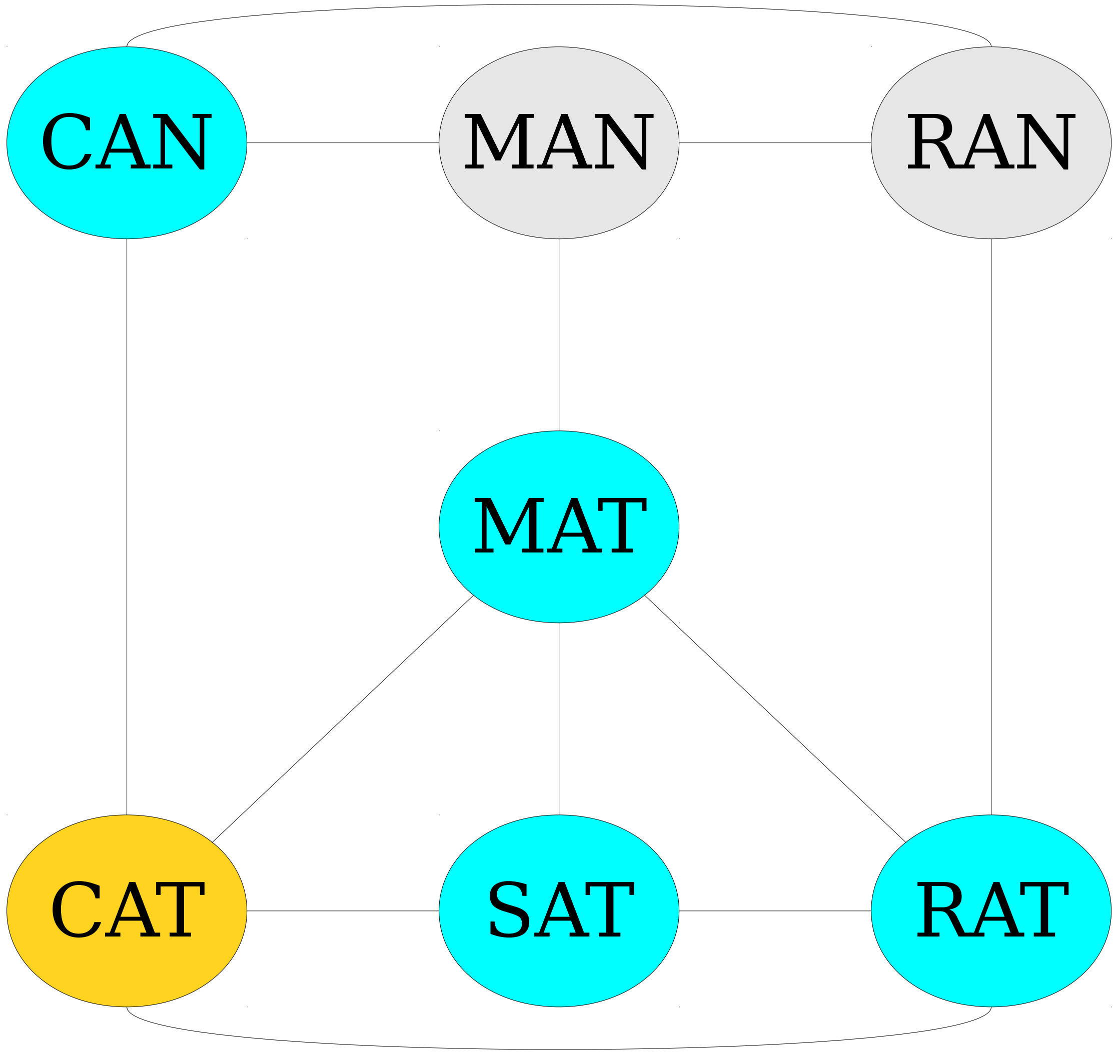
Two nodes are called ***adjacent*** if there is an edge between them.



Two nodes are called ***adjacent*** if there is an edge between them.



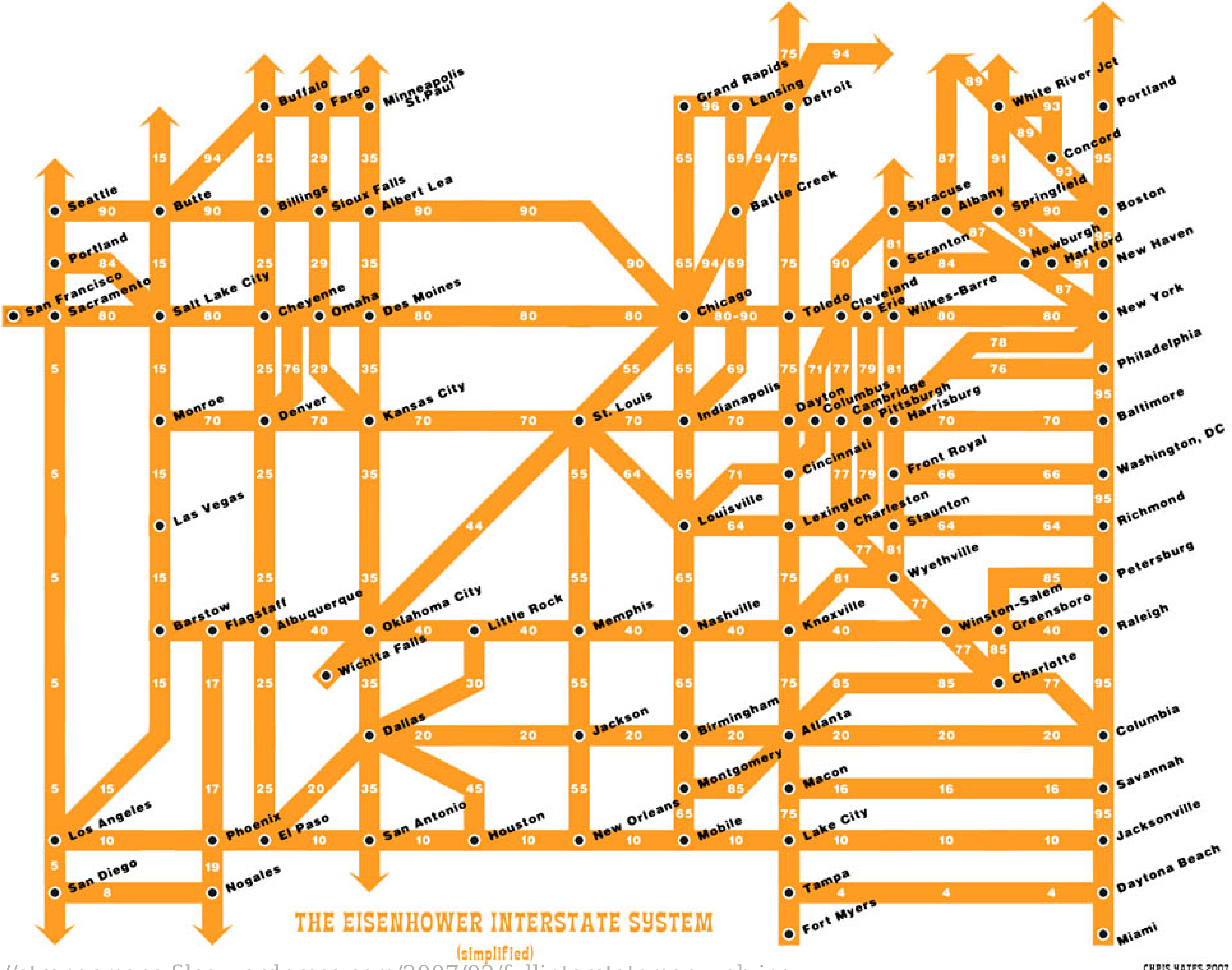
Two nodes are called ***adjacent*** if there is an edge between them.



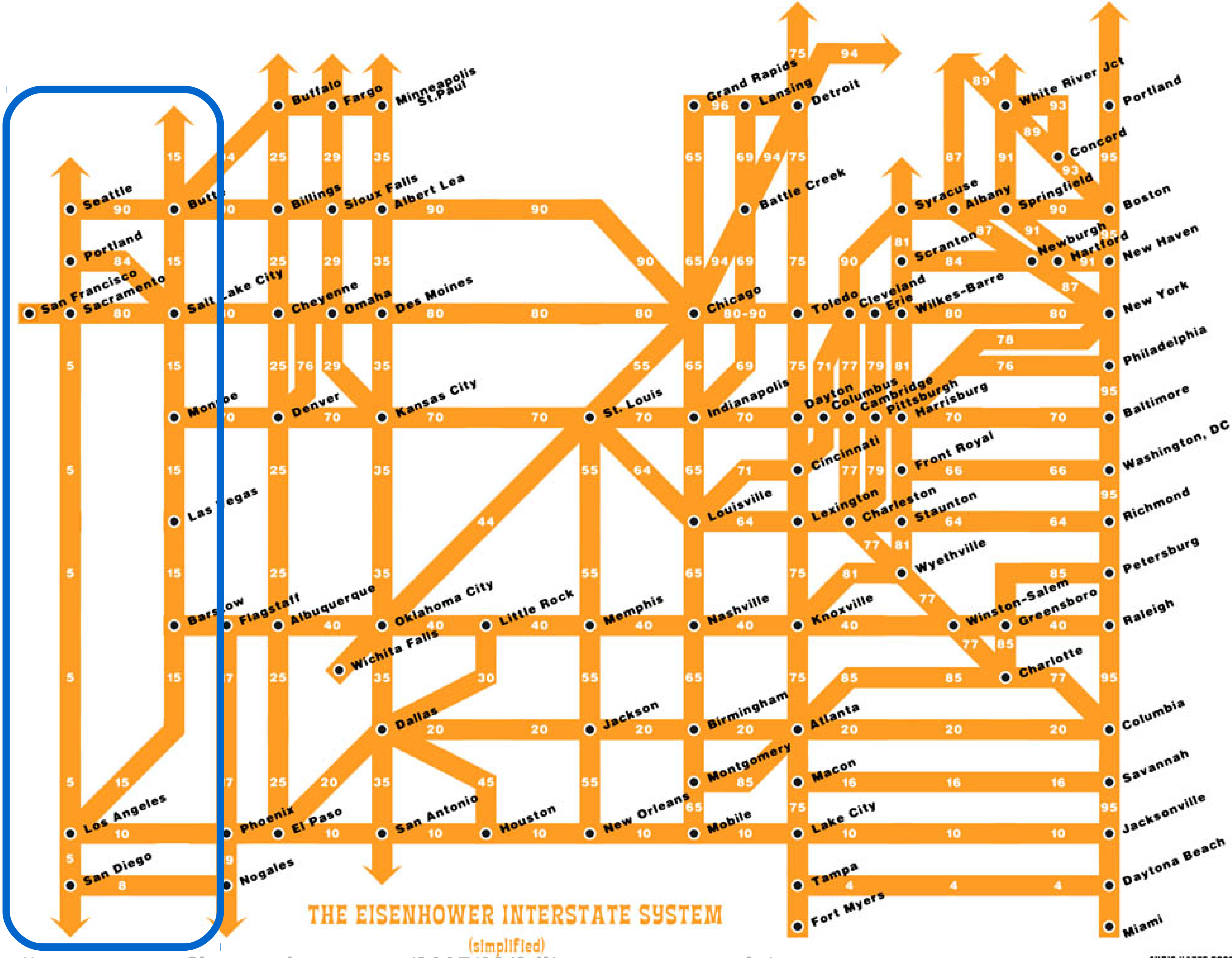
Two nodes are called ***adjacent*** if there is an edge between them.

Using our Formalisms

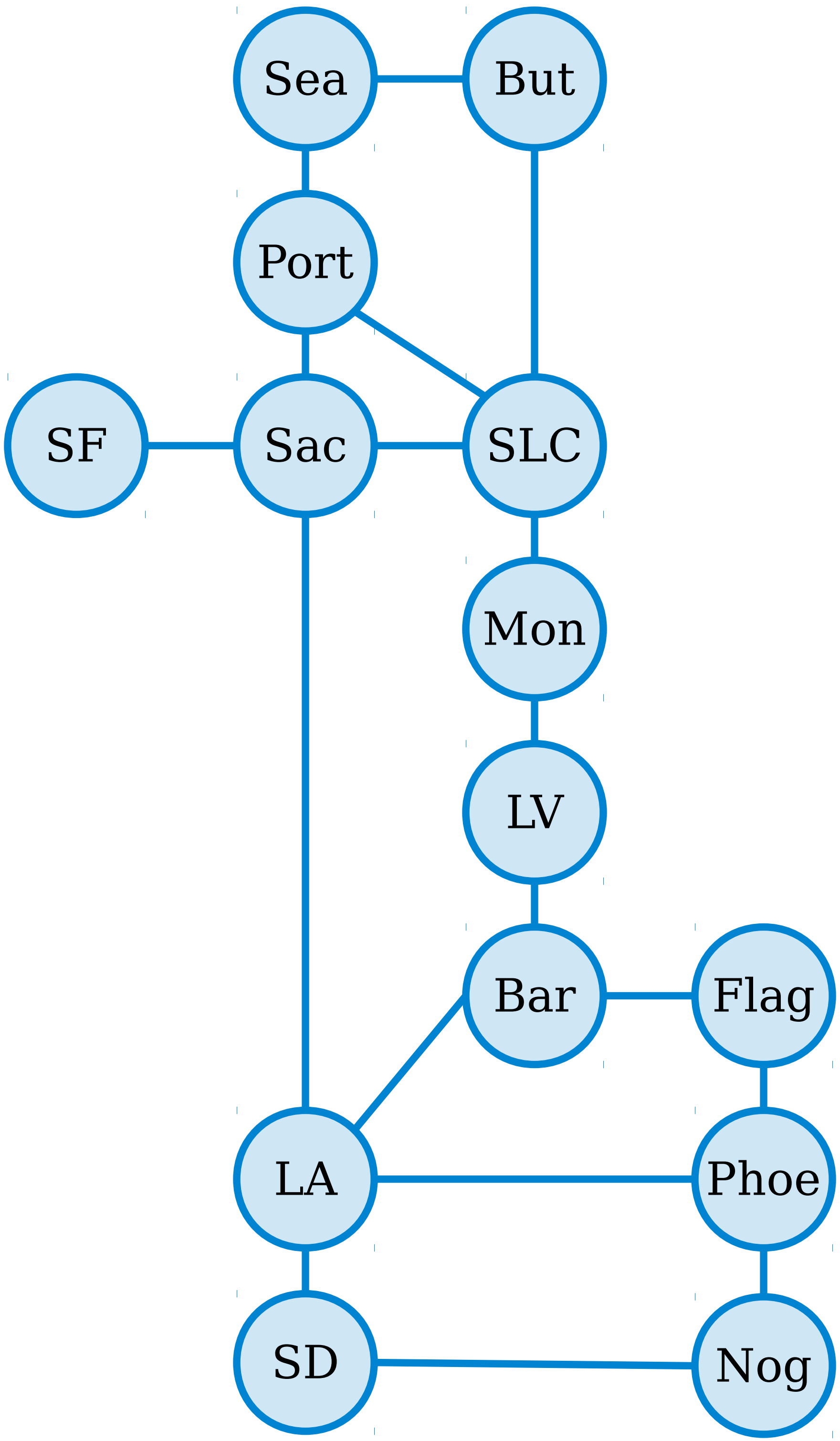
- Let $G = (V, E)$ be a graph.
- Intuitively, two nodes are adjacent if they're linked by an edge.
- Formally speaking, we say that two nodes $u, v \in V$ are ***adjacent*** if $\{u, v\} \in E$.

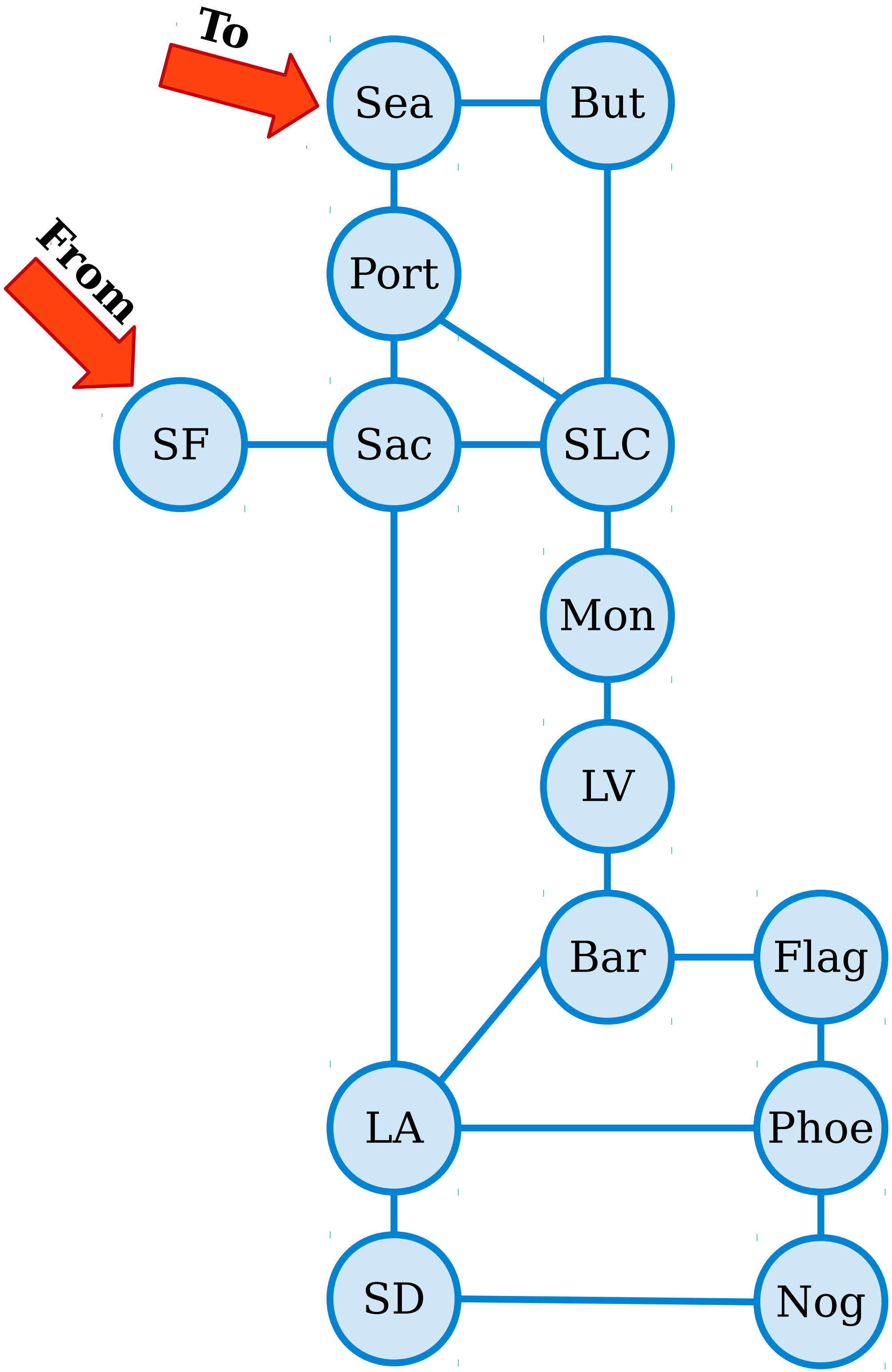


http://strangemaps.files.wordpress.com/2007/02/fullinterstatemap-web.jpg

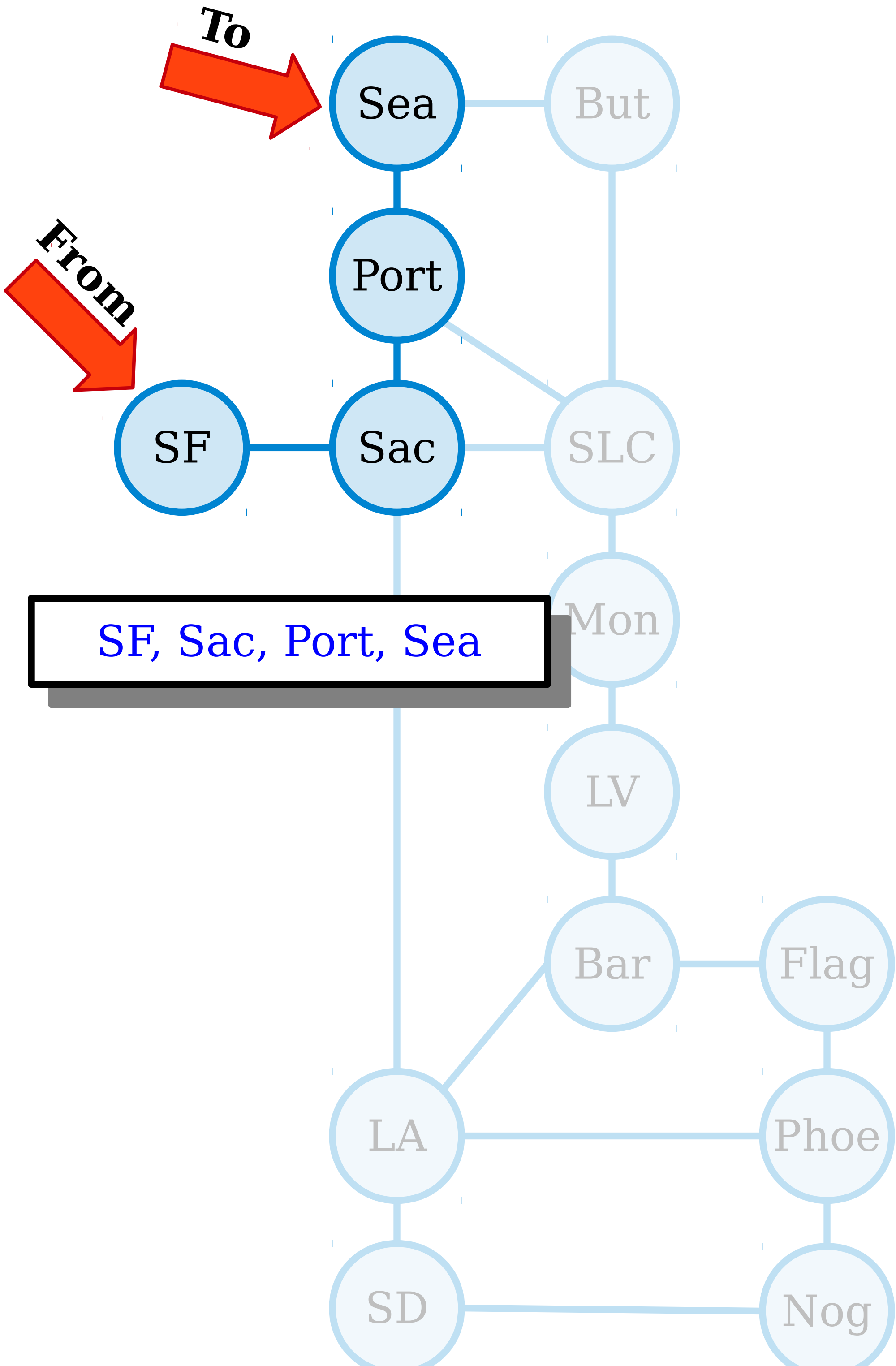


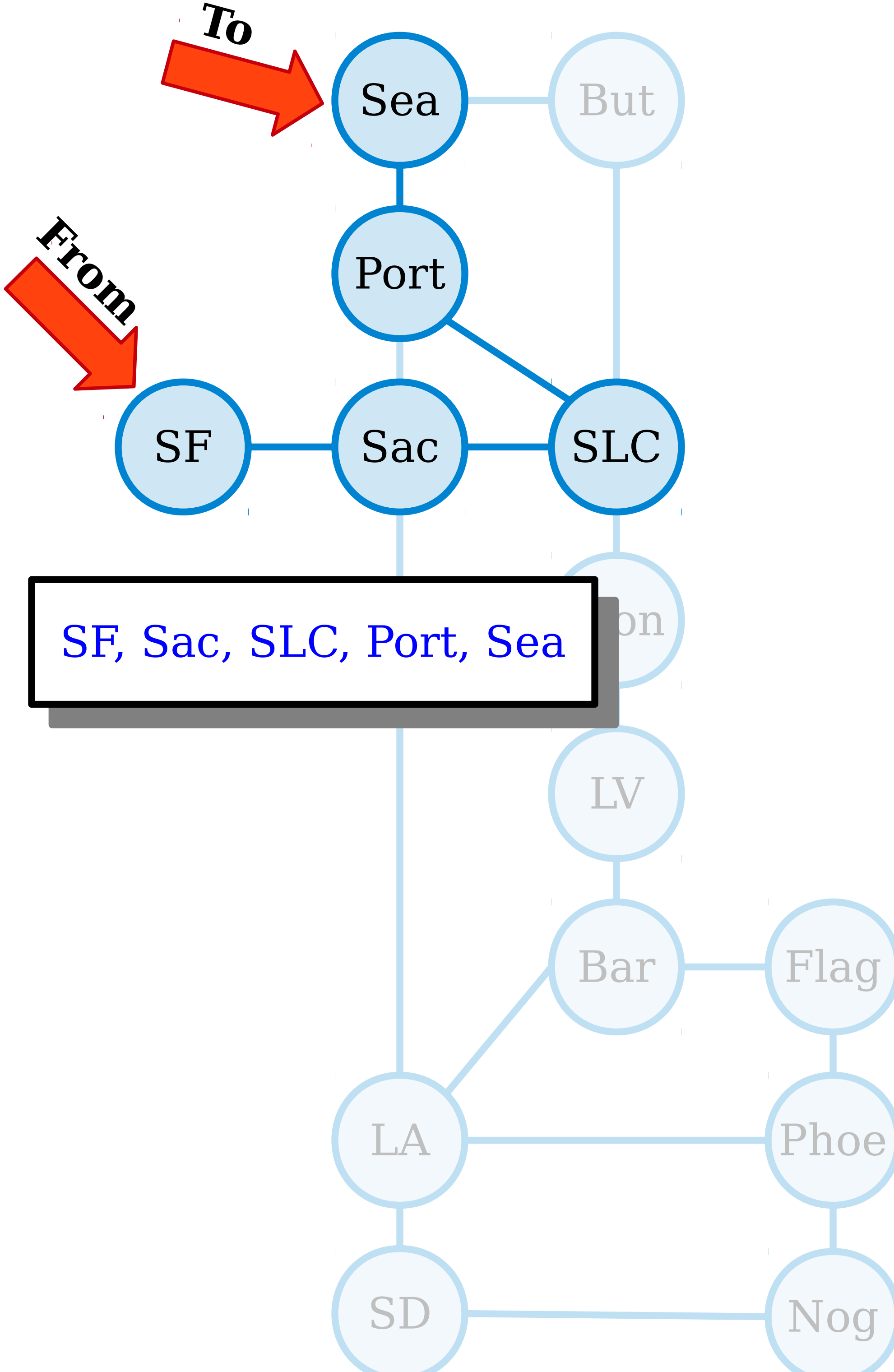
http://strangemaps.files.wordpress.com/2007/02/fullinterstatemap-web.jpg

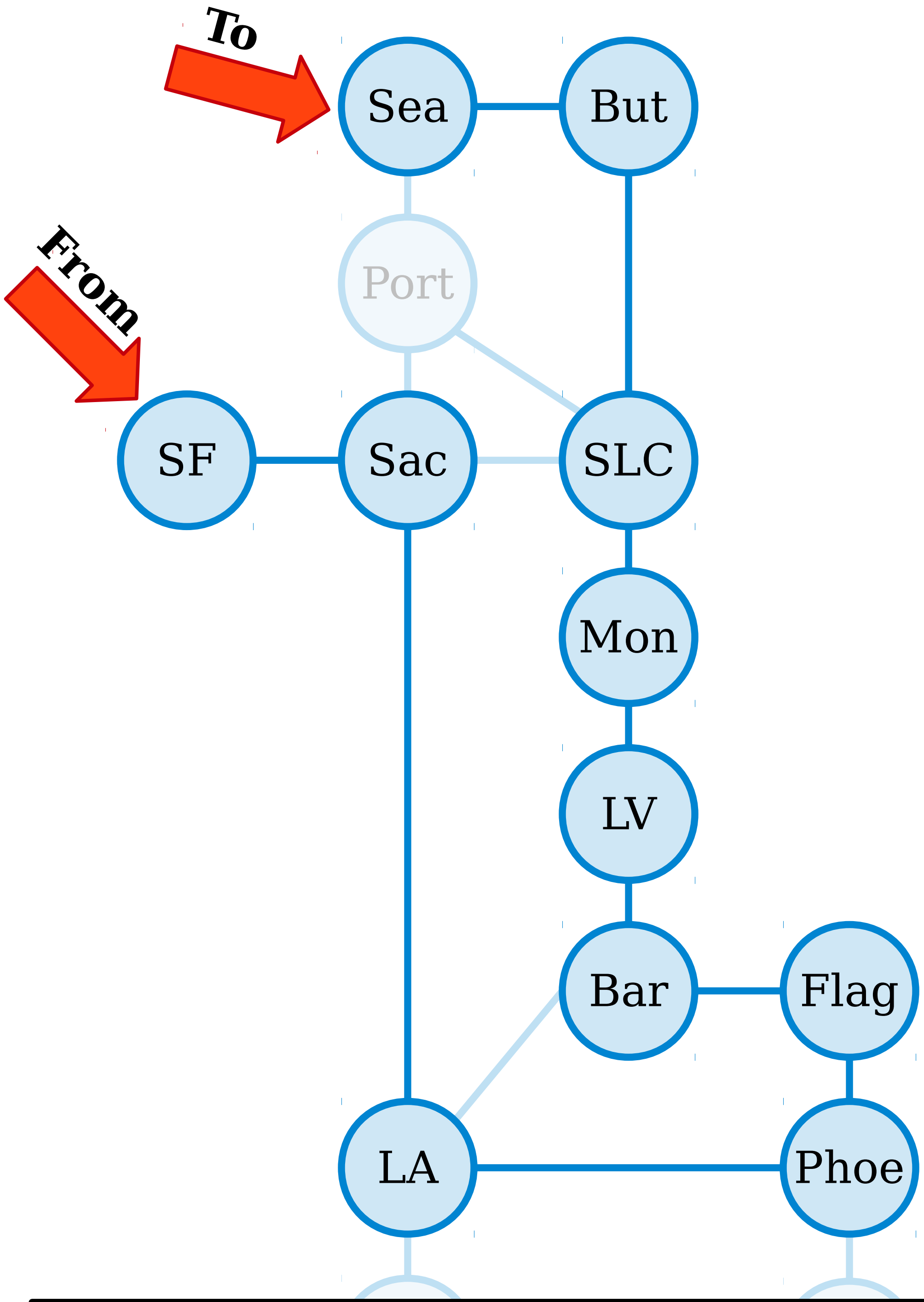

Sea
But
Port
SF
Sac
SLC
Mon
LV
Bar
Flag
LA
Phoe
SD
Nog



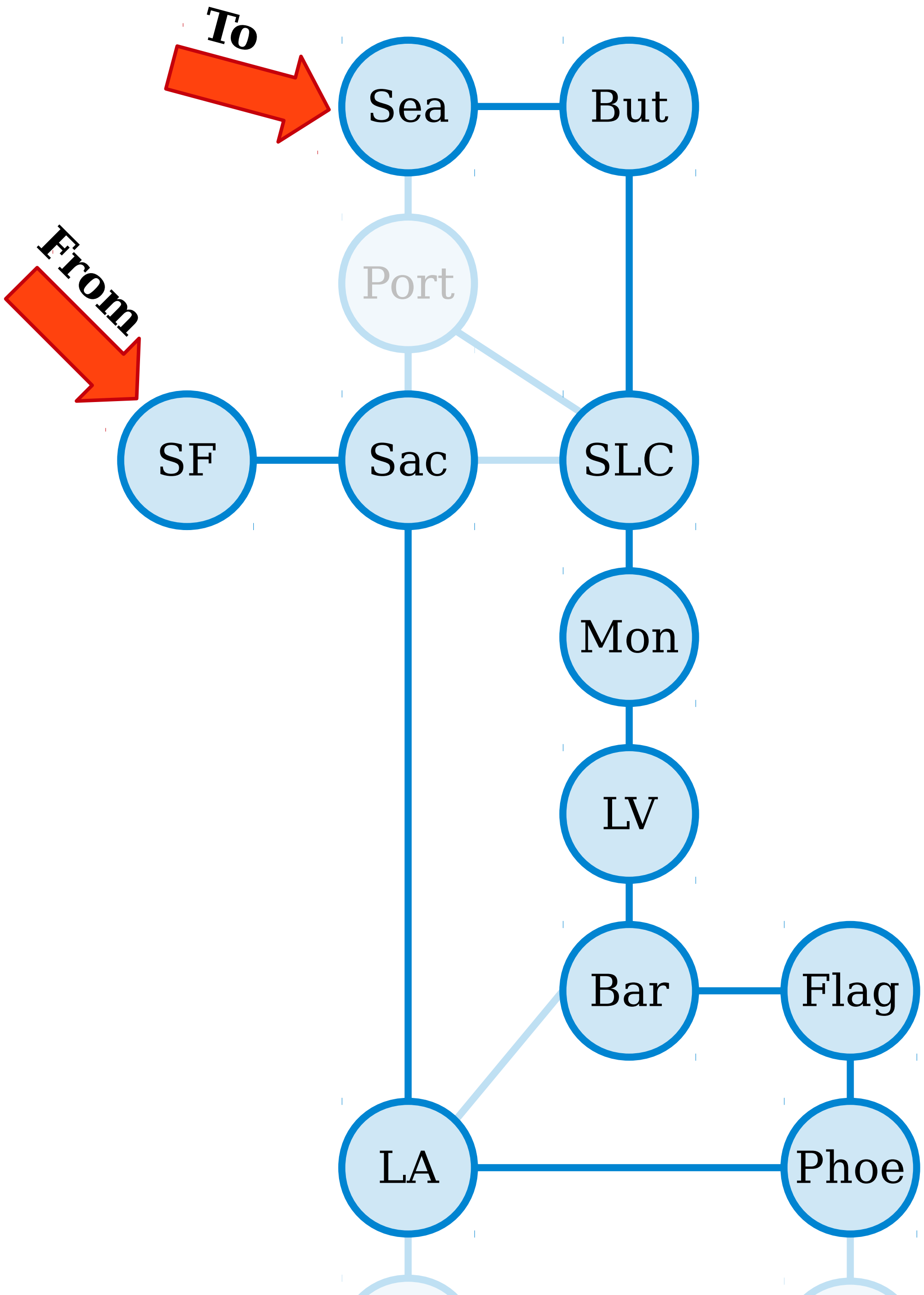
To
From
Sea
But
Port
SF
Sac
SLC
Mon
LV
Bar
Flag
LA
Phoe
SD
Nog


To
From
Sea
But
Port
SF
Sac
SLC
Mon
SF, Sac, Port, Sea
LV
Bar
Flag
LA
Phoe
SD
Nog


To
From
Sea
But
Port
SF
Sac
SLC
SF, Sac, SLC, Port, Sea
LV
Bar
Flag
LA
Phoe
SD
Nog

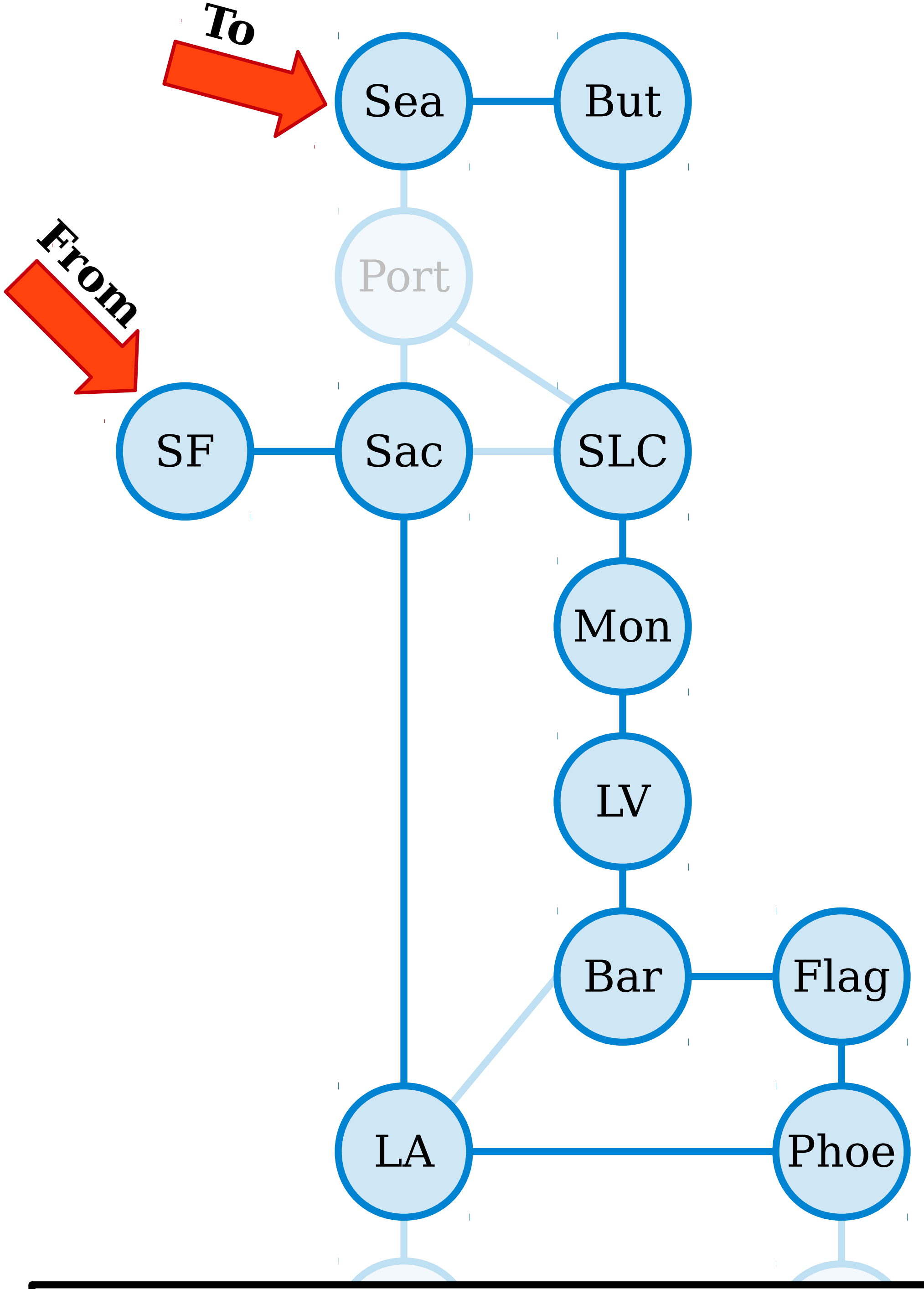


SF, Sac, LA, Phoe, Flag, Bar, LV, Mon, SLC, But, Sea



A ***path*** in a graph $G = (V, E)$ is a sequence of one or more nodes $v_1, v_2, v_3, \ldots, v_n$ such that any two consecutive nodes in the sequence are adjacent.

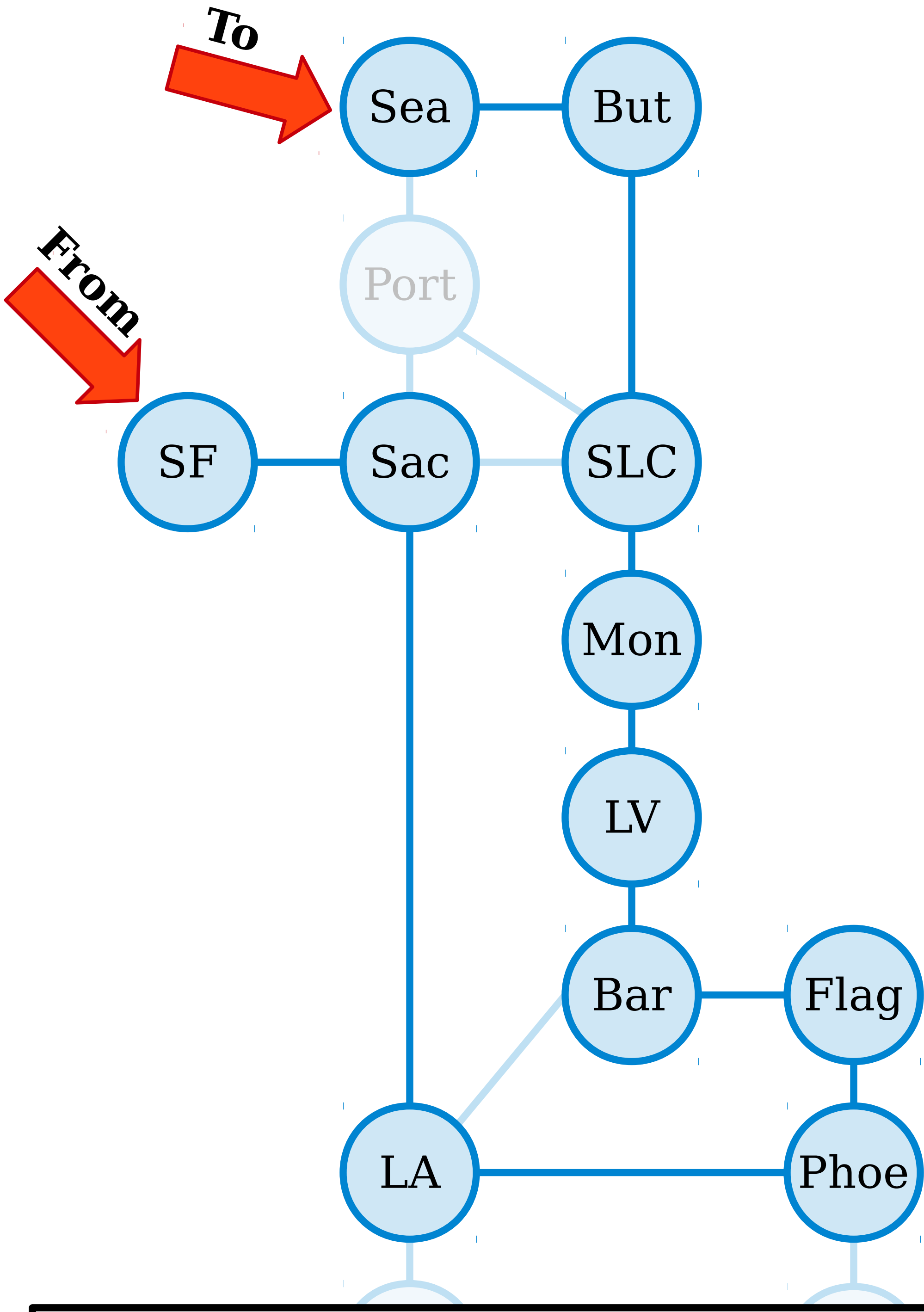
SF, Sac, LA, Phoe, Flag, Bar, LV, Mon, SLC, But, Sea



A ***path*** in a graph $G = (V, E)$ is a sequence of one or more nodes $v_1, v_2, v_3, \ldots, v_n$ such that any two consecutive nodes in the sequence are adjacent.

The ***length*** of the path $v_1, \ldots, v_n$ is $n - 1$.

SF, Sac, LA, Phoe, Flag, Bar, LV, Mon, SLC, But, Sea

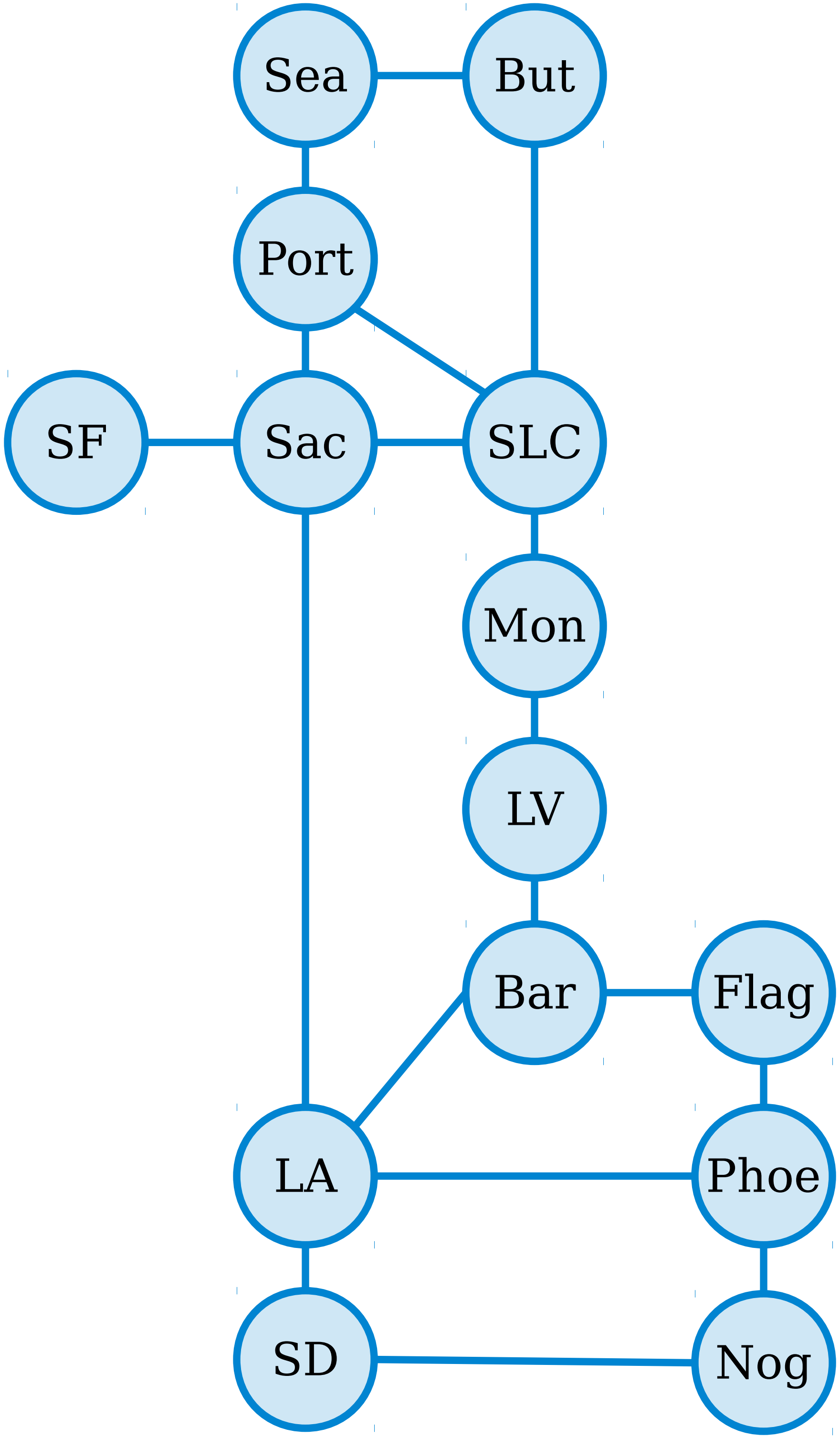


A ***path*** in a graph $G = (V, E)$ is a sequence of one or more nodes $v_1, v_2, v_3, \ldots, v_n$ such that any two consecutive nodes in the sequence are adjacent.

The ***length*** of the path $v_1, \ldots, v_n$ is $n - 1$.

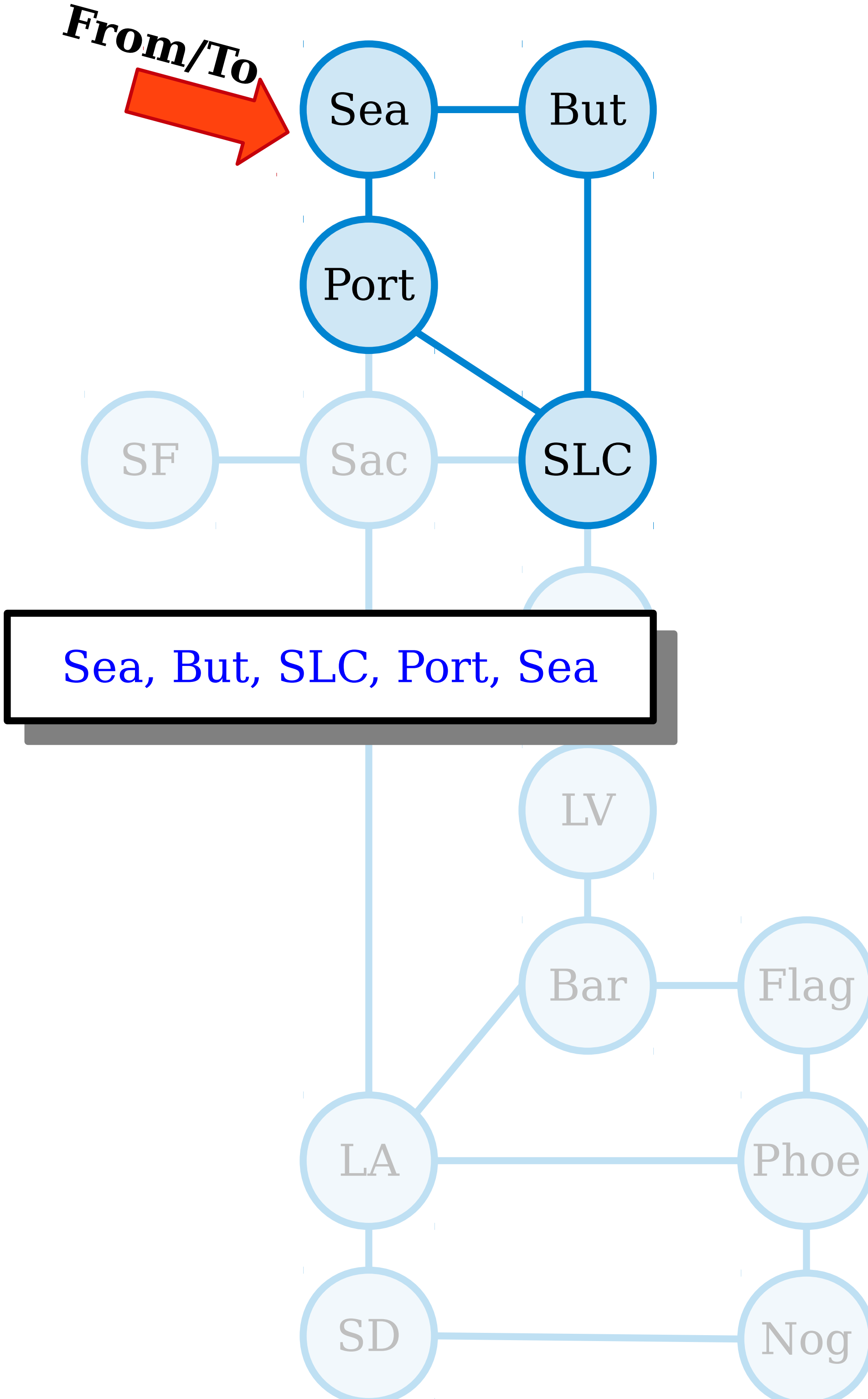
(This path has length 10, but visits 11 cities.)

SF, Sac, LA, Phoe, Flag, Bar, LV, Mon, SLC, But, Sea



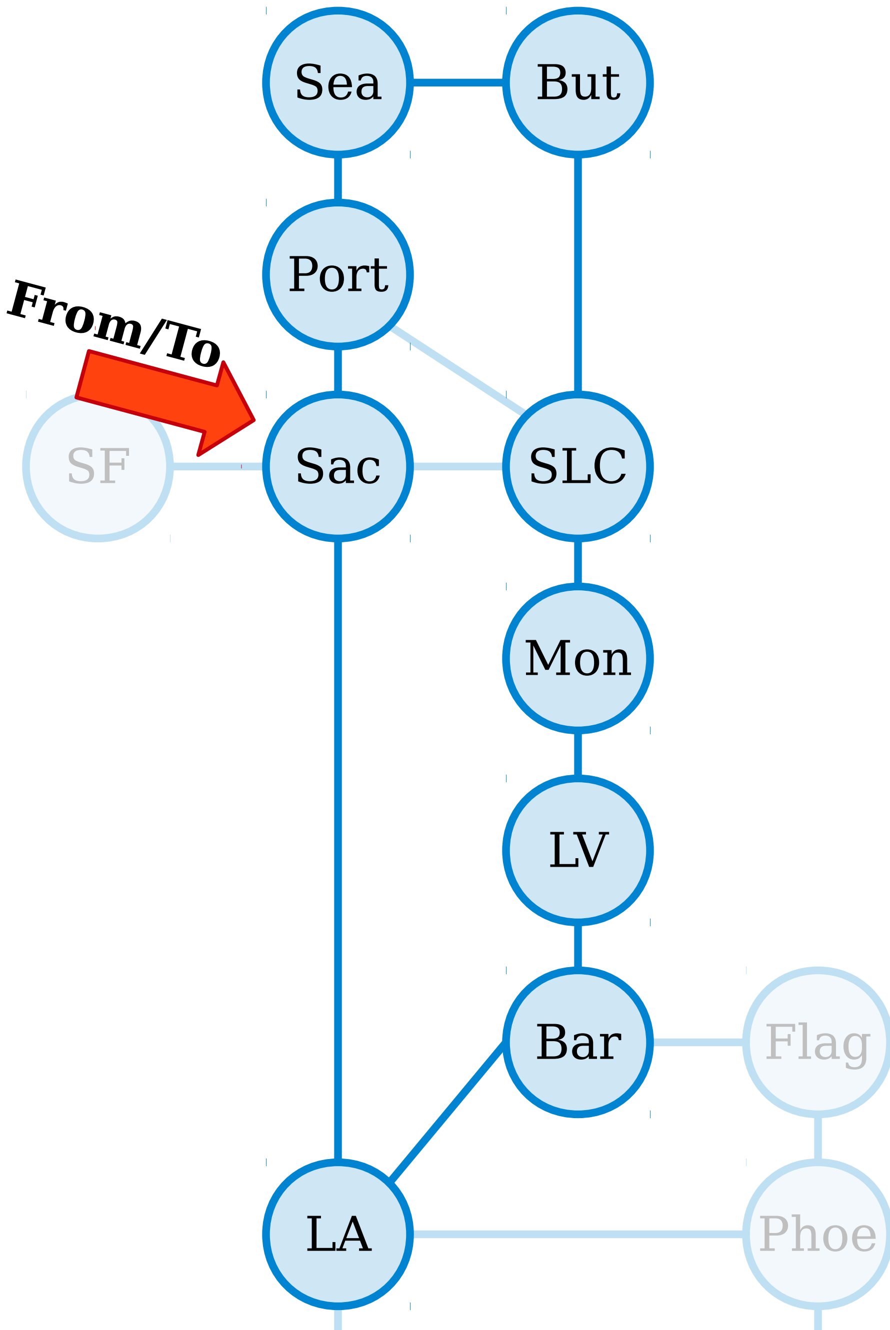
A ***path*** in a graph $G = (V, E)$ is a sequence of one or more nodes $v_1, v_2, v_3, \ldots, v_n$ such that any two consecutive nodes in the sequence are adjacent.

The ***length*** of the path $v_1, \ldots, v_n$ is $n - 1$.



A ***path*** in a graph $G = (V, E)$ is a sequence of one or more nodes $v_1, v_2, v_3, \ldots, v_n$ such that any two consecutive nodes in the sequence are adjacent.

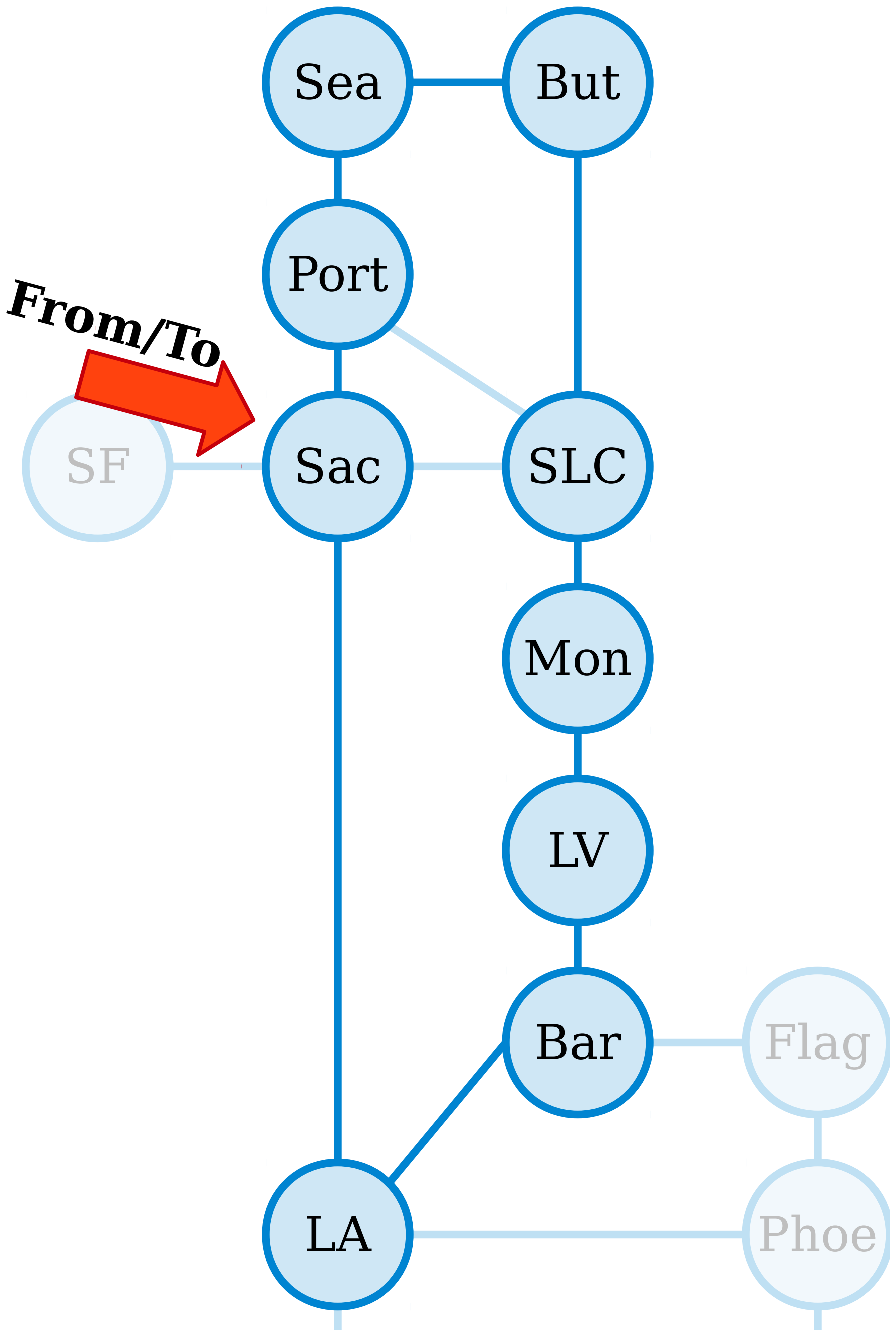
The ***length*** of the path $v_1, \ldots, v_n$ is $n - 1$.



A ***path*** in a graph $G = (V, E)$ is a sequence of one or more nodes $v_1, v_2, v_3, \ldots, v_n$ such that any two consecutive nodes in the sequence are adjacent.

The ***length*** of the path $v_1, \ldots, v_n$ is $n - 1$.

Sac, Port, Sea, But, SLC, Mon, LV, Bar, LA, Sac

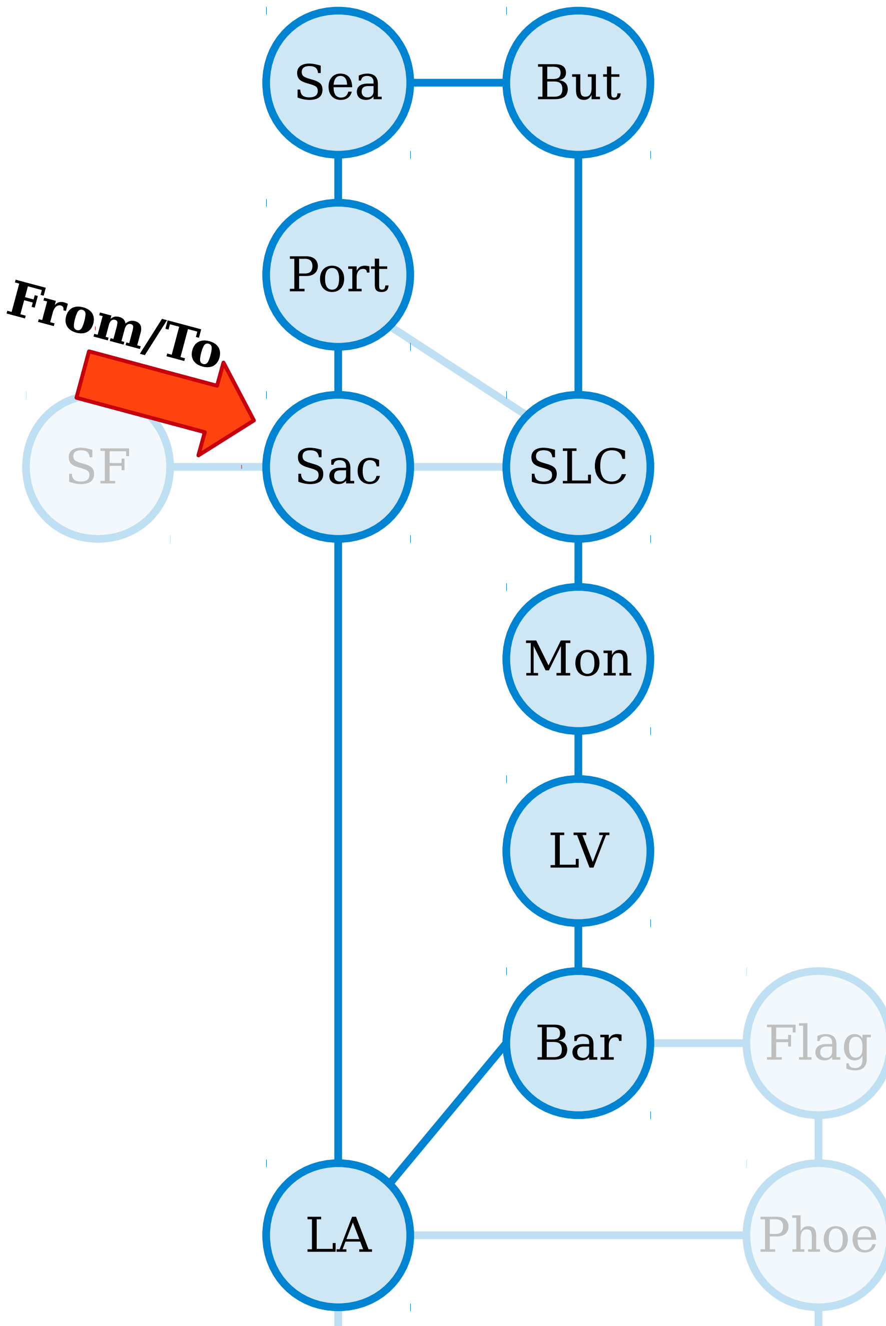


A ***path*** in a graph $G = (V, E)$ is a sequence of one or more nodes $v_1, v_2, v_3, \ldots, v_n$ such that any two consecutive nodes in the sequence are adjacent.

The ***length*** of the path $v_1, \ldots, v_n$ is $n - 1$.

A ***cycle*** in a graph is a path from a node back to itself. (By convention, a cycle cannot have length zero.)

Sac, Port, Sea, But, SLC, Mon, LV, Bar, LA, Sac



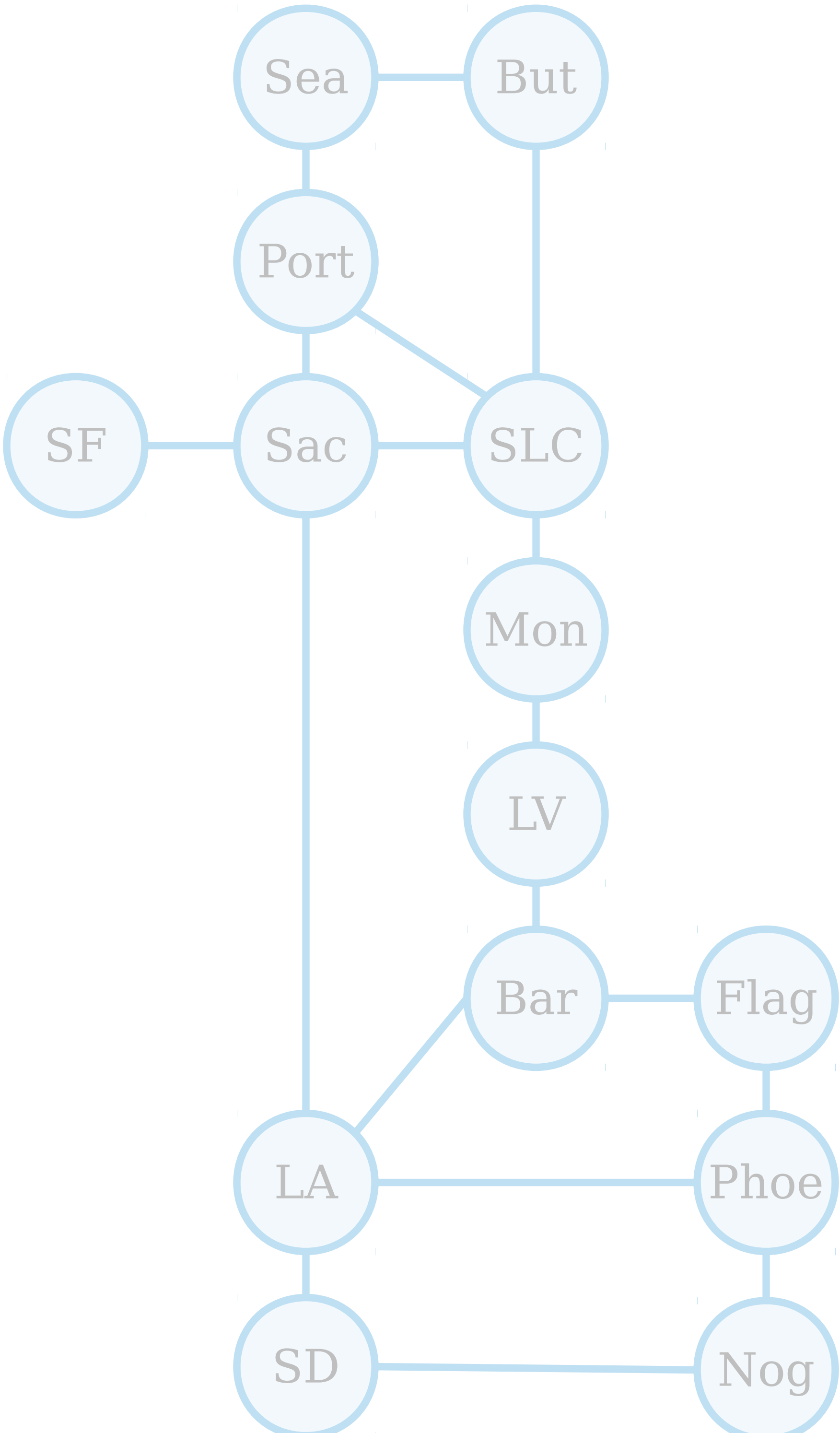
A ***path*** in a graph $G = (V, E)$ is a sequence of one or more nodes $v_1, v_2, v_3, \ldots, v_n$ such that any two consecutive nodes in the sequence are adjacent.

The ***length*** of the path $v_1, \ldots, v_n$ is $n - 1$.

A ***cycle*** in a graph is a path from a node back to itself. (By convention, a cycle cannot have length zero.)

(This cycle has length nine and visits nine different cities.)

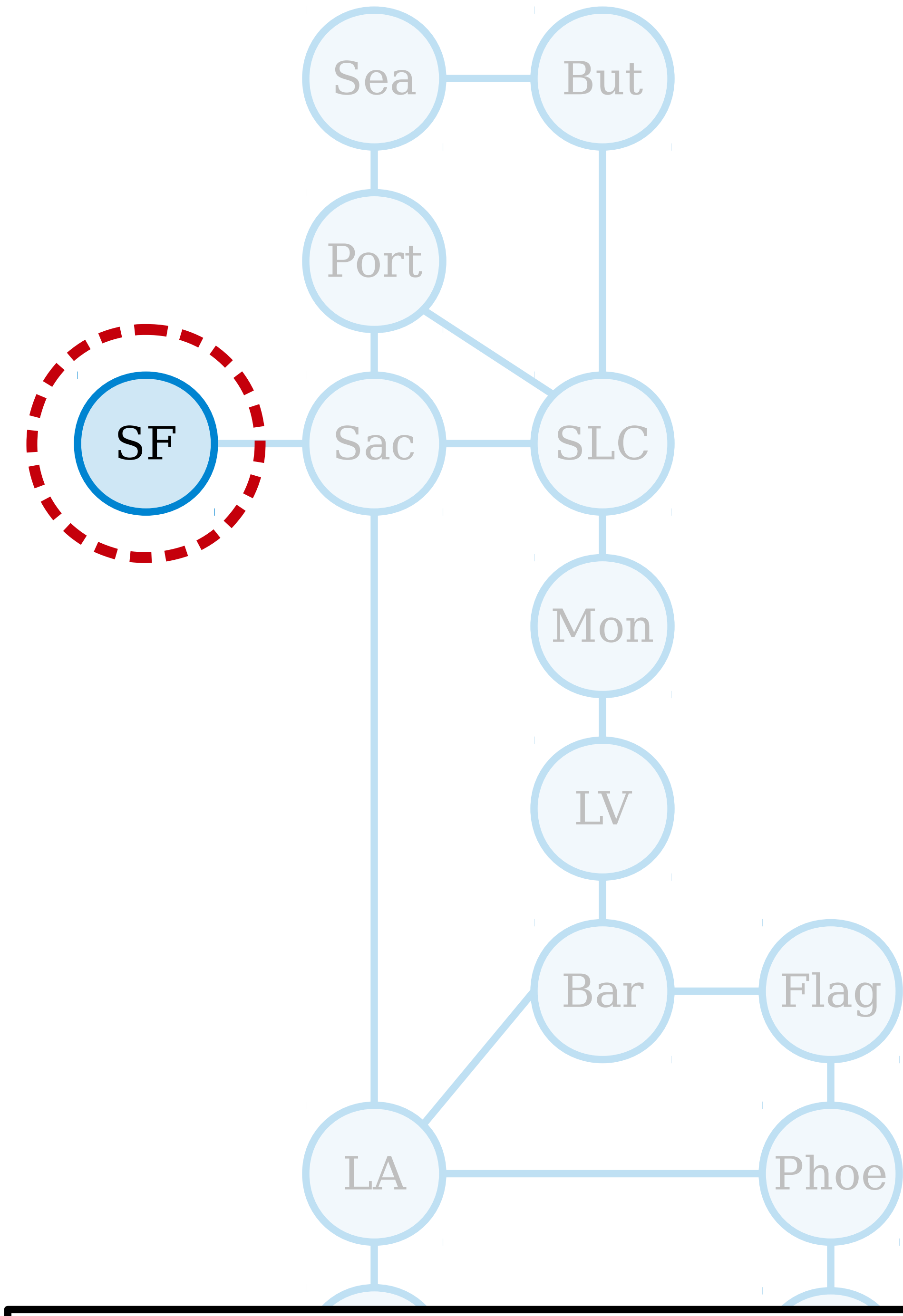
Sac, Port, Sea, But, SLC, Mon, LV, Bar, LA, Sac



A ***path*** in a graph $G = (V, E)$ is a sequence of one or more nodes $v_1, v_2, v_3, \ldots, v_n$ such that any two consecutive nodes in the sequence are adjacent.

The ***length*** of the path $v_1, \ldots, v_n$ is $n - 1$.

A ***cycle*** in a graph is a path from a node back to itself. (By convention, a cycle cannot have length zero.)

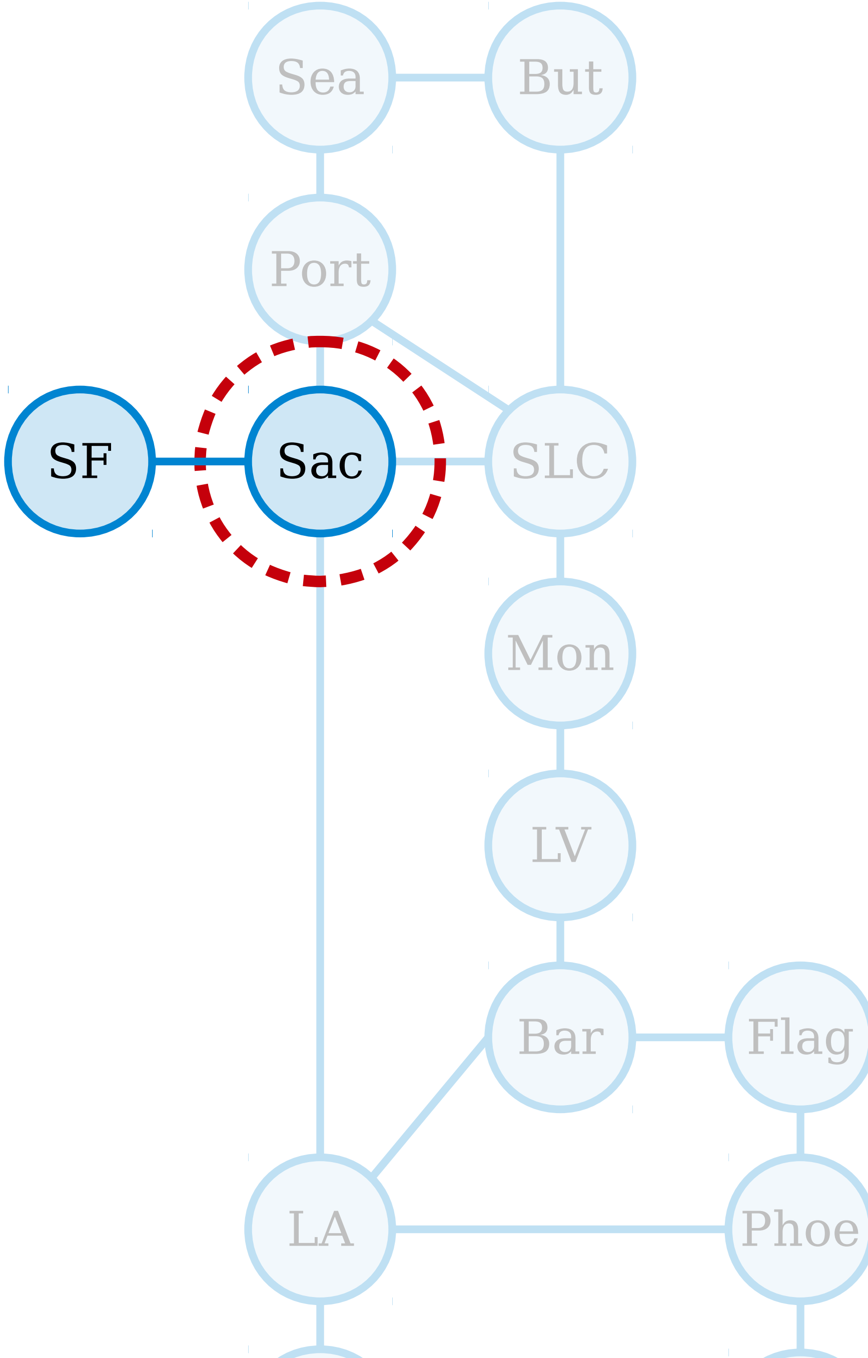


A ***path*** in a graph $G = (V, E)$ is a sequence of one or more nodes $v_1, v_2, v_3, \ldots, v_n$ such that any two consecutive nodes in the sequence are adjacent.

The ***length*** of the path $v_1, \ldots, v_n$ is $n - 1$.

A ***cycle*** in a graph is a path from a node back to itself. (By convention, a cycle cannot have length zero.)

SF

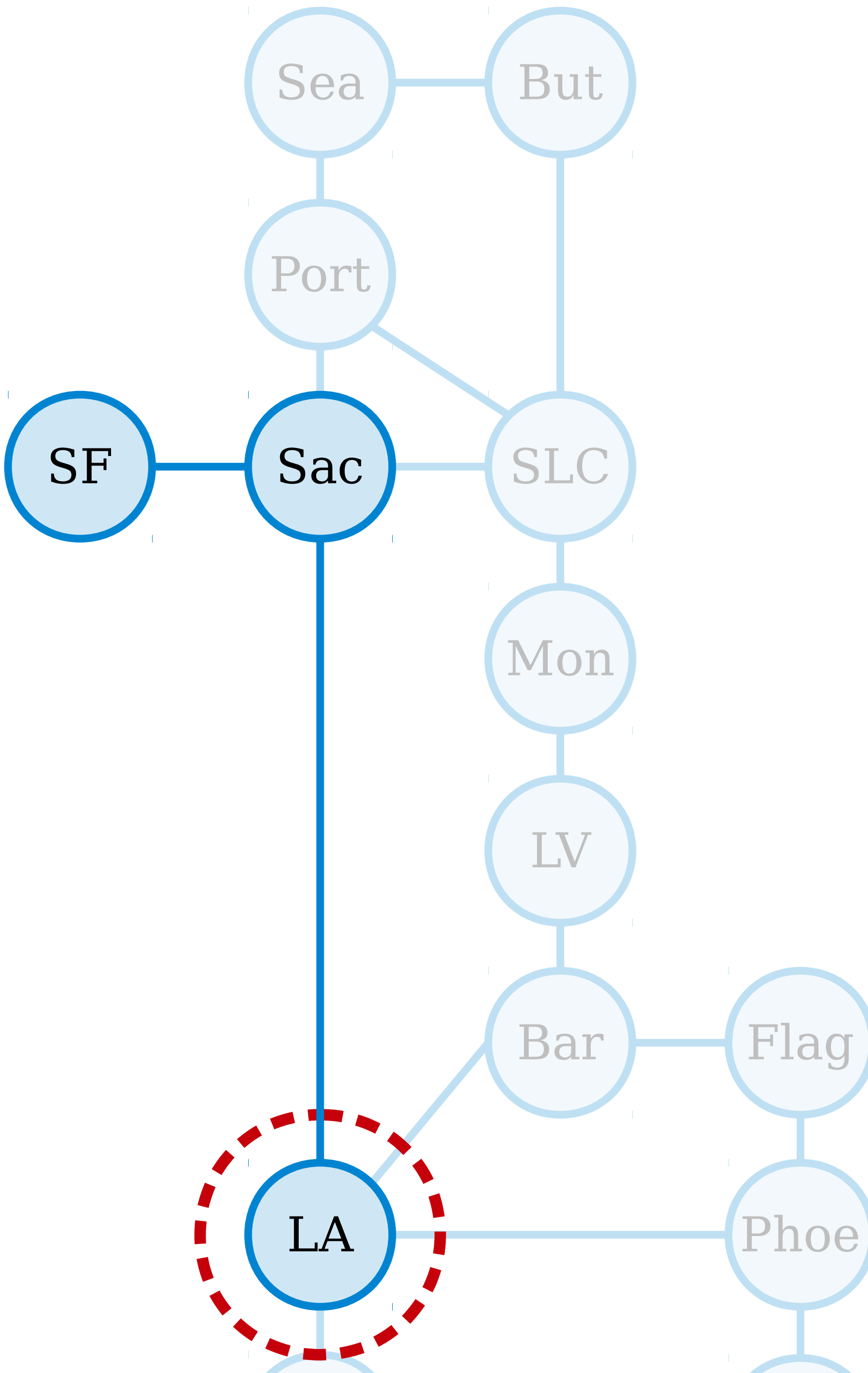


A ***path*** in a graph $G = (V, E)$ is a sequence of one or more nodes $v_1, v_2, v_3, \ldots, v_n$ such that any two consecutive nodes in the sequence are adjacent.

The ***length*** of the path $v_1, \ldots, v_n$ is $n - 1$.

A ***cycle*** in a graph is a path from a node back to itself. (By convention, a cycle cannot have length zero.)

SF, Sac

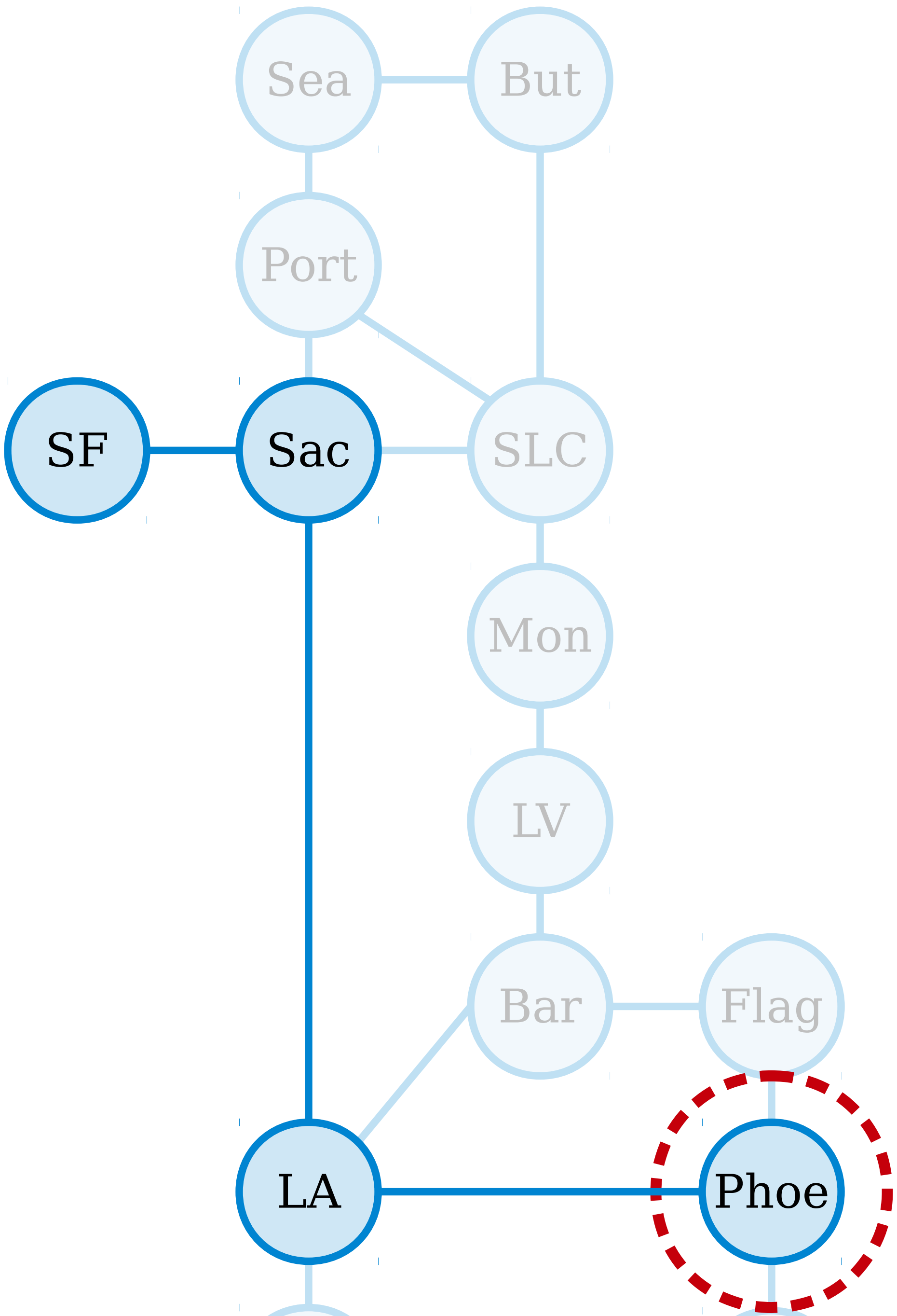


A ***path*** in a graph $G = (V, E)$ is a sequence of one or more nodes $v_1, v_2, v_3, \ldots, v_n$ such that any two consecutive nodes in the sequence are adjacent.

The ***length*** of the path $v_1, \ldots, v_n$ is $n - 1$.

A ***cycle*** in a graph is a path from a node back to itself. (By convention, a cycle cannot have length zero.)

SF, Sac, LA

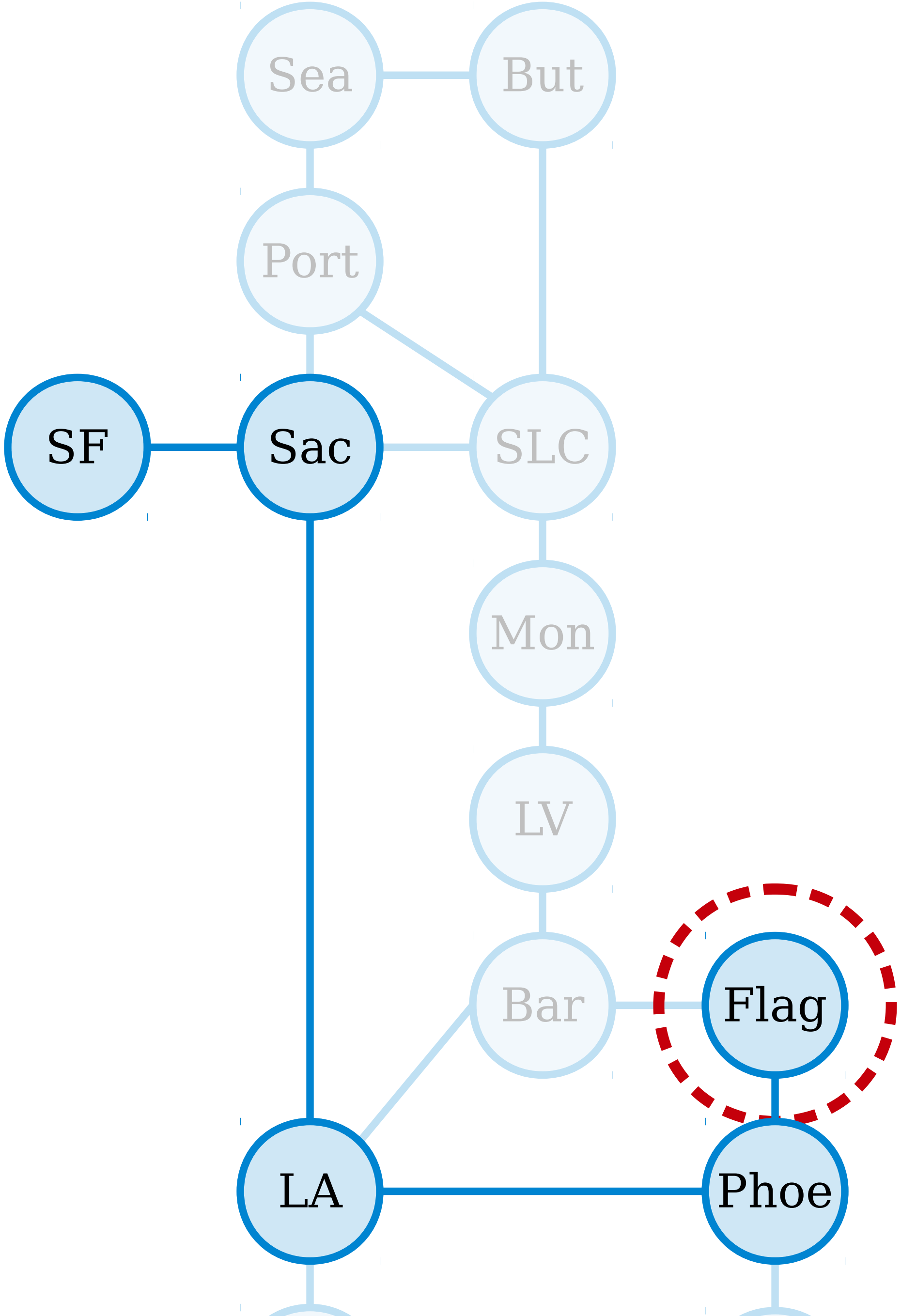


A ***path*** in a graph $G = (V, E)$ is a sequence of one or more nodes $v_1, v_2, v_3, \ldots, v_n$ such that any two consecutive nodes in the sequence are adjacent.

The ***length*** of the path $v_1, \ldots, v_n$ is $n - 1$.

A ***cycle*** in a graph is a path from a node back to itself. (By convention, a cycle cannot have length zero.)

SF, Sac, LA, Phoe

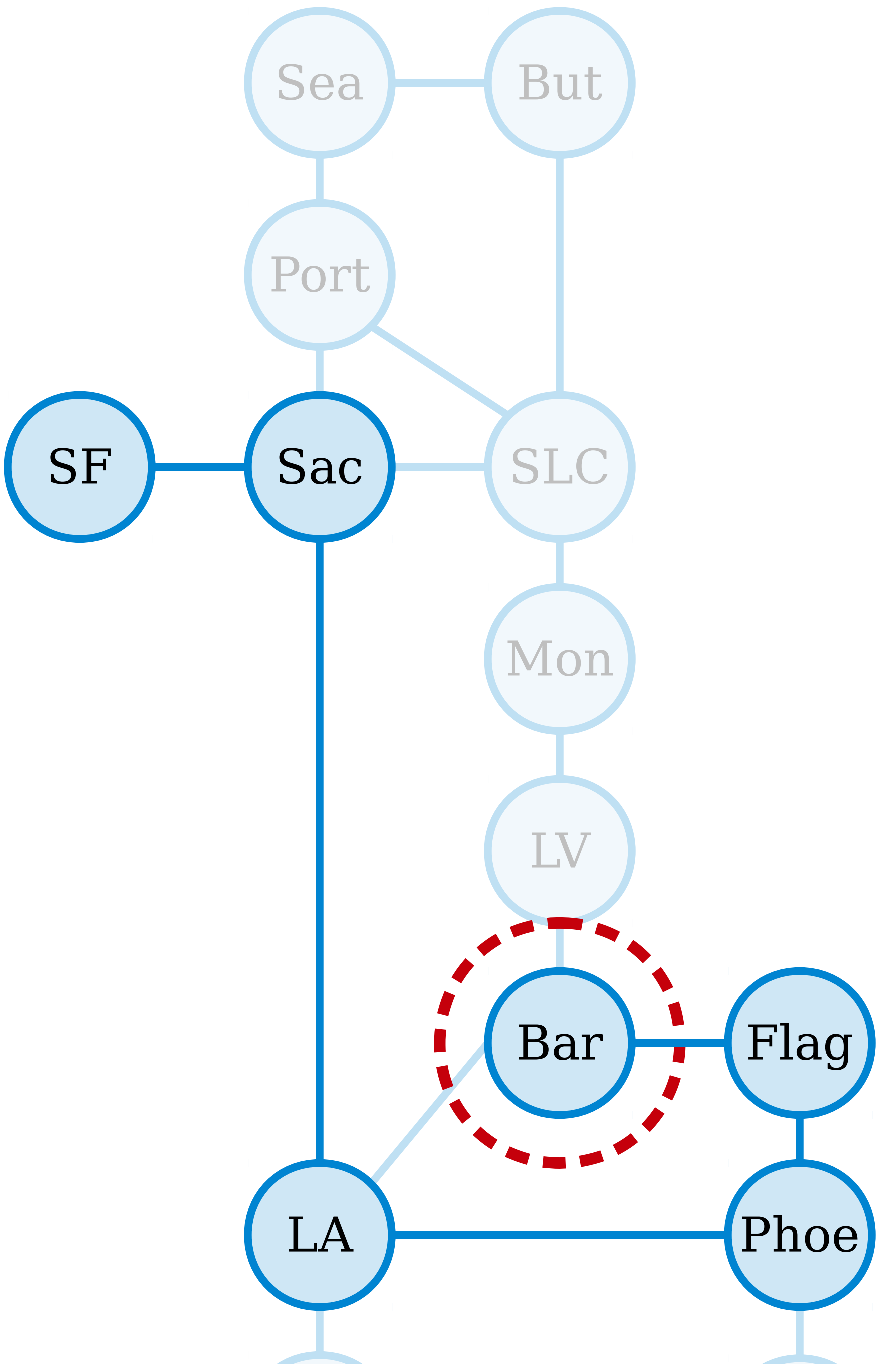


A ***path*** in a graph $G = (V, E)$ is a sequence of one or more nodes $v_1, v_2, v_3, \ldots, v_n$ such that any two consecutive nodes in the sequence are adjacent.

The ***length*** of the path $v_1, \ldots, v_n$ is $n - 1$.

A ***cycle*** in a graph is a path from a node back to itself. (By convention, a cycle cannot have length zero.)

SF, Sac, LA, Phoe, Flag

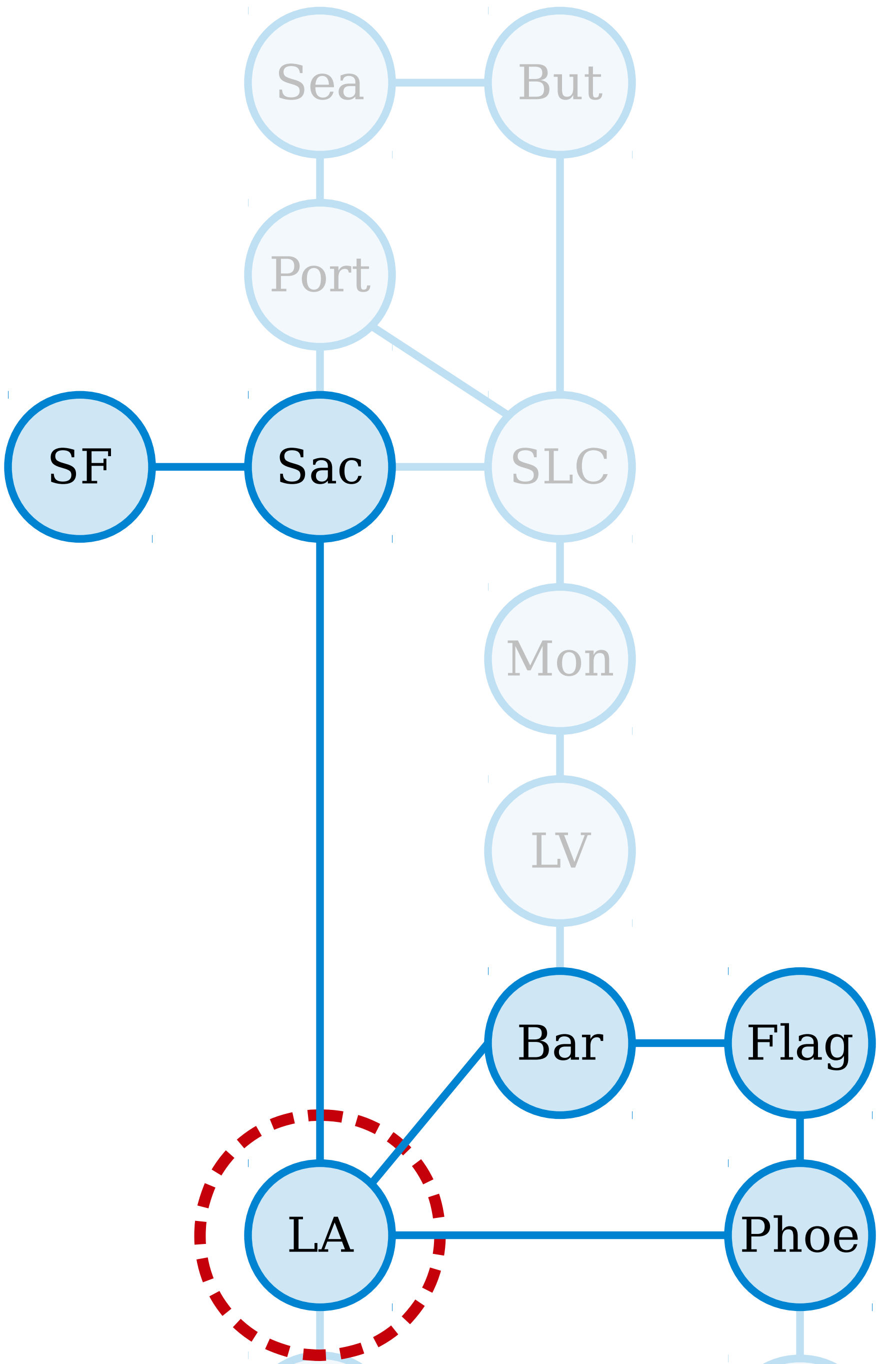


A ***path*** in a graph $G = (V, E)$ is a sequence of one or more nodes $v_1, v_2, v_3, \ldots, v_n$ such that any two consecutive nodes in the sequence are adjacent.

The ***length*** of the path $v_1, \ldots, v_n$ is $n - 1$.

A ***cycle*** in a graph is a path from a node back to itself. (By convention, a cycle cannot have length zero.)

SF, Sac, LA, Phoe, Flag, Bar

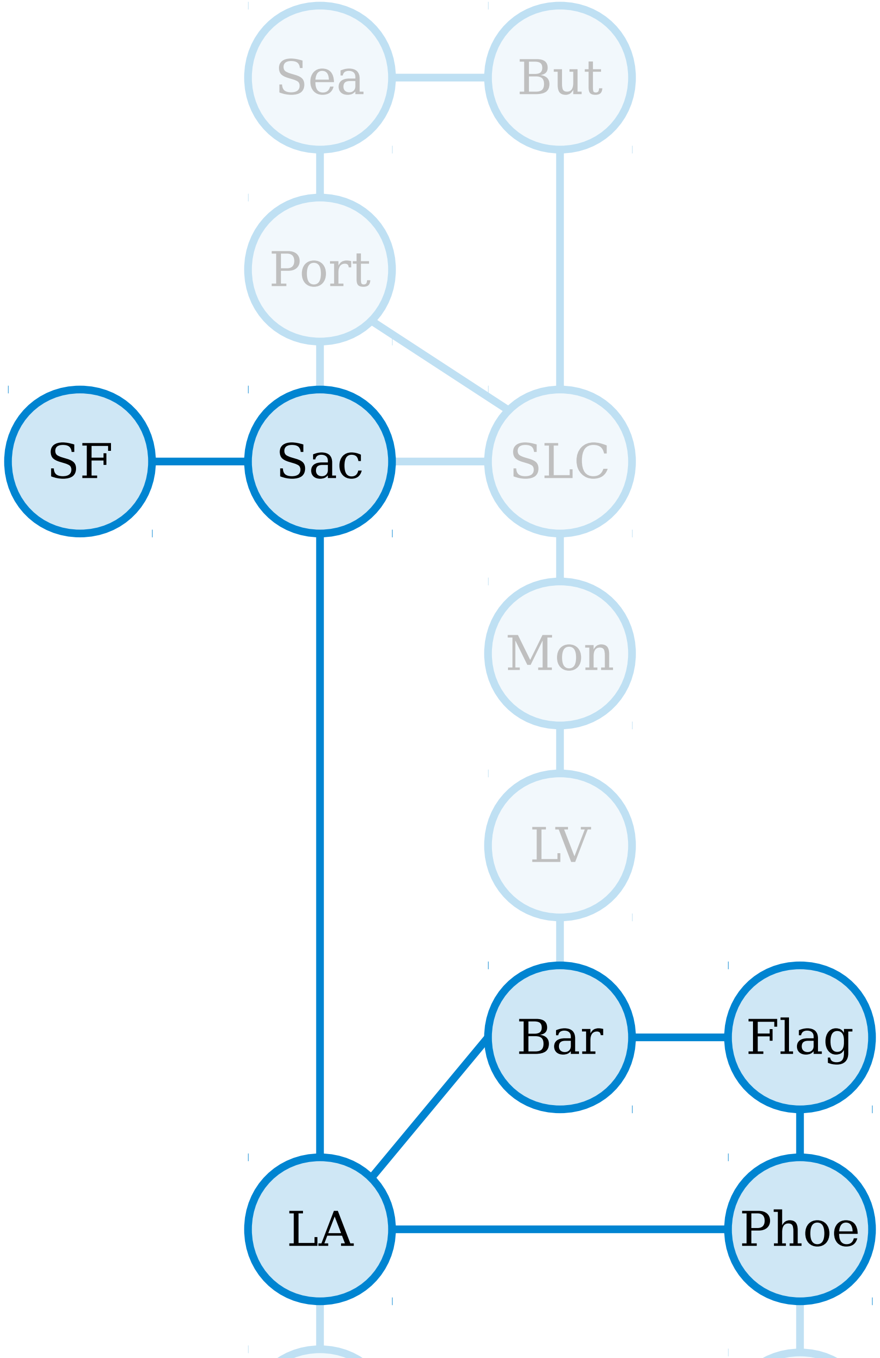


A ***path*** in a graph $G = (V, E)$ is a sequence of one or more nodes $v_1, v_2, v_3, \ldots, v_n$ such that any two consecutive nodes in the sequence are adjacent.

The ***length*** of the path $v_1, \ldots, v_n$ is $n - 1$.

A ***cycle*** in a graph is a path from a node back to itself. (By convention, a cycle cannot have length zero.)

SF, Sac, LA, Phoe, Flag, Bar, LA

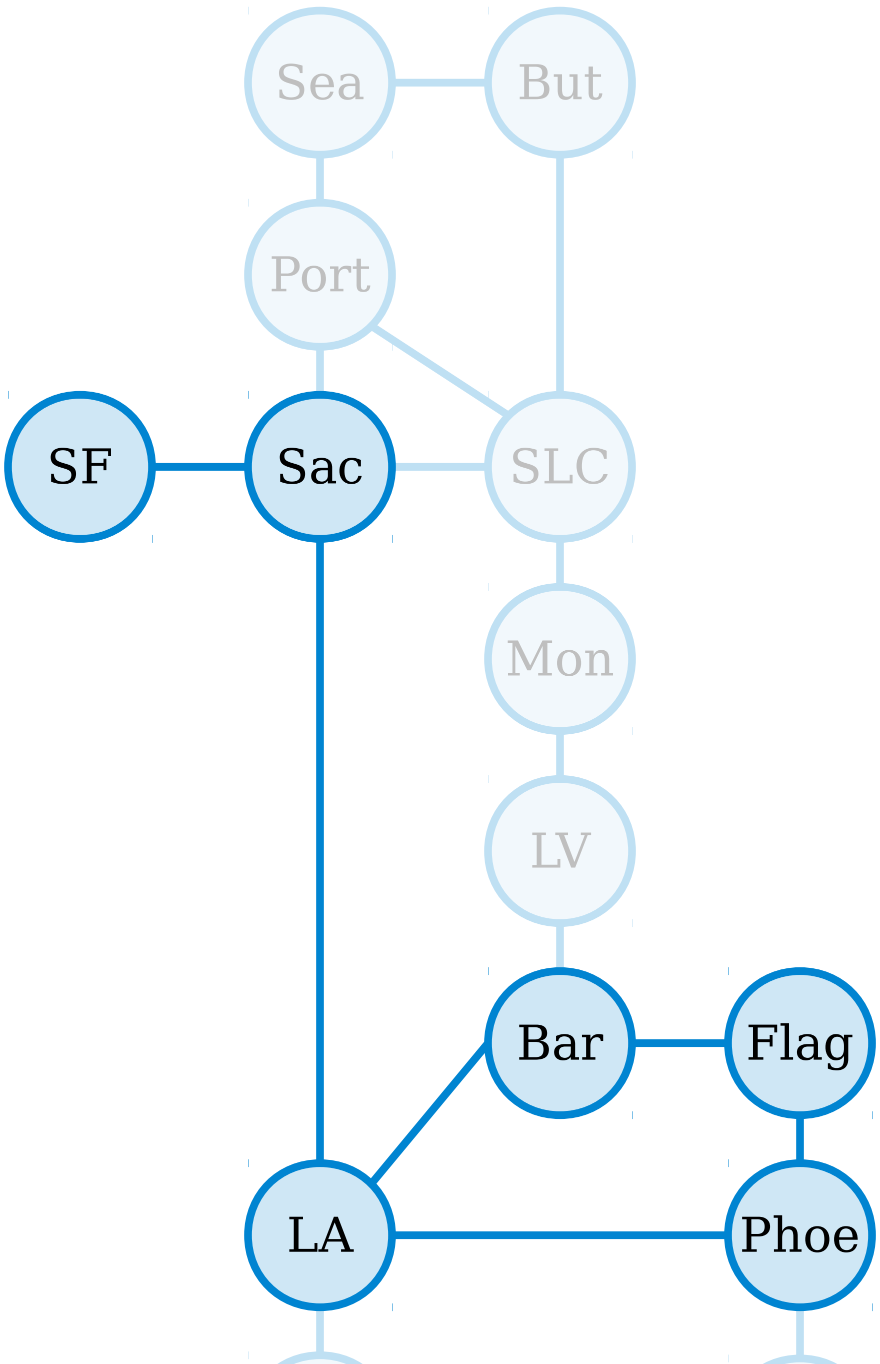


A ***path*** in a graph $G = (V, E)$ is a sequence of one or more nodes $v_1, v_2, v_3, \ldots, v_n$ such that any two consecutive nodes in the sequence are adjacent.

The ***length*** of the path $v_1, \ldots, v_n$ is $n - 1$.

A ***cycle*** in a graph is a path from a node back to itself. (By convention, a cycle cannot have length zero.)

SF, Sac, LA, Phoe, Flag, Bar, LA



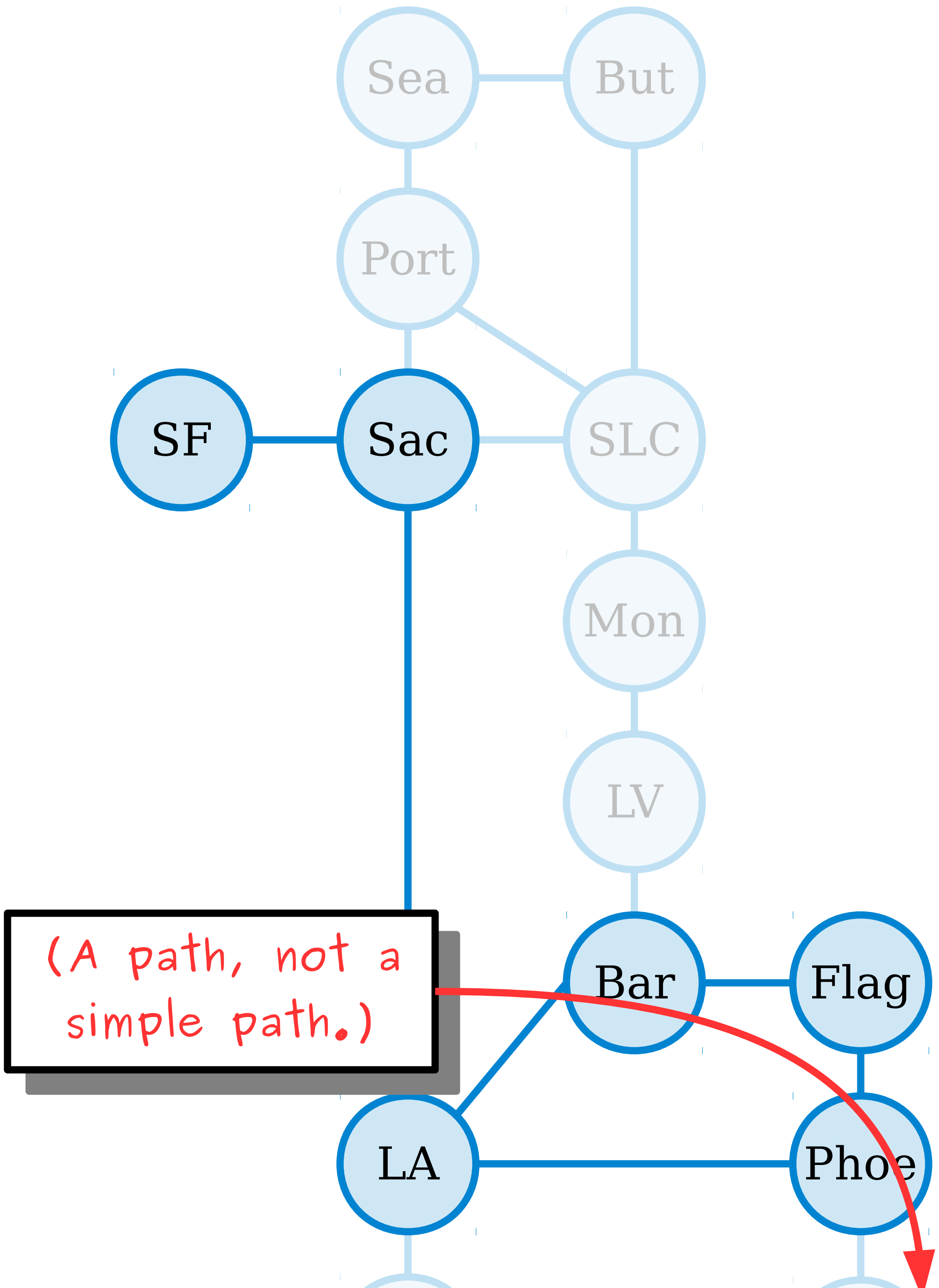
A ***path*** in a graph $G = (V, E)$ is a sequence of one or more nodes $v_1, v_2, v_3, \ldots, v_n$ such that any two consecutive nodes in the sequence are adjacent.

The ***length*** of the path $v_1, \ldots, v_n$ is $n - 1$.

A ***cycle*** in a graph is a path from a node back to itself. (By convention, a cycle cannot have length zero.)

A ***simple path*** in a graph is path that does not repeat any nodes or edges.

SF, Sac, LA, Phoe, Flag, Bar, LA



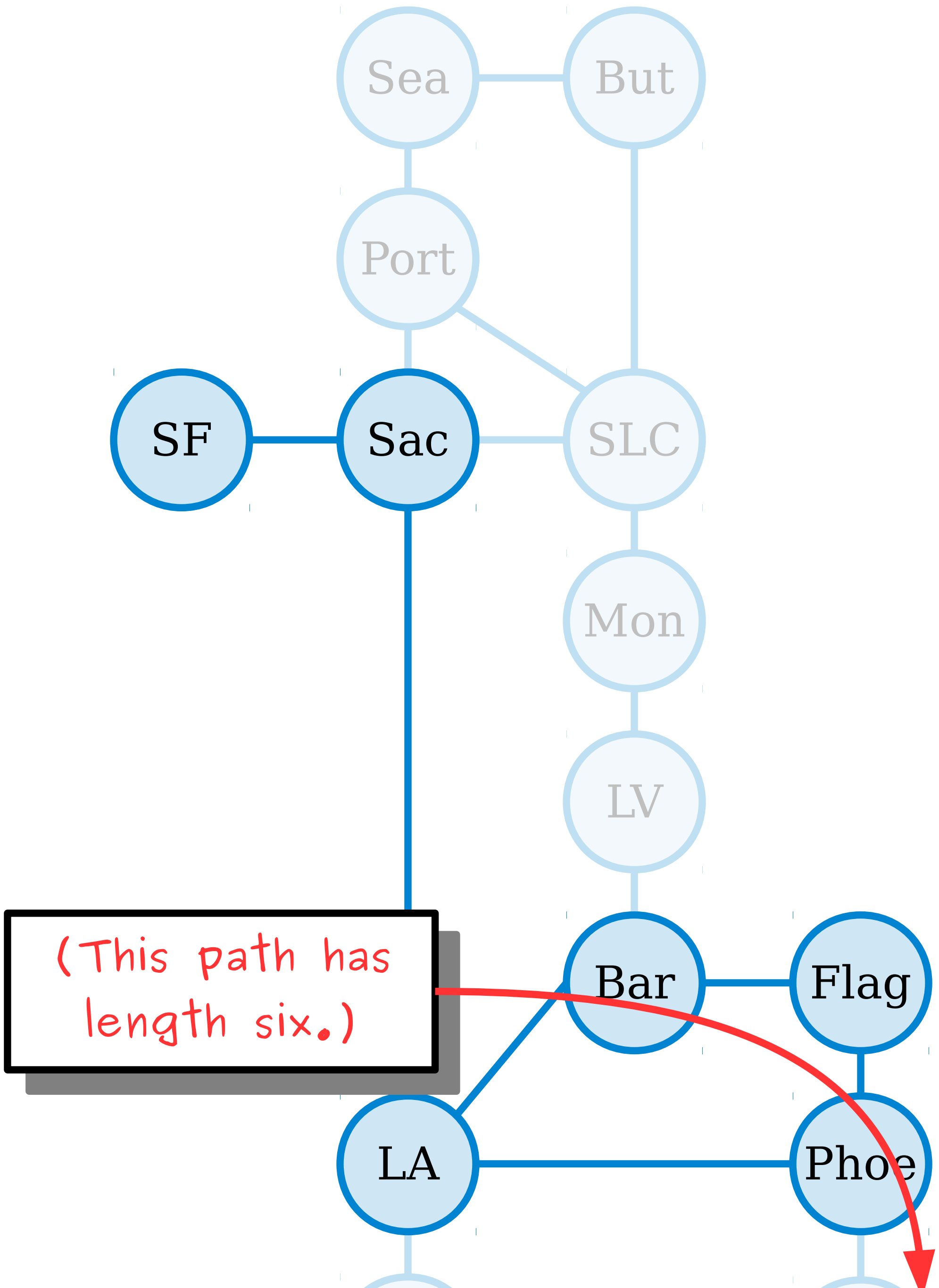
A ***path*** in a graph $G = (V, E)$ is a sequence of one or more nodes $v_1, v_2, v_3, \ldots, v_n$ such that any two consecutive nodes in the sequence are adjacent.

The ***length*** of the path $v_1, \ldots, v_n$ is $n - 1$.

A ***cycle*** in a graph is a path from a node back to itself. (By convention, a cycle cannot have length zero.)

A ***simple path*** in a graph is path that does not repeat any nodes or edges.

SF, Sac, LA, Phoe, Flag, Bar, LA



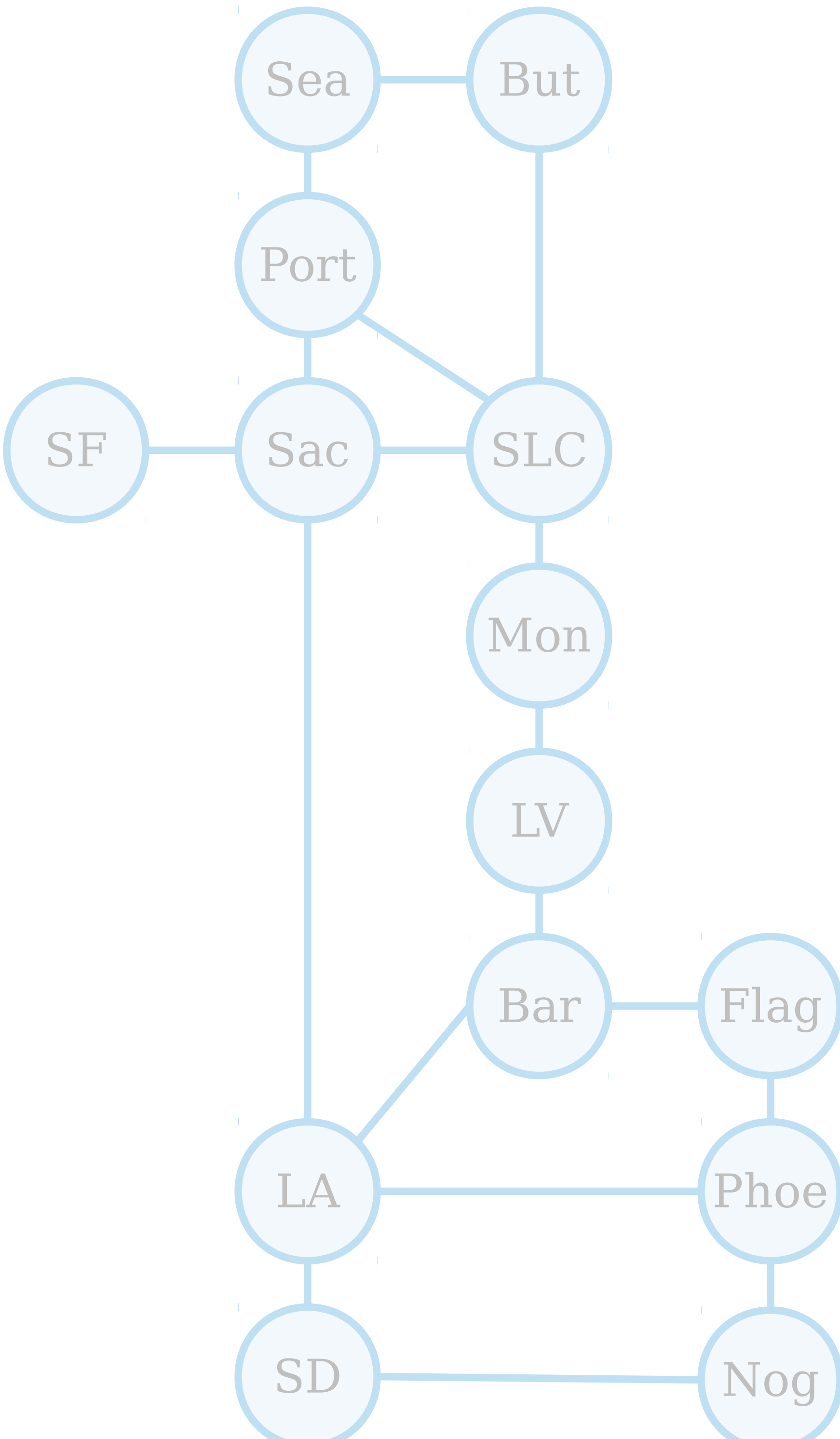
SF, Sac, LA, Phoe, Flag, Bar, LA

A ***path*** in a graph $G = (V, E)$ is a sequence of one or more nodes $v_1, v_2, v_3, \ldots, v_n$ such that any two consecutive nodes in the sequence are adjacent.

The ***length*** of the path $v_1, \ldots, v_n$ is $n - 1$.

A ***cycle*** in a graph is a path from a node back to itself. (By convention, a cycle cannot have length zero.)

A ***simple path*** in a graph is path that does not repeat any nodes or edges.

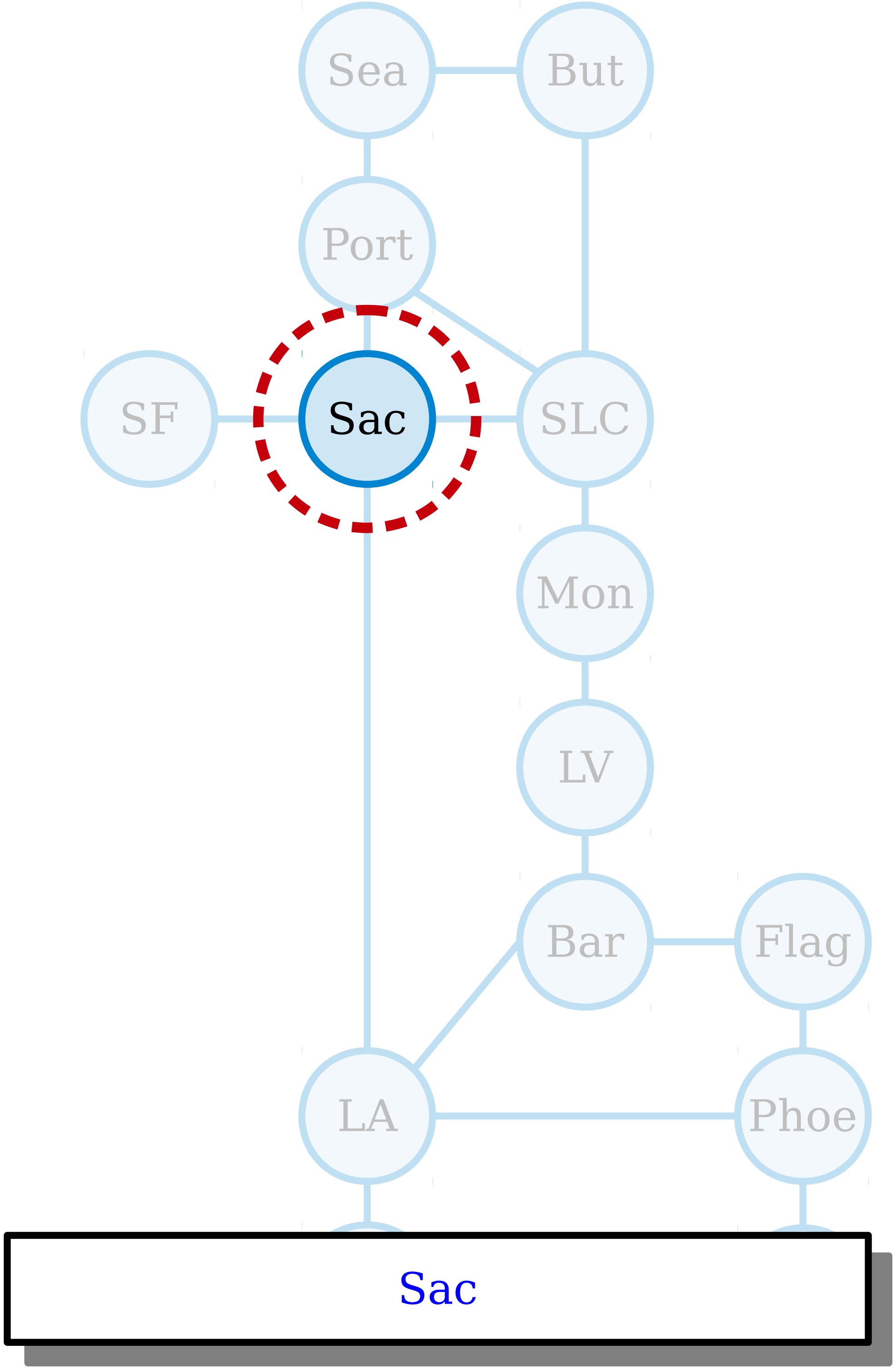


A ***path*** in a graph $G = (V, E)$ is a sequence of one or more nodes $v_1, v_2, v_3, \ldots, v_n$ such that any two consecutive nodes in the sequence are adjacent.

The ***length*** of the path $v_1, \ldots, v_n$ is $n - 1$.

A ***cycle*** in a graph is a path from a node back to itself. (By convention, a cycle cannot have length zero.)

A ***simple path*** in a graph is path that does not repeat any nodes or edges.

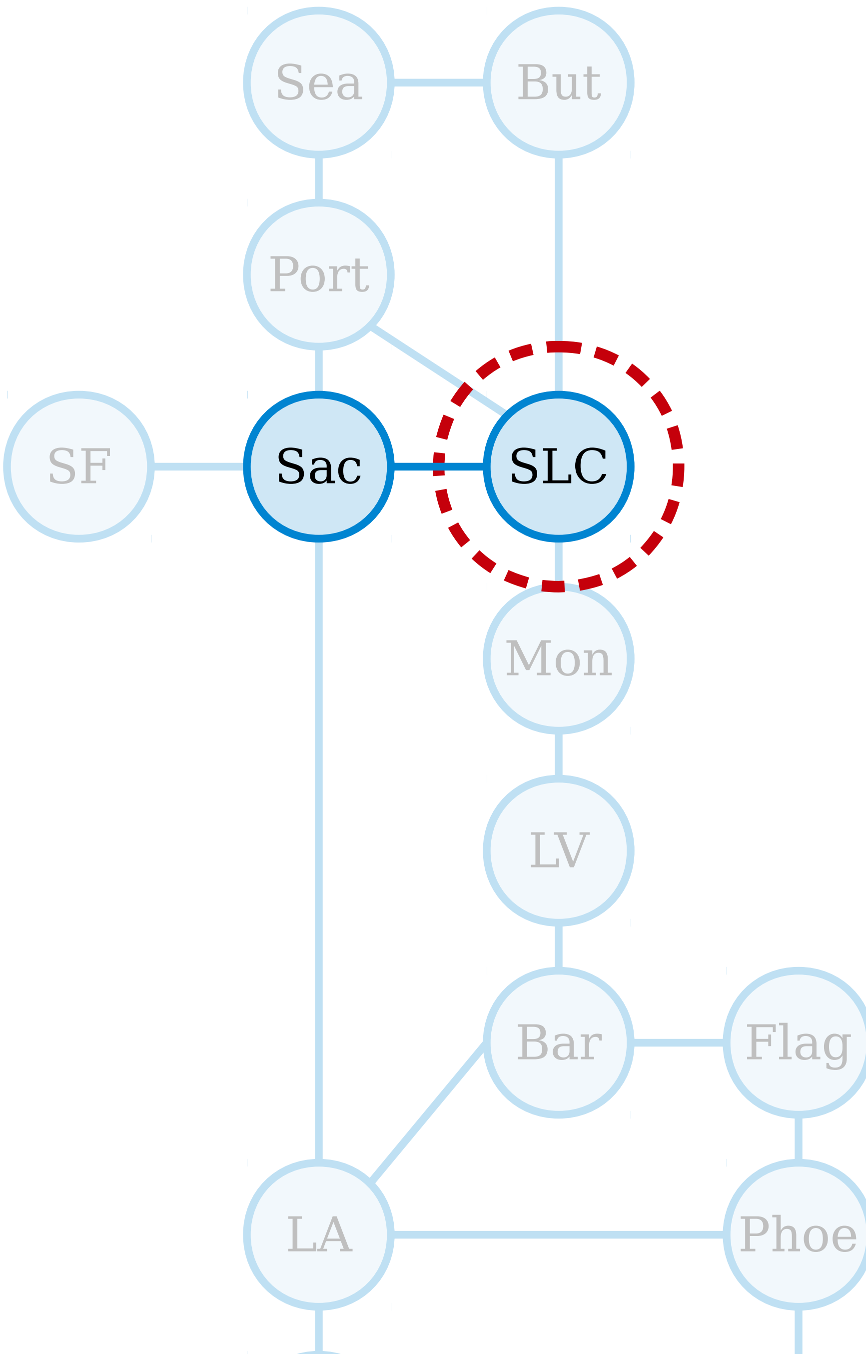


A ***path*** in a graph $G = (V, E)$ is a sequence of one or more nodes $v_1, v_2, v_3, \ldots, v_n$ such that any two consecutive nodes in the sequence are adjacent.

The ***length*** of the path $v_1, \ldots, v_n$ is $n - 1$.

A ***cycle*** in a graph is a path from a node back to itself. (By convention, a cycle cannot have length zero.)

A ***simple path*** in a graph is path that does not repeat any nodes or edges.



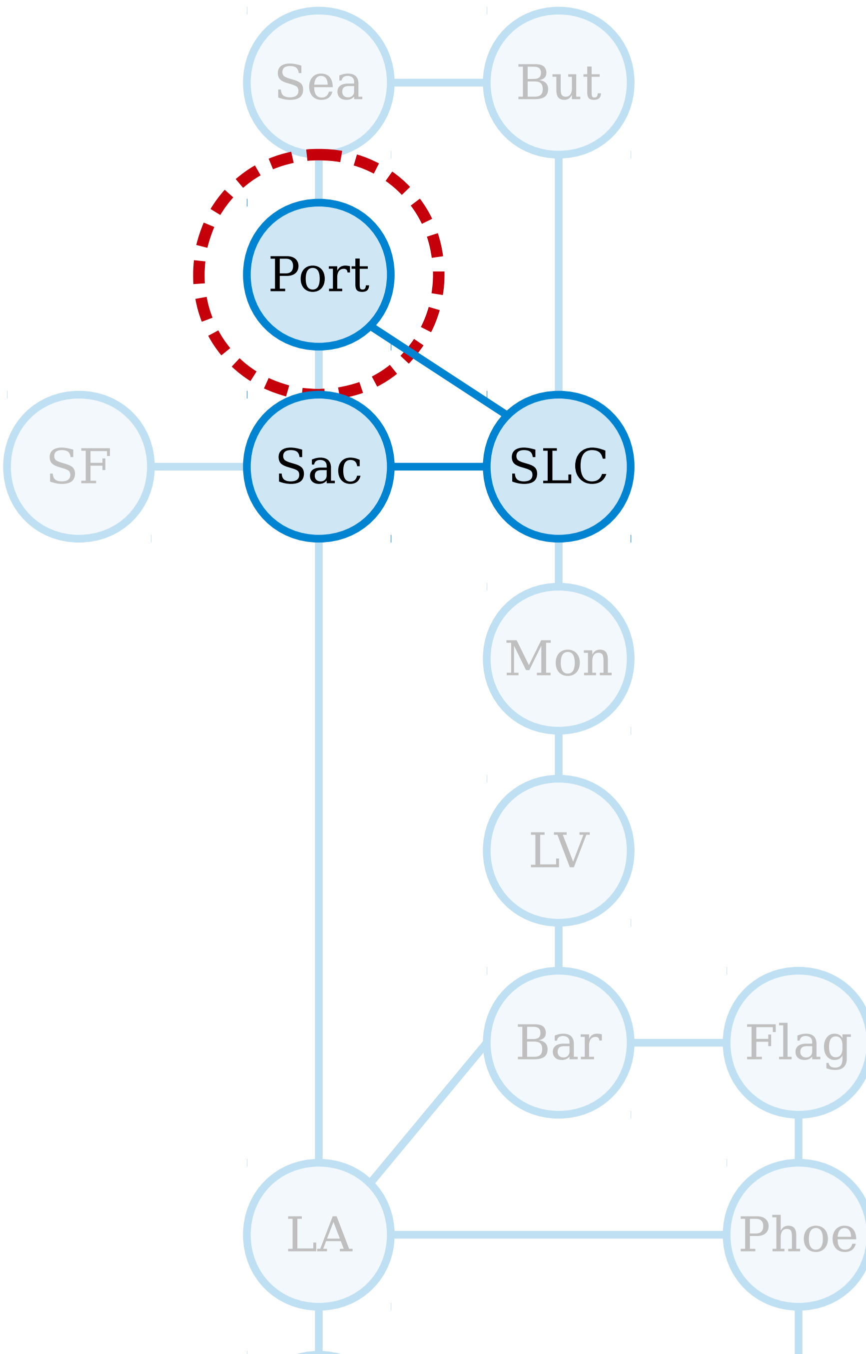
A ***path*** in a graph $G = (V, E)$ is a sequence of one or more nodes $v_1, v_2, v_3, \ldots, v_n$ such that any two consecutive nodes in the sequence are adjacent.

The ***length*** of the path $v_1, \ldots, v_n$ is $n - 1$.

A ***cycle*** in a graph is a path from a node back to itself. (By convention, a cycle cannot have length zero.)

A ***simple path*** in a graph is path that does not repeat any nodes or edges.

Sac, SLC



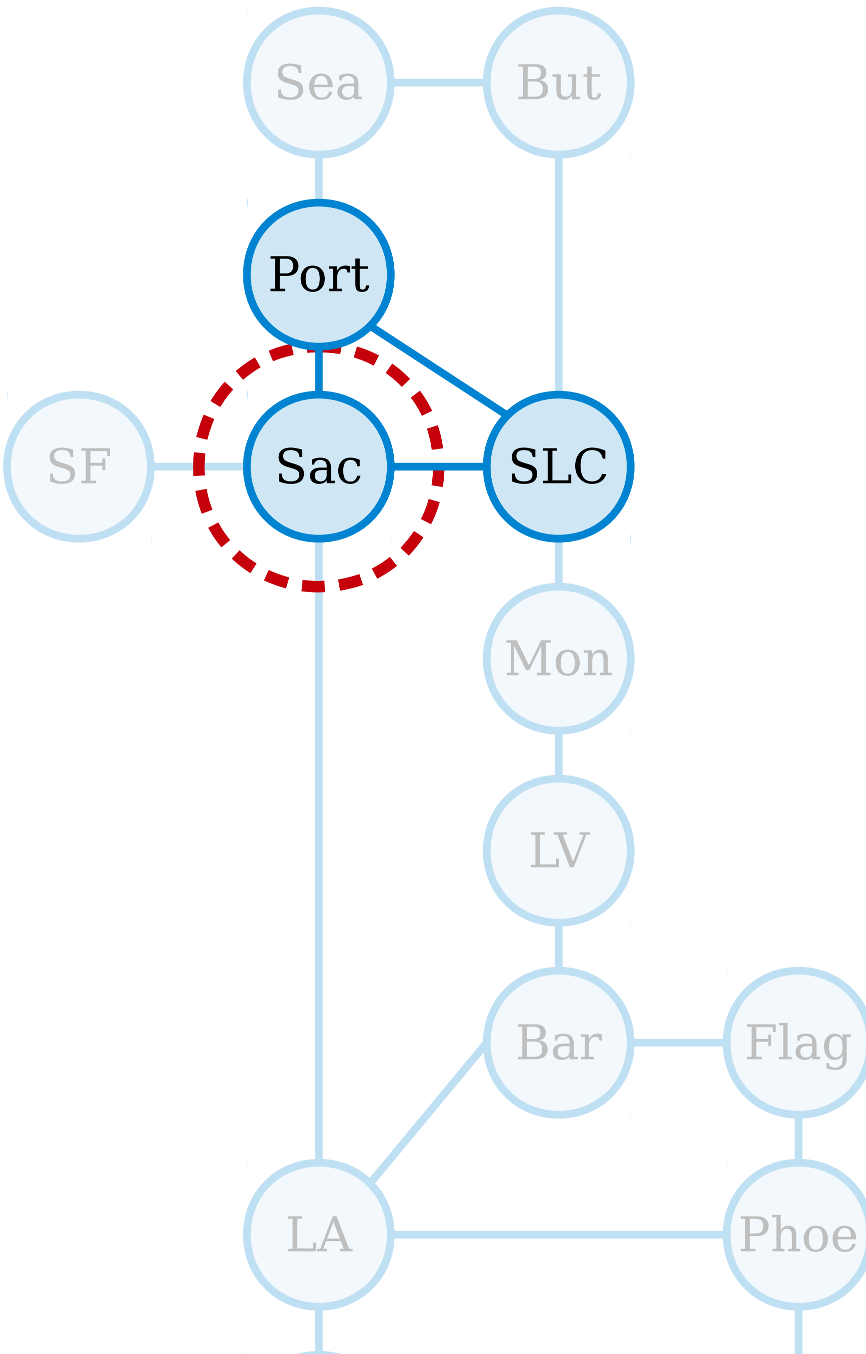
Sac, SLC, Port

A ***path*** in a graph $G = (V, E)$ is a sequence of one or more nodes $v_1, v_2, v_3, \ldots, v_n$ such that any two consecutive nodes in the sequence are adjacent.

The ***length*** of the path $v_1, \ldots, v_n$ is $n - 1$.

A ***cycle*** in a graph is a path from a node back to itself. (By convention, a cycle cannot have length zero.)

A ***simple path*** in a graph is path that does not repeat any nodes or edges.



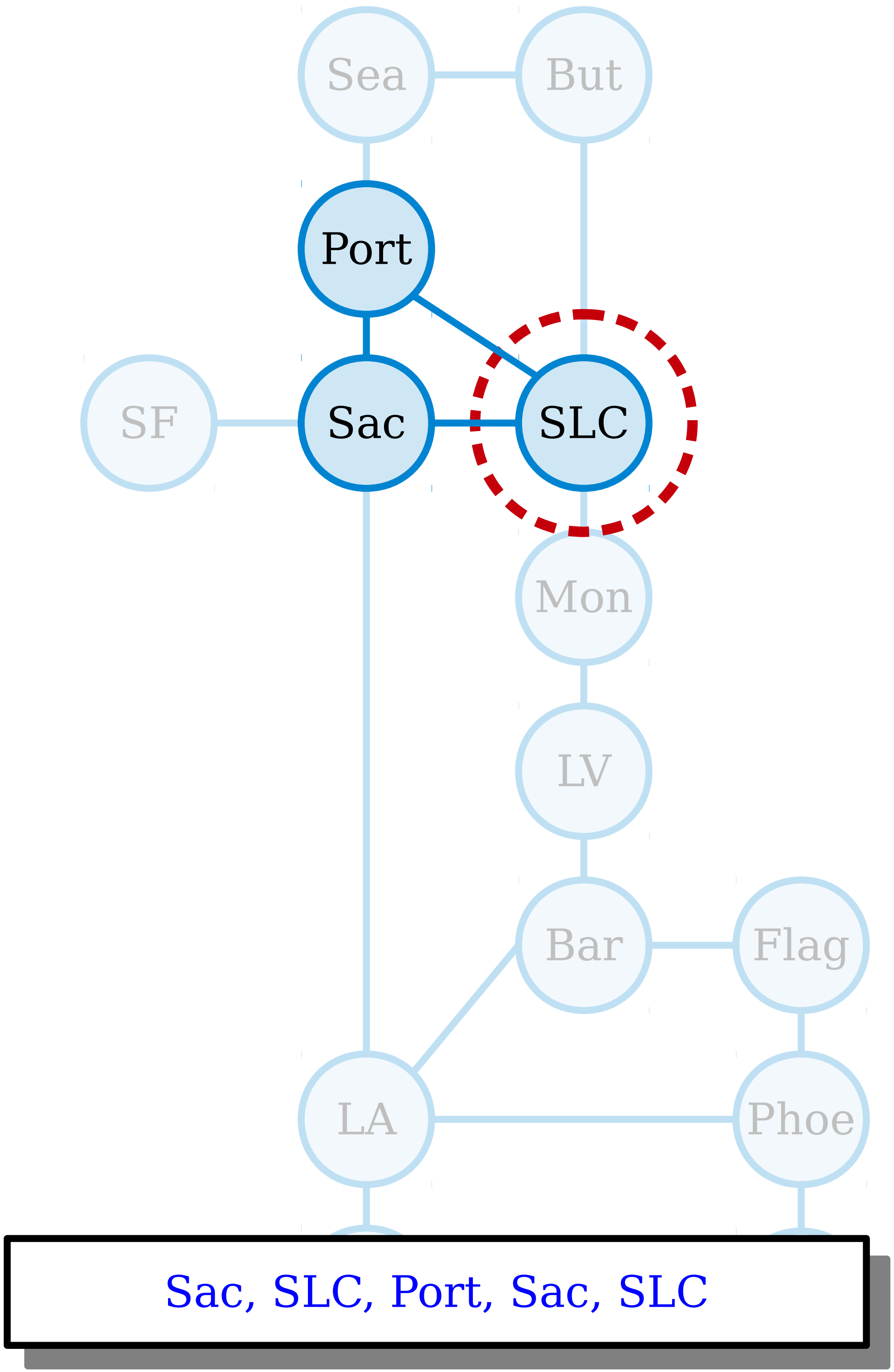
Sac, SLC, Port, Sac

A ***path*** in a graph $G = (V, E)$ is a sequence of one or more nodes $v_1, v_2, v_3, \ldots, v_n$ such that any two consecutive nodes in the sequence are adjacent.

The ***length*** of the path $v_1, \ldots, v_n$ is $n - 1$.

A ***cycle*** in a graph is a path from a node back to itself. (By convention, a cycle cannot have length zero.)

A ***simple path*** in a graph is path that does not repeat any nodes or edges.

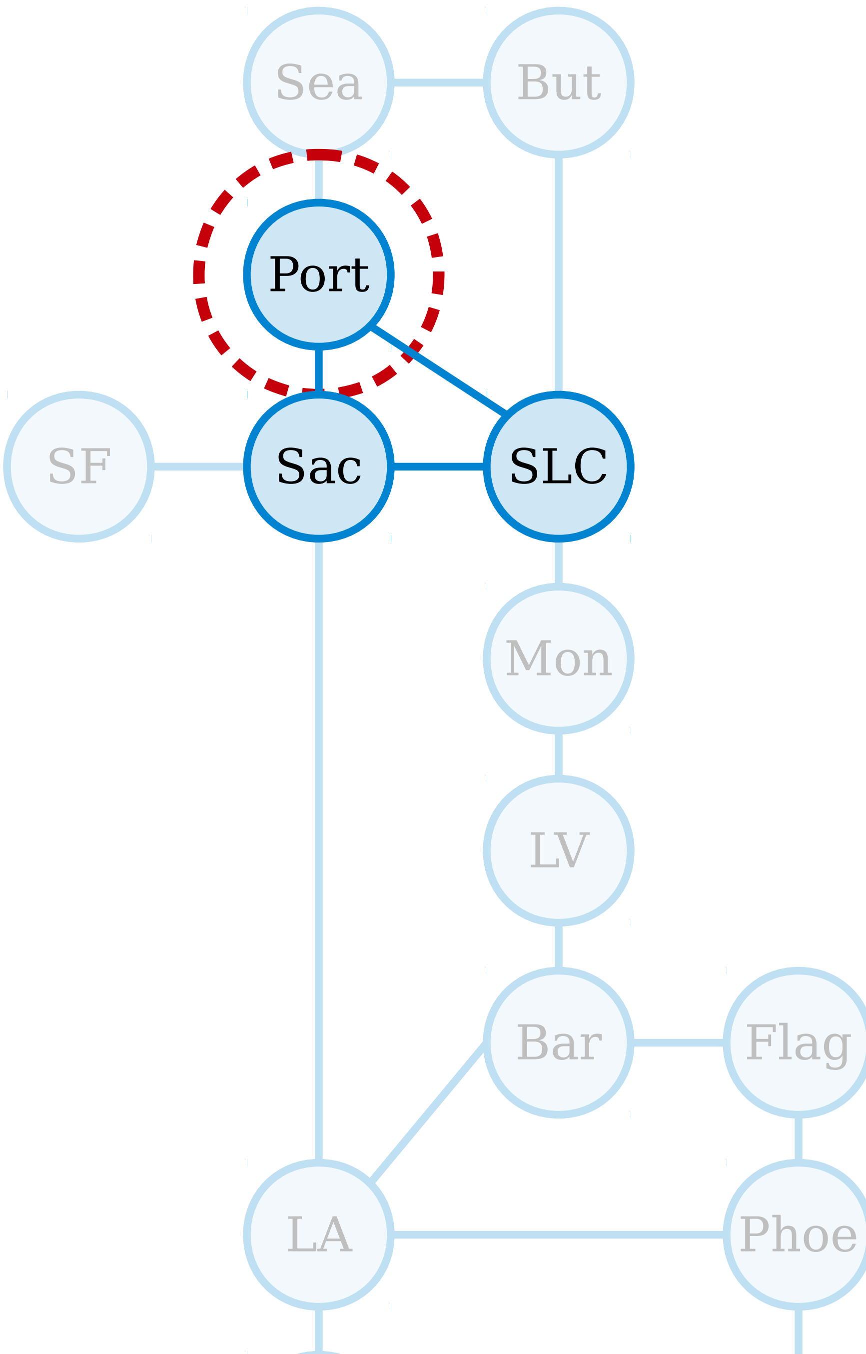


A ***path*** in a graph $G = (V, E)$ is a sequence of one or more nodes $v_1, v_2, v_3, \ldots, v_n$ such that any two consecutive nodes in the sequence are adjacent.

The ***length*** of the path $v_1, \ldots, v_n$ is $n - 1$.

A ***cycle*** in a graph is a path from a node back to itself. (By convention, a cycle cannot have length zero.)

A ***simple path*** in a graph is path that does not repeat any nodes or edges.



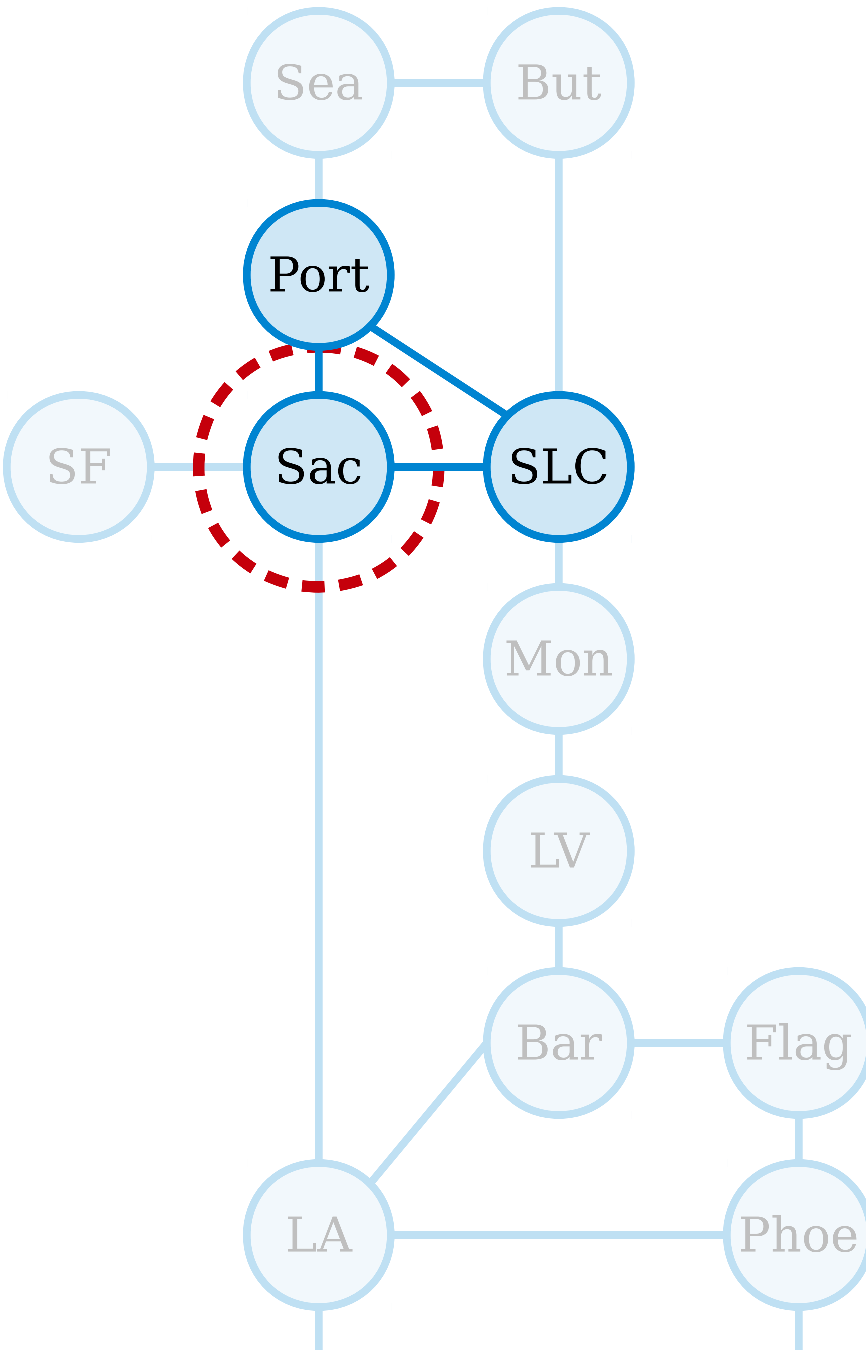
Sac, SLC, Port, Sac, SLC, Port

A ***path*** in a graph $G = (V, E)$ is a sequence of one or more nodes $v_1, v_2, v_3, \ldots, v_n$ such that any two consecutive nodes in the sequence are adjacent.

The ***length*** of the path $v_1, \ldots, v_n$ is $n - 1$.

A ***cycle*** in a graph is a path from a node back to itself. (By convention, a cycle cannot have length zero.)

A ***simple path*** in a graph is path that does not repeat any nodes or edges.



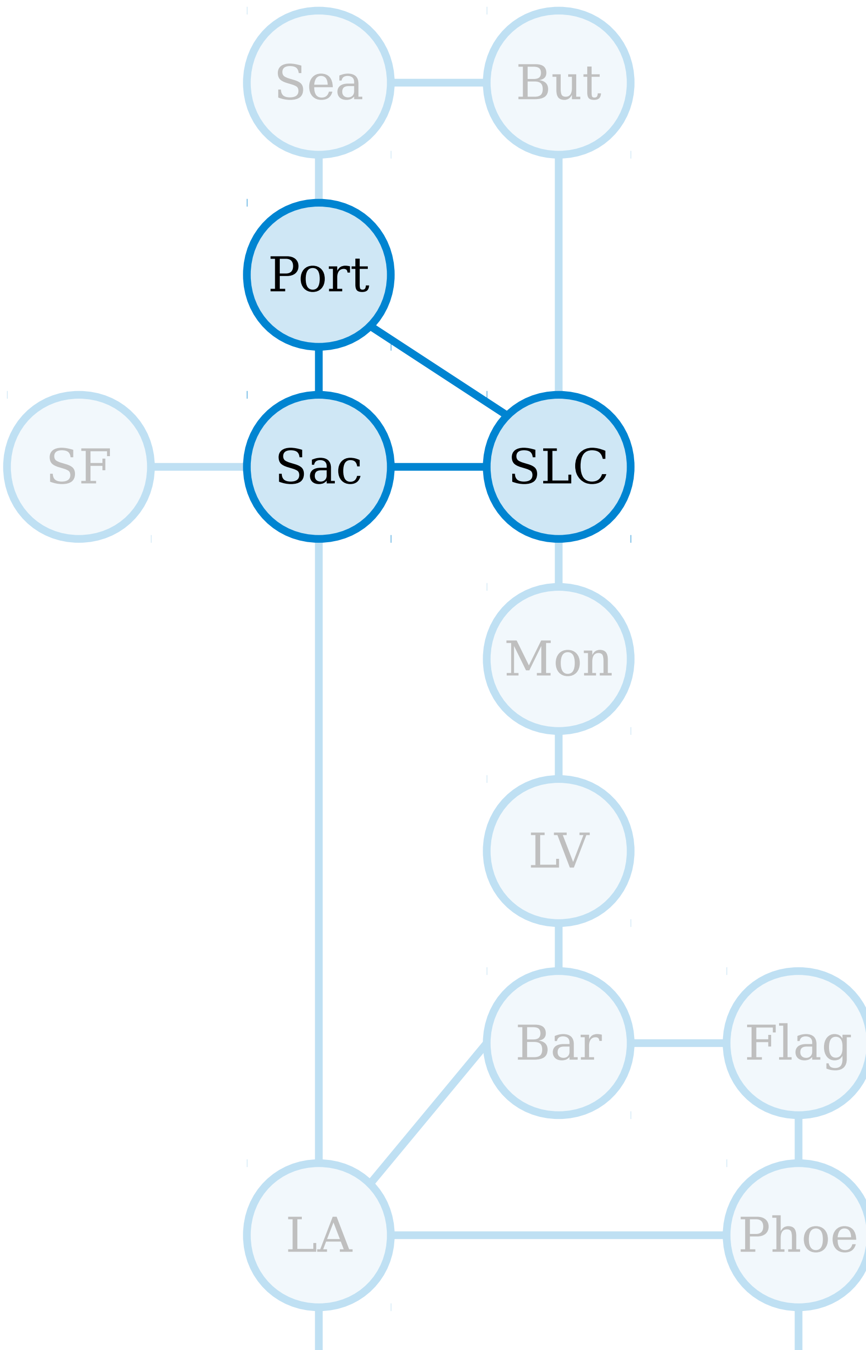
Sac, SLC, Port, Sac, SLC, Port, Sac

A ***path*** in a graph $G = (V, E)$ is a sequence of one or more nodes $v_1, v_2, v_3, \ldots, v_n$ such that any two consecutive nodes in the sequence are adjacent.

The ***length*** of the path $v_1, \ldots, v_n$ is $n - 1$.

A ***cycle*** in a graph is a path from a node back to itself. (By convention, a cycle cannot have length zero.)

A ***simple path*** in a graph is path that does not repeat any nodes or edges.



Sac, SLC, Port, Sac, SLC, Port, Sac

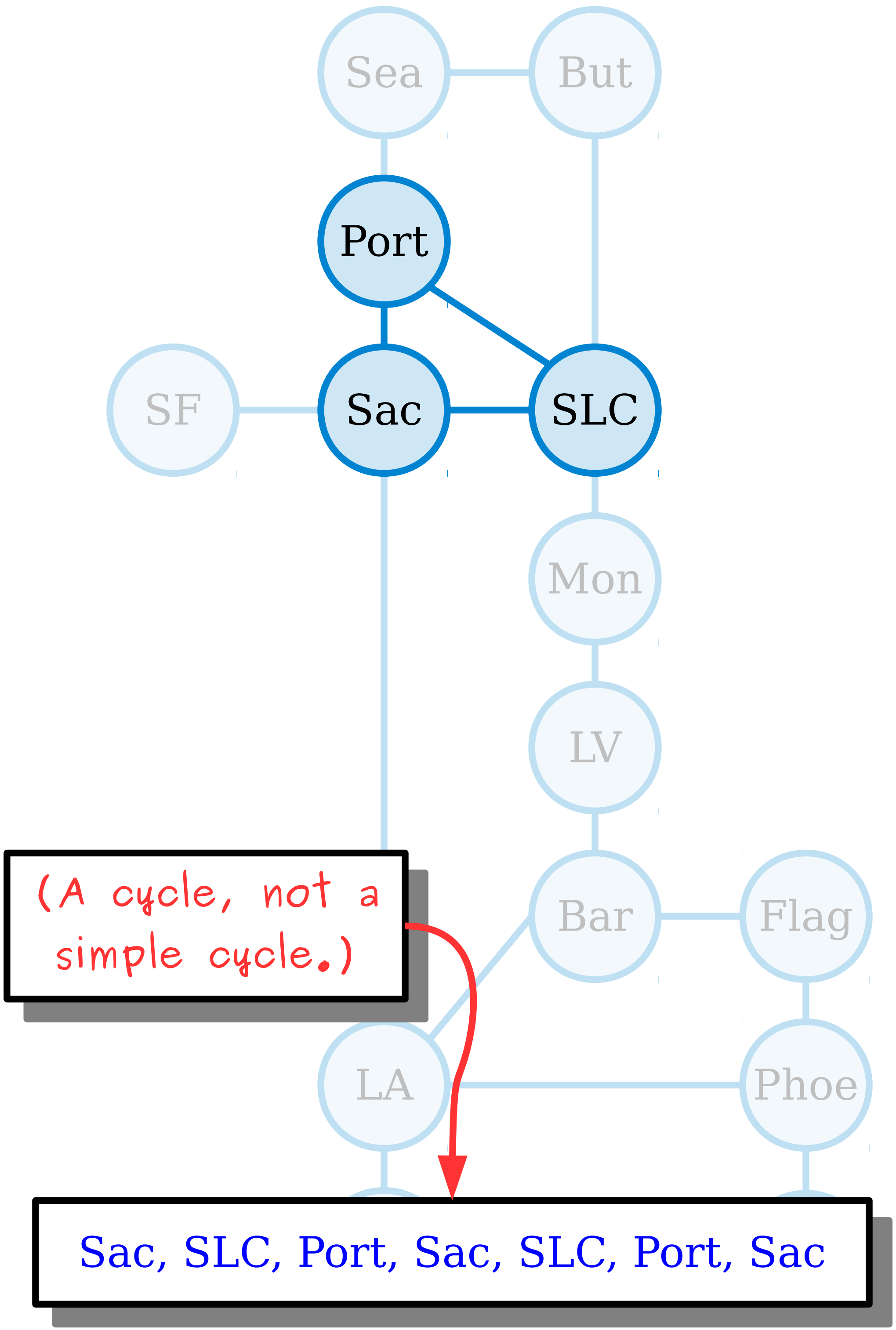
A ***path*** in a graph $G = (V, E)$ is a sequence of one or more nodes $v_1, v_2, v_3, \ldots, v_n$ such that any two consecutive nodes in the sequence are adjacent.

The ***length*** of the path $v_1, \ldots, v_n$ is $n - 1$.

A ***cycle*** in a graph is a path from a node back to itself. (By convention, a cycle cannot have length zero.)

A ***simple path*** in a graph is path that does not repeat any nodes or edges.

A ***simple cycle*** in a graph is cycle that does not repeat any nodes or edges except the first/last node.



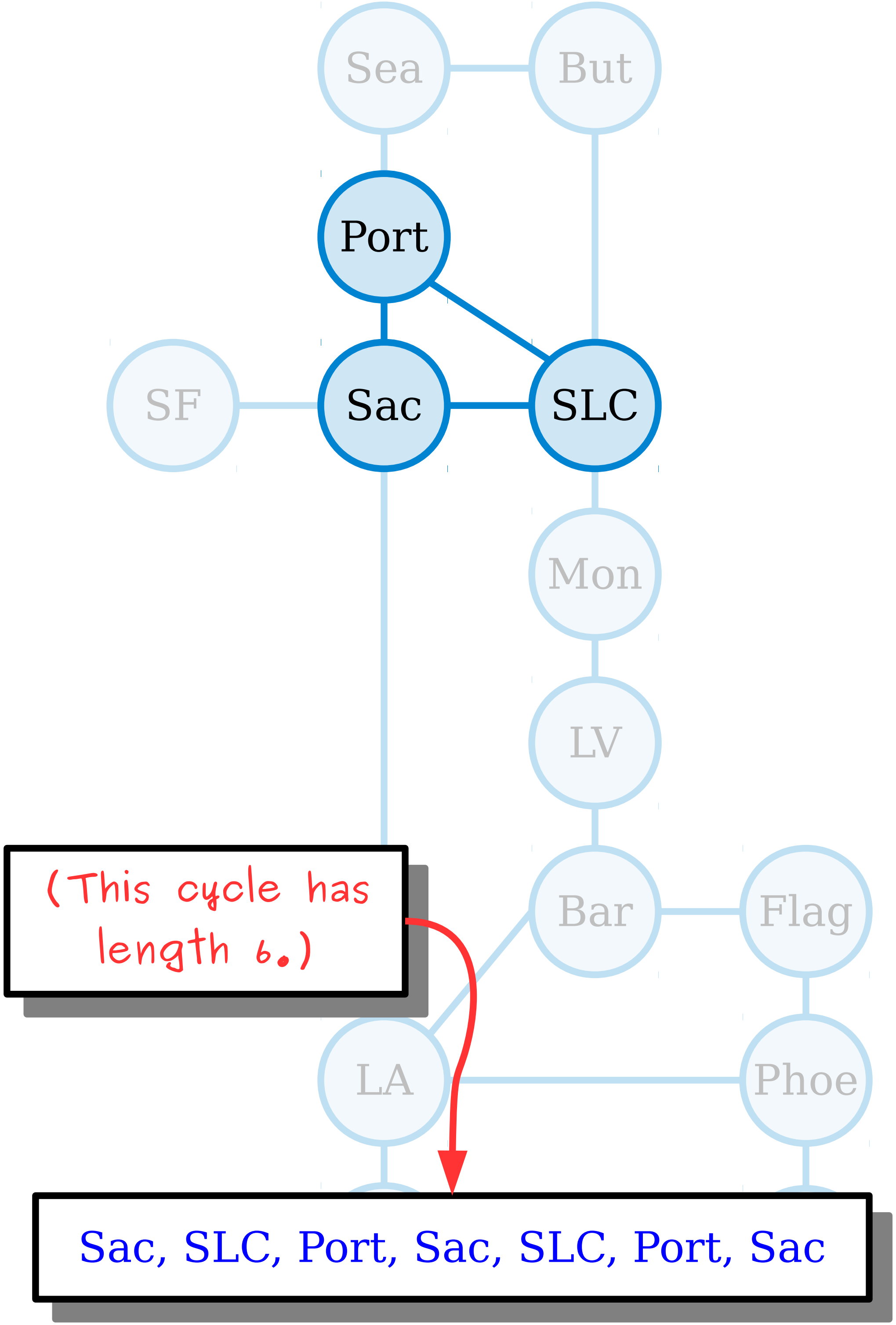
A ***path*** in a graph $G = (V, E)$ is a sequence of one or more nodes $v_1, v_2, v_3, \ldots, v_n$ such that any two consecutive nodes in the sequence are adjacent.

The ***length*** of the path $v_1, \ldots, v_n$ is $n - 1$.

A ***cycle*** in a graph is a path from a node back to itself. (By convention, a cycle cannot have length zero.)

A ***simple path*** in a graph is path that does not repeat any nodes or edges.

A ***simple cycle*** in a graph is cycle that does not repeat any nodes or edges except the first/ last node.



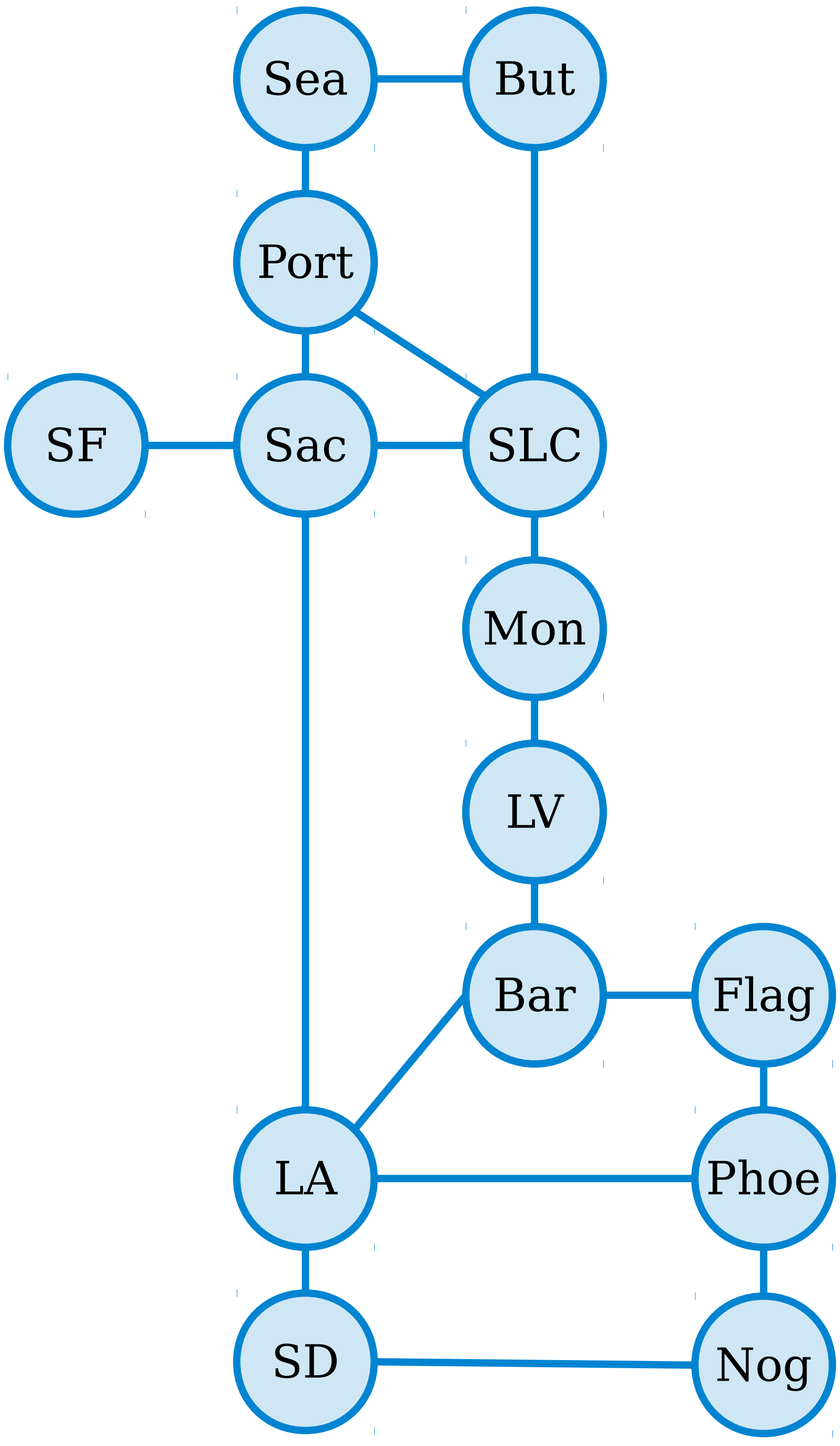
A ***path*** in a graph $G = (V, E)$ is a sequence of one or more nodes $v_1, v_2, v_3, \ldots, v_n$ such that any two consecutive nodes in the sequence are adjacent.

The ***length*** of the path $v_1, \ldots, v_n$ is $n - 1$.

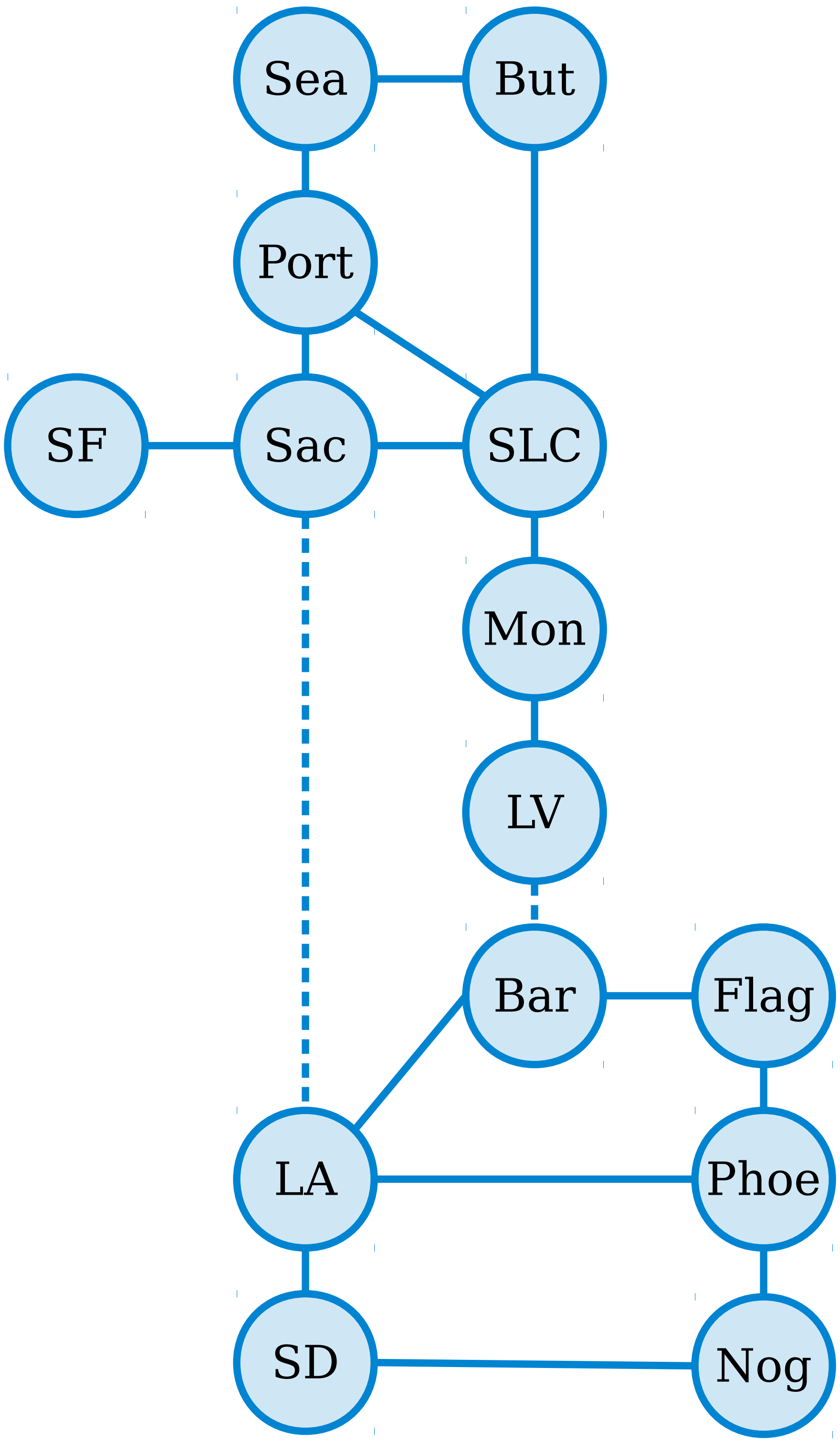
A ***cycle*** in a graph is a path from a node back to itself. (By convention, a cycle cannot have length zero.)

A ***simple path*** in a graph is path that does not repeat any nodes or edges.

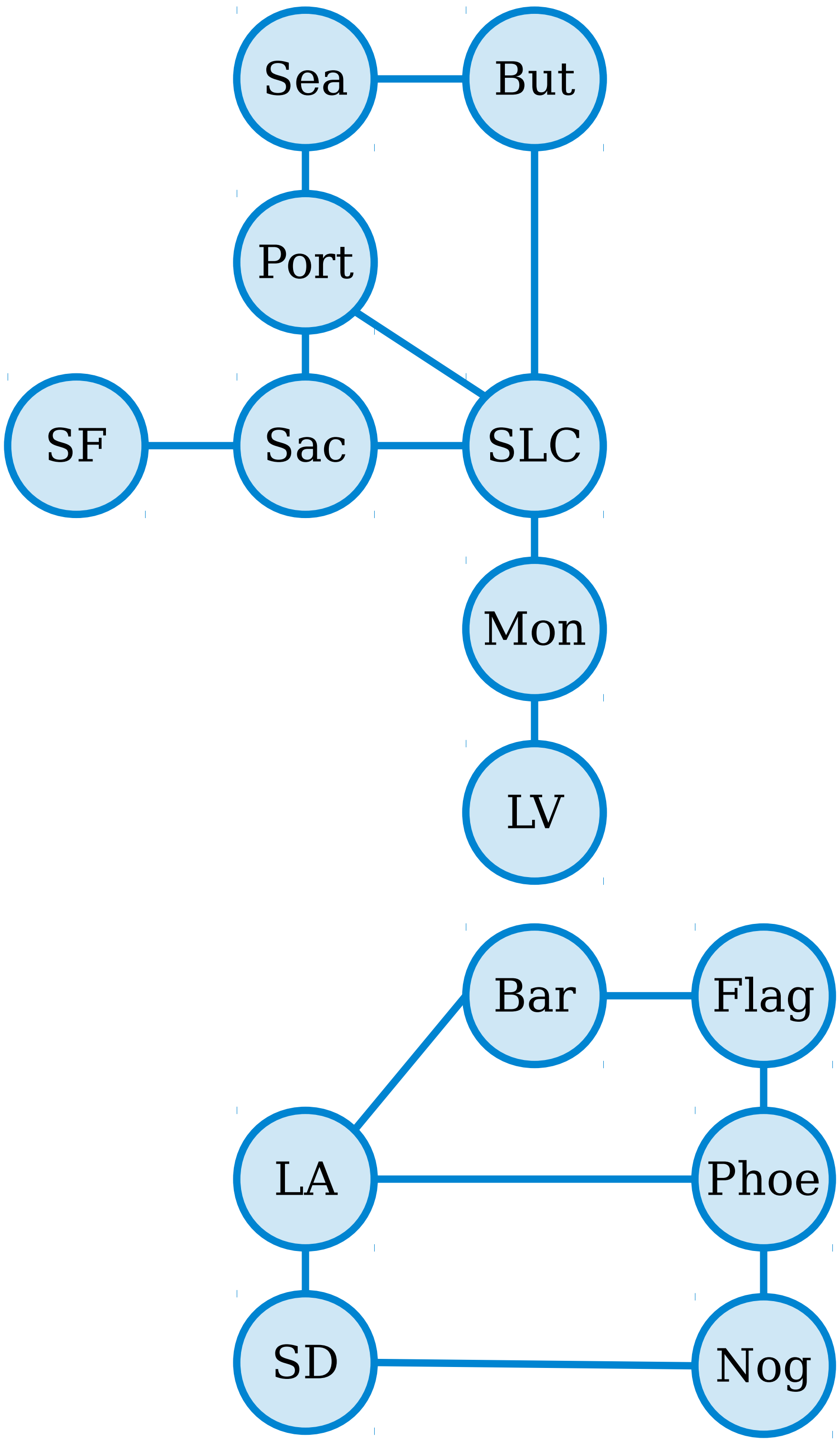
A ***simple cycle*** in a graph is cycle that does not repeat any nodes or edges except the first/last node.



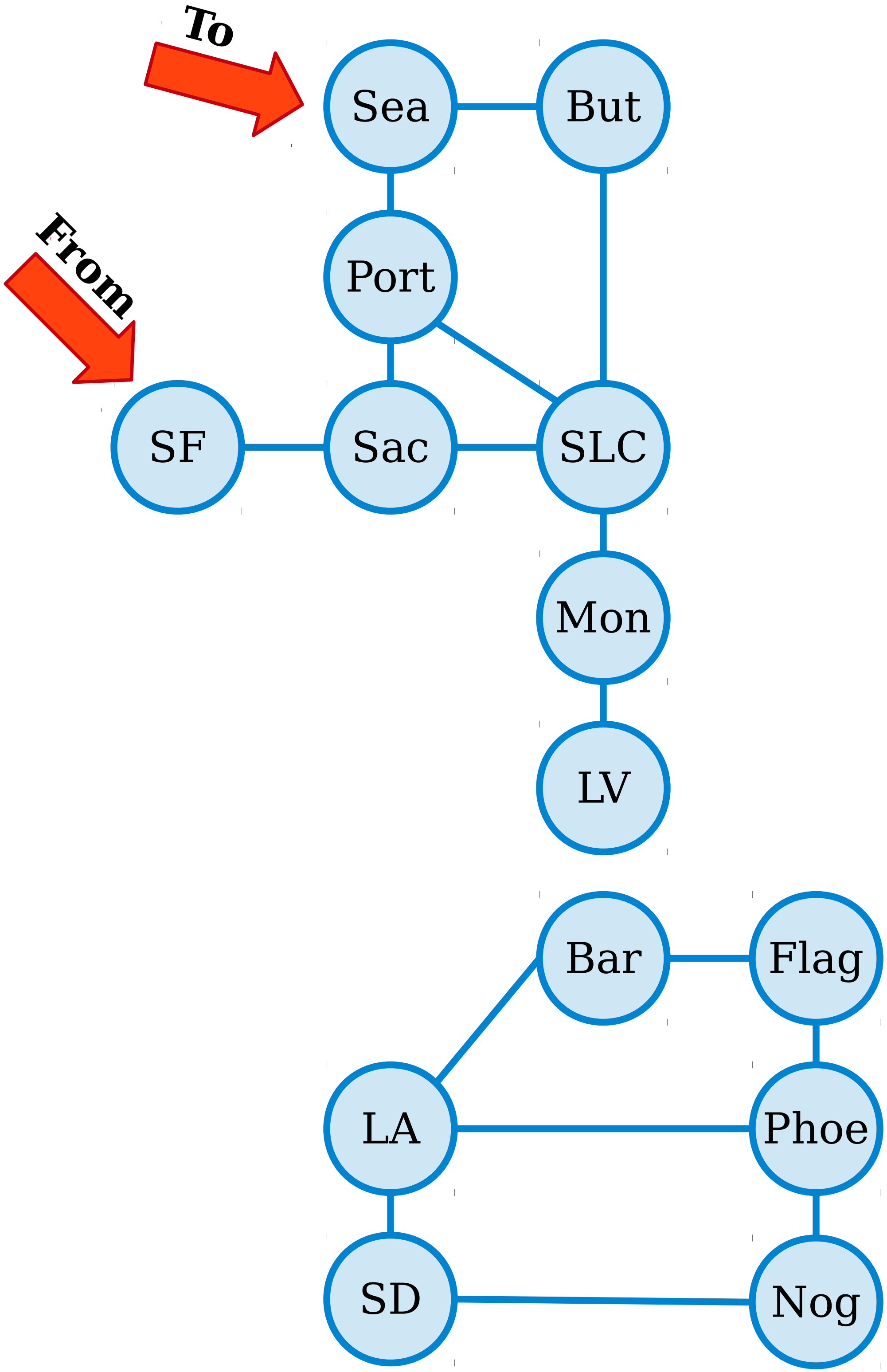
A ***path*** in a graph $G = (V, E)$ is a sequence of one or more nodes $v_1, v_2, v_3, \ldots, v_n$ such that any two consecutive nodes in the sequence are adjacent.



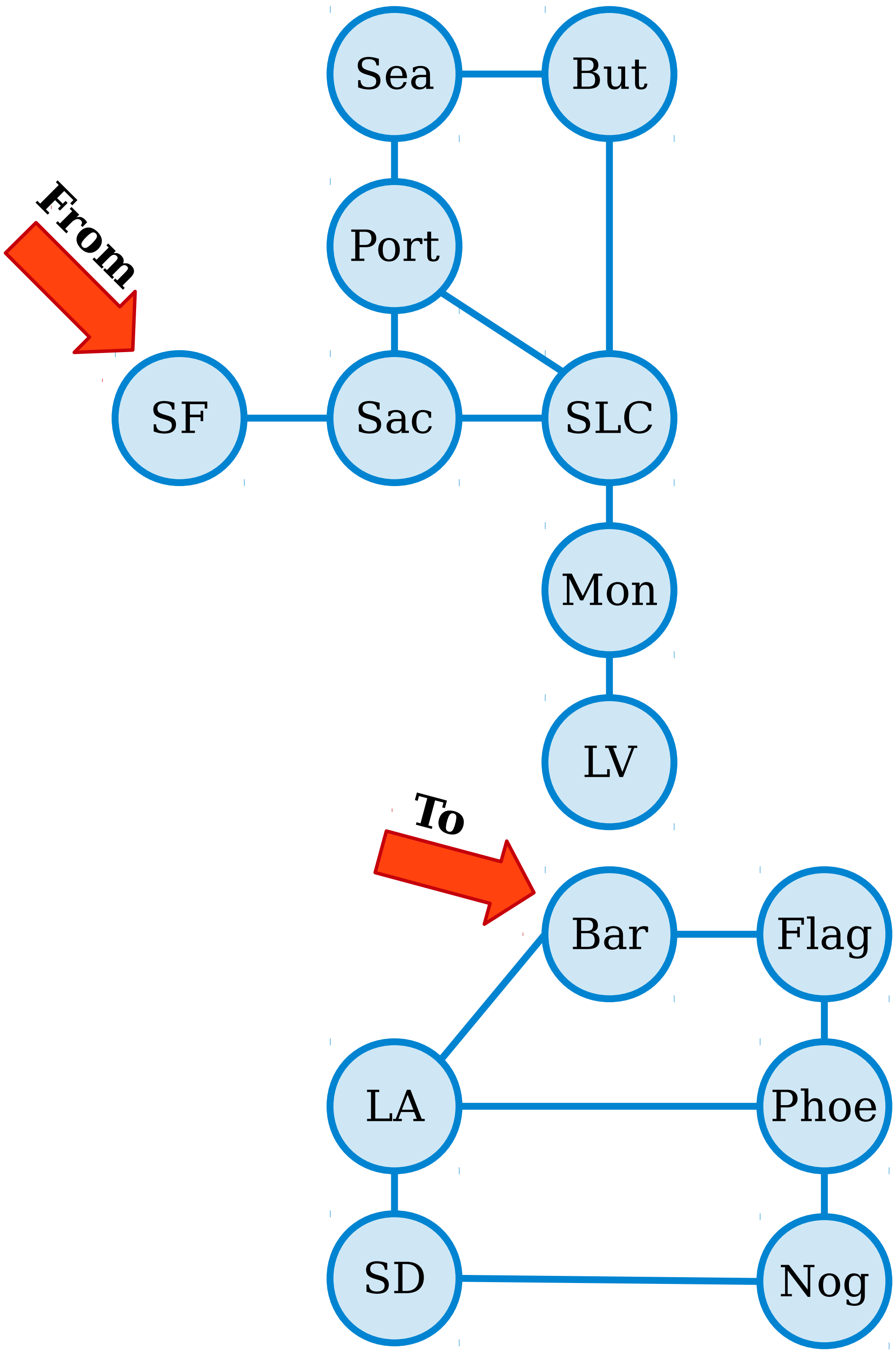
A ***path*** in a graph $G = (V, E)$ is a sequence of one or more nodes $v_1, v_2, v_3, \ldots, v_n$ such that any two consecutive nodes in the sequence are adjacent.



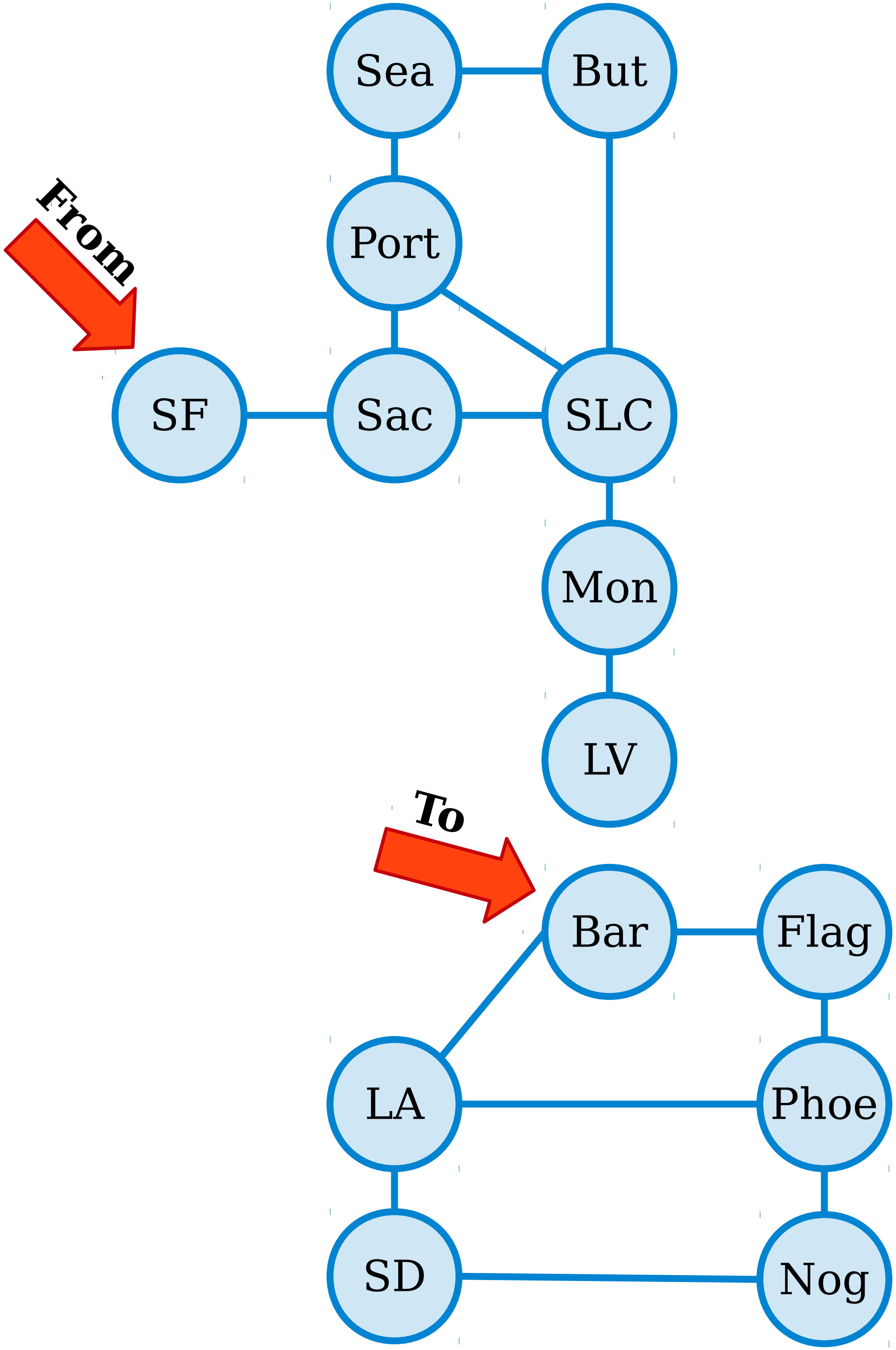
A ***path*** in a graph $G = (V, E)$ is a sequence of one or more nodes $v_1, v_2, v_3, \ldots, v_n$ such that any two consecutive nodes in the sequence are adjacent.



A ***path*** in a graph $G = (V, E)$ is a sequence of one or more nodes $v_1, v_2, v_3, \ldots, v_n$ such that any two consecutive nodes in the sequence are adjacent.

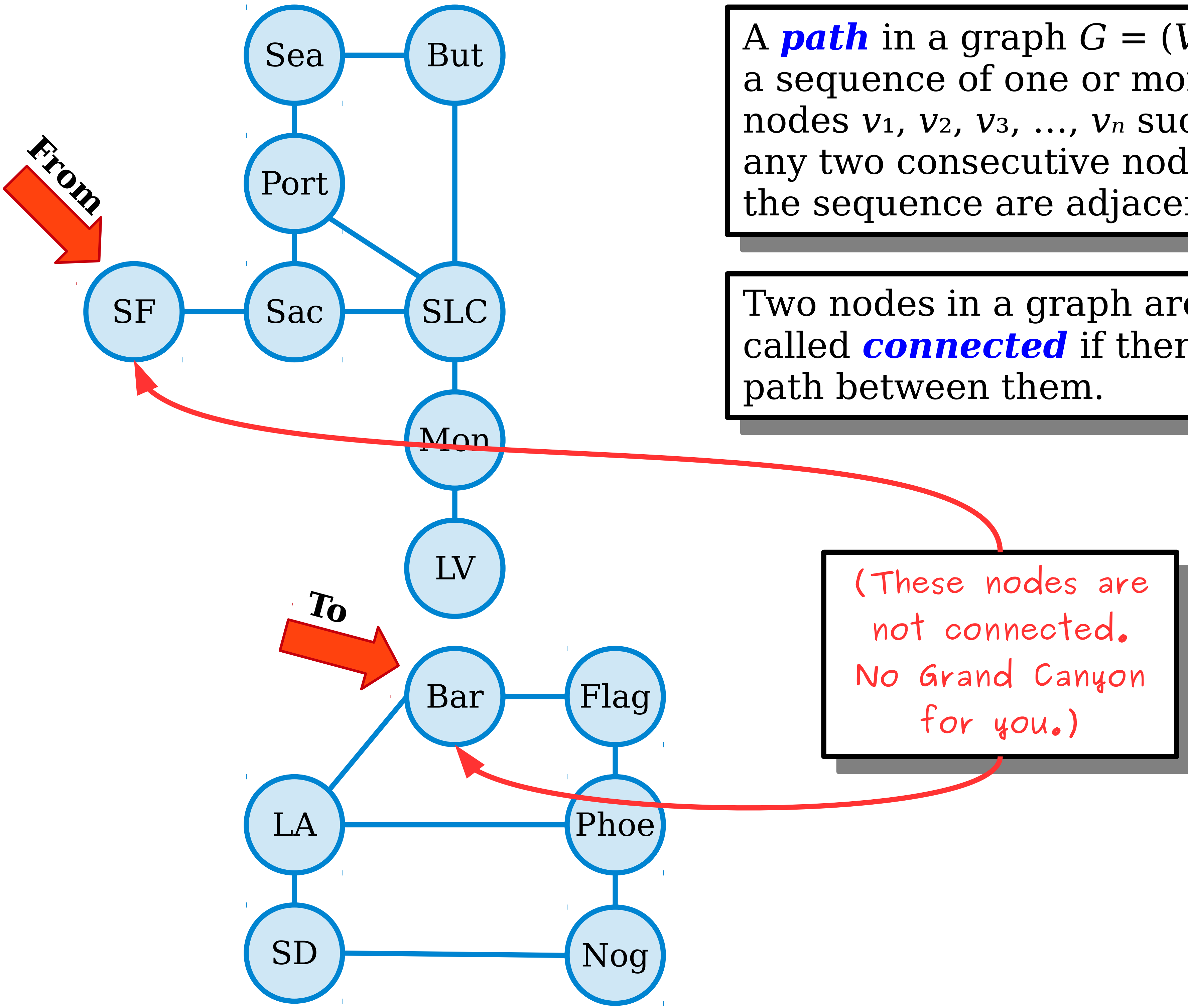


A ***path*** in a graph $G = (V, E)$ is a sequence of one or more nodes $v_1, v_2, v_3, \ldots, v_n$ such that any two consecutive nodes in the sequence are adjacent.



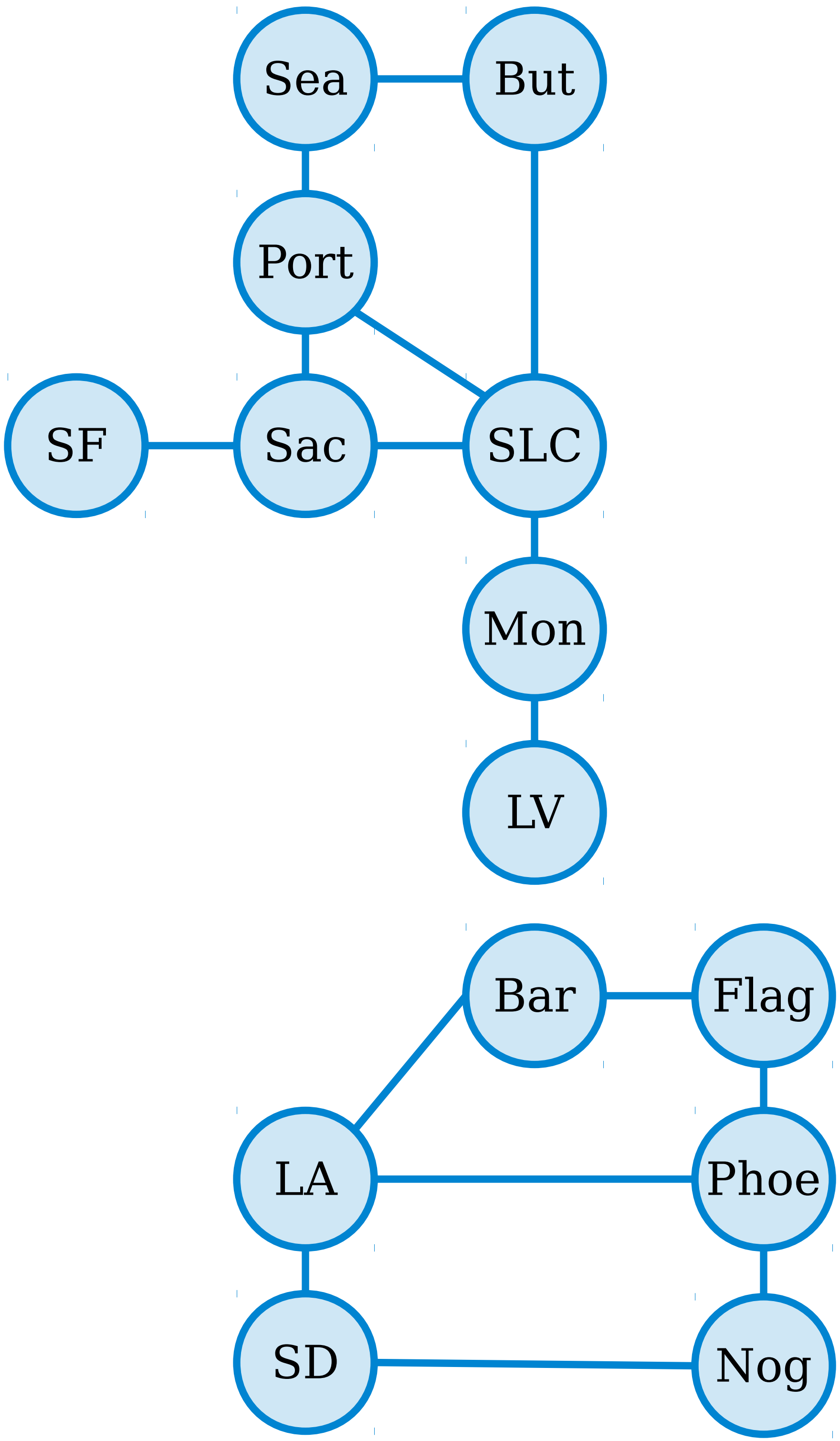
A ***path*** in a graph $G = (V, E)$ is a sequence of one or more nodes $v_1, v_2, v_3, \ldots, v_n$ such that any two consecutive nodes in the sequence are adjacent.

Two nodes in a graph are called ***connected*** if there is a path between them.



A ***path*** in a graph $G = (V, E)$ is a sequence of one or more nodes $v_1, v_2, v_3, \ldots, v_n$ such that any two consecutive nodes in the sequence are adjacent.

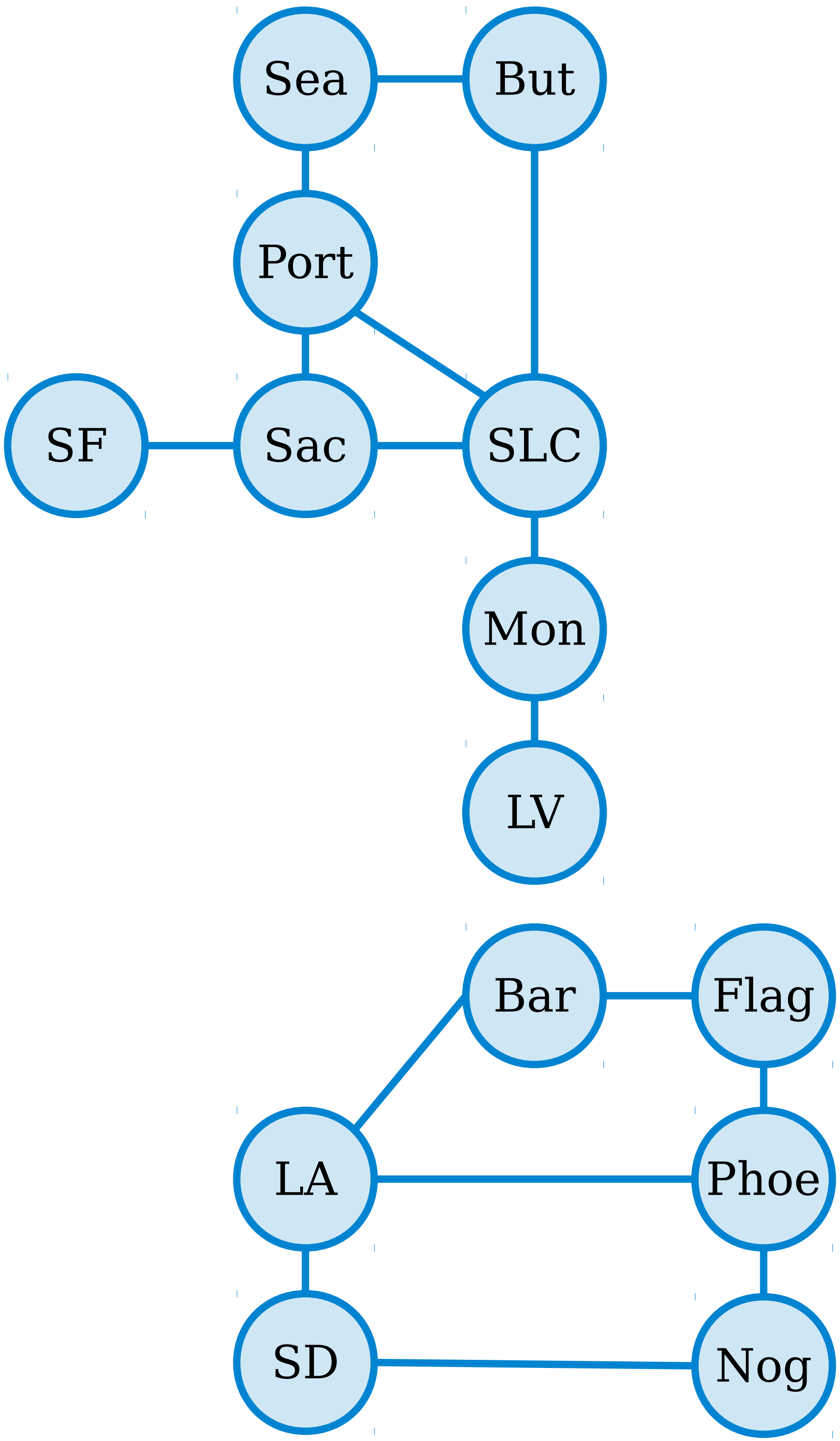
Two nodes in a graph are called ***connected*** if there is a path between them.



A ***path*** in a graph $G = (V, E)$ is a sequence of one or more nodes $v_1, v_2, v_3, \ldots, v_n$ such that any two consecutive nodes in the sequence are adjacent.

Two nodes in a graph are called ***connected*** if there is a path between them.

A graph G as a whole is called ***connected*** if all pairs of nodes in G are connected.



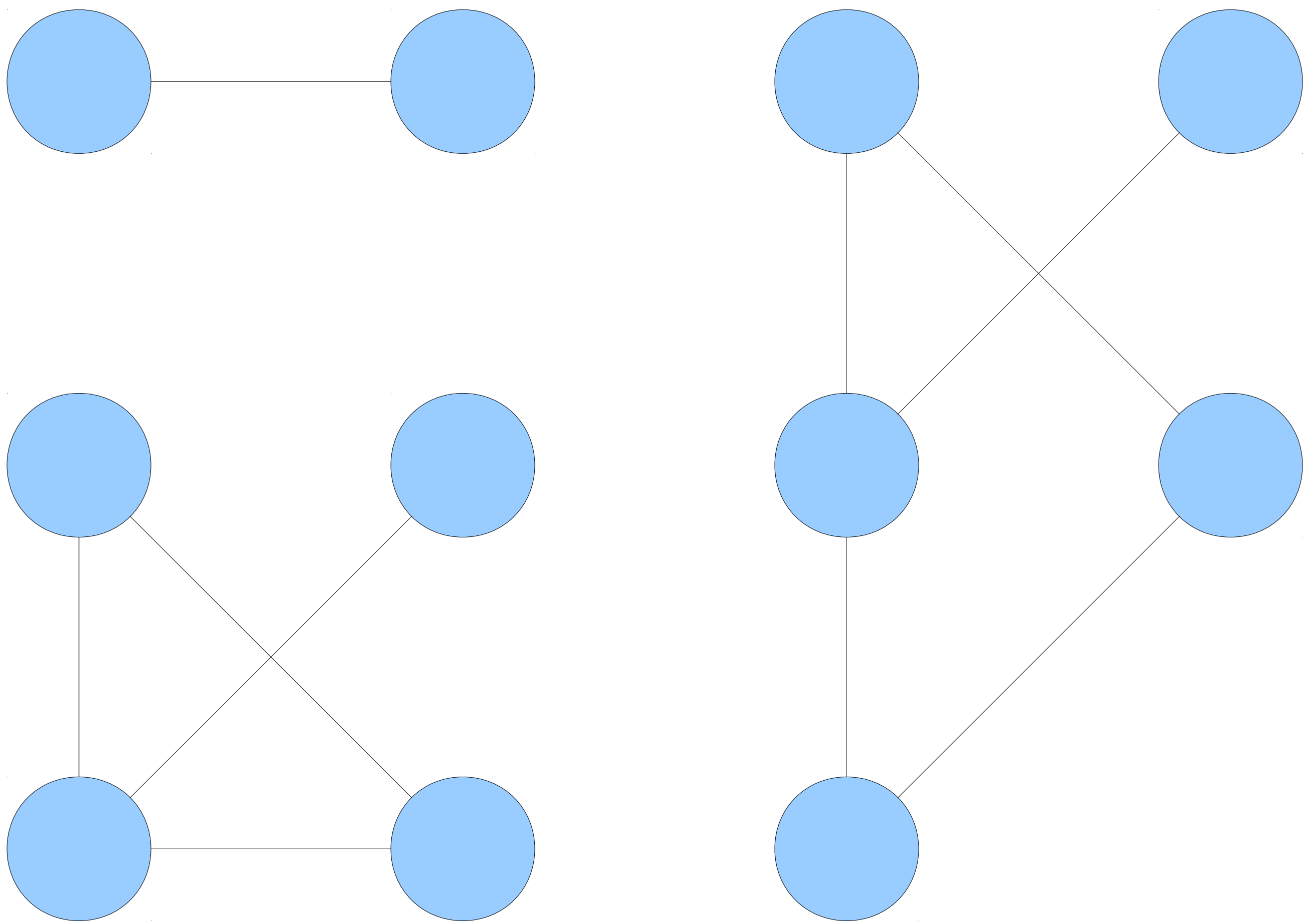
A ***path*** in a graph $G = (V, E)$ is a sequence of one or more nodes $v_1, v_2, v_3, \ldots, v_n$ such that any two consecutive nodes in the sequence are adjacent.

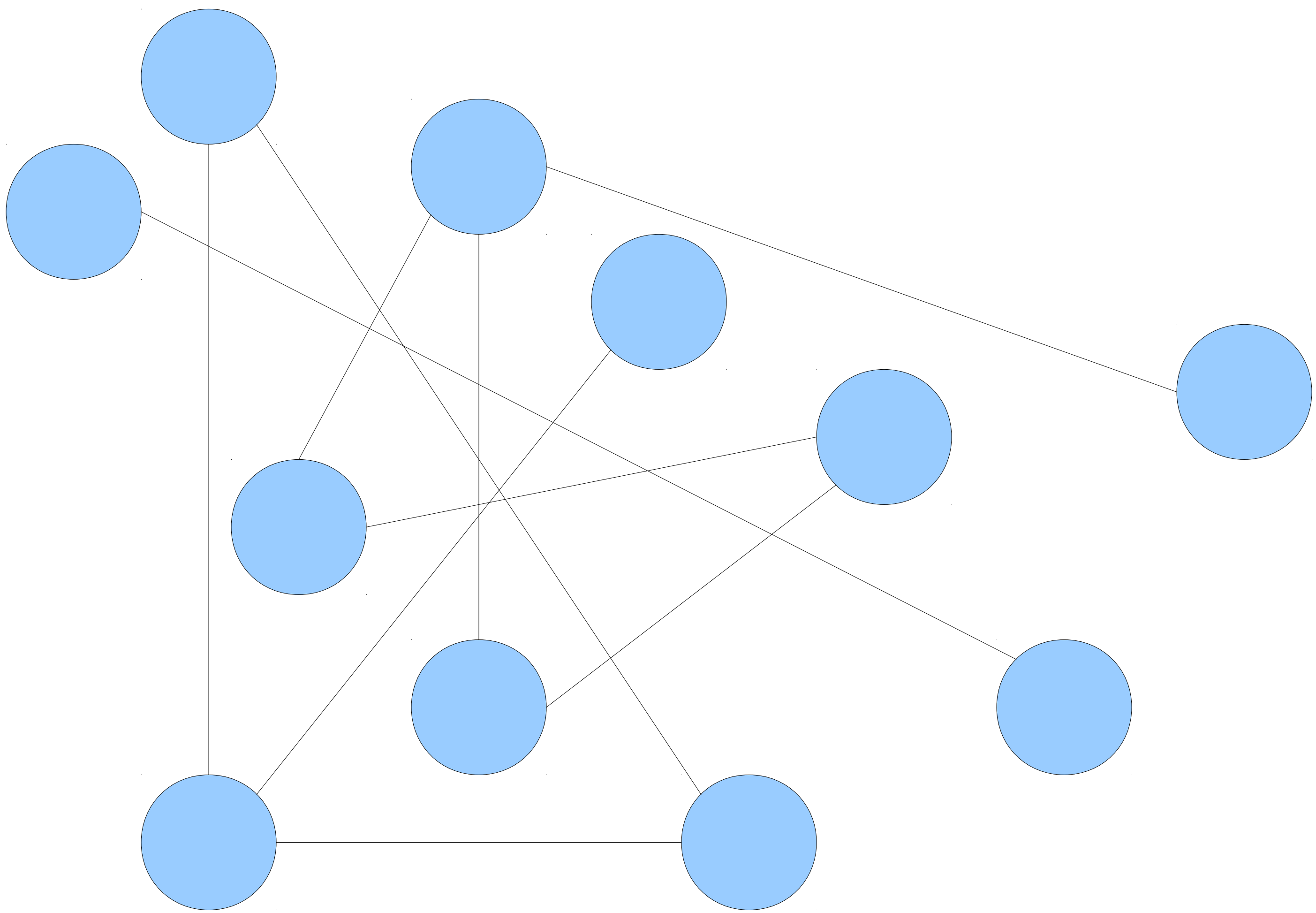
Two nodes in a graph are called ***connected*** if there is a path between them.

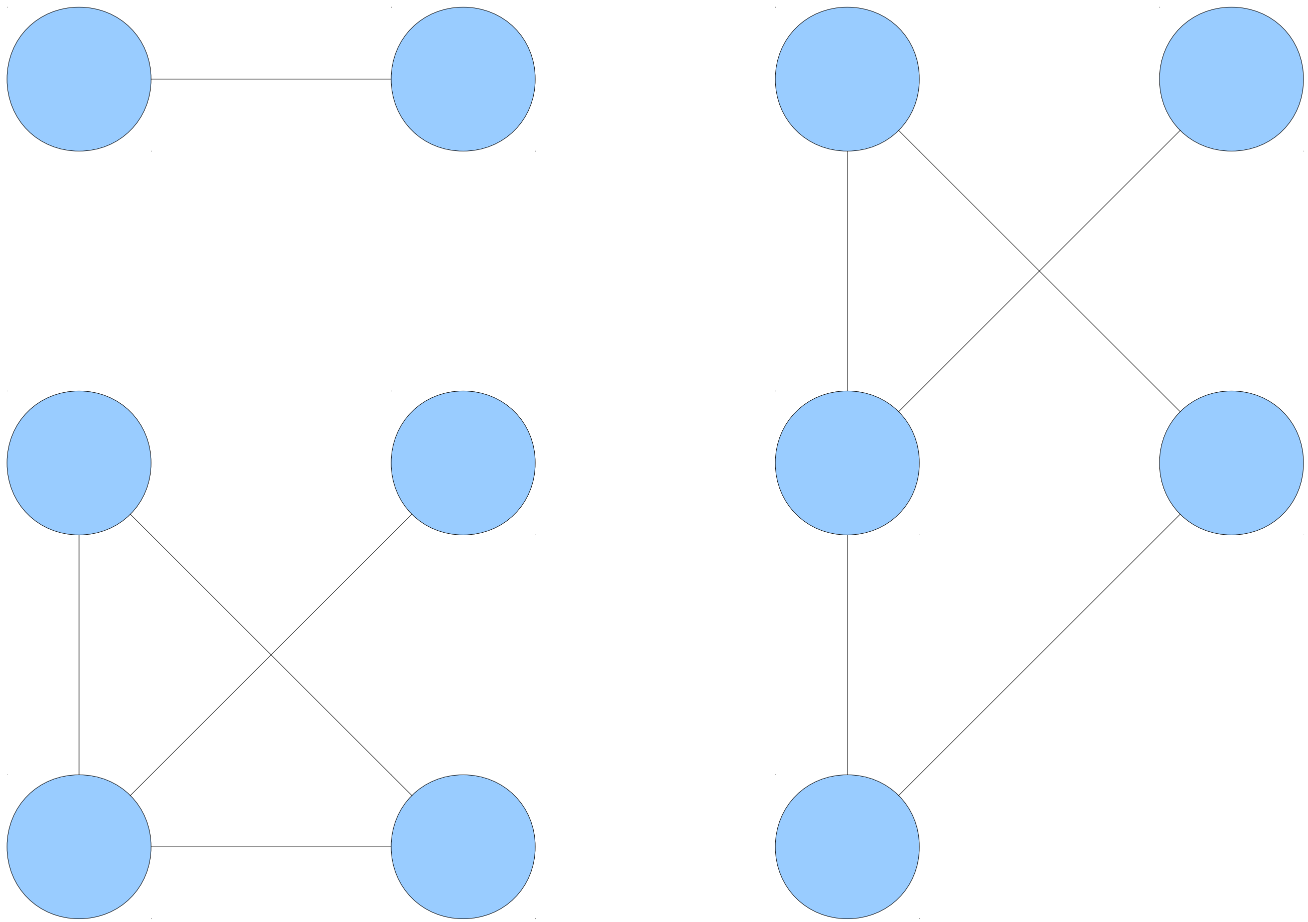
A graph G as a whole is called ***connected*** if all pairs of nodes in G are connected.

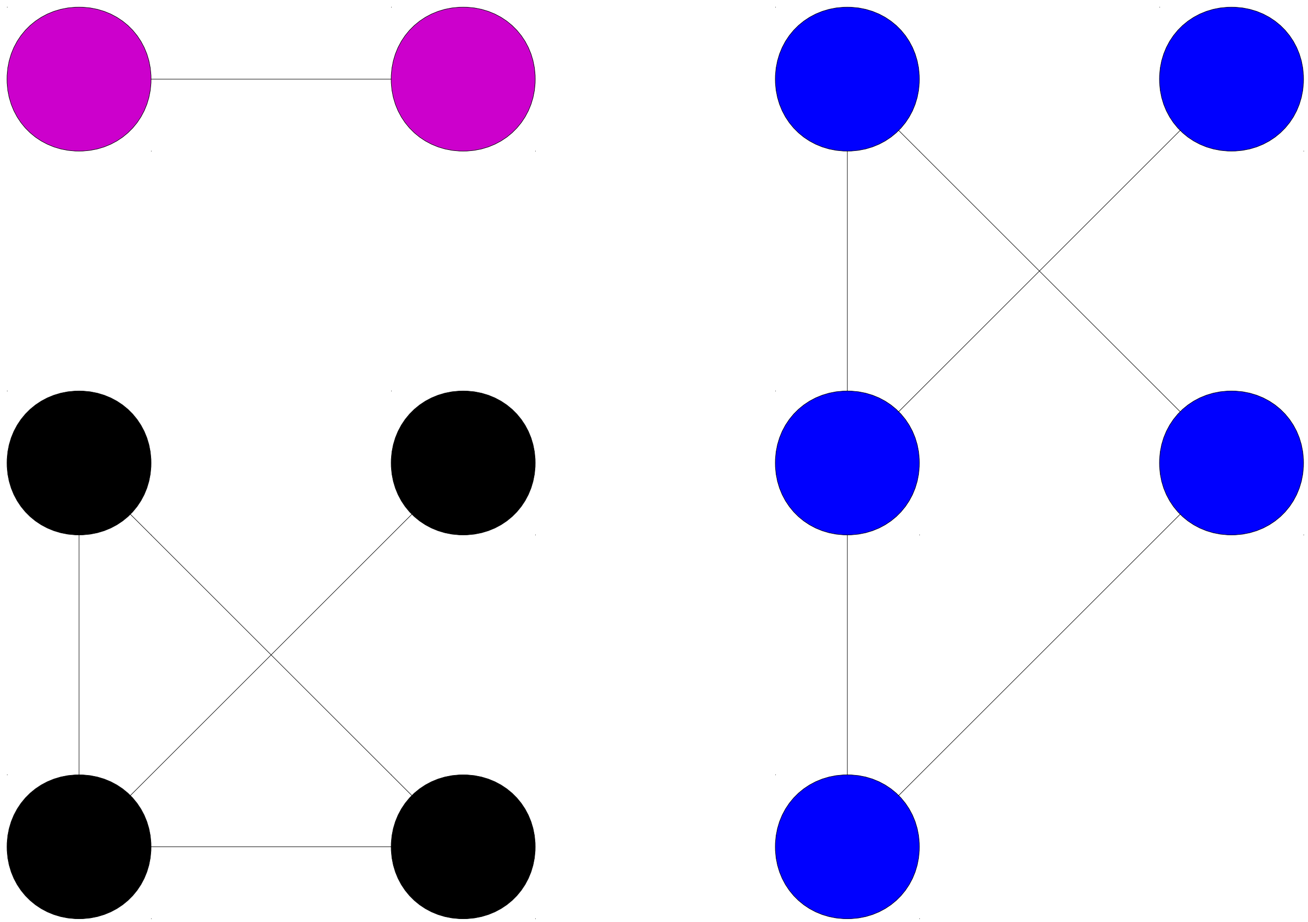
(This graph is not connected.)

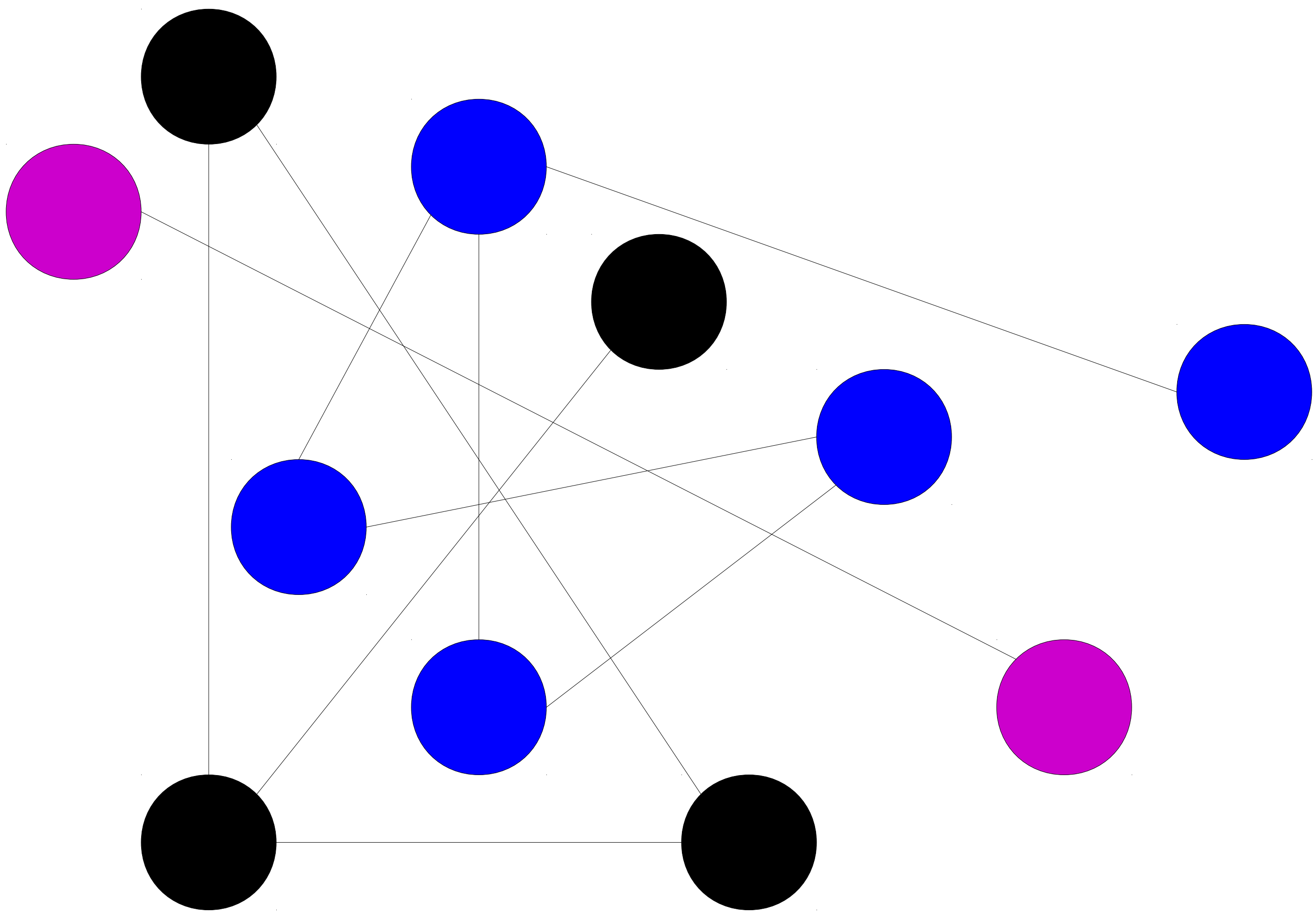
Connected Components

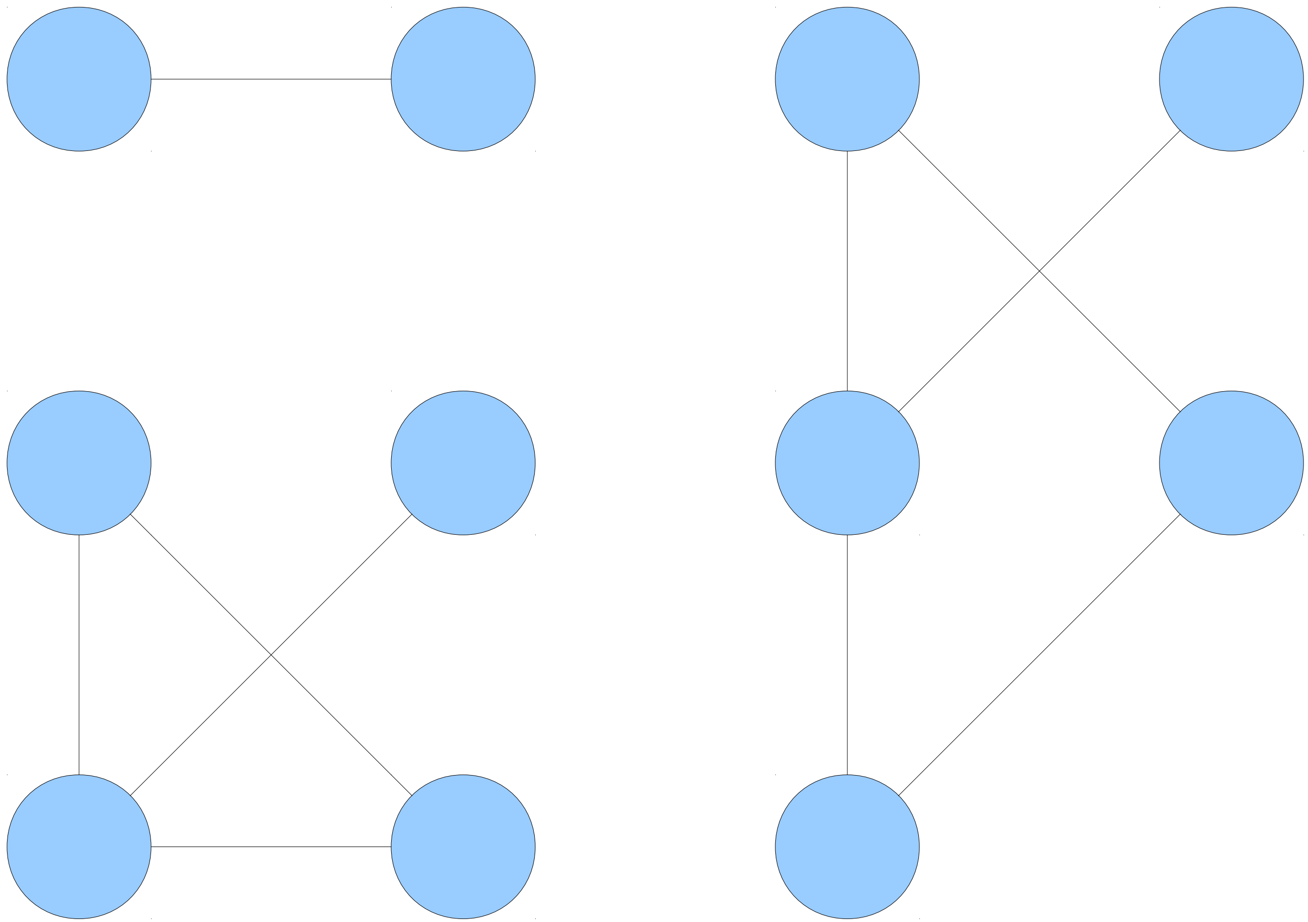


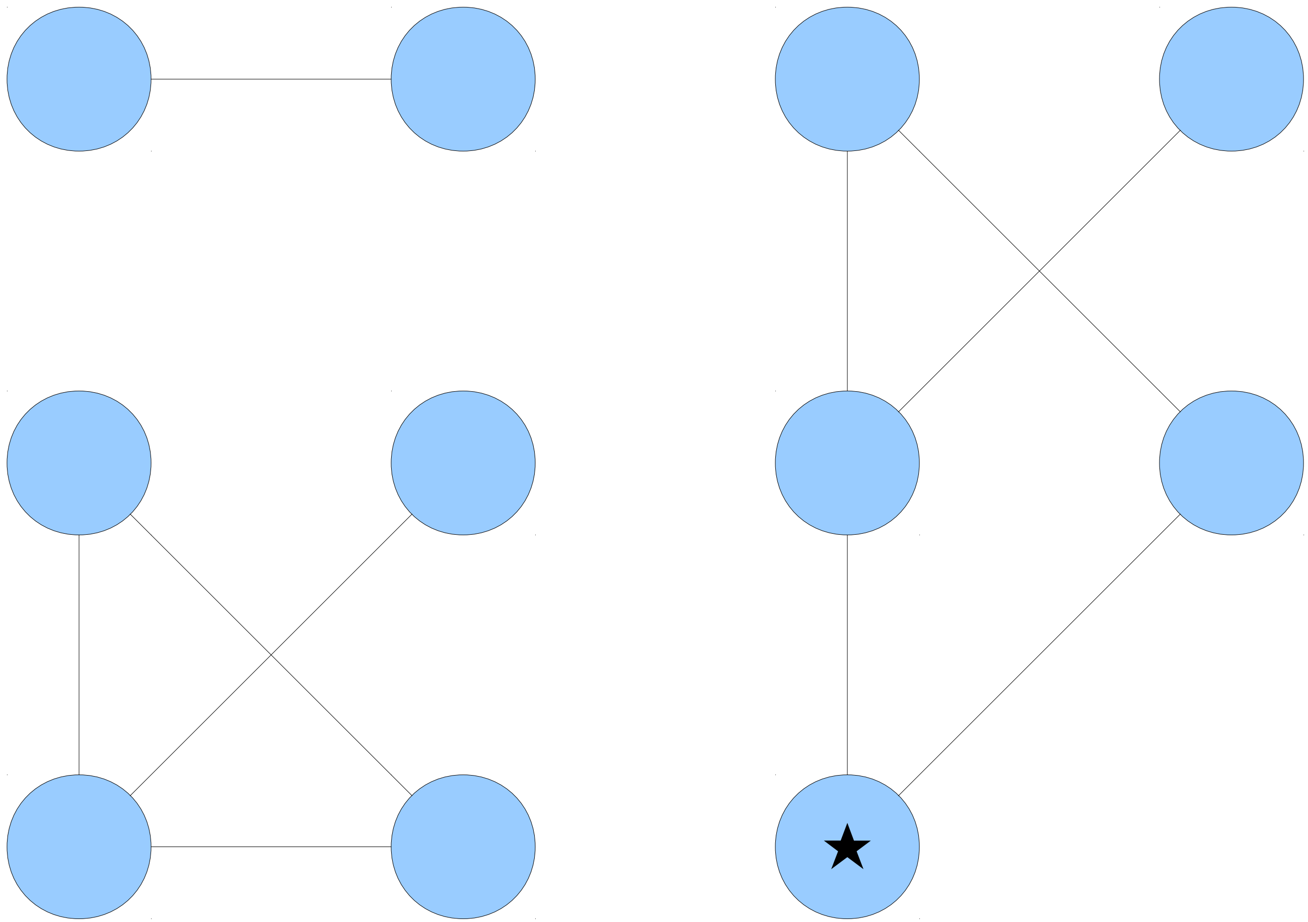


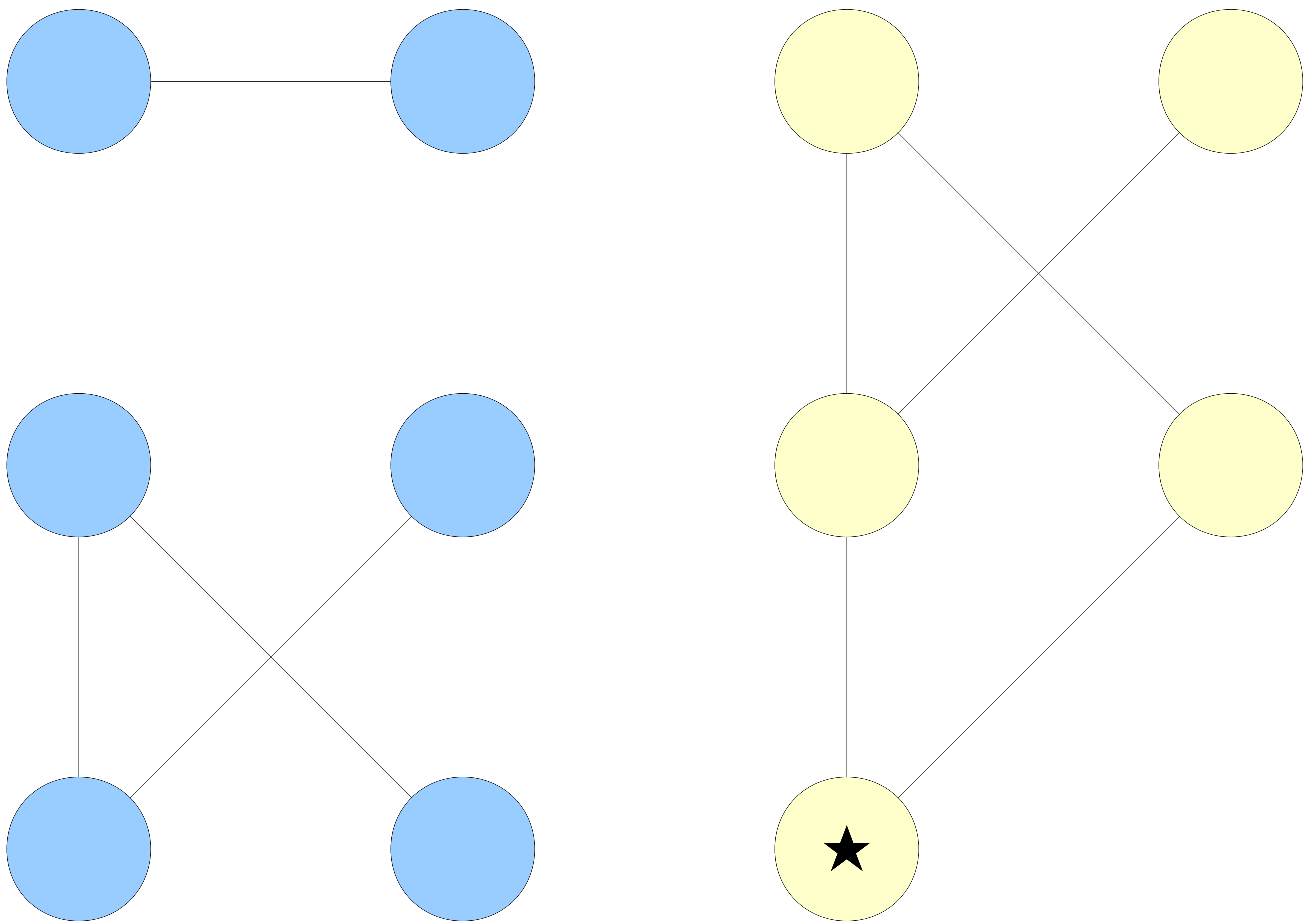












Connected Components

- Let $G = (V, E)$ be a graph. For each $v \in V$, the ***connected component*** containing v is the set

$$[v] = \{ x \in V \mid v \text{ is connected to } x \}$$

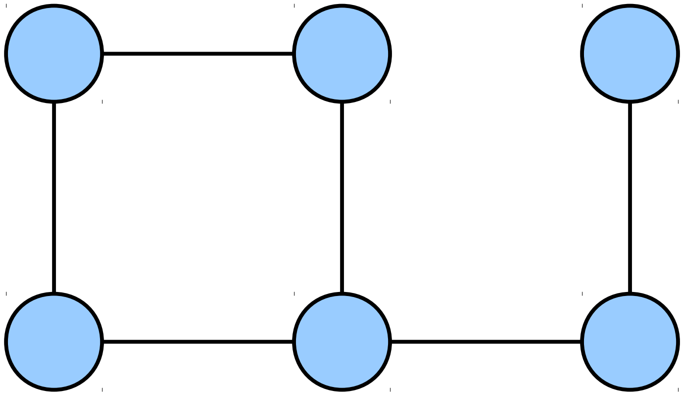
- Intuitively, a connected component is a "piece" of a graph in the sense we just talked about.
- ***Question:*** How do we know that this particular definition of a "piece" of a graph is a good one?
- ***Goal:*** Prove that any graph can be broken apart into different connected components.

We’re trying to reason about some way of partitioning the nodes in a graph into different groups.

What structure have we studied that captures the idea of a partition?

Connectivity

- ***Claim:*** For any graph *G*, the "is connected to" relation is an equivalence relation.
  - Is it reflexive?
  - Is it symmetric?
  - Is it transitive?



Connectivity

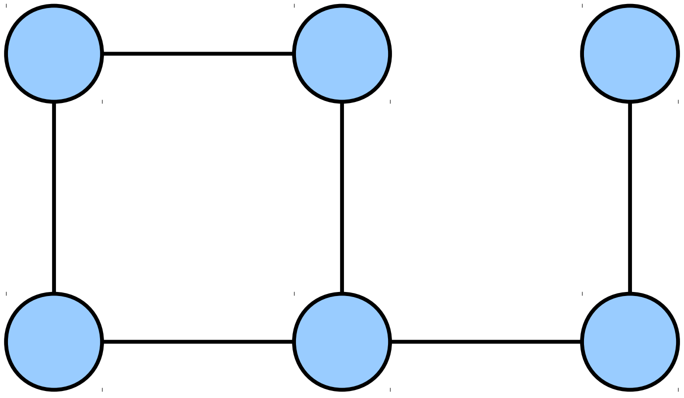
Claim: For any graph *G*, the "is connected to" relation is an equivalence relation.

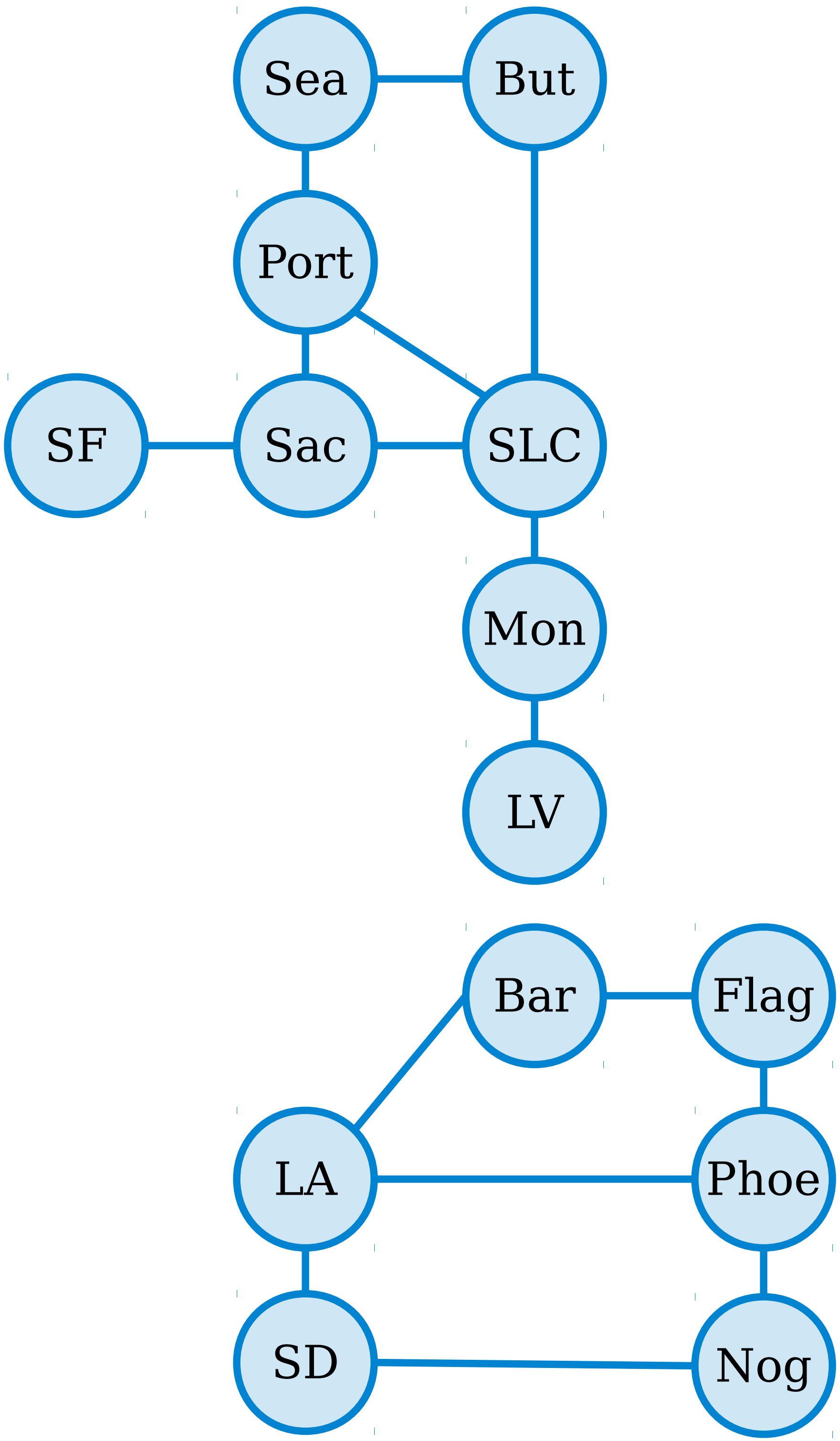
- Is it reflexive?

  Is it symmetric?

  Is it transitive?

$$\forall v \in V.\ Conn(v, v)$$





A ***path*** in a graph $G = (V, E)$ is a sequence of one or more nodes $v_1, v_2, v_3, \ldots, v_n$ such that any two consecutive nodes in the sequence are adjacent.

Two nodes in a graph are called ***connected*** if there is a path between them

Connectivity

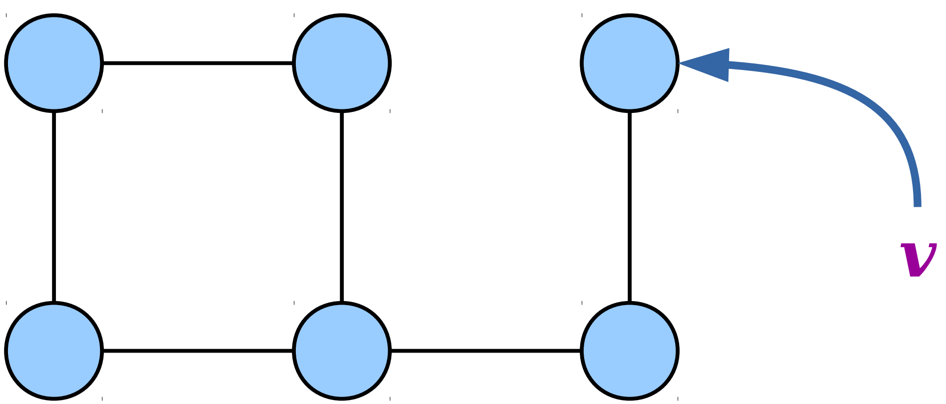
Claim: For any graph G, the "is connected to" relation is an equivalence relation.

- Is it reflexive?

Is it symmetric?

Is it transitive?

$$\forall v \in V.\ Conn(v, v)$$



A ***path*** in a graph $G = (V, E)$ is a sequence of one or more nodes $v_1, v_2, v_3, \ldots, v_n$ such that any two consecutive nodes in the sequence are adjacent.

Connectivity

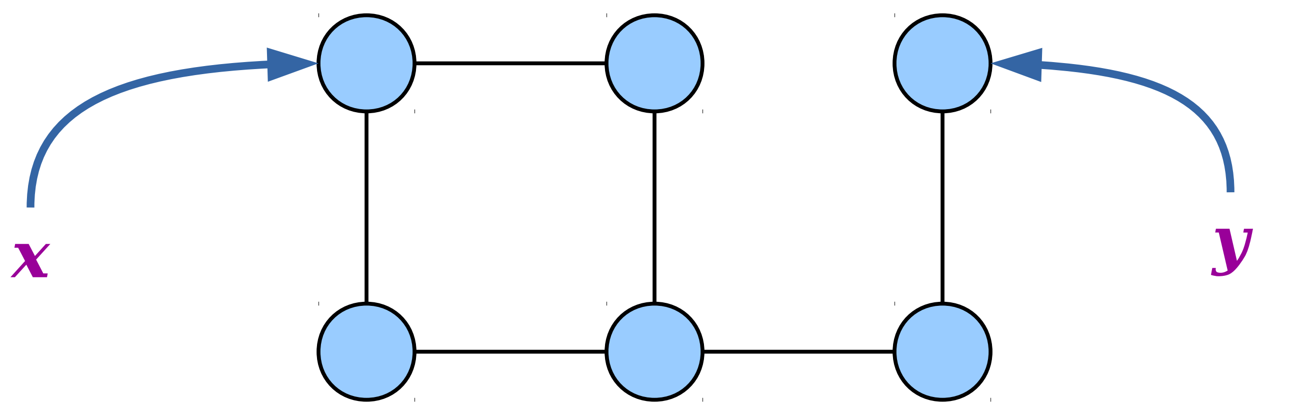
Claim: E
connecte
relation.

$\forall x \in V.\ \forall y \in V.\ (Conn(x, y) \rightarrow Conn(y, x))$

Is it reflexive?

- Is it symmetric?

Is it transitive?



Connectivity

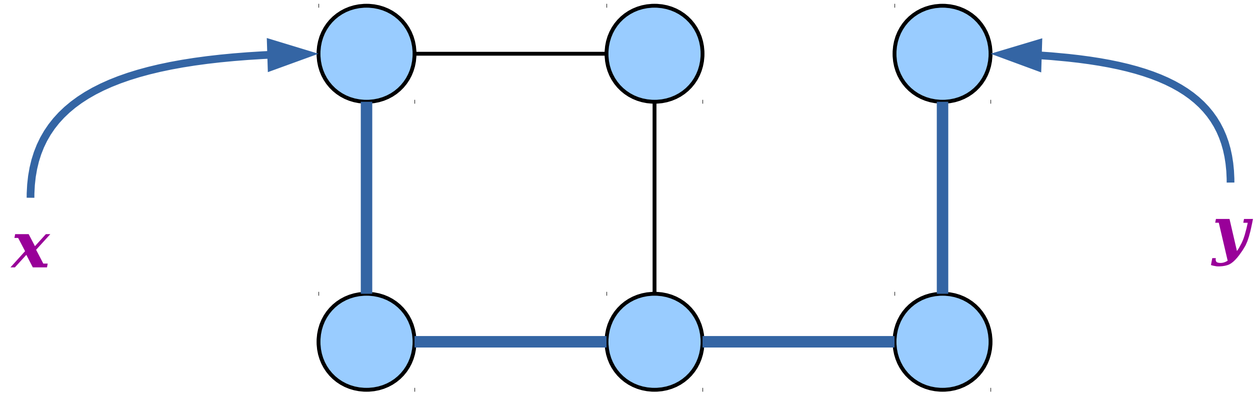
Claim: E
connecte
relation.

$$\forall x \in V.\ \forall y \in V.\ (Conn(x, y) \rightarrow Conn(y, x))$$

Is it reflexive?

- Is it symmetric?

Is it transitive?



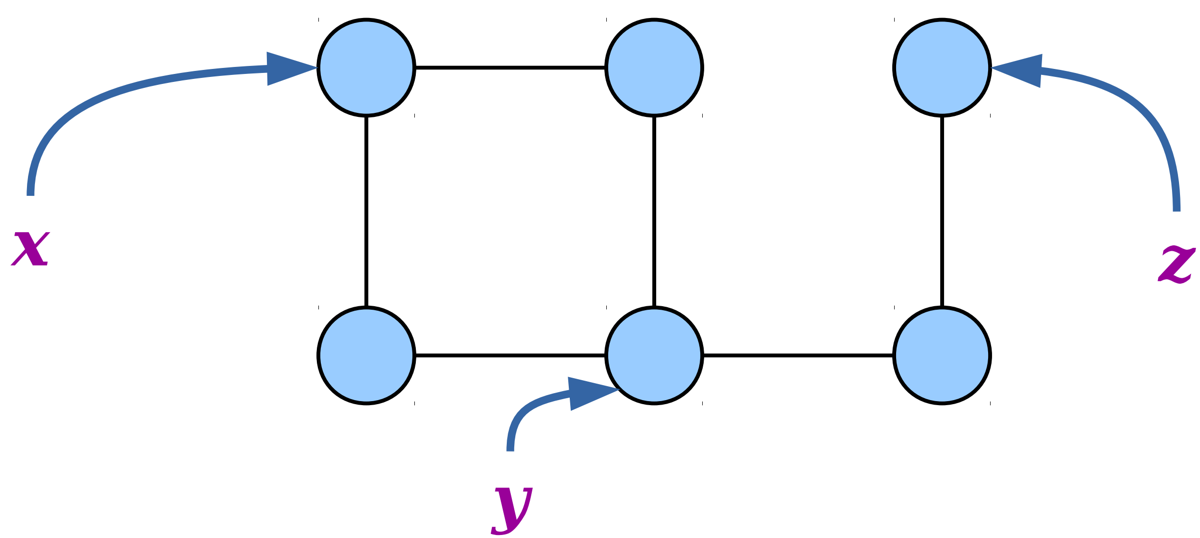
Connectivity

$$\forall x \in V.\ \forall y \in V.\ \forall z \in V.\ (Conn(x, y) \land Conn(y, z) \rightarrow Conn(x, z))$$

Is it reflexive?

Is it symmetric?

- Is it transitive?



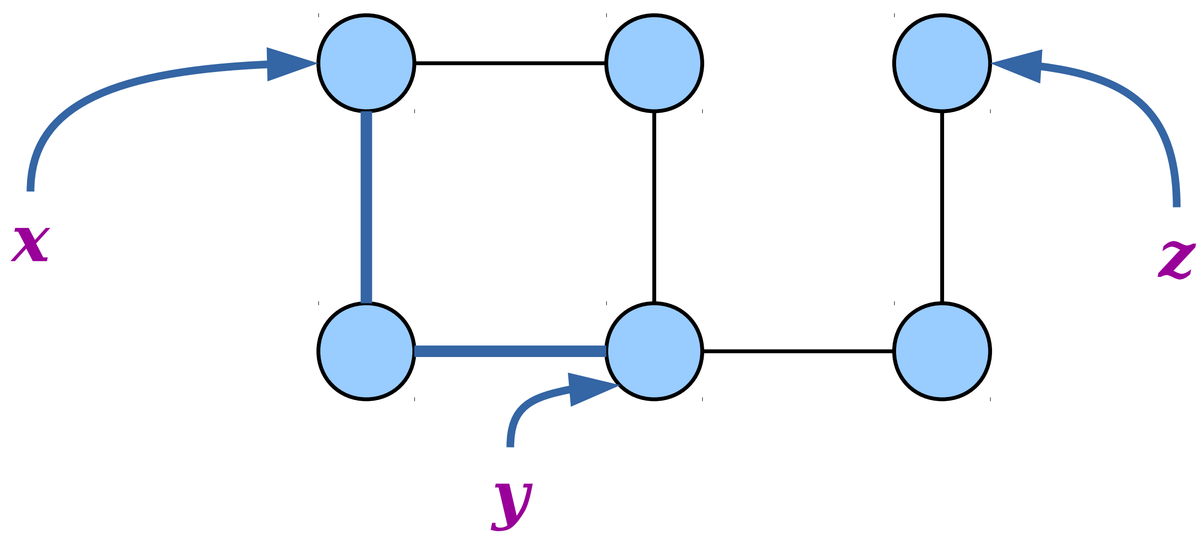
Connectivity

$$\forall x \in V.\ \forall y \in V.\ \forall z \in V.\ (Conn(x, y) \land Conn(y, z) \rightarrow Conn(x, z))$$

Is it reflexive?

Is it symmetric?

- Is it transitive?



Theorem: Let $G = (V, E)$ be a graph. Then the connectivity relation over V is an equivalence relation.

Proof: Consider an arbitrary graph $G = (V, E)$. We will prove that the connectivity relation over V is reflexive, symmetric, and transitive.

To show that connectivity is reflexive, consider any $v \in V$. Then the singleton path v is a path from v to itself. Therefore, v is connected to itself, as required.

To show that connectivity is symmetric, consider any $x, y \in V$ where x is connected to y. We need to show that y is connected to x. Since x is connected to y, there is some path $x, v_1, \ldots, v_n, y$ from x to y. Then $y, v_n, \ldots, v_1, x$ is a path from y back to x, so y is connected to x.

Finally, to show that connectivity is transitive, let $x, y, z \in V$ be arbitrary nodes where x is connected to y and y is connected to z. We will prove that x is connected to z. Since x is connected to y, there is a path $x, u_1, \ldots, u_n, y$ from x to y. Since y is connected to z, there is a path $y, v_1, \ldots, v_k, z$ from y to z. Then the path $x, u_1, \ldots, u_n, y, v_1, \ldots, v_k, z$ goes from x to z. Thus x is connected to z, as required. ■

Putting Things Together

- Earlier, we defined the connected component of a node v to be

$$[v] = \{ x \in V \mid v \text{ is connected to } x \}$$

- Connectivity is an equivalence relation! So what's the equivalence class of a node v with respect to connectivity?

$$[v]_{conn} = \{ x \in V \mid v \text{ is connected to } x \}$$

- ***Connected components are equivalence classes of the connectivity relation!***

Theorem: If $G = (V, E)$ is a graph, then every node in G belongs to exactly one connected component of G.

Proof: Let $G = (V, E)$ be an arbitrary graph and let $v \in V$ be any node in G. The connected components of G are just the equivalence classes of the connectivity relation in G. The Fundamental Theorem of Equivalence Relations guarantees that v belongs to exactly one equivalence class of the connectivity relation. Therefore, v belongs to exactly one connected component in G. ■

Time-Out for Announcements!

Problem Set Three

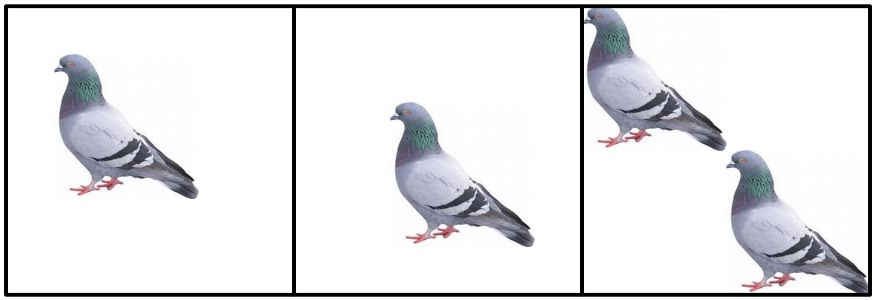
- The checkpoint problems for PS3 were due at 3:00PM today.
  - We'll try to have it graded and returned by Wednesday morning.
- The remaining problems from PS3 are due on Friday at 3:00PM.
  - Have questions? Stop by office hours or ask on Piazza!

Back to CS103!

The Pigeonhole Principle

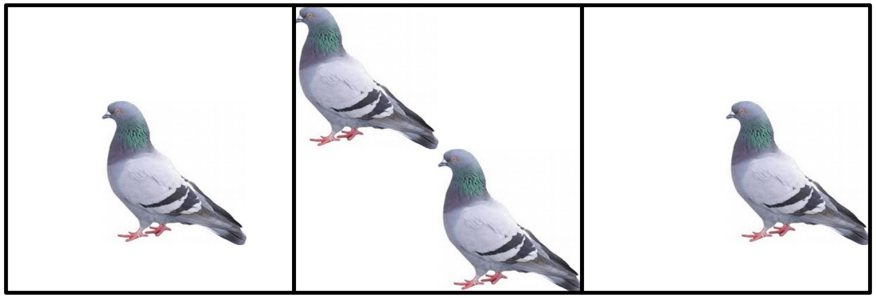
The Pigeonhole Principle

- ***Theorem (The Pigeonhole Principle):*** If m objects are distributed into n bins and $m > n$, then at least one bin will contain at least two objects.



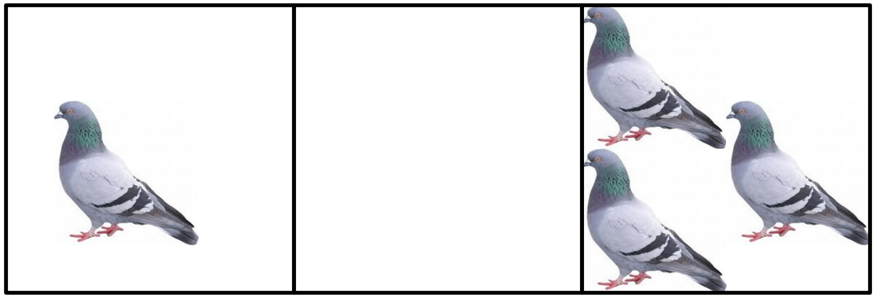
The Pigeonhole Principle

- ***Theorem (The Pigeonhole Principle):*** If m objects are distributed into n bins and $m > n$, then at least one bin will contain at least two objects.



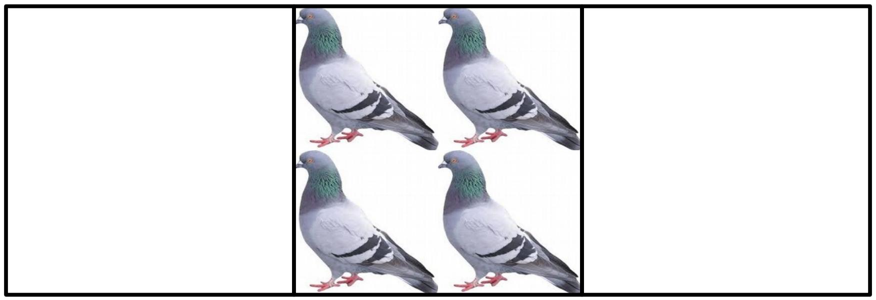
The Pigeonhole Principle

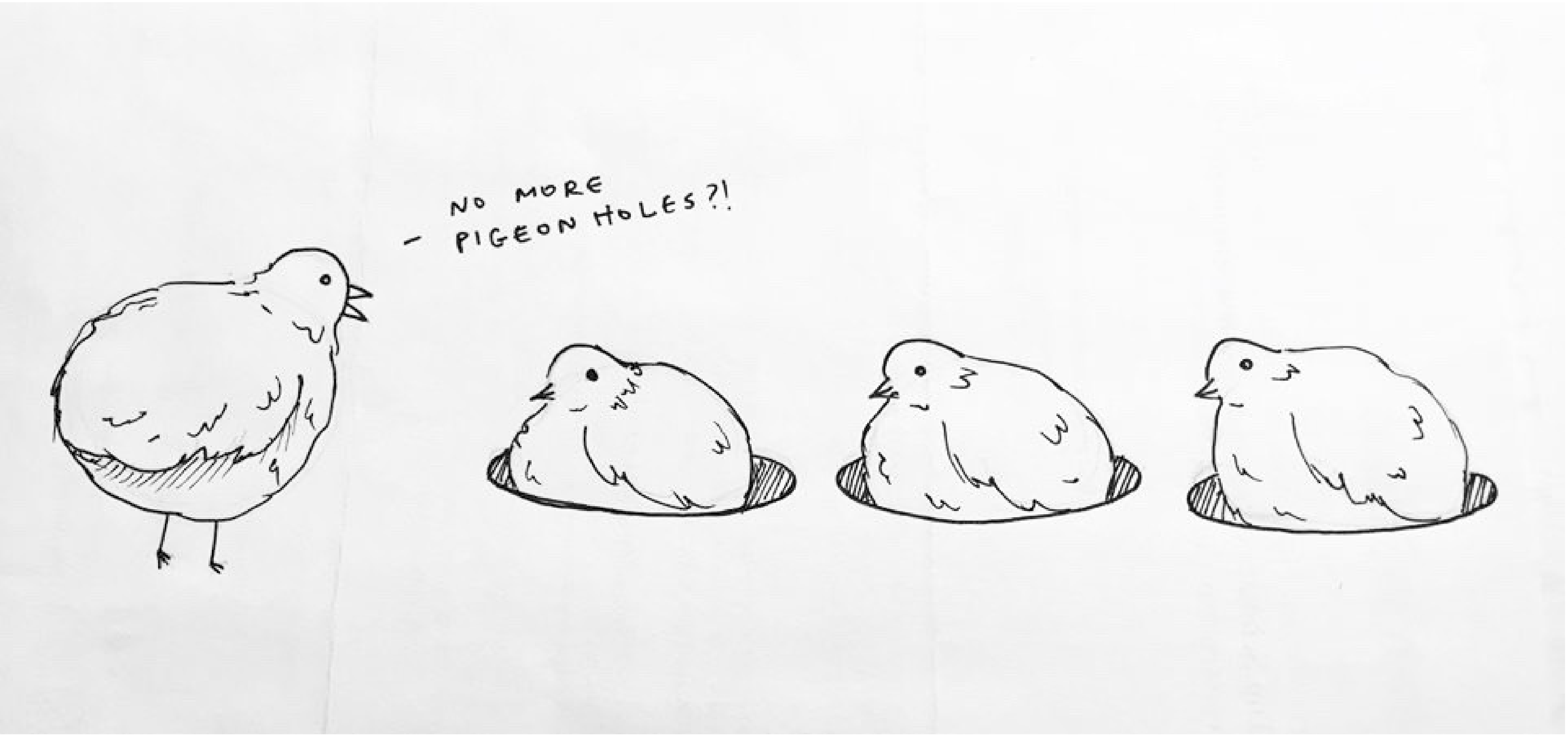
- ***Theorem (The Pigeonhole Principle):*** If m objects are distributed into n bins and $m > n$, then at least one bin will contain at least two objects.



The Pigeonhole Principle

- ***Theorem (The Pigeonhole Principle):*** If m objects are distributed into n bins and $m > n$, then at least one bin will contain at least two objects.





$m = 4, n = 3$

Some Simple Applications

- Any group of 367 people must have a pair of people that share a birthday.
  - 366 possible birthdays (pigeonholes)
  - 367 people (pigeons)
- Two people in San Francisco have the exact same number of hairs on their head.
  - Maximum number of hairs ever found on a human head is no greater than 500,000.
  - There are over 800,000 people in San Francisco.

Proving the Pigeonhole Principle

Theorem: If m objects are distributed into n bins and $m > n$, then there must be some bin that contains at least two objects.

Proof: Suppose for the sake of contradiction that, for some m and n where $m > n$, there is a way to distribute m objects into n bins such that each bin contains at most one object.

Number the bins 1, 2, 3, …, n and let x_i denote the number of objects in bin i. There are m objects in total, so we know that

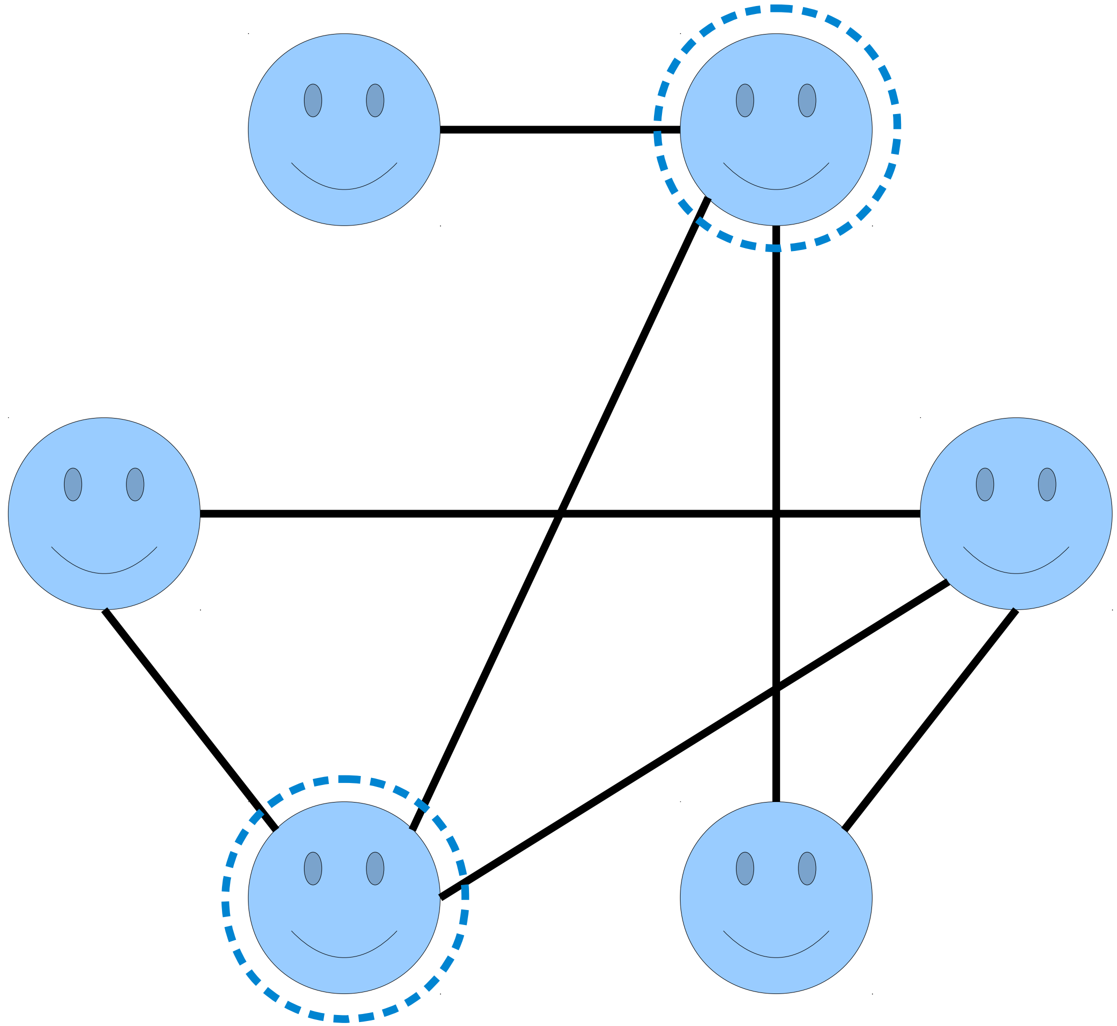
$$m = x_1 + x_2 + \ldots + x_n.$$

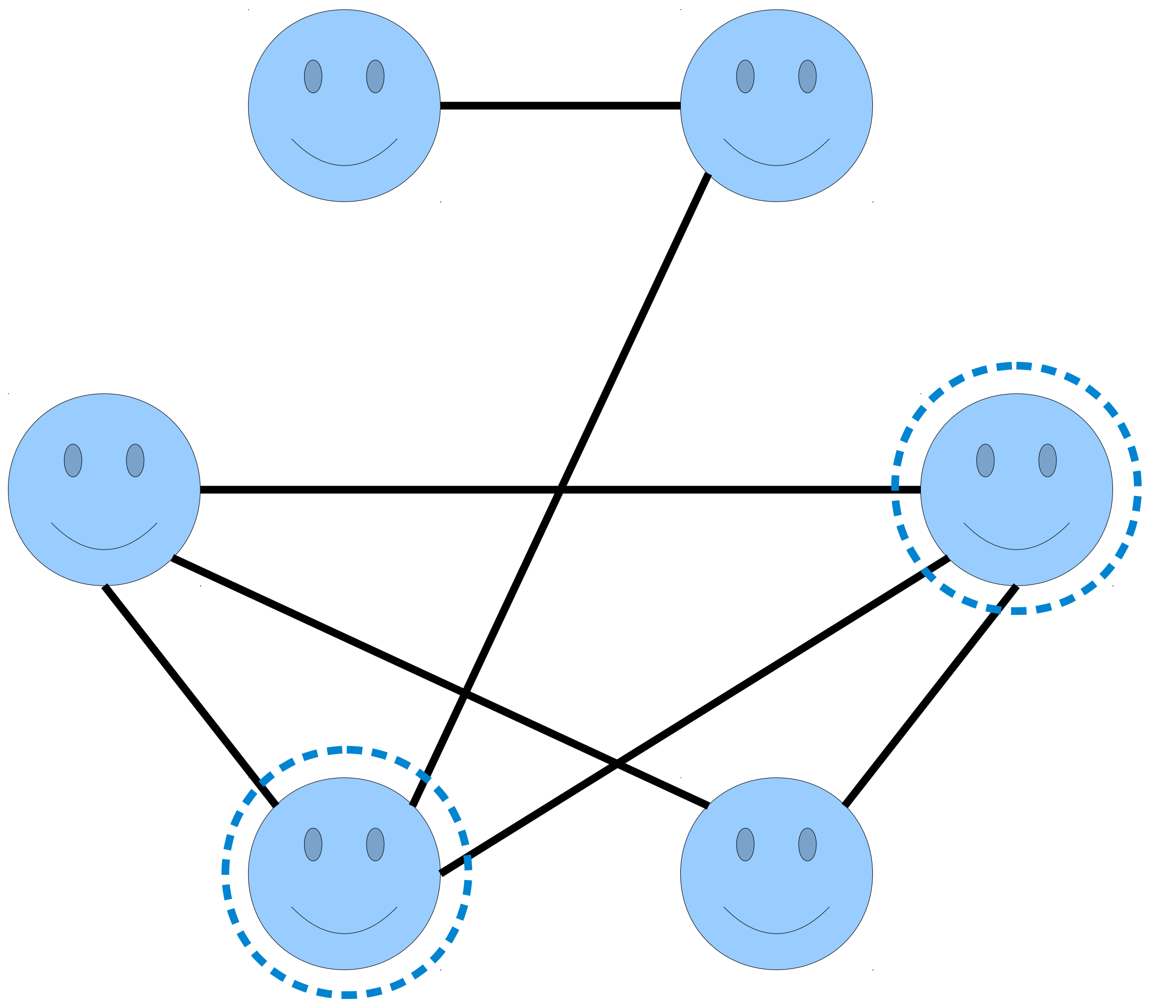
Since each bin has at most one object in it, we know $x_i \leq 1$ for each i. This means that

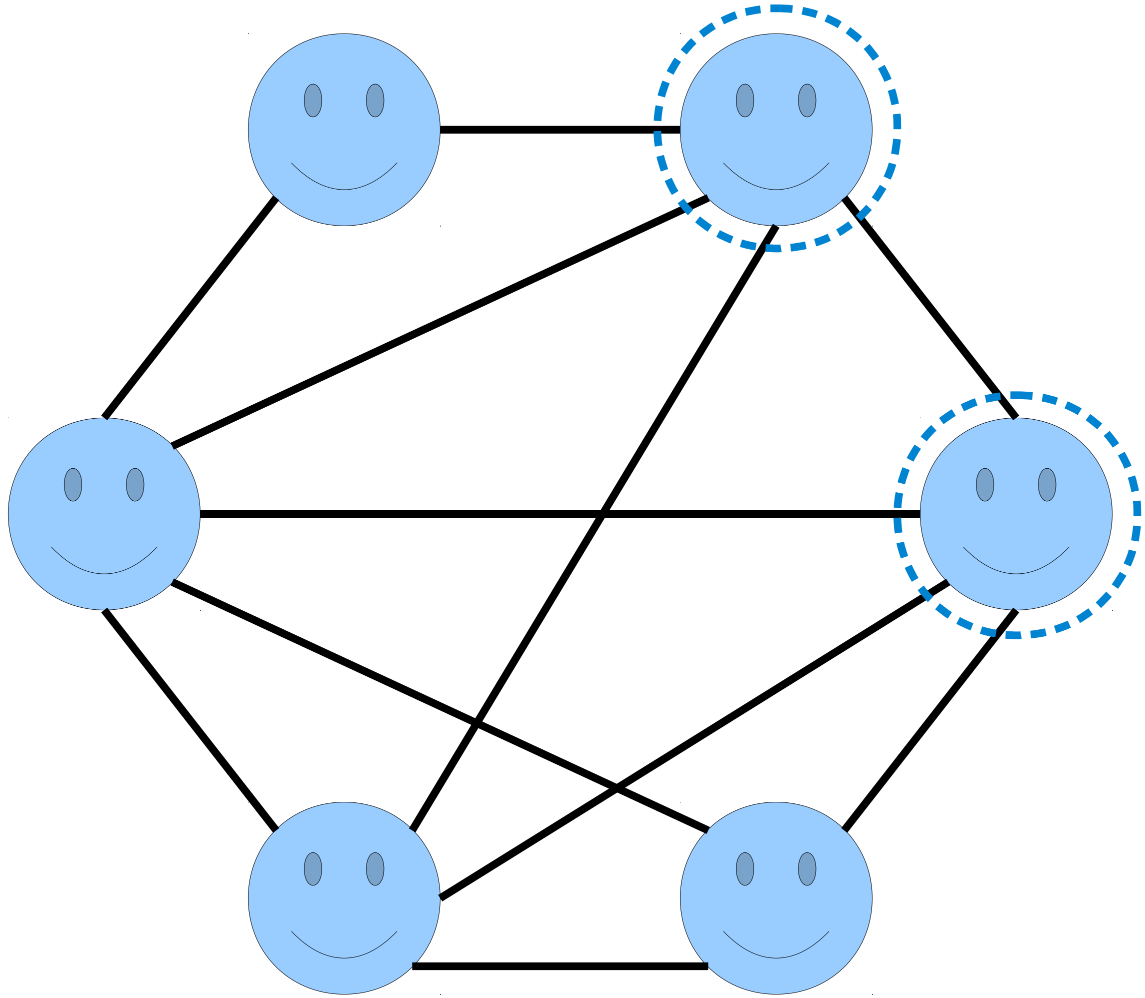
$$\begin{aligned} m &= x_1 + x_2 + \ldots + x_n \\ &\leq 1 + 1 + \ldots + 1 \quad \textit{(n times)} \\ &= n. \end{aligned}$$

This means that $m \leq n$, contradicting that $m > n$. We've reached a contradiction, so our assumption must have been wrong. Therefore, if m objects are distributed into n bins with $m > n$, some bin must contain at least two objects. ■

Pigeonhole Principle Party Tricks

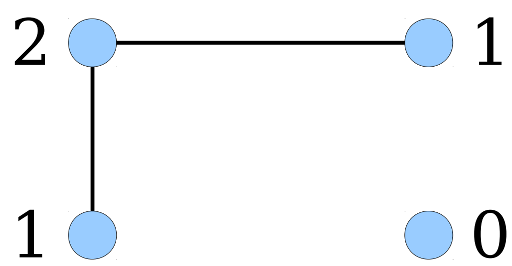


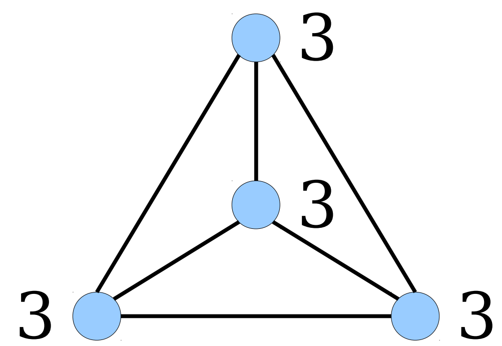




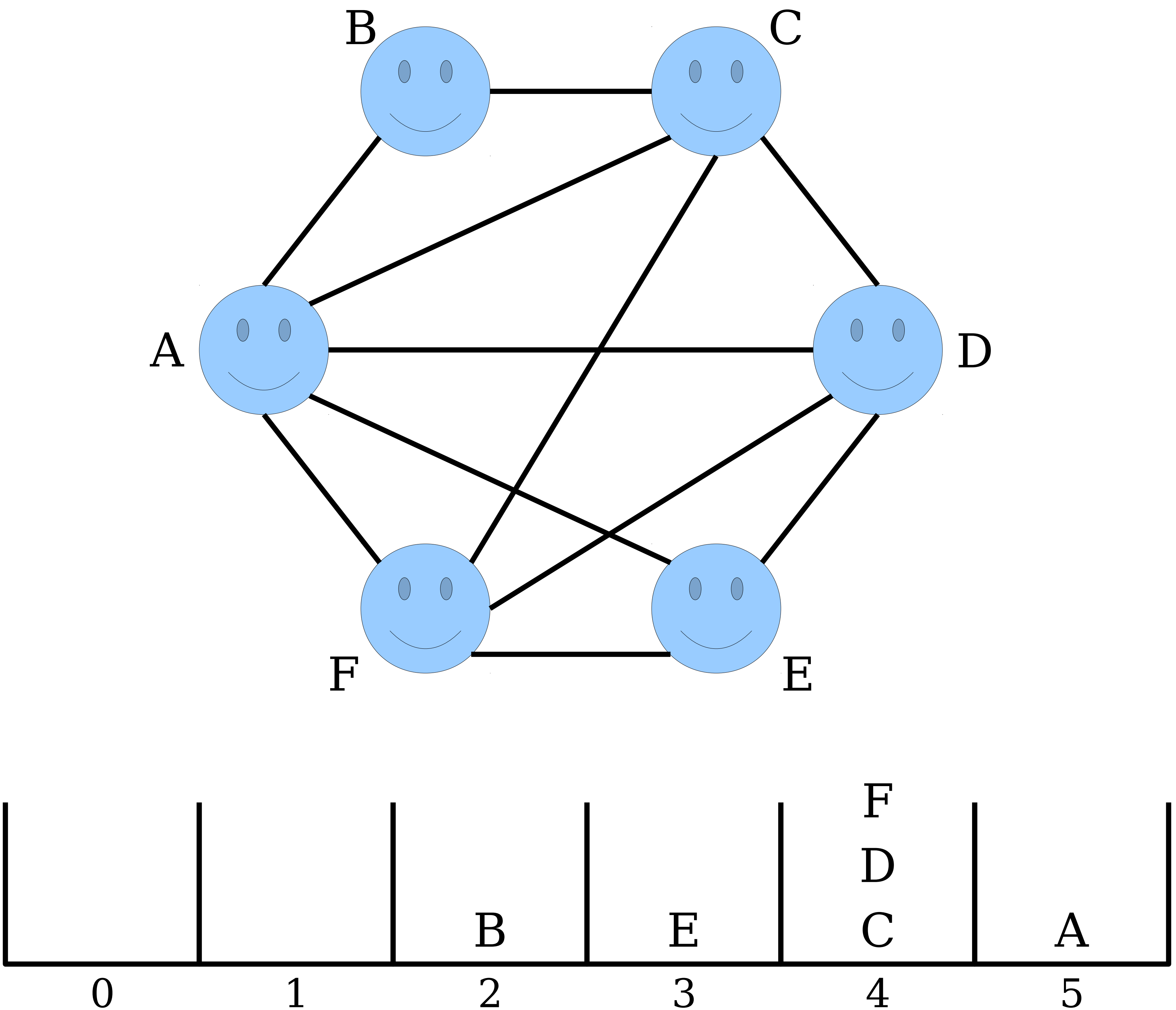
Degrees

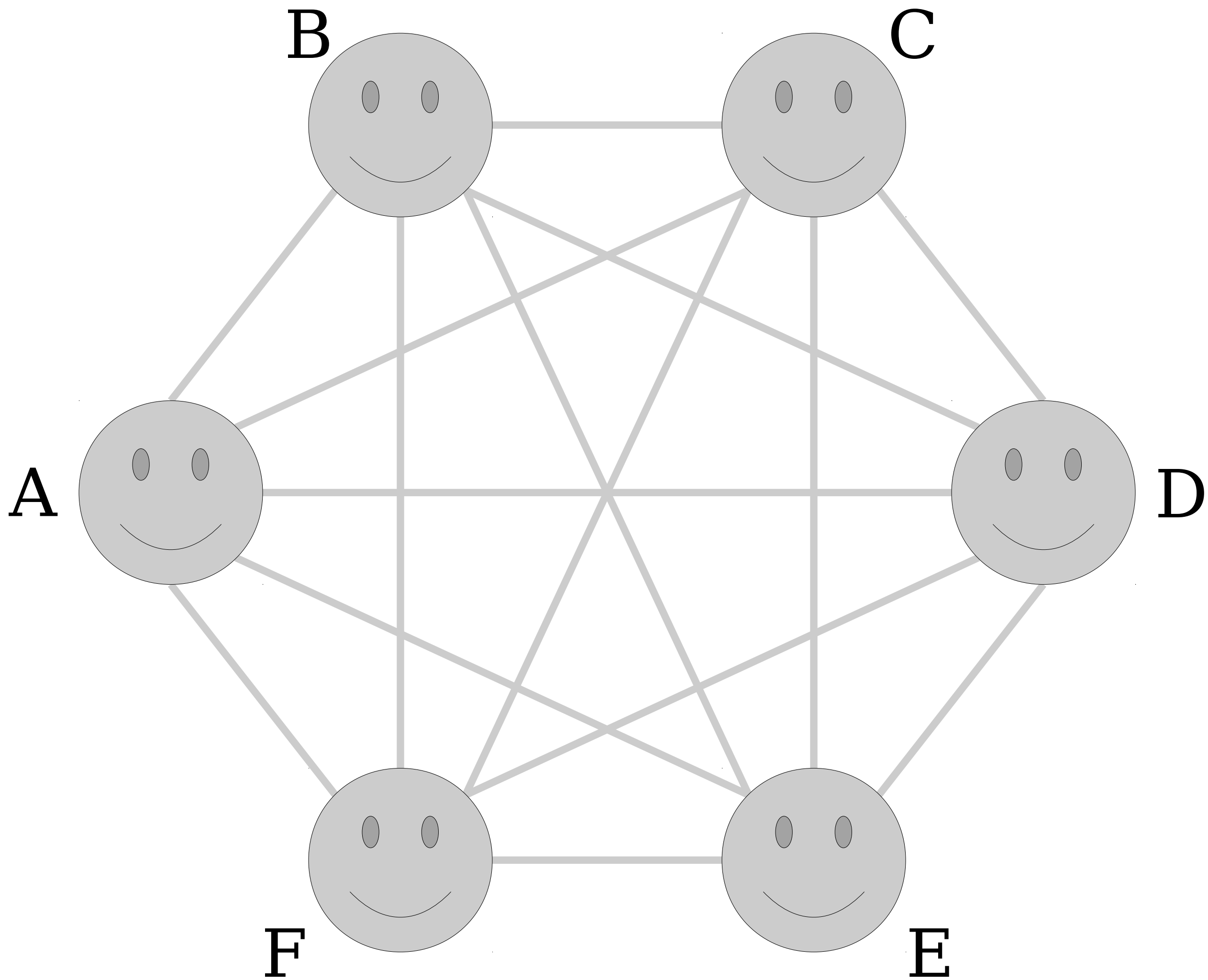
- The ***degree*** of a node *v* in a graph is the number of nodes that *v* is adjacent to.

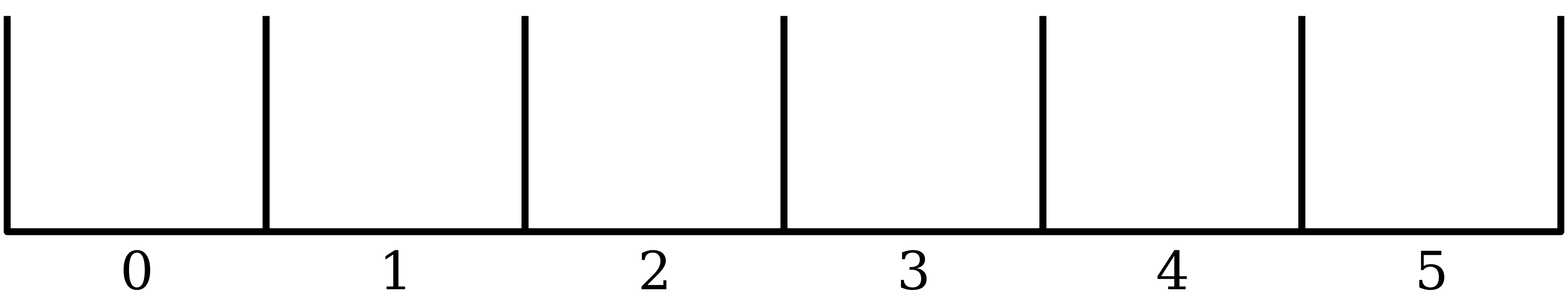


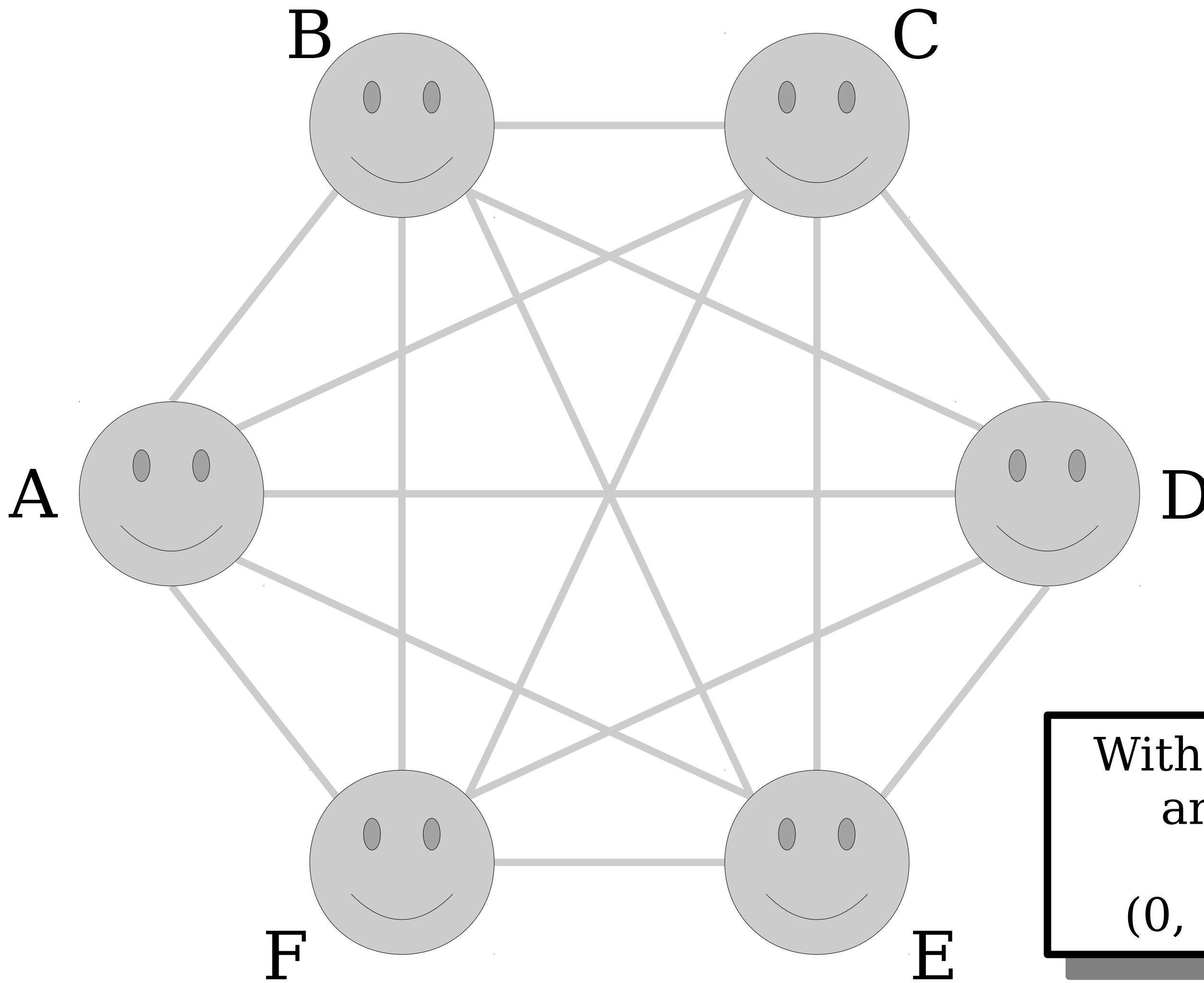


- ***Theorem:*** Every graph with at least two nodes has at least two nodes with the same degree.
    - Equivalently: at any party with at least two people, there are at least two people with the same number of friends at the party.

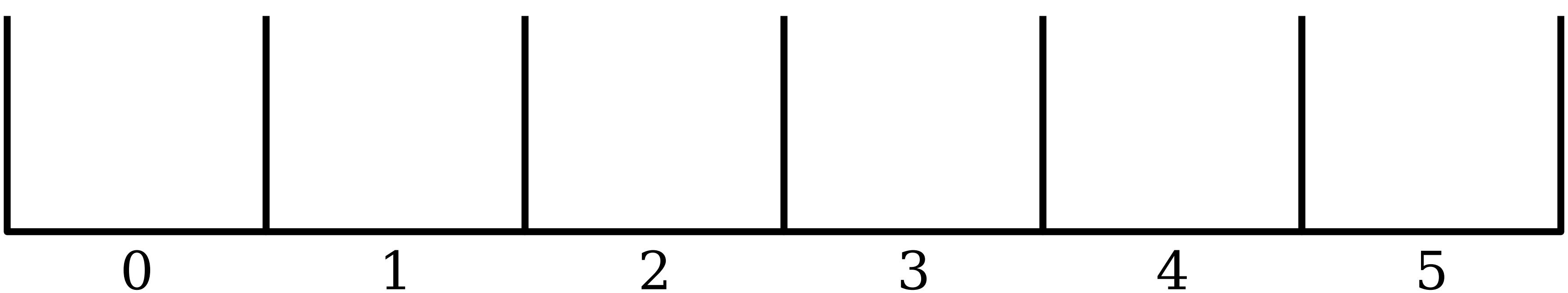

B
C
A
D
F
E
F
D
B
E
C
A
0
1
2
3
4
5

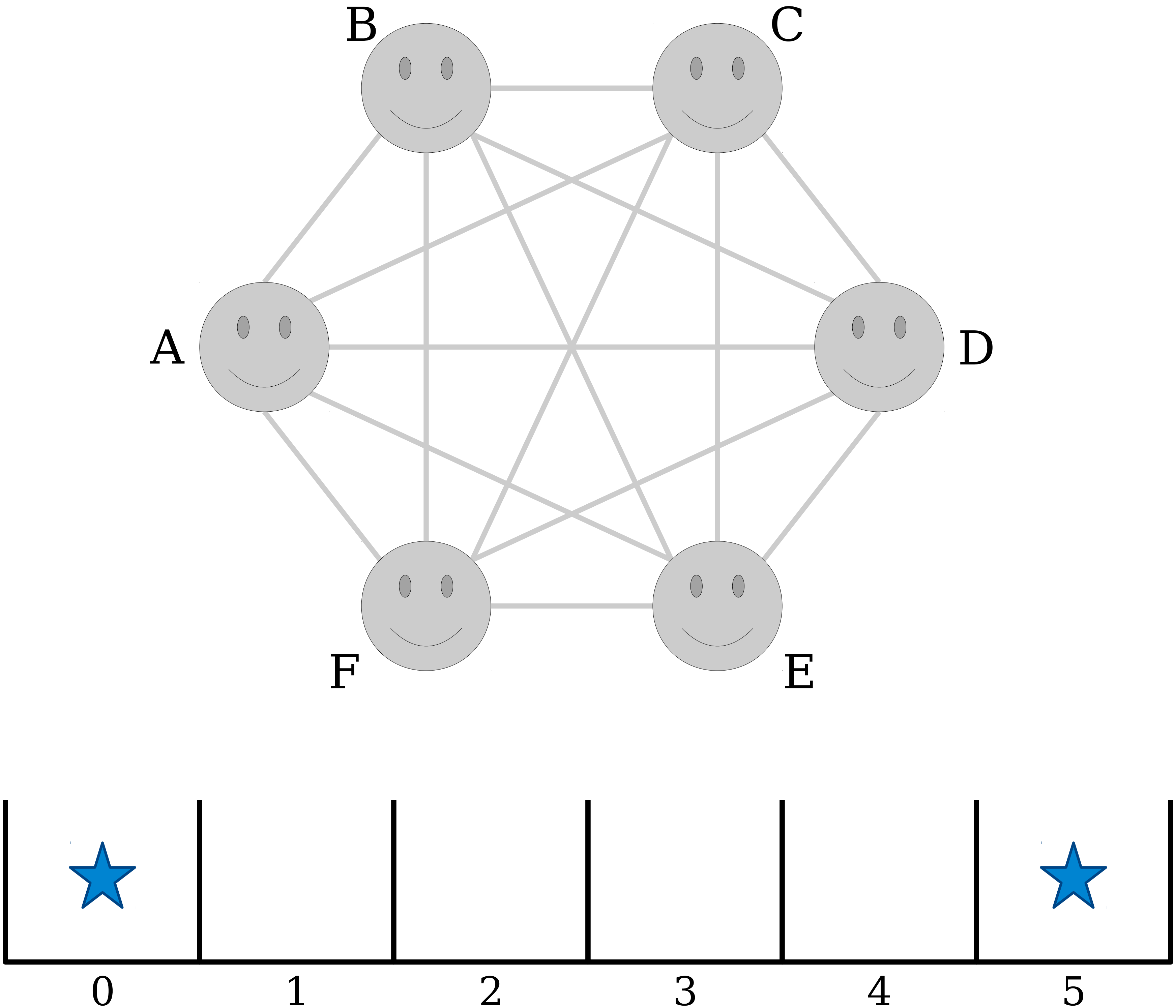

B
C
A
D
F
E

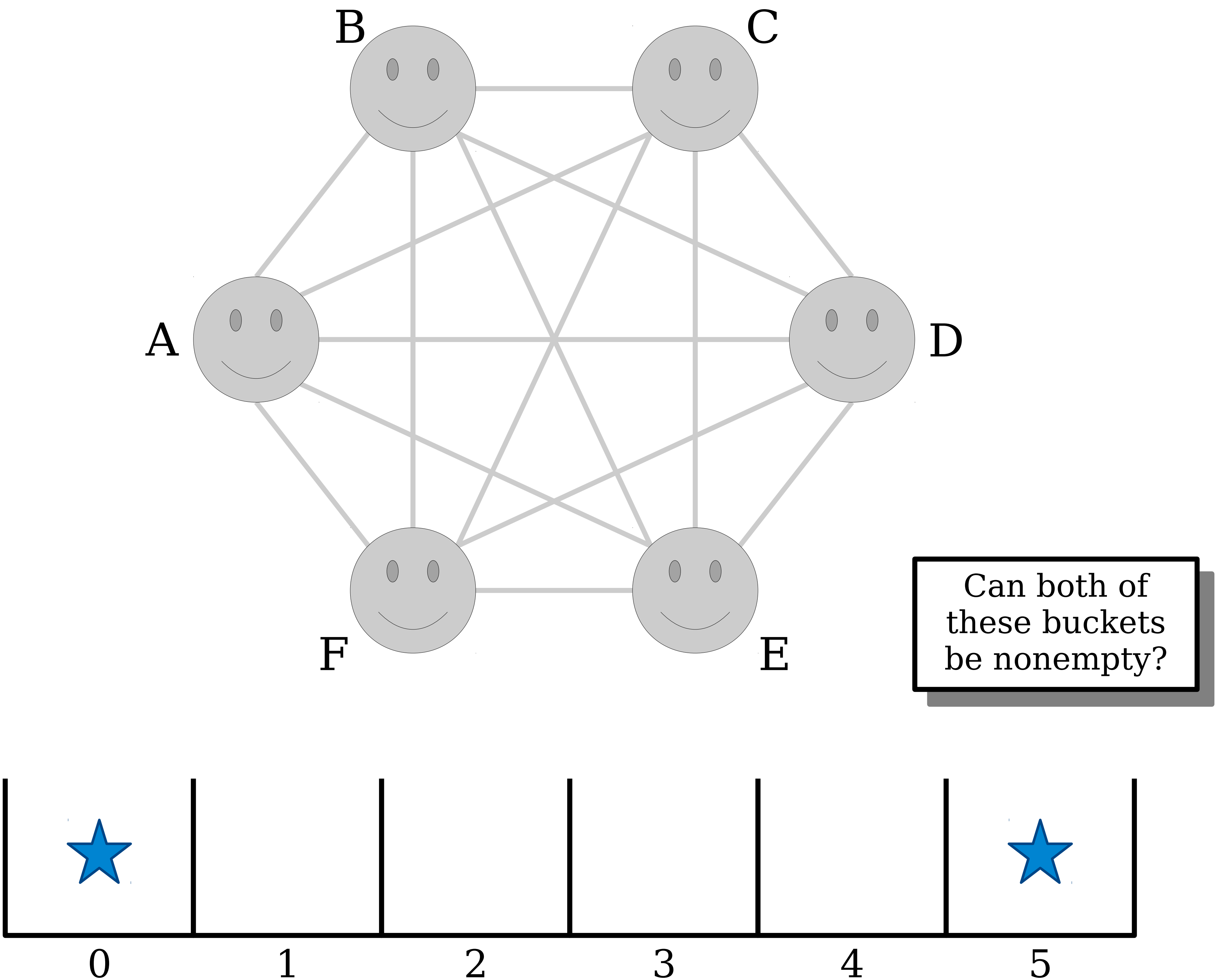

0
1
2
3
4
5

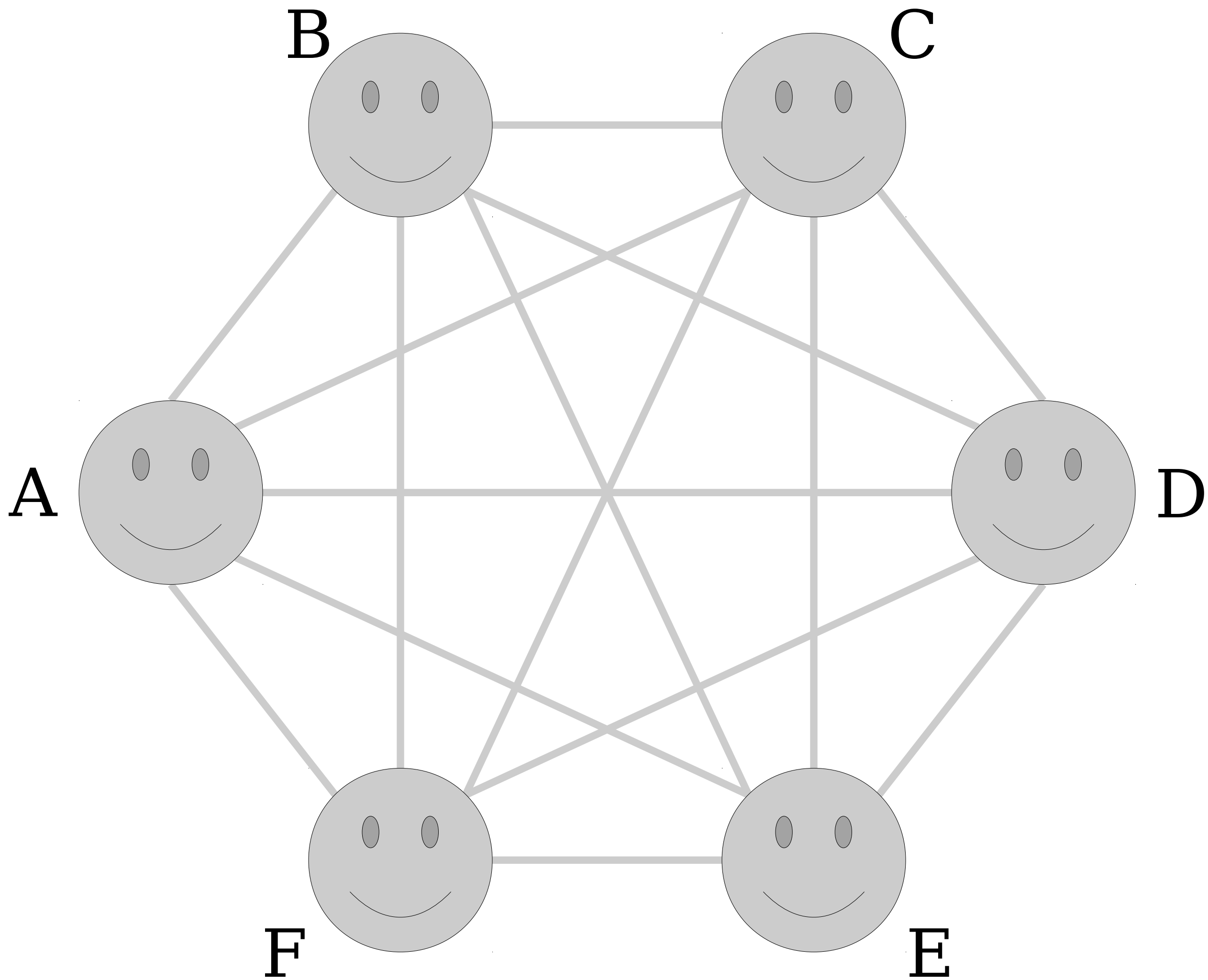


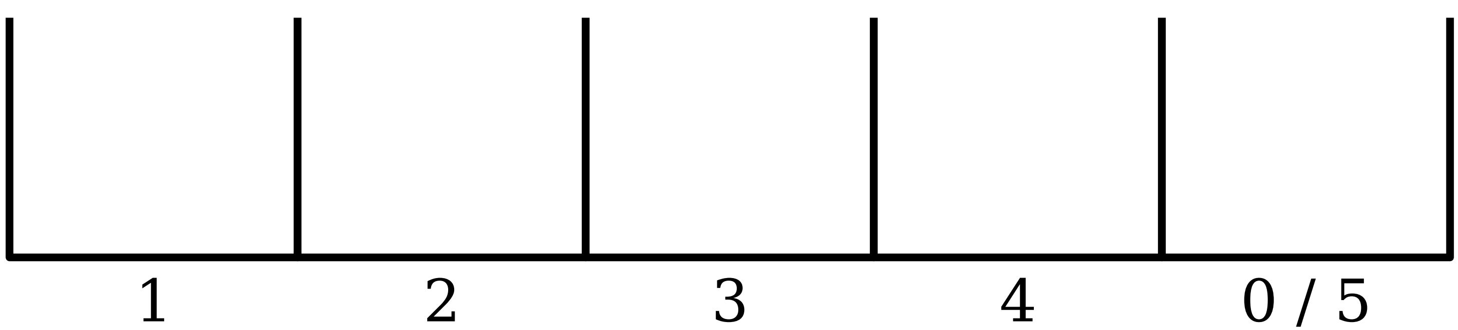
With *n* nodes, there are *n* possible degrees (0, 1, 2, …, *n* – 1)




B
C
A
D
F
E
0
1
2
3
4
5


B
C
A
D
F
E
Can both of
these buckets
be nonempty?
0
1
2
3
4
5


B
C
A
D
F
E


1
2
3
4
0 / 5

Theorem: In any graph with at least two nodes, there are at least two nodes of the same degree.

Theorem: In any graph with at least two nodes, there are at least two nodes of the same degree.

Proof 1:

Theorem: In any graph with at least two nodes, there are at least two nodes of the same degree.

Proof 1: Let G be a graph with $n \geq 2$ nodes.

Theorem: In any graph with at least two nodes, there are at least two nodes of the same degree.

Proof 1: Let G be a graph with $n \geq 2$ nodes. There are n possible choices for the degrees of nodes in G, namely, $0, 1, 2, \ldots,$ and $n - 1$.

Theorem: In any graph with at least two nodes, there are at least two nodes of the same degree.

Proof 1: Let *G* be a graph with $n \geq 2$ nodes. There are *n* possible choices for the degrees of nodes in *G,* namely, 0, 1, 2, …, and $n - 1$.

We claim that G cannot simultaneously have a node *u* of degree 0 and a node *v* of degree $n - 1$:

Theorem: In any graph with at least two nodes, there are at least two nodes of the same degree.

Proof 1: Let *G* be a graph with $n \geq 2$ nodes. There are *n* possible choices for the degrees of nodes in *G*, namely, 0, 1, 2, …, and $n - 1$.

We claim that G cannot simultaneously have a node *u* of degree 0 and a node *v* of degree $n - 1$: if there were such nodes, then node *u* would be adjacent to no other nodes and node *v* would be adjacent to all other nodes, including *u*. (Note that *u* and *v* must be different nodes, since *v* has degree at least 1 and u has degree 0.)

Theorem: In any graph with at least two nodes, there are at least two nodes of the same degree.

Proof 1: Let G be a graph with $n \geq 2$ nodes. There are n possible choices for the degrees of nodes in G, namely, $0, 1, 2, \ldots,$ and $n - 1$.

We claim that G cannot simultaneously have a node u of degree 0 and a node v of degree $n - 1$: if there were such nodes, then node u would be adjacent to no other nodes and node v would be adjacent to all other nodes, including u. (Note that u and v must be different nodes, since v has degree at least 1 and u has degree 0.)

We therefore see that the possible options for degrees of nodes in G are either drawn from $0, 1, \ldots, n - 2$ or from $1, 2, \ldots, n - 1$.

Theorem: In any graph with at least two nodes, there are at least two nodes of the same degree.

Proof 1: Let G be a graph with $n \geq 2$ nodes. There are n possible choices for the degrees of nodes in G, namely, 0, 1, 2, …, and $n - 1$.

We claim that G cannot simultaneously have a node u of degree 0 and a node v of degree $n - 1$: if there were such nodes, then node u would be adjacent to no other nodes and node v would be adjacent to all other nodes, including u. (Note that u and v must be different nodes, since v has degree at least 1 and u has degree 0.)

We therefore see that the possible options for degrees of nodes in G are either drawn from 0, 1, …, $n - 2$ or from 1, 2, …, $n - 1$. In either case, there are n nodes and $n - 1$ possible degrees, so by the pigeonhole principle two nodes in G must have the same degree.

Theorem: In any graph with at least two nodes, there are at least two nodes of the same degree.

Proof 1: Let G be a graph with $n \geq 2$ nodes. There are n possible choices for the degrees of nodes in G, namely, $0, 1, 2, \ldots,$ and $n - 1$.

We claim that G cannot simultaneously have a node u of degree 0 and a node v of degree $n - 1$: if there were such nodes, then node u would be adjacent to no other nodes and node v would be adjacent to all other nodes, including u. (Note that u and v must be different nodes, since v has degree at least 1 and u has degree 0.)

We therefore see that the possible options for degrees of nodes in G are either drawn from $0, 1, \ldots, n - 2$ or from $1, 2, \ldots, n - 1$. In either case, there are n nodes and $n - 1$ possible degrees, so by the pigeonhole principle two nodes in G must have the same degree. ■

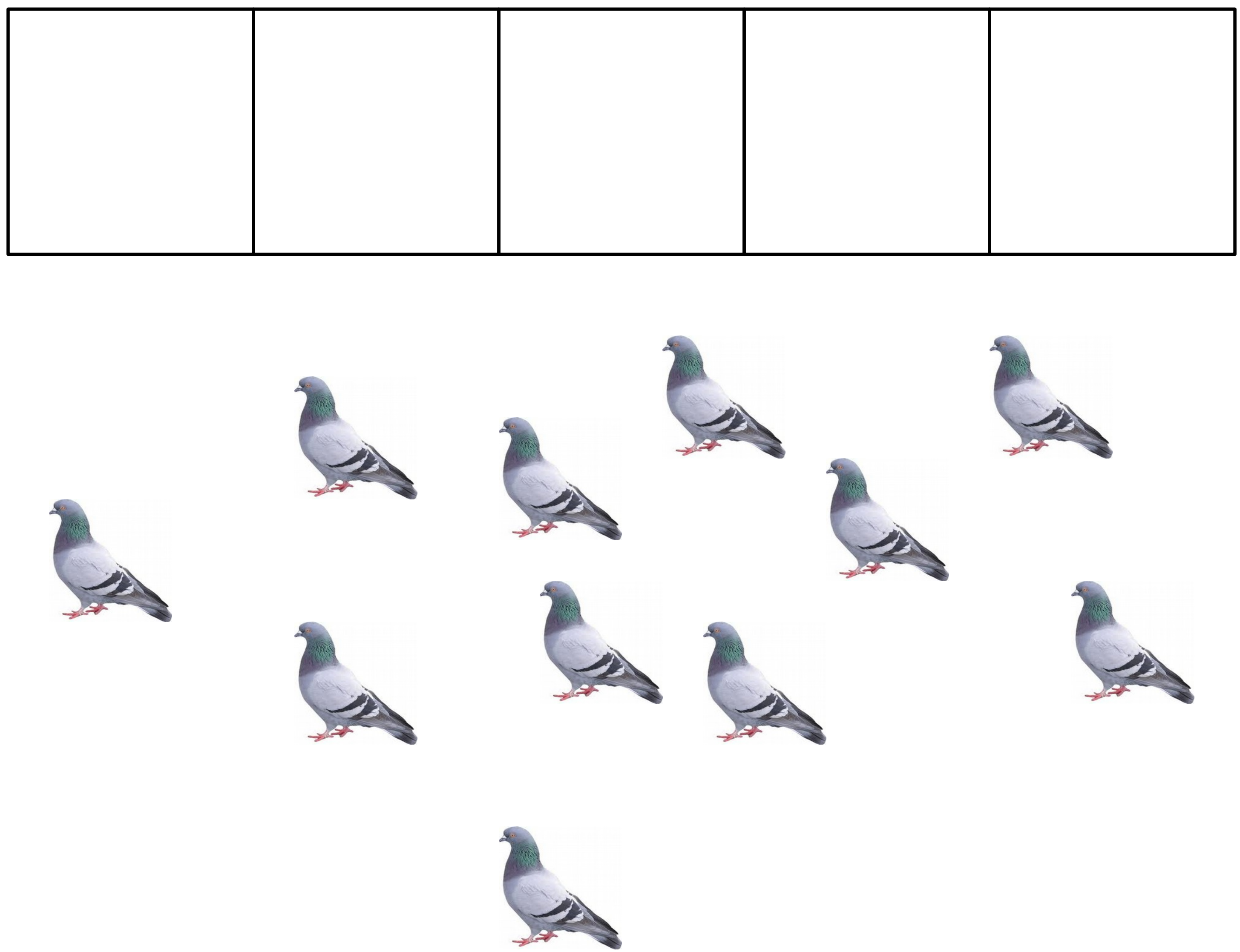
Theorem: In any graph with at least two nodes, there are at least two nodes of the same degree.

Proof 2: Assume for the sake of contradiction that there is a graph G with $n \geq 2$ nodes where no two nodes have the same degree. There are n possible choices for the degrees of nodes in G, namely 0, 1, 2, …, $n - 1$, so this means that G must have exactly one node of each degree. However, this means that G has a node of degree 0 and a node of degree $n - 1$. (These can't be the same node, since $n \geq 2$.) This first node is adjacent to no other nodes, but this second node is adjacent to every other node, which is impossible.

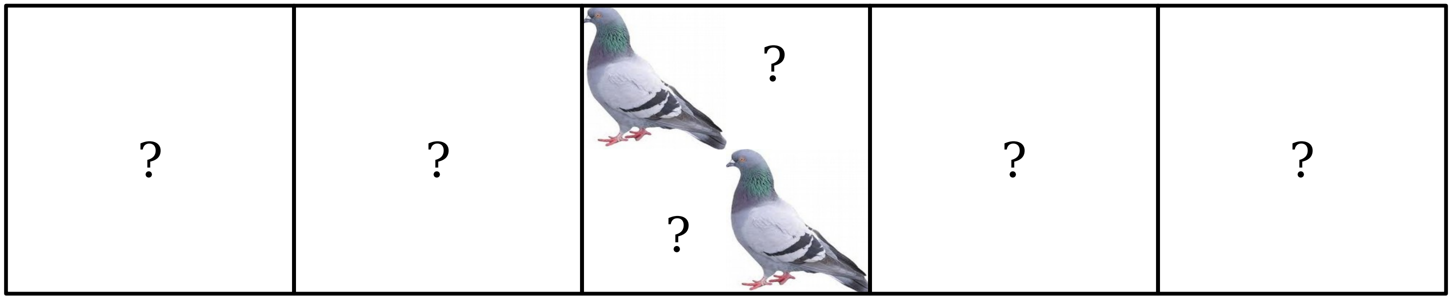
We have reached a contradiction, so our assumption must have been wrong. Thus if G is a graph with at least two nodes, G must have at least two nodes of the same degree. ■

The Generalized Pigeonhole Principle

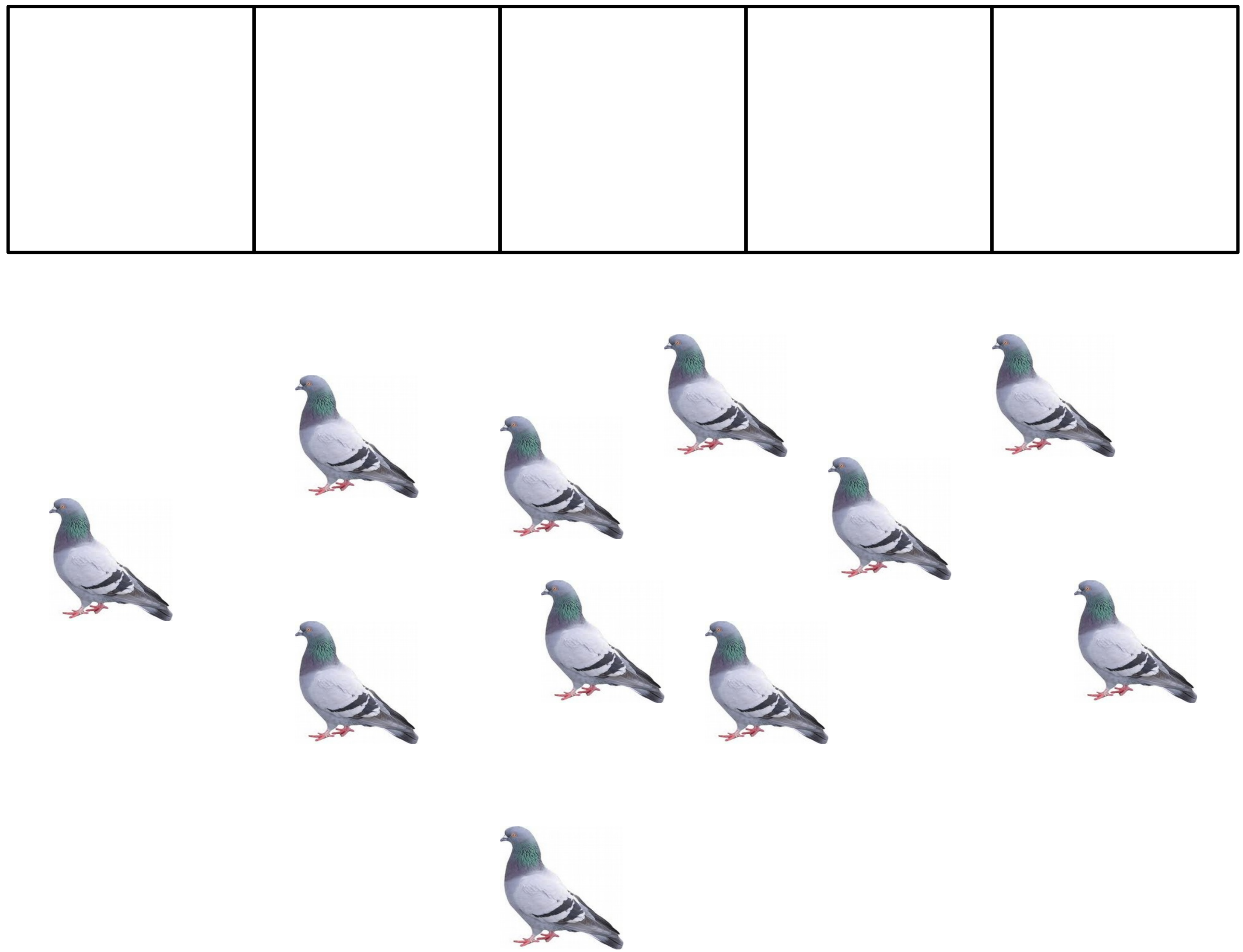
The Pigeonhole Principle



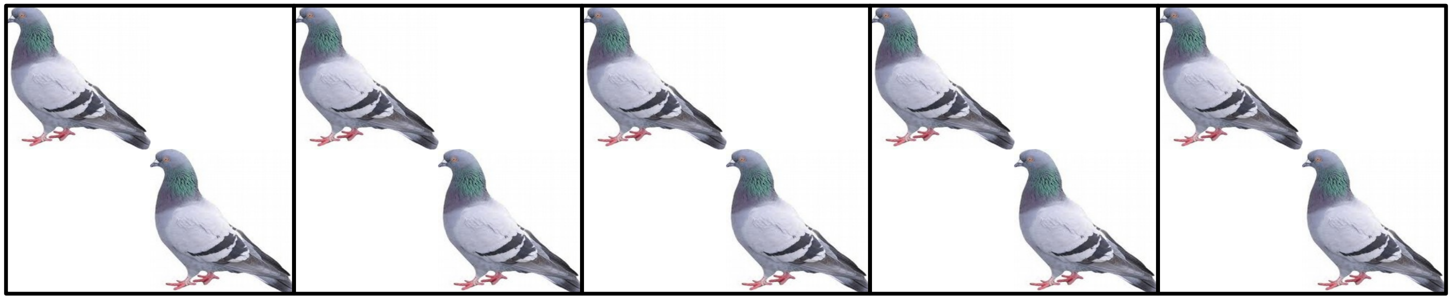
The Pigeonhole Principle



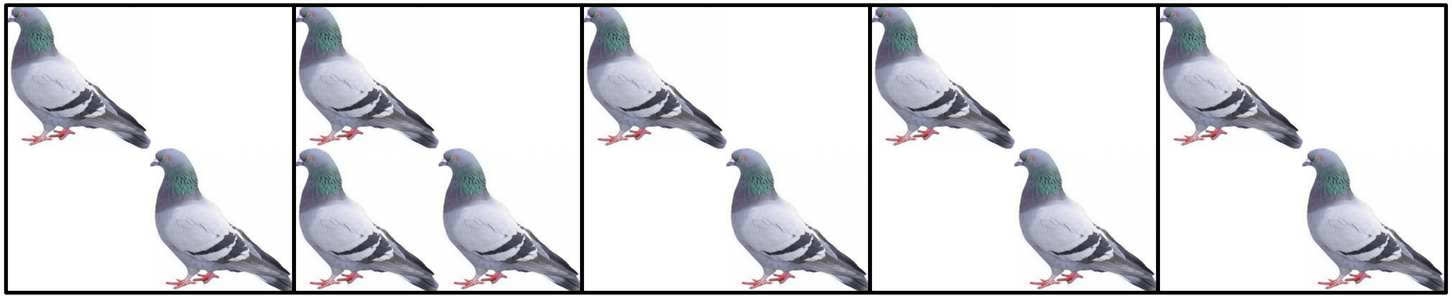
The Pigeonhole Principle



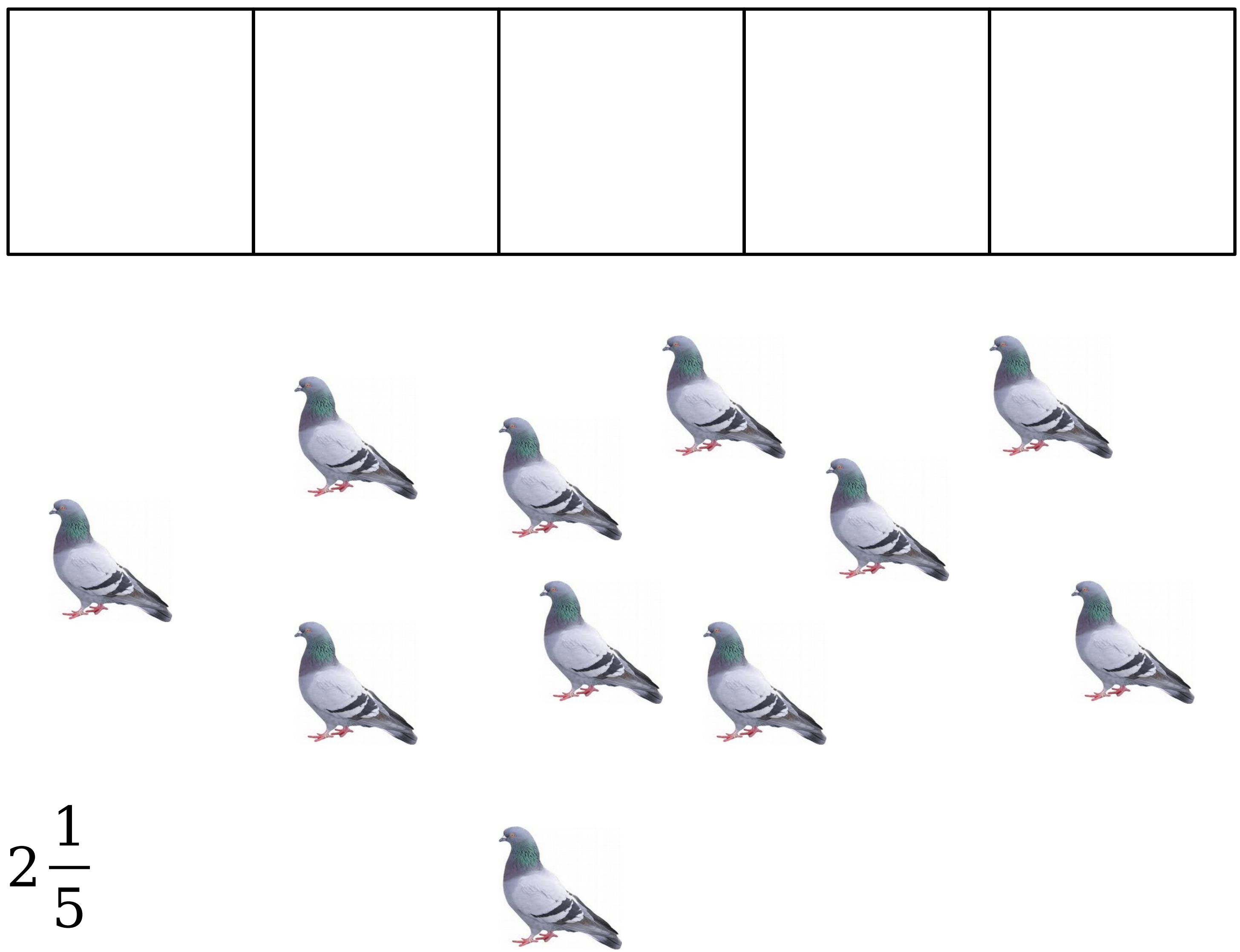
The Pigeonhole Principle



The Pigeonhole Principle



The Pigeonhole Principle

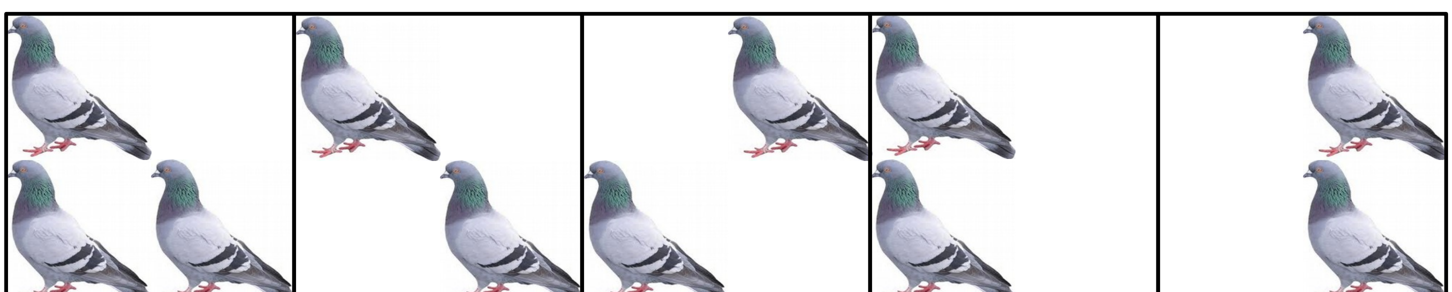


$$\frac{11}{5} = 2\frac{1}{5}$$

A More General Version

- The ***generalized pigeonhole principle*** says that if you distribute *m* objects into *n* bins, then
  - some bin will have at least $\lceil m/n \rceil$ objects in it, and
  - some bin will have at most $\lfloor m/n \rfloor$ objects in it.

$\lceil m/n \rceil$ means "m/n, rounded up."
$\lfloor m/n \rfloor$ means "m/n, rounded down."



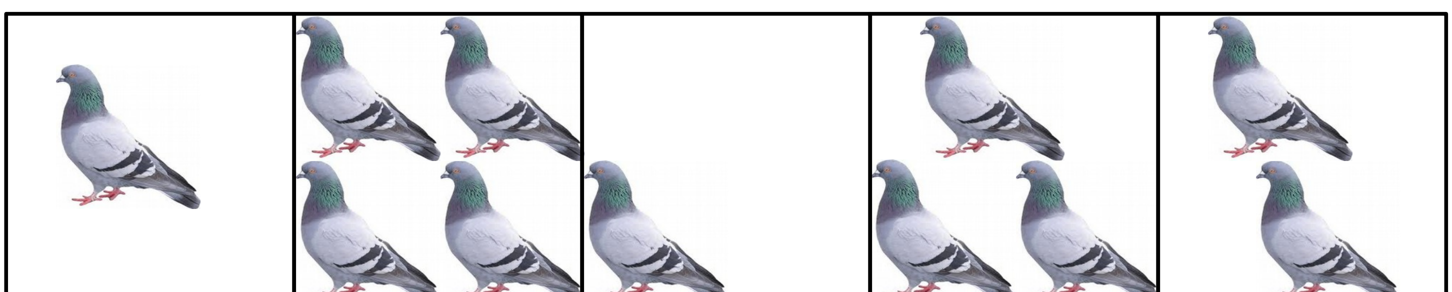
$m = 11$
$n = 5$

$\lceil m / n \rceil = 3$
$\lfloor m / n \rfloor = 2$

A More General Version

- The ***generalized pigeonhole principle*** says that if you distribute m objects into n bins, then
  - some bin will have at least $\lceil m/n \rceil$ objects in it, and
  - some bin will have at most $\lfloor m/n \rfloor$ objects in it.

$\lceil m/n \rceil$ means "m/n, rounded up."
$\lfloor m/n \rfloor$ means "m/n, rounded down."



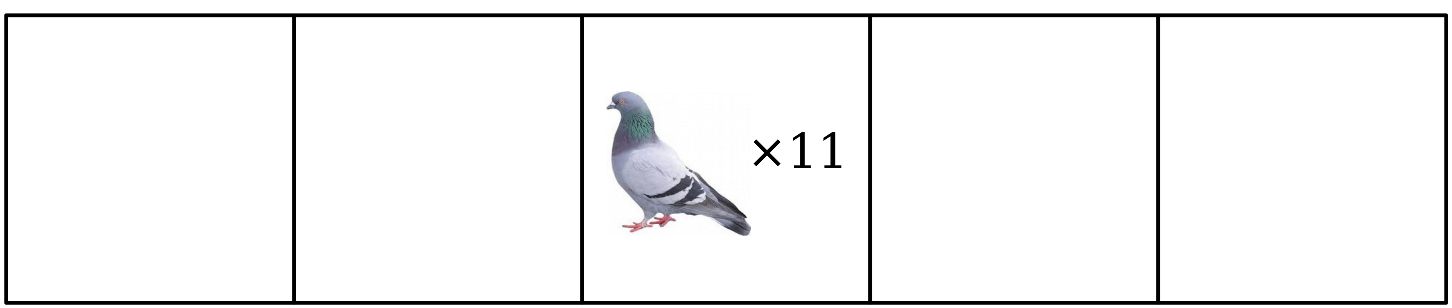
$m = 11$
$n = 5$

$\lceil m / n \rceil = 3$
$\lfloor m / n \rfloor = 2$

A More General Version

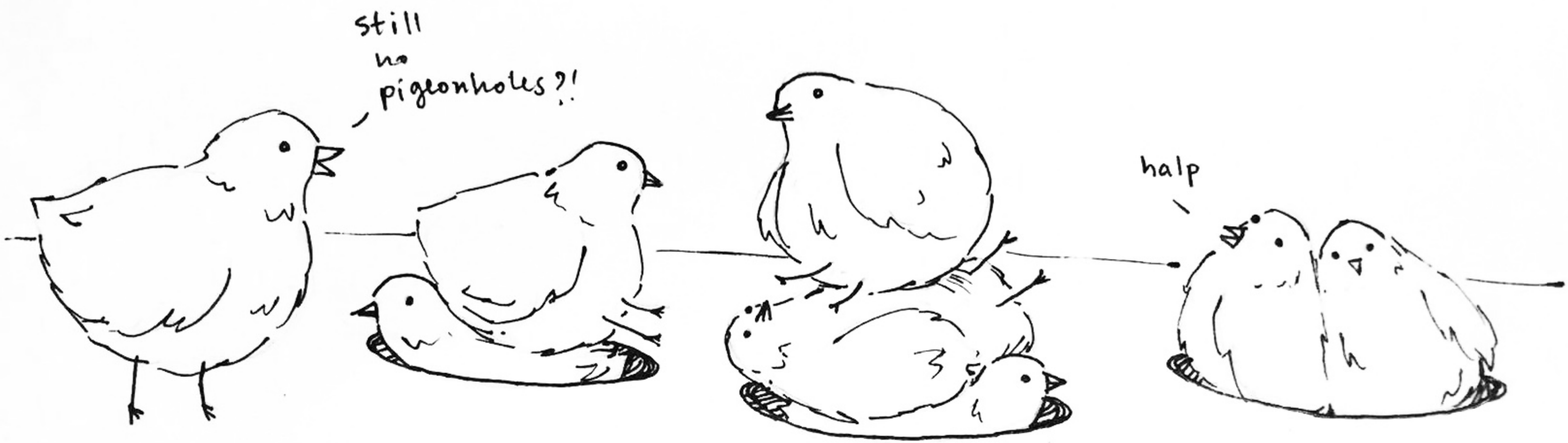
- The ***generalized pigeonhole principle*** says that if you distribute *m* objects into *n* bins, then
  - some bin will have at least $\lceil m/n \rceil$ objects in it, and
  - some bin will have at most $\lfloor m/n \rfloor$ objects in it.

$\lceil m/n \rceil$ means "m/n, rounded up."
$\lfloor m/n \rfloor$ means "m/n, rounded down."



$m = 11$
$n = 5$

$\lceil m / n \rceil = 3$
$\lfloor m / n \rfloor = 2$



$m = 8, n = 3$

Theorem: If m objects are distributed into $n > 0$ bins, then some bin will contain at least $\lceil m/n \rceil$ objects.

Proof: We will prove that if m objects are distributed into n bins, then some bin contains at least m/n objects. Since the number of objects in each bin is an integer, this will prove that some bin must contain at least $\lceil m/n \rceil$ objects.

To do this, we proceed by contradiction. Suppose that, for some m and n, there is a way to distribute m objects into n bins such that each bin contains fewer than m/n objects.

Number the bins 1, 2, 3, …, n and let x_i denote the number of objects in bin i. Since there are m objects in total, we know that

$$m = x_1 + x_2 + \ldots + x_n.$$

Since each bin contains fewer than m/n objects, we see that $x_i < m/n$ for each i. Therefore, we have that

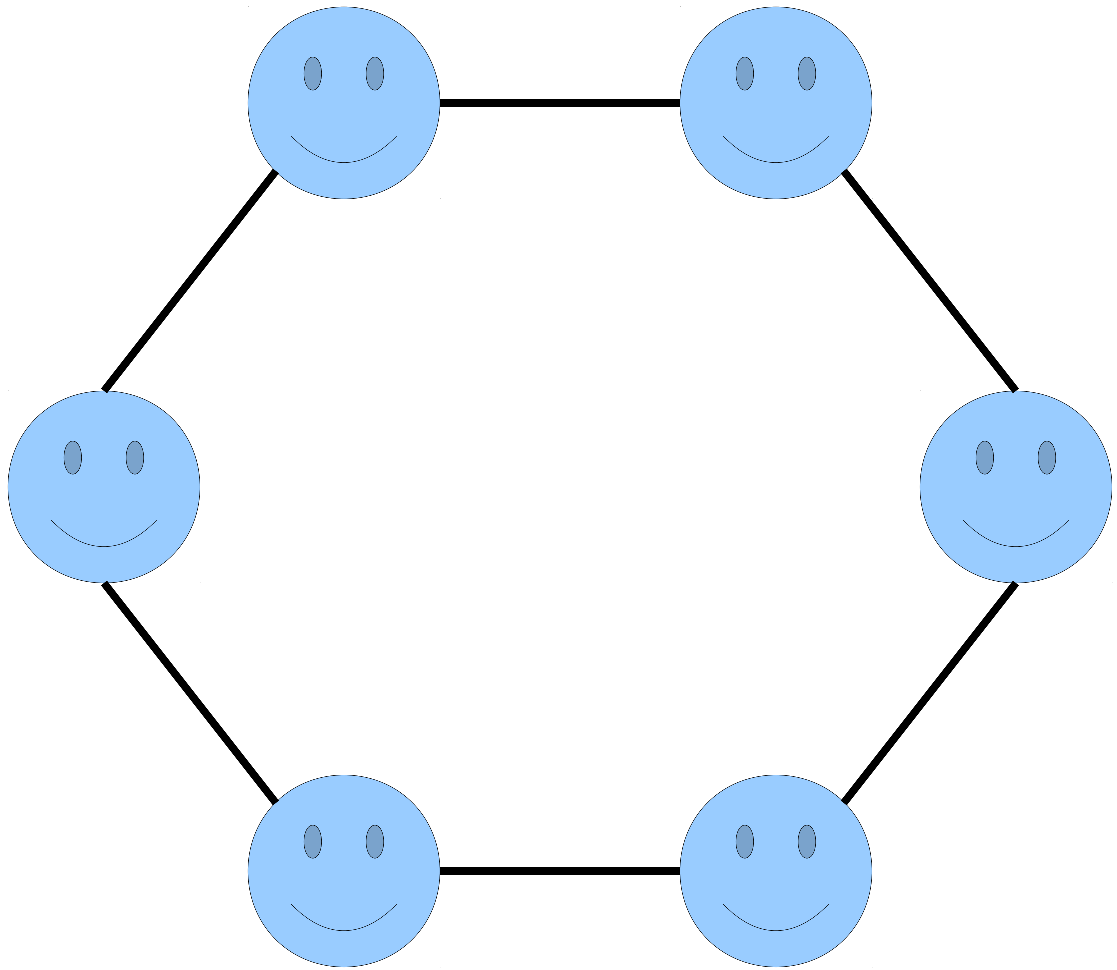
$$\begin{aligned} m &= x_1 + x_2 + \ldots + x_n \\ &< m/n + m/n + \ldots + m/n \quad (n \text{ times}) \\ &= m. \end{aligned}$$

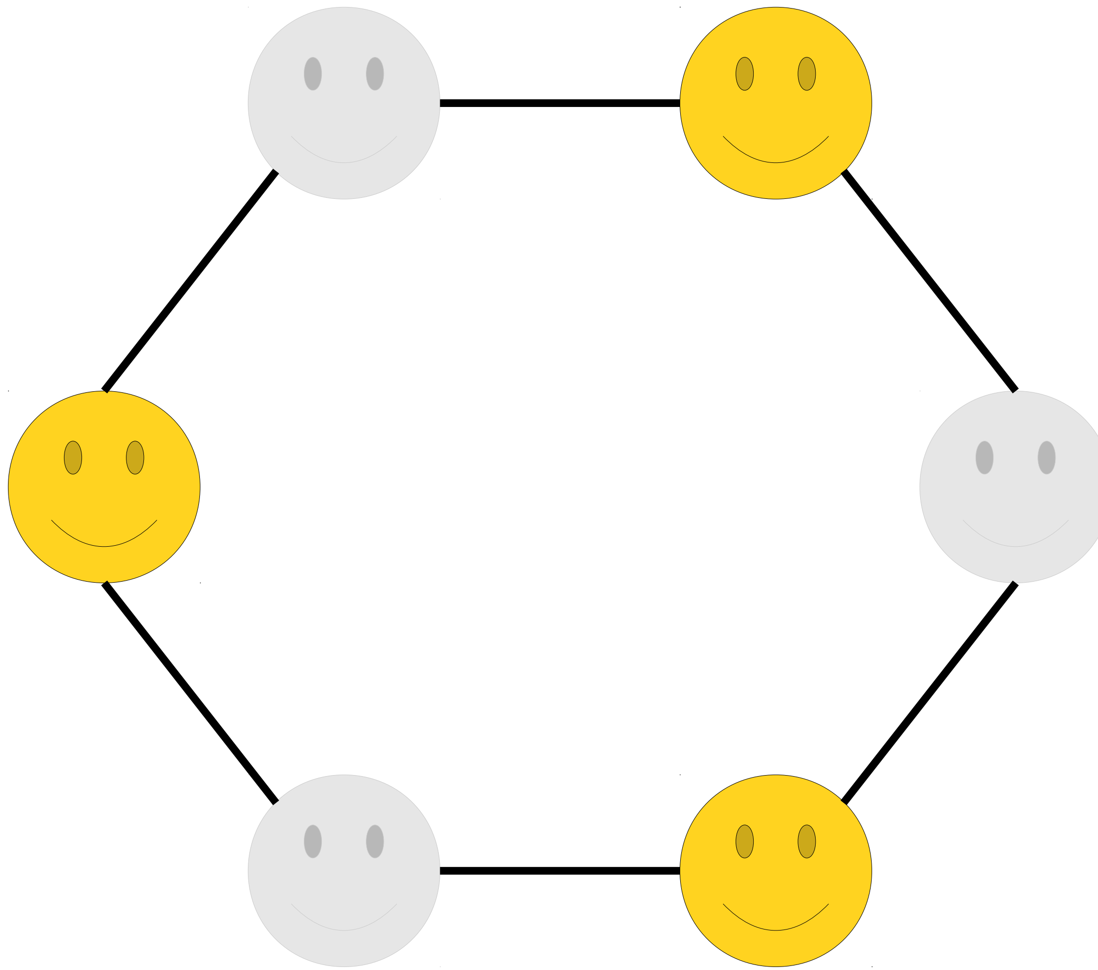
But this means that $m < m$, which is impossible. We have reached a contradiction, so our initial assumption must have been wrong. Therefore, if m objects are distributed into n bins, some bin must contain at least $\lceil m/n \rceil$ objects. ■

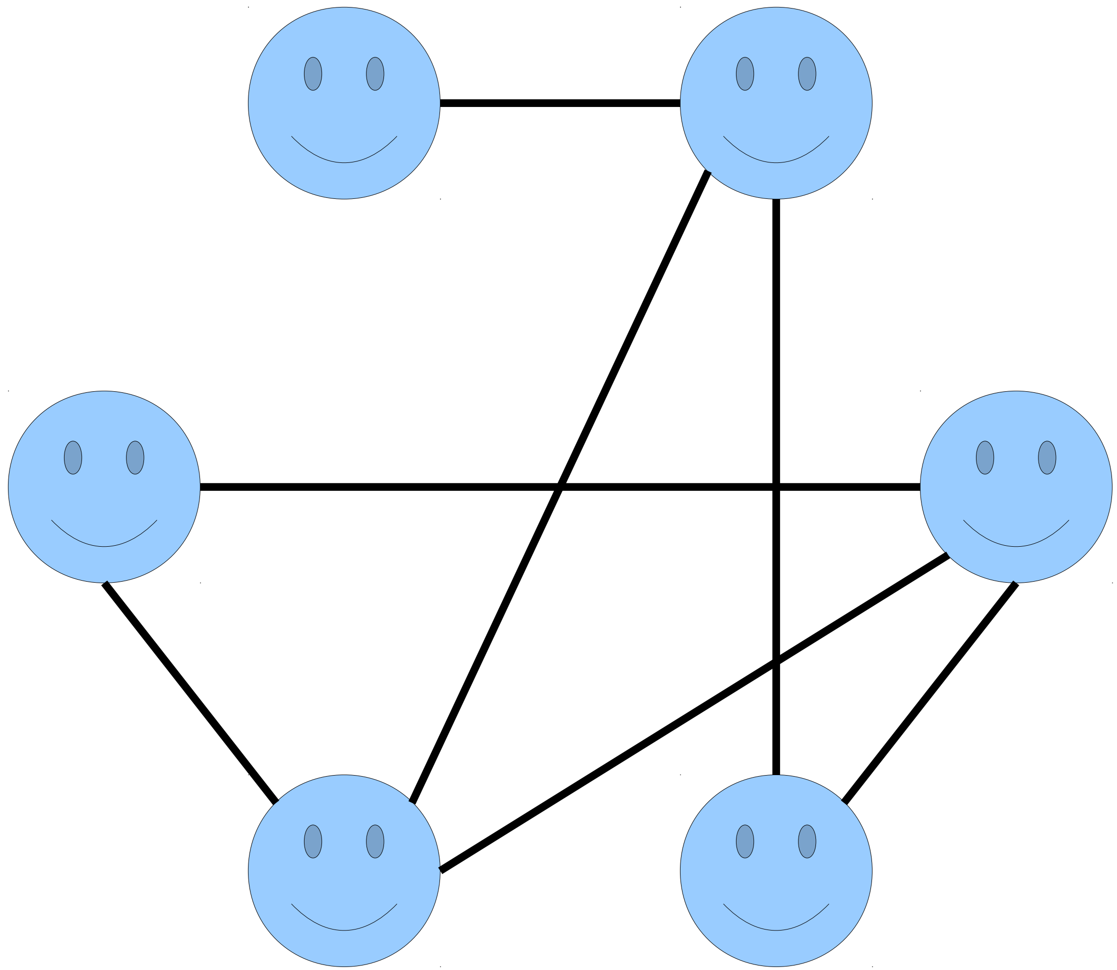
An Application: Friends and Strangers

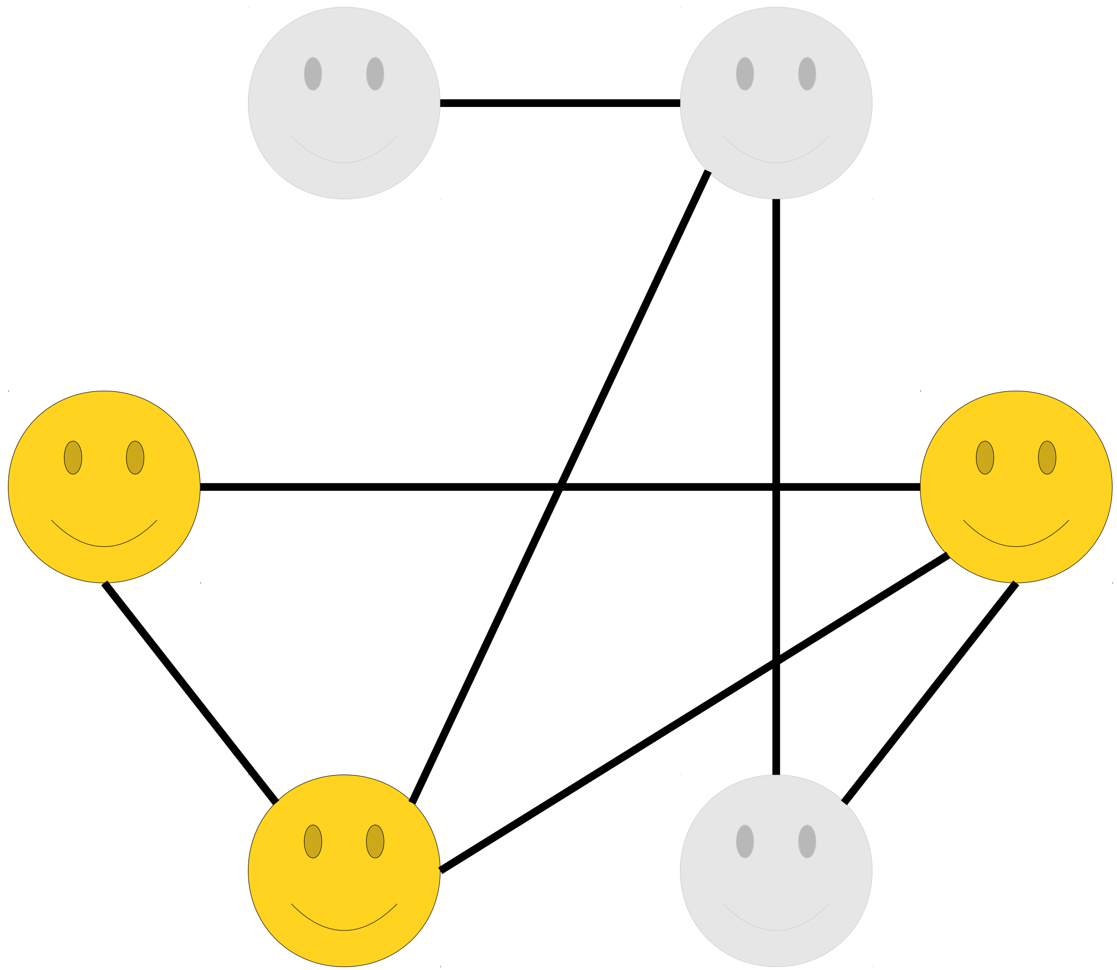
Friends and Strangers

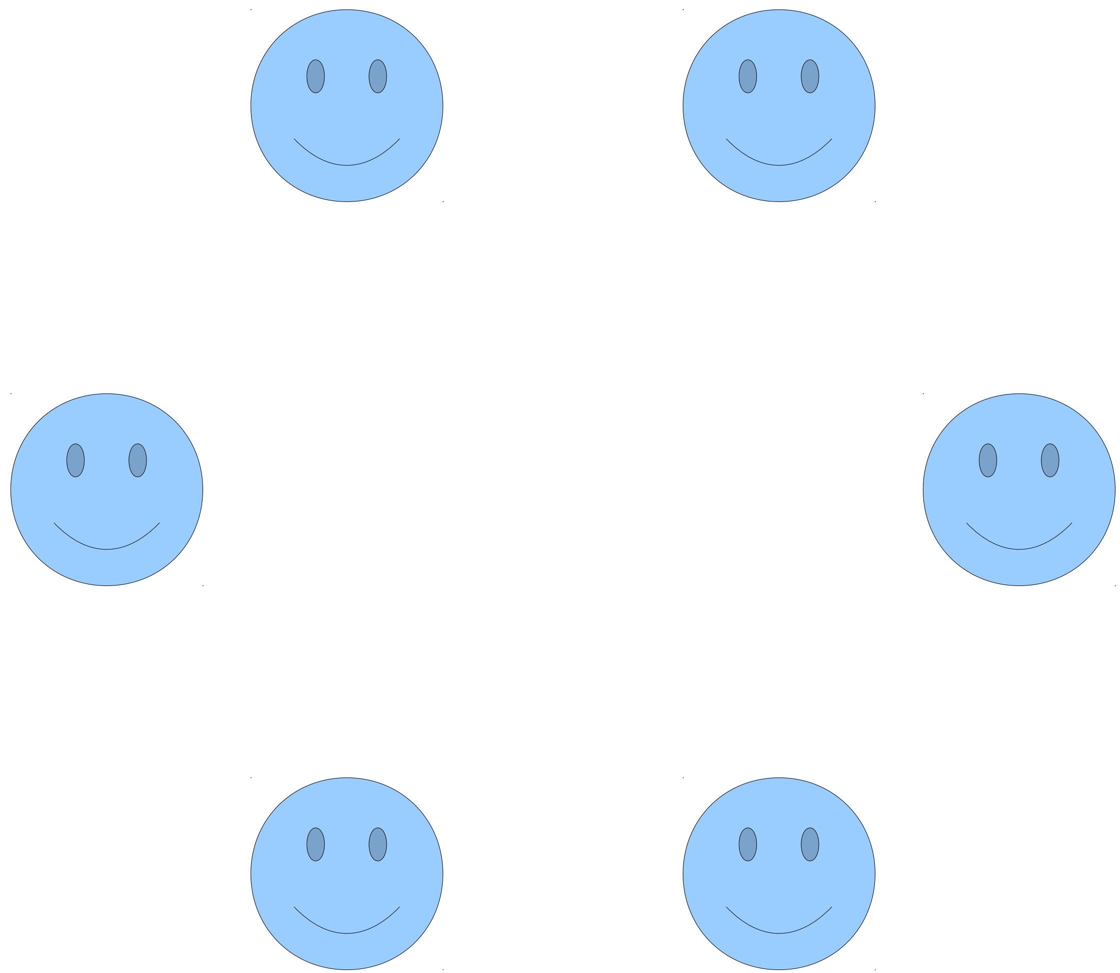
- Suppose you have a party of six people. Each pair of people are either friends (they know each other) or strangers (they do not).
- ***Theorem:*** Any such party must have a group of three mutual friends (three people who all know one another) or three mutual strangers (three people, none of whom know any of the others).

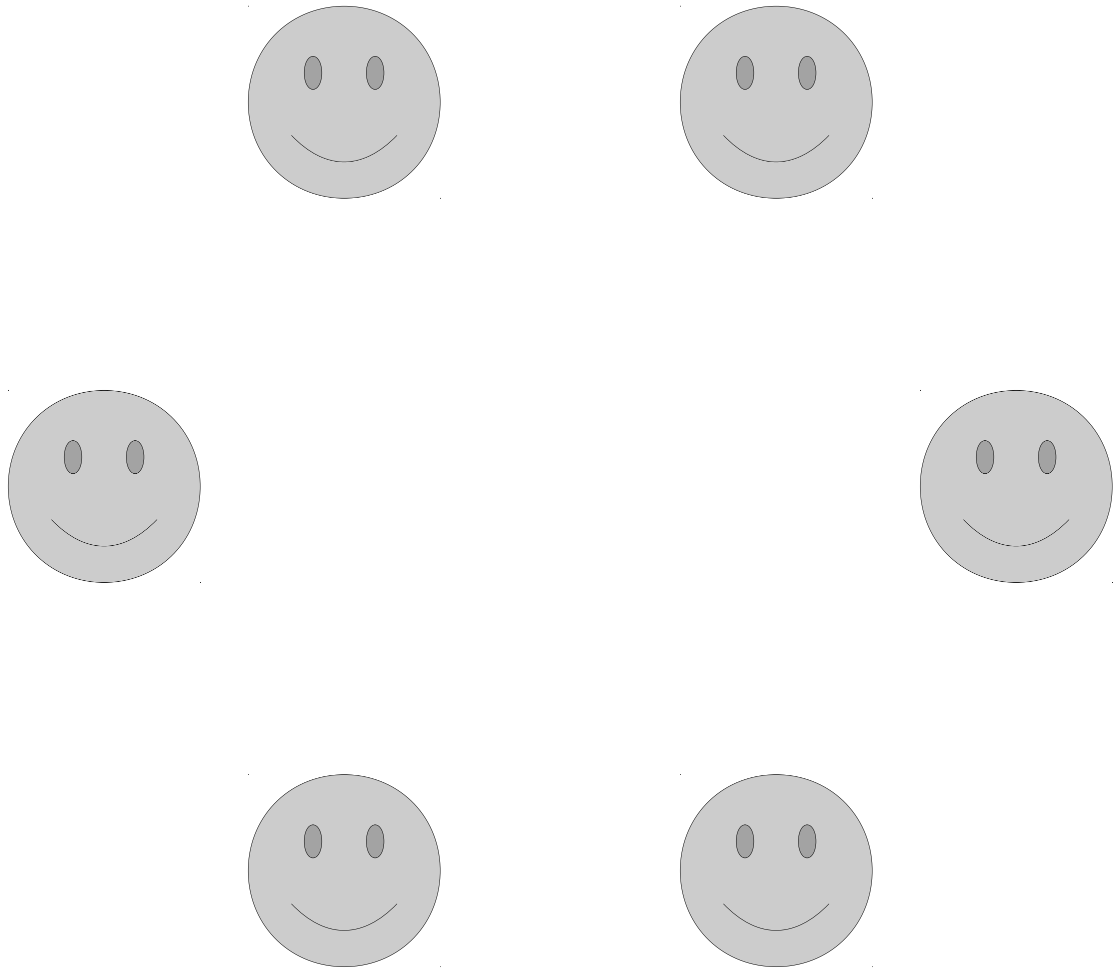


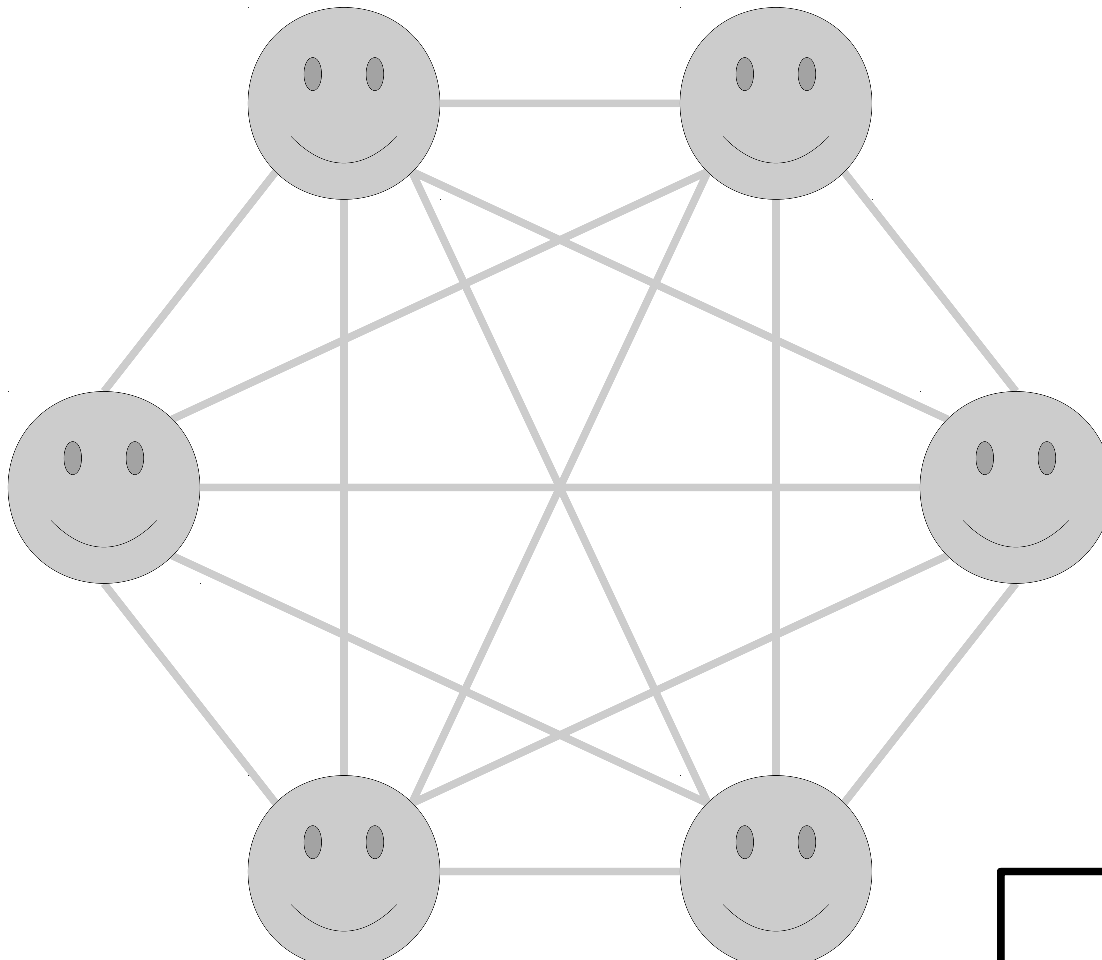




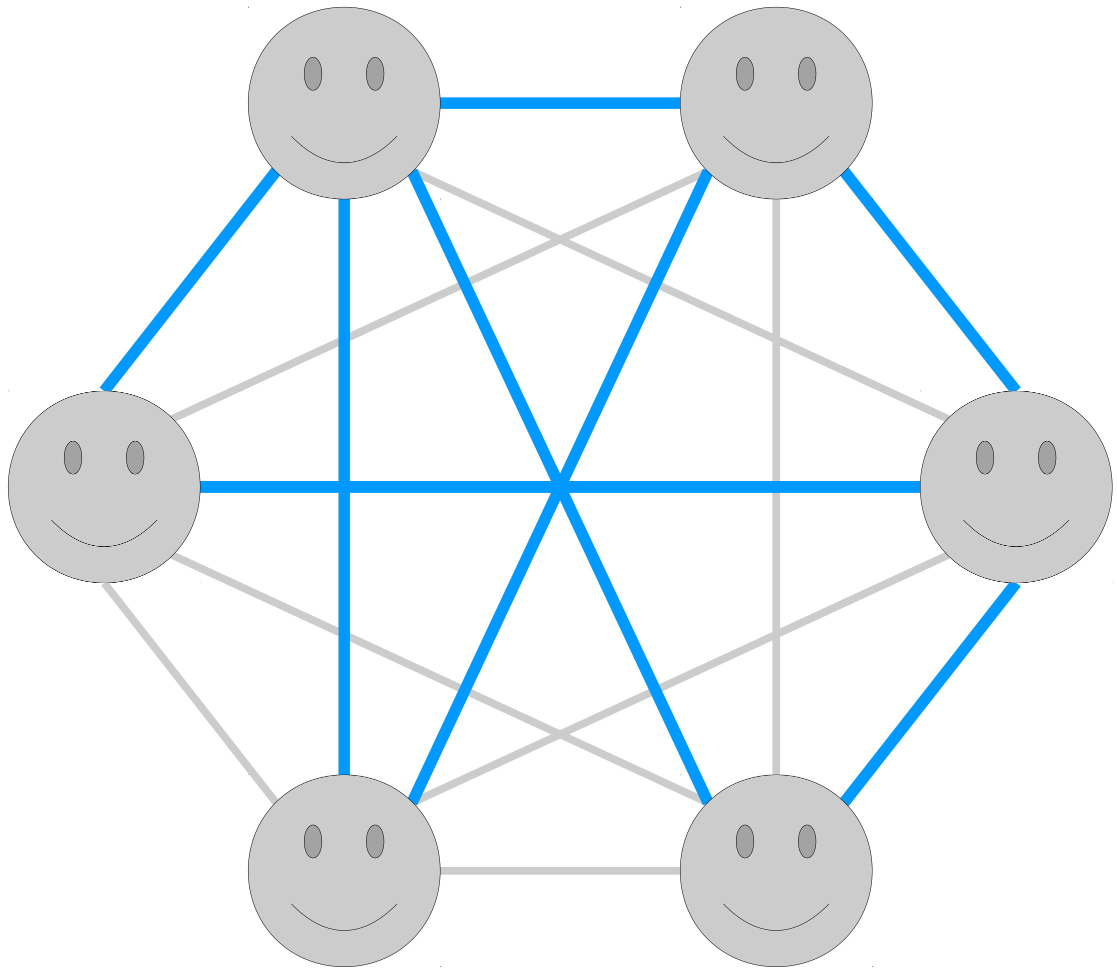


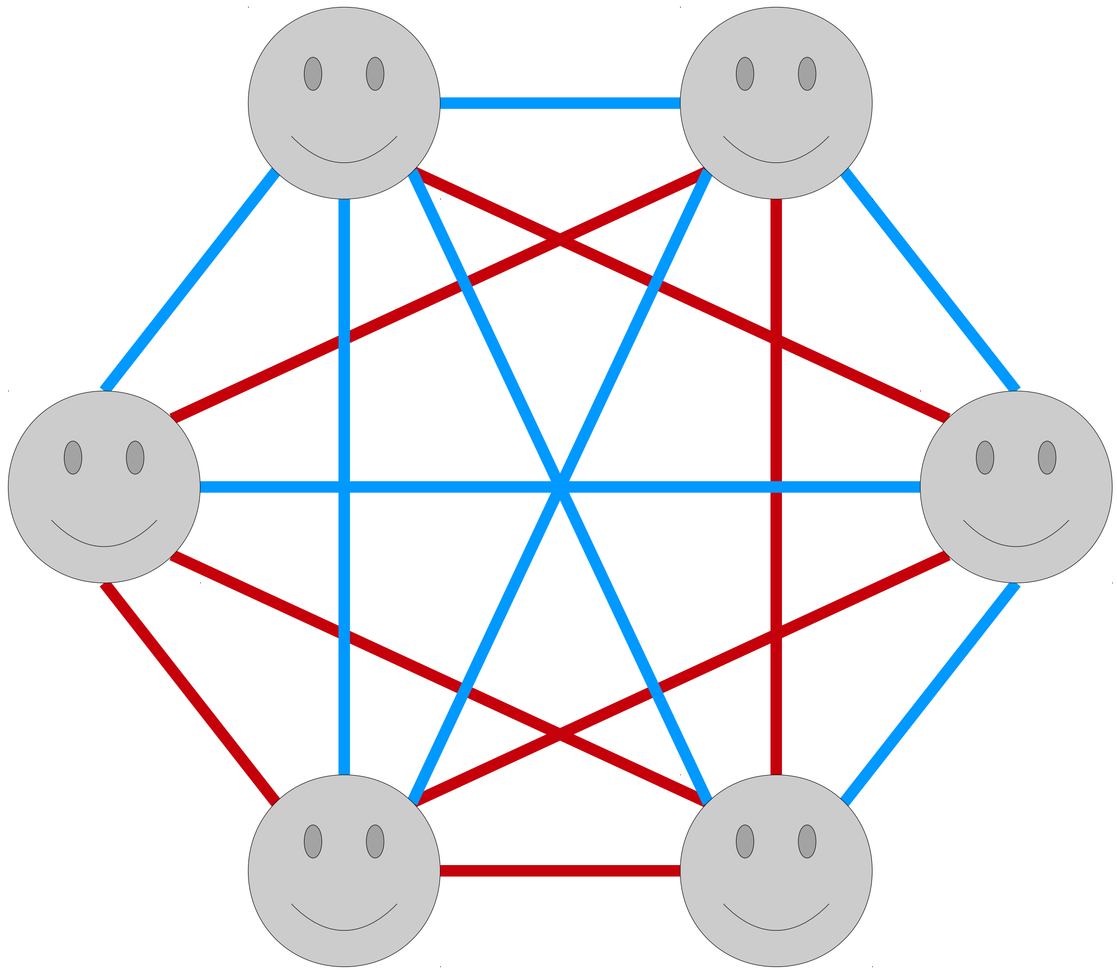


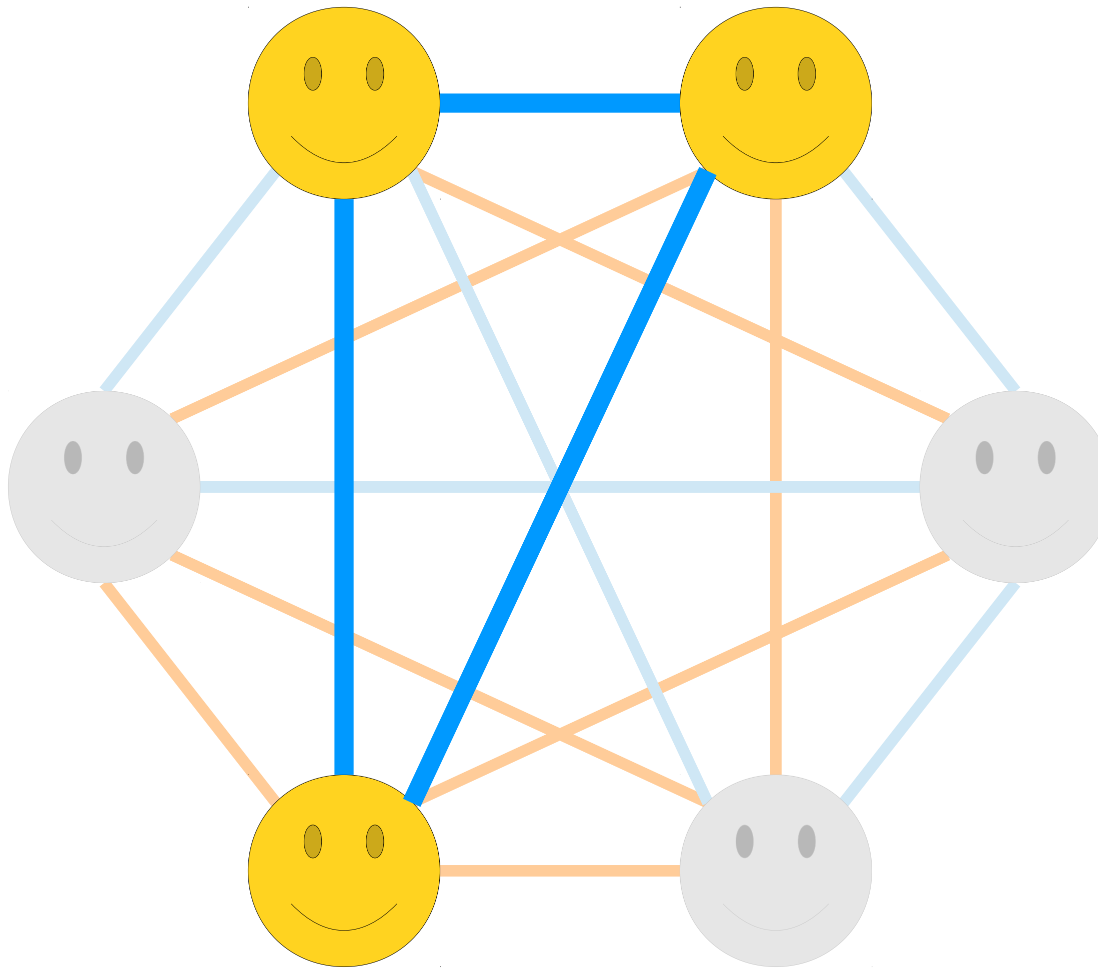


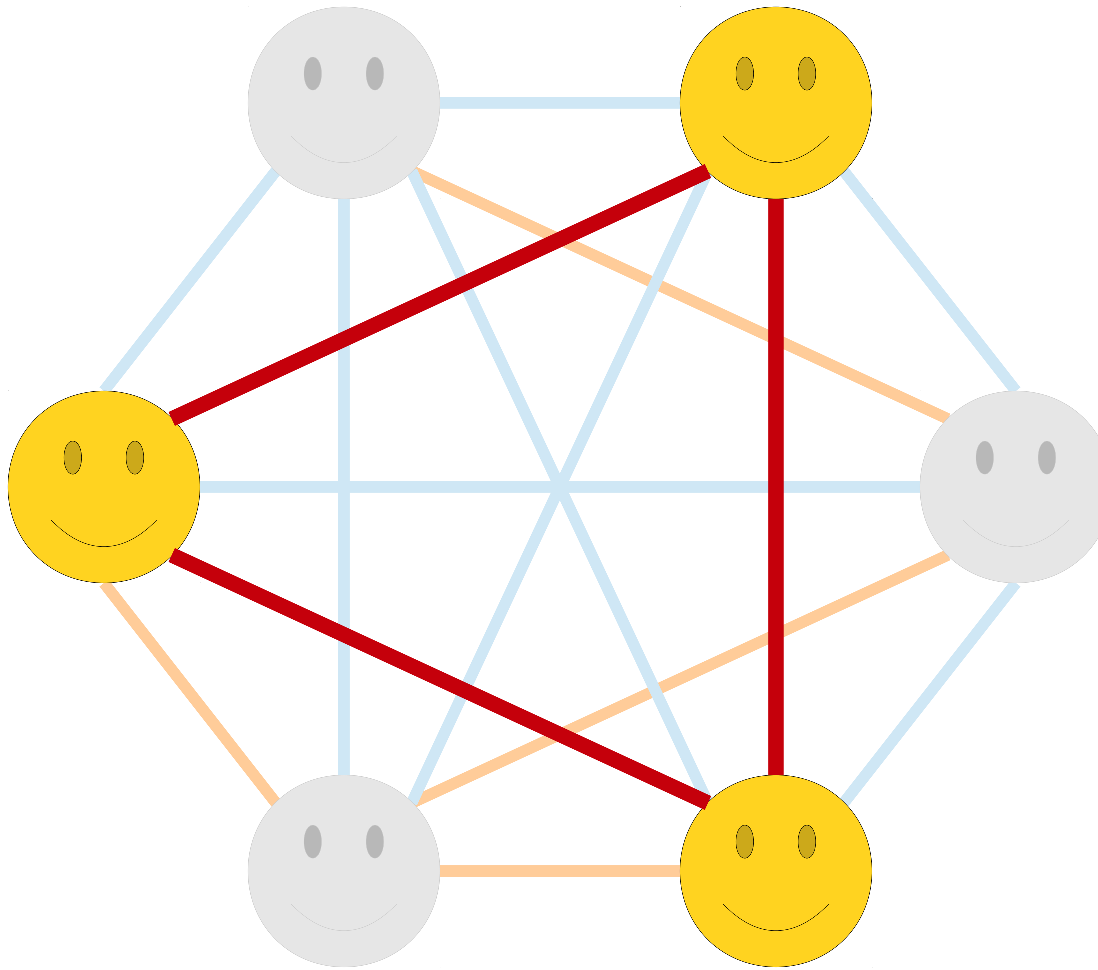


This graph is called a 6-clique, by the way.







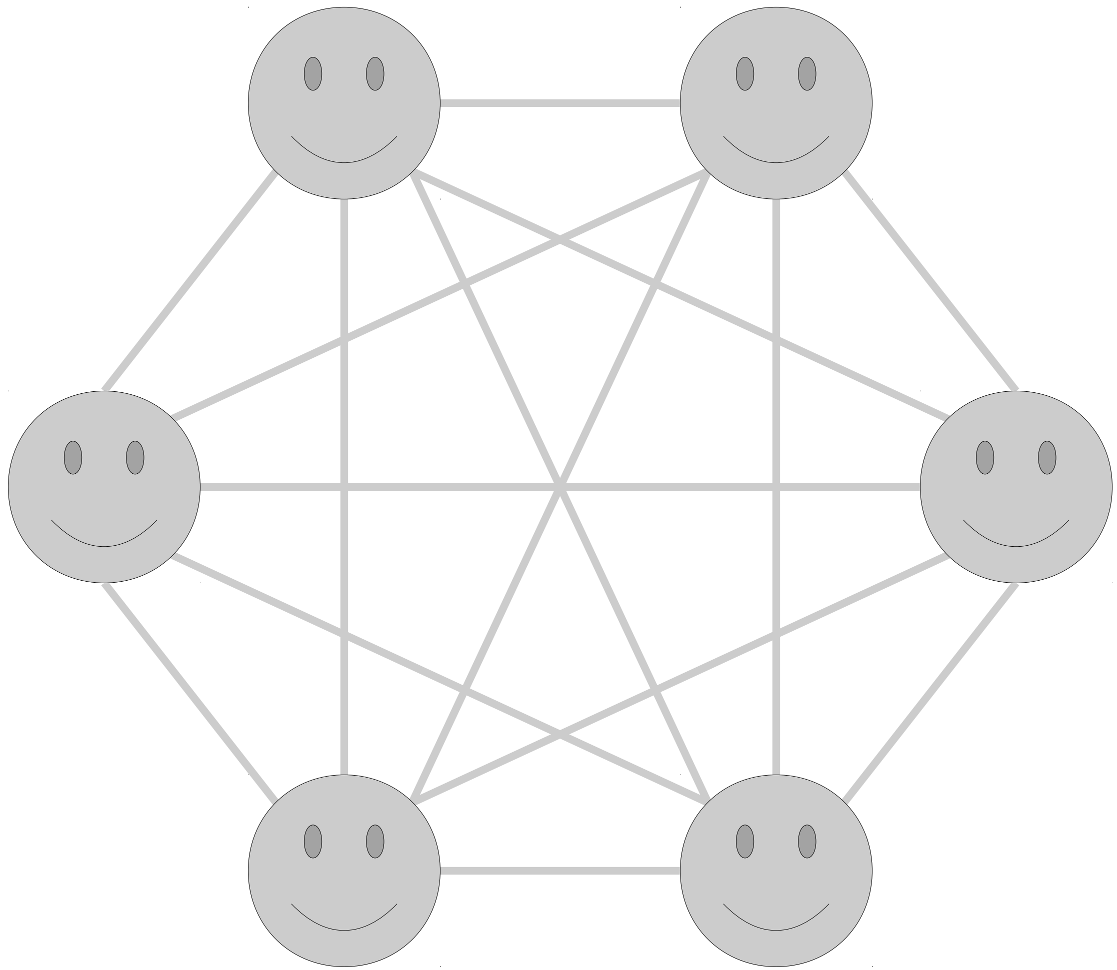


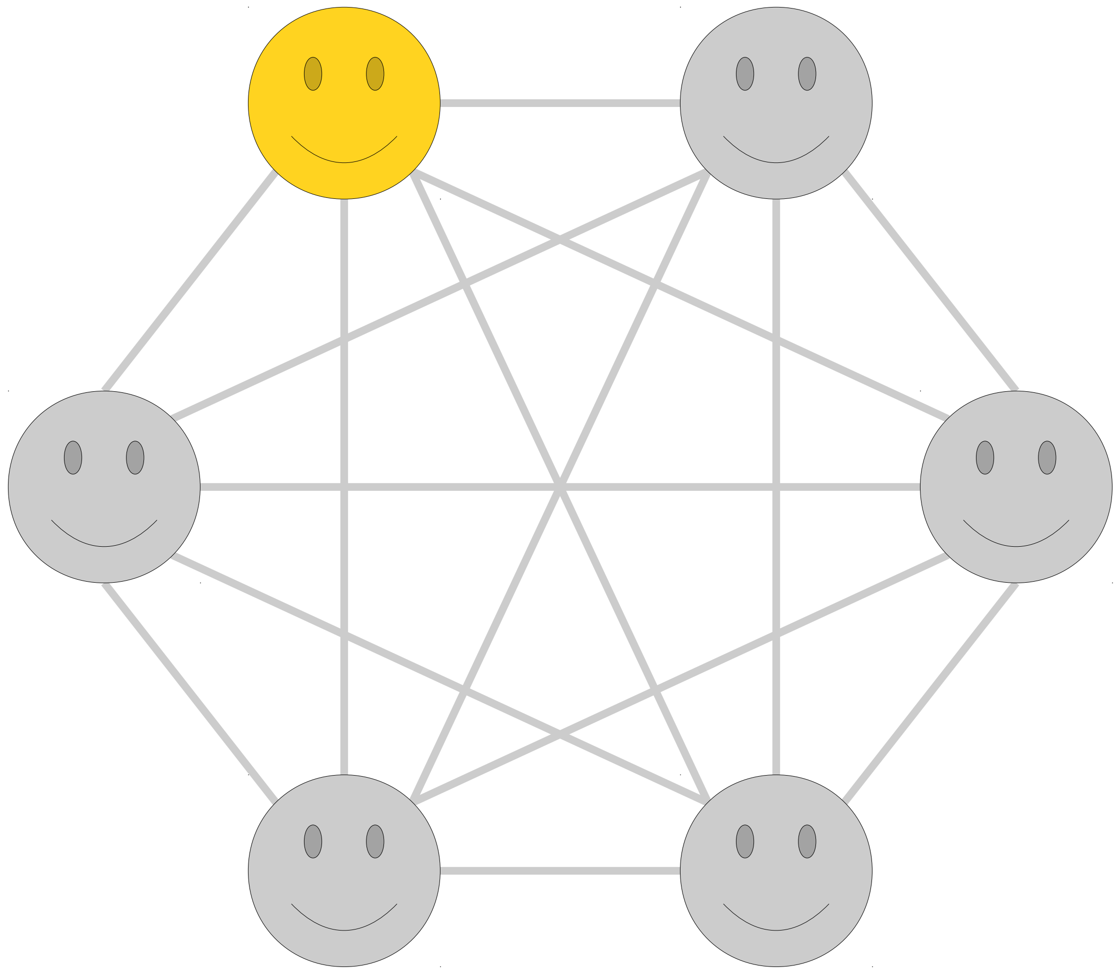
Friends and Strangers Restated

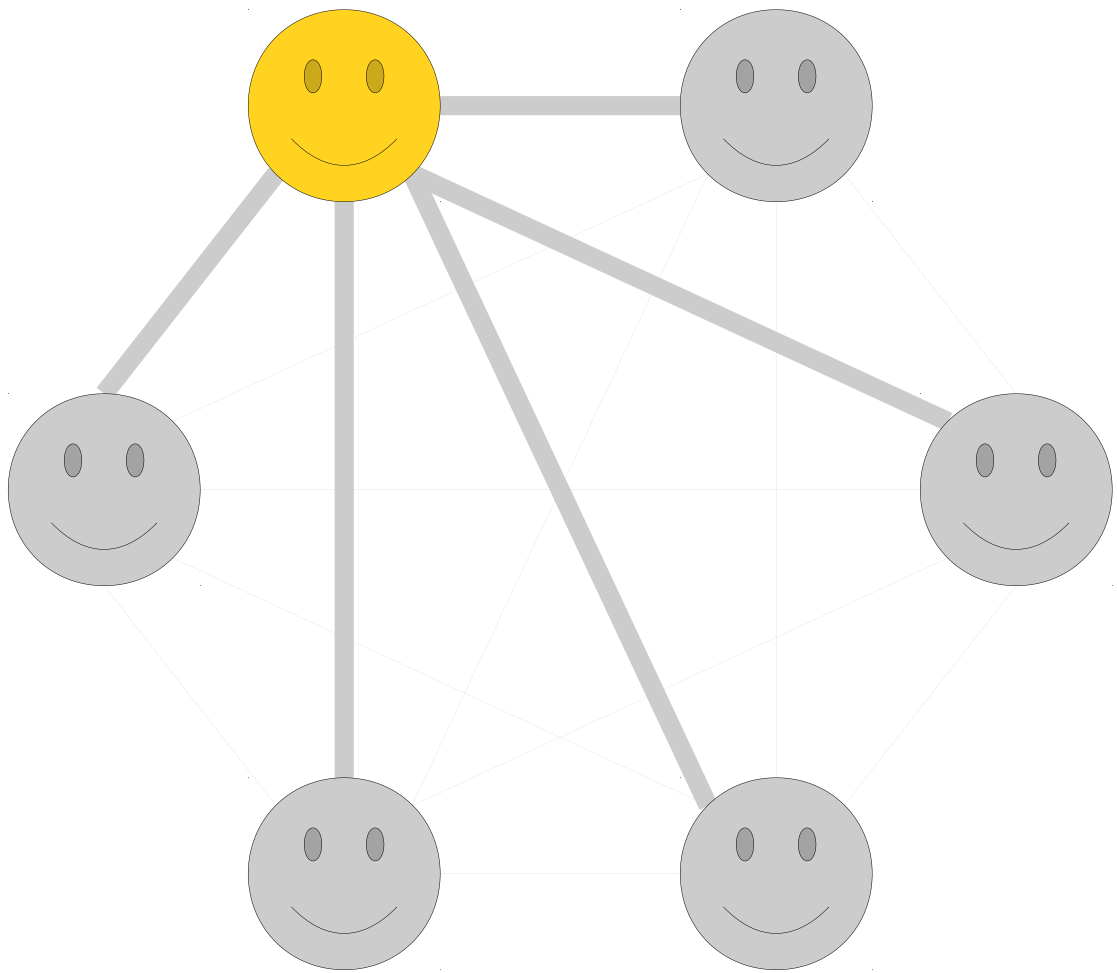
- From a graph-theoretic perspective, the Theorem on Friends and Strangers can be restated as follows:

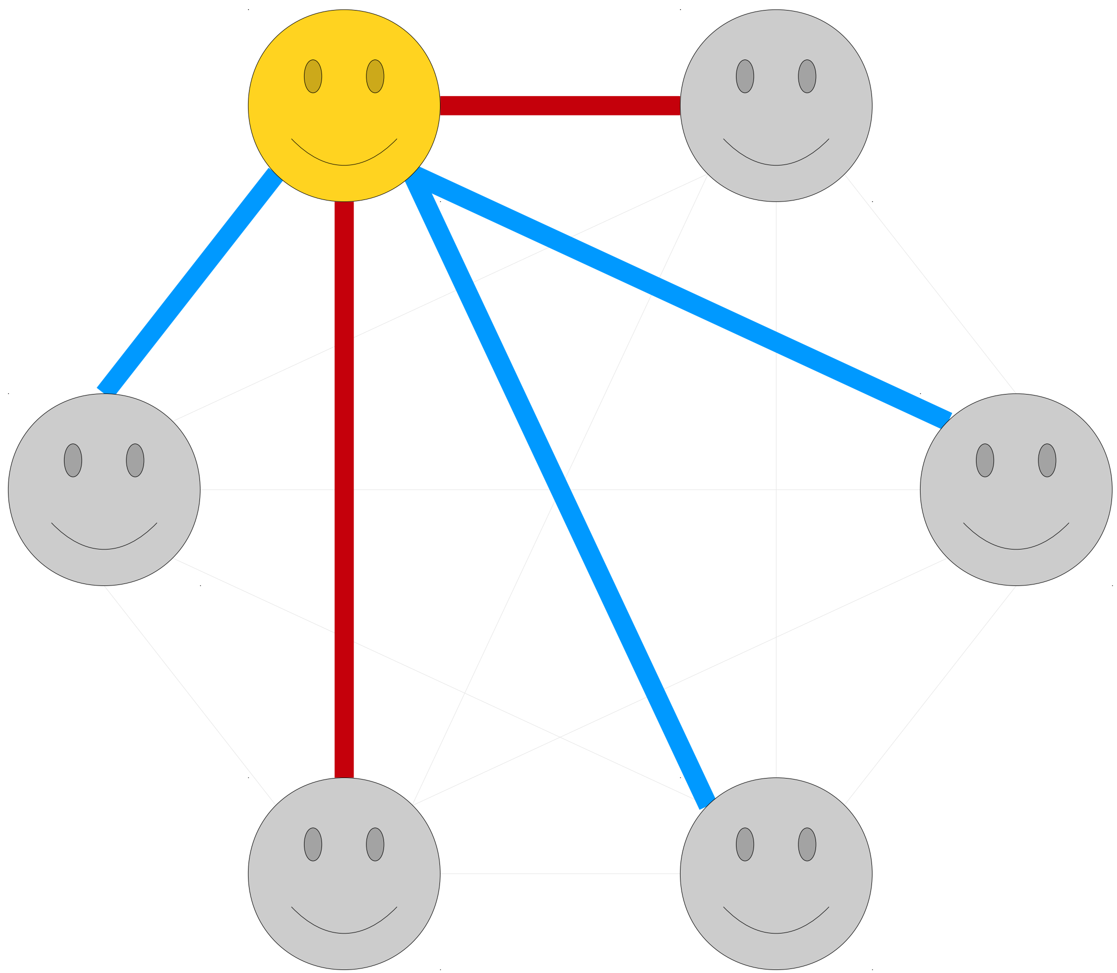
  Theorem: Any 6-clique whose edges are colored red and blue contains a red triangle or a blue triangle (or both).

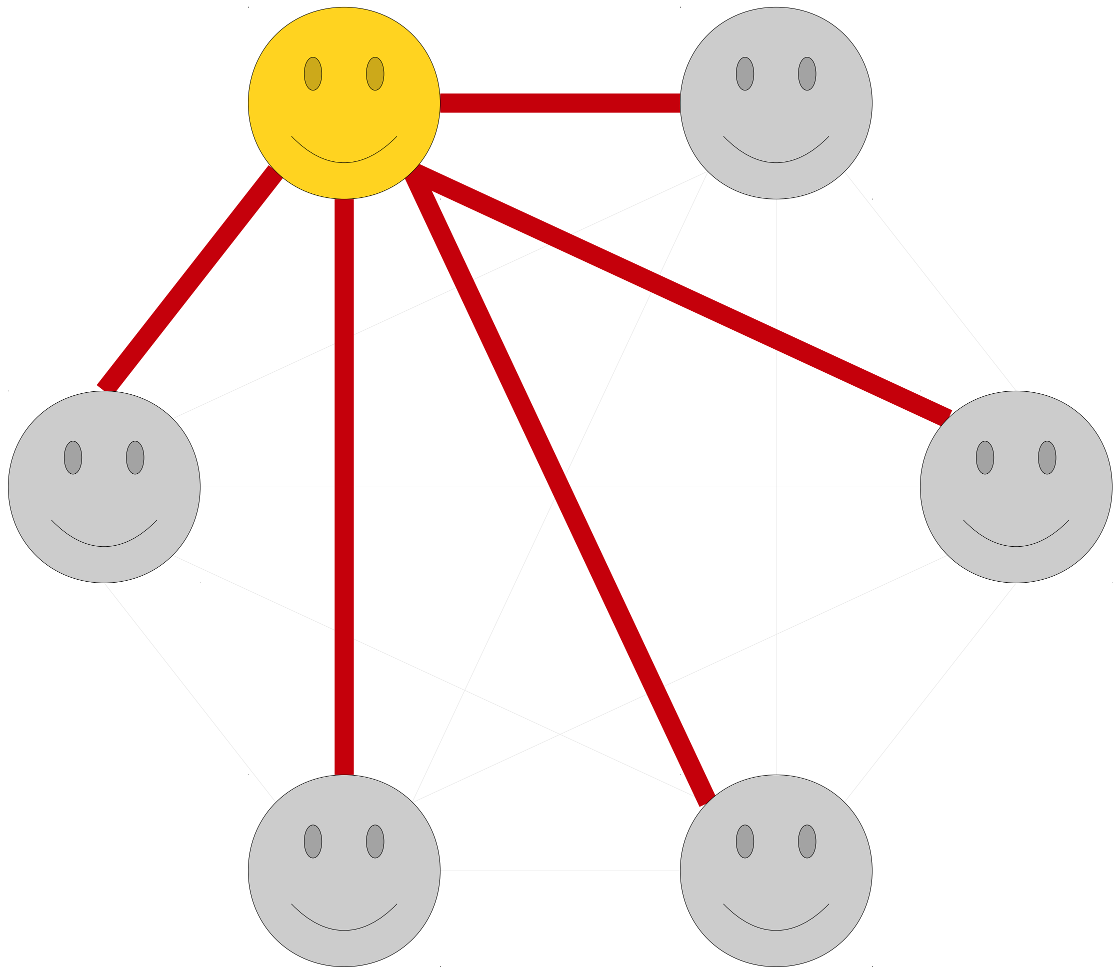
- How can we prove this?

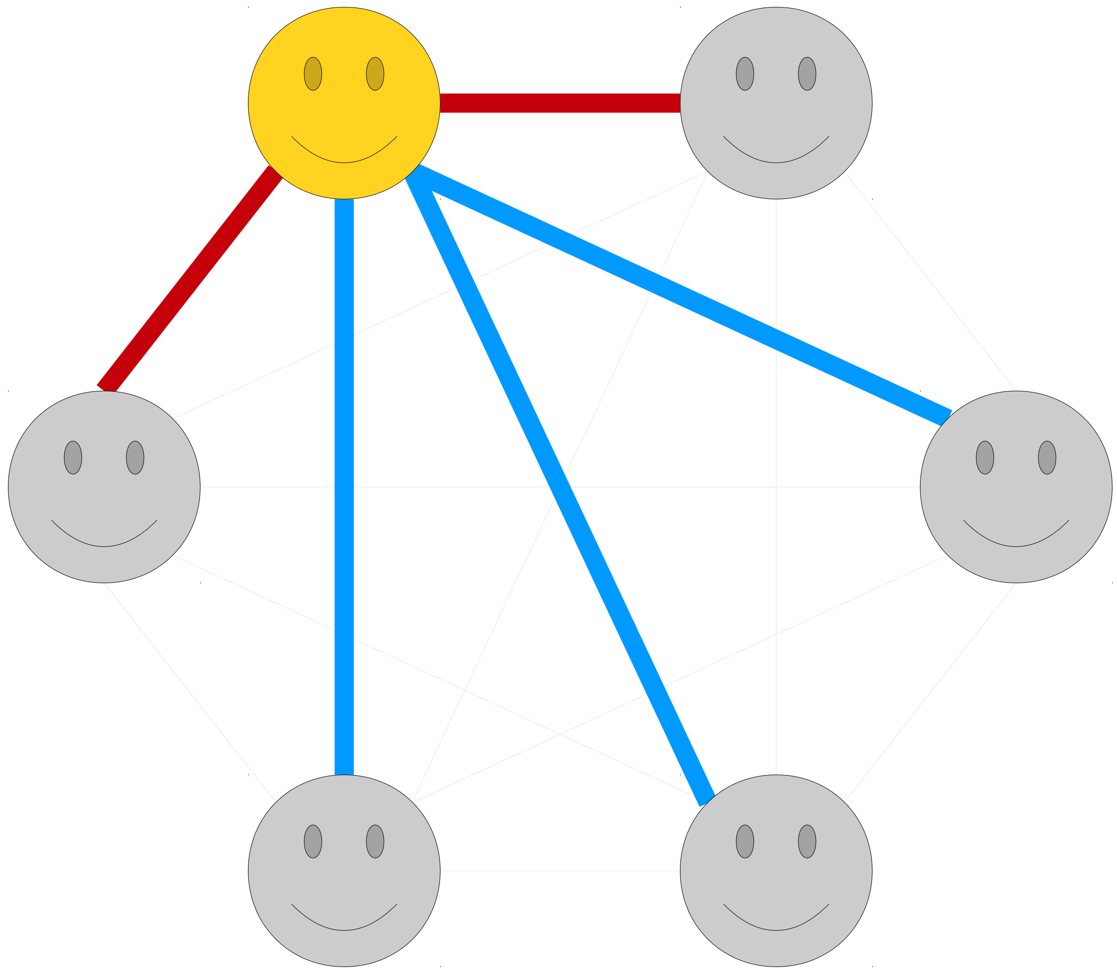


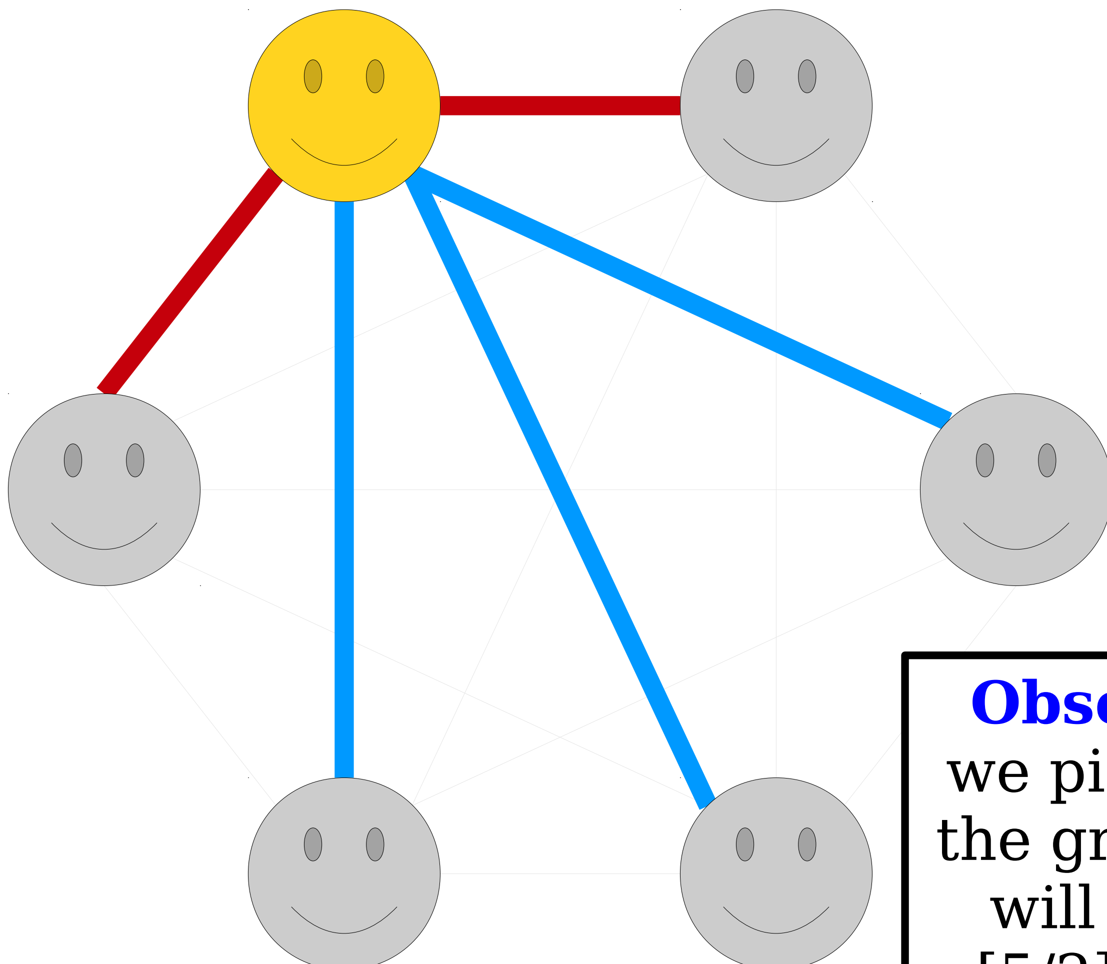




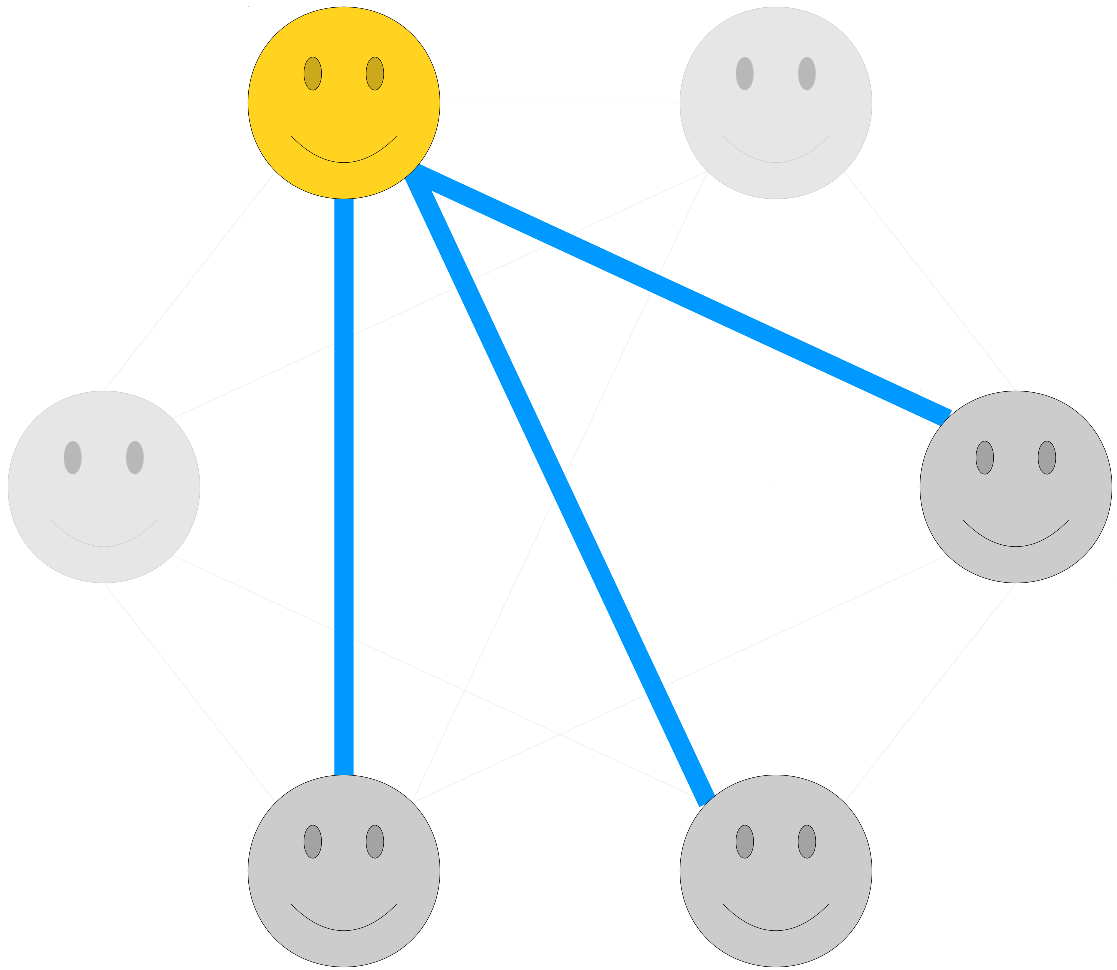


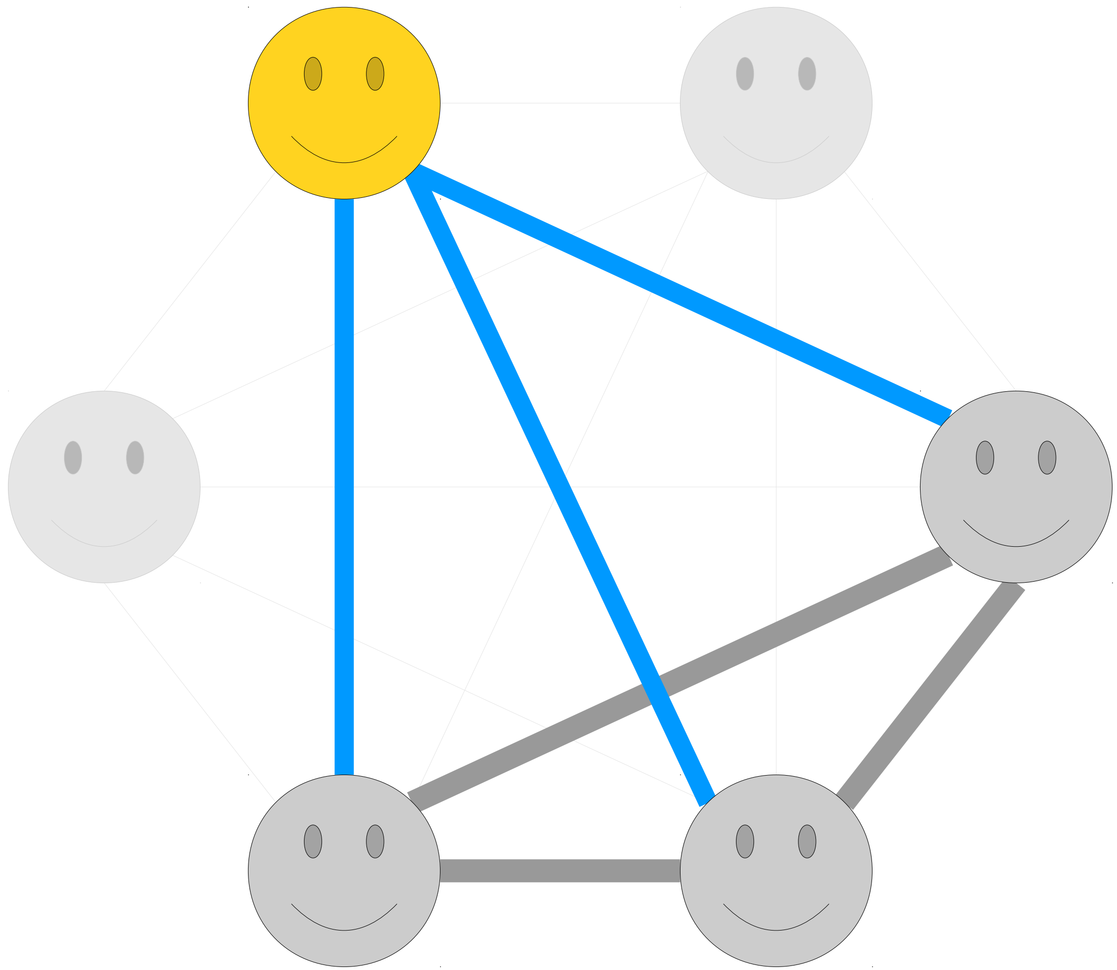


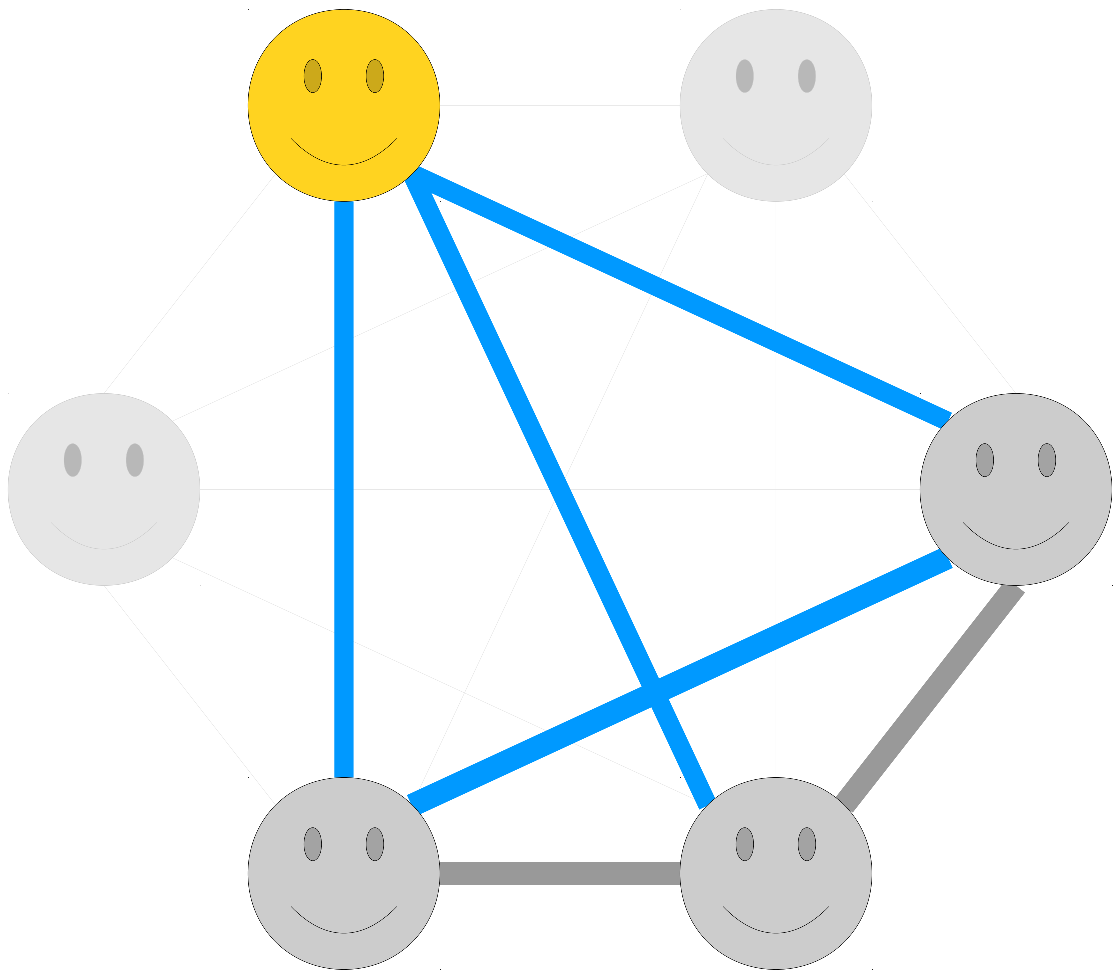


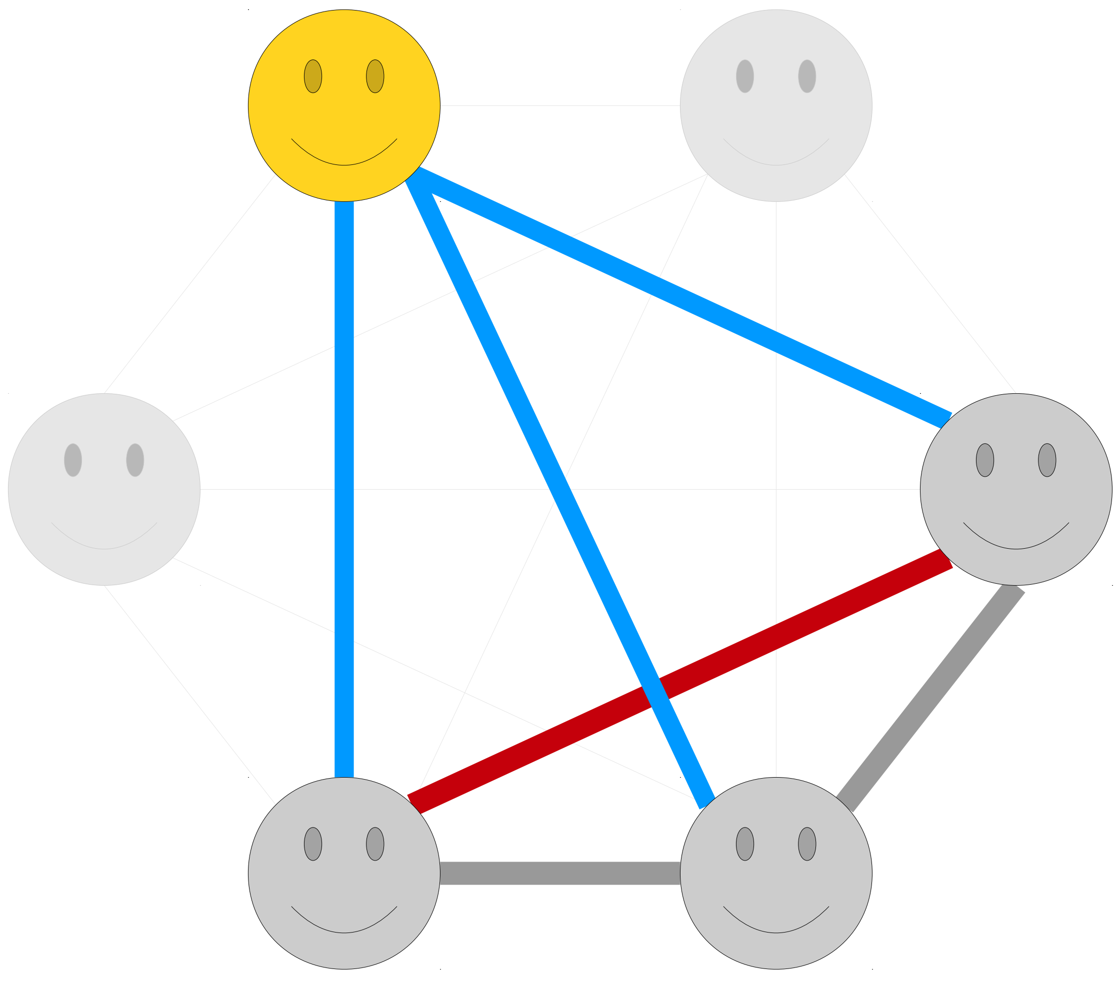


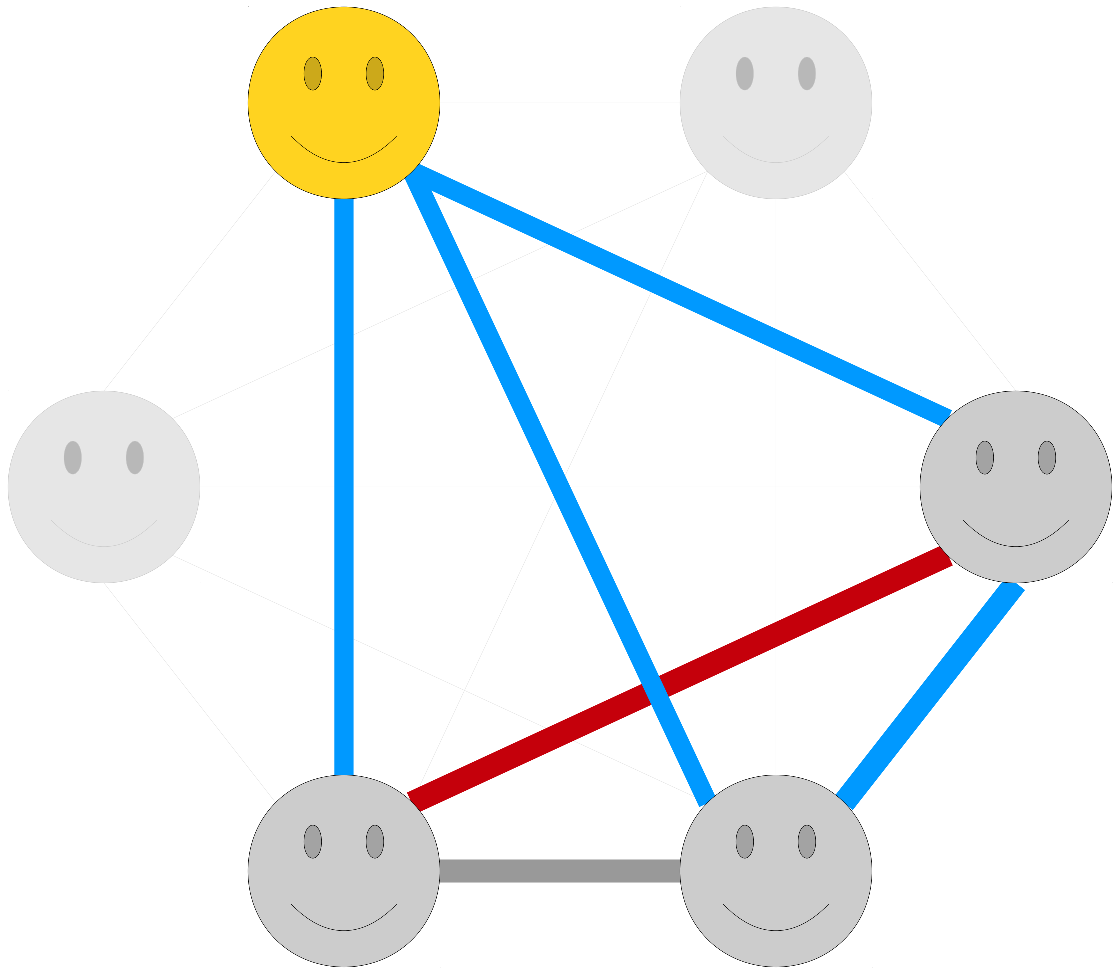
Observation 1: If we pick any node in the graph, that node will have at least ⌈5/2⌉ = 3 edges of the same color incident to it.

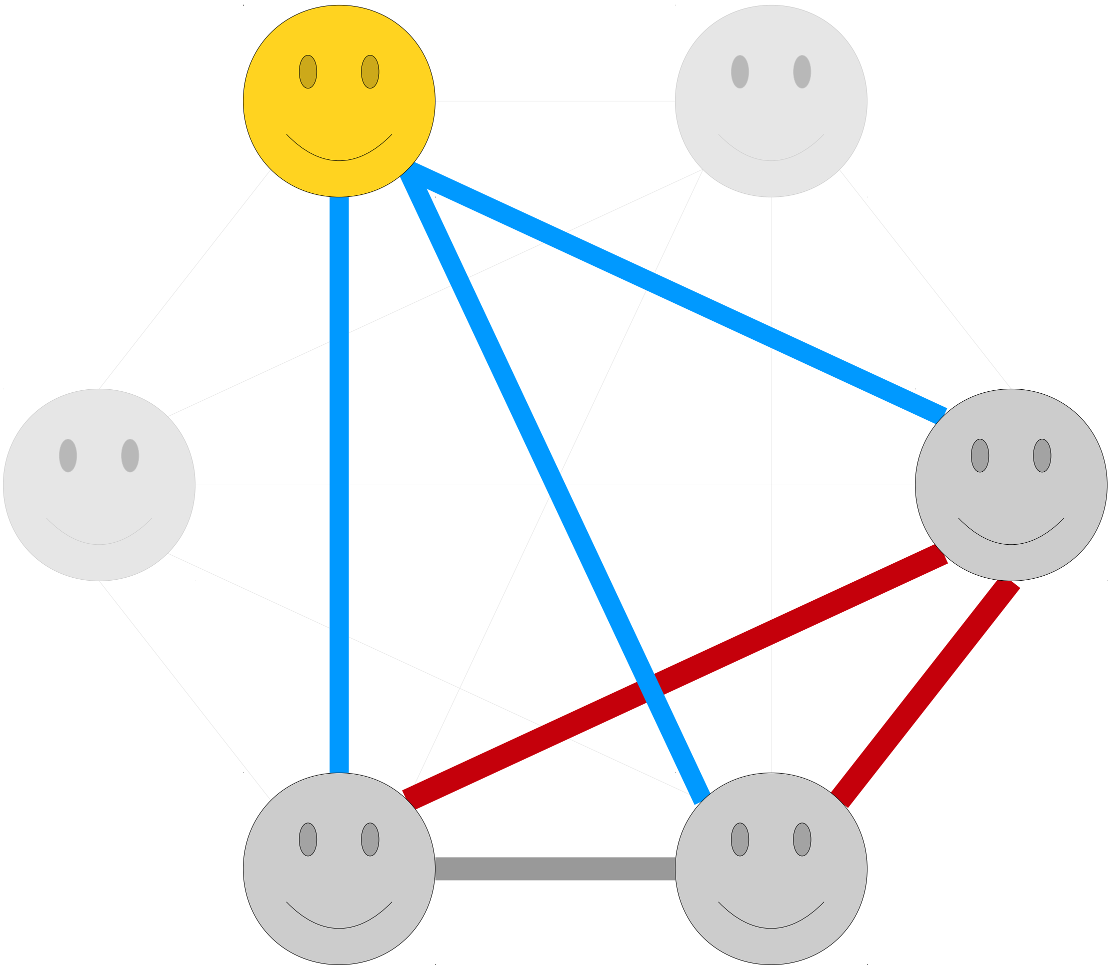


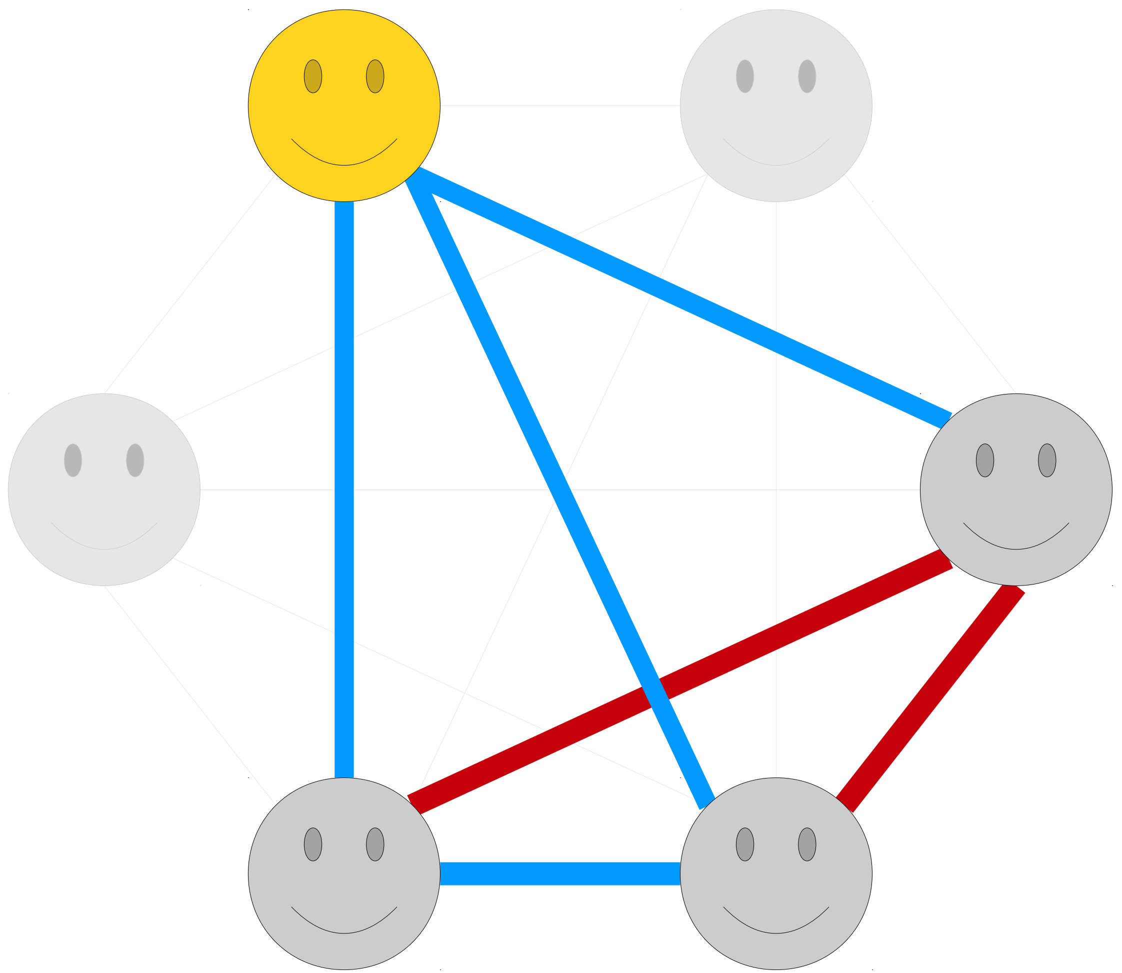


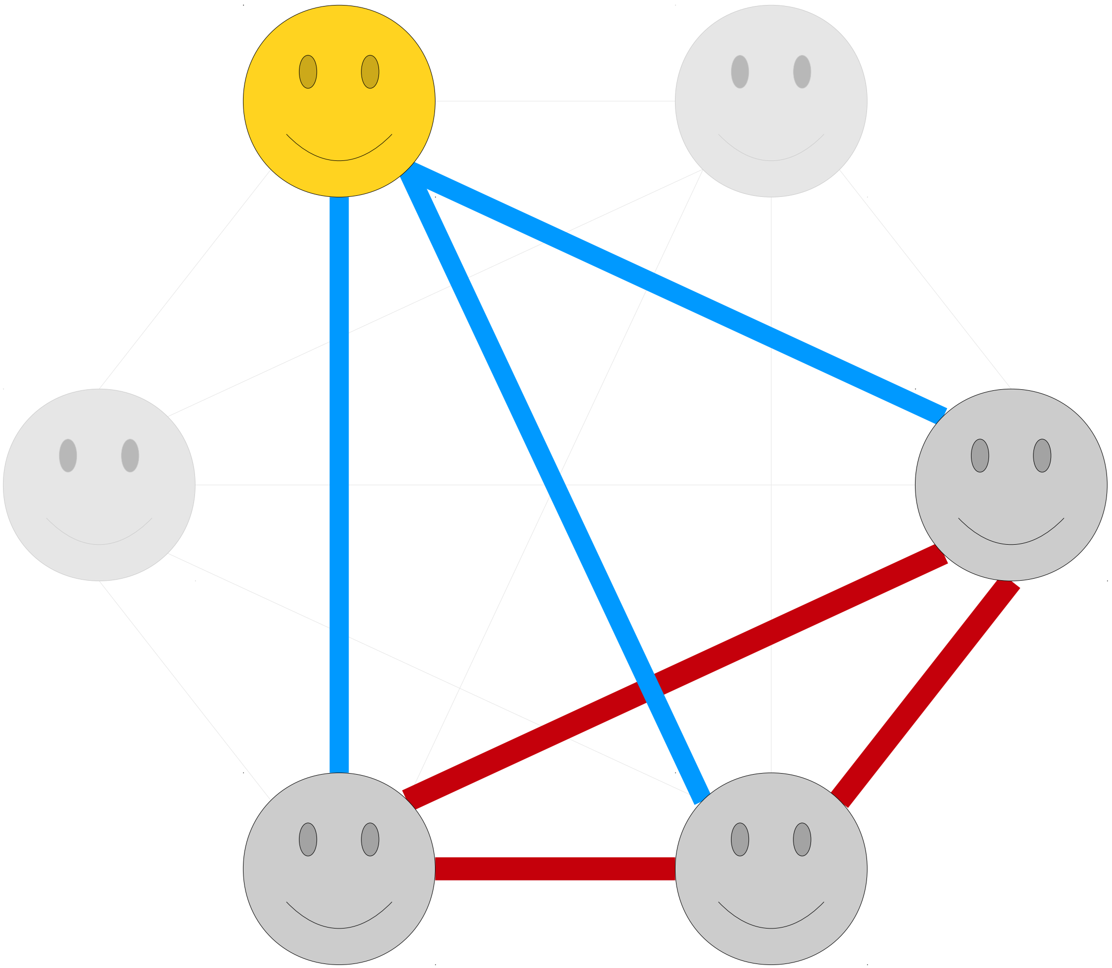












Theorem: Consider a 6-clique in which every edge is colored either red or blue. Then there must be a triangle of red edges, a triangle of blue edges, or both.

Proof: Color the edges of the 6-clique either red or blue arbitrarily. Let x be any node in the 6-clique. It is incident to five edges and there are two possible colors for those edges. Therefore, by the generalized pigeonhole principle, at least $\lceil 5/2 \rceil = 3$ of those edges must be the same color. Call that color c_1 and let the other color be c_2.

Let r, s, and t be three of the nodes adjacent to node x along an edge of color c_1. If any of the edges $\{r, s\}$, $\{r, t\}$, or $\{s, t\}$ are of color c_1, then one of those edges plus the two edges connecting back to node x form a triangle of color c_1. Otherwise, all three of those edges are of color c_2, and they form a triangle of color c_2. Overall, this gives a red triangle or a blue triangle, as required. ■

Ramsey Theory

- The proof we did is a special case of a broader result.
- ***Theorem (Ramsey's Theorem):*** For any natural number n, there is a smallest natural number $R(n)$ such that if the edges of an $R(n)$-clique are colored red or blue, the resulting graph will contain either a red n-clique or a blue n-clique.
  - Our proof was that $R(3) \leq 6$.
- A more philosophical take on this theorem: true disorder is impossible at a large scale, since no matter how you organize things, you're guaranteed to find some interesting substructure.

A Little Math Puzzle

“In a group of $n > 0$ people …

- 90% of those people enjoyed *Get Out*,
- 80% of those people enjoyed *Lady Bird*,
- 70% of those people enjoyed *Arrival*, and
- 60% of those people enjoyed *Zootopia*.

No one enjoyed all four movies. How many people enjoyed at least one of *Get Out* and *Arrival*?”

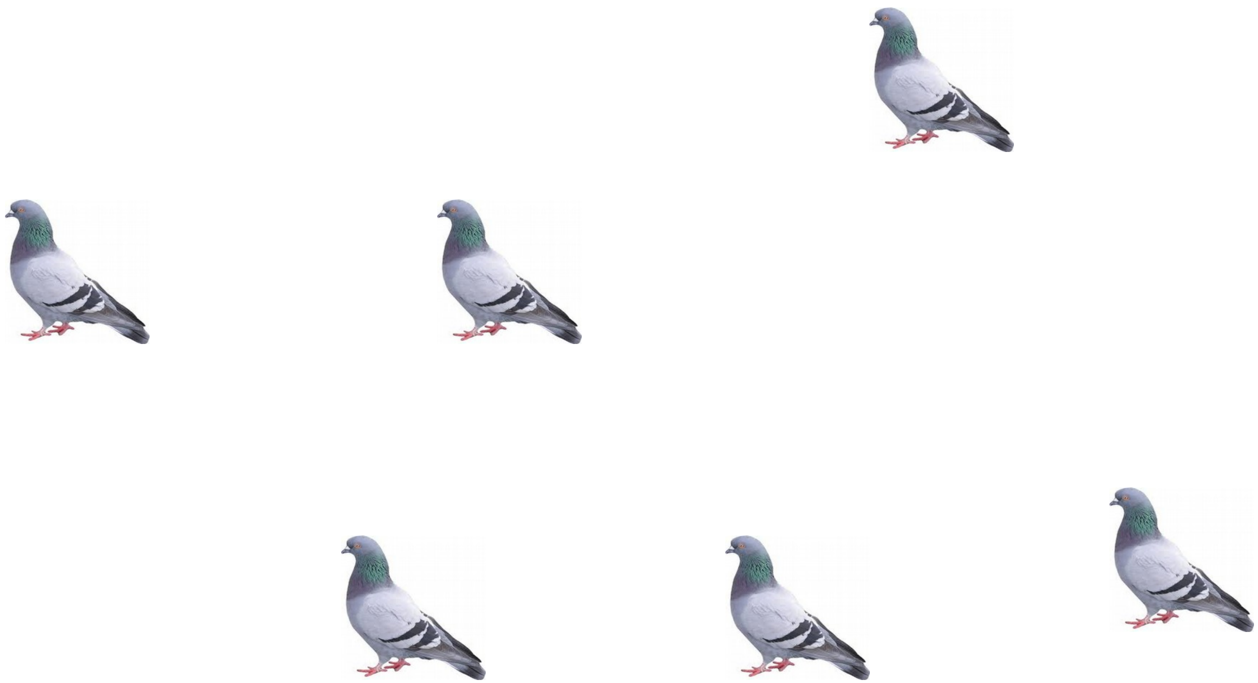
(Adapted from here.)

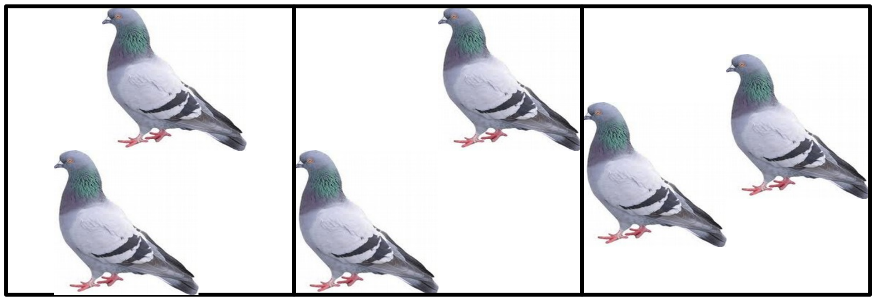
Other Pigeonhole-Type Results

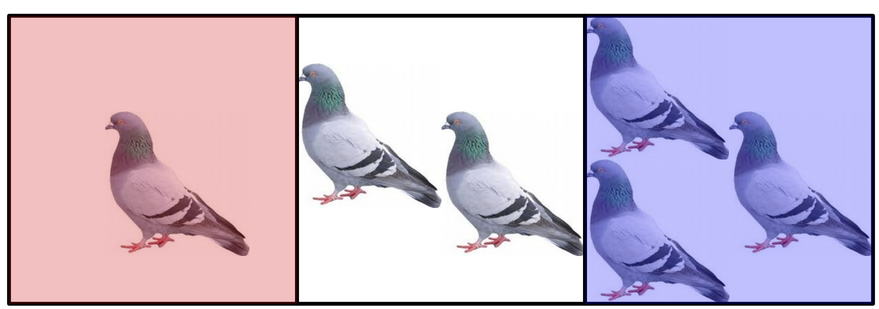
If m objects are distributed into n boxes, then [condition] holds.

If m objects are distributed into n boxes, then some box is loaded to at least the average m/n, and some box is loaded to at most the average m/n.

If m objects are distributed into n boxes, then [condition] holds.







Theorem: If m objects are distributed into n bins, then there is a bin containing more than m/n objects if and only if there is a bin containing fewer than m/n objects.

Lemma: If m objects are distributed into n bins and there are no bins containing more than m/n objects, then there are no bins containing fewer than m/n objects.

Lemma: If m objects are distributed into n bins and there are no bins containing more than m/n objects, then there are no bins containing fewer than m/n objects.

Proof: Assume for the sake of contradiction that m objects are distributed into n bins such that no bin contains more than m/n objects, yet some bin has fewer than m/n objects.

Lemma: If m objects are distributed into n bins and there are no bins containing more than m/n objects, then there are no bins containing fewer than m/n objects.

Proof: Assume for the sake of contradiction that m objects are distributed into n bins such that no bin contains more than m/n objects, yet some bin has fewer than m/n objects.

For simplicity, denote by x_i the number of objects in bin i.

Lemma: If m objects are distributed into n bins and there are no bins containing more than m/n objects, then there are no bins containing fewer than m/n objects.

Proof: Assume for the sake of contradiction that m objects are distributed into n bins such that no bin contains more than m/n objects, yet some bin has fewer than m/n objects.

For simplicity, denote by x_i the number of objects in bin i. Without loss of generality, assume that bin 1 has fewer than m/n objects, meaning that $x_1 < m/n$.

Lemma: If m objects are distributed into n bins and there are no bins containing more than m/n objects, then there are no bins containing fewer than m/n objects.

Proof: Assume for the sake of contradiction that m objects are distributed into n bins such that no bin contains more than m/n objects, yet some bin has fewer than m/n objects.

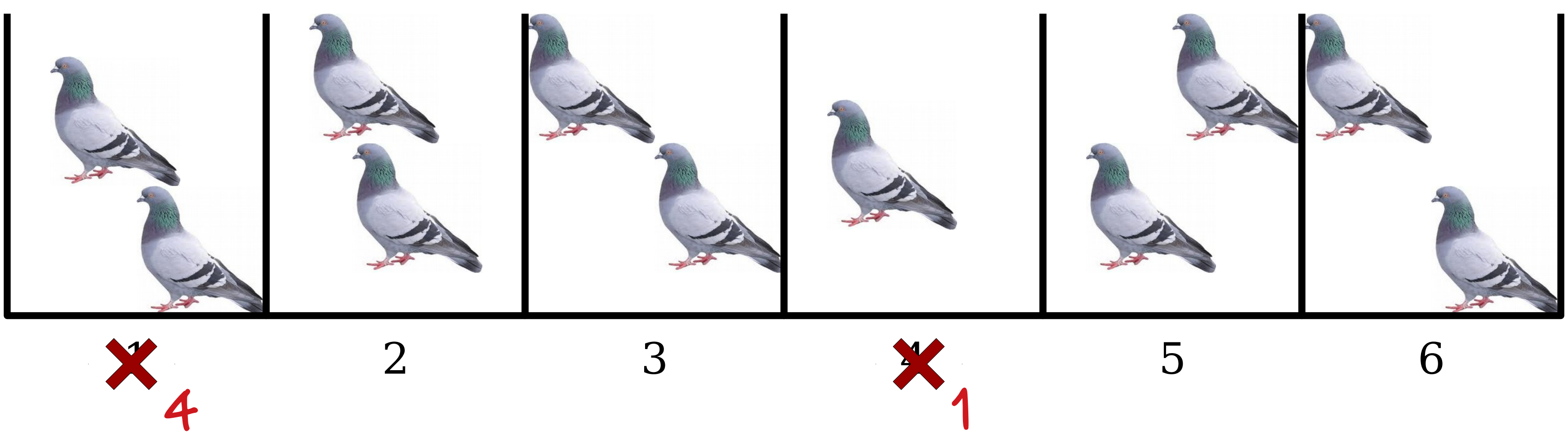
For simplicity, denote by x_i the number of objects in bin i. Without loss of generality, assume that bin 1 has fewer than m/n objects, meaning that $x_1 < m/n$.

Lemma: If m objects are distributed into n bins and there are no bins containing more than $^m/_n$ objects, then there are no bins containing fewer than $^m/_n$ objects.

Proof: Assume for the sake of contradiction that m objects are distributed into n bins such that no bin contains more than $^m/_n$ objects, yet some bin has fewer than $^m/_n$ objects.

For simplicity, denote by x_i the number of objects in bin i. Without loss of generality, assume that bin 1 has fewer than $^m/_n$ objects, meaning that $x_1 < {}^m/_n$.

This magic phrase means "we get to pick how we're labeling things anyway, so if it doesn't work out, just relabel things."



Lemma: If m objects are distributed into n bins and there are no bins containing more than m/n objects, then there are no bins containing fewer than m/n objects.

Proof: Assume for the sake of contradiction that m objects are distributed into n bins such that no bin contains more than m/n objects, yet some bin has fewer than m/n objects.

For simplicity, denote by x_i the number of objects in bin i. Without loss of generality, assume that bin 1 has fewer than m/n objects, meaning that $x_1 < m/n$.

Lemma: If m objects are distributed into n bins and there are no bins containing more than m/n objects, then there are no bins containing fewer than m/n objects.

Proof: Assume for the sake of contradiction that m objects are distributed into n bins such that no bin contains more than m/n objects, yet some bin has fewer than m/n objects.

For simplicity, denote by x_i the number of objects in bin i. Without loss of generality, assume that bin 1 has fewer than m/n objects, meaning that $x_1 < m/n$. Adding up the number of objects in each bin tells us that

$$m = x_1 + x_2 + x_3 + \ldots + x_n$$

Lemma: If m objects are distributed into n bins and there are no bins containing more than m/n objects, then there are no bins containing fewer than m/n objects.

Proof: Assume for the sake of contradiction that m objects are distributed into n bins such that no bin contains more than m/n objects, yet some bin has fewer than m/n objects.

For simplicity, denote by x_i the number of objects in bin i. Without loss of generality, assume that bin 1 has fewer than m/n objects, meaning that $x_1 < m/n$. Adding up the number of objects in each bin tells us that

$$\begin{aligned} m &= x_1 + x_2 + x_3 + \ldots + x_n \\ &< m/n + x_2 + x_3 + \ldots + x_n \end{aligned}$$

Lemma: If m objects are distributed into n bins and there are no bins containing more than m/n objects, then there are no bins containing fewer than m/n objects.

Proof: Assume for the sake of contradiction that m objects are distributed into n bins such that no bin contains more than m/n objects, yet some bin has fewer than m/n objects.

For simplicity, denote by x_i the number of objects in bin i. Without loss of generality, assume that bin 1 has fewer than m/n objects, meaning that $x_1 < m/n$. Adding up the number of objects in each bin tells us that

$$
\begin{aligned}
m &= x_1 + x_2 + x_3 + \ldots + x_n \\
&< m/n + x_2 + x_3 + \ldots + x_n \\
&\leq m/n + m/n + m/n + \ldots + m/n.
\end{aligned}
$$

Lemma: If m objects are distributed into n bins and there are no bins containing more than m/n objects, then there are no bins containing fewer than m/n objects.

Proof: Assume for the sake of contradiction that m objects are distributed into n bins such that no bin contains more than m/n objects, yet some bin has fewer than m/n objects.

For simplicity, denote by x_i the number of objects in bin i. Without loss of generality, assume that bin 1 has fewer than m/n objects, meaning that $x_1 < m/n$. Adding up the number of objects in each bin tells us that

$$\begin{aligned} m &= x_1 + x_2 + x_3 + \ldots + x_n \\ &< m/n + x_2 + x_3 + \ldots + x_n \\ &\leq m/n + m/n + m/n + \ldots + m/n. \end{aligned}$$

This third step follows because each remaining bin has at most m/n objects.

Lemma: If m objects are distributed into n bins and there are no bins containing more than m/n objects, then there are no bins containing fewer than m/n objects.

Proof: Assume for the sake of contradiction that m objects are distributed into n bins such that no bin contains more than m/n objects, yet some bin has fewer than m/n objects.

For simplicity, denote by x_i the number of objects in bin i. Without loss of generality, assume that bin 1 has fewer than m/n objects, meaning that $x_1 < m/n$. Adding up the number of objects in each bin tells us that

$$\begin{aligned} m &= x_1 + x_2 + x_3 + \dots + x_n \\ &< m/n + x_2 + x_3 + \dots + x_n \\ &\le m/n + m/n + m/n + \dots + m/n. \end{aligned}$$

This third step follows because each remaining bin has at most m/n objects. Grouping the n copies of the m/n term here tells us that

$$\begin{aligned} m &< m/n + m/n + m/n + \dots + m/n \\ &= m. \end{aligned}$$

Lemma: If m objects are distributed into n bins and there are no bins containing more than m/n objects, then there are no bins containing fewer than m/n objects.

Proof: Assume for the sake of contradiction that m objects are distributed into n bins such that no bin contains more than m/n objects, yet some bin has fewer than m/n objects.

For simplicity, denote by x_i the number of objects in bin i. Without loss of generality, assume that bin 1 has fewer than m/n objects, meaning that $x_1 < m/n$. Adding up the number of objects in each bin tells us that

$$\begin{aligned} m &= x_1 + x_2 + x_3 + \ldots + x_n \\ &< m/n + x_2 + x_3 + \ldots + x_n \\ &\leq m/n + m/n + m/n + \ldots + m/n. \end{aligned}$$

This third step follows because each remaining bin has at most m/n objects. Grouping the n copies of the m/n term here tells us that

$$\begin{aligned} m &< m/n + m/n + m/n + \ldots + m/n \\ &= m. \end{aligned}$$

But this means $m < m$, which is impossible.

Lemma: If m objects are distributed into n bins and there are no bins containing more than m/n objects, then there are no bins containing fewer than m/n objects.

Proof: Assume for the sake of contradiction that m objects are distributed into n bins such that no bin contains more than m/n objects, yet some bin has fewer than m/n objects.

For simplicity, denote by x_i the number of objects in bin i. Without loss of generality, assume that bin 1 has fewer than m/n objects, meaning that $x_1 < m/n$. Adding up the number of objects in each bin tells us that

$$\begin{aligned} m &= x_1 + x_2 + x_3 + \ldots + x_n \\ &< m/n + x_2 + x_3 + \ldots + x_n \\ &\leq m/n + m/n + m/n + \ldots + m/n. \end{aligned}$$

This third step follows because each remaining bin has at most m/n objects. Grouping the n copies of the m/n term here tells us that

$$\begin{aligned} m &< m/n + m/n + m/n + \ldots + m/n \\ &= m. \end{aligned}$$

But this means $m < m$, which is impossible. We've reached a contradiction, so our assumption was wrong, so if m objects are distributed into n bins and no bin has more than m/n objects, no bin has fewer than m/n objects either.

Lemma: If m objects are distributed into n bins and there are no bins containing more than m/n objects, then there are no bins containing fewer than m/n objects.

Proof: Assume for the sake of contradiction that m objects are distributed into n bins such that no bin contains more than m/n objects, yet some bin has fewer than m/n objects.

For simplicity, denote by x_i the number of objects in bin i. Without loss of generality, assume that bin 1 has fewer than m/n objects, meaning that $x_1 < m/n$. Adding up the number of objects in each bin tells us that

$$\begin{aligned} m &= x_1 + x_2 + x_3 + \ldots + x_n \\ &< m/n + x_2 + x_3 + \ldots + x_n \\ &\leq m/n + m/n + m/n + \ldots + m/n. \end{aligned}$$

This third step follows because each remaining bin has at most m/n objects. Grouping the n copies of the m/n term here tells us that

$$\begin{aligned} m &< m/n + m/n + m/n + \ldots + m/n \\ &= m. \end{aligned}$$

But this means $m < m$, which is impossible. We've reached a contradiction, so our assumption was wrong, so if m objects are distributed into n bins and no bin has more than m/n objects, no bin has fewer than m/n objects either. ■

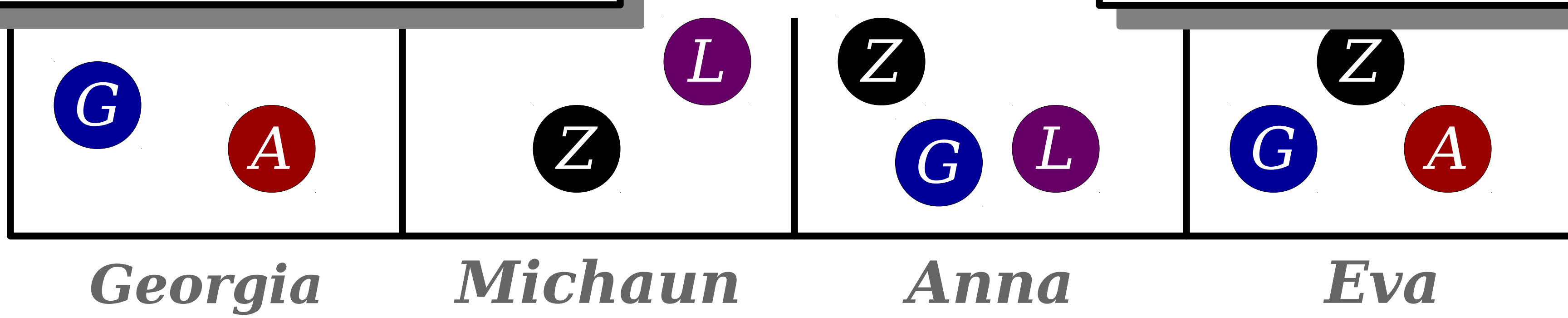
"In a group of $n > 0$ people …

- 90% of those people enjoyed ***Get Out***,
- 80% of those people enjoyed ***Lady Bird***,
- 70% of those people enjoyed ***Arrival***, and
- 60% of those people enjoyed ***Zootopia***.

No one enjoyed all four movies. How many people enjoyed at least one of *Get Out* and *Arrival*?"

Insight 1: Model movie preferences as balls (movies) in bins (people).

Insight 2: There are *n* total bins, one for each person.



"In a group of $n > 0$ people …

- 90% of those people enjoyed ***Get Out***,
- 80% of those people enjoyed ***Lady Bird***,
- 70% of those people enjoyed ***Arrival***, and
- 60% of those people enjoyed ***Zootopia***.

No one enjoyed all four movies. How many people enjoyed at least one of *Get Out* and *Arrival*?"

$$.9n + .8n + .7n + .6n$$

$$= 3n$$

Insight 3: There are $3n$ balls being distributed into n bins.

Insight 4: The average number of balls in each bin is 3.

"In a group of $n > 0$ people ...

- 90% of those people enjoyed ***Get Out***,
- 80% of those people enjoyed ***Lady Bird***,
- 70% of those people enjoyed ***Arrival***, and
- 60% of those people enjoyed ***Zootopia***.

No one enjoyed all four movies. How many people enjoyed at least one of *Get Out* and *Arrival*?"

Insight 5: No one enjoyed more than three movies…

Insight 6: … so no one enjoyed fewer than three movies …

Insight 7: … so everyone enjoyed exactly three movies.

"In a group of $n > 0$ people …

- 90% of those people enjoyed ***Get Out***,
- 80% of those people enjoyed ***Lady Bird***,
- 70% of those people enjoyed ***Arrival***, and
- 60% of those people enjoyed ***Zootopia***.

No one enjoyed all four movies. How many people enjoyed at least one of *Get Out* and *Arrival*?"

Insight 8: You have to enjoy at least one of these movies to enjoy three of the four movies.

Conclusion: Everyone liked at least one of these two movies!

Theorem: In the scenario described here, all n people enjoyed at least one of *Get Out* and *Arrival*.

"In a group of $n > 0$ people …

- 90% of those people enjoyed ***Get Out***,
- 80% of those people enjoyed ***Lady Bird***,
- 70% of those people enjoyed ***Arrival***, and
- 60% of those people enjoyed ***Zootopia***.

No one enjoyed all four movies. How many people enjoyed at least one of *Get Out* and *Arrival*?"

Theorem: In the scenario described here, all n people enjoyed at least one of *Get Out* and *Arrival*.

Proof: Suppose there is a group of n people meeting these criteria.

"In a group of $n > 0$ people …

- 90% of those people enjoyed ***Get Out***,
- 80% of those people enjoyed ***Lady Bird***,
- 70% of those people enjoyed ***Arrival***, and
- 60% of those people enjoyed ***Zootopia***.

No one enjoyed all four movies. How many people enjoyed at least one of *Get Out* and *Arrival*?"

Theorem: In the scenario described here, all n people enjoyed at least one of *Get Out* and *Arrival*.

Proof: Suppose there is a group of n people meeting these criteria. We can model this problem by representing each person as a bin and each time a person enjoys a movie as a ball.

"In a group of $n > 0$ people …

- 90% of those people enjoyed ***Get Out***,
- 80% of those people enjoyed ***Lady Bird***,
- 70% of those people enjoyed ***Arrival***, and
- 60% of those people enjoyed ***Zootopia***.

No one enjoyed all four movies. How many people enjoyed at least one of *Get Out* and *Arrival*?"

Theorem: In the scenario described here, all n people enjoyed at least one of *Get Out* and *Arrival*.

Proof: Suppose there is a group of n people meeting these criteria. We can model this problem by representing each person as a bin and each time a person enjoys a movie as a ball. The number of balls is

$$.9n + .8n + .7n + .6n = 3n,$$

and since there are n people, there are n bins.

"In a group of $n > 0$ people …

- 90% of those people enjoyed ***Get Out***,
- 80% of those people enjoyed ***Lady Bird***,
- 70% of those people enjoyed ***Arrival***, and
- 60% of those people enjoyed ***Zootopia***.

No one enjoyed all four movies. How many people enjoyed at least one of *Get Out* and *Arrival*?"

Theorem: In the scenario described here, all n people enjoyed at least one of *Get Out* and *Arrival*.

Proof: Suppose there is a group of n people meeting these criteria. We can model this problem by representing each person as a bin and each time a person enjoys a movie as a ball. The number of balls is

$$.9n + .8n + .7n + .6n = 3n,$$

and since there are n people, there are n bins. Since no person liked all four movies, no bin contains more than $3 = {}^{3n}/_{n}$ balls, so by our earlier theorem we see that no bin contains fewer than three balls.

"In a group of $n > 0$ people …

- 90% of those people enjoyed ***Get Out***,
- 80% of those people enjoyed ***Lady Bird***,
- 70% of those people enjoyed ***Arrival***, and
- 60% of those people enjoyed ***Zootopia***.

No one enjoyed all four movies. How many people enjoyed at least one of *Get Out* and *Arrival*?"

Theorem: In the scenario described here, all n people enjoyed at least one of *Get Out* and *Arrival*.

Proof: Suppose there is a group of n people meeting these criteria. We can model this problem by representing each person as a bin and each time a person enjoys a movie as a ball. The number of balls is

$$.9n + .8n + .7n + .6n = 3n,$$

and since there are n people, there are n bins. Since no person liked all four movies, no bin contains more than $3 = {}^{3n}/_{n}$ balls, so by our earlier theorem we see that no bin contains fewer than three balls. Therefore, each bin contains exactly three balls.

"In a group of $n > 0$ people …

- 90% of those people enjoyed ***Get Out***,
- 80% of those people enjoyed ***Lady Bird***,
- 70% of those people enjoyed ***Arrival***, and
- 60% of those people enjoyed ***Zootopia***.

No one enjoyed all four movies. How many people enjoyed at least one of *Get Out* and *Arrival*?"

Theorem: In the scenario described here, all n people enjoyed at least one of *Get Out* and *Arrival*.

Proof: Suppose there is a group of n people meeting these criteria. We can model this problem by representing each person as a bin and each time a person enjoys a movie as a ball. The number of balls is

$$.9n + .8n + .7n + .6n = 3n,$$

and since there are n people, there are n bins. Since no person liked all four movies, no bin contains more than $3 = {}^{3n}/_{n}$ balls, so by our earlier theorem we see that no bin contains fewer than three balls. Therefore, each bin contains exactly three balls.

Now suppose for the sake of contradiction that someone didn't enjoy *Get Out* and didn't enjoy *Arrival*.

"In a group of $n > 0$ people …

- 90% of those people enjoyed ***Get Out***,
- 80% of those people enjoyed ***Lady Bird***,
- 70% of those people enjoyed ***Arrival***, and
- 60% of those people enjoyed ***Zootopia***.

No one enjoyed all four movies. How many people enjoyed at least one of *Get Out* and *Arrival*?"

Theorem: In the scenario described here, all n people enjoyed at least one of *Get Out* and *Arrival*.

Proof: Suppose there is a group of n people meeting these criteria. We can model this problem by representing each person as a bin and each time a person enjoys a movie as a ball. The number of balls is

$$.9n + .8n + .7n + .6n = 3n,$$

and since there are n people, there are n bins. Since no person liked all four movies, no bin contains more than $3 = {}^{3n}/_{n}$ balls, so by our earlier theorem we see that no bin contains fewer than three balls. Therefore, each bin contains exactly three balls.

Now suppose for the sake of contradiction that someone didn't enjoy *Get Out* and didn't enjoy *Arrival*. This means they could enjoy at most two of the four movies, contradicting that each person enjoys exactly three.

"In a group of $n > 0$ people …

- 90% of those people enjoyed ***Get Out***,
- 80% of those people enjoyed ***Lady Bird***,
- 70% of those people enjoyed ***Arrival***, and
- 60% of those people enjoyed ***Zootopia***.

No one enjoyed all four movies. How many people enjoyed at least one of *Get Out* and *Arrival*?"

Theorem: In the scenario described here, all n people enjoyed at least one of *Get Out* and *Arrival*.

Proof: Suppose there is a group of n people meeting these criteria. We can model this problem by representing each person as a bin and each time a person enjoys a movie as a ball. The number of balls is

$$.9n + .8n + .7n + .6n = 3n,$$

and since there are n people, there are n bins. Since no person liked all four movies, no bin contains more than $3 = {}^{3n}/_{n}$ balls, so by our earlier theorem we see that no bin contains fewer than three balls. Therefore, each bin contains exactly three balls.

Now suppose for the sake of contradiction that someone didn't enjoy *Get Out* and didn't enjoy *Arrival*. This means they could enjoy at most two of the four movies, contradicting that each person enjoys exactly three.

We've reached a contradiction, so our assumption was wrong and each person enjoyed at least one of *Get Out* and *Arrival*. ■

"In a group of $n > 0$ people …

- 90% of those people enjoyed ***Get Out***,
- 80% of those people enjoyed ***Lady Bird***,
- 70% of those people enjoyed ***Arrival***, and
- 60% of those people enjoyed ***Zootopia***.

No one enjoyed all four movies. How many people enjoyed at least one of *Get Out* and *Arrival*?"

Going Further

- The pigeonhole principle can be used to prove a *ton* of amazing theorems. Here's a sampler:
  - There is always a way to fairly split rent among multiple people, even if different people want different rooms. *(Sperner's lemma)*
  - You and a friend can climb any mountain from two different starting points so that the two of you maintain the same altitude at each point in time. *(Mountain-climbing theorem)*
  - If you model coffee in a cup as a collection of infinitely many points and then stir the coffee, some point is always where it initially started. *(Brower's fixed-point theorem)*
  - A complex process that doesn't parallelize well must contain a large serial subprocess. *(Mirksy's theorem)*
  - Any positive integer *n* has a nonzero multiple that can be written purely using the digits 1 and 0. *(Doesn't have a name, but still cool!)*