Graph Theory

Part Three

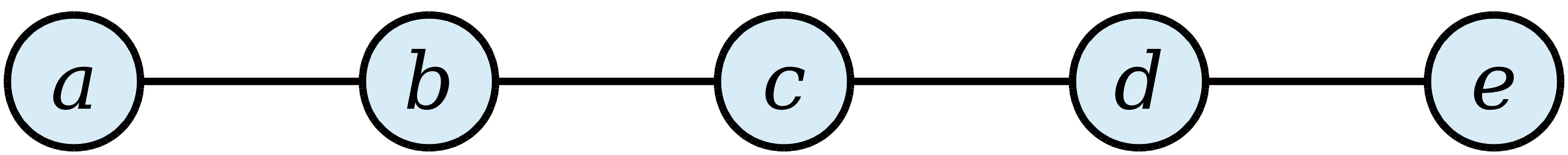
Outline for Today

- ***The Pigeonhole Principle***
  - A simple yet surprisingly effective fact.
- ***Graph Theory Party Tricks***
  - Cool tricks to try at your next group meeting.
- ***A Little Movie Puzzle***
  - Who watched what?

Recap from Last Time

Adjacency and Reachability

- Two nodes in a graph are called ***adjacent*** if there's an edge between them.
- Two nodes in a graph are called ***reachable*** if there's a path between them.

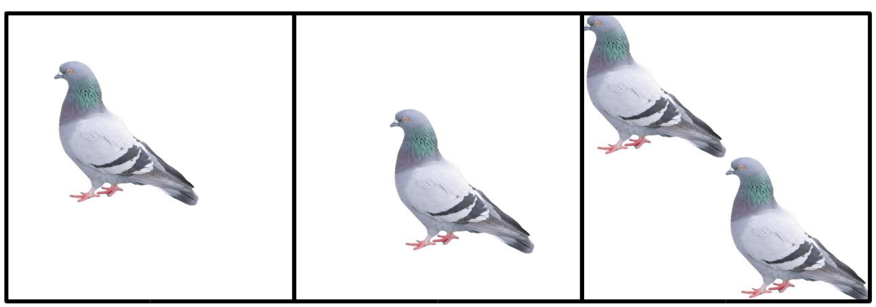


New Stuff!

The Pigeonhole Principle

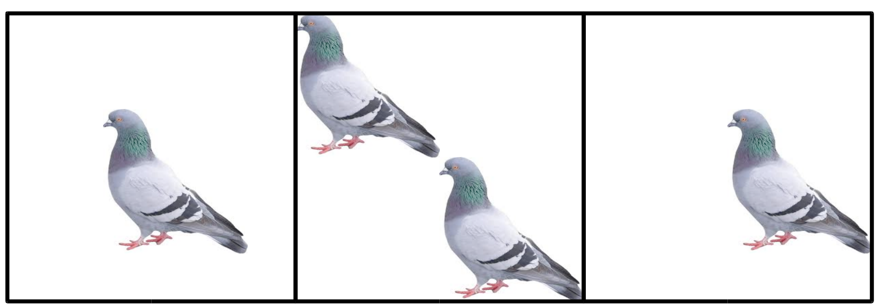
The Pigeonhole Principle

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.



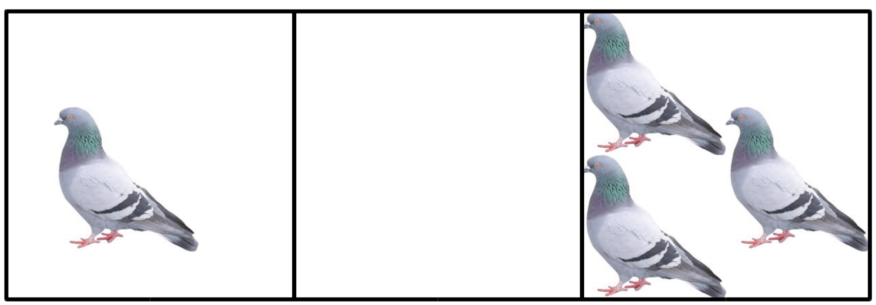
The Pigeonhole Principle

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.



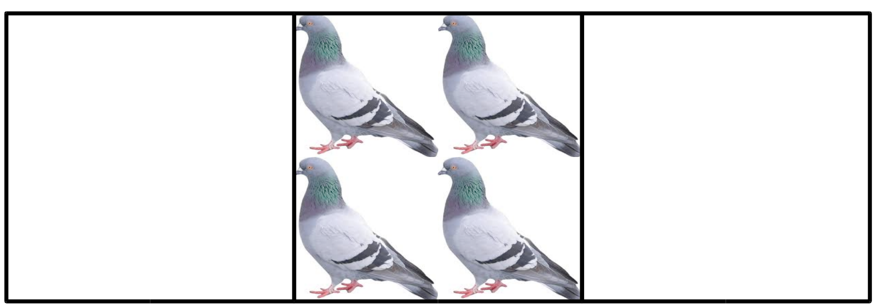
The Pigeonhole Principle

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.



$m = 4, n = 3$

Thanks to Amy Liu for this awesome drawing!

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.

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.

Let A and B be finite sets (sets whose cardinalities are natural numbers) and assume $|A| > |B|$. Which of the following statements are true?

(1) If $f : A \to B$, then f is injective.
(2) If $f : A \to B$, then f is not injective.
(3) If $f : A \to B$, then f is surjective.
(4) If $f : A \to B$, then f is not surjective.

Answer at

https://cs103.stanford.edu/pollev

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

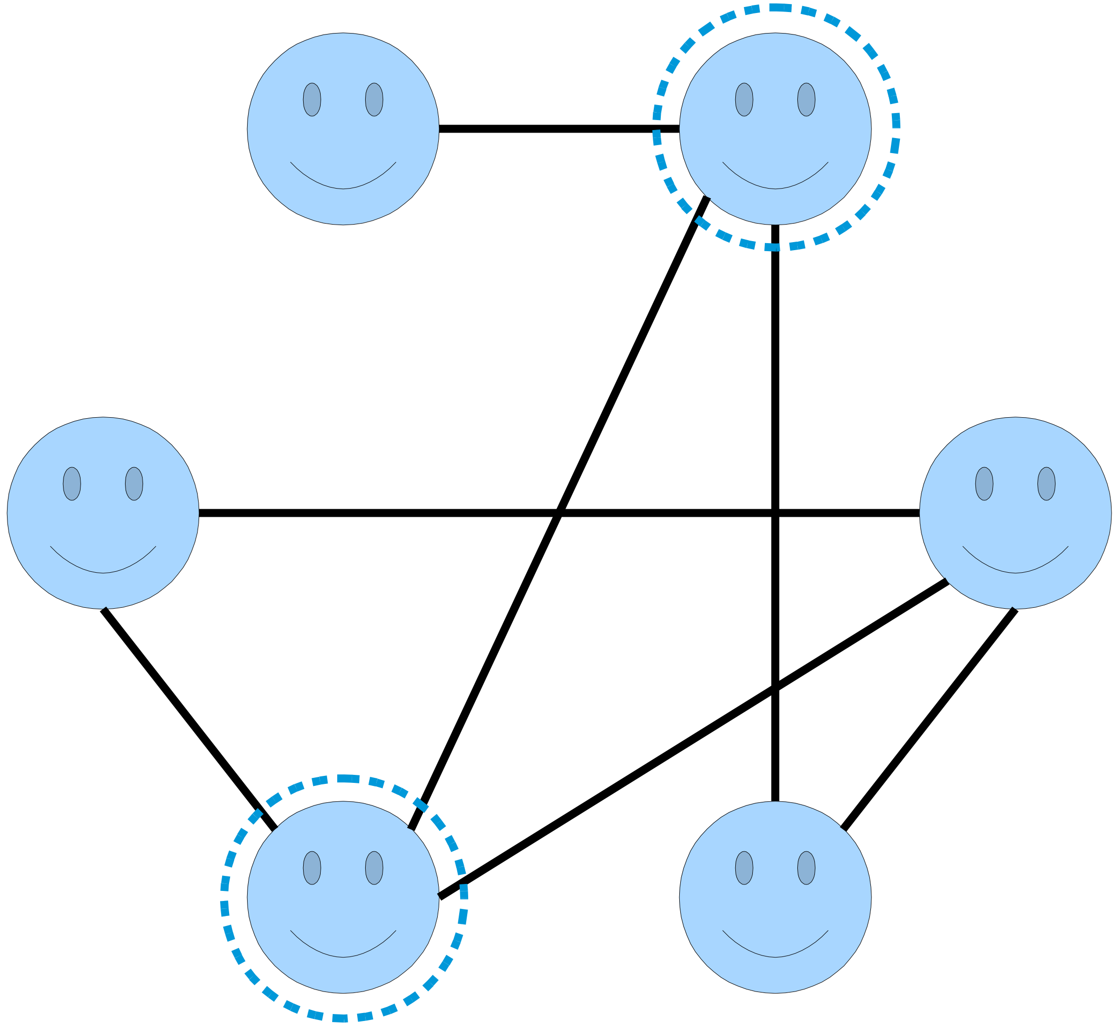
$$m = x_1 + x_2 + \ldots + x_n.$$

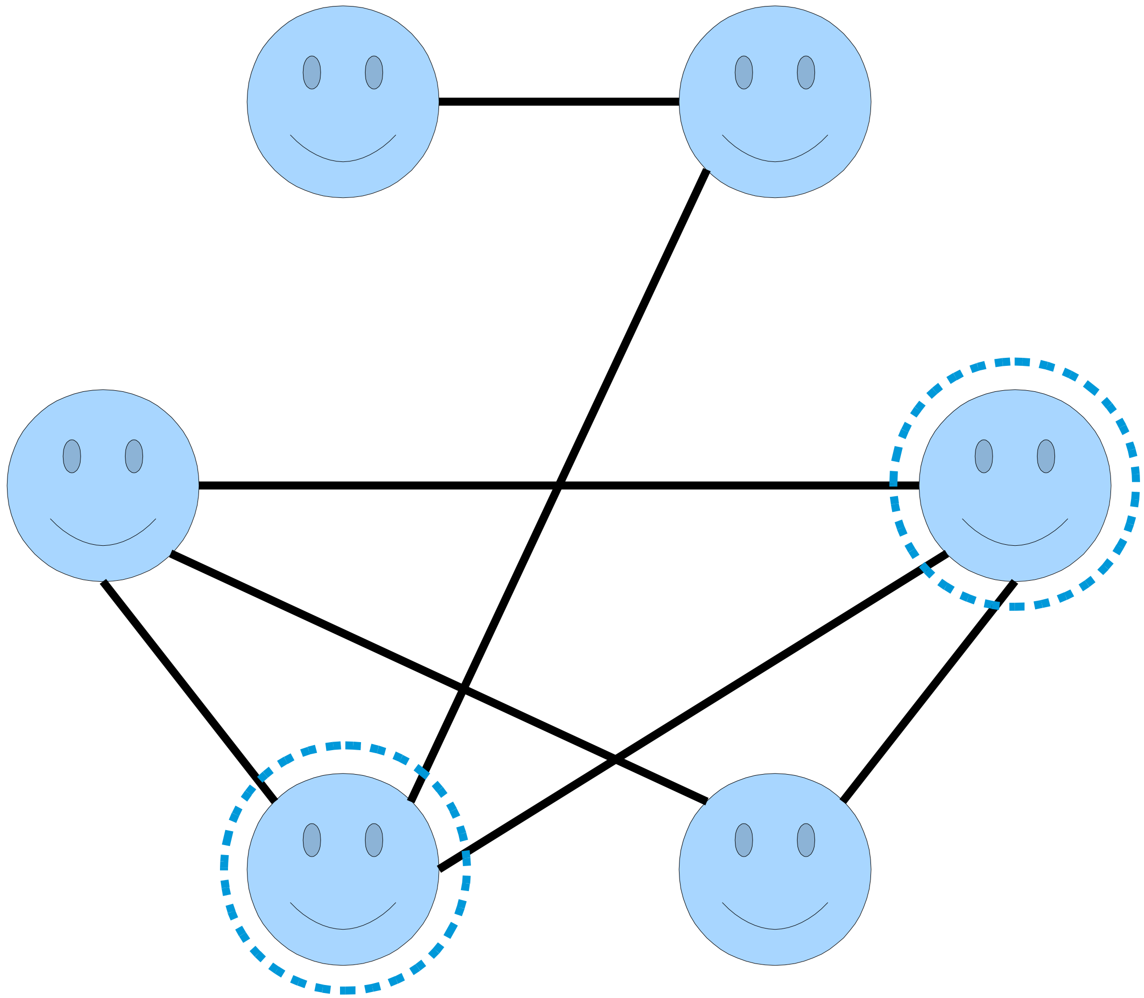
Since each bin has at most one object in it, we know $x_i \le 1$ for each i. This means that

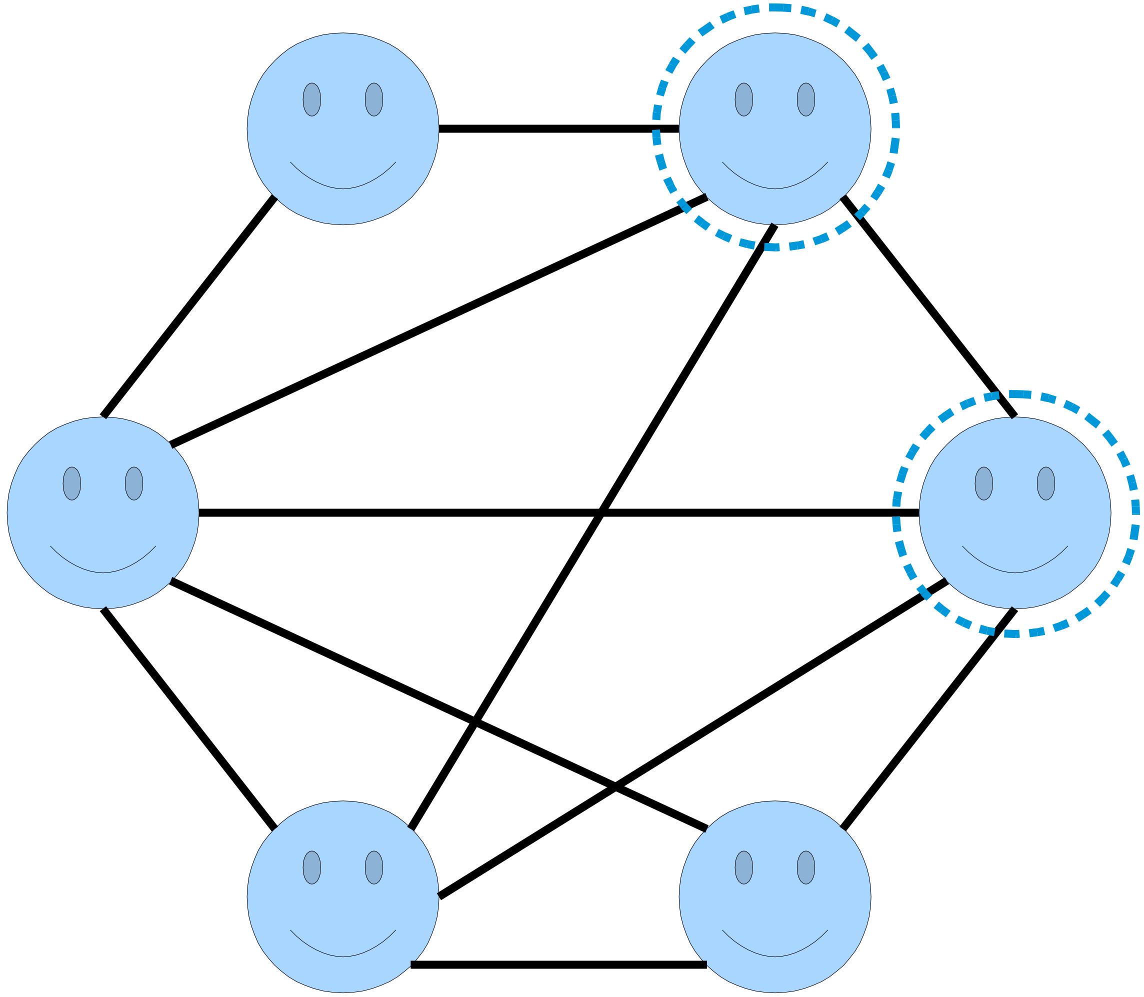
$$\begin{aligned} m &= x_1 + x_2 + \ldots + x_n \\ &\le 1 + 1 + \ldots + 1 \quad \textit{(n times)} \\ &= n. \end{aligned}$$

This means that $m \le n$, 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

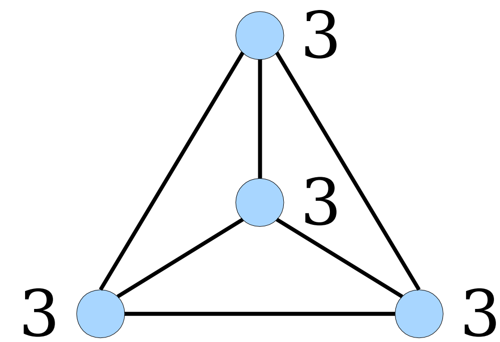




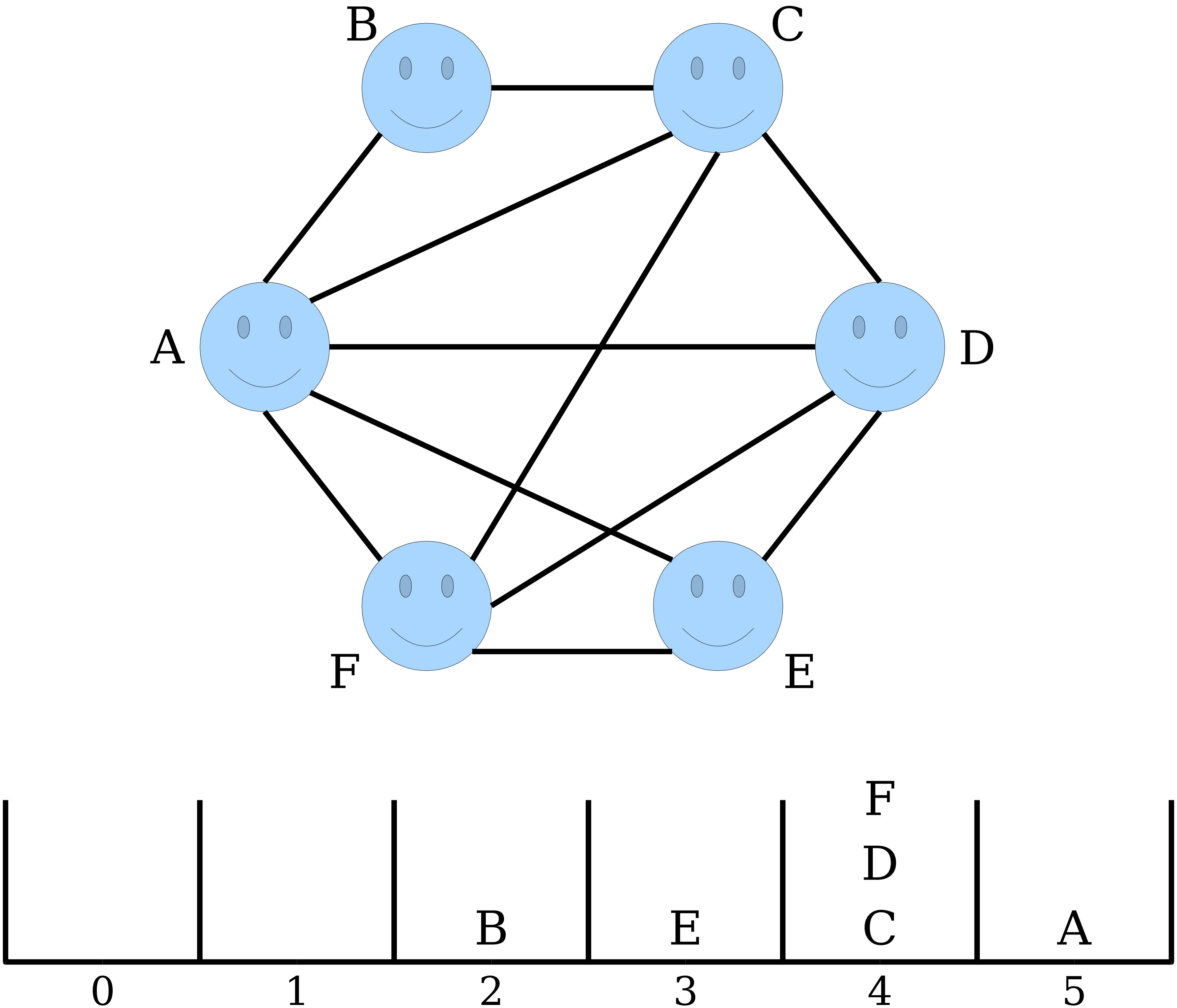


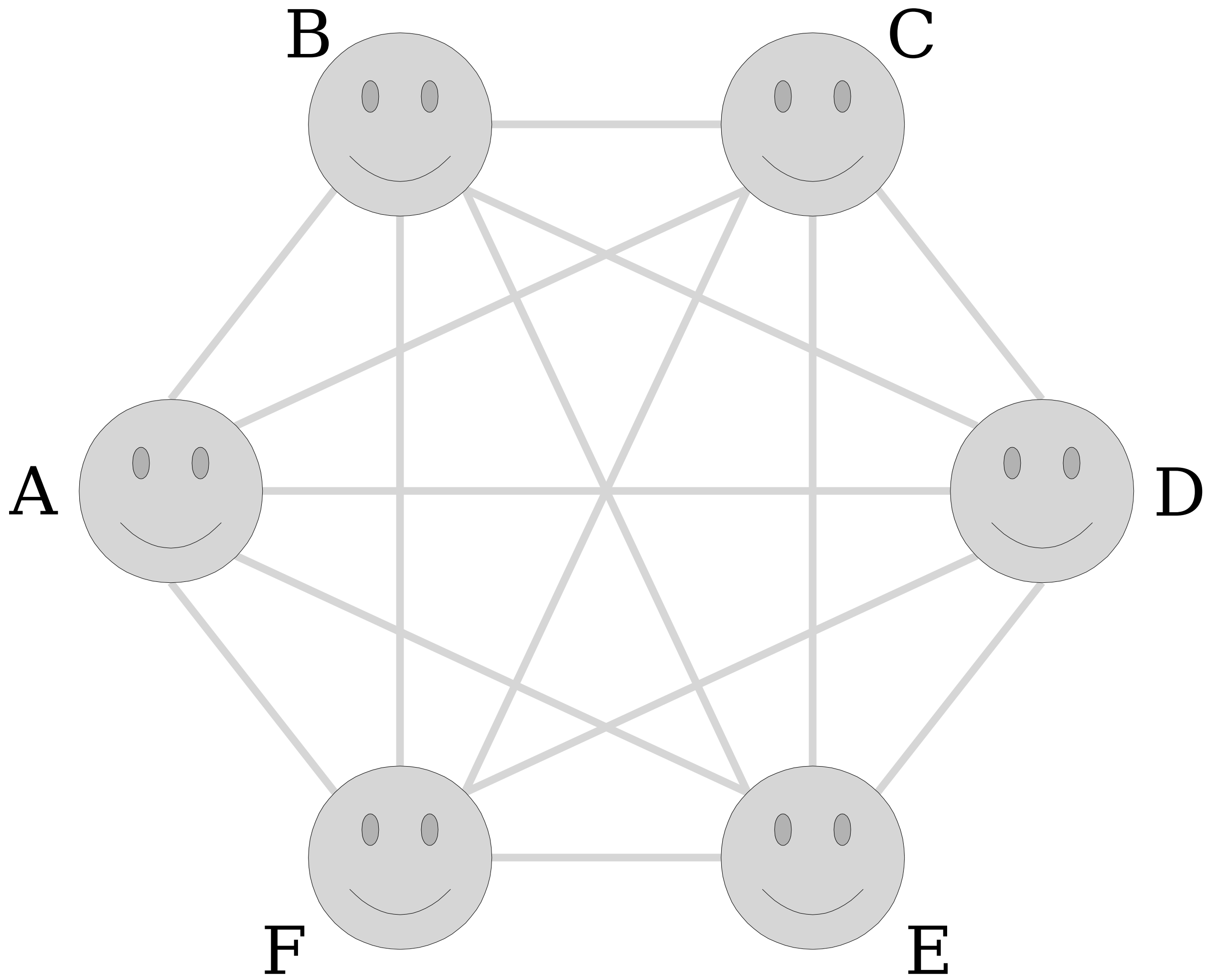
Degrees

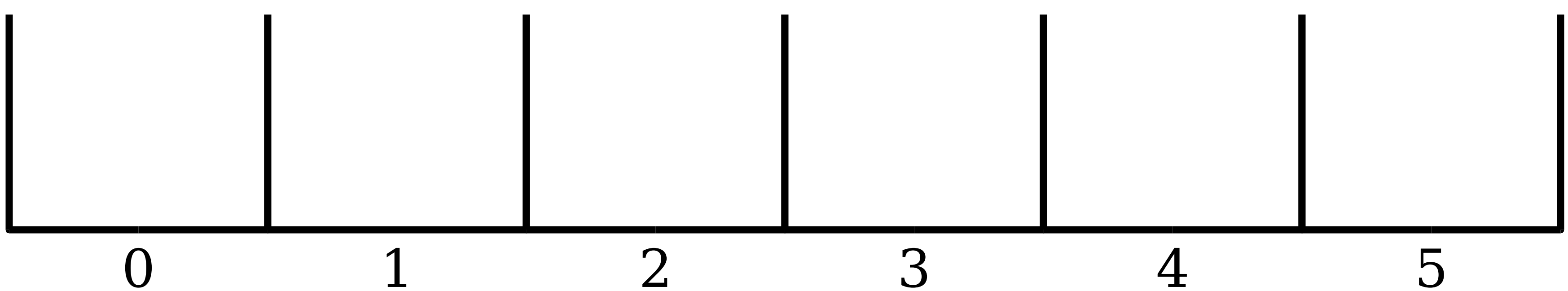
- The ***degree*** of a node *v* in a graph is the number of nodes that *v* is adjacent to.

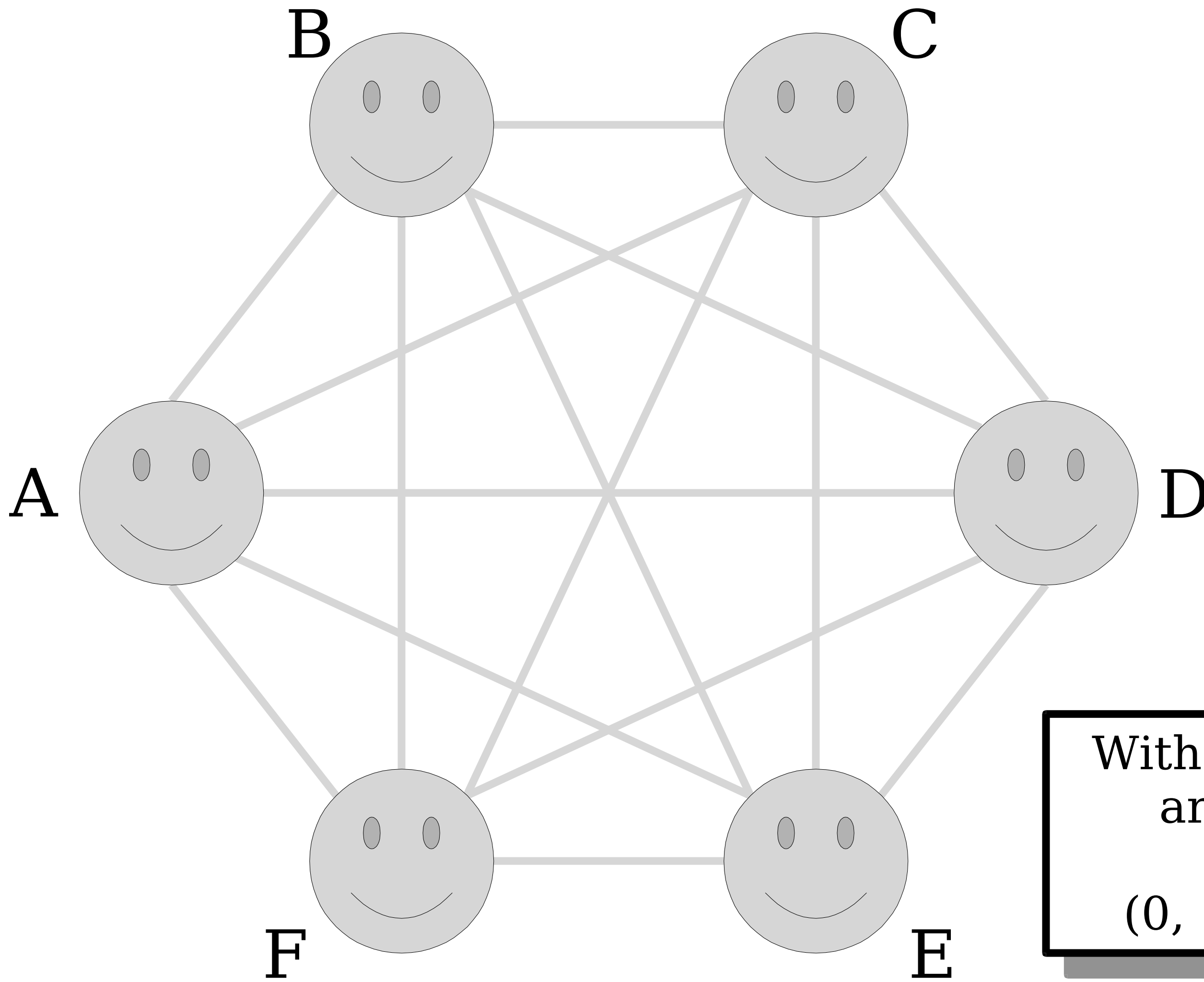


- ***Theorem:*** Every graph with at least two nodes has at least two nodes with the same degree.
  - Equivalently: at any party with at least two people, there are at least two people with the same number of friends at the party.

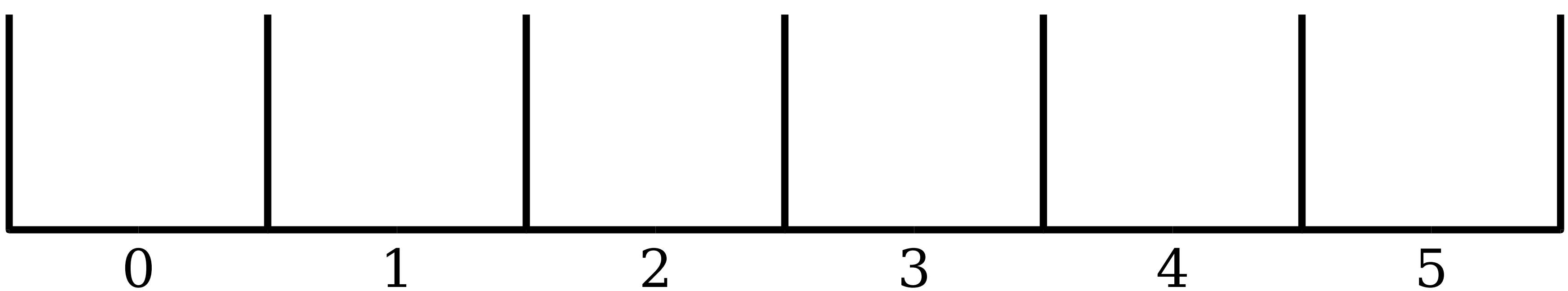

B
C
A
D
F
E
F
D
B
E
C
A
0
1
2
3
4
5

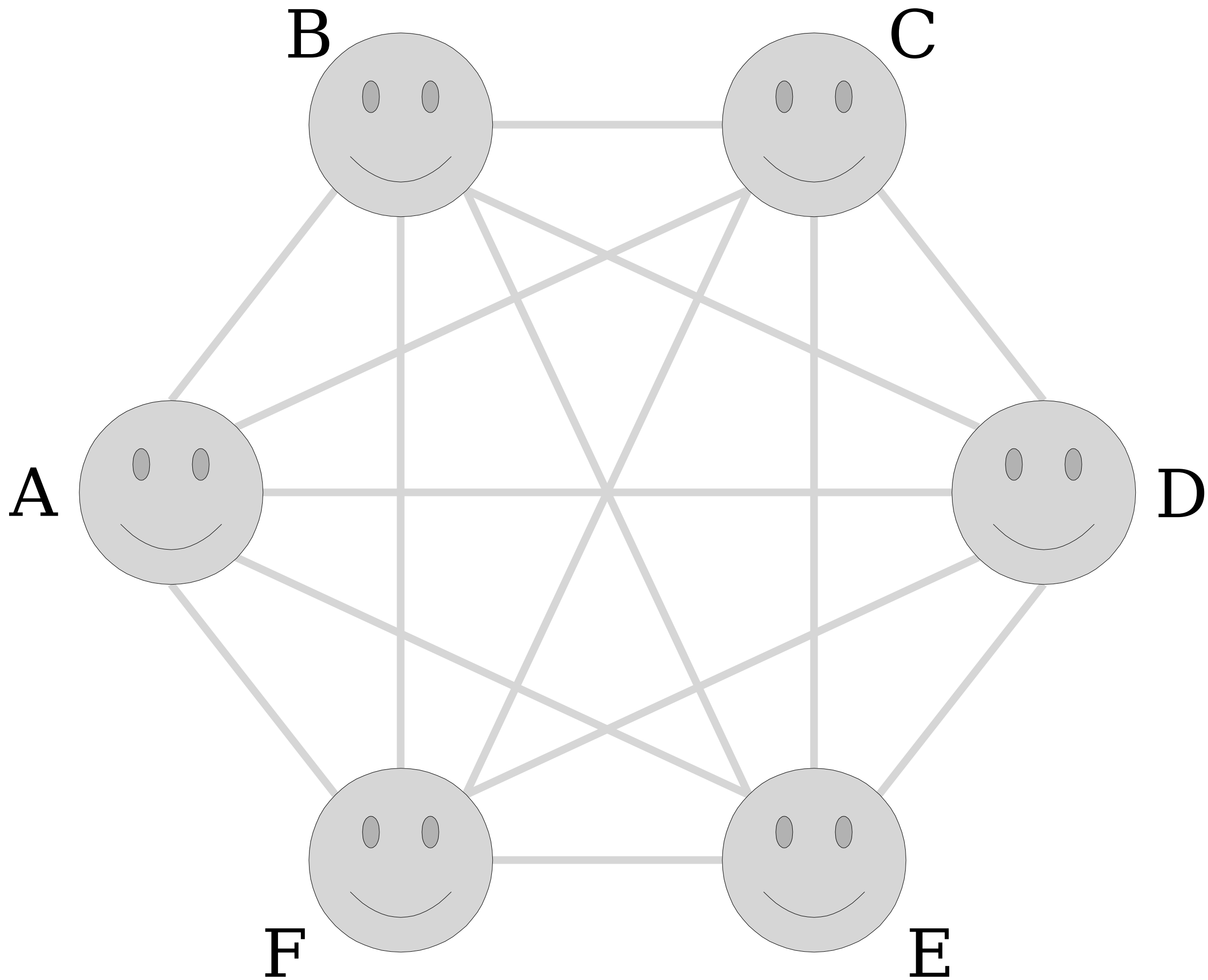

B
C
A
D
F
E

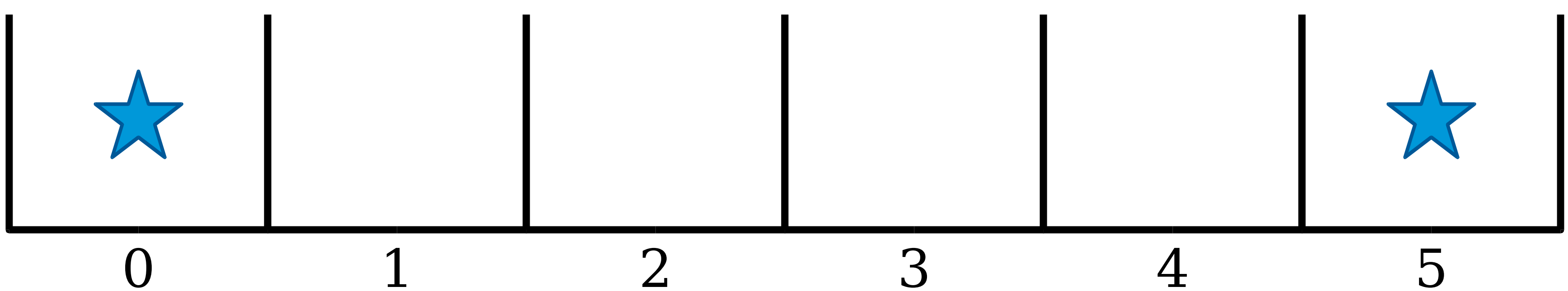

0
1
2
3
4
5

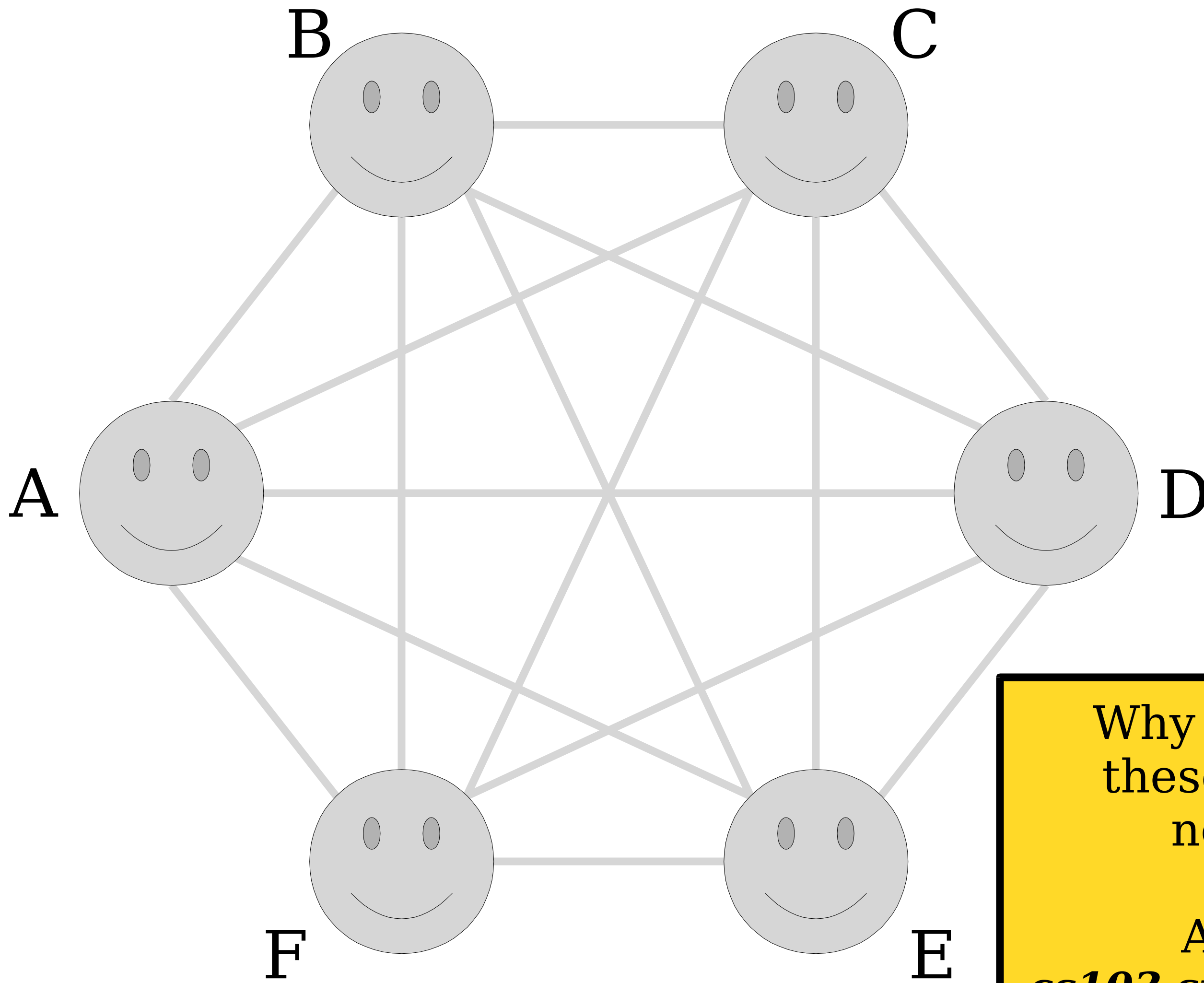


With *n* nodes, there are *n* possible degrees (0, 1, 2, …, *n* – 1)



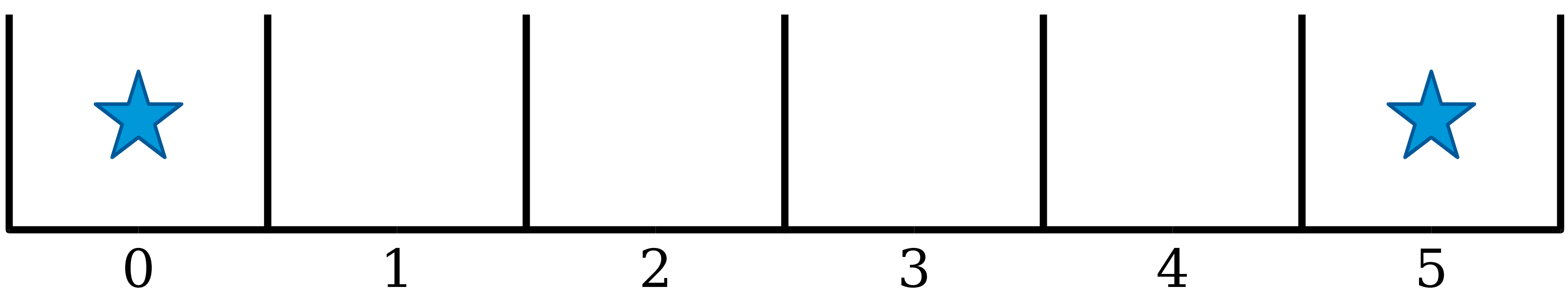

B
C
A
D
F
E

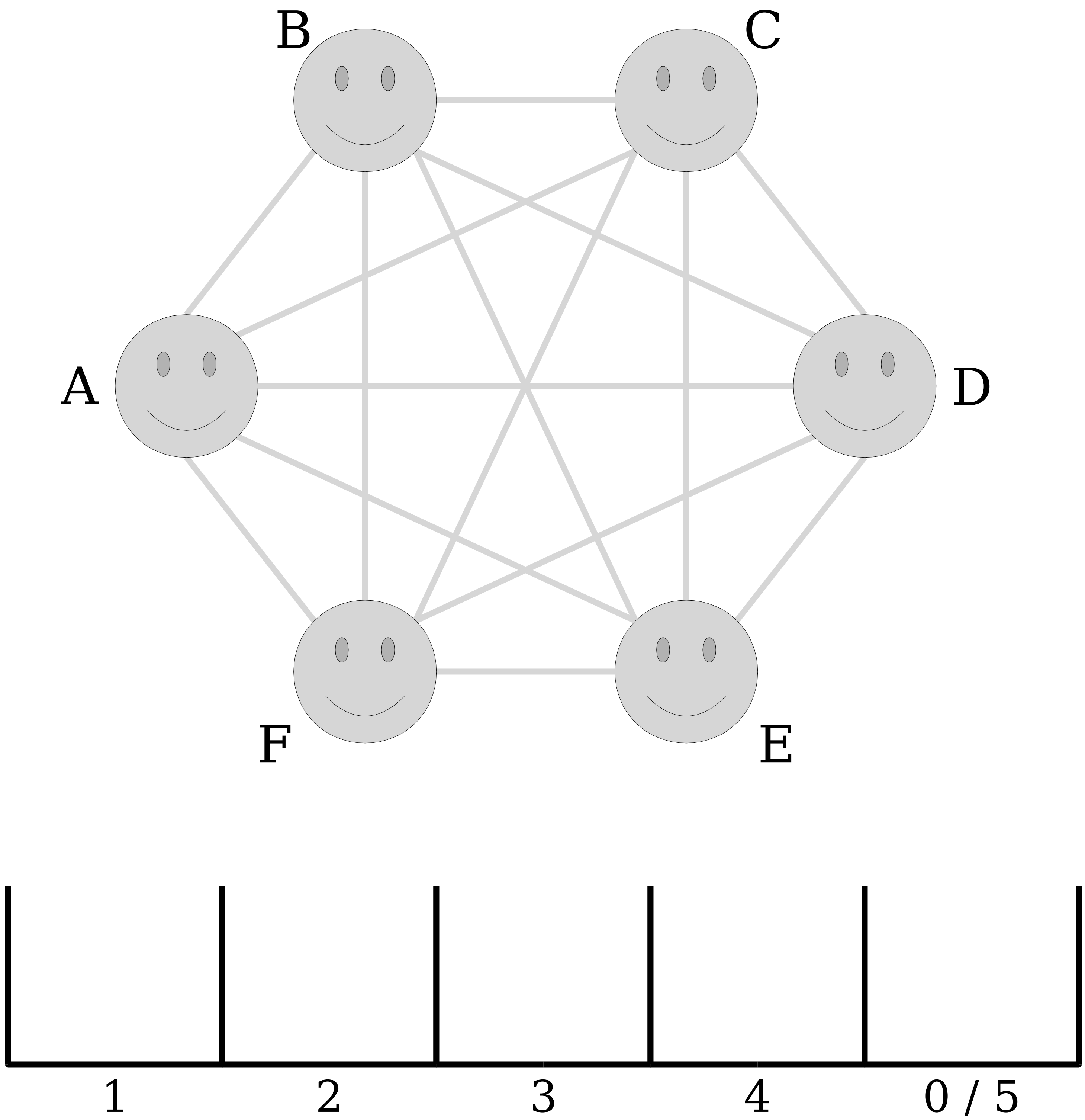

0
1
2
3
4
5



Why can't both of these buckets be nonempty?

Answer at ***cs103.stanford.edu/pollev***




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, …, 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, $0, 1, 2, \ldots,$ 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.)

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, …, $n - 1$.

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 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. ■

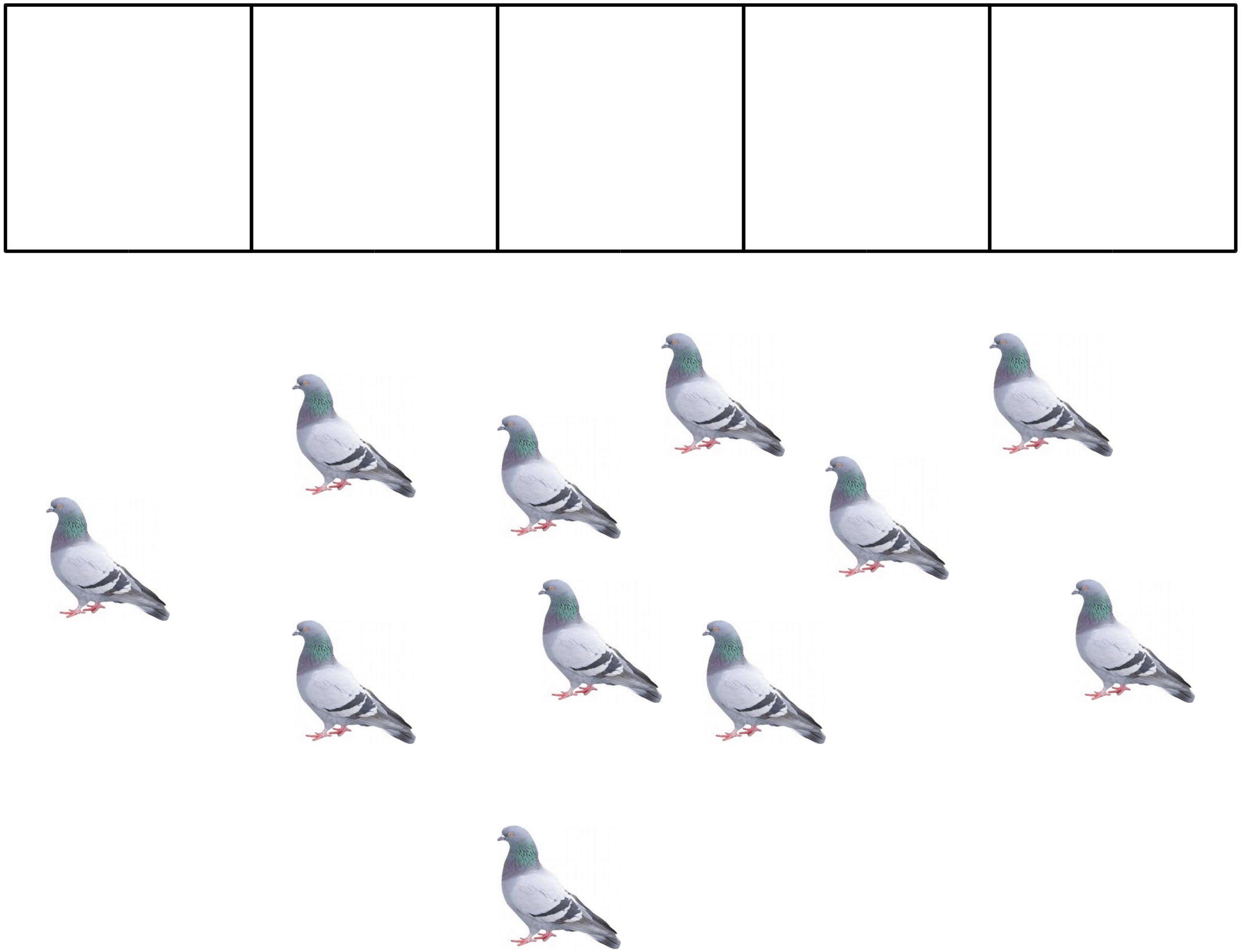
Theorem: In any graph with at least two nodes, there are at least two nodes of the same degree.

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.

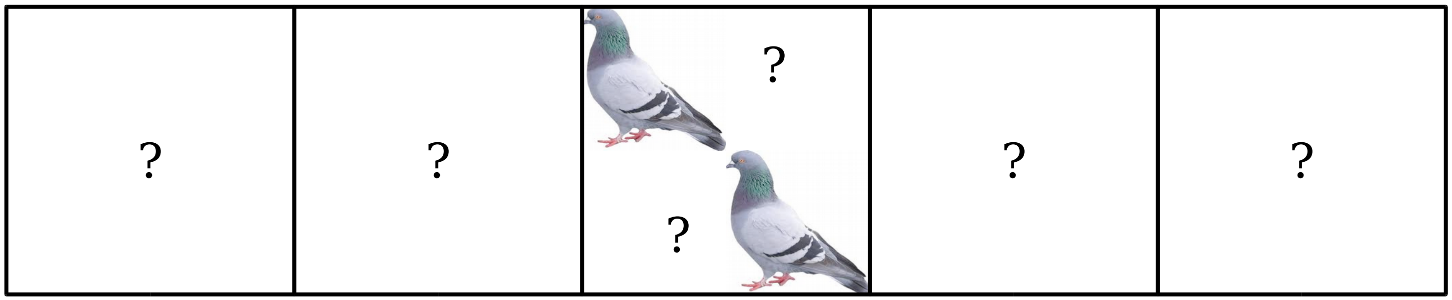
We have reached a contradiction, so our assumption must have been wrong. Thus if G is a graph with at least two nodes, G must have at least two nodes of the same degree. ■

The Generalized Pigeonhole Principle

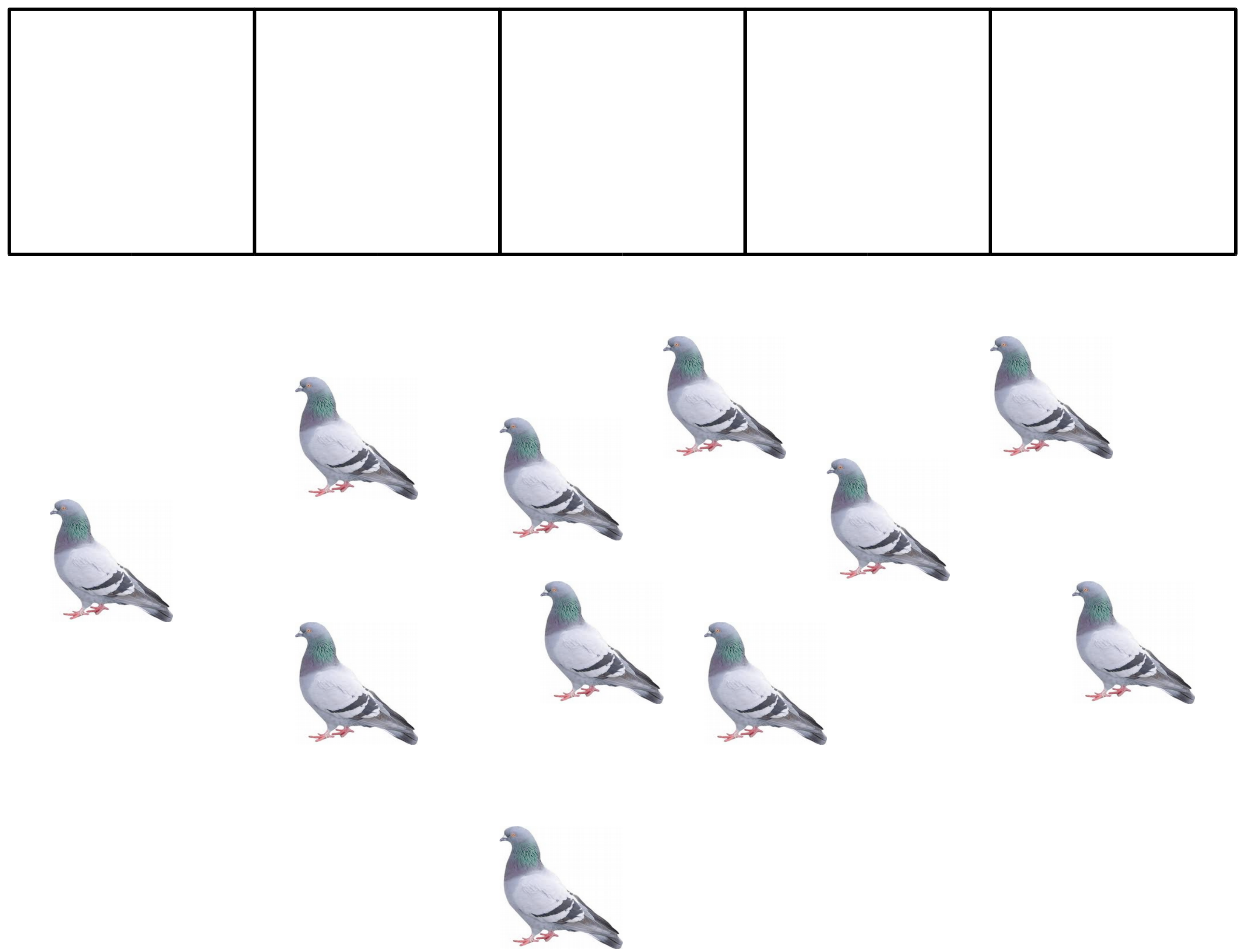
The Pigeonhole Principle



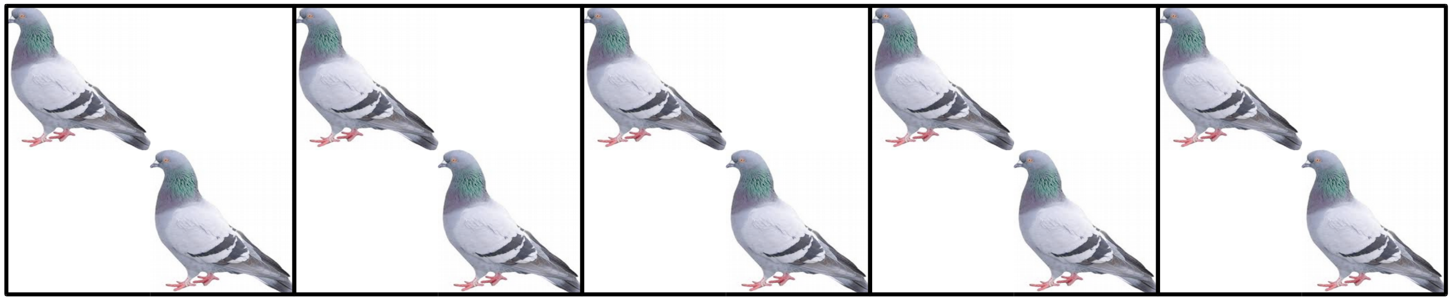
The Pigeonhole Principle



The Pigeonhole Principle

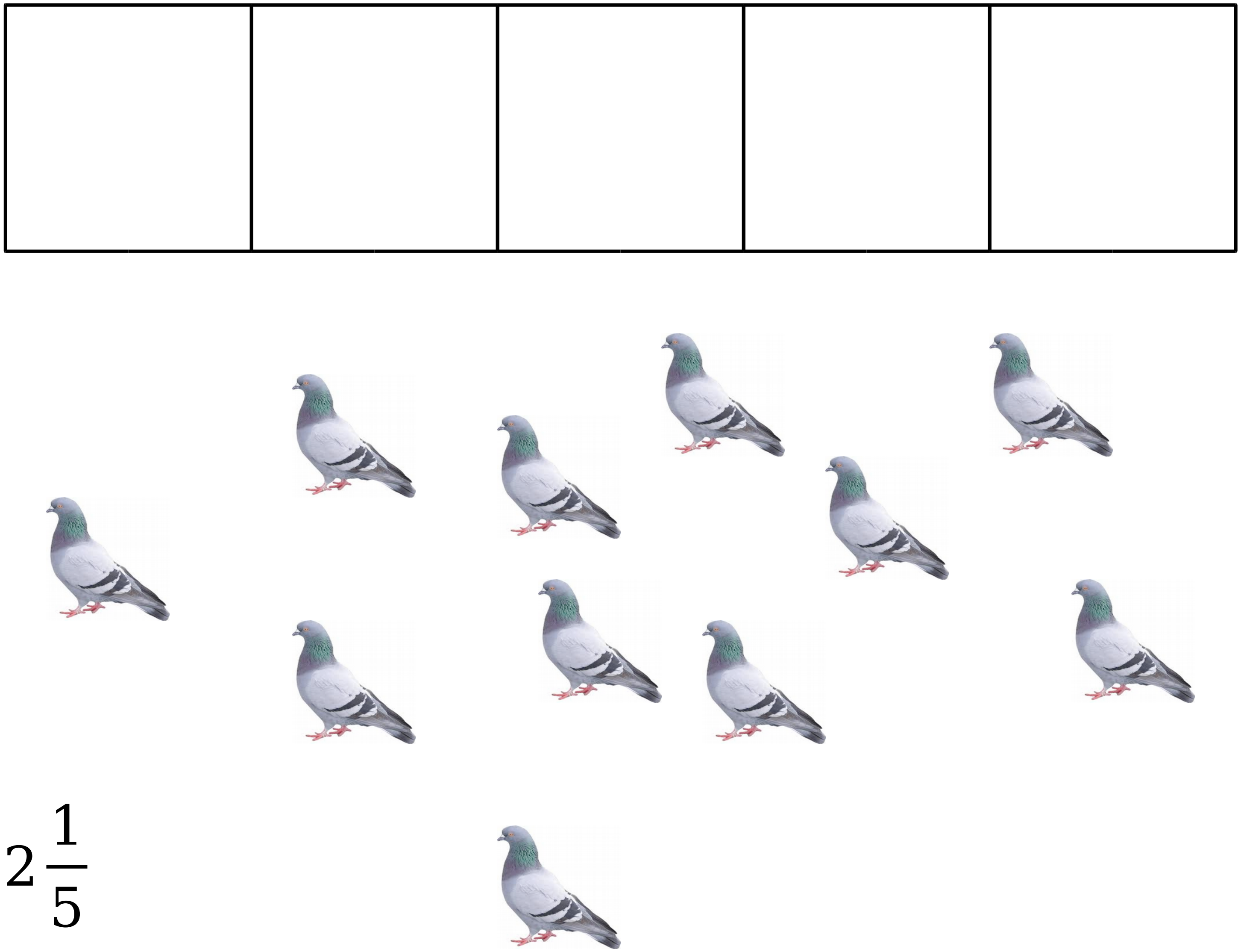


The Pigeonhole Principle



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."

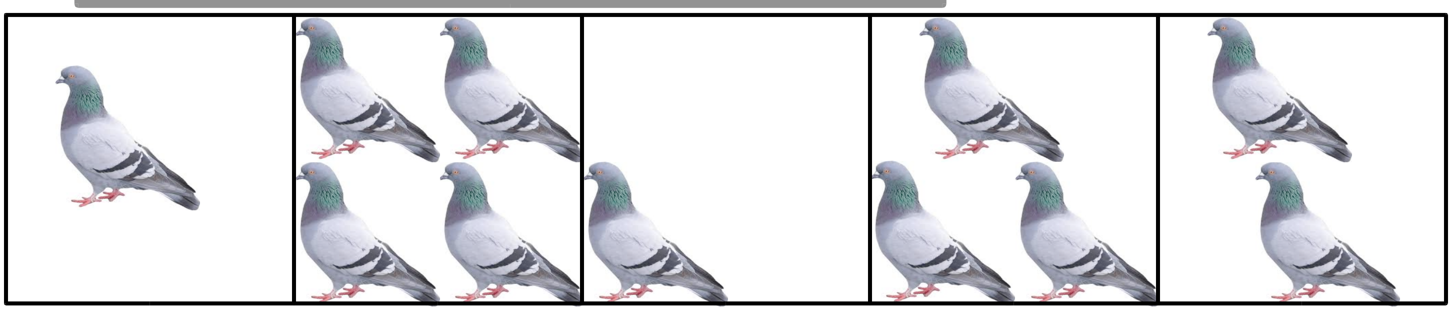
$m = 11$
$n = 5$

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

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."



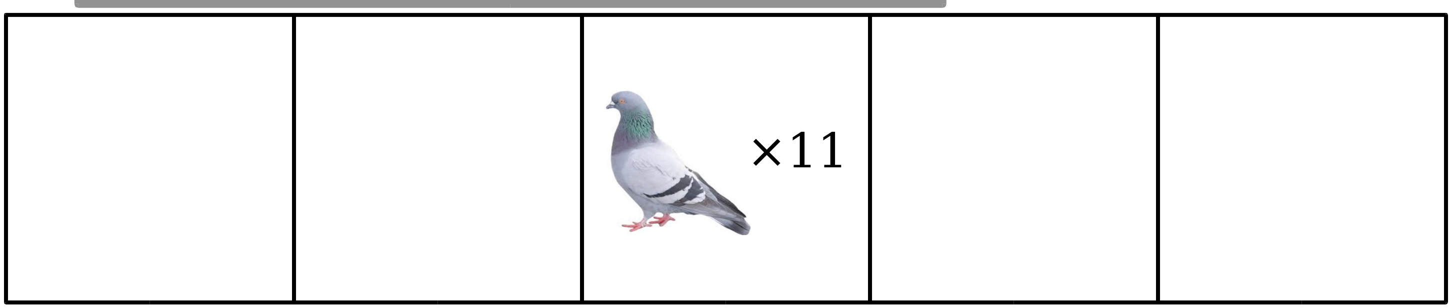
$m = 11$
$n = 5$

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

A More General Version

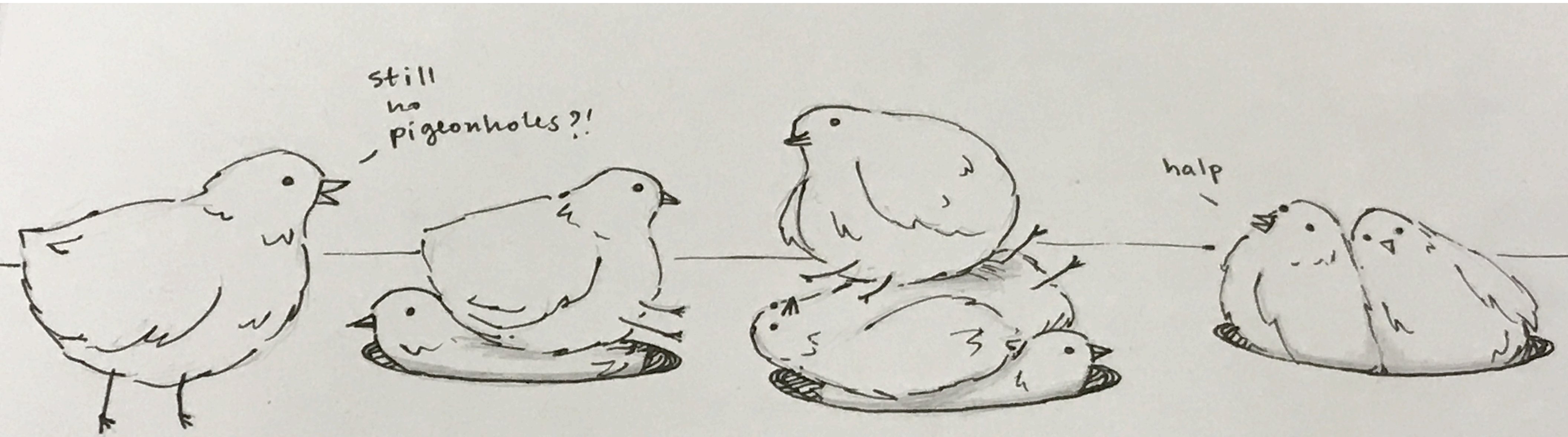
- The ***generalized pigeonhole principle*** says that if you distribute m objects into n bins, then
  - some bin will have at least $\lceil m/n \rceil$ objects in it, and
  - some bin will have at most $\lfloor m/n \rfloor$ objects in it.

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



$m = 11$
$n = 5$

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



$m = 8, n = 3$

Thanks to Amy Liu for this awesome drawing!

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

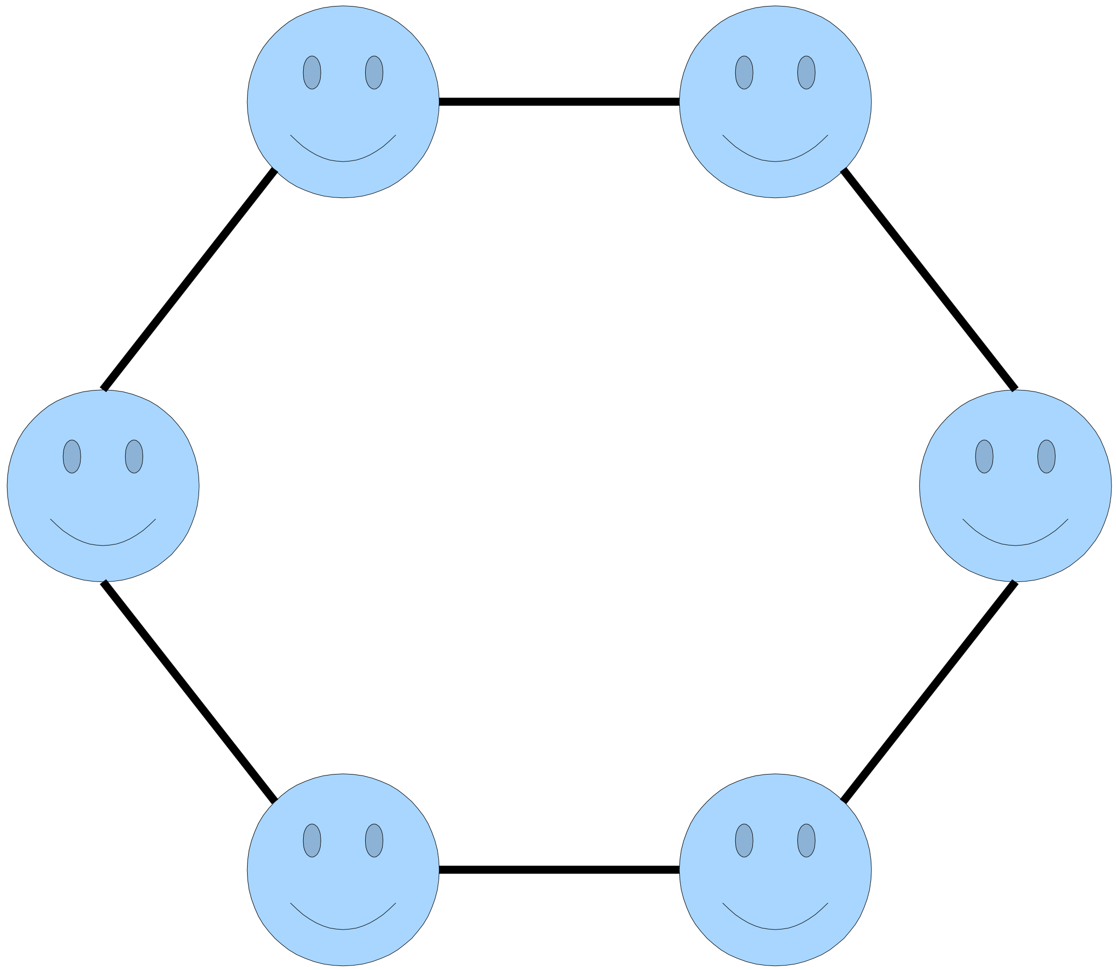
$$\begin{aligned} m &= x_1 + x_2 + \ldots + x_n \\ &< m/n + m/n + \ldots + m/n \quad \textit{(n times)} \\ &= m. \end{aligned}$$

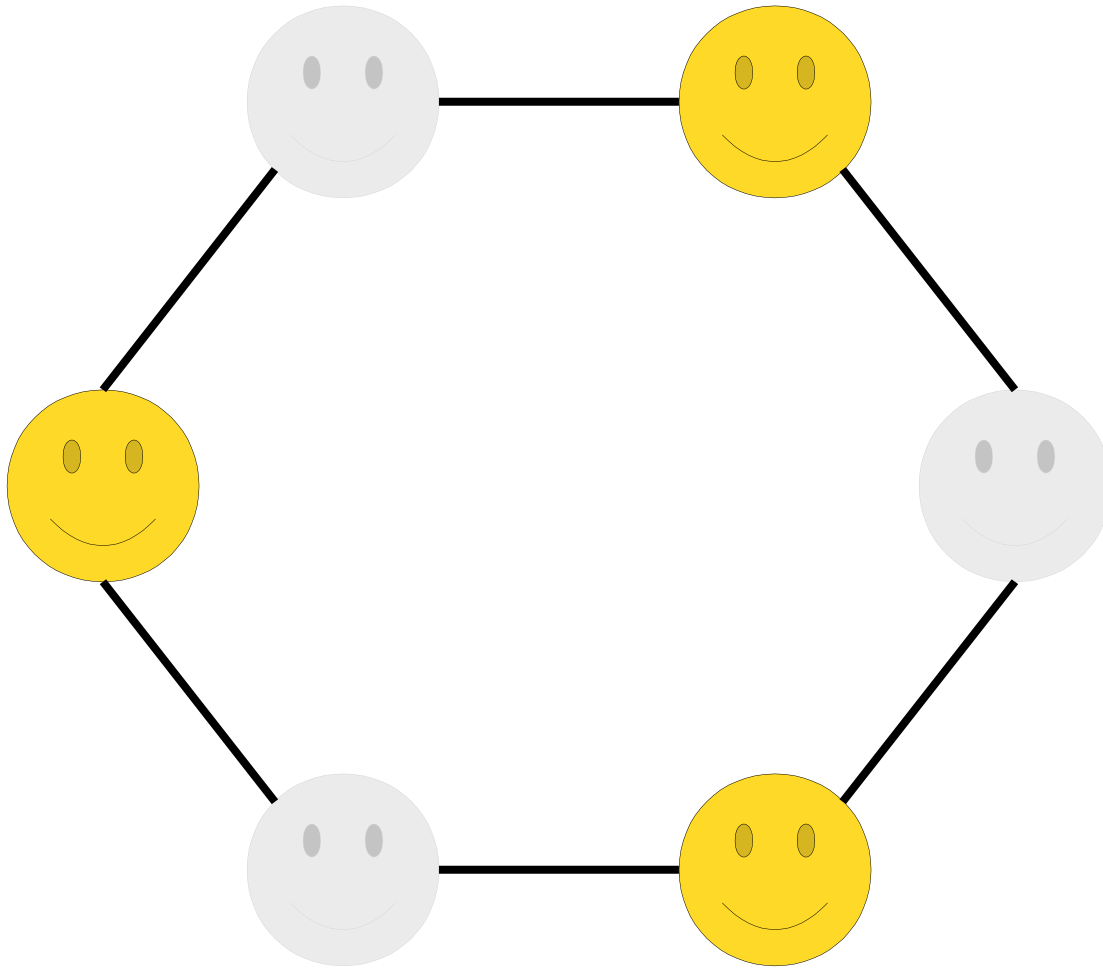
But this means that $m < m$, which is impossible. We have reached a contradiction, so our initial assumption must have been wrong. Therefore, if m objects are distributed into n bins, some bin must contain at least $\lceil m/n \rceil$ objects. ■

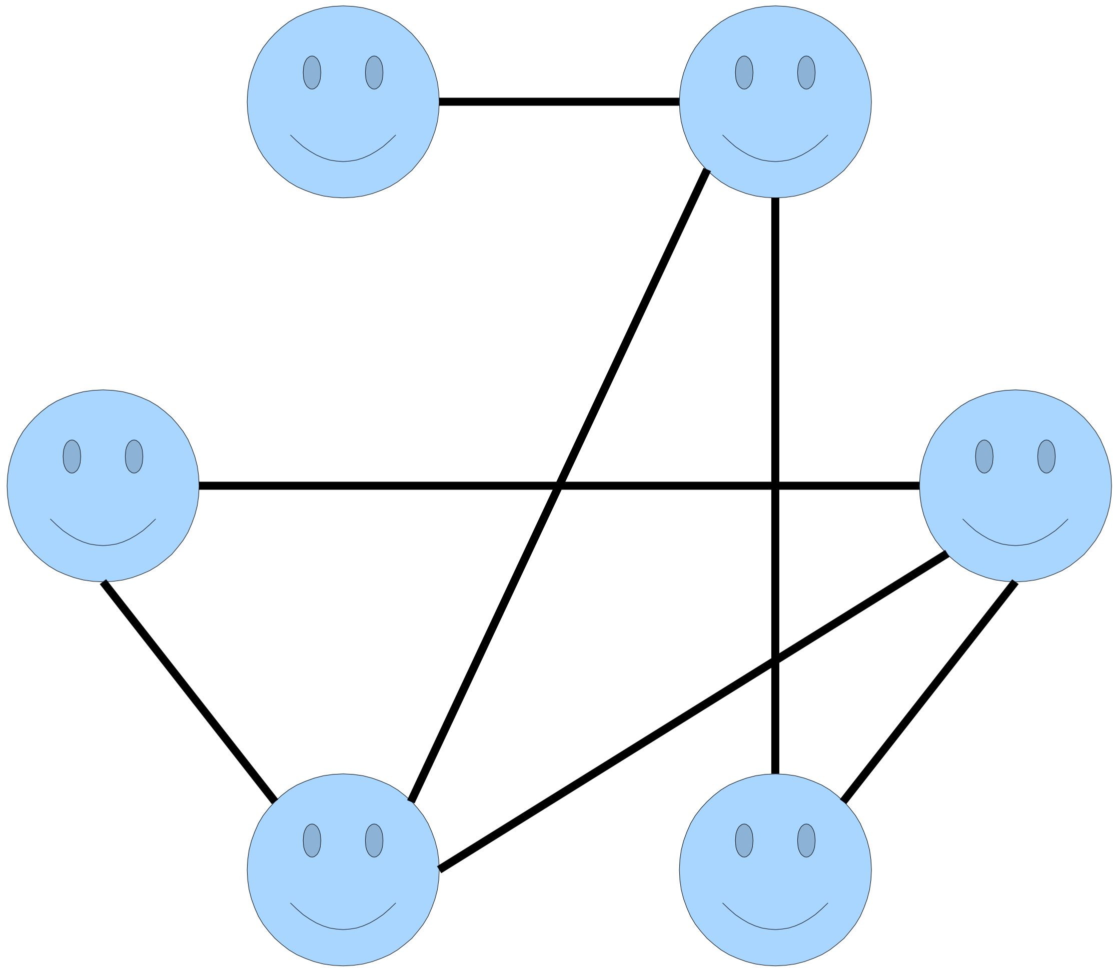
An Application: Friends and Strangers

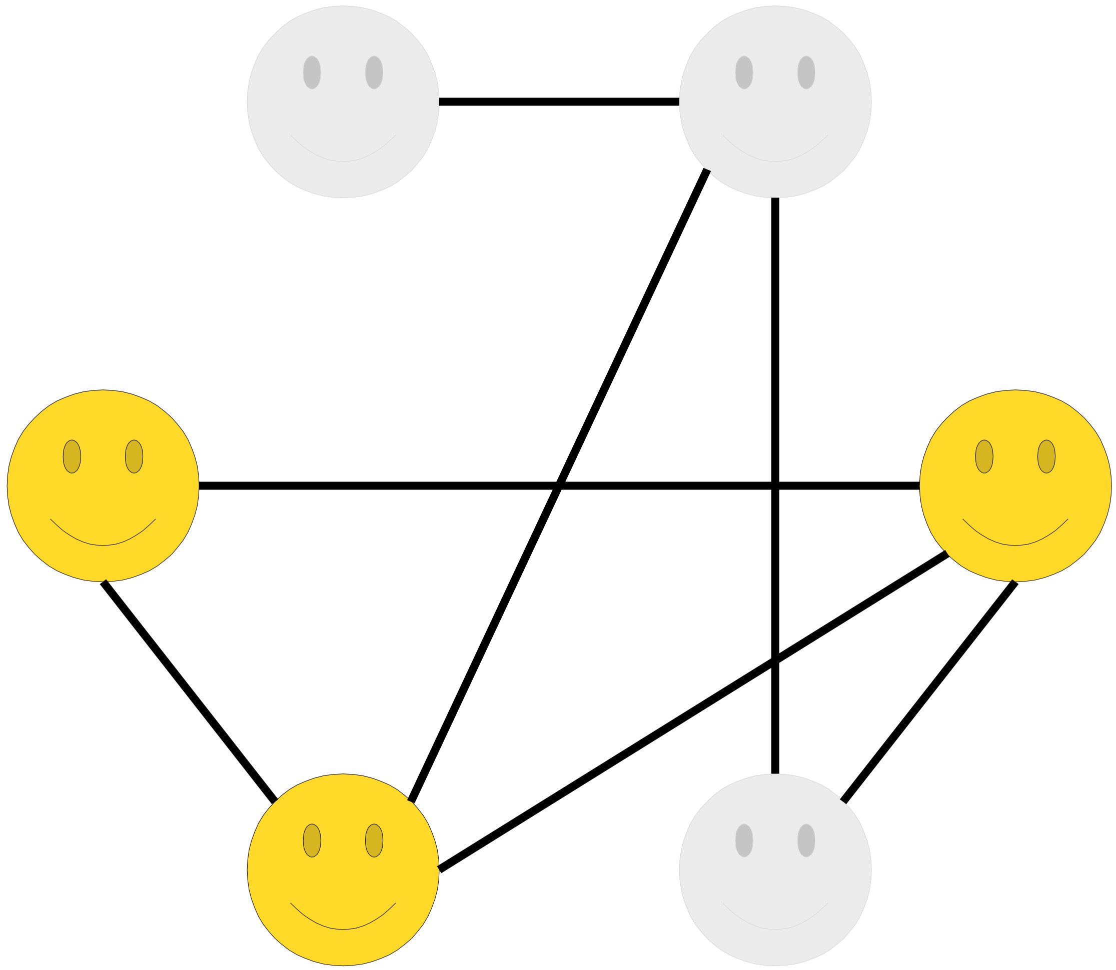
Friends and Strangers

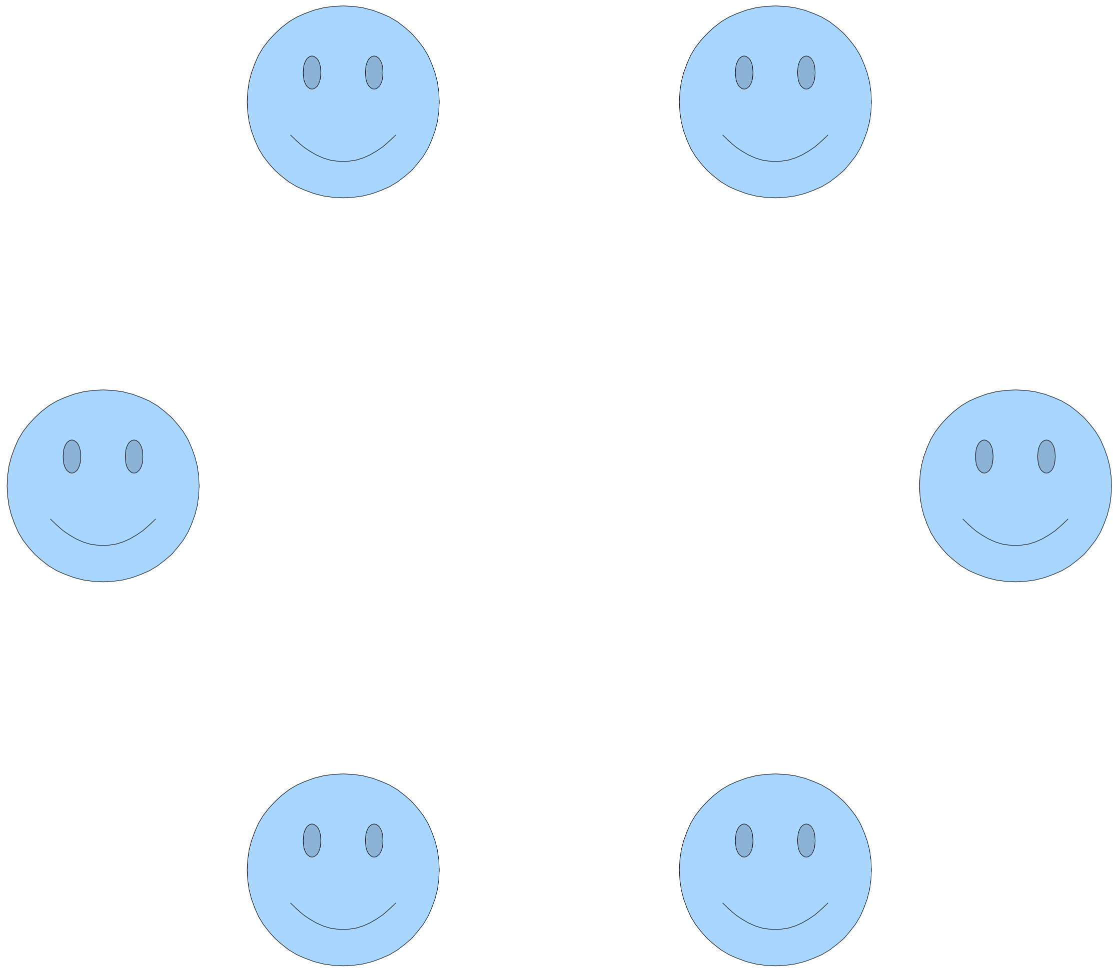
- Suppose you have a party of six people. Each pair of people are either friends (they know each other) or strangers (they do not).
- ***Theorem:*** Any such party must have a group of three mutual friends (three people who all know one another) or three mutual strangers (three people, none of whom know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

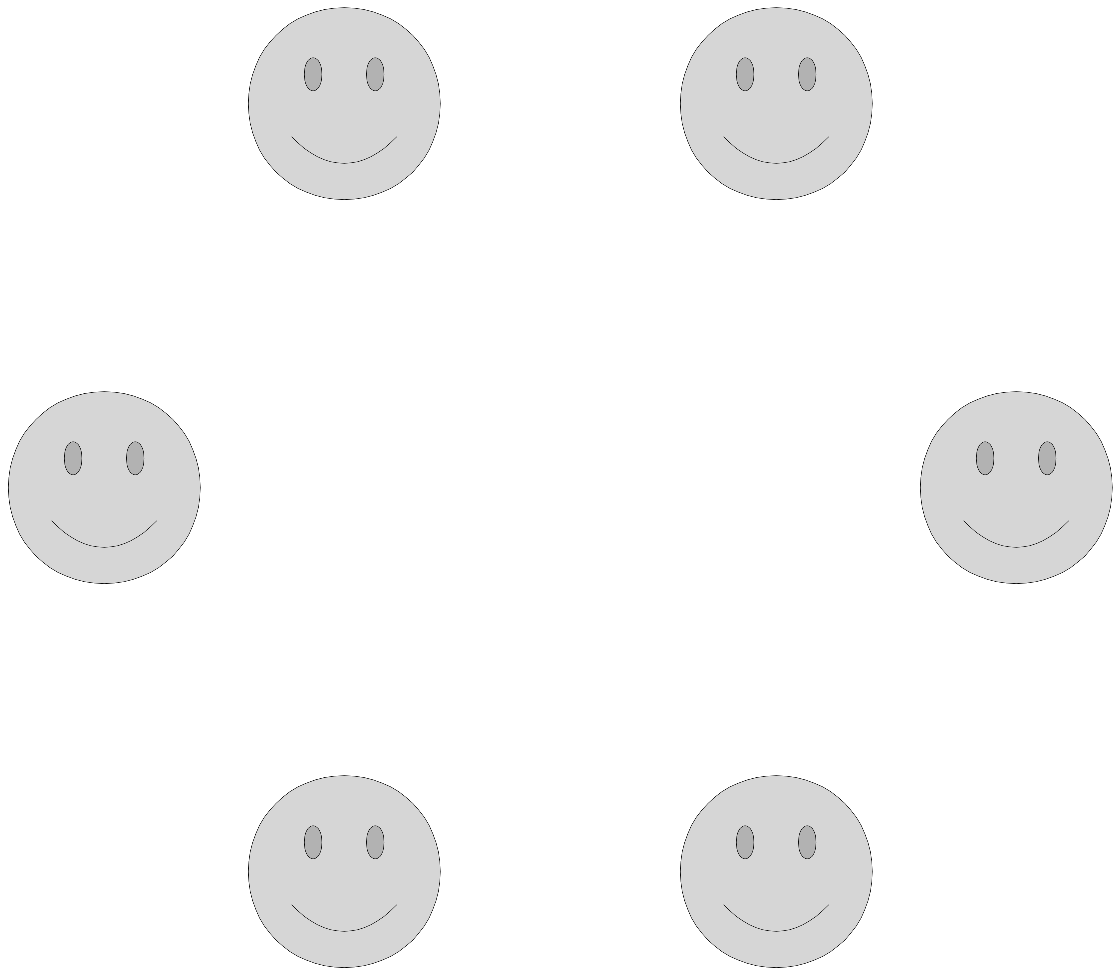


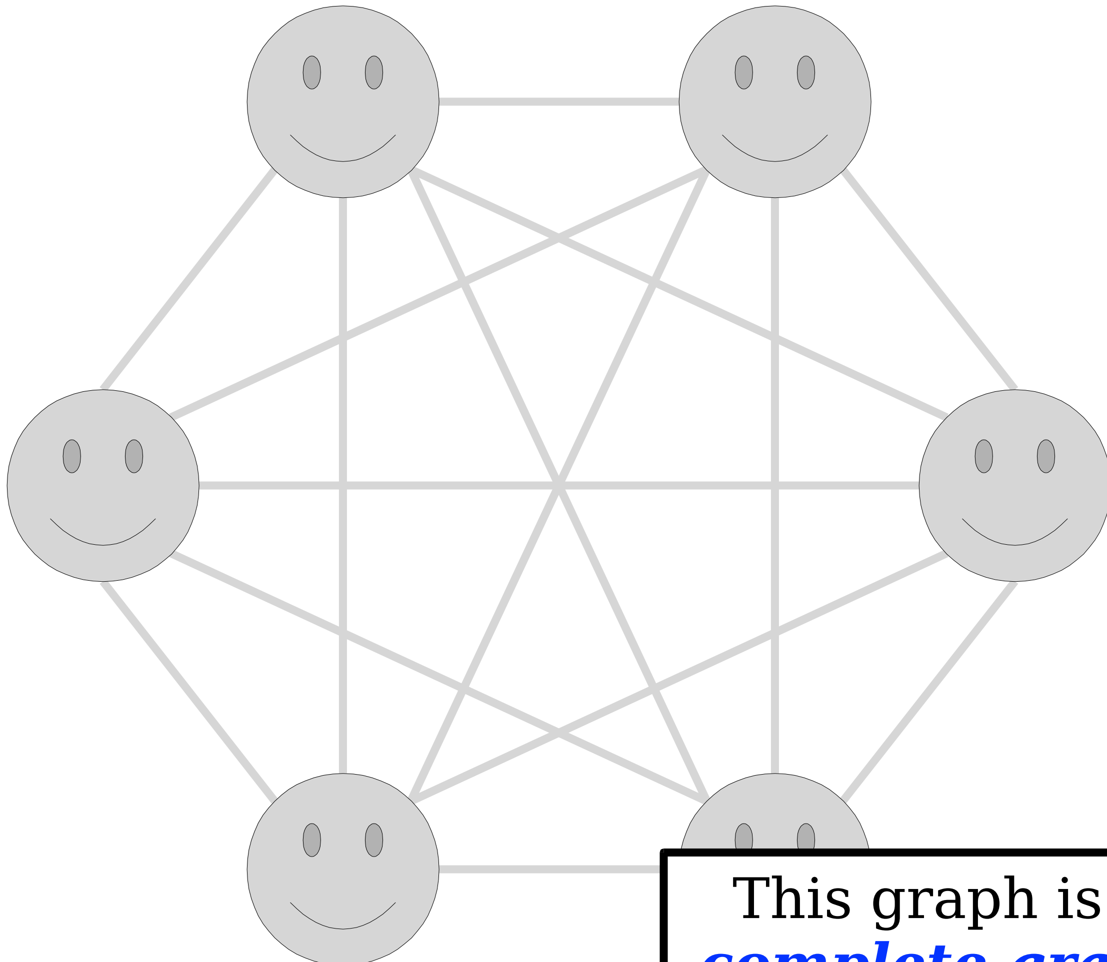




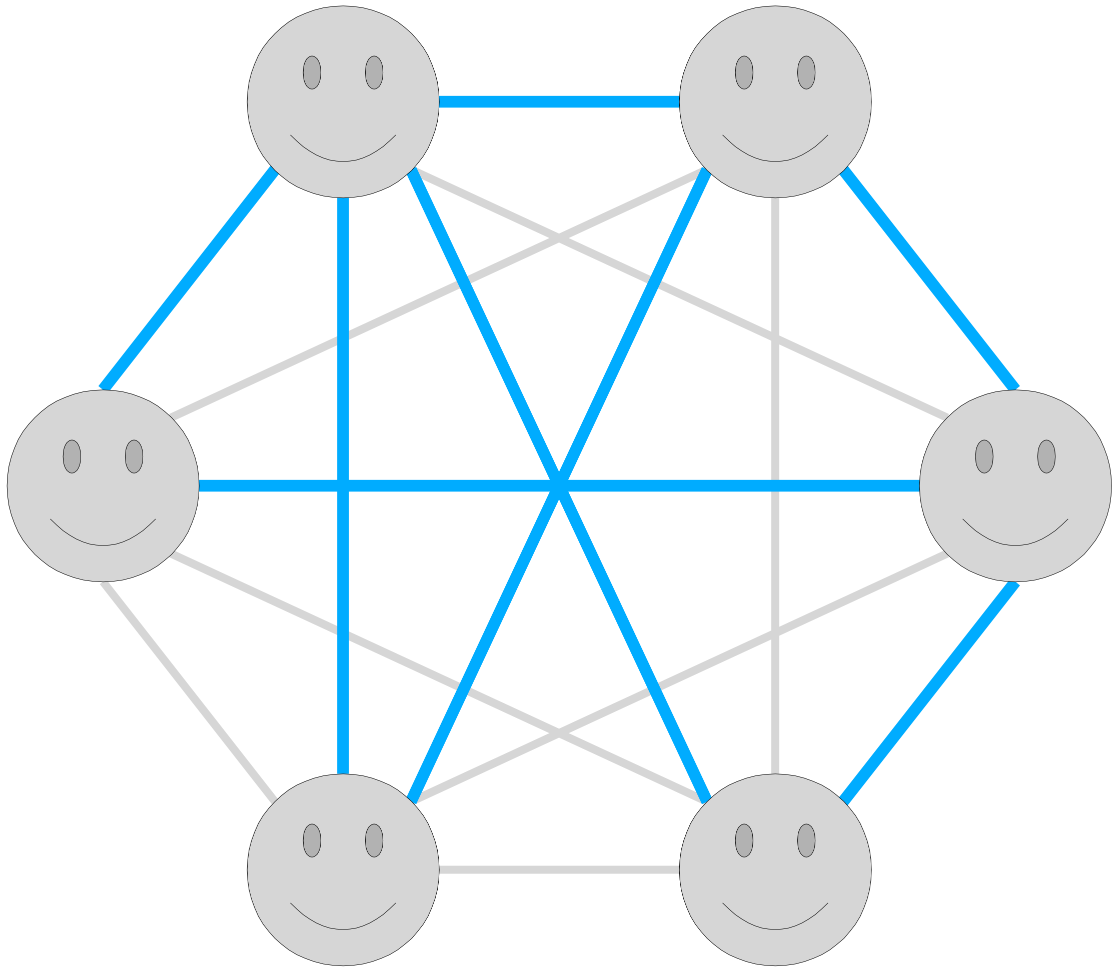


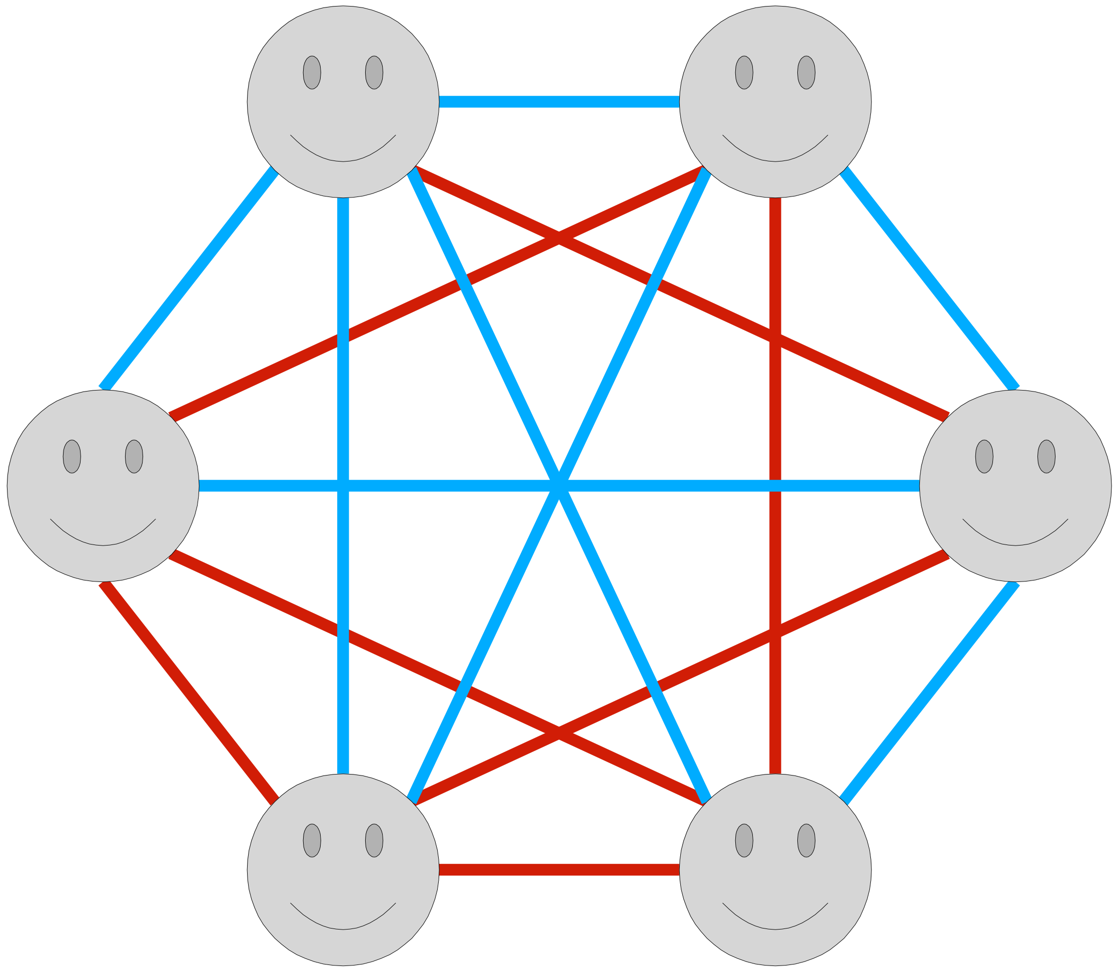


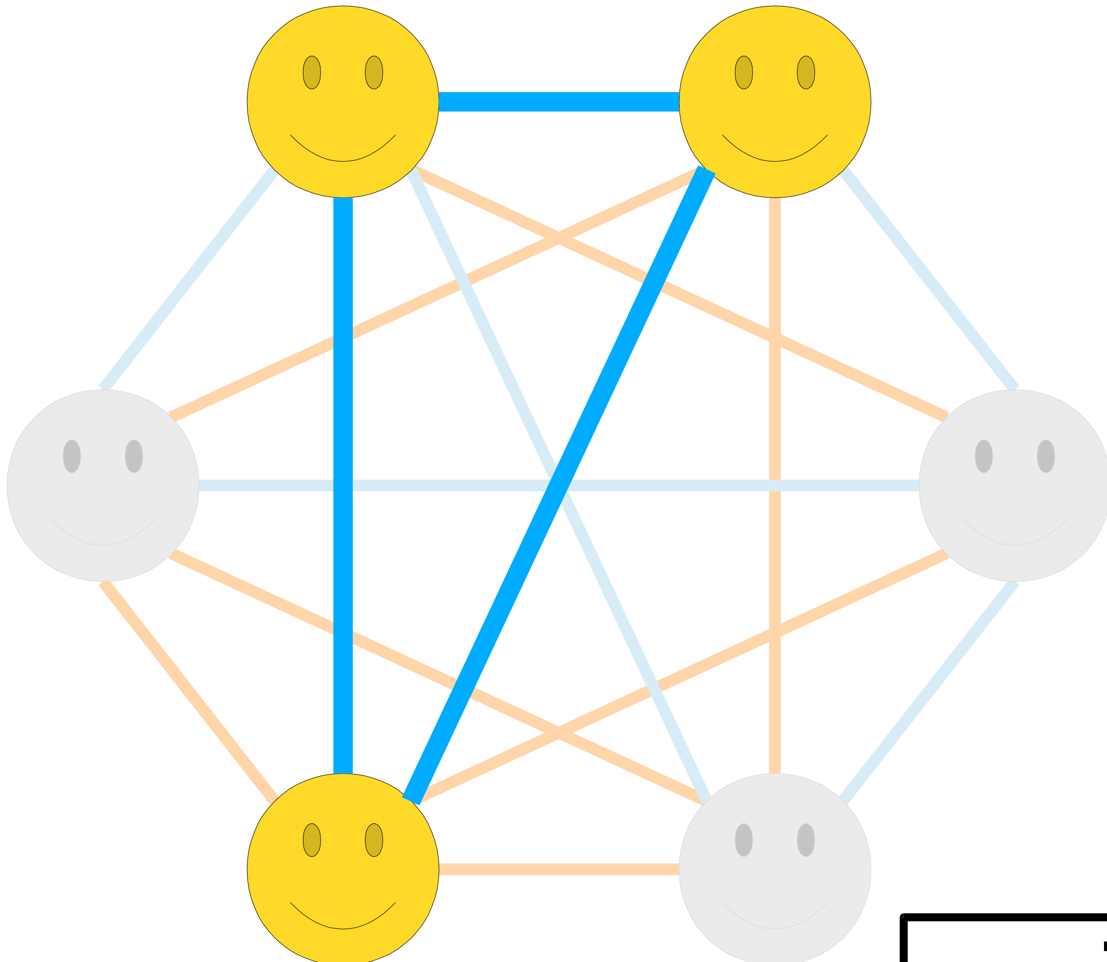




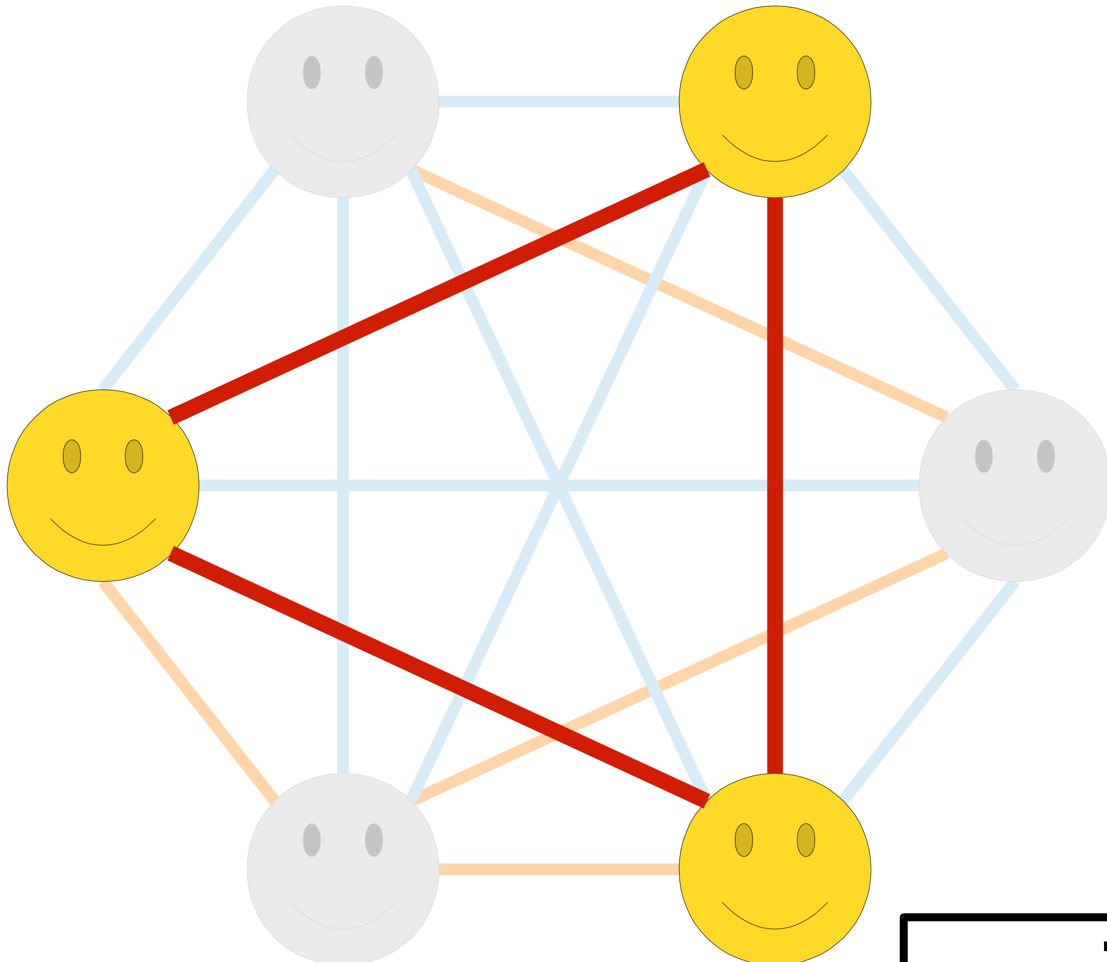
This graph is called $\boldsymbol{K_6}$, the ***complete graph of order 6***. More generally, the graph K_n consists of n mutually adjacent nodes.







This is a ***monochrome*** (one-color) copy of K_3.



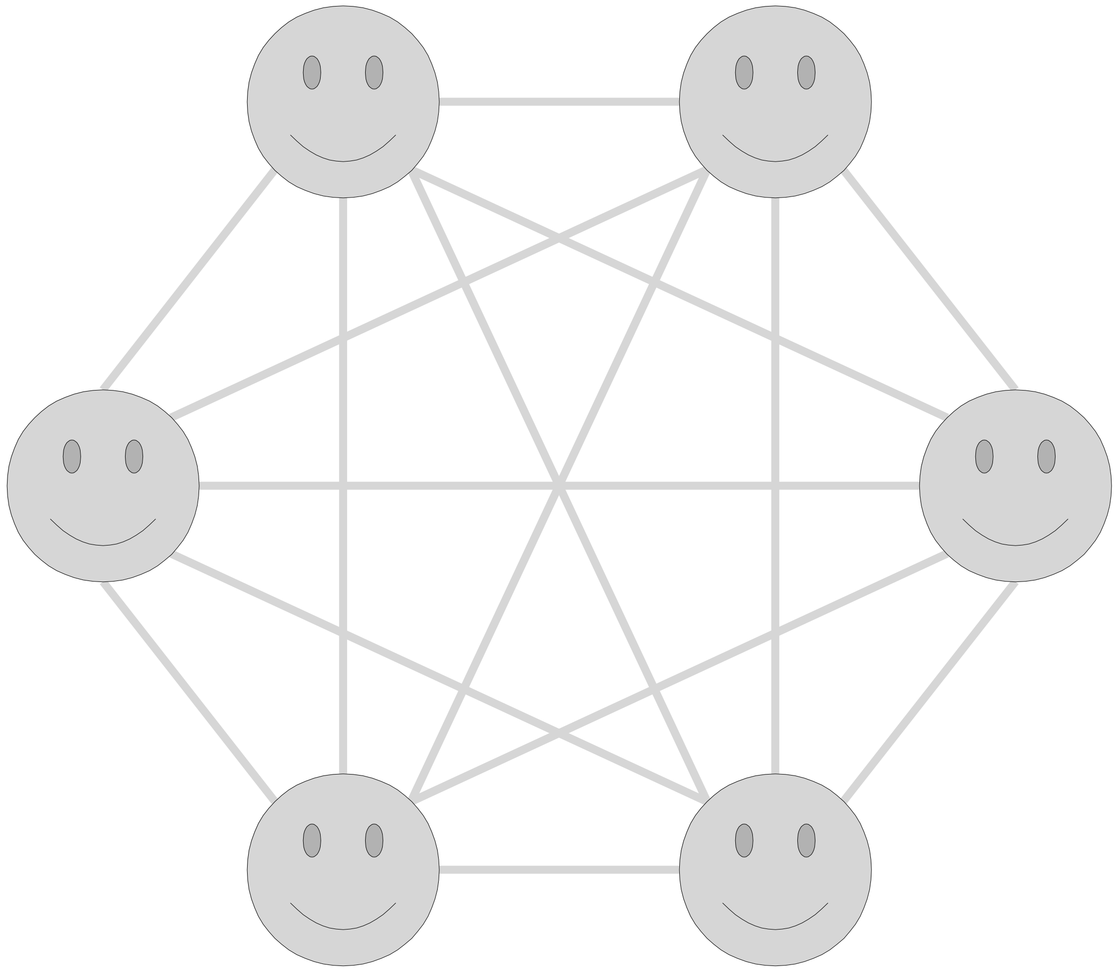
This is a ***monochrome*** (one-color) copy of K_3.

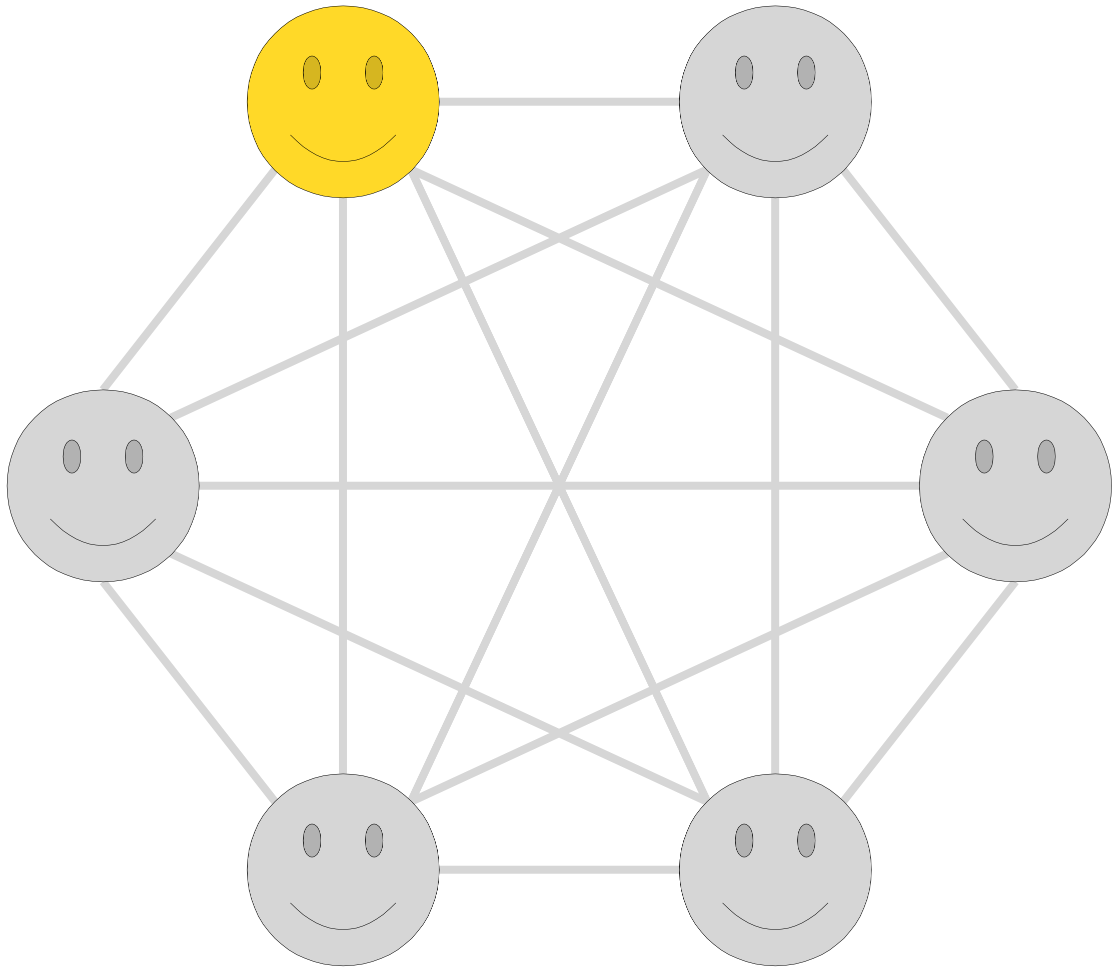
Friends and Strangers Restated

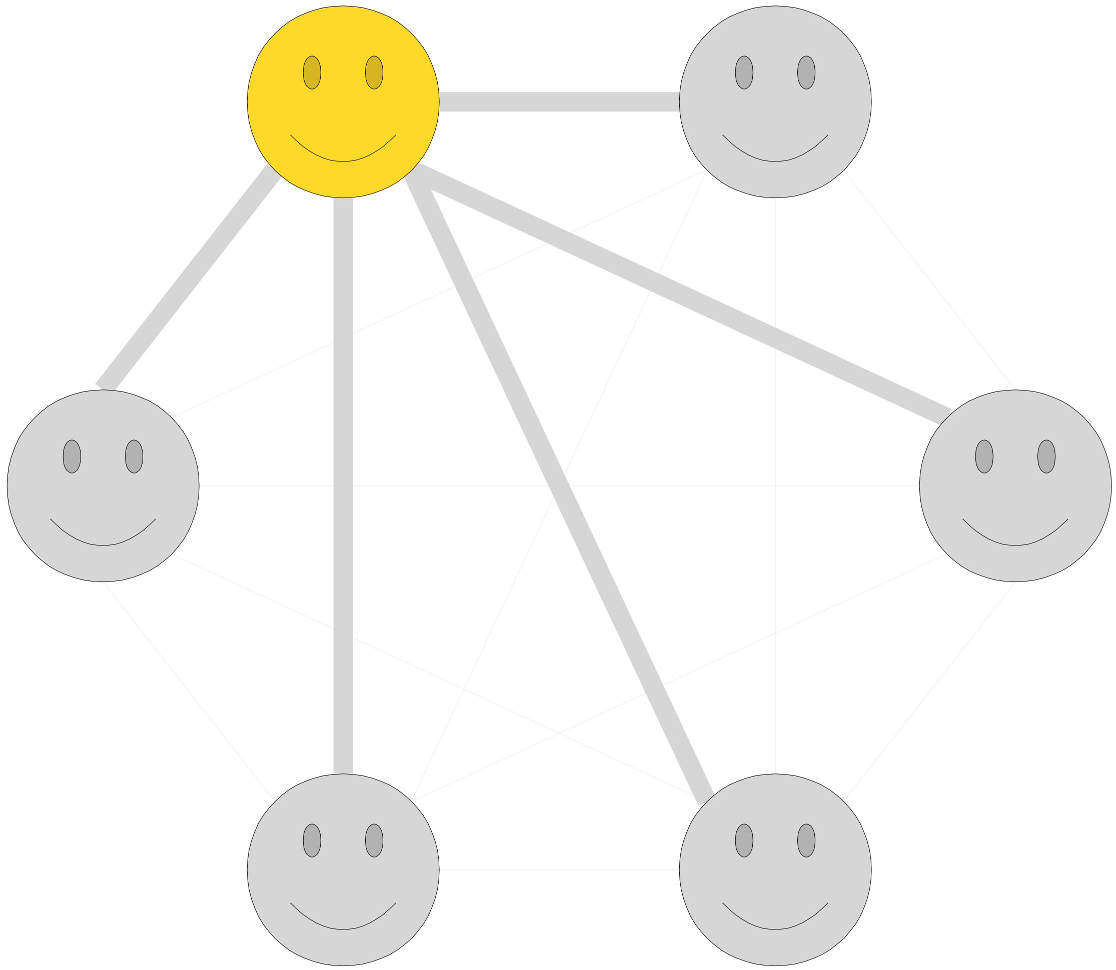
- From a graph-theoretic perspective, the Theorem on Friends and Strangers can be restated as follows:

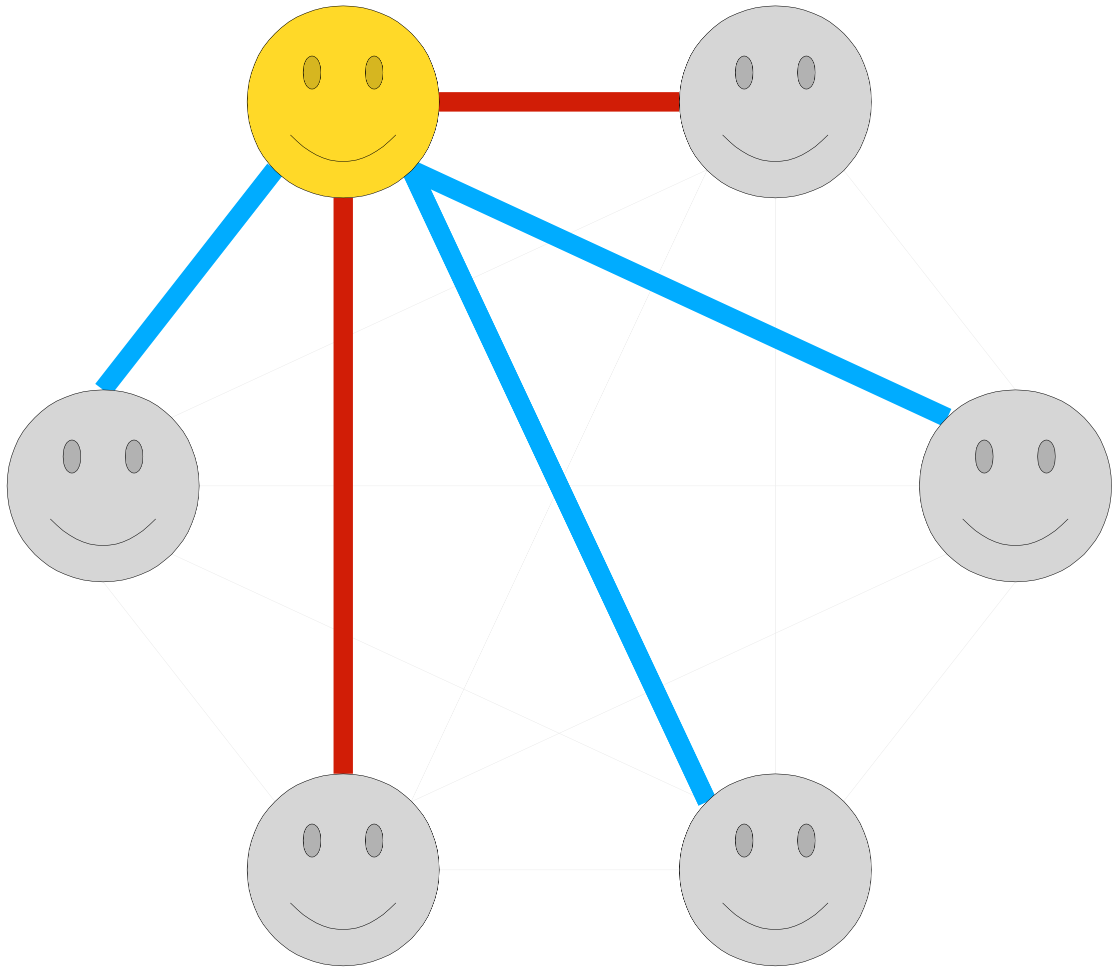
  Theorem: Color every edge of K_6 either red or blue. The resulting graph always contains a monochrome copy of K_3.

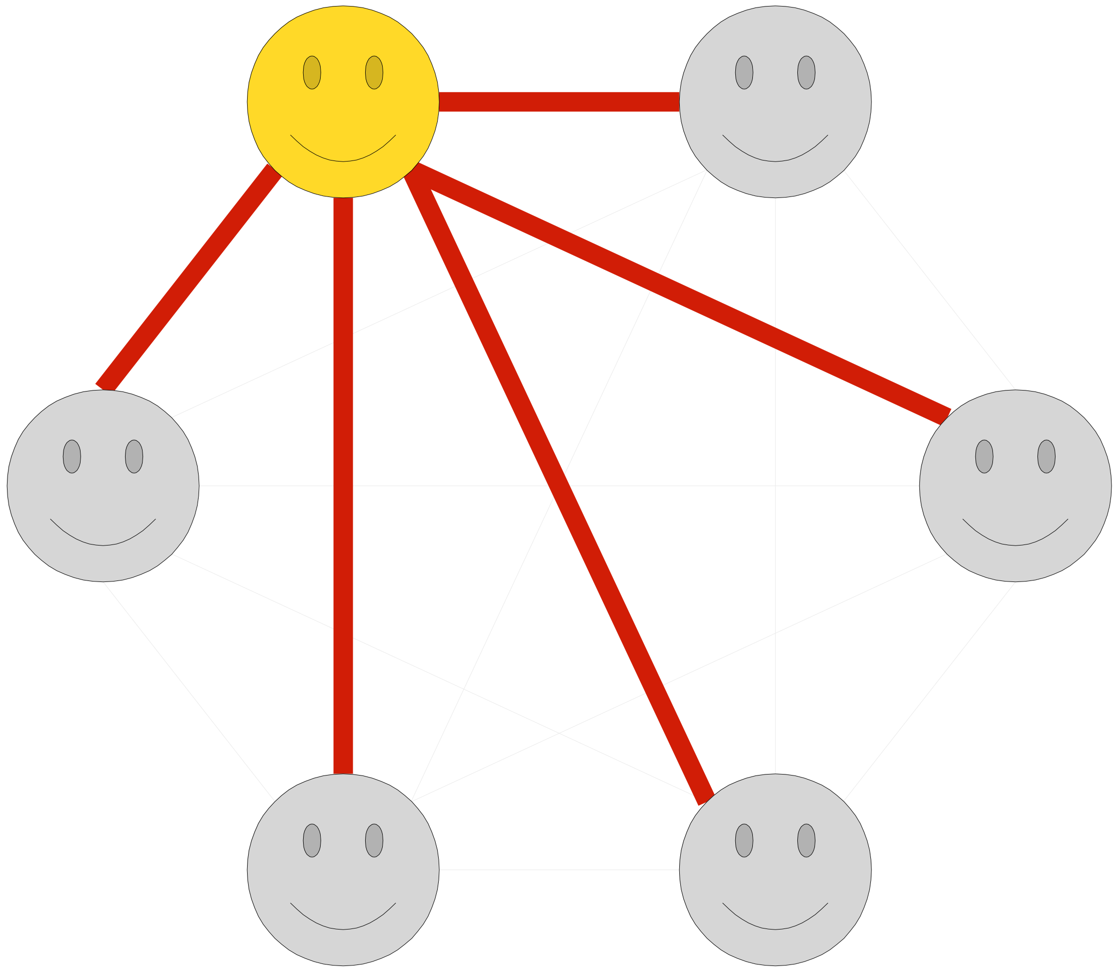
- How can we prove this?

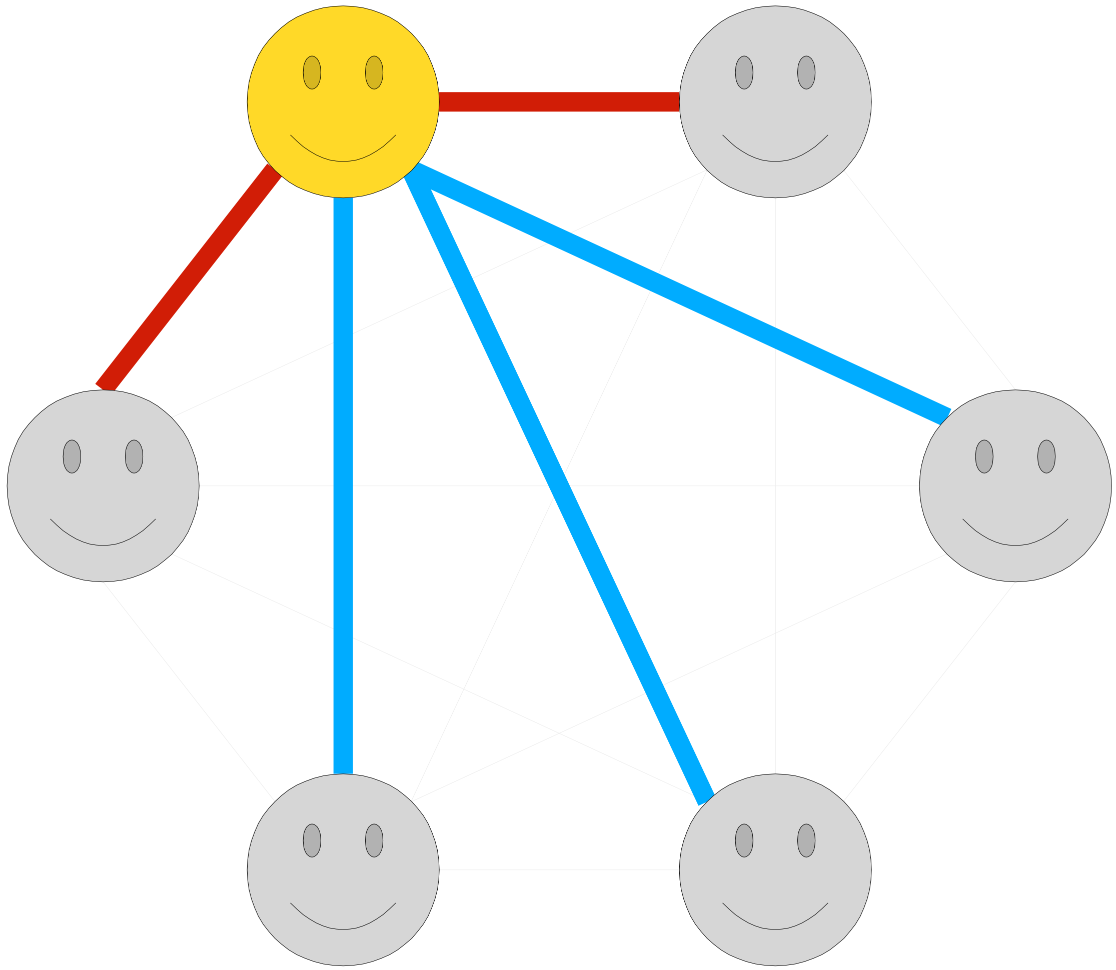


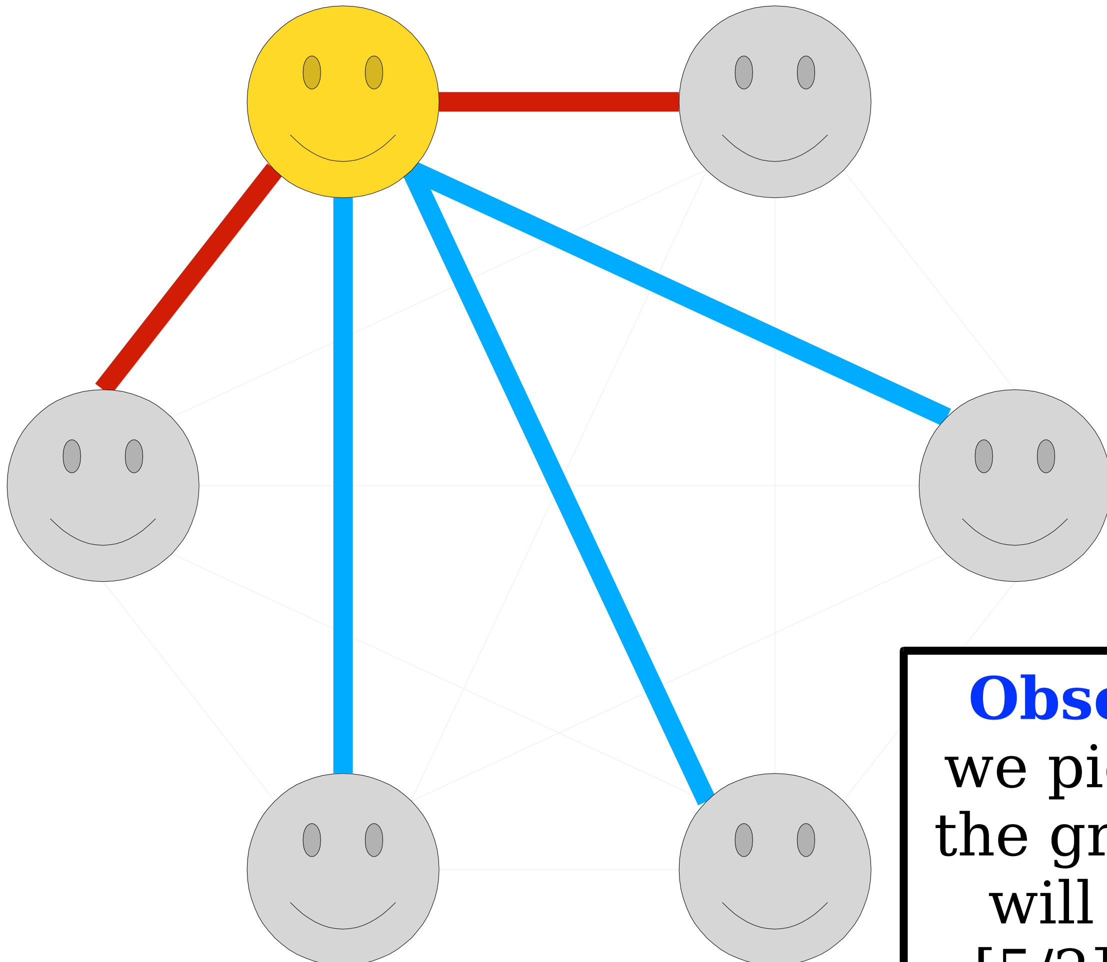




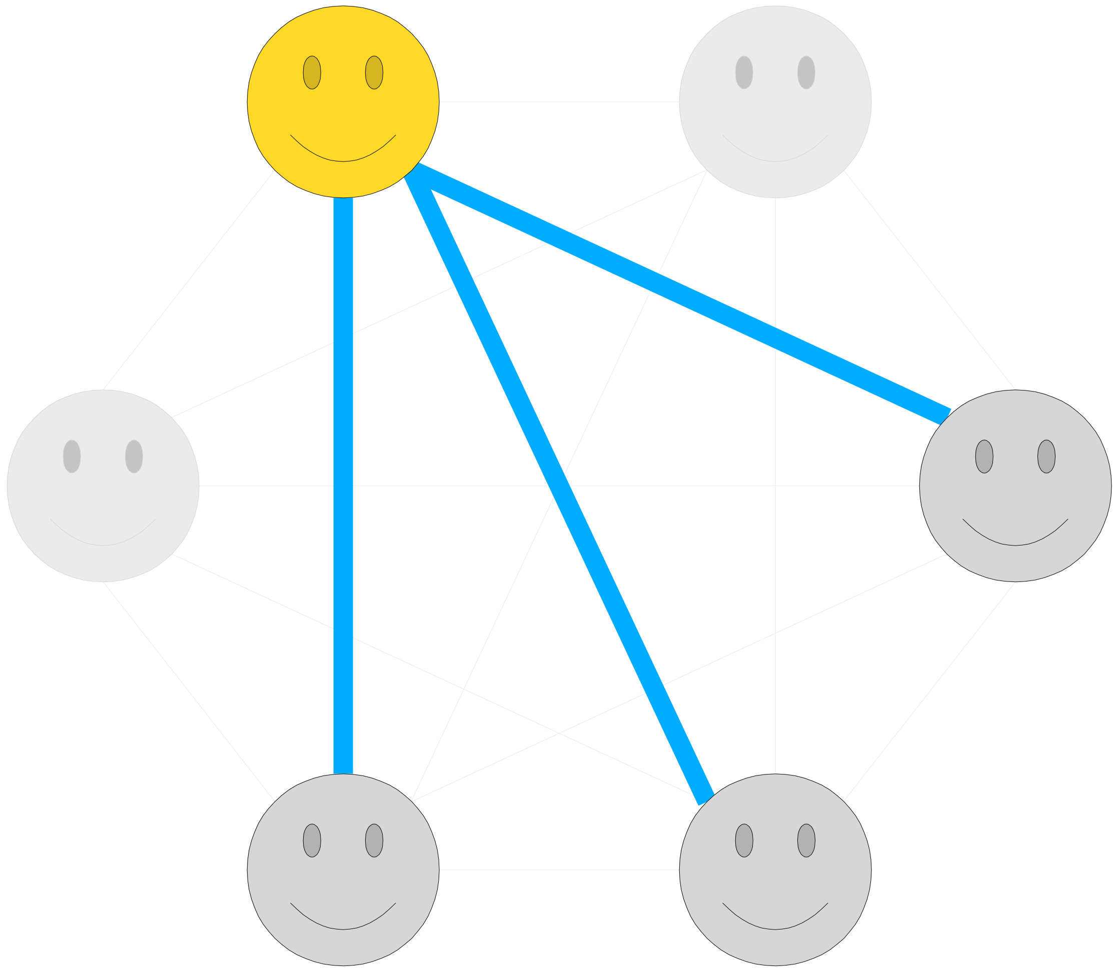


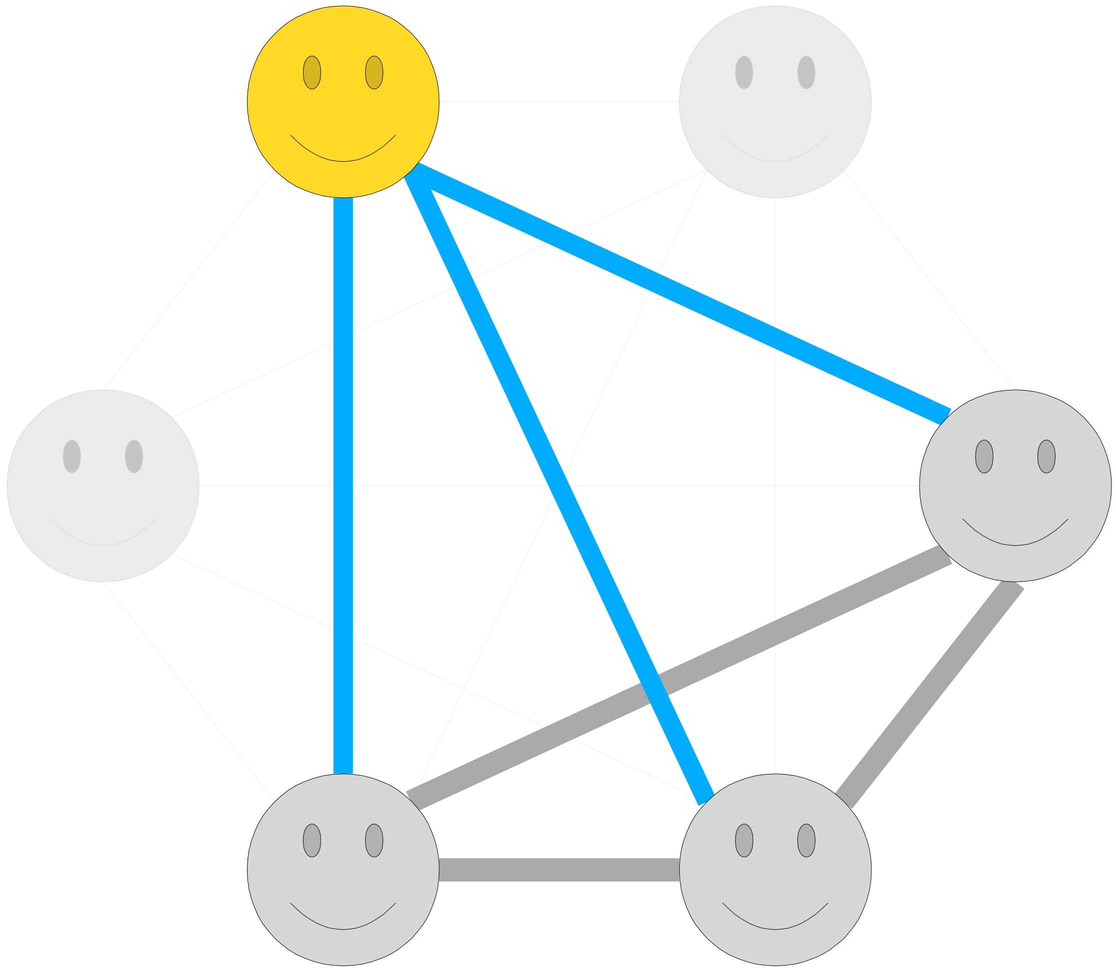


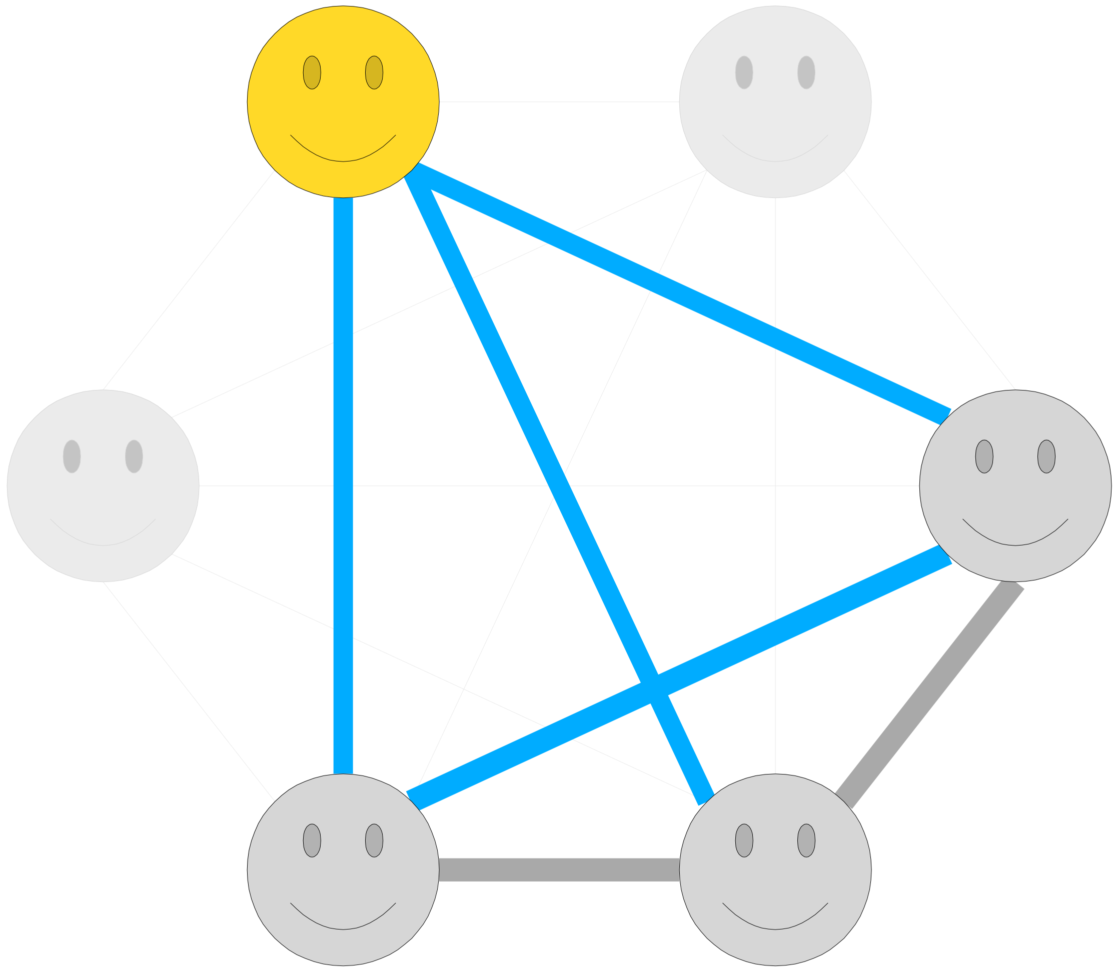


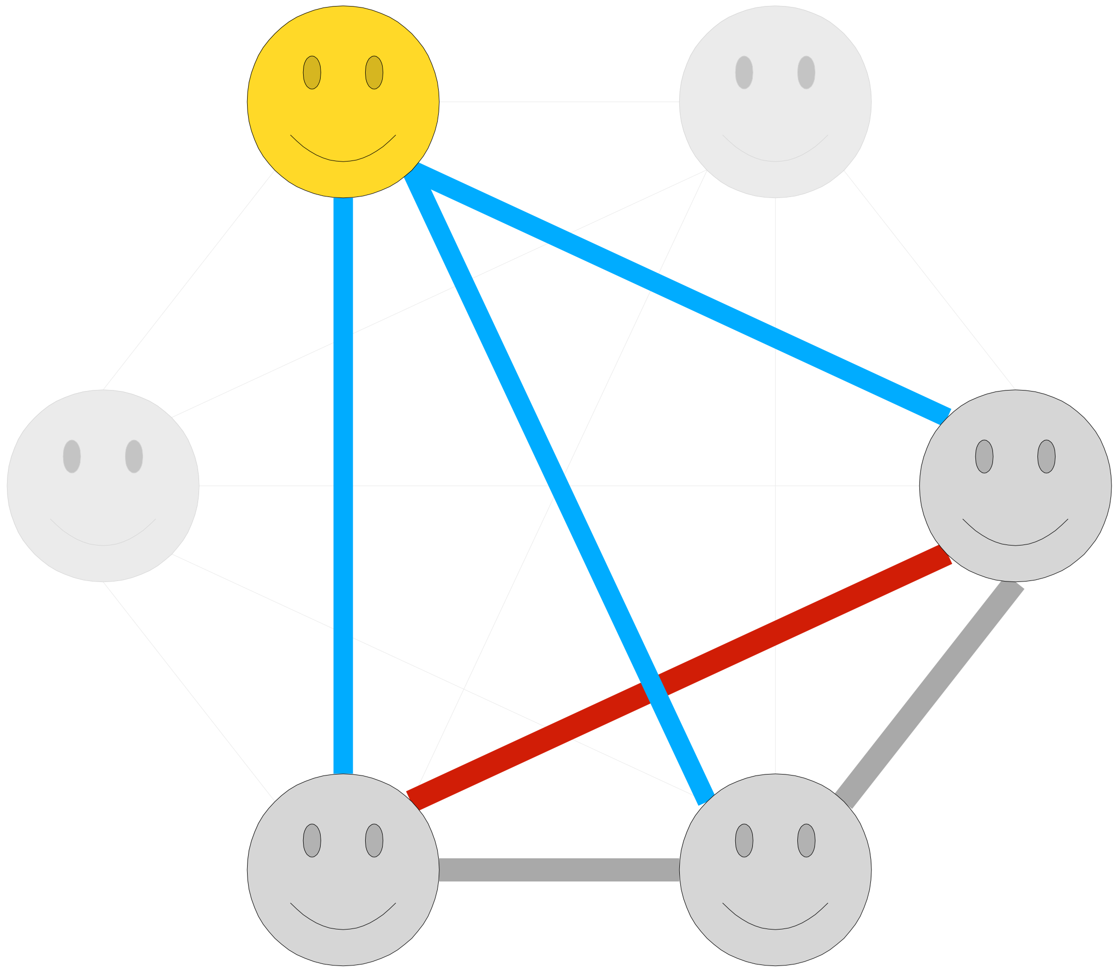


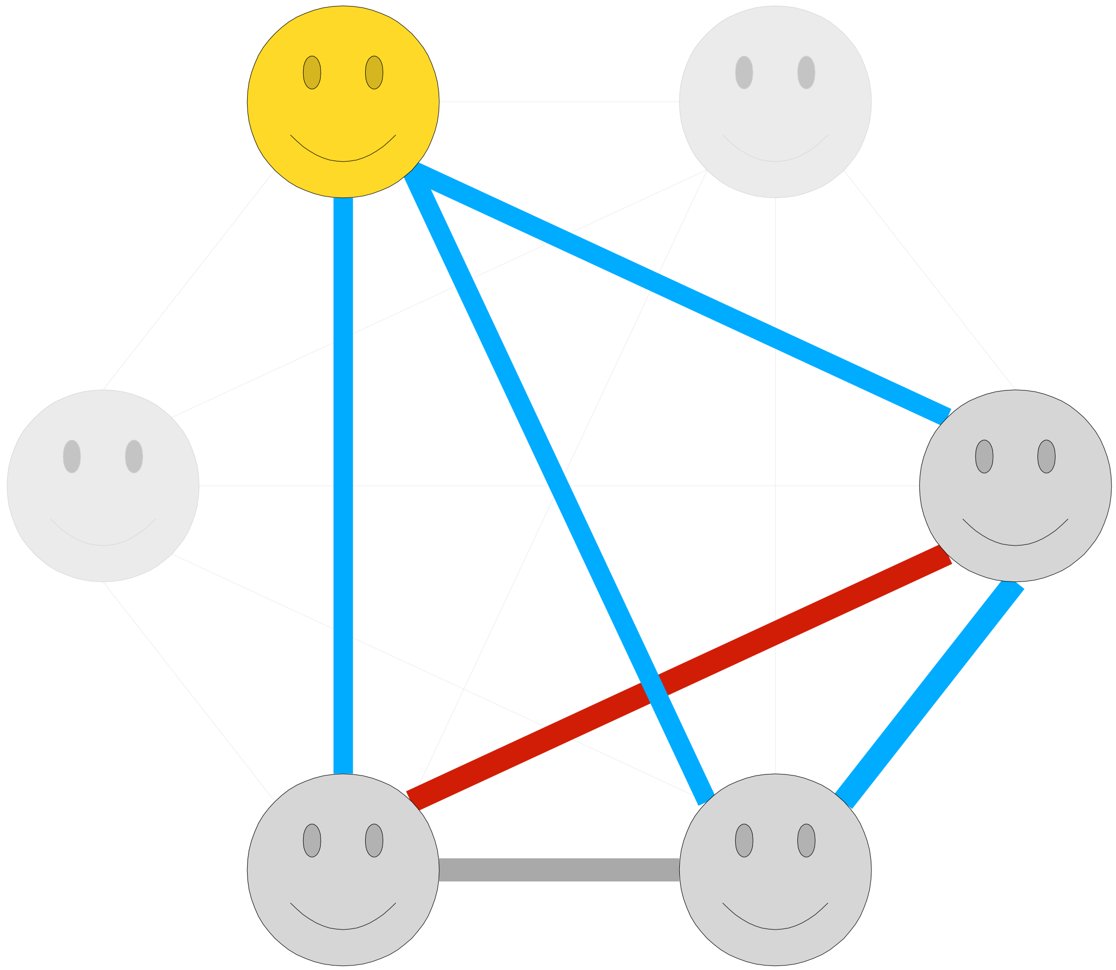
Observation 1: If we pick any node in the graph, that node will have at least ⌈5/2⌉ = 3 edges of the same color incident to it.

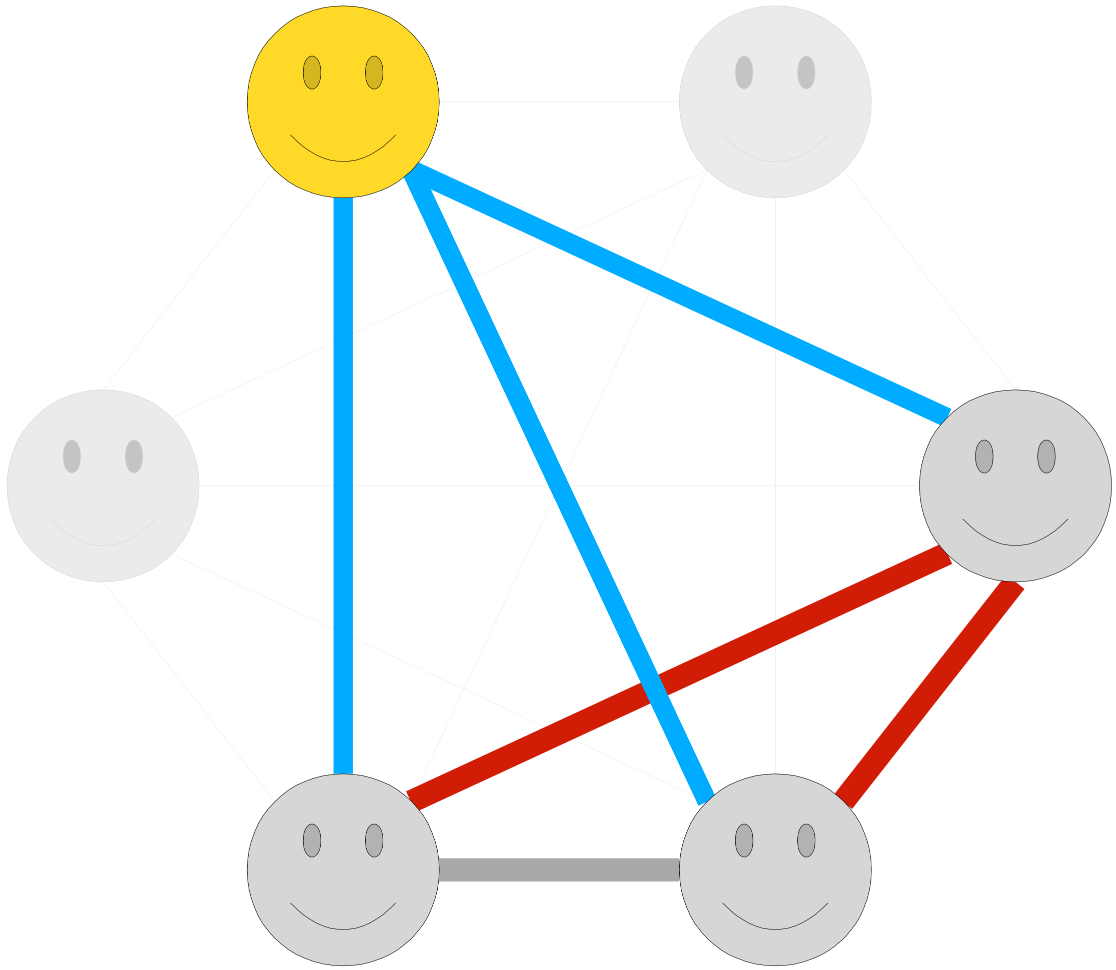


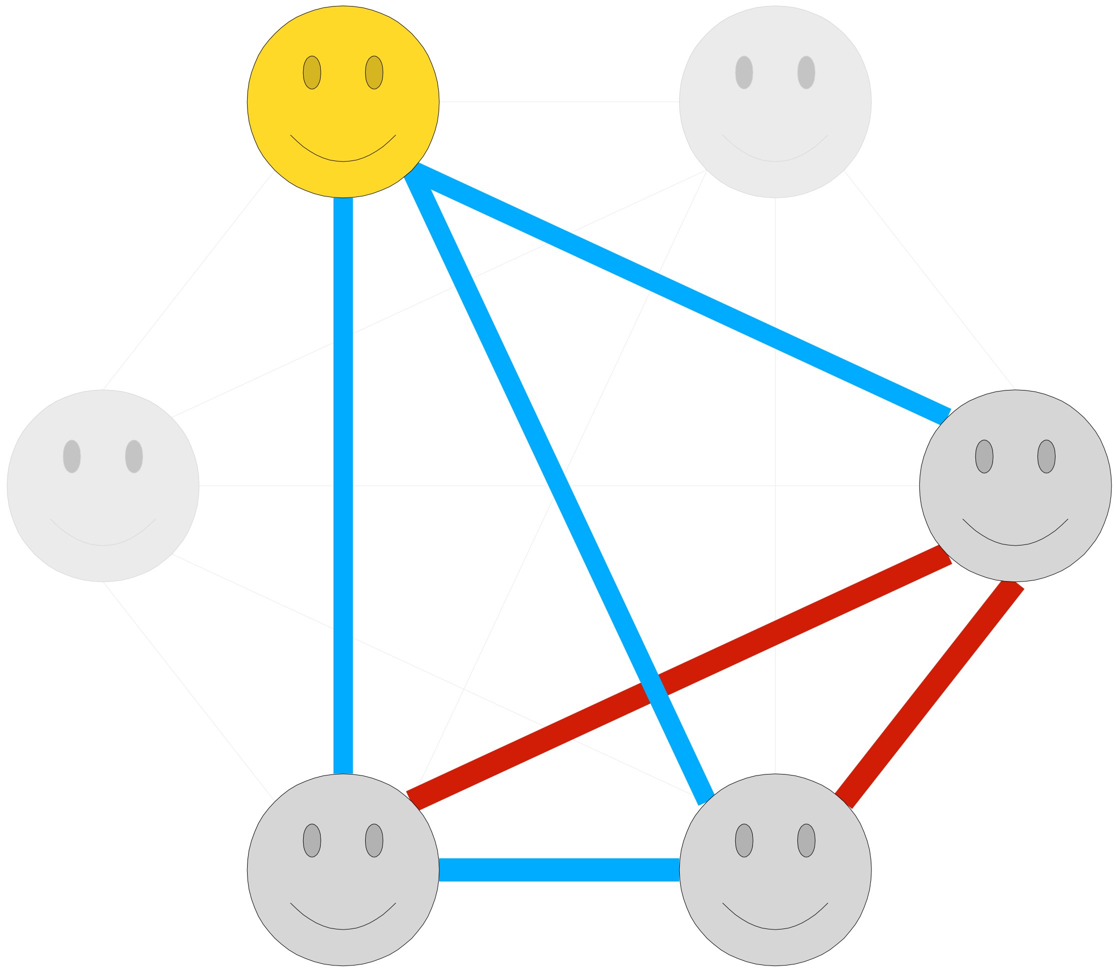


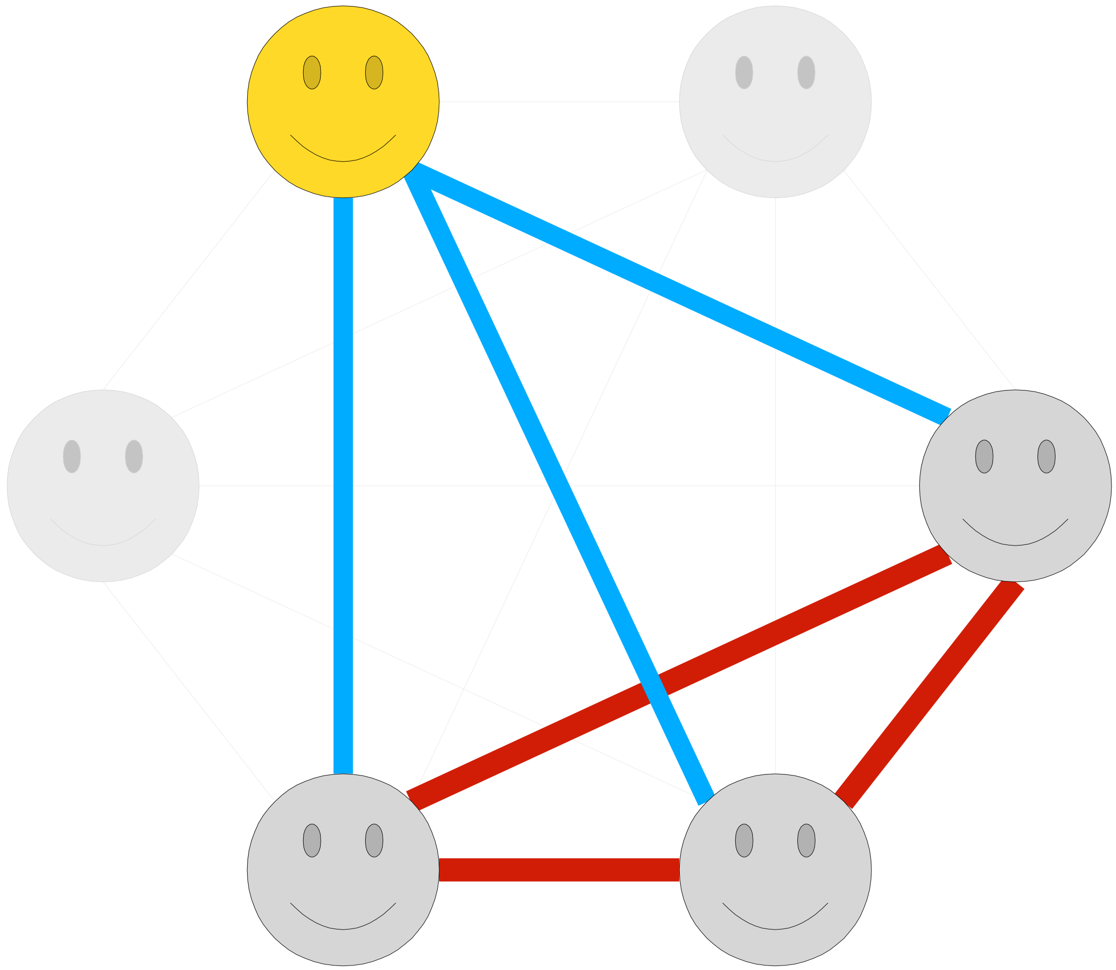












Theorem: Color each edge of K_6 red or blue. The resulting graph contains a monochrome copy of K_3.

Proof: We need to show that the colored K_6 contains a red copy of K_3 or a blue copy of K_3.

Pick some node x from K_6. 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. Without loss of generality, assume those edges are blue.

Let r, s, and t be three of the nodes adjacent to node x along a blue edge. If any of the edges $\{r, s\}$, $\{r, t\}$, or $\{s, t\}$ are blue, then one of those edges plus the two edges connecting back to node x form a blue K_3. Otherwise, all three of those edges are red, and they form a red K_3. Overall, this gives a red K_3 or a blue K_3, as required. ■

Ramsey Theory

- This proof is a special case of a broader family of results called ***Ramsey theory***.
- ***Theorem (Ramsey):*** For any natural number s, there is a number $R(s)$ such that
  - for all $n < R(s)$, there's a way to color the edges of K_n red and blue so there are no monochrome copies of K_s, and
  - for all $n \geq R(s)$, every way of coloring the edges of K_n red and blue always has a monochrome copy of K_s.
- Take Math 108 (combinatorics) to learn more!
- A more philosophical (and less literal) 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.

The Game of Sim

- Here's a game you can play with two players.
  - One players plays as red, the other as blue.
  - Begin with six disconnected points.
  - Each turn, a player draws a line of their color.
  - The first to make a triangle of their color loses.
- The theorem we just proved means the game can't end in a draw: someone must win and someone must lose.
- The strategy is more subtle than it looks. Try playing this with a friend to see why!

Time-Out for Announcements!

Problem Sets

- Problem Set Three was due today at 3:00PM.
  - You can use a late day to extend the deadline to Saturday at 3:00PM if you’d like.
- Problem Set Four goes out today. It’s due next Friday at 3:00PM.
  - It’s all about graphs and graph theory, and you’ll see some really cool results!
  - Because the midterm is on Tuesday, we’ve made this problem set shorter than the previous problem sets.

Midterm Logistics

- Our first midterm is next Tuesday from 7PM – 10PM.
- Best of luck on the exam – ***you can do this!*** We're all cheering you on.
- There's plenty of extra problems online if you're looking to get some additional practice.
- Feel free to ask questions on EdStem over the weekend.

My Advice

- ***Do*** block out some dedicated time to work through practice problems.
- ***Do*** get the TAs to review your answers to those problems; ask privately on EdStem.
- ***Do*** take some time this weekend to take a walk, smell the rosemary bushes on campus, and watch the bees buzz.
- ***Don't*** pull an all-nighter studying for the exam.
- ***Don't*** skip meals or alter your daily routine to fit in time for studying.
- ***Don't*** panic. You can do this!

Back to CS103!

A Little Math Puzzle

"In a group of $n > 0$ people …

- 90% of those people enjoyed *CODA,*
- 80% of those people enjoyed *Nomadland,*
- 70% of those people enjoyed *Parasite,* and
- 60% of those people enjoyed *Knives Out*.

No one enjoyed all four movies. How many people enjoyed at least one of *CODA* and *Parasite*?"

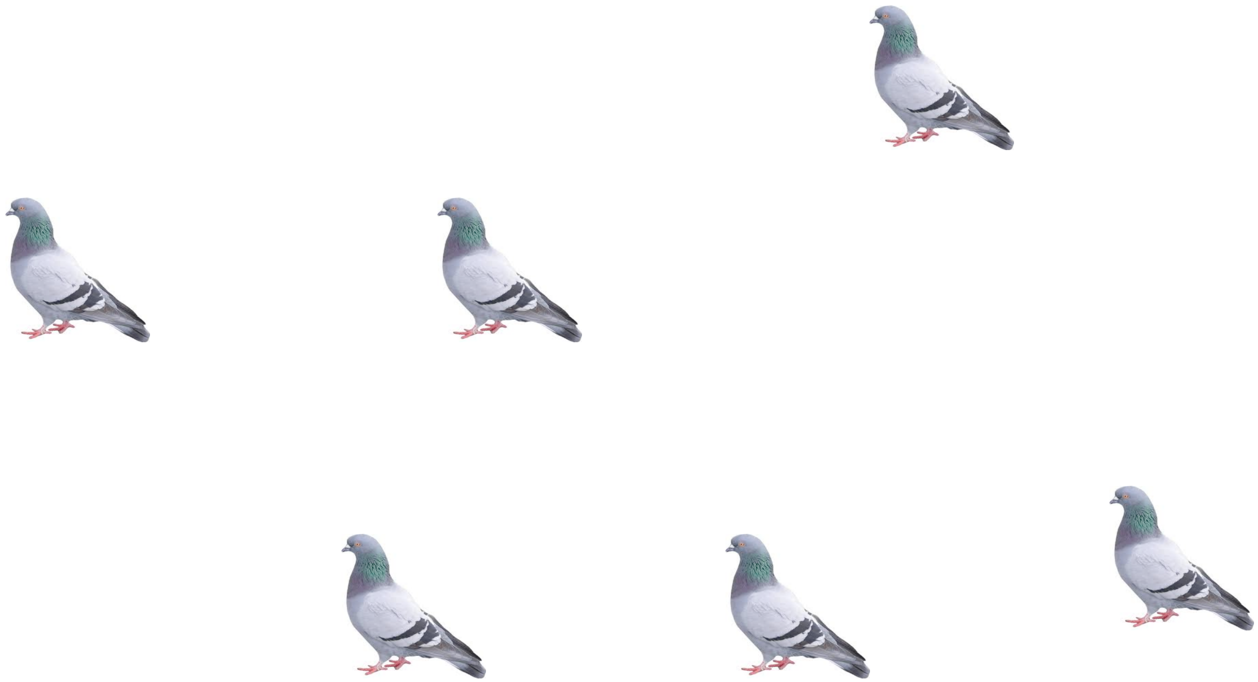
(Adapted from here.)

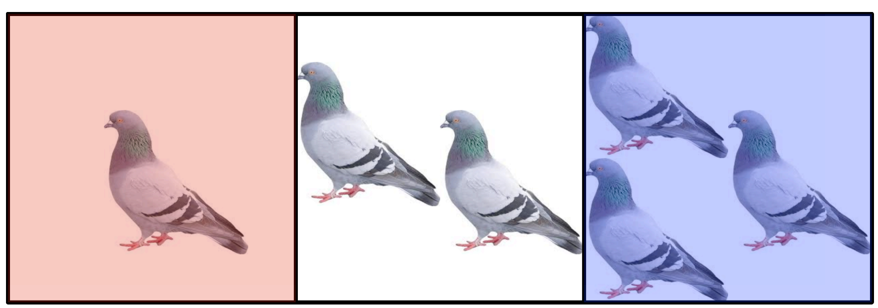
Other Pigeonhole-Type Results

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.

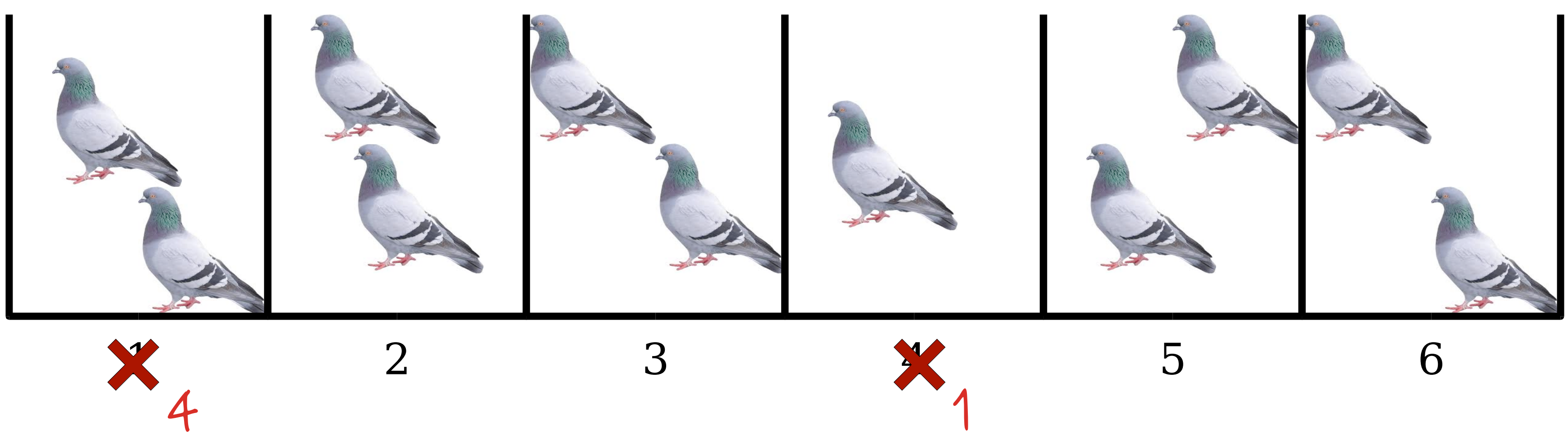
For simplicity, denote by x_i the number of objects in bin i. Without loss of generality, assume that bin 1 has fewer than m/n objects, meaning that $x_1 < m/n$.

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

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

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

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

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

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

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

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

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

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. ■

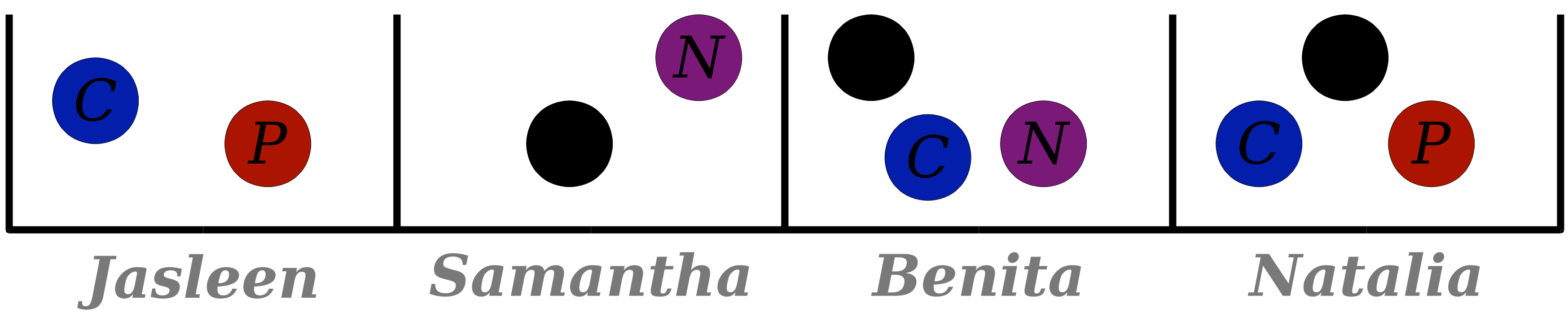
"In a group of $n > 0$ people …

- 90% of those people enjoyed ***CODA***,
- 80% of those people enjoyed ***Nomadland***,
- 70% of those people enjoyed ***Parasite***, and
- 60% of those people enjoyed ***Knives Out***.

No one enjoyed all four movies. How many people enjoyed at least one of *CODA* and *Parasite*?"

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 ***CODA***,
- 80% of those people enjoyed ***Nomadland***,
- 70% of those people enjoyed ***Parasite***, and
- 60% of those people enjoyed ***Knives Out***.

No one enjoyed all four movies. How many people enjoyed at least one of *CODA* and *Parasite*?"

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

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 ***CODA***,
- 80% of those people enjoyed ***Nomadland***,
- 70% of those people enjoyed ***Parasite***, and
- 60% of those people enjoyed ***Knives Out***.

No one enjoyed all four movies. How many people enjoyed at least one of *CODA* and *Parasite*?"

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 ***CODA***,
- 80% of those people enjoyed ***Nomadland***,
- 70% of those people enjoyed ***Parasite***, and
- 60% of those people enjoyed ***Knives Out***.

No one enjoyed all four movies. How many people enjoyed at least one of *CODA* and *Parasite*?"

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 *CODA* and *Parasite*.

"In a group of $n > 0$ people …

- 90% of those people enjoyed ***CODA***,
- 80% of those people enjoyed ***Nomadland***,
- 70% of those people enjoyed ***Parasite***, and
- 60% of those people enjoyed ***Knives Out***.

No one enjoyed all four movies. How many people enjoyed at least one of *CODA* and *Parasite*?"

Theorem: In the scenario described here, all n people enjoyed at least one of *CODA* and *Parasite*.

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

"In a group of $n > 0$ people …

- 90% of those people enjoyed ***CODA***,
- 80% of those people enjoyed ***Nomadland***,
- 70% of those people enjoyed ***Parasite***, and
- 60% of those people enjoyed ***Knives Out***.

No one enjoyed all four movies. How many people enjoyed at least one of *CODA* and *Parasite*?"

Theorem: In the scenario described here, all n people enjoyed at least one of *CODA* and *Parasite*.

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 ***CODA***,
- 80% of those people enjoyed ***Nomadland***,
- 70% of those people enjoyed ***Parasite***, and
- 60% of those people enjoyed ***Knives Out***.

No one enjoyed all four movies. How many people enjoyed at least one of *CODA* and *Parasite*?"

Theorem: In the scenario described here, all n people enjoyed at least one of *CODA* and *Parasite*.

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 ***CODA***,
- 80% of those people enjoyed ***Nomadland***,
- 70% of those people enjoyed ***Parasite***, and
- 60% of those people enjoyed ***Knives Out***.

No one enjoyed all four movies. How many people enjoyed at least one of *CODA* and *Parasite*?"

Theorem: In the scenario described here, all n people enjoyed at least one of *CODA* and *Parasite*.

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 ***CODA***,
- 80% of those people enjoyed ***Nomadland***,
- 70% of those people enjoyed ***Parasite***, and
- 60% of those people enjoyed ***Knives Out***.

No one enjoyed all four movies. How many people enjoyed at least one of *CODA* and *Parasite*?"

Theorem: In the scenario described here, all n people enjoyed at least one of *CODA* and *Parasite*.

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 ***CODA***,
- 80% of those people enjoyed ***Nomadland***,
- 70% of those people enjoyed ***Parasite***, and
- 60% of those people enjoyed ***Knives Out***.

No one enjoyed all four movies. How many people enjoyed at least one of *CODA* and *Parasite*?"

Theorem: In the scenario described here, all n people enjoyed at least one of *CODA* and *Parasite*.

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 *CODA* and didn't enjoy *Parasite*.

"In a group of $n > 0$ people …

- 90% of those people enjoyed ***CODA***,
- 80% of those people enjoyed ***Nomadland***,
- 70% of those people enjoyed ***Parasite***, and
- 60% of those people enjoyed ***Knives Out***.

No one enjoyed all four movies. How many people enjoyed at least one of *CODA* and *Parasite*?"

Theorem: In the scenario described here, all n people enjoyed at least one of *CODA* and *Parasite*.

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 *CODA* and didn't enjoy *Parasite*. 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 ***CODA***,
- 80% of those people enjoyed ***Nomadland***,
- 70% of those people enjoyed ***Parasite***, and
- 60% of those people enjoyed ***Knives Out***.

No one enjoyed all four movies. How many people enjoyed at least one of *CODA* and *Parasite*?"

Theorem: In the scenario described here, all n people enjoyed at least one of *CODA* and *Parasite*.

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 *CODA* and didn't enjoy *Parasite*. 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 *CODA* and *Parasite*. ■

"In a group of $n > 0$ people …

- 90% of those people enjoyed ***CODA***,
- 80% of those people enjoyed ***Nomadland***,
- 70% of those people enjoyed ***Parasite***, and
- 60% of those people enjoyed ***Knives Out***.

No one enjoyed all four movies. How many people enjoyed at least one of *CODA* and *Parasite*?"

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!)*

More to Explore

- Interested in more about graphs and the pigeonhole principle? Check out…
  - … ***Math 107*** (Graph Theory), a deep dive into graph theory.
  - … ***Math 108*** (Combinatorics), which explores a bunch of results pertaining to graphs and counting things.
  - … ***CS161*** (Algorithms), which explores algorithms for computing important properties of graphs.
  - … ***CS224W*** (Deep Learning on Graphs), which uses a mix of mathematical and statistical techniques to explore graphs.
- Happy to chat about this in person if you'd like.

Next Time

- ***Mathematical Induction***
  - Reasoning about stepwise processes!
- ***Applications of Induction***
  - To numbers!
  - To anticounterfeiting!
  - To puzzles!