Graph Theory

Part Three

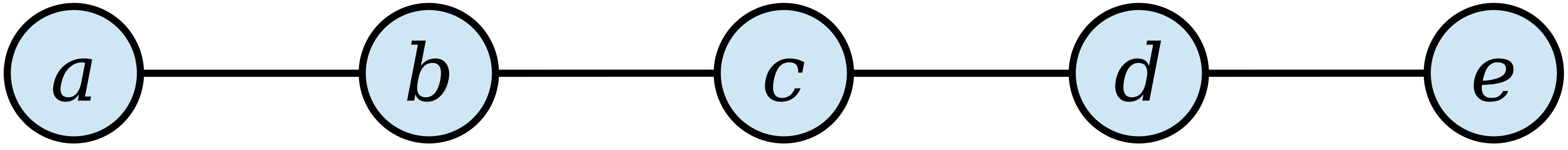
Outline for Today

- ***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.

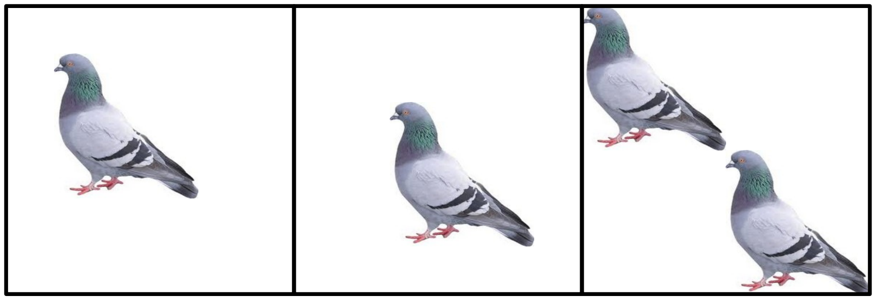


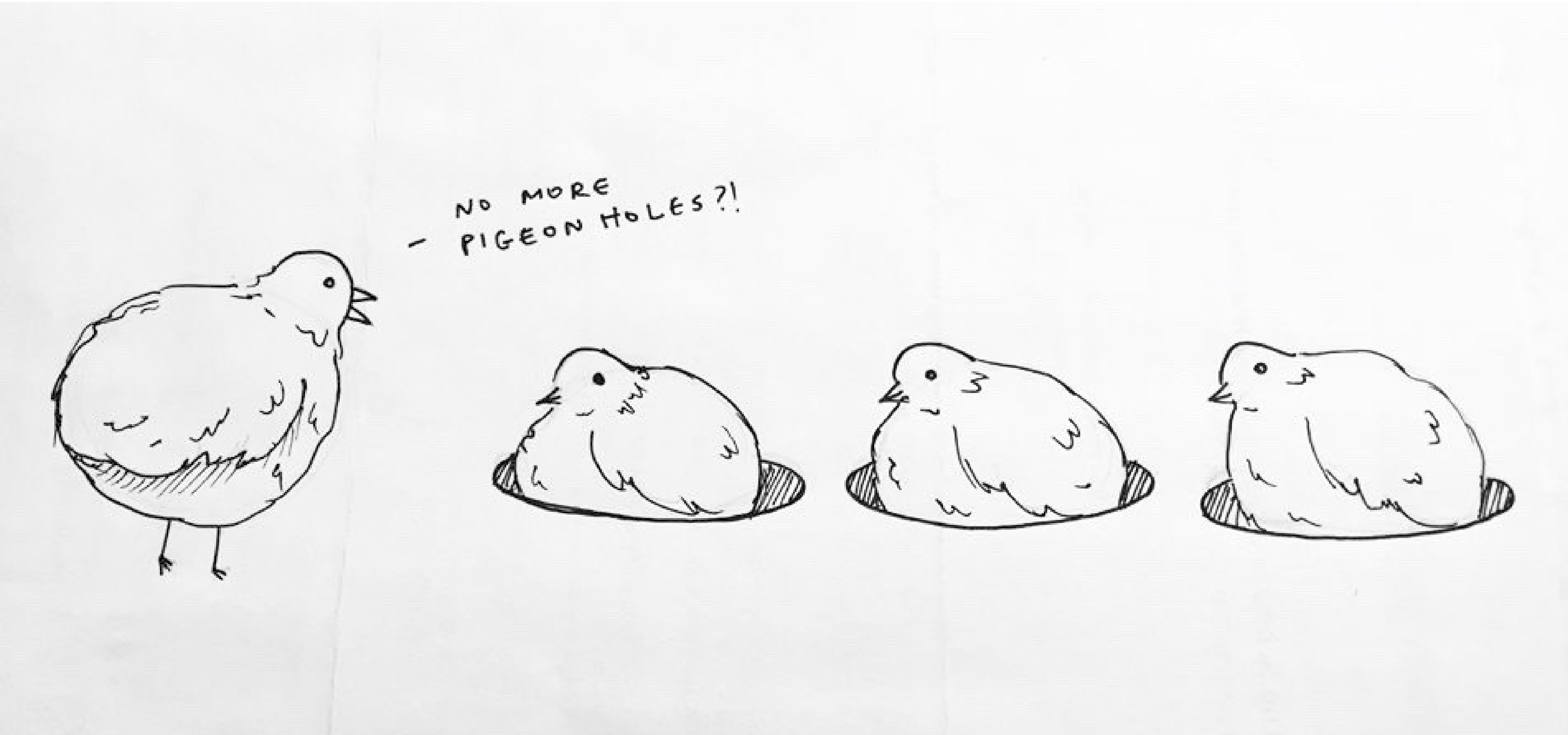
New Stuff!

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.





$m = 4, n = 3$

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 $|A| > |B|$. How many 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.

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

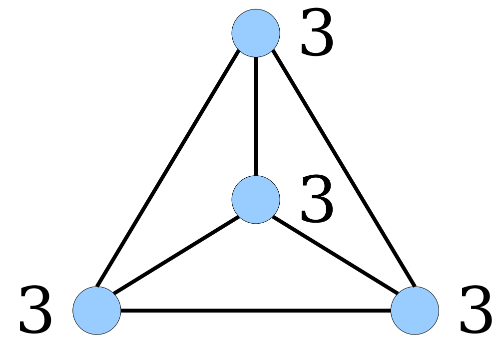
$$\begin{aligned} m &= x_1 + x_2 + \ldots + x_n \\ &\le 1 + 1 + \ldots + 1 \quad (n \text{ times}) \\ &= n. \end{aligned}$$

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. ■

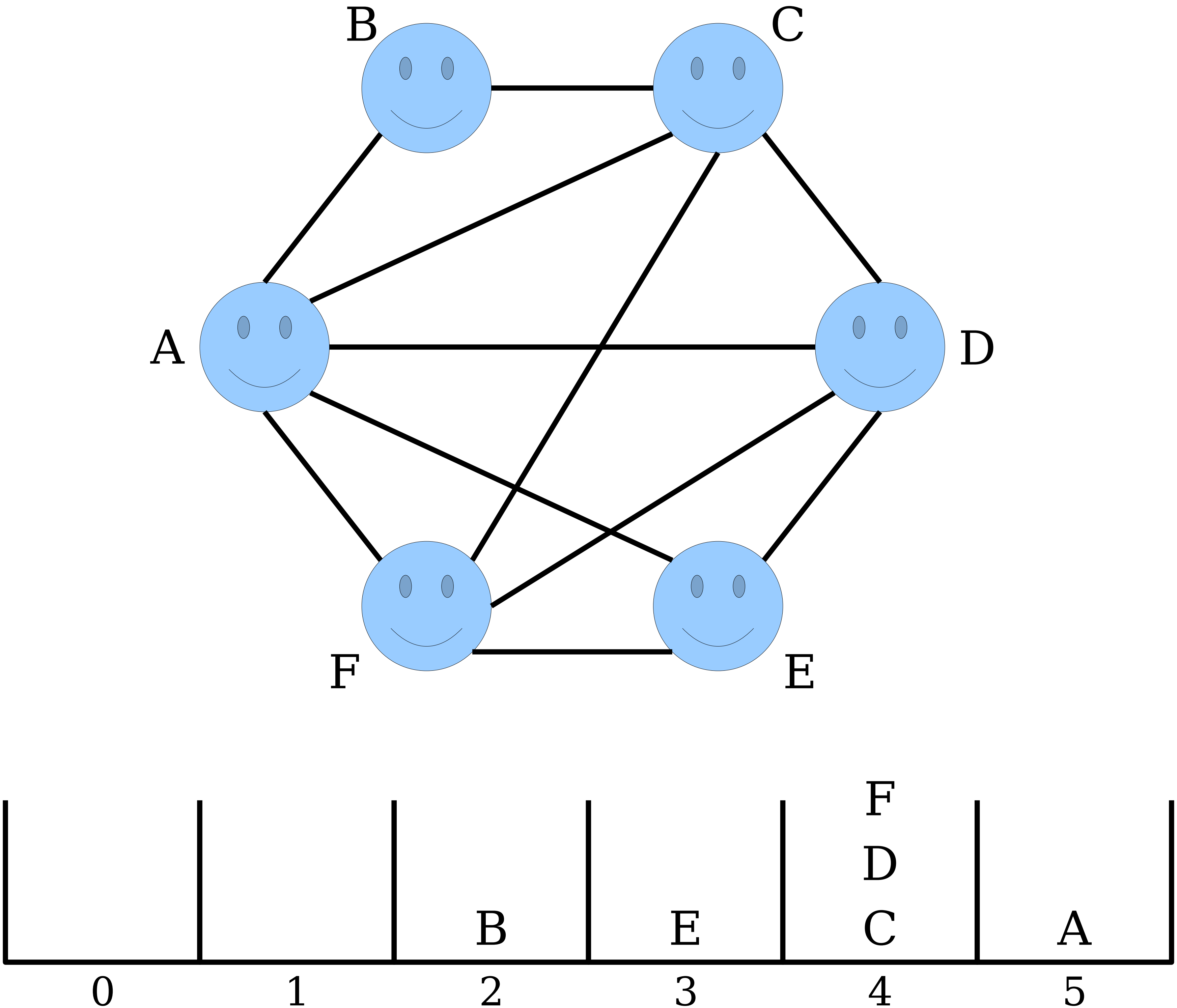
Pigeonhole Principle Party Tricks

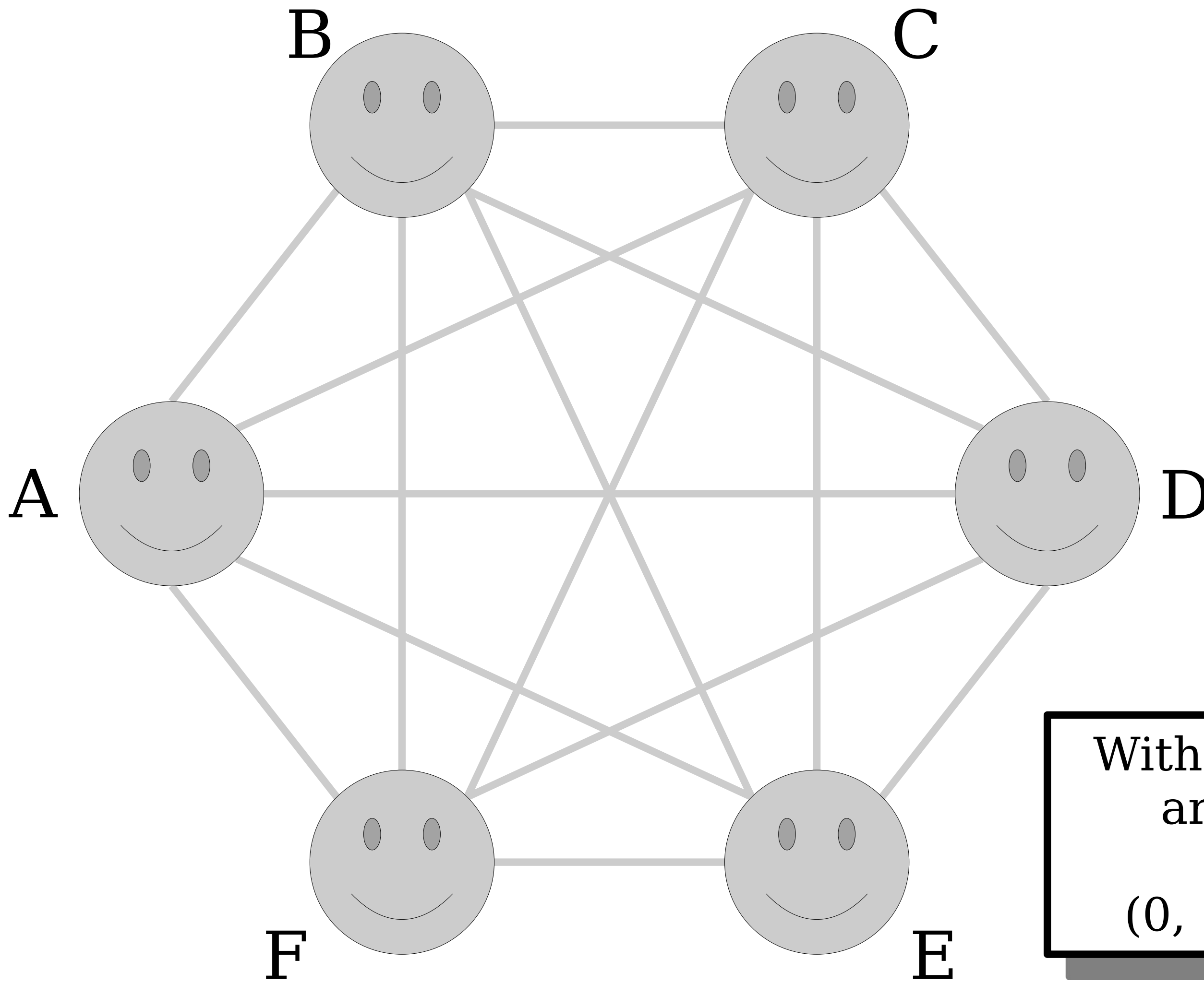
Degrees

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

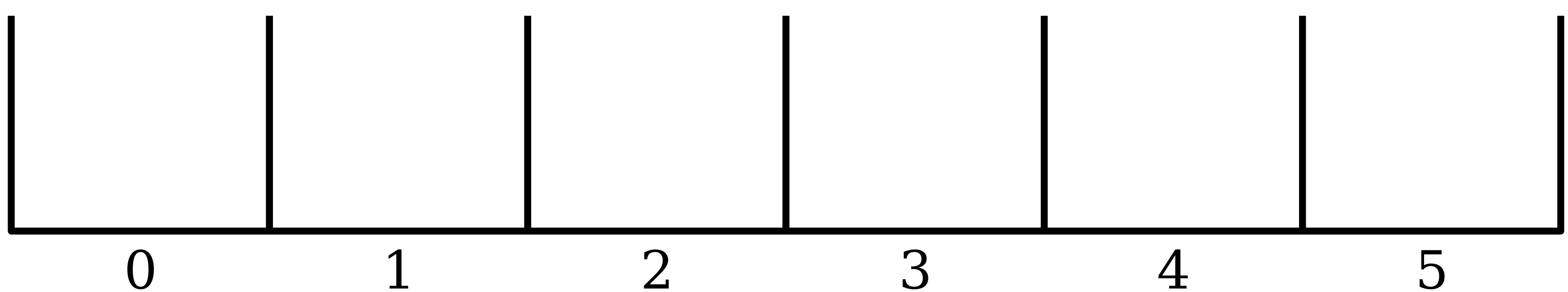


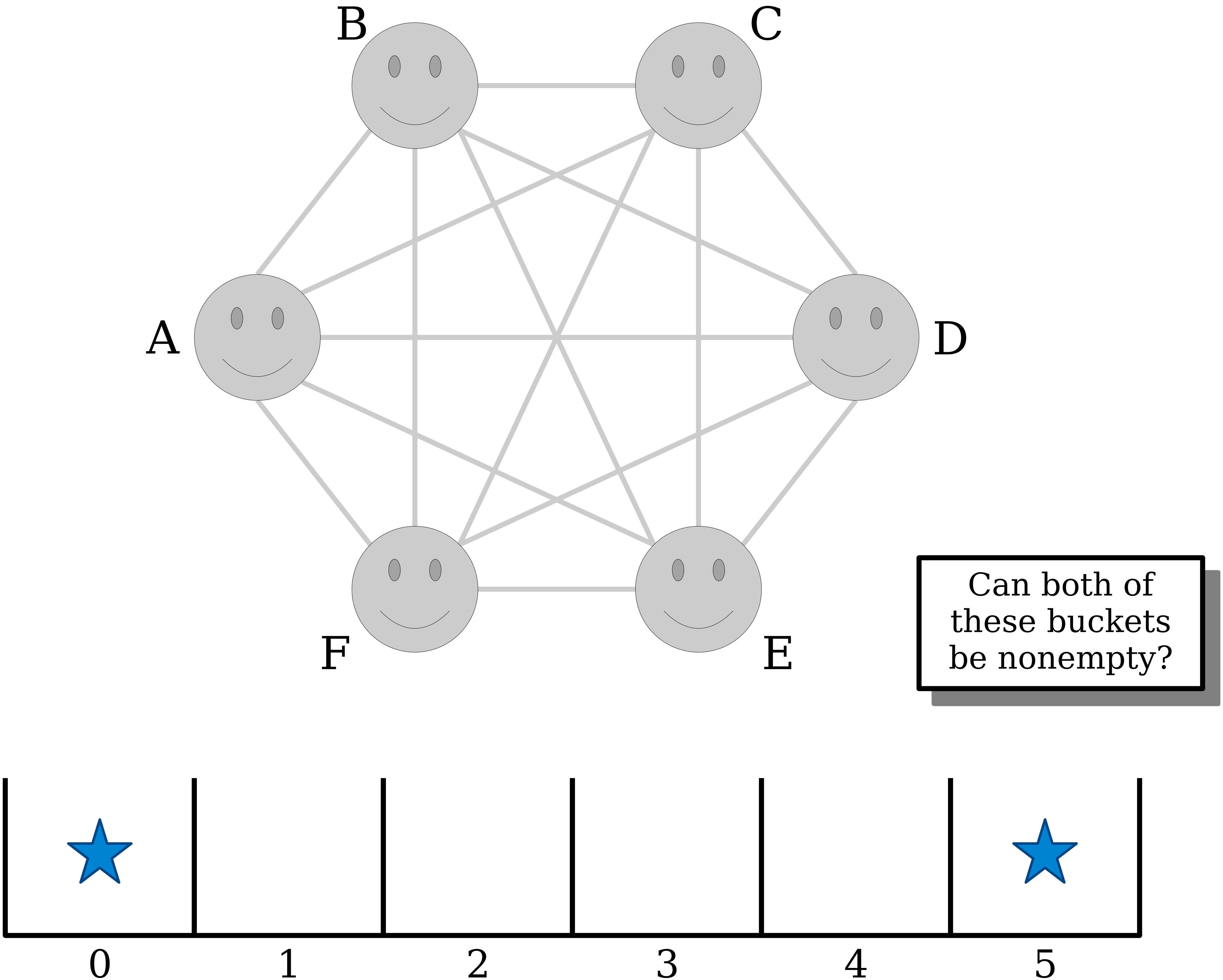
- ***Theorem:*** Every graph with at least two nodes has at least two nodes with the same degree.
  - Equivalently: at any party with at least two people, there are at least two people with the same number of friends at the party.

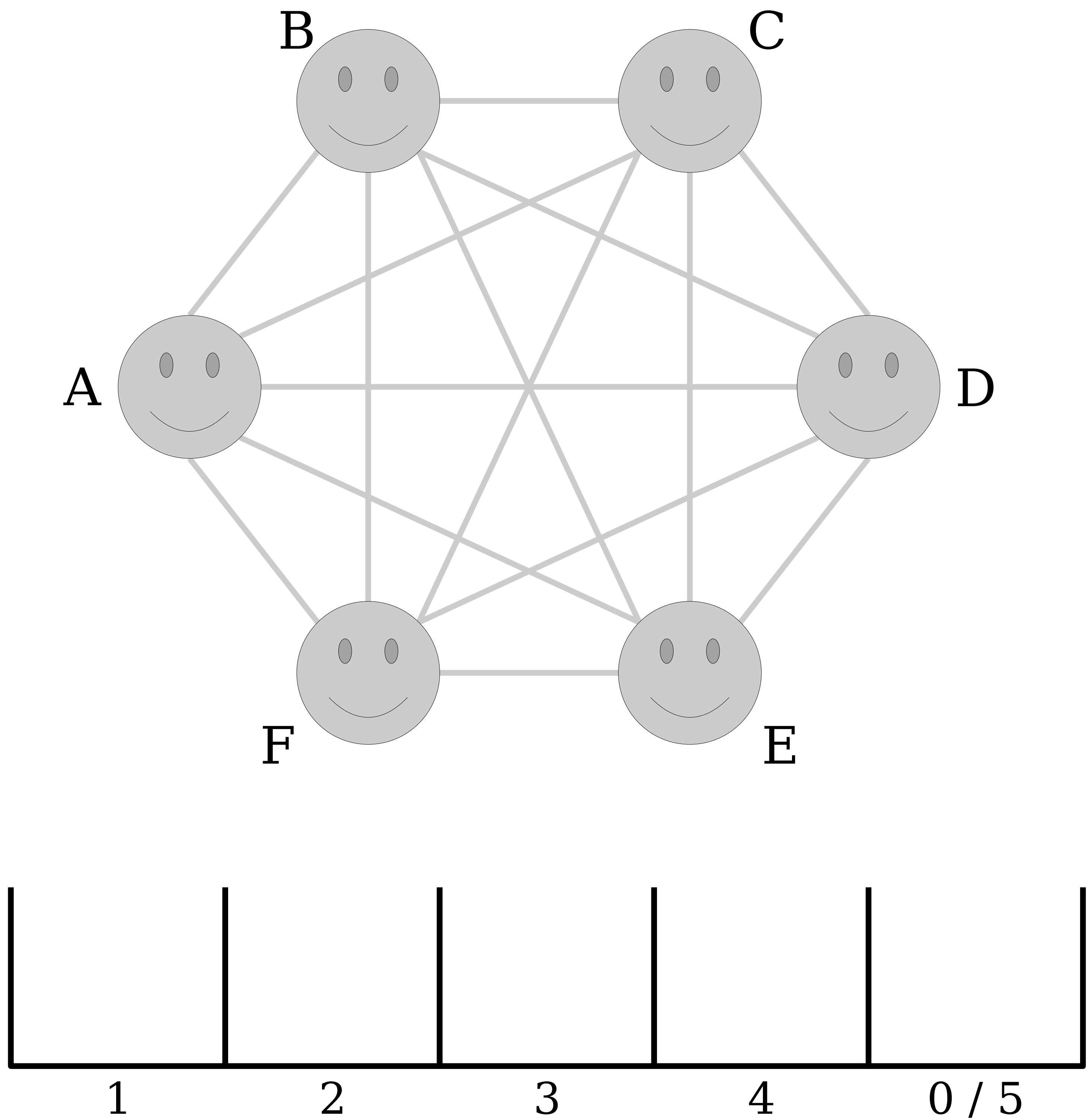

B
C
A
D
F
E
F
D
B
E
C
A
0
1
2
3
4
5



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




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.

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.

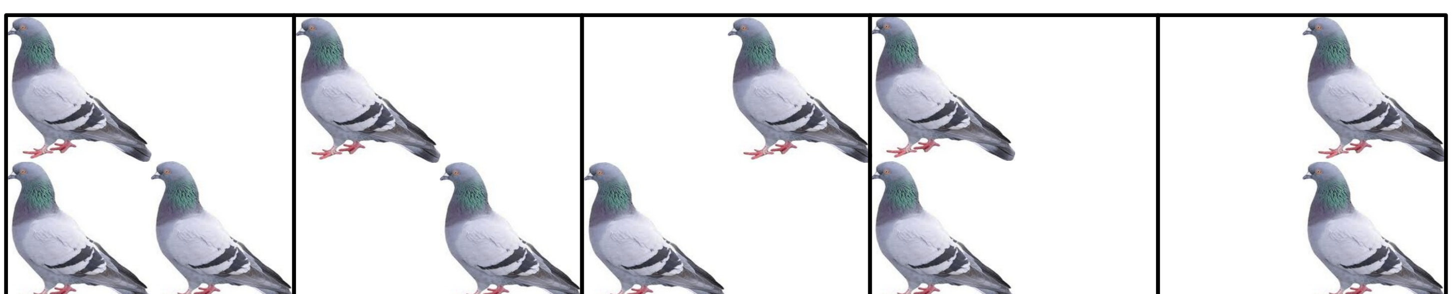
We have reached a contradiction, so our assumption must have been wrong. Thus if G is a graph with at least two nodes, G must have at least two nodes of the same degree. ■

The Generalized Pigeonhole Principle

A More General Version

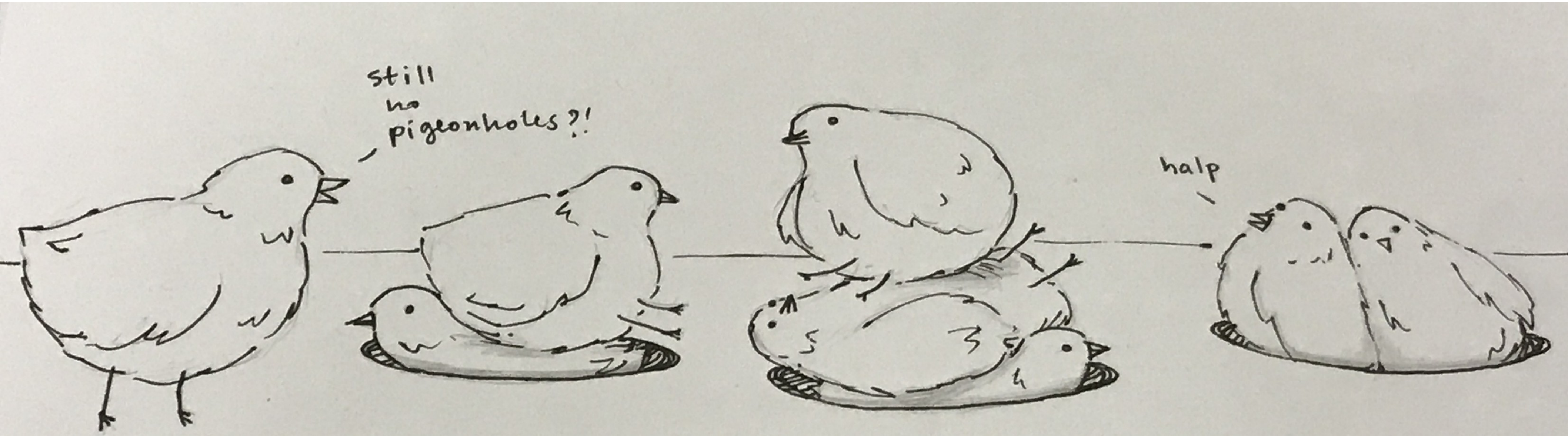
- The ***generalized pigeonhole principle*** says that if you distribute m objects into n bins, then
  - some bin will have at least $\lceil {}^{m}/_{n} \rceil$ objects in it, and
  - some bin will have at most $\lfloor {}^{m}/_{n} \rfloor$ objects in it.

$\lceil {}^{m}/_{n} \rceil$ 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

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

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

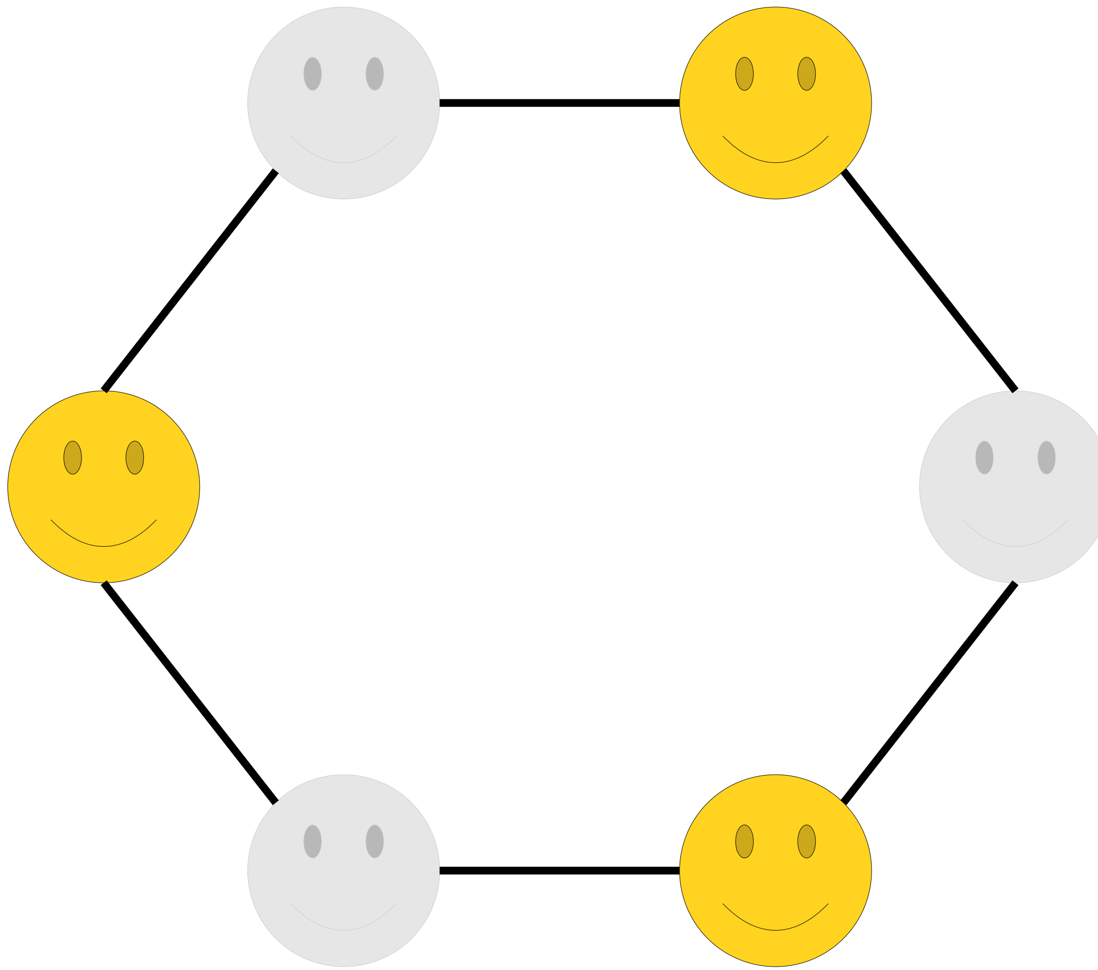
$$\begin{aligned} m &= x_1 + x_2 + \dots + x_n \\ &< m/n + m/n + \dots + m/n \quad (n \text{ times}) \\ &= m. \end{aligned}$$

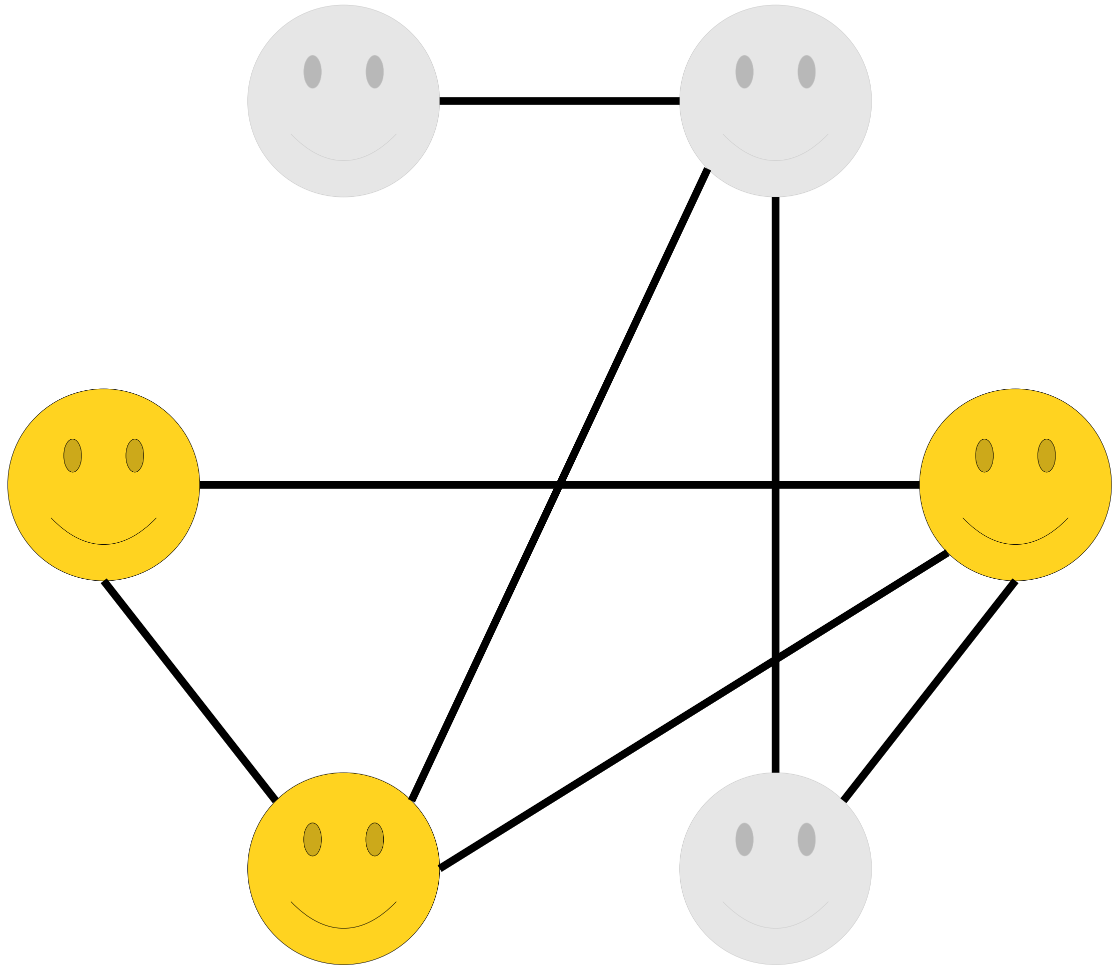
But this means that $m < m$, which is impossible. We have reached a contradiction, so our initial assumption must have been wrong. Therefore, if m objects are distributed into n bins, some bin must contain at least $\lceil m/n \rceil$ objects. ■

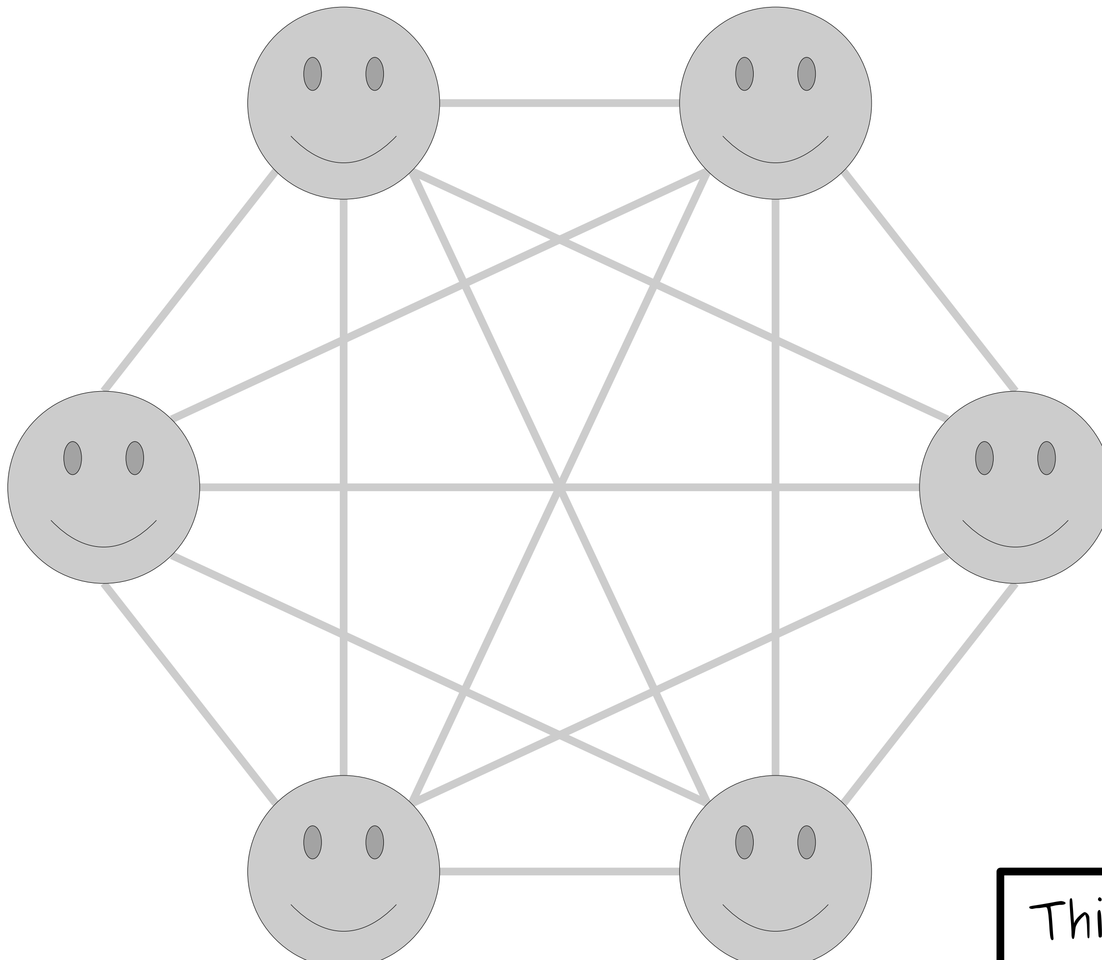
An Application: Friends and Strangers

Friends and Strangers

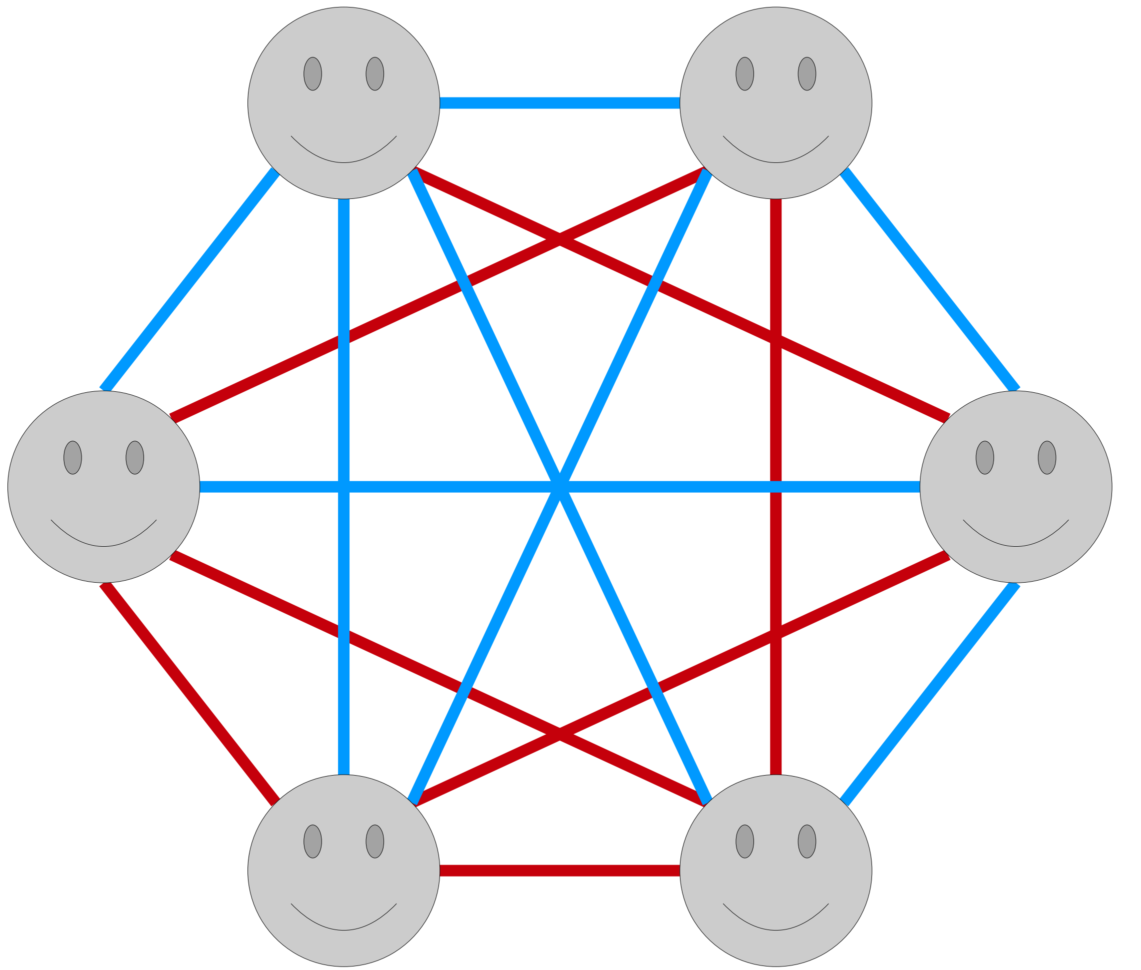
- Suppose you have a party of six people. Each pair of people are either friends (they know each other) or strangers (they do not).
- ***Theorem:*** Any such party must have a group of three mutual friends (three people who all know one another) or three mutual strangers (three people, none of whom know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

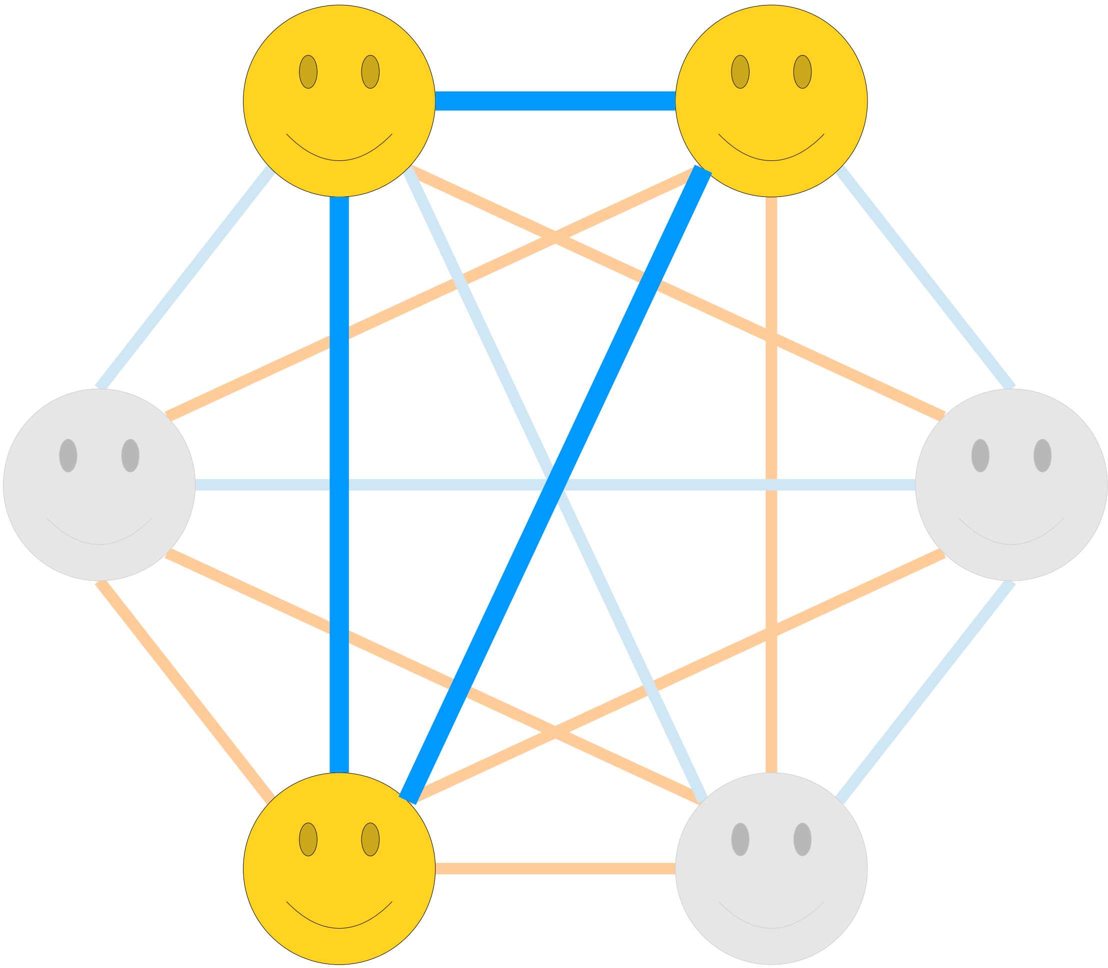


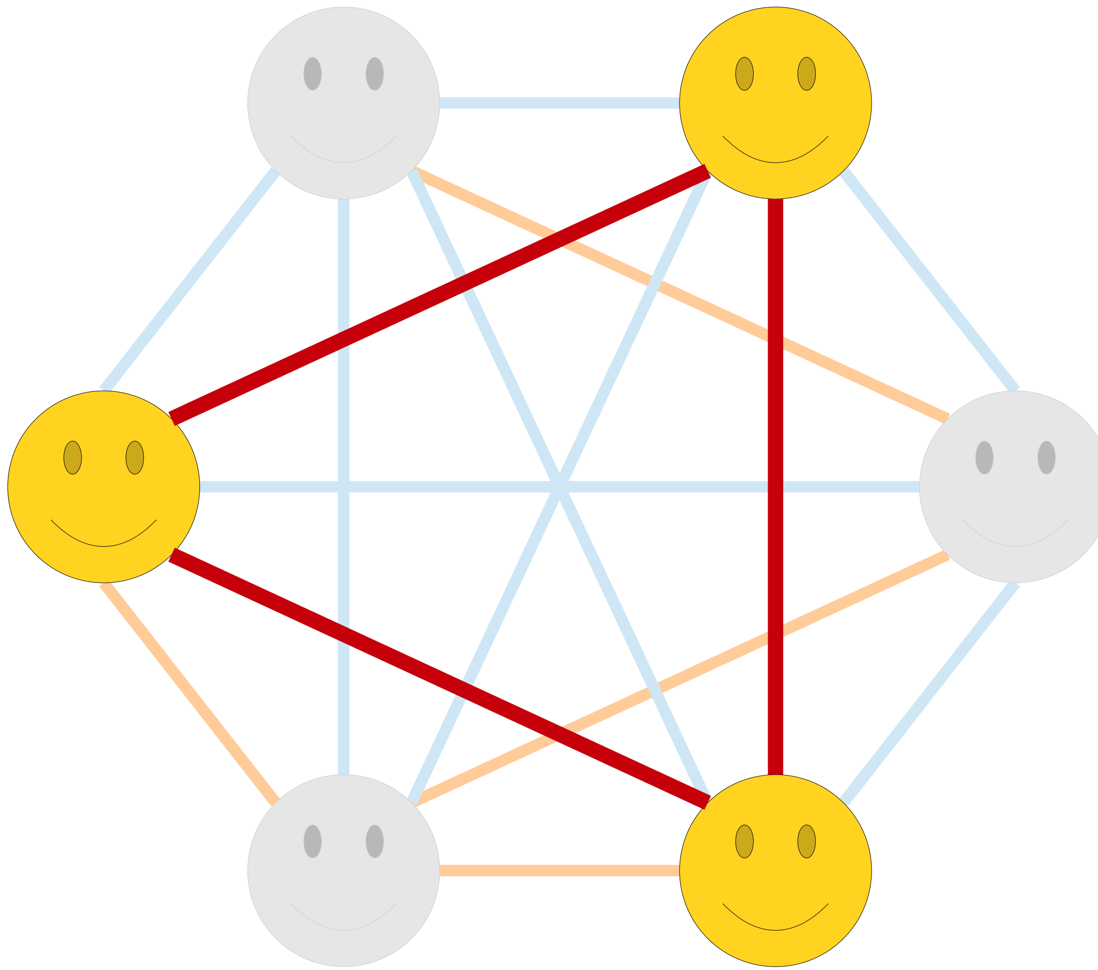




This graph is called a 6-clique, by the way.





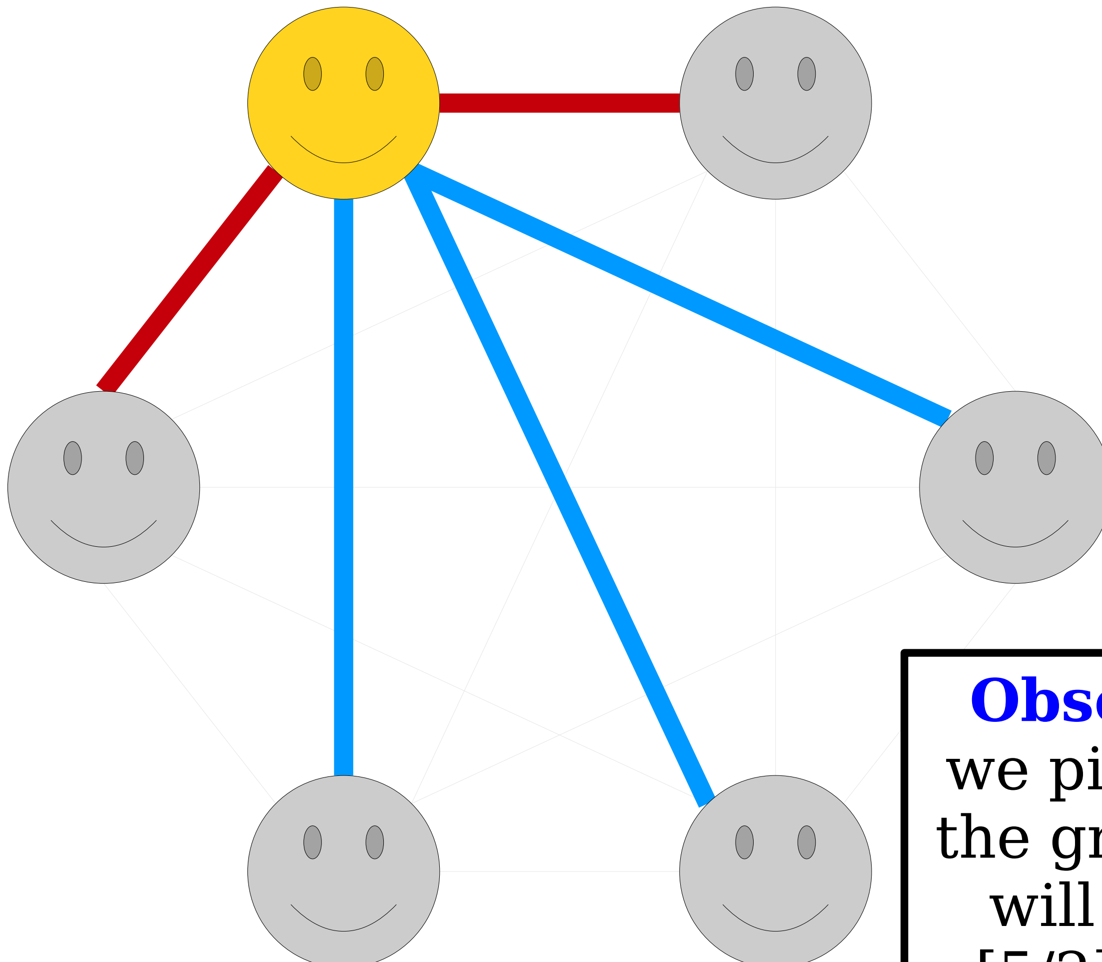


Friends and Strangers Restated

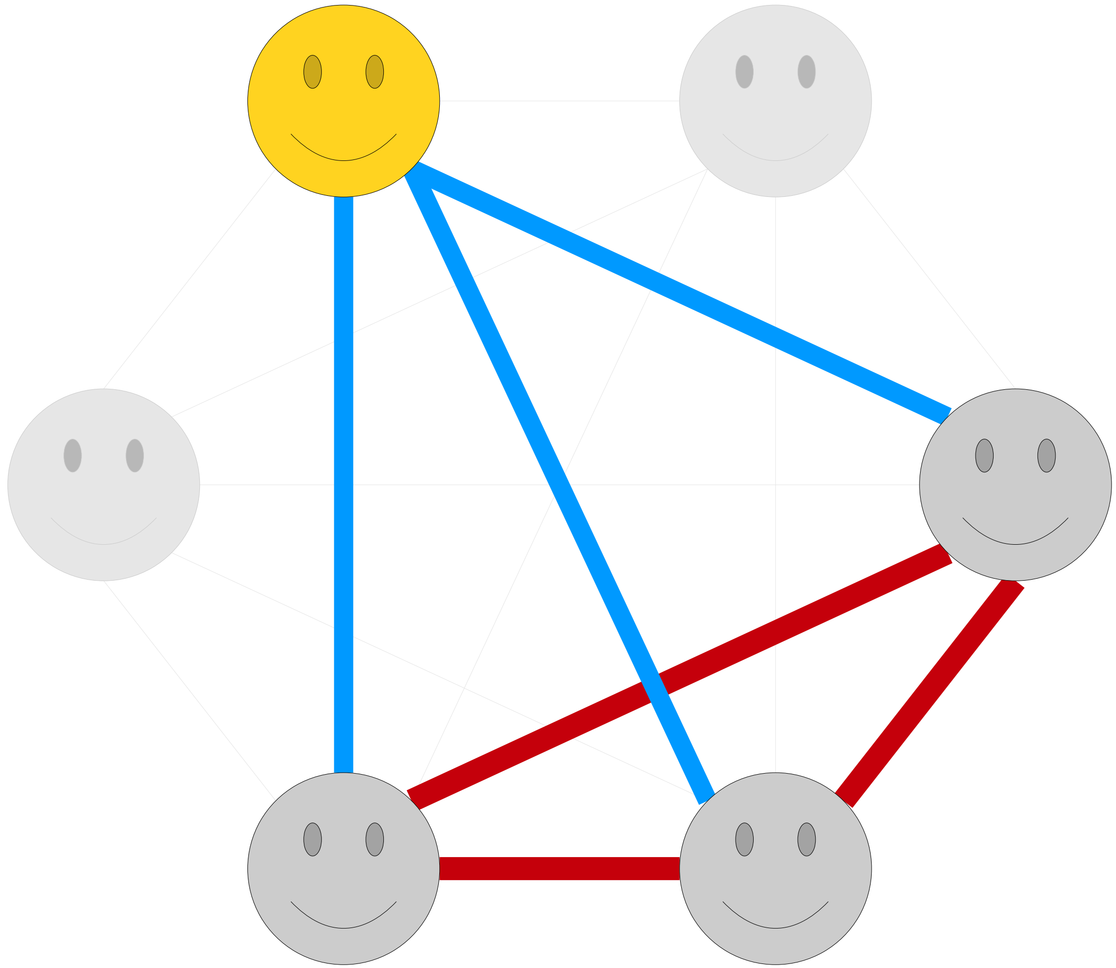
- From a graph-theoretic perspective, the Theorem on Friends and Strangers can be restated as follows:

  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: We need to show that the colored 6-clique contains a red triangle or a blue triangle.

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. 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 triangle. Otherwise, all three of those edges are red, and they form a red triangle. Overall, this gives a red triangle or a blue triangle, as required. ■

Ramsey Theory

- The theorem we just proved 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.

Time-Out for Announcements!

Midterm 1

- Our first midterm exam went out today at 2:30PM Pacific. It's due Sunday at 2:30PM Pacific.
- The exam is open-book, open-note, open-internet, and closed-other-humans.
  - You can ask us clarifying questions about what the problems are asking on EdStem, as long as you post privately.
  - You cannot communicate with other humans about this exam, search online for answers to the questions, or solicit answers from other people.
- ***You can do this.*** Best of luck on the exam!

Problem Sets

- Problem Set Three was due today at 2:30PM.
- Problem Set Four goes out today. It's due next Friday at 2:30PM.
  - It's all about graphs and graph theory, and you'll see some really cool results!
  - Because the midterm goes out today and is due Sunday, we don't expect you to look at this one until Monday. We've made it shorter than the previous problem sets.

Your Questions

"What is a class that you struggled the most with at Stanford and how did you pull it through? (ignore if you never struggled with classes)"

Oddly enough, that would be CS103! Pretty much everything in that class was totally new to me, I was surrounded by a bunch of people who had way more experience than I did, and I hadn't yet learned how to take advantage of the resources given to me.

In retrospect, I was approaching the class the wrong way. I was trying to solve everything on my own and was spending hours upon hours turning over the problems in my head without drawing pictures, writing things down, or even considering going to office hours. I thought that was how you were "supposed" to be doing things. I also had atrocious proof style (e.g. proofs that spanned three handwritten pages) because, at the time, no one was pointing out to me that I wasn't supposed to do that.

Now having taught CS103 for a decade, it's amazing to see how much I learned in that class. I definitely didn't get much of what we covered, and it took a few years for me to really get how to think about some of the topics. I still wonder how much more I would have learned if I hadn't been siloed off from other students and if I had been more strategic about asking for big-picture advice.

“Favorite albums of all time?”

In unsorted order:

- “Nurture” by Porter Robinson. It’s a reflection on hitting a creative/artistic block and trying to figure out how to find the motivation to move forward.
- “Little Big” and “Little Big II: Dreams of a Mechanical Man” by Aaron Parks. Exceptional modern jazz that gets better after every listen.
- “Live at the Quick” by Béla Fleck and the Flecktones. Fantastic live album showcasing a mix of performers having a great time.
- “Random Access Memories” by Daft Punk. Daft Punk put so much effort into getting everything in this album to sound just right over a huge range of styles.
- “City Folk” by James Farm. Stellar jazz album that never gets old.
- “MTV Unplugged in New York” by Nirvana. There’s a reason this group had such an outsized impact on the direction of popular music.
- “The Rise and Fall of Ziggy Stardust and the Spiders from Mars” by David Bowie. Starts strong and never lets up.
- “Remain in Light” by Talking Heads. This album doesn’t sound like anything else I’ve ever heard and has many standout tracks.
- “Kind of Blue” by Miles Davis. Wow, what an album. Everything here ages well.

Back to CS103!

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.

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.

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. ■

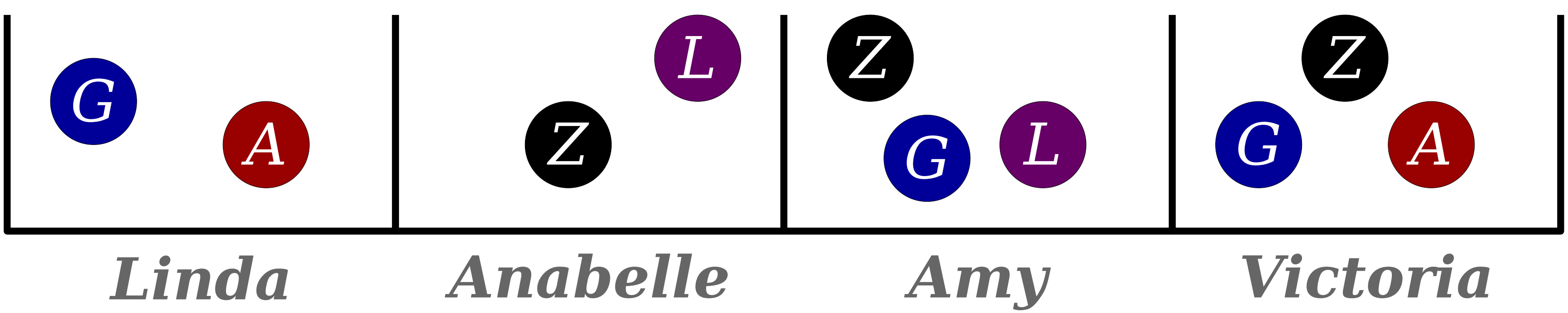
"In a group of $n > 0$ people …

- 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*.

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!)*

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

- ***No class on Monday – take a break!***
- ***Then, when we get back:***
  - ***Mathematical Induction***
    - Reasoning about stepwise processes!
  - ***Applications of Induction***
    - To numbers!
    - To anticounterfeiting!
    - To puzzles!