1. Reviewing Graph Definitions

Graphs come with a lot of specific terminology. Here are some questions to review the terms.

a. What does it mean for two nodes to be adjacent in a graph?

> Two nodes are adjacent if they're linked by an edge. Formally, if $G = (V, E)$ is a graph, then nodes $u, v \in V$ are adjacent if $\{u, v\} \in E$.

b. Can undirected graphs have self-loops? Can directed graphs? Why/why not?

> Undirected graphs cannot have self-loops, because then the edge representing the self-loop would look like $\{v, v\}$, which is equivalent to $\{v\}$, a set containing one node, instead of the required two nodes.
>
> Directed graphs can have self-loops, because their edges are ordered pairs instead of sets, which means having repeated elements is OK.

c. What is a walk in a graph? How do you measure the length of a walk? What is a closed walk?

> A walk in a graph is a sequence of nodes (an ordered list where repeats are allowed) where any two nodes next to each other in the sequence are adjacent in the graph.
>
> The length of a walk is the number of nodes in the list, minus one. You can think of this as the number of edges followed by the walk.
>
> A closed walk is a walk where the first and last node are the same. A closed walk can't have length 0, i.e. it has to have multiple nodes.

d. What is a path in a graph? How is it different from a walk? Can a path be a closed walk?

> A path in a graph is a walk without any repeated nodes. (By the definition of edges, this also means no edges followed by the path are repeated.)
>
> A path is a type of walk. Paths have a condition on no repeats, so they cannot be closed walks.

e. What is a cycle in a graph? How is it different from a path?

> A cycle in a graph is a closed walk that doesn't repeat any nodes except the very first/last node. (By the definition of edges, this also means no edges followed by the cycle are repeated.)
>
> A cycle is not a type of path. A cycle is different from a path because a cycle has a

repeated node (the first/last) and a path has no repeats.

f. What does it mean for two nodes to be connected in a graph? (This is also sometimes called being reachable.) How is this different from being adjacent?

Two nodes are connected if there's a path that starts at the first node and ends at the second node.

g. Is it possible for two nodes in a graph to be adjacent but not connected?

No, this is not possible. If two nodes are adjacent, then those two nodes together form a path.

h. Is it possible for two nodes in a graph to be connected but not adjacent?

Yes, this is possible. For a simple example, consider a graph with three nodes a, b, c, and two edges, one between a and b and one between b and c. Then, a and c are connected, since (a, b, c) forms a path between them. But a and c are not adjacent, since there's no edge between them.

i. What does it mean for a graph to be connected?

For any two nodes in the graph, those nodes must be connected.

j. What is a connected component in a graph?

A connected component is a non-empty set of nodes where (1) any two nodes in the set are connected and (2) any node in the set is not connected to any node outside the set. You can also think of a connected component as one node, and every node connected to that node.

k. How many connected components does each node in a graph belong to?

Each node in a graph belongs to exactly one connected component in that graph.

l. What is meant by the degree of a node in an undirected graph? What about the in-degree and the out-degree in a directed graph?

In an undirected graph, the degree of a node is the number of edges touching it. Equivalently, it's the number of nodes that the node is adjacent to.

In a directed graph, the in-degree of a node is the number of incoming edges touching it, and the out-degree is the number of outgoing edges coming from it.

2. Applying Definitions on Graphs

Many proofs on graphs deal with properties expressed in first-order logic. This problem demonstrates some strategies you can use when interpreting new formal definitions.

Given a graph $G = (V, E)$, G is called **triangle-free** if the following property holds:

$$\forall u \in V.\ \forall v \in V.\ \forall w \in V.\ \big((\{u, v\} \in E \land \{v, w\} \in E) \to \{u, w\} \notin E\big)$$

Given a graph $G = (V, E)$ and a specific node $u \in V$, let the **neighborhood set** of u be the set $\{v \in V \mid (u, v) \in E\}$.

Finally, here's the formal definition of an **independent set** again: In a graph $G = (V, E)$, a set $I \subseteq V$ is an independent set if the following property holds:

$$\forall u \in I.\ \forall v \in I.\ \{u, v\} \notin E$$

We'll prove the following property of graphs: For any triangle-free graph $G = (V, E)$ and node $u \in V$, the neighborhood set of u is an independent set.

a. One strategy for approaching definitions is to start by *breaking down the structure of the definition.* We'll do this with the triangle-free definition.

(1) In the definition of triangle-free, do the variables u, v, and w represent nodes or edges of the graph?

> They are nodes, since they are elements in V.

(2) Is the definition universally or existentially quantified?

> Universally quantified.

(3) If you wanted to check if a graph *is* triangle-free, what would you have to do: find a specific counterexample, or check a property holds for all possible choices of u, v, w?

> Since the formal definition is universally quantified, we would have to check that the property holds for all possible choices of u, v, w. In other words, we'd have to look at every combination of three nodes in the graph and make sure that if there is an edge between u and v and an edge between v and w, then there is no edge between u and w.

(4) If you wanted to check if a graph is *not* triangle-free, what would you have to do: find a specific counterexample, or check a property holds for all possible choices of u, v, w? (Hint: Take the negation of the definition.)

Since the negation of the definition is existentially quantified, we would have to find a specific counterexample. Just finding three nodes u, v, and w where there are edges between all three is enough to show that a graph is not triangle-free.

(5) Explain the triangle-free property in words.

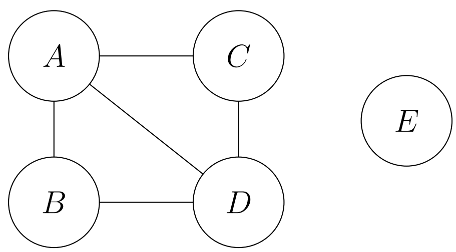
Here are some ways to explain it:

A triangle-free graph has no cycles of length 3.

A triangle-free graph has no copies of K_3 in it.

A triangle-free graph has no trios of nodes where all of them are adjacent to each other.

b. Another strategy is to *try small examples to explore the definition.* We'll do this with the neighborhood set definition. Here's a small graph we can use to try this out.



(1) Could a node ever be in its own neighborhood set? Why or why not?

No, because its neighborhood set is a set of nodes that are adjacent to it. In an undirected graph, a node cannot be adjacent to itself. (Equivalently, we cannot have self-loops.)

(2) What is the neighborhood set of C? How about D? And E?

To check your answers, you should be able to point to each node that is in your neighborhood set and justify to yourself why it meets the criteria on the right-hand-side of the set-builder notation. You should also be able to point to every node that is *not* in your neighborhood set and justify to yourself why it doesn't meet those criteria.

The neighborhood set of C is $\{A, D\}$.

The neighborhood set of D is $\{A, B, C\}$.

The neighborhood set of E is $\varnothing$. There are no nodes adjacent to E.

(3) Explain what a neighborhood set is in words.

> The neighborhood set of a node is the set of nodes adjacent to it.

c. Bonus: let's try to connect the definitions along the lines of the proof.

(1) Which of the neighborhood sets you found in (b)(2) are independent sets?

> C's neighborhood set, $\{A, D\}$, is not an independent set, since A and D are adjacent.
>
> D's neighborhood set, $\{A, B, C\}$, is not an independent set since A and C are adjacent and A and B are also adjacent.
>
> The empty set is an independent set (the definition is vacuously true.)

(2) The graph from part (b) is not triangle-free. What is the smallest number of edges you have to remove to make it triangle-free? (Note: We will not be talking about removing edges from a graph in the proof. This is just to get a look at an actually triangle-free graph.)

> We can make this graph triangle-free by removing just one edge: the edge between A and D.
>
> Notice how this would change the neighborhood sets we found in the previous part. C's neighborhood set would still be $\{A, D\}$, but this is now an independent set because they are not connected by an edge any more. Meanwhile, D's neighborhood set would become $\{B, C\}$, since A and D are no longer linked by an edge. This is an independent set since B and C are not adjacent.

d. With a better intuitive understanding of the definitions, let's move on to *setting up the proof.*

Here's the theorem we're trying to prove again: For any triangle-free graph $G = (V, E)$ and node $u \in V$, the neighborhood set of u is an independent set.

(1) What should our assume and want-to-show be? How should the graph G and node u be picked: can they be picked arbitrarily by the reader, or should you give a specific example?

> The reader should pick the graph G and node u, since the statement is universally quantified.
>
> We will assume that G is triangle-free and prove that the neighborhood set of u is an independent set.

(2) Write out the definitions of the properties in your assume and want-to-show columns. (Since we have a specific node u mentioned in the problem, you'll probably want to

change the variable names in the definitions.)

Assume: G is a triangle-free graph, and u is a node in G. The definition of triangle-free is:

$$\forall a \in V.\ \forall b \in V.\ \forall c \in V.\ \big((\{a,b\} \in E \land \{b,c\} \in E) \to \{a,c\} \notin E\big)$$

We also know the neighborhood set of u is the set of nodes $\{v \in V \mid \{u,v\} \in E\}$. Let's give this set a name, N, so we can refer to it later.

Want to show: N is an independent set. The definition of an independent set as applied to N would be:

$$\forall x \in N.\ \forall y \in N. \{x,y\} \notin E$$

(3) We are assuming a universally quantified statement and trying to prove a universally quantified statement. Which variables should we introduce — corresponding to the variables in the statement we're assuming, or corresponding to the statement we're trying to show?

We should introduce variables corresponding to the statement we're trying to show. In other words, we should introduce variables corresponding to x and y, the elements of N from our independent set definition above.

Knowing that G is triangle-free means that, *given* any three nodes a, b, c in V, we can say something about the edges linking them. But we should *not* introduce three variables corresponding to nodes in V.

e. After setting up the proof, work it out using the two-column organizer and write your proof.

Proof: Let $G = (V, E)$ be an arbitrary triangle-free graph, let u be an arbitrary node in V, and let N be the neighborhood set of u. We'll show that N is an independent set.

To do so, pick two nodes x, y from N. We'll show that there is no edge $\{x,y\}$ in E. Notice that because x and y are in N, we know that $\{x,u\} \in E$ and $\{y,u\} \in E$. Then, because G is triangle-free, we know that $\{x,y\} \notin E$, which is what we wanted to show. ■

3. Applying the Pigeonhole Principle

a. Say there are 8,000 undergrads at Stanford. There are 366 possible birthdays. Fill in the blanks with the largest possible number guaranteed by the Pigeonhole Principle:

(1) There must be at least __________ undergrad(s) who have the same birthday as one another.

> We can use the generalized pigeonhole principle, with the "objects" being undergrads and the "bins" being birthdays. By the generalized pigeonhole principle, there must be some birthday with at least $\lceil \frac{\text{number of undergrads}}{\text{number of birthdays}} \rceil$ undergrads having that birthday, which works out to 22 people.

(2) There must be at least __________ undergrad(s) who were born on February 29.

> Here, the pigeonhole principle doesn't actually apply! The generalized pigeonhole principle says *some* bin (or birthday) has to have at least or at most some number of people in it; it doesn't guarantee *which* bin/birthday has that many people. It could be the case that no undergrads were born on February 29. So we are not guaranteed any larger than 0, which is the answer.

b. Fill in the blank with the smallest possible number guaranteed by the Pigeonhole Principle:

(1) There is some day of the year with at most __________ undergrad(s) who have that birthday.

> The part of the generalized pigeonhole principle that applies here is: given 8,000 people and 366 birthdays, there is some birthday with at most $\lfloor \frac{8000}{366} \rfloor$ people having that birthday, which works out to 21 people. (It's probably February 29, but the Pigeonhole Principle doesn't guarantee anything about that.)

(2) There are 7 days of the week. If we have a group of __________ people, we are guaranteed that two of them were born on the same day of the week.

> The pigeonhole principle says that if we have a number of people *strictly* greater than the number of days of the week, then at least two of them share a day of the week. So the answer is 8 people, which is the smallest number greater than 7.

(3) There are 26 letters of the alphabet. If we have a group of __________ people, we are guaranteed that at least 5 of them have the same first initial.

> The part of the generalized pigeonhole principle that applies here is: given m people and 26 first initials, there is some first initial shared by at least $\lceil \frac{m}{26} \rceil$ of them. The smallest number of people that works is $4 \cdot 26 + 1$, which works out to 105 people. If we had chosen the next smallest number, $4 \cdot 26 = 104$, then we would only be guaranteed that at least 4 people had the same first initial.

c. Suppose you pick 11 numbers from the set $\{n \in \mathbb{N} \mid 1 \leq n \leq 20\}$. We'll prove that out of those 11 numbers, there must be at least one pair of numbers whose difference is exactly 10.

- First, try this exercise yourself: write down eleven numbers, and say which ones differ by exactly 10. Then, write down ten numbers without including any pairs of numbers that differ by exactly 10. (There are many ways to do this!)

  What happens when you try to add an eleventh number to your list of ten? Can you generalize this into a claim about having too many "pigeons" for your "holes"?

It's possible to write down 10 numbers where none of them differ by exactly 10. For example: 1, 2, 3, 4, 5, 6, 7, 8, 9, 10. Or: 1, 12, 3, 14, 5, 16, 7, 18, 9, 20. However, once we add an 11th number, we run into an issue: no matter which remaining number we pick, it will differ by exactly 10 with another number that we already have! Since we want to show something about the number of pairs of numbers with difference 10, it would be helpful to treat the 10 pairs of numbers as our "holes" and our 11 choices of numbers as "pigeons".

Proof: Consider the ten pairs of numbers $\{1, 11\}, \{2, 12\}, \{3, 13\}, ..., \{10, 20\}$ as our "bins", and our choices of numbers as objects. For example, if we chose the number 2, we would put that object in the $\{2, 12\}$ bin. Since we have 11 objects/choices and 10 bins/categories, by the pigeonhole principle, some bin/category must have two numbers in it.