GRAPH THEORY

Graph isomorphism

Definition 19.1 Two graphs $G_1 = (V_1, E_1)$ and $G_2 = (V_2, E_2)$ are isomorphic, write $G_1 \simeq G_2$, when there is a bijection $\beta : V_1 \rightarrow V_2$ such that for any two distinct $u_1, v_1 \in V_1$ we have $u_1 - v_1 \in E_1$ iff $\beta(u_1) - \beta(v_1) \in E_2$.

Examples of isomorphic, and non-isomorphic graphs will be given in class.

We encourage you to develop a good intuition for graph isomorphism. This intuition should convince you without a tedious proof, for example that when $G_1 \simeq G_2$ then $G_1$ is connected iff $G_2$ is connected.

Recall the definition of the complete graph on $n$ vertices, which we denote by $K_n$. A graph $G$ is said to be complete when it is isomorphic to $K_n$ for some $n$. The isomorphism is simply renaming the vertices, but the crucial property that there is an edge between any two vertices hold in any complete graph.

Subgraphs

So far we have defined connected components as blocks of a partition on $V$, therefore as sets of vertices. In fact, the same term is used for the closely associated subgraph.

Definition 19.2 A graph $G_1 = (V_1, E_1)$ is a subgraph of the graph $G_2 = (V_2, E_2)$ when $V_1 \subseteq V_2$ and $E_1 \subseteq E_2$. (Beware: not all pairs of such subsets form graphs!)

If $G = (V, E)$ is a graph and $V' \subseteq V$ is a set consisting of some of $G$’s nodes, the subgraph of $G$ induced by $V'$ is the graph $G' = (V', E')$ where $E'$ consists of all the edges of $G$ whose both endpoints are in $V'$.

We have defined a connected component as a maximally connected set of nodes $C \subseteq V$. We shall use the same name for the subgraph induced by $C$.

The vertices and edges of a walk form a subgraph, which in some sense “corresponds” to that walk. Note that in the walk, which is a sequence, vertices and even edges may occur multiple times.
However, in the subgraph definition (as in any graph definition) vertices and edges occur exactly once. Hence, multiple walks may correspond, in general, to the same subgraph.

Similarly, the vertices and edges of a path form a subgraph. Because vertices and hence, edges, only occur once in a path we can reconstruct the path uniquely out of the corresponding subgraph. Thus we define:

**Definition 19.3** For any \( n \in \mathbb{N} \), we call a subgraph \( S \) a path of length \( n \) when \( S \cong P_{n+1} \) (the path graph on \( n+1 \) vertices).

**Counting paths** Consider the path \( u-v-w \). According to the definition (paths are sequences of nodes, etc.), \( w-v-u \) is a different path. This seems a bit awkward in undirected graphs: since there is no direction on edges, what is the meaning of direction on paths?

Not being concerned with direction must be done carefully, however. Here is the problem with identifying paths just by their sets of node:

**Example 19.4** Consider \( K_4 \), the complete graph on 4 nodes, \( \{1, 2, 3, 4\} \). Find two paths that have the same set of nodes, but different sets of edges, therefore, they correspond to different subgraphs.

**Solution:** Consider the following paths 1-2-3-4 and 1-3-2-4. They have the same set of nodes, \( V = \{1, 2, 3, 4\} \), but different sets of edges: \( E_1 = \{1-2, 2-3, 3-4\} \) and \( E_2 = \{1-3, 3-2, 2-4\} \) hence they correspond to the distinct subgraphs: \((V, E_1)\) and \((V, E_2)\).

Therefore, when we count paths in a graph, we will count the subgraphs that “are” paths (according to Definition 19.3). In particular, when we count paths of length \( n \) we will count subgraphs isomorphic with \( P_{n+1} \).

**Example 19.5** How many paths of length 2 are there in \( C_3 \)? How about in \( C_4 \)?

**Solution:** Recall that \( C_3 = (\{1, 2, 3\}, \{1-2, 2-3, 3-1\}) \). And recall that we count how many subgraphs of \( C_2 \) are isomorphic to \( P_3 \).

Any two edges form a path of length 2 since they must have an endpoint in common. Since any path of length 2 is made of two edges that have an endpoint in common there is a bijection in \( C_3 \) between paths of length 2 and sets of two edges. There are \( \binom{3}{2} = 3 \) of the latter hence there are 3 paths of length 2. As subgraphs all three have the same set of vertices, \( \{1, 2, 3\} \) and they have the following sets of edges: \( \{1-2, 2-3\} \), \( \{2-1, 1-3\} \), and \( \{1-3, 3-2\} \).

In \( C_4 = (\{1, 2, 3, 4\}, \{1-2, 2-3, 3-4, 4-1\}) \) we still have a bijection between paths of length 2 and sets of two edges that have an endpoint in common, but not any two edges do have an endpoint in common. So we count differently.
Note that there is a bijection between paths of length 2 and the vertices of $C_4$ by associating with each path of length 2 its middle vertex. Therefore there are 4 paths of length 2.

For conciseness we describe them as follows: $1\rightarrow 2\rightarrow 3$, $2\rightarrow 1\rightarrow 4$, $4\rightarrow 3\rightarrow 2$, $1\rightarrow 4\rightarrow 3$, i.e., we write, for example, $1\rightarrow 2\rightarrow 3$ instead of $\{(1,2,3),\{1-2,2-3\}\}$, etc. Note that this same subgraph could have also be described as $3\rightarrow 2\rightarrow 1$. ■

**Example 19.6** How many paths are there in $P_n$ (the path graph on $n \geq 1$ vertices)?

**Solution:** The paths of length 0 correspond to subgraphs consisting of a single node. There are exactly $n$ of these. This also the answer when $n = 1$.

Now assume $n \geq 2$. We observe that in $P_n = ([1..n],\{1-2,\ldots,(n-1)-n\})$ the paths of length $\geq 1$ are one-to-one correspondence with the pairs of nodes $(i,j)$ such that $i < j$. Indeed, $i$ and $j$ are the endpoints of the path, which also contains all the the nodes and edges in-between. One way to count these is in steps: in the first step we choose $i \in [1..(n-1)]$ and in the second step we choose $j \in [(i+1)..n]$, for a total of

$$\sum_{i=1}^{n-1} (n-i) = \sum_{i=1}^{n-1} i = \frac{(n-1)n}{2}$$

Another way is to see that such pairs are in one-to-one correspondence with subsets of size 2 of $[1..n]$ and we get the same answer: $\binom{n}{2}$. Therefore, for $n \geq 2$, there are $n + \binom{n}{2}$ paths in $P_n$. ■

**Cycles**

**Definition 19.7** A **closed walk** is a walk in which the first and the last vertex are the same. A **cycle** is a closed walk of length at least 3 in which all nodes are pairwise distinct, except for the last and the first.

The **length** of the cycle is the length of the closed walk.

Hence, walks of length 0 are not cycles. Walks of the form $u-v-u$ are not cycles either. And, there are no closed walks of length 1. Hence, a cycle cannot have length 0, 1, or 2.

**Counting cycles** A cycle has as many nodes as edges but since there is no cycle of length $\leq 2$, we need at least three distinct nodes and thus three distinct edges.

\[1\text{Warning! Some textbooks call “cycle” what we call “closed walk” and they call “simple cycle” what we call “cycle”}.\]
We can have multiple closed walks “corresponding” to the same subgraphs, however note that a cycle can “almost” be reconstructed from the corresponding subgraph, the difference being only that we can start the cycle sequence at any of its vertices. Thus, we give the following:

**Definition 19.8** For any \( n \in \mathbb{N} \), we call a subgraph \( S \) a cycle of length \( n \) when \( S \cong C_n \) (the cycle graph on \( n \) vertices).

Again we want to count cycles in such a way that direction does not matter. Again counting just sets of nodes that form a cycle does not work, e.g., see the following:

**Example 19.9** Consider \( K_4 \), the complete graph on 4 nodes, \( \{1, 2, 3, 4\} \). Find two cycles that have the same nodes, but different edges (and thus we want to count them as different cycles).

**Solution:** We describe them as closed walks: 1-2-3-4-1 and 1-3-2-4-1. (Can you find more?) ■

Therefore, when we count cycles in a graph, we will count the subgraphs that “are” cycles (by the definition above). In particular, when we count cycles of length \( n \) we will count subgraphs isomorphic with \( C_n \).

**Trees and forests**

**Definition 19.10** A graph in which there are no cycles is called acyclic. A graph that is both connected and acyclic is called a tree. Consequently, an acyclic graph is also called a forest since all its connected components are trees!

We skip the tedious proof that when \( G_1 \cong G_2 \) then \( G_1 \) is acyclic iff \( G_2 \) is acyclic. Since we have remarked that the same holds for connectivity, it follows that \( G_1 \) is a tree iff \( G_2 \) is a tree.

**Proposition 19.11** If \( G = (V, E) \) is acyclic (a forest) then

\[
|E| = |V| - |CC|
\]

In particular, \( G = (V, E) \) is a tree then \( |E| = |V| - 1 \).

**Proof:** Actually, we prove the result for trees and then we derive the one for forests by adding up the respective equations for each connected component.

A leaf in a graph is a node of degree 1. Let’s use “edgy” for the negation of edgeless. An edgy graph has at least one edge.
Lemma 19.12 Every edgy tree has at least one leaf (actually, at least two!).

Proof: (of Lemma 19.12). Let $G = (V, E)$ be an edgy tree. Consider the set $L \subseteq \mathbb{N}$ of lengths of paths (not walks!) in $G$. Since paths cannot have length more than $|E|$, $L$ is finite. WOP implies that $L$ has a greatest element, $m$. Because $G$ is edgy, $m \geq 1$.

Therefore, there exists in $G$ a path of maximum length and that path has at least two nodes (it cannot be the path of length 0).

We have an opportunity to define an important concept:

Definition 19.13 A path is called maximal if it cannot be extended with another edge at either end and still remain a path.

A path of maximum length is of course maximal. However, we can have maximal paths that are not of maximum length (example in class).

Now back to the proof of the lemma. We claim that the (distinct) endpoints of a maximal path are leaves. This follows by contradiction (in class).

Lemma 19.14 Removing a leaf from an edgy tree leaves again a tree.

Proof: (of Lemma 19.14). Indeed, the resulting graph is still acyclic. It is also still connected because the only paths affected have the removed leaf as an endpoint.

Back to the proof of the proposition. With these two lemmas we can prove, by induction on $n$ that for any $n$, for any tree, if the tree has $n$ vertices then it has $n - 1$ edges.

Base Case: $n = 1$. Then $m = 0$ and the proposition is true.

Induction Step: Let $k$ be an arbitrary natural number $\geq 1$. Assume (IH) that every tree with $k$ vertices has $k - 1$ edges.

We want to prove that a tree, $G$, with $k + 1$ vertices has $k$ edges.

Note that $k + 1 \geq 2$ so $G$ has at least two distinct nodes. There must be a path between these two nodes and this path must have at least one edge.

Therefore $G$ is an edgy tree. By Lemma 19.12, $G$ has at least one leaf (actually two, but we don’t need this here). Let $G'$ be the graph obtained by removing this leaf. By Lemma 19.14, $G'$ is still a tree. It has $k + 1 - 1 = k$ nodes, so we can apply the IH. Therefore, $G'$ has $k - 1$ edges.

When we removed the leaf from $G$ we removed exactly one edge, because the degree of the leaf is 1. Hence $G$ has one more edge that $G'$, i.e., it has $k - 1 + 1 = k$ edges.