NDMI011: Combinatorics and Graph Theory 1

Lecture #9
Triangle-free graphs and graphs without a C_4 subgraph. Cayley's formula

Irena Penev

1 Graphs without K_3 as a subgraph

A graph is said to be *triangle-free* if it does not contain K_3 as a subgraph. Equivalently, a graph G is triangle-free if $\omega(G) \leq 2$.

The following theorem is a special case of "Turán's theorem."[1]

Mantel's theorem. *Let n be a positive integer. Then*

(a) any triangle-free graph on n vertices has at most $\lfloor \frac{n}{2} \rfloor \lceil \frac{n}{2} \rceil = \lfloor \frac{n^2}{4} \rfloor$ edges;

(b) there exists a triangle-free graph on n vertices that has precisely $\lfloor \frac{n}{2} \rfloor \lceil \frac{n}{2} \rceil = \lfloor \frac{n^2}{4} \rfloor$ edges.

Proof. First, let us check that $\lfloor \frac{n}{2} \rfloor \lceil \frac{n}{2} \rceil = \lfloor \frac{n^2}{4} \rfloor$. If n is even, then this is obvious. If n is odd, then there exists a non-negative integer k such that $n = 2k + 1$, we compute

$$\lfloor \tfrac{n}{2} \rfloor \lceil \tfrac{n}{2} \rceil \quad = \quad \lfloor \tfrac{2k+1}{2} \rfloor \lceil \tfrac{2k+1}{2} \rceil \quad = \quad k(k+1) \quad = \quad k^2 + k$$

and

$$\lfloor \tfrac{n^2}{4} \rfloor \quad = \quad \lfloor \tfrac{(2k+1)^2}{4} \rfloor \quad = \quad \lfloor \tfrac{4k^2+4k+1}{4} \rfloor \quad = \quad k^2 + k,$$

and we deduce that $\lfloor \frac{n}{2} \rfloor \lceil \frac{n}{2} \rceil = \lfloor \frac{n^2}{4} \rfloor$.

For (b), we observe that the complete bipartite graph $K_{\lfloor n/2 \rfloor, \lceil n/2 \rceil}$ is triangle-free[2] and has precisely n vertices and $\lfloor \frac{n}{2} \rfloor \lceil \frac{n}{2} \rceil$ edges.

It remains to prove (a). We assume inductively that the claim holds for graphs on fewer than n vertices, i.e. that for all positive integers $\widetilde{n} < n$, any

[1] Turán's theorem gives a formula for the maximum number of edges in any K_n-free graph. We omit the details.

[2] Indeed, all bipartite graphs are triangle free.

triangle-free graph on $\widetilde{n}$ vertices has at most $\lfloor \frac{\widetilde{n}^2}{4} \rfloor$ edges. It is clear that the theorem holds for $n = 1$ and $n = 2$. So, we assume that $n \geq 3$, we fix a triangle-free graph G on n vertices, and we show that G has at most $\lfloor \frac{n^2}{4} \rfloor$ edges. If G has no edges, then this is obvious. So assume that G has at least one edge, say uv. Then $G \setminus \{u, v\}$ is triangle-free and has $n - 2$ vertices, and so by the induction hypothesis, it has at most $\lfloor \frac{(n-2)^2}{4} \rfloor$ edges. Further, since G is triangle-free, a vertex in $V(G) \setminus \{u, v\}$ can be adjacent to at most one of u, v, and so the number of edges between $\{u, v\}$ and $V(G) \setminus \{u, v\}$ is at most $|V(G) \setminus \{u, v\}| = n - 2$. Since the edges of G are precisely the edges of $G \setminus \{u, v\}$, plus the edges between $\{u, v\}$ and $V(G) \setminus \{u, v\}$, plus the edge uv, we see that

$$\begin{aligned} |E(G)| &\leq \lfloor \tfrac{(n-2)^2}{4} \rfloor + (n-2) + 1 \\ &= \lfloor \tfrac{n^2-4n+4}{4} \rfloor + n - 1 \\ &= \lfloor \tfrac{n^2}{4} \rfloor, \end{aligned}$$

which is what we needed to show. □

2 Graphs without C_4 as a subgraph

In what follows, we will use the Cauchy-Schwarz inequality (below).

The Cauchy-Schwarz inequality. *All real numbers $a_1, \dots, a_n, b_1, \dots, b_n$ satisfy*

$$\Big(\sum_{i=1}^{n} a_i b_i \Big)^2 \leq \Big(\sum_{i=1}^{n} a_i^2 \Big) \Big(\sum_{i=1}^{n} b_i^2 \Big).$$

Proof. Omitted. □

An *isolated vertex* is a vertex that has no neighbors.

Theorem 2.1. *Let $n \in \mathbb{N}$. Any graph on n vertices that does not contain C_4 as a subgraph has at most $\frac{1}{2}(n + n^{3/2})$ edges.*

Proof. Let G be a graph on n vertices, and assume that G does not contain C_4 as a subgraph. Clearly, we may assume that G has no isolated vertices.[3] Let $d_1, \dots, d_n$ be the degrees of the vertices of G;[4] since G is has no isolated veetices, we see that $d_1, \dots, d_n \geq 1$.

Let $M = \{(v, A) \mid v \in V(G), A \in \binom{N_G(v)}{2}\}$.[5] Now, we will count the number of elements of M in two ways.

[3]Why?

[4]The d_i's are not necessarily distinct; d_i is the degree of the i-th vertex of G.

[5]In other words, M is the set of all ordered pairs $(v, \{u_1, u_2\})$ such that $v \in V(G)$, and $u_1, u_2 \in V(G)$ are two distinct neighbors of v. Note also that $(v, \{u_1, u_2\}) \in M$ if and only if u_1, v, u_2 is a (not necessarily induced) two-edge path of G. So, $|M|$ is in fact the number of (not necessarily induced) two-edge paths in G.

First, for each $v \in V(G)$, there are precisely $\binom{d_G(v)}{2}$ choices of A such that $(v, A) \in M$. So, $|M| = \sum\limits_{v \in V(G)} \binom{d_G(v)}{2} = \sum\limits_{i=1}^{n} \binom{d_i}{2}$.

We now bound $|M|$ above, as follows. Note that the second coordinate of any element of M is simply an element of $\binom{V(G)}{2}$; since $|V(G)| = n$, there are at most $\binom{n}{2}$ choices for the second coordinate of an element of M. On the other hand, since G contains no C_4 as a subgraph, we see that no two distinct elements of M have the same second coordinate. Indeed, suppose that (v_1, A) and (v_2, A) are distinct elements of M; we then set $A = \{u_1, u_2\}$, we and observe that v_1, u_1, v_2, u_2, v_1 is a (not necessarily induced) C_4 in G, a contradiction. So, $|M| \leq \binom{n}{2}$.

We now have that

$$\sum_{i=1}^{n} \binom{d_i}{2} \quad \leq \quad \binom{n}{2}.$$

Obviously, $\binom{n}{2} \leq \frac{n^2}{2}$, and since $d_1, \dots, d_n \geq 1$, we see that $\binom{d_i}{2} \geq \frac{(d_i-1)^2}{2}$ for all $i \in \{1, \dots, n\}$; consequently,

$$\sum_{i=1}^{n} \frac{(d_i-1)^2}{2} \quad \leq \quad \sum_{i=1}^{n} \binom{d_i}{2} \quad \leq \quad \binom{n}{2} \quad \leq \quad \frac{n^2}{2},$$

and it follows that

$$\sum_{i=1}^{n} (d_i - 1)^2 \quad \leq \quad n^2.$$

We now compute:

$$\begin{aligned}
\sum_{i=1}^{n} (d_i - 1) \quad &= \quad \sum_{i=1}^{n} (d_i - 1) \cdot 1 \\
&\leq \quad \sqrt{\sum_{i=1}^{n} (d_i - 1)^2} \sqrt{\sum_{i=1}^{n} 1^2} && \text{by the Cauchy-Schwarz inequality} \\
&= \quad \sqrt{\sum_{i=1}^{n} (d_i - 1)^2} \sqrt{n} \\
&\leq \quad \sqrt{n^2} \sqrt{n} && \text{because } \sum_{i=1}^{n} (d_i - 1)^2 \leq n^2 \\
&= \quad n^{3/2}.
\end{aligned}$$

It now follows that

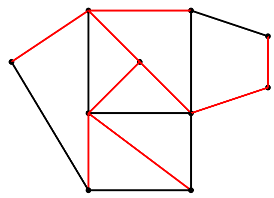
$$|E(G)| \quad = \quad \tfrac{1}{2} \sum_{i=1}^{n} d_i \quad = \quad \tfrac{1}{2}\Big(n + \sum_{i=1}^{n} (d_i - 1)\Big) \quad \leq \quad \tfrac{1}{2}(n + n^{3/2}),$$

which is what we needed to show. □

3 Cayley's formula for the number of spanning trees of a complete graph

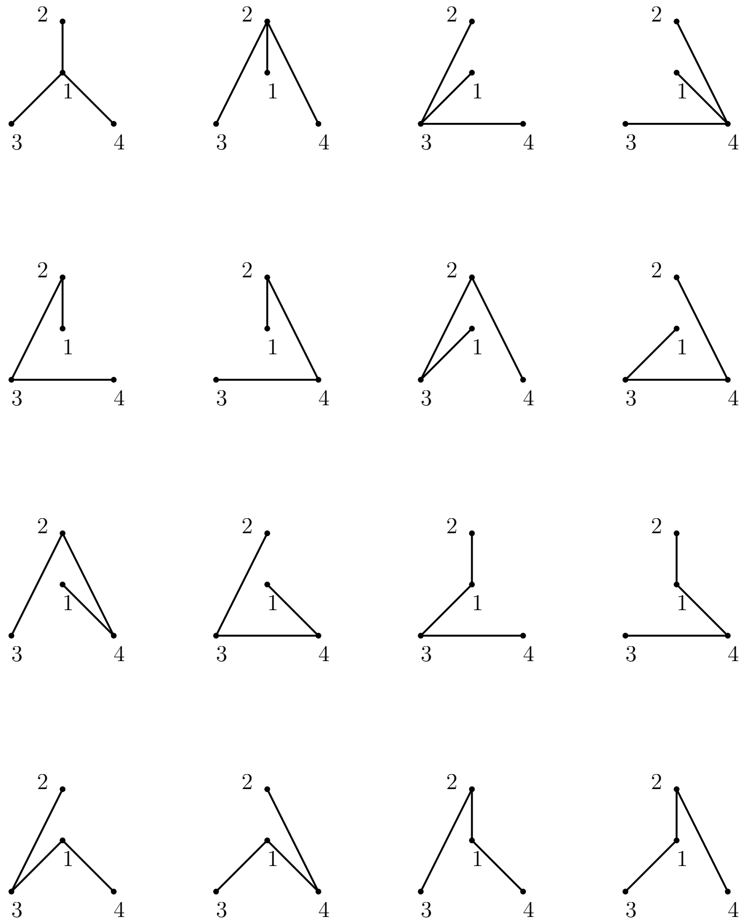
Recall that a *forest* is an acyclic graph (i.e. a graph that has no cycles), and a *tree* is a connected forest. A *leaf* in a graph G is a vertex of degree one, i.e. a vertex that has exactly one neighbor. In what follows, we will use the well-known fact that every tree on at least two vertices has a leaf.[6] It is clear that if v is a leaf of a tree T, then $T \setminus v$ is still a tree.

A *spanning tree* of a connected graph G is a tree T that is a subgraph of G, and satisfies $V(T) = V(G)$. An example is given below (the edges of the spanning tree are in red).



Now, suppose we are given a labeled complete graph on n ($n \geq 2$) vertices (say, with vertices labeled $1, \dots, n$). We would like to count the number of spanning trees in this graph; equivalently, we would like to count the number of trees on the vertex set $\{1, \dots, n\}$. There is only one spanning tree for K_2, and it is easy to see that there are three spanning trees for K_3. For K_4, there are 16 spanning trees, represented below (only the edges of the trees are represented; the remaining edges of the K_4 are not shown).

[6]In fact, every tree on at least two vertices has at least two leaves. Let us prove this. Suppose that T is a tree on at least two vertices, and let $P = p_1, \dots, p_t$ be a path of maximum length in T. Since T has at least one edge (because it is connected and has at least two vertices), we know that $t \geq 2$. We claim that p_1 and p_t are leaves of T; by symmetry, it suffices to show that p_1 is a leaf. Obviously, p_1 is adjacent to p_2 in T. Further, if p_1 were adjacent to some p_i with $i \in \{3, \dots, t\}$, then $p_1, p_2, \dots, p_i, p_1$ would be a cycle in T, contrary to the fact that T is a tree. Finally, if p_1 were adjacent to some vertex $v \in V(T) \setminus \{p_1, \dots, p_t\}$, then the path $v, p_1, \dots, p_t$ would contradict the maximality of P. So, p_2 is the only neighbor of p_1 in T, and it follows that p_1 is a leaf of T.



Our goal in this section is to prove the following.

Cayley's formula. *For all $n \geq 2$, the number of spanning trees of K_n is n^{n-2}.*

There are a number of known proofs of Cayley's formula; here, we give the one that uses the so called "Prüfer codes."

We will show that for all finite sets $S \subseteq \mathbb{N}$ with $|S| \geq 2$, the number of trees on the vertex set S is $|S|^{|S|-2}$ (see Lemma 3.4). Obviously, this will immediately imply Cayley's formula, since the number of spanning trees of K_n is equal to the number of trees on the vertex set $\{1, \dots, n\}$.

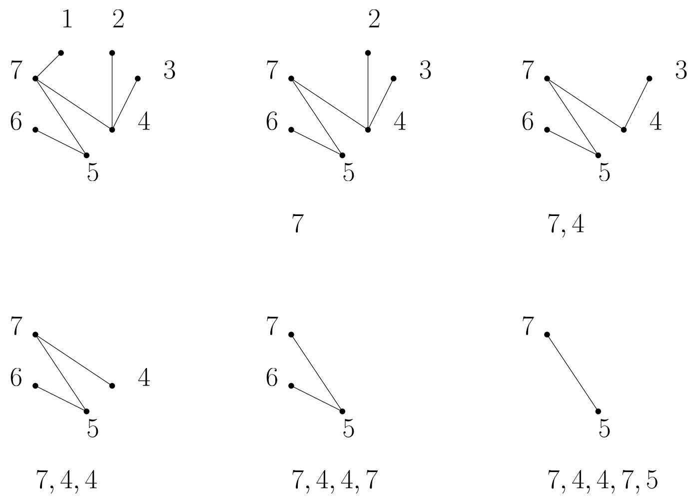
To simplify terminology, we will say that a tree is an *integer tree* if all its vertices are positive integers.[7] We now define the *Prüfer code* of integer trees on at least two vertices recursively, as follows:

- for any integer tree T on exactly two vertices, the Prüfer code of T, denoted by $P(T)$, is the empty sequence;

[7] Note, however, that this is **not** standard terminology. (There is no standard terminology for such trees.) We simply use the term "integer tree" as convenient shorthand in this section.

- for any integer tree T on at least three vertices, we define the Prüfer code of T to be $P(T) := a_i, P(T \setminus i)$, where i is the smallest leaf of T, and a_i is the unique neighbor of i in T.[8]

An example is given below (the Prüfer code of the tree in the top left corner is $7, 4, 4, 7, 5$, and the procedure for finding it is shown below).



Now, our goal is to show that given a set $S \subseteq \mathbb{N}$ (with $|S| = n \geq 2$), the function $T \mapsto P(T)$ is a bijection between the set of all trees with vertex set S, and the set of all sequences of length $n - 2$ with terms in S.[9]

Lemma 3.1. *If T is an integer tree on at least two vertices, then every non-leaf of T appears in $P(T)$, and none of the leaves do.*

Proof. We prove the lemma by induction on the number of vertices of the integer tree T. If T is a 2-vertex integer tree, then both its vertices are leaves, and by definition, $P(T)$ is the empty sequence; so the lemma is true for 2-vertex integer trees. Now, fix an integer $n \geq 2$, and assume inductively that the lemma holds for integer trees on n vertices. Let T be an integer tree on $n + 1$ vertices. Let i be the smallest leaf of T, and let a_i be the unique neighbor of i. Since T is connected and has at least three vertices, adjacent vertices cannot both be leaves of T, and so a_i is a non-leaf of T. By construction, $P(T) = a_i, P(T \setminus i)$, and so the non-leaf a_i of T appears in $P(T)$, whereas the leaf i of T does not. Note that for $v \in V(T) \setminus \{i, a_i\}$, we have that $d_T(v) = d_{T \setminus i}(v)$, and so each vertex of T other than i and a_i is a leaf in T if and only if it is a leaf in $P(T \setminus i)$. The result now follows from the induction hypothesis. □

Lemma 3.2. *If two integer trees have the same vertex set and the same Prüfer code, then they are identical.*

[8] So, $P(T)$ is obtained by adding a_i to the front of $P(T \setminus i)$.

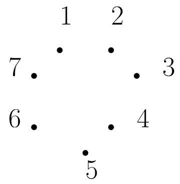
[9] Obviously, there are precisely n^{n-2} such sequences.

Proof. We proceed by induction on the number of vertices. There is only one tree on a fixed two-element vertex set, and so the lemma clearly holds for 2-vertex integer trees. Now, fix an integer $n \geq 2$, and suppose inductively that the lemma is true for integer trees on n vertices. Let $S \subseteq \mathbb{N}$ with $|S| = n+1$, and let T_1 and T_2 be trees with vertex-set S and identical Prüfer code P. P is of length $n-1$, and so at least two members of S do not appear in P; let i be the smallest integer in S that does not appear in P. Let a_i be the first term of P, and let P_i be obtained from P by deleting its first term. By Lemma 3.1, i is the smallest leaf of both T_1 and T_2, and a_i is the unique neighbor of i in both T_1 and T_2. Further, we have that $P(T_1 \setminus i) = P(T_2 \setminus i) = P_i$, and so by the induction hypothesis, $T_1 \setminus i = T_2 \setminus i$. Since i has the same neighborhood in T_1 and in T_2, it follows that $T_1 = T_2$. □

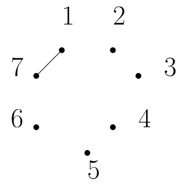
Lemma 3.3. *If $n \geq 2$ is an integer, and if $S \subseteq \mathbb{N}$ with $|S| = n$, then every sequence of length $n-2$, all of whose terms are in S, is the Prüfer code of some tree with vertex-set S.*

Proof. We proceed by induction on n. Suppose first that $S \subseteq \mathbb{N}$ satisfies $|S| = 2$, and let P be a sequence of length $2-2=0$, all of whose terms are in S. Then P is the empty sequence. Let T be the unique tree on the vertex-set S. Then $P(T) = P$. Now, fix an integer $n \geq 2$, and suppose inductively that the lemma is true for some $n \geq 2$. We need to show that it holds for $n+1$. Let $S \subseteq \mathbb{N}$ be such that $|S| = n+1$, and let P be a sequence of length $n-1$, all of whose terms are in S. Let i be the smallest member of S that does not appear in P, and let a_i be the first term of P. Let P_i be the sequence obtained by deleting the first term from P. By the induction hypothesis, there is a tree T_i with vertex-set $S \setminus \{i\}$ and Prüfer code P_i. Let T be the tree obtained by adding the vertex i to T_i, and making i adjacent to a_i and to no other vertex of T_i. Now $P(T) = P$. This completes the induction. □

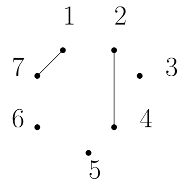
Note that the proof of Lemma 3.3 in fact gives us a recipe for "decoding" a given Prüfer code, i.e. for finding the tree to which the code corresponds. For an integer $n \geq 2$, an n-element set $S \subseteq \mathbb{N}$, and an $(n-2)$-term sequence P, with terms in S, we proceed as follows. If $n \geq 3$, then we let i be the smallest element of S that is not in P, and we let a_i be the first term of P. We make i and a_i adjacent, we delete i from S, and we delete the first term of P. We repeat the process until S only has two elements left, and P is the empty sequence. At this point, we make the last two remaining elements of S adjacent. An example is given below: the tree on the vertex set $S = \{1, 2, 3, 4, 5, 6, 7\}$ whose Prüfer code is $7, 4, 4, 7, 5$ is the tree on the bottom of the picture (e is the empty sequence).



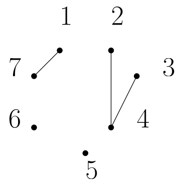
$P = 7, 4, 4, 7, 5$
$S = \{1, 2, 3, 4, 5, 6, 7\}$



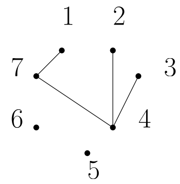
$P = 4, 4, 7, 5$
$S = \{2, 3, 4, 5, 6, 7\}$



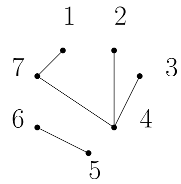
$P = 4, 7, 5$
$S = \{3, 4, 5, 6, 7\}$



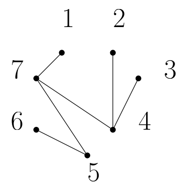
$P = 7, 5$
$S = \{4, 5, 6, 7\}$



$P = 5$
$S = \{5, 6, 7\}$



$P = e$
$S = \{5, 7\}$



Putting Lemmas 3.2 and 3.3 together, we obtain the following.

Lemma 3.4. *Let $n \geq 2$ be an integer, and let $S \subseteq \mathbb{N}$ be such that $|S| = n$. Then the number of trees on the vertex set S is n^{n-2}.*

Proof. By Lemmas 3.2 and 3.3, the mapping $T \mapsto P(T)$ is a bijection from the set of all integer trees on the vertex set S to the set of $(n-2)$-term sequences, all of whose terms are elements of S. There are precisely n^{n-2} sequences of length $n-2$, with terms in S, and it follows that there are precisely n^{n-2} trees on the vertex set S. $\square$

Cayley's formula follows immediately from Lemma 3.4, since the number of spanning trees of K_n is precisely the number of trees on the vertex set $\{1, \ldots, n\}$.

3.1 Cayley's formula via determinants

In this subsection, we give (without proof) a formula for computing the number of spanning trees of **any** graph on the vertex set $\{1, \ldots, n\}$.

Suppose that $n \geq 2$ is an integer, and that G is a graph on the vertex set

$\{1, \ldots, n\}$. Then the *Laplacian* of G is the matrix $Q = [q_{i,j}]_{n\times n}$ given by

$$q_{i,j} = \begin{cases} d_G(i) & \text{if } \; i = j \\ \\ -1 & \text{if } \; i \neq j \text{ and } ij \in E(G) \\ \\ 0 & \text{if } \; i \neq j \text{ and } ij \notin E(G) \end{cases}$$

We now need some notation. Suppose $A = [a_{i,j}]_{n\times m}$ is a matrix, and suppose $i \in \{1, \ldots, n\}$ and $j \in \{1, \ldots, m\}$; then $A_{i,j}$ is the matrix obtained from A by deleting the i-th row and j-th column. In particular, $A_{1,1}$ is the matrix obtained from A by deleting the first row and first column.

Theorem 3.5. *Let $n \geq 2$ be an integer, let G be any graph on the vertex set $\{1, \ldots, n\}$, and let Q be the Laplacian of G. Then the number of spanning trees of G is precisely* $\det(Q_{1,1})$.

Proof. Omitted. □

Example 3.6. *Using Theorem 3.5, prove Cayley's formula.*

Solution. Fix an integer $n \geq 2$, and consider the complete graph on the vertex set $\{1, \ldots, n\}$. Then the Laplacian of this graph is the $n \times n$ matrix

$$Q = \begin{bmatrix} n-1 & -1 & -1 & \ldots & -1 \\ -1 & n-1 & -1 & \ldots & -1 \\ -1 & -1 & n-1 & \ldots & -1 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ -1 & -1 & -1 & \ldots & n-1 \end{bmatrix}_{n\times n}.$$

The matrix $Q_{1,1}$ has exactly the same form, only it is of size $(n-1) \times (n-1)$:

$$Q_{1,1} = \begin{bmatrix} n-1 & -1 & -1 & \ldots & -1 \\ -1 & n-1 & -1 & \ldots & -1 \\ -1 & -1 & n-1 & \ldots & -1 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ -1 & -1 & -1 & \ldots & n-1 \end{bmatrix}_{(n-1)\times(n-1)}.$$

We now compute the determinant of $Q_{1,1}$:

$$\det(Q_{1,1}) = \begin{vmatrix} n-1 & -1 & -1 & \dots & -1 \\ -1 & n-1 & -1 & \dots & -1 \\ -1 & -1 & n-1 & \dots & -1 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ -1 & -1 & -1 & \dots & n-1 \end{vmatrix}_{(n-1)\times(n-1)}$$

$$\stackrel{(*)}{=} \begin{vmatrix} n-1 & -1 & -1 & \dots & -1 \\ -n & n & 0 & \dots & 0 \\ -n & 0 & n & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ -n & 0 & 0 & \dots & n \end{vmatrix}_{(n-1)\times(n-1)}$$

$$\stackrel{(**)}{=} \begin{vmatrix} 1 & -1 & -1 & \dots & -1 \\ 0 & n & 0 & \dots & 0 \\ 0 & 0 & n & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \dots & n \end{vmatrix}_{(n-1)\times(n-1)}$$

$$\stackrel{(***)}{=} n^{n-2},$$

where (*) is obtained by subtracting the first row from all the subsequent ones, (**) is obtained by adding to the first column the sum of all subsequent ones, and (***) is obtained by multiplying the diagonal entries of the upper triangular matrix that we obtained. By Theorem 3.5, we now have that the number of spanning trees of K_n is precisely n^{n-2}, which proves Cayley's formula. □