A SIMPLE PROOF OF THE CAYLEY FORMULA USING RANDOM GRAPHS

SCOTT WU, RAY LI, ANDREW HE, AND STEVEN HAO

Abstract. We present a nice result on the probability of a cycle occurring in a randomly generated graph. We then provide some extensions and applications, including the proof of the famous Cayley formula, which states that the number of labeled trees on \( n \) vertices is \( n^{n-2} \). 

1. Proof of Cayley’s Theorem

Lemma 1.1. Let \( G \) be a graph with \( n \) vertices, initially empty, and let \( p \in [0, 1] \). For each vertex \( v \in G \), independently perform the following process: Pick a vertex \( w \) uniformly at random from the vertices of \( G \). With probability \( p \), call \( v \) good and create a directed edge from \( v \) to \( w \). The probability that \( G \) contains a directed cycle of any kind, including self-loops, is \( p \).

Proof. We proceed by strong induction on \( n \). If \( n = 1 \), the result is trivial.

Consider a graph \( G \) on \( n \geq 2 \) vertices generated as stated above. Suppose that \( k \) of the \( n \) vertices are good. We then claim that the probability that a cycle exists in the graph is \( \frac{k}{n} \).

If \( k = n \), then the graph clearly contains a cycle, so the probability is \( 1 = \frac{n}{n} \). If \( k = 0 \), the graph clearly does not contain a cycle, so the probability is 0. Now suppose that \( 0 < k < n \). Note that any cycle in the graph must consist only of good vertices. Therefore, the probability of the existence of a cycle in \( G \) is equal to the probability of the existence of a cycle in the subgraph \( G' \) consisting of the \( k \) good vertices. Each vertex in \( G' \) has probability \( \frac{k}{n} \) of having an outedge leading to a vertex in \( G' \). By the inductive hypothesis, \( G' \), and therefore \( G \), has probability \( \frac{k}{n} \) of containing a cycle, as desired.

We have shown that if \( k \) of the \( n \) vertices are good, the probability of a cycle existing is \( \frac{k}{n} \). Thus, the expected probability of a cycle is \( E \left[ \frac{k}{n} \right] \), which is \( \frac{E[k]}{n} = \frac{E[n]}{n} = p \) by linearity of expectation. Thus, the probability of creating a cycle is \( p \) for \( n \) vertices, and the induction is complete. \( \square \)

Theorem 1.2. (Cayley’s Formula) For any positive integer \( n \), the number of trees on \( n \) labeled vertices is exactly \( n^{n-2} \).

Proof. Assume \( n \geq 2 \). If \( n = 1 \) the proof is trivial.

Apply Lemma 1.1 on a graph of \( n - 1 \) vertices with \( p = \frac{n-1}{n} \). Each vertex is equally likely to have a directed edge to a specified vertex or to have no outedge at all, so there are \( n^{n-1} \) possible directed graphs, each equally probable. Call one such graph special if it does not contain any cycles. Note that special graphs must be forests of rooted trees with edges pointing towards the root in each tree; furthermore, every such forest of rooted trees is a special graph.
We construct a mapping from special graphs to trees on \( n \) vertices rooted at the \( n \)th vertex. For each special graph, construct an additional, \( n \)th vertex in the graph. For each of the original \( n-1 \) vertices that does not have an outedge, create a directed edge leading to the \( n \)th vertex. This gives a tree with all edges directed at the \( n \)th vertex (i.e. a tree rooted at the \( n \)th vertex).

Note that this mapping is injective, as the vertices with edges directed toward the \( n \)th vertex are precisely those without outedges in the original graph, while all edges are preserved. On the other hand, the mapping is surjective over trees rooted at the \( n \)th vertex, because we can just direct all edges toward the root and remove the root and its edges. This must form a forest of rooted trees.

Thus, a bijection exists between such directed graphs and rooted undirected trees on \( n \) vertices. Furthermore, the number of trees rooted at a particular vertex (the \( n \)th vertex in this case) on \( n \) labeled vertices is simply the number of trees on \( n \) vertices, we need only to count the number of directed graphs defined above.

Since there are \( n^{n-1} \) possible graphs, and exactly \( 1 - p = \frac{1}{n} \) of the graphs correspond to labeled trees with \( n \) vertices, there are a total of \( n^{n-2} \) trees on \( n \) vertices, as desired.

\[ \square \]

2. Generalizations

In addition to the above result, we found two nice generalizations of the lemma.

**Corollary 2.1.** Consider a variant of the graph generation process in Lemma 1.1. For each vertex \( v \), independently perform the following process: Pick a vertex \( w \) uniformly at random from the vertices of \( G \). With probability \( p_v \), call \( v \) good and draw a directed edge between \( v \) and \( w \). The probability of a cycle in the graph is equal to the average of the probabilities of each vertex being good, or \( \frac{\sum p_v}{n} \).

**Proof.** As shown in the proof of the lemma, the probability of a cycle existing is equal to \( \frac{1}{n} \) of the expected number of good vertices in the graph, or \( \frac{\sum p_v}{n} \). \[ \square \]

**Corollary 2.2.** Consider another variant of the graph generation in Lemma 1.1. Let \( \pi \) be a probability distribution over the vertices. For each vertex \( v \), independently perform the following process: Pick a vertex \( w \) from the vertices of \( G \) according to \( \pi \). With probability \( p \), call that vertex good and draw a directed edge between \( v \) and \( w \). The probability of a cycle in the graph is equal to \( p \).

**Proof.** Proceed with strong induction as in the lemma. We wish to show that the probability of a cycle in the graph is equal to \( \frac{1}{n} \) of the expected number of good vertices. The \( k = 0 \) and \( k = n \) cases remain the same. For all other values of \( k \), note that the probability of each node in \( G' \) having its outedge lead to a node in \( G' \) is \( \sum_{v \in G'} \pi(v) \). However, since each node has equal probability of being good, the expected value of this probability, and therefore the expected value of a cycle in \( G' \) by the inductive hypothesis, is \( k(\frac{1}{n})(\sum_{v \in G} \pi(v)) = \frac{k}{n} \), as desired. \[ \square \]

3. Acknowledgements

We would like to thank Po Shen Loh, Yan Zhang, and others for their guidance.
References

[1] A. Cayley. A theorem on trees. Quart. J. Math, 23:376–378, 1889.