Elaboration of "Enumerating Graphs and Brownian Motion"

Risan

July 31, 2024

This report is an elaboration of the paper [1], which gave a connection between the graph enumeration problem and Brownian motion.

For $n \geq 1$ and $k \geq 0$, we denote the number of connected graph $G = (V, E)$ such that $V = \{1, \dots, n\}$ vertices and $|E| = n-1+k$ as $c(n, k)$. Note that, it is not up to isomorphism. Two graphs are considered the same if and only if the have the same set of vertices and set of edges. For example, the line graph with $E = \{\{1, 2\}, \{2, 3\}\}$ and $E = \{\{1, 3\}, \{3, 2\}\}$ are counted as different graphs, and $c(3, 0) = 3$.

From here, we will only consider the graphs where the vertices are $\{1, \dots, n\}$ for some $n > 0$.

1 Breadth-first search (BFS)

Let $G = (V, E)$ be a simple undirected graph where V denotes the finite set of vertices and E denotes the finite set of edges. We will consider the following graph traversal algorithm to traverse G starting from vertex 1.

Algorithm 1 Breadth-first search (BFS)

```
 1: Q ← empty queue
 2: push(Q, 1)
 3: while Q not empty do
 4:     v ← front(Q)
 5:     pop(Q)
 6:     for w adjacent to v in the order from smallest to largest do
 7:         if w has never been pushed to Q then
 8:             push(Q, w)
 9:         end if
10:     end for
11: end while
```

Note that the order on line 6 is very important. This detail is very easy to miss because it is not explicitly mentioned in the original paper that a certain order needs to be fixed. We cannot compute Proposition 1 without this. Moreover, not every kind of ordering leads to Proposition 1. For example, if $f(w, v)$ denotes the size of the component in which w belongs to after the edge $\{w, v\}$ is removed from the graph, and the ordering of line 6 is from smallest to largest according to f (with vertex label as tie breaker), then $G = (\{1, 2, 3, 4\}, \{\{1, 2\}, \{1, 3\}, \{2, 4\}\})$ and $G = (\{1, 2, 3, 4\}, \{\{1, 2\}, \{1, 3\}, \{3, 4\}\})$ will be traversed by BFS with the order of $1, 3, 2, 4$ and $1, 2, 3, 4$ respectively, which is not what Proposition 1 wants.

For example, let us consider a tree $G = (V, E)$ where $V = \{1, \dots, 7\}$ and $E = \{\{1, 3\}, \{1, 5\}, \{1, 6\}, \{2, 7\}, \{3, 4\}, \{3, 7\}\}$. Then, if we combine line 4 and 5 in the BFS algorithm into single pop(Q,v), then the order of push and pop operations in the above algorithm will be: push(Q,1), pop(Q,1), push(Q, 3), push(Q, 5), push(Q, 6), pop(Q,3), push(Q, 4), push(Q, 7), pop(Q,5), pop(Q,6), pop(Q,4), pop(Q,7), push(Q, 2).

In an execution of BFS for a graph G, we say that a vertex v finds a set of vertices S if and only if the set S are the vertices in the consecutive push calls after v is popped from Q. For example, in the above example, vertex 1 finds $\{3, 5, 6\}$, vertex 3 finds $\{4, 7\}$, vertex 5, 6, and 4 find $\emptyset$, and vertex 7 finds $\{2\}$.

We also denote the group of operations from line 4 to 10 in the BFS algorithm as processing v. This includes popping v from the queue and consecutive pushing of all vertices adjacent to v that have not been in the queue before.

For a graph $G = (V, E)$, let $\mathcal{X}(G) = (x_t)_{1 \leq t \leq |V|}$ and $\mathcal{Q}(G) = (q_t)_{0 \leq t \leq |V|}$. For $1 \leq t \leq |V|$, let x_t be the size of the set of vertices found by the t-th popped vertex (not the vertex t) and let q_t be the size of the queue Q in the BFS algorithm right after processing t-th popped vertex. We also set $q_0 = 1$. Clearly, $q_t = q_{t-1} + x_t - 1$ when $t > 0$.

In the above example, we have $(x_t)_{1\le t\le 7} = (3,2,0,0,0,1,0)$ and $(q_t)_{0\le t\le 7} = (1,3,4,3,2,1,1,0)$. Moreover, we have the following necessary and sufficient conditions which is due to line 3 of the BFS algorithm above

$$q_{|V|} = 0 \text{ and } q_i > 0 \text{ for all } 0 < i < |V|.$$

For a graph G, we denote $\pi(G) = (V, E')$ to be the tree where E' consists of the edges that we get by connecting a vertex with all the vertices that are found by it during BFS.

2 Trees

The following proposition gives the number of trees that shares the same $\mathcal{X}$ with a given tree T.

Proposition 1. *Let T be a tree with n vertices and let $(x_t)_{1\le t\le n} = \mathcal{X}(T)$. There are exactly*

$$\frac{(n-1)!}{\prod_{1\le i\le n} x_i!}$$

trees such that each tree t satisfies $\mathcal{X}(T) = \mathcal{X}(t)$.

Proof. Let $\mathcal{T}$ be the set of such trees and let $\mathcal{P}$ be the set of tuples $(p_t)_{1\le t\le n}$ of disjoint tuples $\mathcal{P} \subset \{1,\dots,n\}$ such that $p_1 \cup \dots \cup p_n = \{2,\dots,n\}$ and $p_{t,j} < p_{t,j+1}$ for any $1 \le j < |p_t|$.

Consider the map $m : \mathcal{T} \to \mathcal{P}$ such that a tree in $\mathcal{T}$ is mapped to the tuple p such that p_t is the tuple that corresponds to the set of vertices found by the t-th popped vertex during an execution of BFS. Moreover, given $p \in \mathcal{P}$, let L be the result of concatenating p in the order of its index. Then $m^{-1}(p)$ is given by a tree that is the result of relabeling the original tree T such that the i-th vertex popped by the BFS is labelled with L_i. Hence, the map m is a bijection. Moreover, the size of $\mathcal{P}$ is clearly

$$\prod_{1\le i\le n} \binom{n-1-(x_1+\dots+x_{i-1})}{x_i} = \prod_{1\le i\le n} \frac{(n-1-(x_1+\dots+x_{i-1}))!}{(n-1-(x_1+\dots+x_i))!x_i!}$$

which gives a telescopic product that results to the proposition. □

3 Random trees

In this section, we will find random variables that corresponds to random trees.

Proposition 2. *Let $X_1,\dots,X_n$ be i.i.d. random variables where each has Poisson distribution (`https://en.wikipedia.org/wiki/Poisson_distribution`) with mean 1. In other words, $\mathbb{P}(X_i = x) = \frac{e^{-1}}{x!}$. We also define $Q_0 = 1$ and $Q_t = Q_{t-1} + X_t - 1$ for $1 \le t \le n$. For $(x_t)_{1\le t\le n}$ be such that $x_t \ge 0$,*

$$\mathbb{P}(X_1 = x_1,\dots,X_n = x_n \mid Q_0 = 0, Q_1,\dots,Q_n > 0)$$

is the probability of choosing a tree t randomly with uniform distribution from the set of all trees with n vertices such that $\mathcal{X}(t) = (x_t)_{1\le t\le n}$.

Proof. Suppose that $(x_t)_{1\le t\le n} = \mathcal{X}(T)$ for some tree T. Then,

$$\begin{aligned}\mathbb{P}(X_1 = x_1,\dots,X_n = x_n \mid Q_n = 0, Q_1,\dots,Q_{n-1} > 0) &= \frac{\mathbb{P}(X_1 = x_1,\dots,X_n = x_n, Q_n = 0, Q_1,\dots,Q_{n-1} > 0)}{\mathbb{P}(Q_n = 0, Q_1,\dots,Q_{n-1} > 0)} \\ &= \frac{\mathbb{P}(X_1 = x_1,\dots,X_n = x_n)}{\mathbb{P}(Q_n = 0, Q_1,\dots,Q_{n-1} > 0)}\end{aligned}$$

because of the property of $\mathcal{Q}(T)$. Therefore,

$$\mathbb{P}(X_1 = x_1,\dots,X_n = x_n \mid Q_n = 0, Q_1,\dots,Q_{n-1} > 0) = \frac{e^{-n}}{\mathbb{P}(Q_n = 0, Q_1,\dots,Q_{n-1} > 0)\prod_{1\le i\le n} x_i!}$$

which is proportional to the probability of choosing a tree t randomly with uniform distribution from the set of all trees with n vertices such that $\mathcal{X}(t) = \mathcal{X}(T)$ as shown in Proposition 1.

Moreover, for any $(x_t)_{1\leq t\leq n}$ with no tree T such that $\mathcal{X}(T) = (x_t)_{1\leq t\leq n}$, the necessary and sufficient condition for $\mathcal{Q}(T)$ given in the first section implies that

$$\mathbb{P}(X_1 = x_1, \dots, X_n = x_n \mid Q_n = 0, Q_1, \dots, Q_n > 0) = 0.$$

This and above means the support (`https://en.wikipedia.org/wiki/Support_(mathematics)`) of X_t is that for trees, as stated in the original paper.

Therefore, the fact that $\mathbb{P}(X_1 = x_1, \dots, X_n = x_n \mid Q_n = 0, Q_1, \dots, Q_n > 0)$ is proportional to the probability of choosing the corresponding random tree as described above implies that their probabilities are actually exactly the same. □

Corollary 1. *By construction, it follows that $\mathbb{P}(Q_0 = q_0, \dots, Q_n = q_n \mid Q_n = 0, Q_1, \dots, Q_{n-1} > 0)$ gives the probability of choosing a tree t uniformly at random from the set of all trees with n vertices such that $\mathcal{Q}(t) = (q_t)_{0\leq t\leq n}$.*

4 General graphs

First, we will consider the following proposition.

Proposition 3. *Let $G = (V, E)$ be a graph. A pair of vertices $\{u, v\}$ satisfies $\pi\big((V, E \cup \{u, v\})\big) = T$ if and only if one is in the queue right after the first of the two vertices being processed during BFS.*

Proof. Let $G' = (V, E \cup \{u, v\})$. Suppose that u is processed before v during BFS of G. Suppose that v is found by a. If $a = u$, then $\{u, v\} \in E$ and v is in the queue when u is processed, so $\pi(G) = \pi(G')$. Suppose that a is processed before u. Then, v is in the queue before u is processed, and hence v is still in the queue when u is processed. That also means that in the BFS of G', v will still be found by a first and hence $\pi(G') = \pi(G)$. Suppose that a is processed after u. In this case, v is not in the queue when u is processed, and vice-versa. But, v will be found by u in $\pi(G')$. Therefore, $\pi(G) \neq \pi(G')$. □

Corollary 2. *Let $G = (V, E)$ be a graph and K be a set of pair of vertices. Then, $\pi(G') = \pi(G)$ where $G' = (V, E \cup K)$ if and only if K contains pair of vertices where one is in the queue right after the first of the two vertices being processed during BFS.*

Proof. It is clear by repeatedly applying Proposition 3. □

Let $M = \sum_{t=1}^{n-1}(Q_t - 1)$. From the previous section and Corollary 2, M corresponds to the number of edges that can possibly be added to a tree T such that $\pi(T) = T$. Therefore,

$$\mathbb{E}\left[M\right] = \frac{c(n,1)}{c(n,0)}$$

and

$$\mathbb{E}\left[\binom{M}{k}\right] = \frac{c(n,k)}{c(n,0)}$$

where $\binom{M}{k}$ corresponds to the number of choice of k edges from M possible edges to be added to a tree T such that $\pi(T) = T$.

5 Brownian motion

Because $X_1, \dots, X_n$ are i.i.d., it is clear that $X_1 - 1, \dots, X_n - 1$ are also i.i.d. Moreover, $X_1 - 1, \dots, X_n - 1$ have mean 0 and variance 1. Now, we define $f(\frac{t}{n}) = \frac{Q_t}{\sqrt{n}}$ for $0 \leq t \leq n$ and interpolate linearly between the values to give a curve $f(s)$ on $[0, 1]$. By Donsker's theorem, as $n \to \infty$, this curve approaches Brownian motion. The condition that this curve starts and ends with zero and everywhere else nonnegative means that it is a Brownian excursion. I saw many resources on this part, but one of the most direct ones is given in `https://en.wikipedia.org/wiki/Donsker%27s_theorem`.

By definition, we have

$$Mn^{-3/2} = \sum_{t=1}^{n-1} \frac{1}{n}\left(\frac{Q_t}{\sqrt{n}} - \frac{1}{\sqrt{n}}\right).$$

When $n \to \infty$, this summation approaches

$$L = \int_0^1 f(t)dt.$$

Therefore, as $n \to \infty$,

$$Mn^{-3/2} \to L$$

in distribution and

$$\mathbb{E}[M] \sim n^{3/2}\mathbb{E}[L].$$

With what seems to be a well-known results that

$$\binom{n}{k} \sim \frac{n^k}{k!},$$

we conclude that

$$\mathbb{E}\left[\binom{M}{k}\right] \sim \frac{n^{\frac{3k}{2}}}{k!}\mathbb{E}[L^k]$$

which gives an asymptotic formula for $\frac{c(n,k)}{c(n,0)}$, with $c(n,0) = n^{n-2}$ is given by the Cayley's theorem. The computations of above expectations are well studied, for example in [2].

References

[1] Spencer, John, *Enumerating Graphs and Brownian Motion*, Communications on Pure and Applied Mathematics 50, 1998, pp.291-294.

[2] Louchard, G., *The Brownian excursion area: a numerical analysis*, Computers & Mathematics with Applications 10, 1984, pp.413-417.