Probability Wrapup; Intro to Graph Theory

Robert Y. Lewis

CS 0220 2023

April 19, 2023

Overview

Coupon collecting

Sometimes a company will run a coupon-collecting promotion. For example, each time you make a purchase, you get a coupon at random. If you collect all n coupons, you win a large prize.

How many purchases should you expect to make to win (collect all n coupons) if coupon types are given uniformly at random at each purchase?

Solving it

Consider a sequence of coupons that has all n types at the end. We want to know the *length* of this sequence.

Let X_i be the sequence of coupons from right *after* the $(i-1)$th unique coupon was received until the ith unique coupon is received.

Examples:

$$\underbrace{3}_{X_1}\ \underbrace{4}_{X_2}\ \underbrace{5}_{X_3}\ \underbrace{3\ \ 3\ \ 2}_{X_4}\ \ \underbrace{2\ \ 5\ \ 1}_{X_5}$$

The length of the entire coupon sequence is the sum of the lengths of the X_is: $1+1+1+3+3=9$.

The *expected* length of the entire sequence is the sum of the expected lengths of the X_is.

What's Xk?

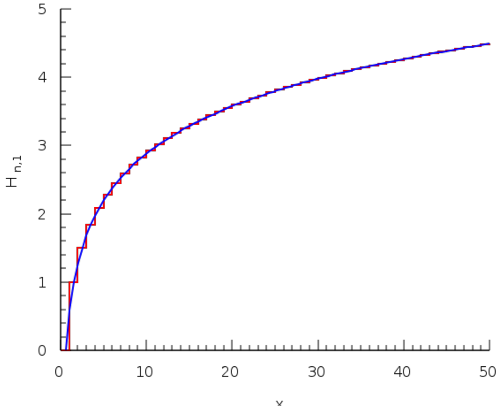
What's the length of X_k? At the beginning of segment k, we have collected $k-1$ different coupons from the n possibilities. Thus, the probability of getting one of the new ones is $p = \frac{n-k+1}{n}$.

What's the expected number of tries until a new one is found? $\frac{n}{n-k+1}$ by the "mean time to failure" analysis!

$\mathbb{E}[\sum_{k=1}^{n} \text{length}(X_k)]$	coup col
$= \sum_{k=1}^{n} \mathbb{E}[\text{length}(X_k)]$	lin of exp
$= \sum_{k=1}^{n} n/(n-k+1)$	mean fail time
$= n \sum_{i=1}^{n} 1/i$	$i = n-k+1$
$= nH(n)$	defn Harmonic number
$\approx n \ln(n)$	prop of H

Aside: harmonic numbers

$$H(n) = \sum_{i=1}^{n} \frac{1}{i} \approx \ln(n)$$



Examples

- People to poll before having representatives of each birthday: $365 \times H(365) \approx 2365$
- Ice skates before you have one of each: $2 \times H(2) = 3$
- Die rolls until you have one of each: $6 \times H(6) = 14.7$
- Number of randomly assigned pigeons until each hole is filled: $n \times H(n) \approx n \ln(n)$

Three-sided coin

If we want a uniform distribution over two options, we can flip a coin (H/T). If we want a uniform distribution over four options, we can flip two coins (HH/HT/TH/TT). What if we want a uniform distribution over three options?

We could flip two coins and say it's option 1 if both heads, option 2 if both tails, and option 3 if mismatch. Problem? Yes. The probability of mismatch is $\Pr[HT] + \Pr[TH] = 1/2$. Not $1/3$–$1/3$–$1/3$.

HH	1/4
HT	1/4
TH	1/4
TT	1/4

Repeat the trial

We could flip two coins and say it's option 1 if HH, option 2 if TT, option 3 if HT, and *do over* if TH. Problem? Maybe. There is an infinite number of outcomes...

Procedure selects option 1 if:

- HH on the first trial
- or TH on the first trial and HH on the second trial
- or TH on the first two trials and HH on the third trial
- ...

Pr(option 1)
$= 1/4 + 1/4 \times 1/4 + 1/4 \times 1/4 \times 1/4 + \ldots$
$= \sum_{i=1}^{\infty} 1/4^i$
$= 1/4 \times \sum_{i=0}^{\infty} 1/4^i$
$= 1/4 \left(\frac{1}{1-1/4}\right)$
$= 1/4 \times 4/3 = 1/3.$

Aside: Geometric sum

$$
\begin{aligned}
x &= \sum_{i=0}^{\infty} p^i && \text{the sum we want} \\
x &= p^0 + p^1 + p^2 + p^3 + \dots && \text{expand} \\
px &= p^1 + p^2 + p^3 + p^4 + \dots && \text{multiply by } p \\
p^0 + px &= p^0 + p^1 + p^2 + p^3 + p^4 + \dots && \text{add } p^0 \\
p^0 + px &= x && \text{defn of } x \\
p^0 &= x - px && \text{subtract } px \\
1 &= x(1-p) && \text{factor/simplify} \\
\frac{1}{1-p} &= x && \text{divide by } 1-p
\end{aligned}
$$

Infinite sample space

$\mathcal{S} = \{HH, HT, TT, TH : HH, TH : HT, TH : TT, TH : TH : HH, TH : TH : HT, TH : TH : TT, \ldots\}$
$= \{(TH)^n : HH, (TH)^n : HT, (TH)^n : TT | n \in \mathbb{N}\}$

The probability space is:

$\Pr((TH)^n : HH) = \Pr((TH)^n : HT) = \Pr((TH)^n : TT) = 1/4^{n+1}$.

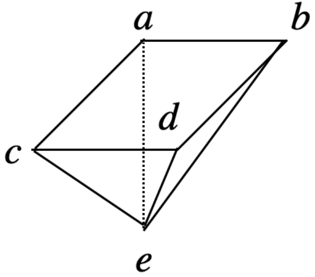
Note: $\sum_{n=0}^{\infty} 3 \times 1/4^{n+1} = 3/4 \times \frac{1}{1-1/4} = 3/4 \times 4/3 = 1$.

Non-negative and sums to one, valid probability space!

Graphs and polyhedra

A *graph* consists of a set of vertices (nodes) and edges (links) that connect them.

This terminology comes from descriptions of flat-sided solid shapes—polyhedra.



$\{a, b, c, d, e\}$ are the vertices, $\{\{a, b\}, \{a, c\}, \{c, d\}, \{b, d\}, \{a, e\}, \{b, e\}, \{c, e\}, \{d, e\}\}$ are the edges.

Examples of graphs

A graph can be viewed as a symmetric relation. Examples of edges:

- Two people that have been in direct contact.
- Two intersections connected by a road.
- Two accounts that are Facebook friends.
- Two gates connected by a wire.
- Two translators that speak a shared language.
- Two atoms bonded together.

(There are also directed graphs where the pairs are ordered. But we’re not talking about them.)

Graph definition

Definition: A *graph* G consists of a set of vertices $V(G)$ and edges $E(G)$.

$E(G) \subseteq \{\{u, v\} \mid u, v \in V(G), u \neq v\}$.

u adjacent to v in G iff $\{u, v\} \in E(G)$.

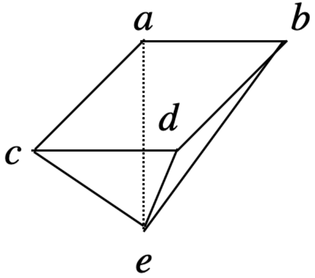
The “adjacent to” relation is reflexive? No. Symmetric? Yes! Transitive? Depends. Empty is graph is! Otherwise, no, because symmetry and non-reflexive.

Degree

Definition: The *degree* of vertex $u \in V(G)$ is

$$\deg(u) = |\{v \mid \{u, v\} \in E(G)\}| .$$

In words? The number of vertices adjacent to u. The number of edges that include u.

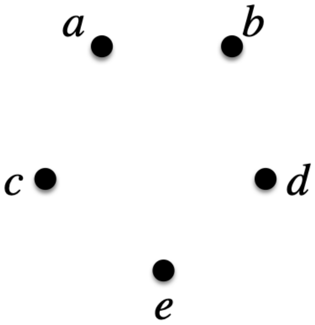


All of the vertices of this graph have the same degree. Yes or no?

Empty graph

Often convenient to let $n = |V(G)|$ and $m = |E(G)|$.

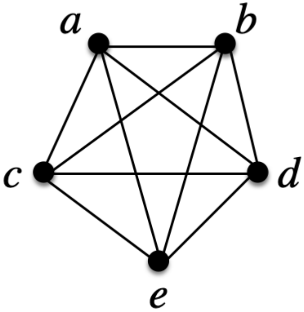
We insist that $n > 0$. But, if $m = 0$, we call it the *empty graph* on n nodes.



Complete graph

If, $\forall u, v \in V(G), u \neq v \rightarrow \{u, v\} \in E(G)$, we say G is a *complete graph* (or a clique). How do we read that in English? There is an edge between every two distinct vertices. Also known as K_n.

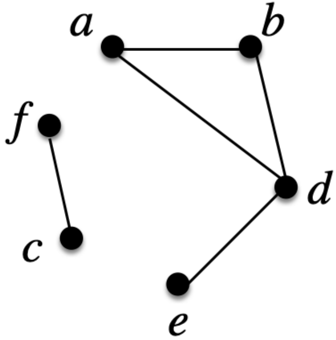
$m = \binom{n}{2} = n(n-1)/2$.



Path

A length-k path p in G is

- an element of $V(G)^{k+1}$
- such that $\forall i \in [1, k], \{p^i, p^{i+1}\} \in E(G)$.
- The path is *simple* if there is no $i \neq i'$ such that $p^i = p^{i'}$.

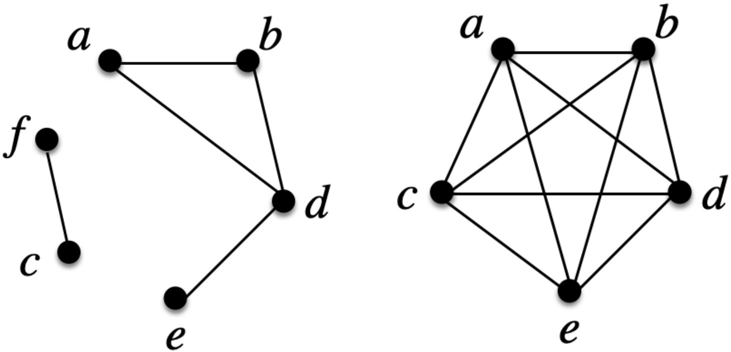


Paths? (a, b, d, a)? Yes (not simple). (b, a, e)? No, missing edge.

Connected graph

A pair of vertices u and v is *connected* if there exists a length-k path such that $p^1 = u$ and $p^{k+1} = v$.

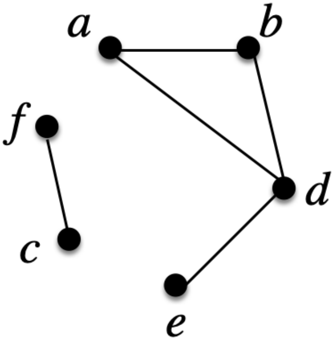
A graph G is connected if all pairs of vertices in $V(G)$ are connected: for all $u, v \in V(G)$, there is a (u, v)-path.



Cycle

A length-k cycle p in G is

- a length-k path in G
- where $p^{k+1} = p^1$.



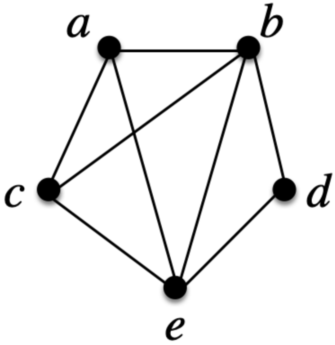
Cycles? (a, b, d, a)? Yes. (b, a, e, b)? No, missing edge. (e, d, b, d, e)? Yes. (f, c, f)? Yes.

Edge and vertex sets of paths

For a length-k path p, let $E(p) = \{\{p^i, p^{i+1}\} \mid i \in [1, k]\}$.

Let $V(p) = \{p^i \mid i \in [1, k+1]\}$. These are p's edge and vertex sets.

Find two cycles p and q such that $V(p) = V(q)$ but $E(p) \neq E(q)$.

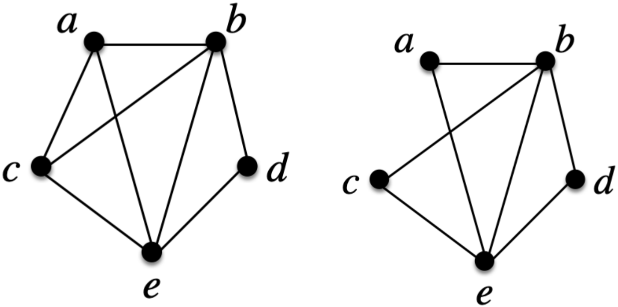


Tours on Graphs

Let G be a graph with n vertices and m edges. Let p be a length-k cycle in G.

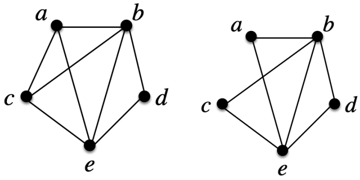
Definition: p is an *Eulerian tour* (edge tour) of G if $E(p) = E(G)$ and $k = m$.

Definition: p is a *Hamiltonian tour* (vertex tour) of G if $V(p) = V(G)$ and $k = n$.



Eulerian implies even

If a graph has an Eulerian tour, deg(v) must be even for all vertices.

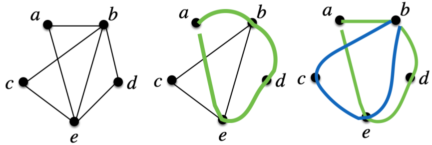


Proof sketch: Let p be an Eulerian tour of G. Every time the tour visits a vertex, it must arrive and depart (on fresh edges). Thus, each appearence of v on the cycle accounts for two edges that include v. In total, an even number.

Because it's an Eulerian cycle, every edge of the graph appears exactly once. Thus, every vertex of the graph is adjacent to an even number of other vertices.

Even implies Eulerian

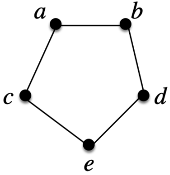
If all vertices in a connected graph have even degree, it has an Eulerian path (!).



Pick a vertex. Pick an arbitrary unused edge (if there is one) to go to another vertex. This process can only get stuck when we've returned to the starting vertex. That's the only vertex with an odd number of unused edges. If not an Eulerian tour, there's some vertex that we visited that has unused edges. That's because the graph is connected. Pick such a vertex and start this process again, rerouting our original route through this side route. By the earlier argument, the side route has to return to its starting point. Process must complete and can only complete with an Eulerian tour.

Cycle graph

If an n-node graph G has a cycle p that is both an Eulerian tour *and* a Hamiltonian tour, we call the graph a cycle and write it C_n.



Equivalent definition: C_n is connected and $\forall v \in V(G), \deg(v) = 2$.

Number of edges of C_n is n.

For what values of n, if any, does $K_n = C_n$?

Puzzles

- When does C_n have an Eulerian/Hamiltonian tour?
- When does K_n have an Eulerian/Hamiltonian tour?
- How many simple paths in K_n of length $n-1$?
- How many simple paths in C_n of length $n-1$?