15-251

Some

~~Great~~ Theoretical Ideas

~~in~~ Computer Science

for

This Course…

Probability Refresher

What's a Random Variable?

A Random Variable is a real-valued function on a sample space S

$E[X+Y] = E[X] + E[Y]$

Probability Refresher

What does this mean: $E[X \mid A]$?

Is this true:

$$\Pr[\, A \,] = \Pr[\, A \mid B \,] \Pr[\, B \,] + \Pr[\, A \mid \overline{B} \,] \Pr[\, \overline{B} \,]$$

Yes!

Similarly:

$$E[\, X \,] = E[\, X \mid A \,] \Pr[\, A \,] + E[\, X \mid \overline{A} \,] \Pr[\, \overline{A} \,]$$

Air Marshal Problem

Every passenger has an assigned seat

There are n-1 passengers and n seats

Before the passengers board, an air marshal sits on a random seat

When a passenger enters the plane, if their assigned seat is taken, they pick a seat at random

What is the probability that the last passenger to enter the plane sits in their assigned seat?

Several years ago Berkeley faced a law suit …

1. % of male applicants admitted to graduate school was 10%
2. % of female applicants admitted to graduate school was 5%

Grounds for discrimination?

SUIT

Berkeley did a survey of its departments to find out which ones were at fault

Every department was more likely to admit a female than a male

$$\frac{\text{\#of females accepted to department X}}{\text{\#of female applicants to department X}} > \frac{\text{\#of males accepted to department X}}{\text{\#of male applicants to department X}}$$

How can this be ?

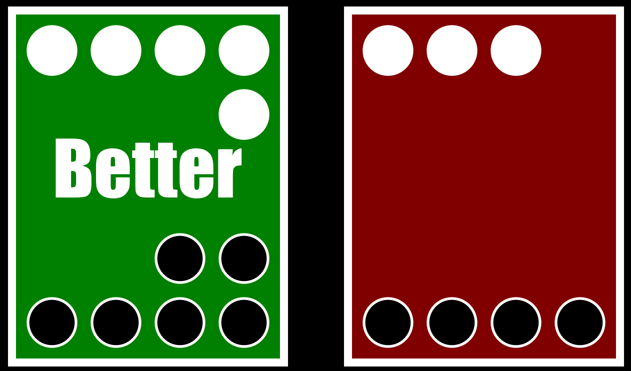
Answer

Women tended to apply to departments that admit a smaller percentage of their applicants

	Women		Men	
Dept	Applied	Accepted	Applied	Accepted
A	99	4	1	0
B	1	1	99	10
total	100	5	100	10

A single summary statistic
(such as an **average**, or a **median**)
may not summarize the data well !

Try to get a white ball

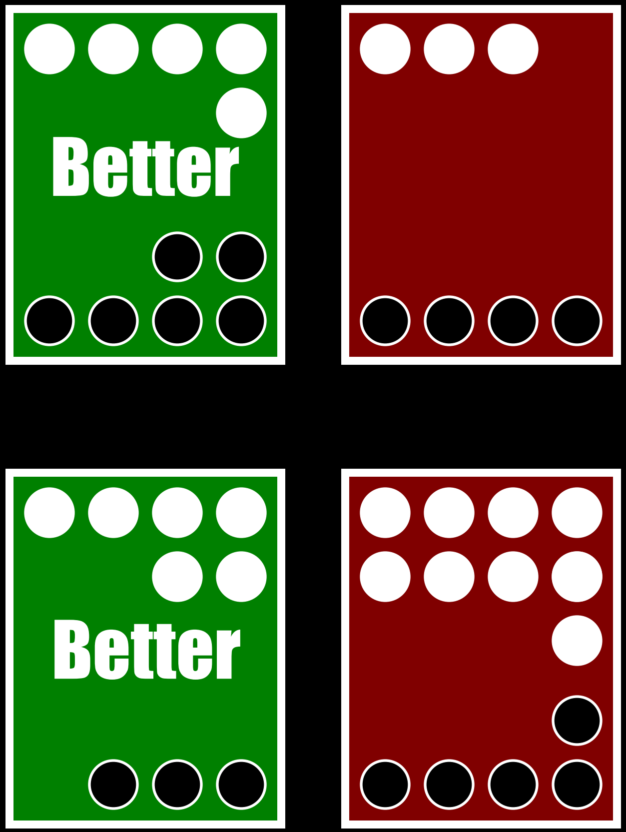


Choose one box and pick a random ball from it.

Max the chance of getting a white ball…

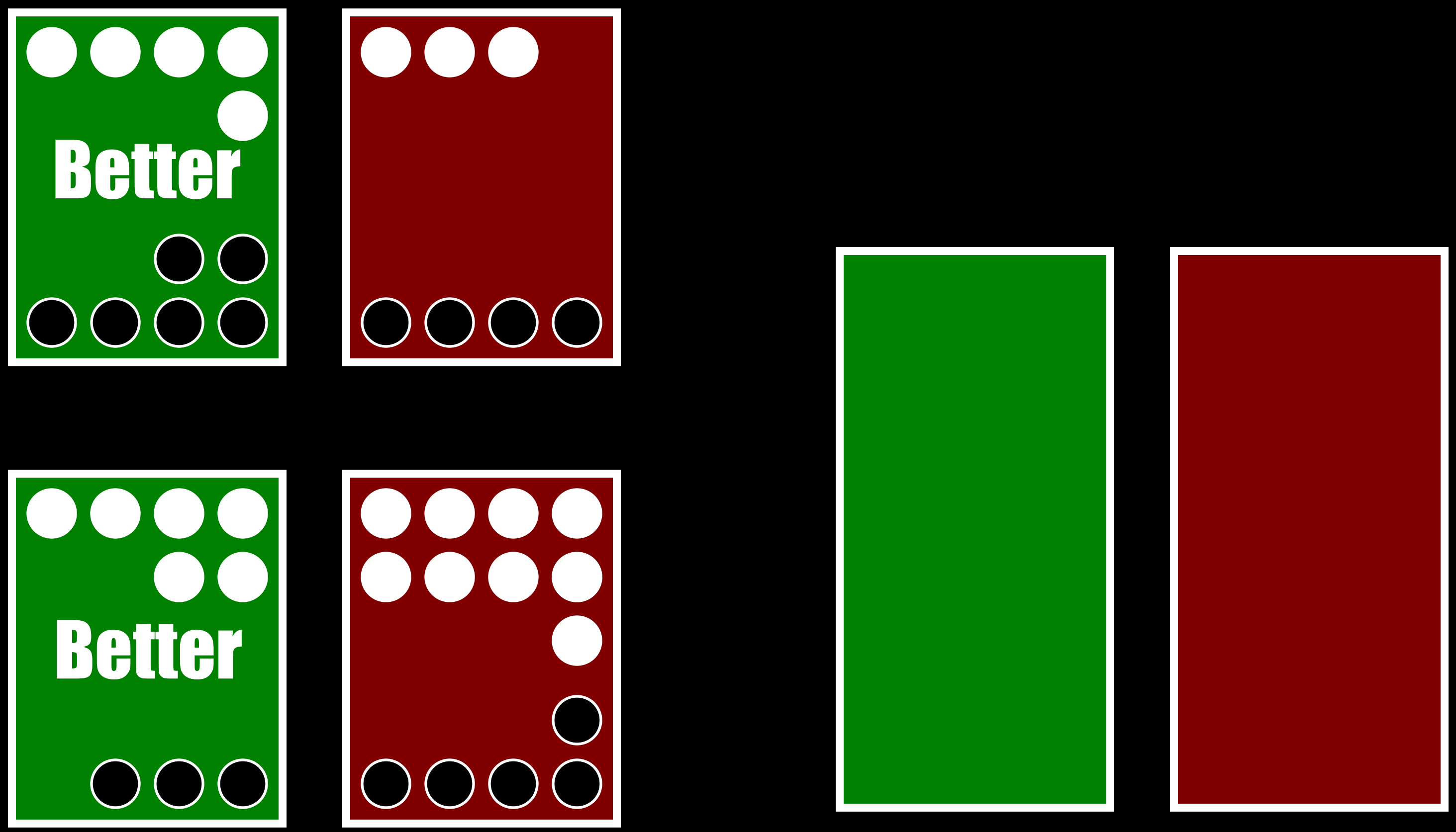
5/11 > 3/7

Try to get a white ball

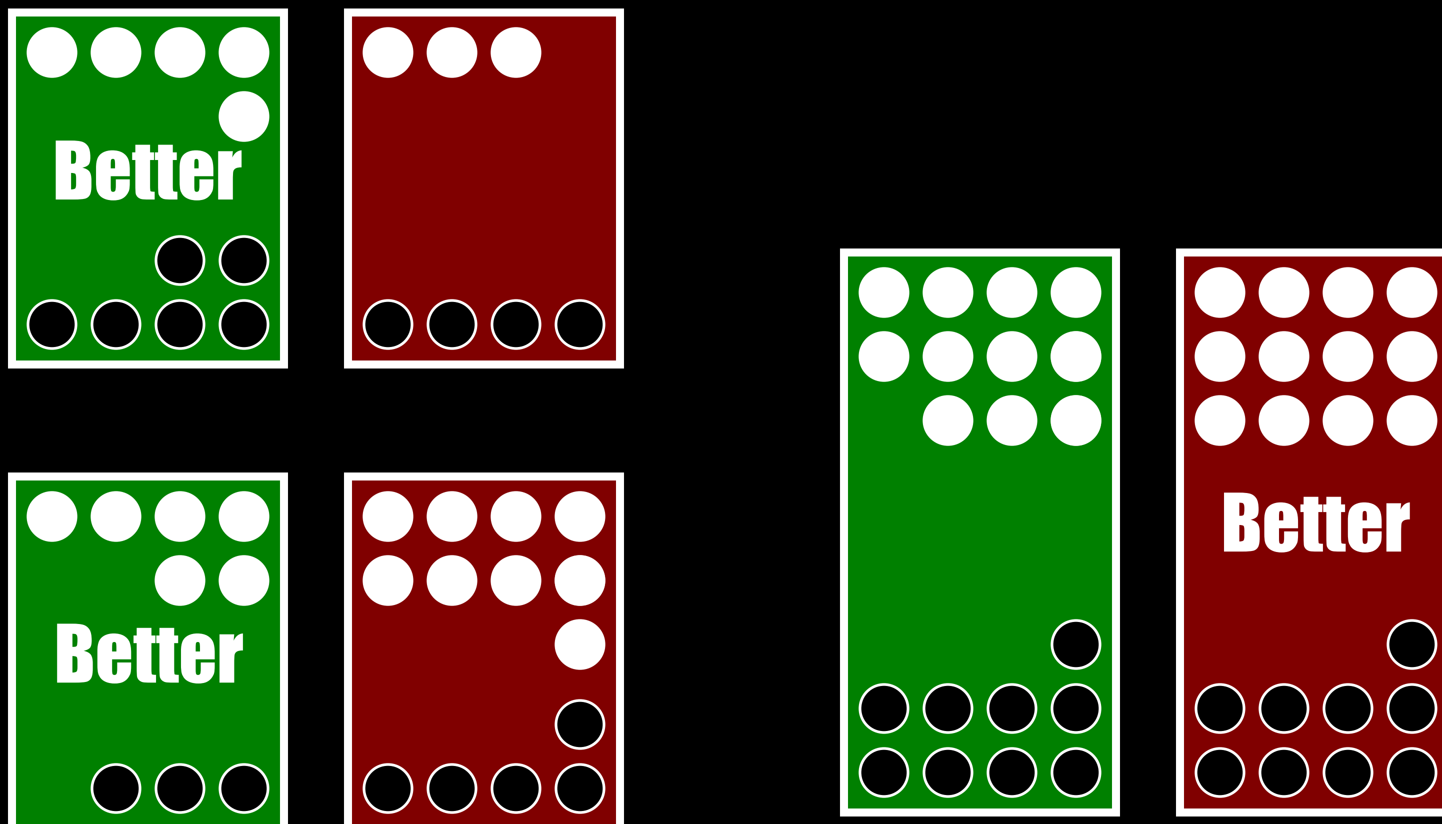


6/9 > 9/14

Try to get a white ball



Try to get a white ball



$11/20 < 12/21$!!!

Simpson's Paradox

Arises all the time…

Be careful when you interpret numbers

Department of Transportation requires that each month all airlines report their "on-time record"

$$\frac{\text{\# of on-time flights landing at nation's 30 busiest airports}}{\text{\# of total flights into those airports}}$$

http://www.bts.gov/programs/oai/

Different airlines serve different airports with different frequency

An airline sending most of its planes into fair weather airports will crush an airline flying mostly into foggy airports

It can even happen that an airline has a better record at each airport, but gets a worse overall rating by this method.

	Alaska airlines		America West	
	% on time	# flights	% on time	# flights
LA	88.9	559	85.6	811
Phoenix	94.8	233	92.1	5255
San Diego	91.7	232	85.5	448
SF	83.1	605	71.3	449
Seattle	85.8	2146	76.7	262
OVERALL	86.7	3775	89.1	7225

Alaska Air beats America West at each airport but America West has a better overall rating!

Geometric Random Variable

Flip a coin with probability p of heads

Let X = number of times the coin has to be flipped until we get a heads

$$
\begin{aligned}
E[X] = \; & E[X \mid \text{flip 1 is H}]\Pr[\text{flip 1 is H}] + \\
& E[X \mid \text{flip 1 is T}]\Pr[\text{flip 1 is T}]
\end{aligned}
$$

$$= p + E[X+1](1-p)$$

$$= p + E[X](1-p) + (1-p)$$

$$E[X] = 1/p$$

CMULand

All CMU SCS students fly off to space and colonize the moon

Faced with the problem that there are more men than women, the authorities impose a new rule:

When having kids, stop after you have a girl

Will the number of new girls be higher than the number of new boys?

What if the rule is: stop after having two girls?

Coupon Collector's Problem

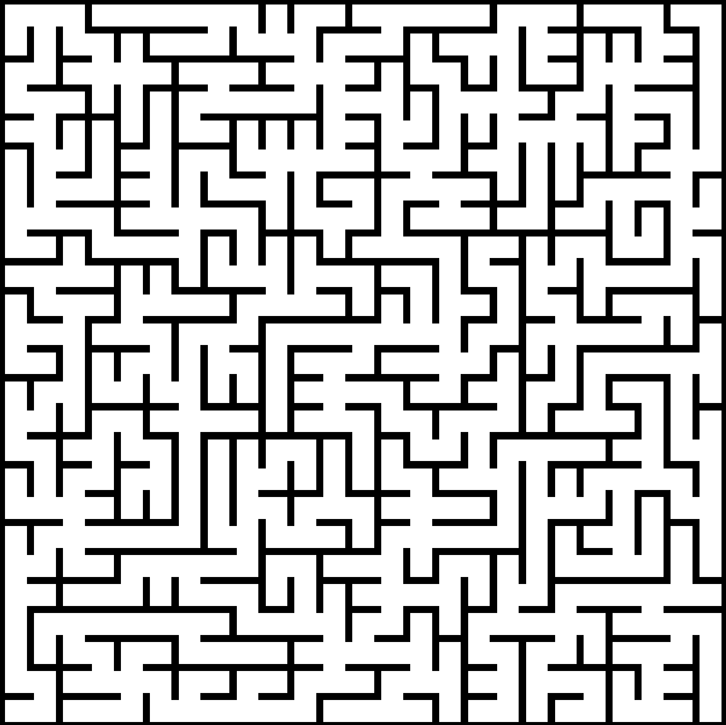
There are n types of coupons in cereal boxes

Want to collect them all

On average, how many cereal boxes do I have to buy to get them all?

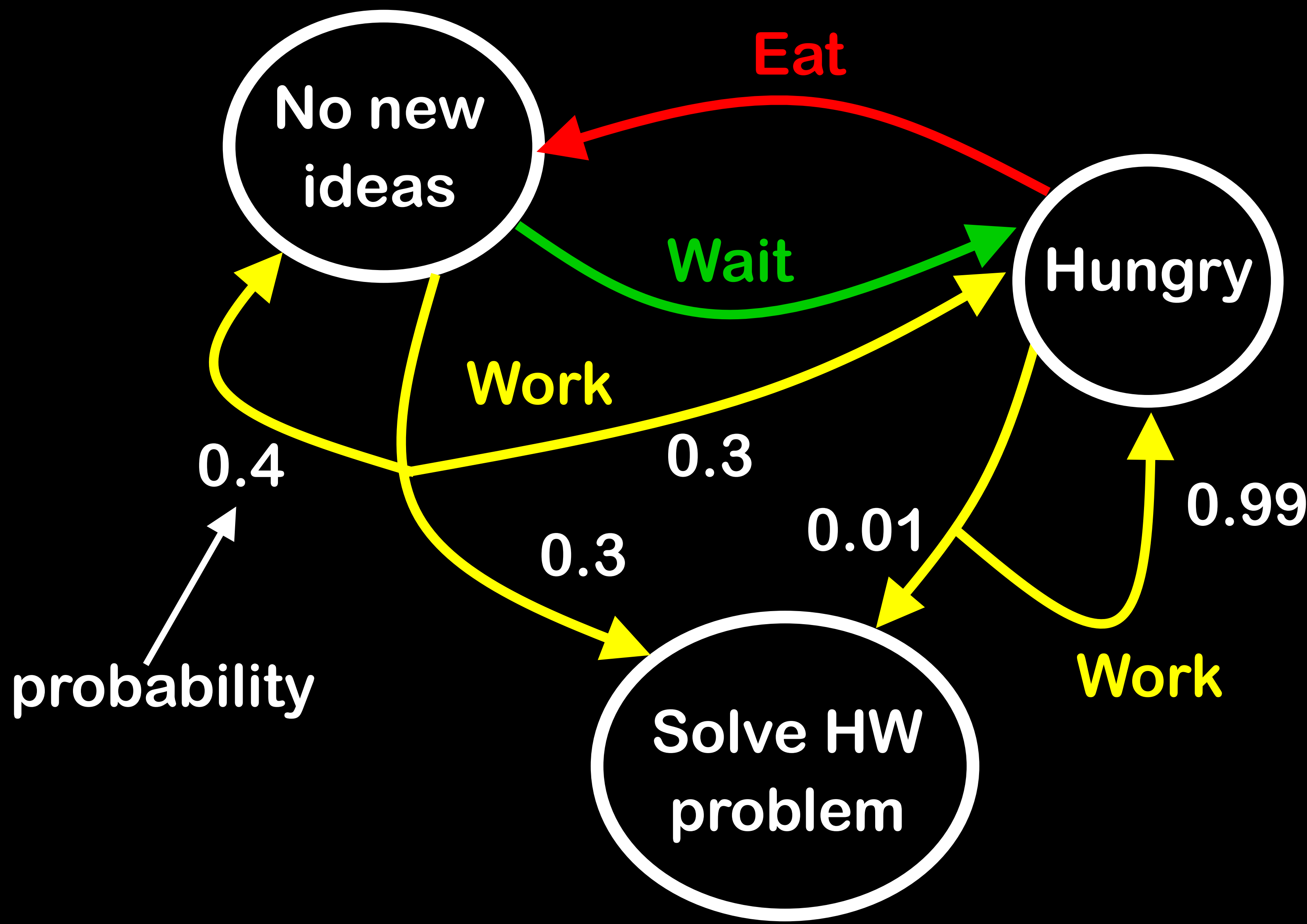
Random Walks

Lecture 12 (February 19, 2009)

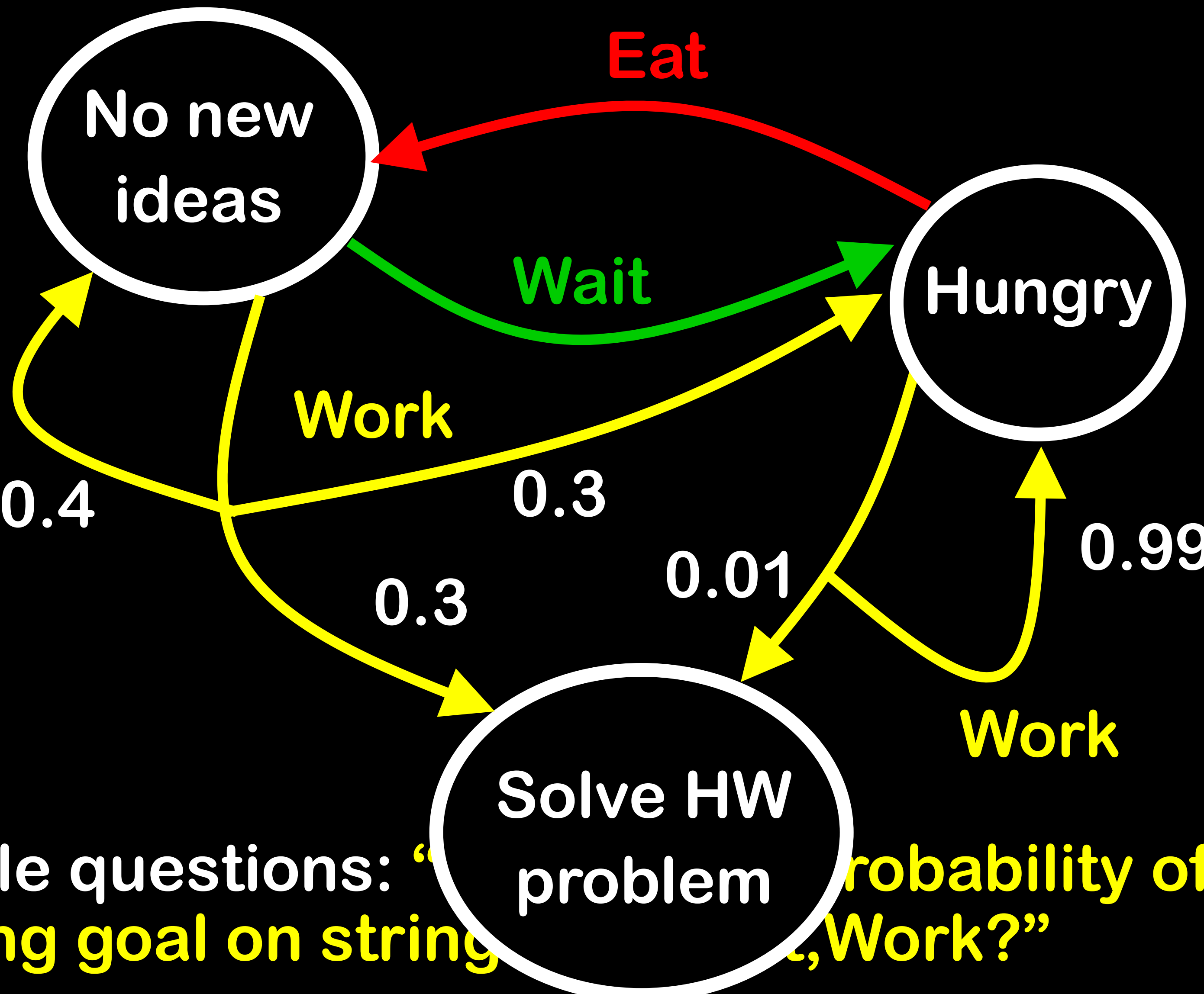


How to walk home drunk

Abstraction of Student Life

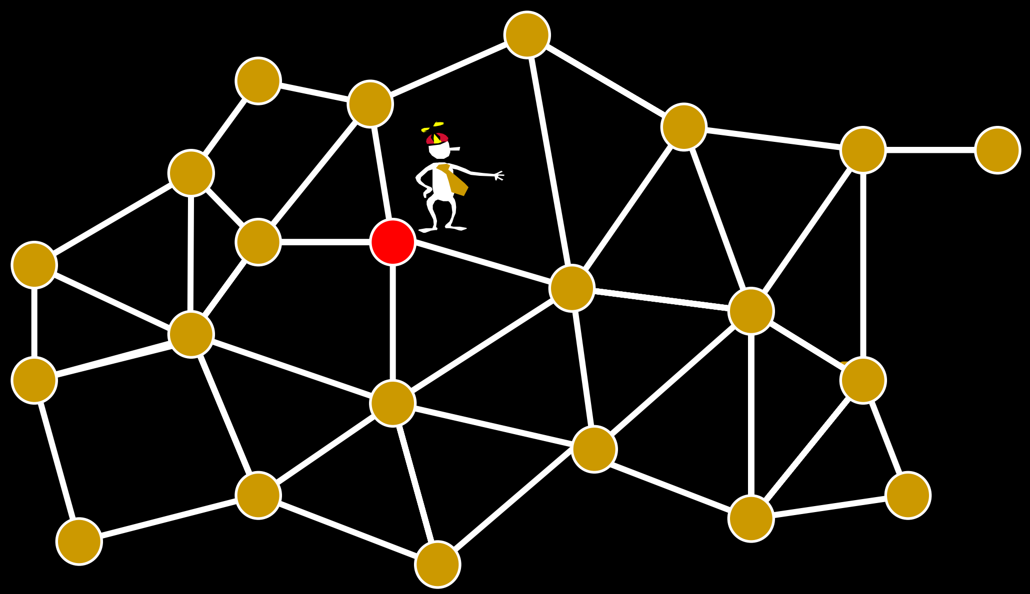


Abstraction of Student Life



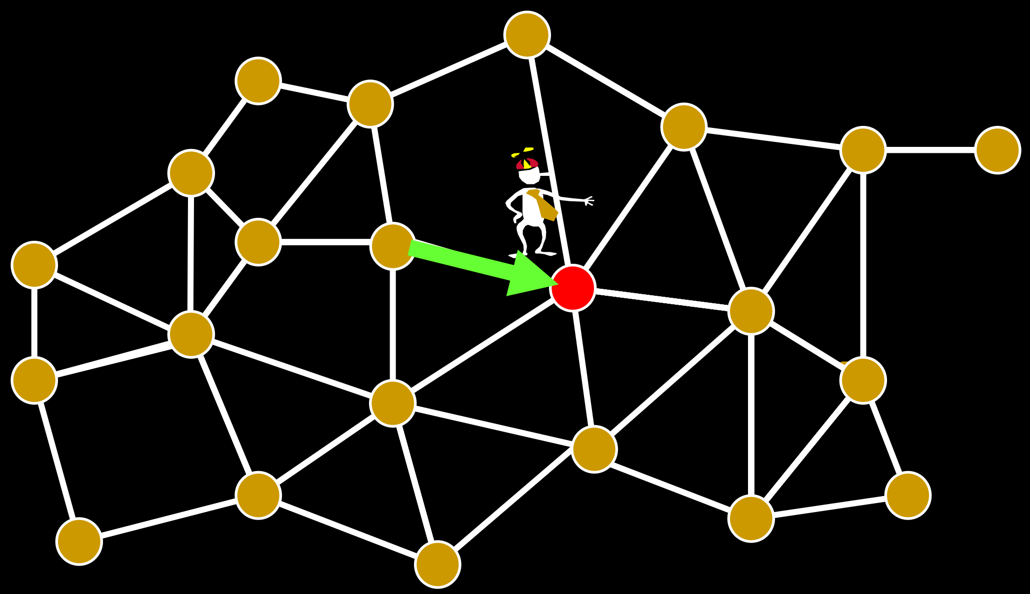
Example questions: “… robability of reaching goal on string … ,Work?”

Simpler:
Random Walks on Graphs



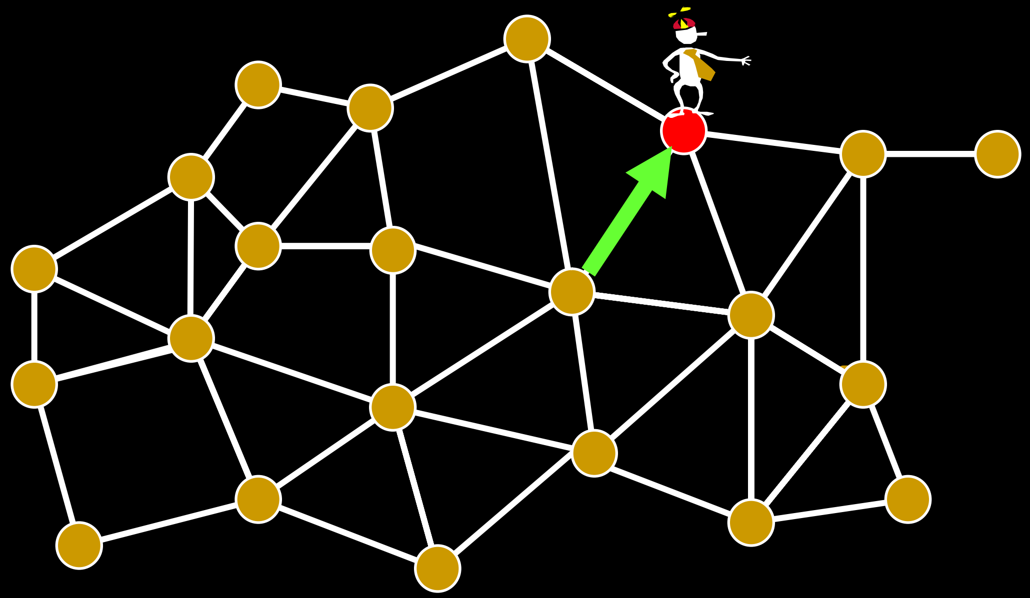
At any node, go to one of the neighbors of the node with equal probability

Simpler:
Random Walks on Graphs



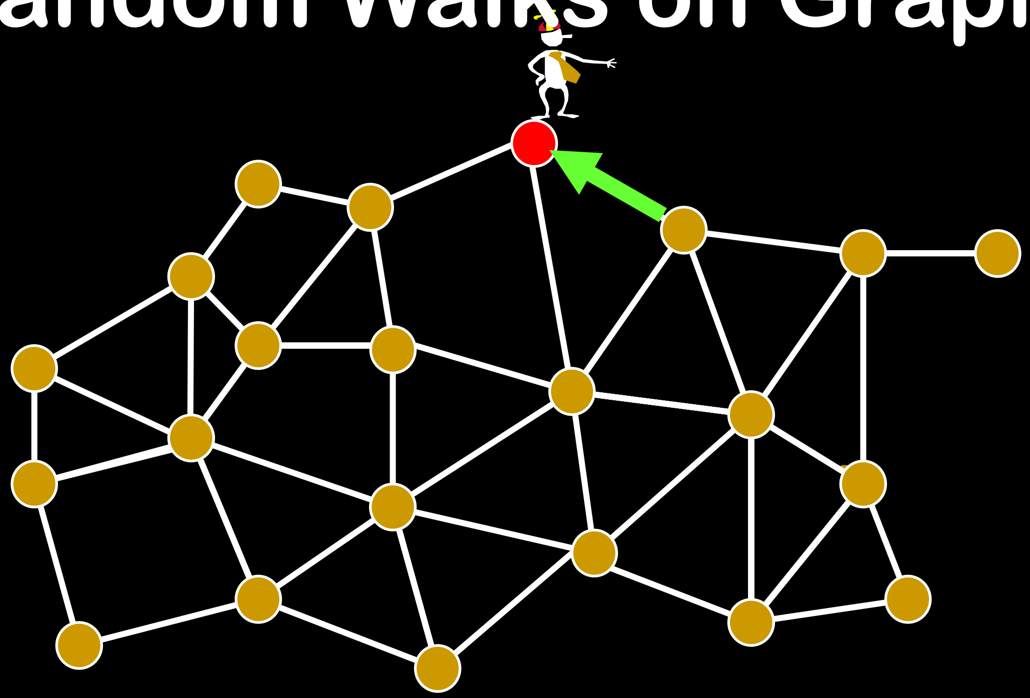
At any node, go to one of the neighbors of the node with equal probability

Simpler: Random Walks on Graphs



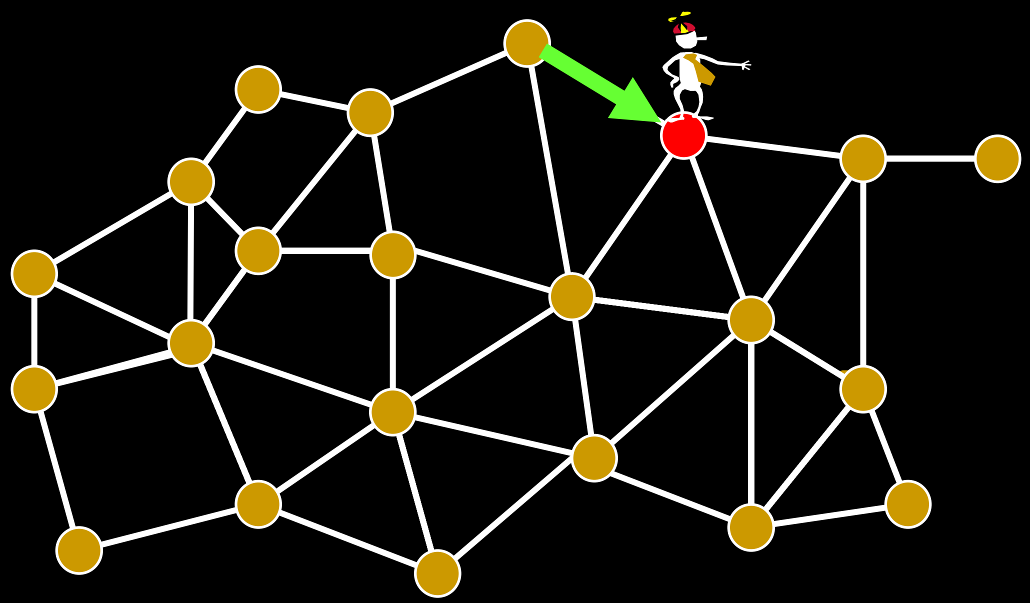
At any node, go to one of the neighbors of the node with equal probability

Simpler:
Random Walks on Graphs



At any node, go to one of the neighbors of the node with equal probability

Simpler:
Random Walks on Graphs

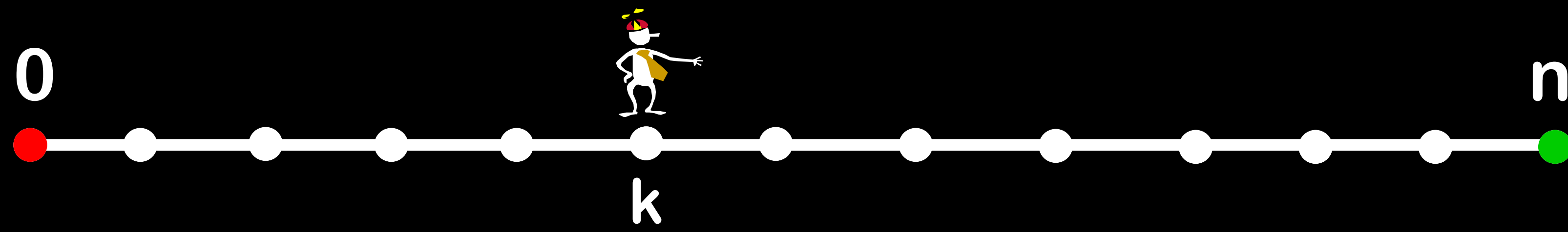


At any node, go to one of the neighbors of the node with equal probability

Random Walk on a Line

You go into a casino with $k, and at each time step, you bet $1 on a fair game

You leave when you are broke or have $n



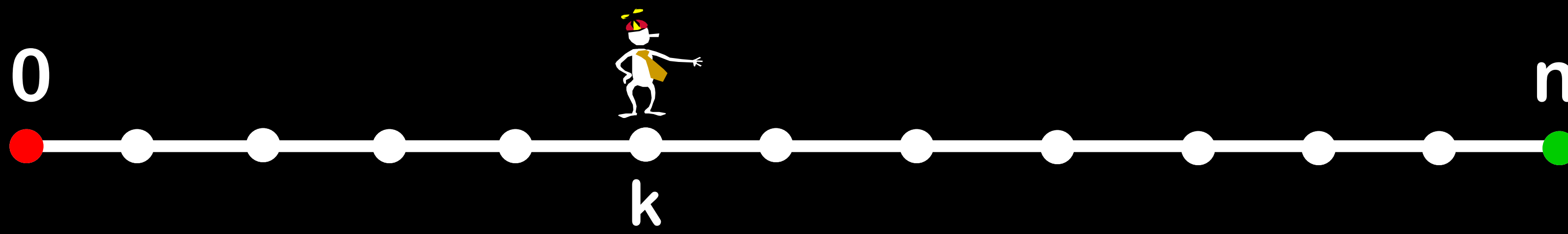
Question 1: what is your expected amount of money at time t?

Let X_t be a R.V. for the amount of $$$ at time t

Random Walk on a Line

You go into a casino with $k, and at each time step, you bet $1 on a fair game

You leave when you are broke or have $n



$X_t = k + \delta_1 + \delta_2 + \ldots + \delta_t,$

(δ_i is RV for change in your money at time i)

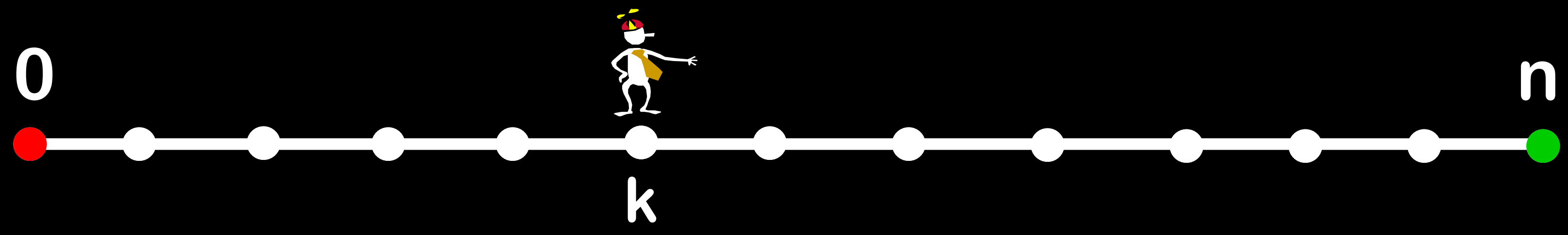
$E[\delta_i] = 0$

So, $E[X_t] = k$

Random Walk on a Line

You go into a casino with $k, and at each time step, you bet $1 on a fair game

You leave when you are broke or have $n



Question 2: what is the probability that you leave with $n?

Random Walk on a Line

Question 2: what is the probability that you leave with **\$n**?

$E[X_t] = k$

$$E[X_t] = E[X_t \mid X_t = 0] \times \Pr(X_t = 0)$$
$$+ E[X_t \mid X_t = n] \times \Pr(X_t = n)$$
$$+ E[X_t \mid \text{neither}] \times \Pr(\text{neither})$$

$$k = n \times \Pr(X_t = n)$$
$$+ (\text{something}_t) \times \Pr(\text{neither})$$

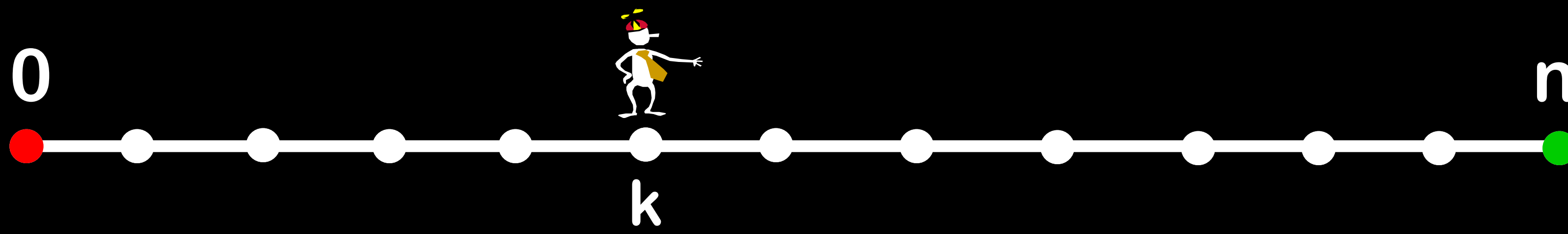
As $t \to \infty$, $\Pr(\text{neither}) \to 0$, also $\text{something}_t < n$

Hence $\Pr(X_t = n) \to k/n$

Another Way To Look At It

You go into a casino with \$k, and at each time step, you bet \$1 on a fair game

You leave when you are broke or have \$n

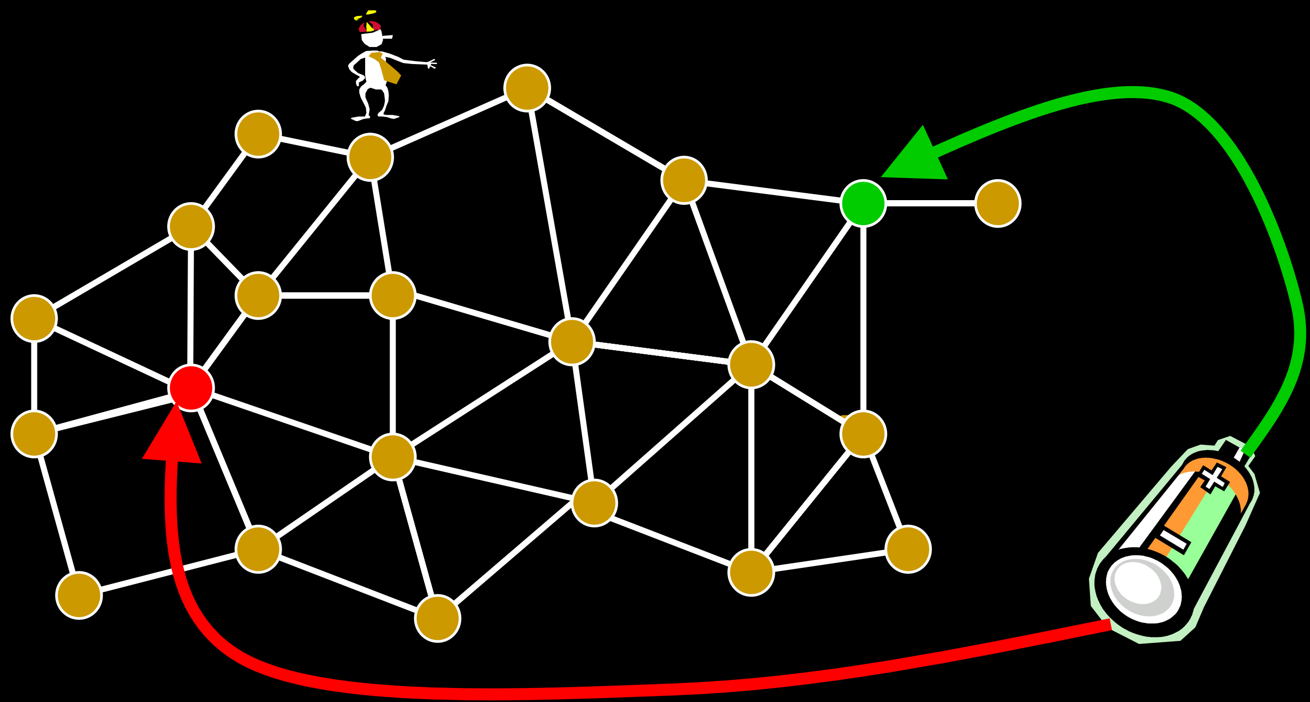


Question 2: what is the probability that you leave with \$n?

= probability that I hit green before I hit red

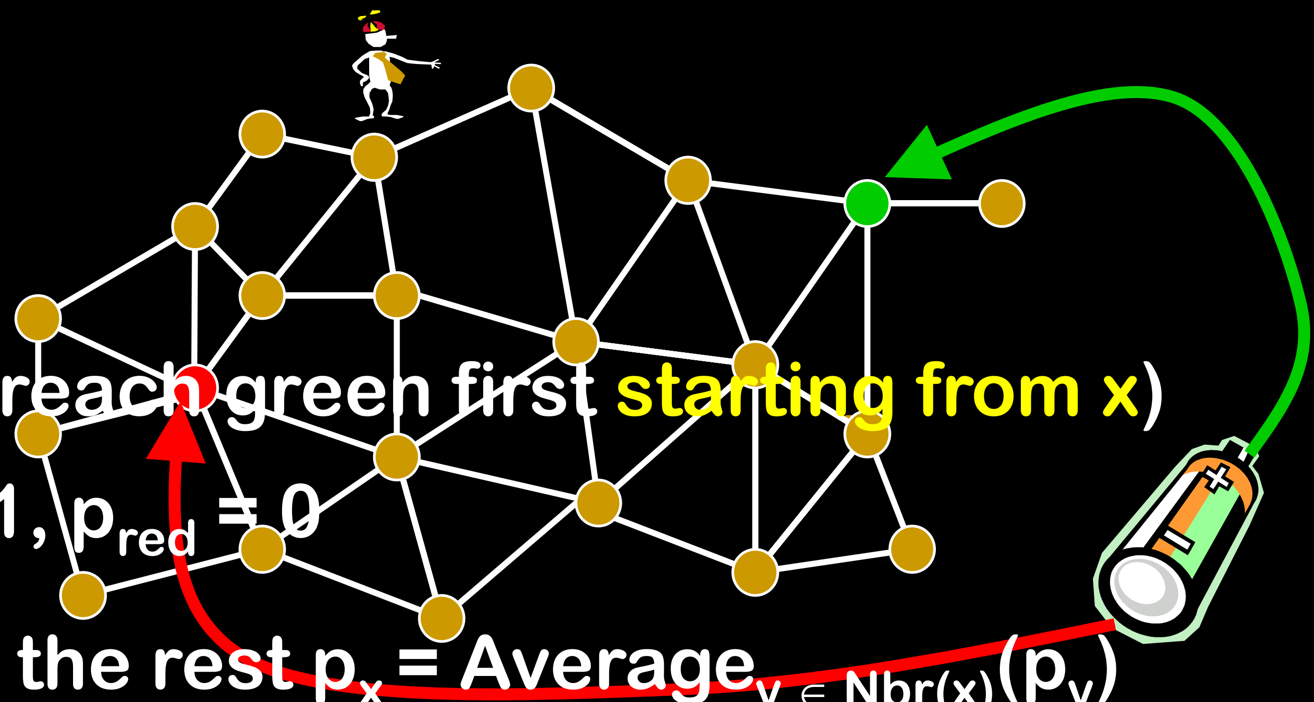
Random Walks and Electrical Networks

What is chance I reach green before red?



Same as voltage if edges are resistors and we put 1-volt battery between green and red

Random Walks and Electrical Networks



p_x = Pr(reach green first starting from x)

$p_{green} = 1$, $p_{red} = 0$

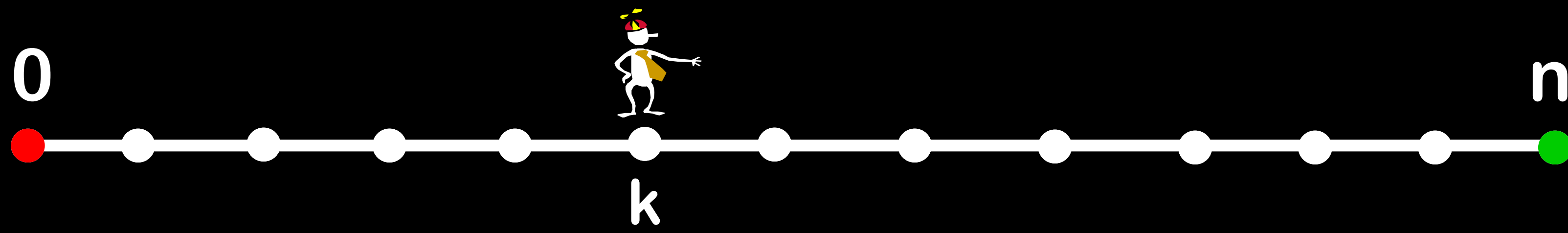
And for the rest $p_x = \text{Average}_{y \in Nbr(x)}(p_y)$

Same as equations for voltage if edges all have same resistance!

Another Way To Look At It

You go into a casino with \$k, and at each time step, you bet \$1 on a fair game

You leave when you are broke or have \$n

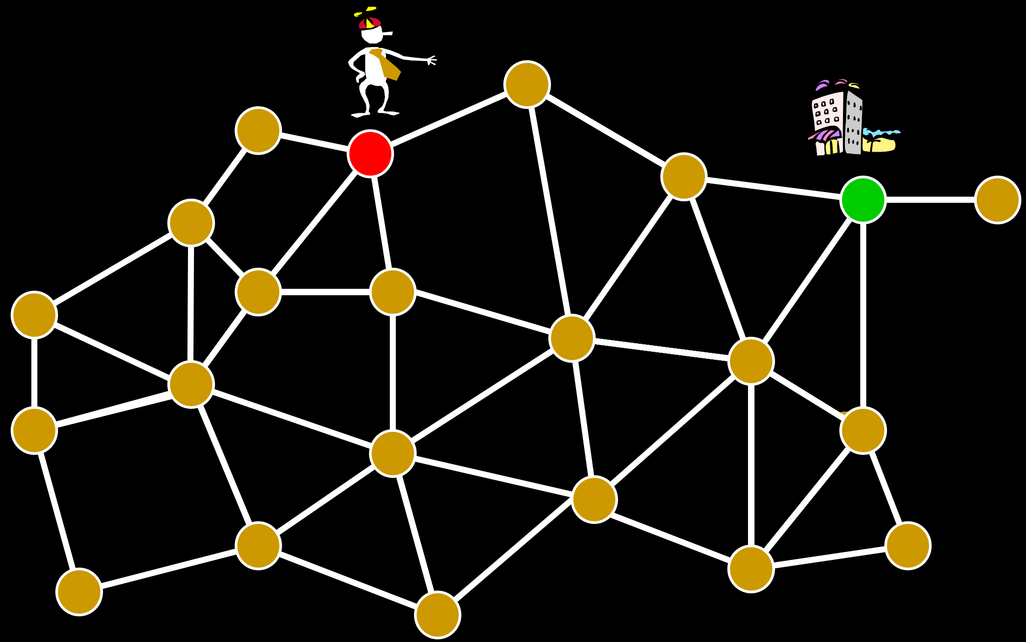


Question 2: what is the probability that you leave with \$n?

voltage(k) = k/n

= Pr[hitting n before 0 starting at k] !!!

Getting Back Home



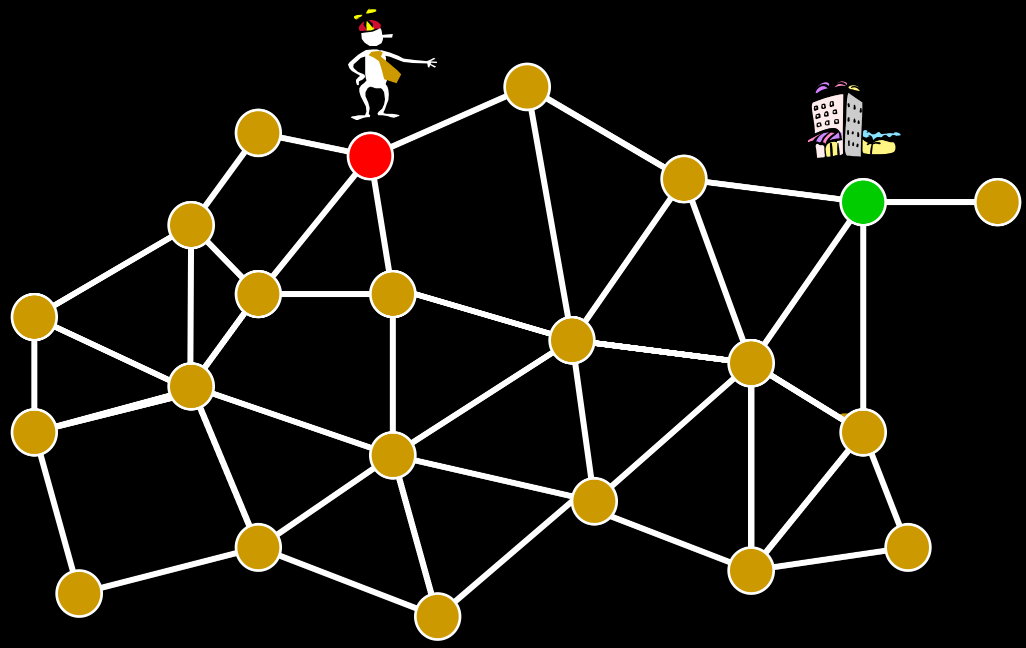
Lost in a city, you want to get back to your hotel

How should you do this?

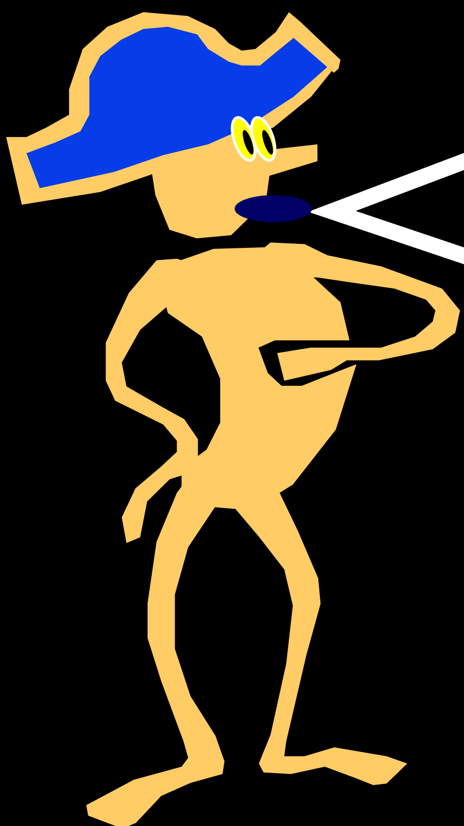
Depth First Search!

Requires a good memory and a piece of chalk

Getting Back Home



How about walking randomly?



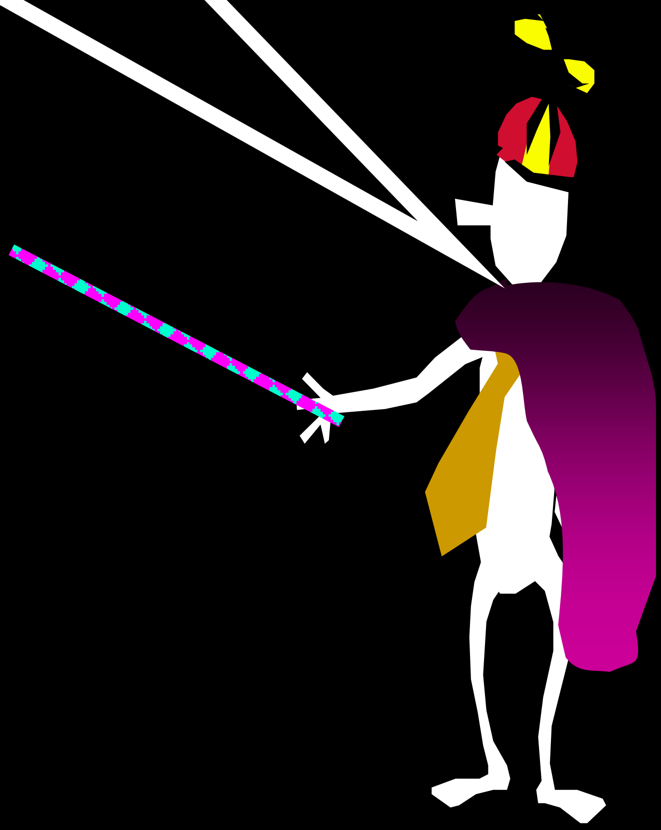
Will this work?

Is Pr[reach home] = 1?

When will I get home?

What is
E[time to reach home]?

Pr[will reach home] = 1



We Will Eventually Get Home

Look at the first n steps

There is a non-zero chance p_1 that we get home

Also, $p_1 \geq (1/n)^n$

Suppose we fail

Then, wherever we are, there is a chance $p_2 \geq (1/n)^n$ that we hit home in the next n steps from there

Probability of failing to reach home by time kn

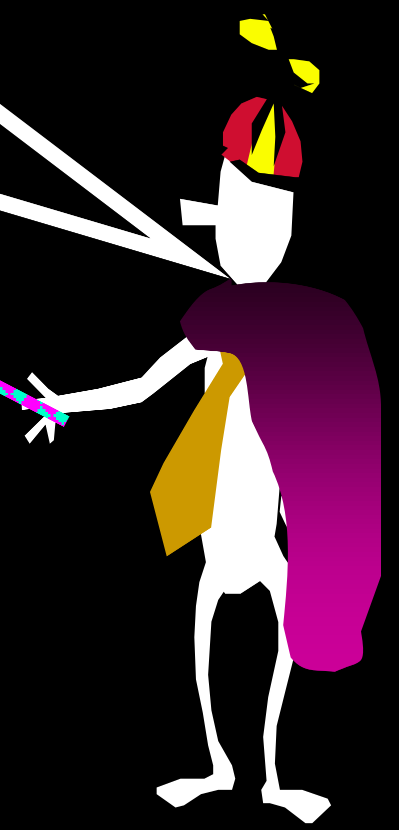
$$= (1 - p_1)(1 - p_2) \ldots (1 - p_k) \rightarrow 0 \text{ as } k \rightarrow \infty$$

Furthermore:

If the graph has
n nodes and m edges, then

E[time to visit all nodes]
$\leq 2m \times (n-1)$

E[time to reach home] is at
most this



Actually, we get home pretty fast…

Chance that we don't hit home by $(2k)2m(n-1)$ steps is $(\tfrac{1}{2})^k$

A Simple Calculation

True of False:

If the average income of people is $100 then

more than 50% of the people can be earning more than $200 each

False! else the average would be higher!!!

Markov's Inequality

If X is a non-negative r.v. with mean E[X], then

$$\Pr[\, X > 2\, E[X] \,] \leq \tfrac{1}{2}$$

$$\Pr[\, X > k\, E[X] \,] \leq 1/k$$

Andrei A. Markov

Markov's Inequality

Non-neg random variable X has expectation A = E[X]

$$A = E[X] = E[X \mid X > 2A]\ \Pr[X > 2A] + E[X \mid X \le 2A]\ \Pr[X \le 2A]$$

$$\ge E[X \mid X > 2A]\ \Pr[X > 2A] \quad \text{(since X is non-neg)}$$

Also, $E[X \mid X > 2A] > 2A$

$$\Rightarrow A \ge 2A \times \Pr[X > 2A]$$

$$\Rightarrow \tfrac{1}{2} \ge \Pr[X > 2A]$$

$\Pr[\, X > k \times \text{expectation}\,] \le 1/k$

Actually, we get home pretty fast…

Chance that we don't hit home by $(2k)2m(n-1)$ steps is $(\frac{1}{2})^k$

An Averaging Argument

Suppose I start at u

E[time to hit all vertices | start at u] $\leq$ C(G)

Hence, by Markov's Inequality:
Pr[time to hit all vertices > 2C(G) | start at u] $\leq$ ½

So Let's Walk Some Mo!

Pr [time to hit all vertices > 2C(G) | start at u] $\leq$ ½

Suppose at time 2C(G), I'm at some node
with more nodes still to visit

Pr [haven't hit all vertices in 2C(G) <u>more</u> time
| start at v] $\leq$ ½

Chance that you failed both times $\leq$ ¼ = $(½)^2$

Hence,
Pr[havent hit everyone in time k × 2C(G)] $\leq$ $(½)^k$

Hence, if we know that

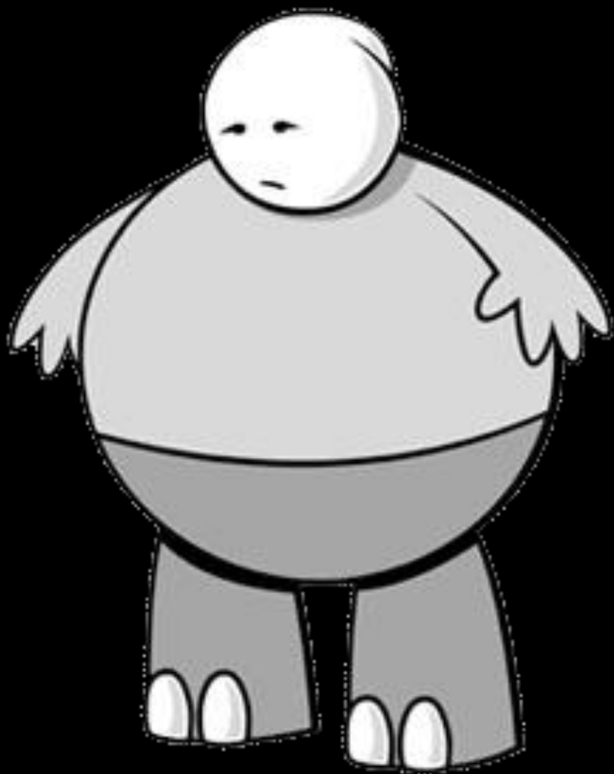
Expected Cover Time

$C(G) < 2m(n-1)$

then

Pr[home by time $4k\ m(n-1)$]

$\geq 1 - (½)^k$



Here's What You Need to Know…

Random Walk in a Line

Cover Time of a Graph

Markov's Inequality