Great Theoretical Ideas In Computer Science

Matt Streeter
(A. Gupta, J. Lafferty)

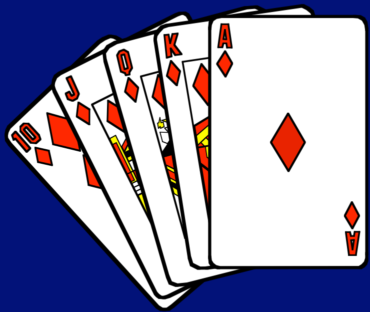
CS 15-251 Spring 2006

Lecture 23

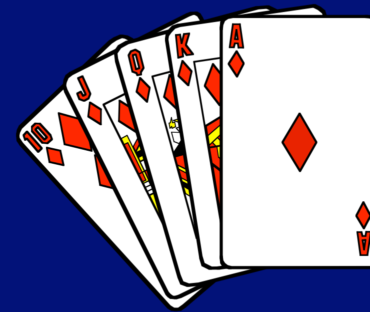
Apr. 11, 2006

Carnegie Mellon University

Probability III: The Probabilistic Method



Five Volunteers...



Mario	2/10
Eric 1	1/10
Eric 2	2/10
Mike	1/10
Robert	2/10

Two Equivalent Facts

Fact 1: in any five-card hand, some pair has the same suit.

Fact 2: in any five card hand, Pr(random pair has same suit) > 0

Fact 1 ⇔ Fact 2.

Proof

Let n = #(same-suit pairs)

Let A = event "random pair has same suit"

$\Pr(A) = n/10$

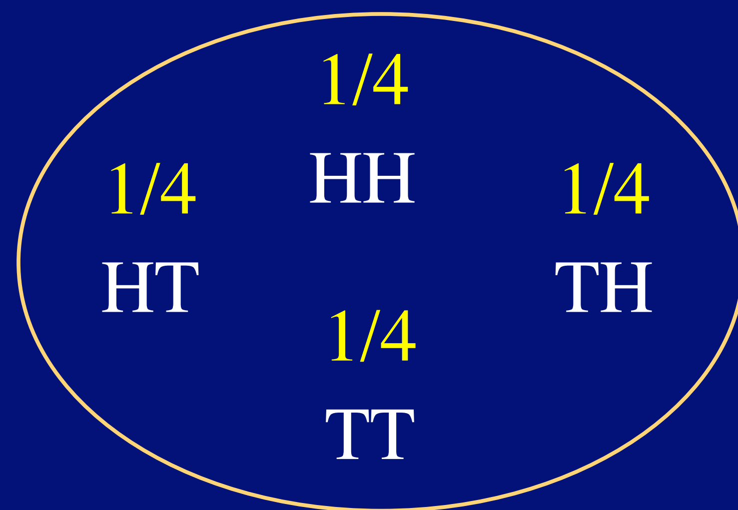
So, n is always > 0 iff. $\Pr(A)$ is always > 0

Review

Random Variables

- An event is a subset of S.
- A Random Variable (RV) is a (real-valued) function on S.

Example: flip two coins



- S = {HH,HT,TH,TT}
- Event A: first coin is heads
- Random Variable X: the number of heads

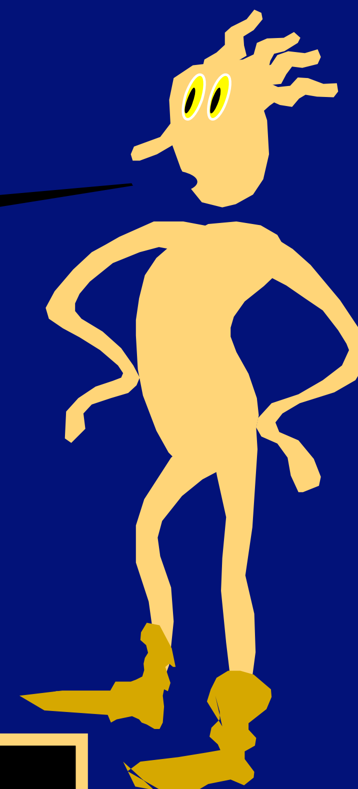
E.g., X(HH)=2, X(HT)=1.

Two Views

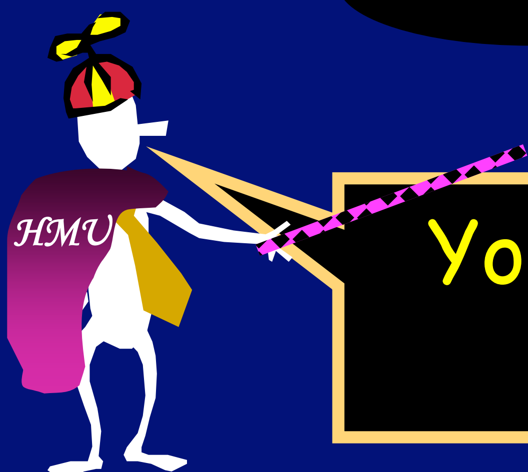
D	S	X	Distrib on X
1/4	TT		0 --- 1/4
1/4	TH		1 --- 1/2
1/4	HT		2 --- 1/4
1/4	HH		

It's a floor wax and a dessert topping

It's a function on
the sample space S.



It's a variable with a
probability distribution
on its values.



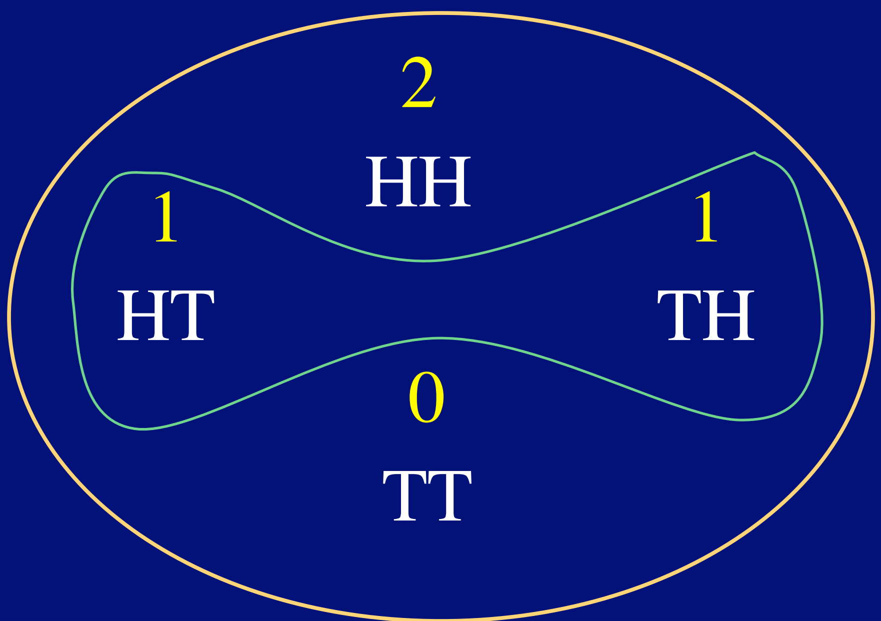
You should be comfortable
with both views.

Definition: expectation

The expectation, or expected value of a random variable X is

E.g, 2 coin flips,
X = # heads.

What is E[X]?



Two Views

D	S		Distrib on X
1/4	TT	X	0 --- 1/4
1/4	TH		1 --- 1/2
1/4	HT		2 --- 1/4
1/4	HH		

E[X] = (1/4)*0 + (1/4)*1 + (1/4)*1 + (1/4)*2 = 1

E[X] = (1/4)*0 + (1/2)*1 + (1/4)*2 = 1

Linearity of Expectation

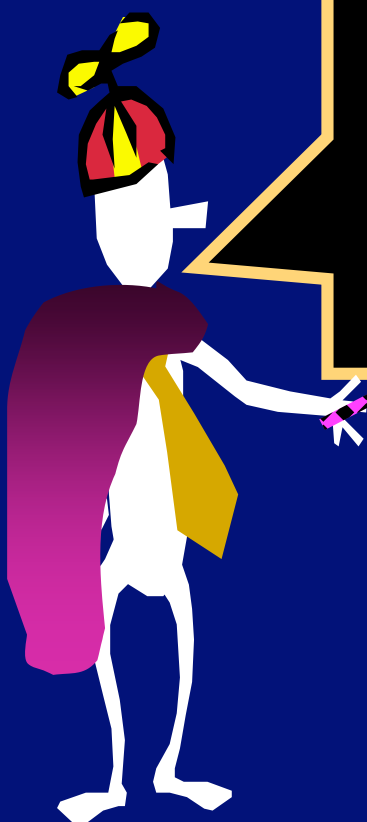
If $Z = X+Y$, then

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

Even if X and Y are not independent.

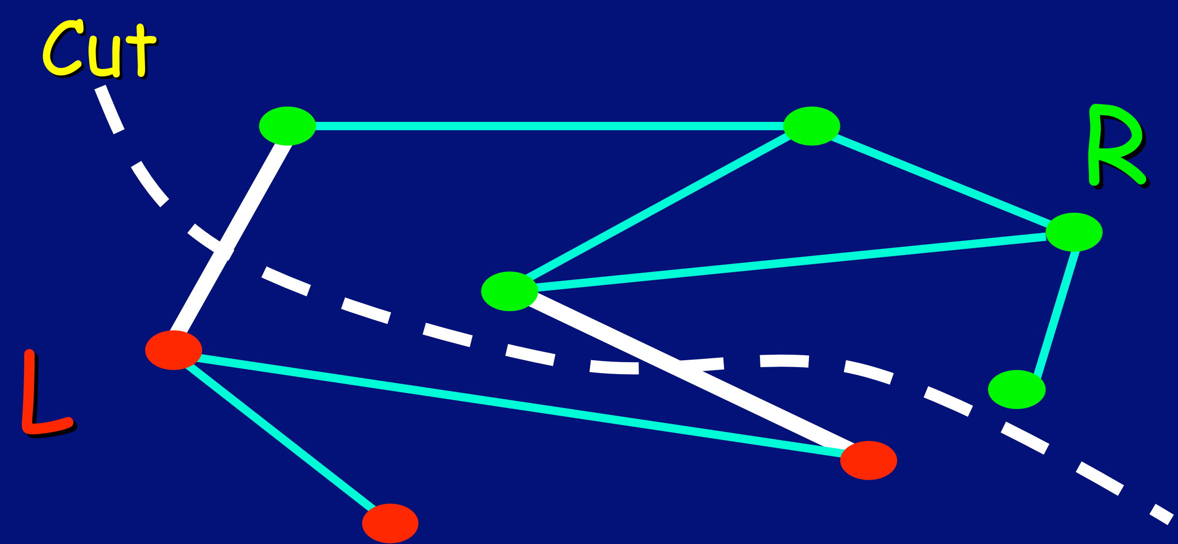
New Topic: The probabilistic method

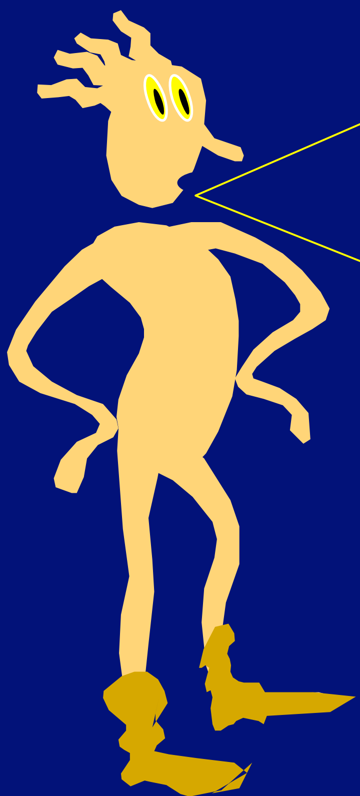
Use a probabilistic argument to prove a non-probabilistic mathematical theorem.



Definition: A cut in a graph.

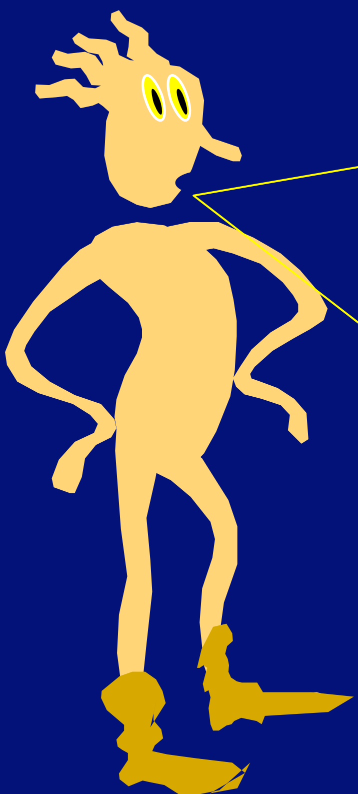
A *cut* is a partition of the vertices of a graph into two sets: L and R. We say that an edge *crosses the cut* if it goes from a vertex in L to a vertex in R.





Theorem:

In any graph, there exists a cut such that at least half the edges cross the cut.



Theorem:

In any graph, there exists a cut such that at least half the edges cross the cut.

How are we going to prove this?

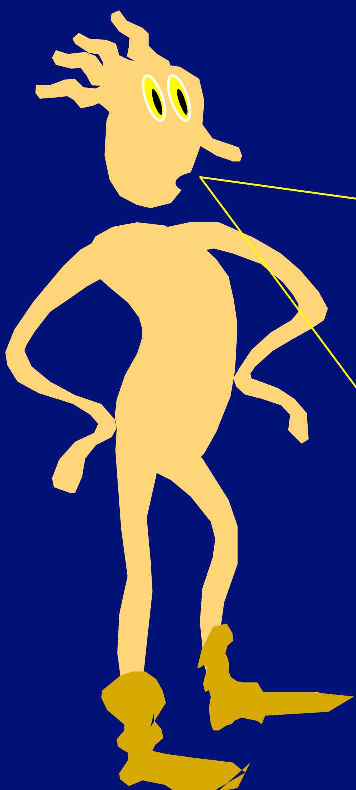
Will show that if we pick a cut at random, the expected number of edges crossing is e/2.

Why does this prove the theorem?

Pigeonhole principle: Given n boxes and $m > n$ objects, at least one box must contain more than one object.

Letterbox principle: If the average number of letters per box is a, then some box will have at least a letters. (Similarly, some box has at most a.)

If $E[X] = v$, then $\Pr(X \geq v) > 0$.



Pick a cut uniformly at random. I.e., for each vertex flip a fair coin to see if it should be in L or R.

E[#of edges crossing cut] = e/2

The sample space of all possible cuts must contain at least one cut that at least half the edges cross: if not, the average number of edges would be less than half!

Theorem:

In any graph, there exists a cut such that at least half the edges cross the cut.

Proof: Pick a cut uniformly at random. I.e., for each vertex flip a fair coin to determine if it is in L or R.

Let X_a be the indicator R.V. for the event that edge $a = \{u,v\}$ crosses the cut. What is $E[X_a]$?

u	v	X_a
L	L	0
L	R	1
R	L	1
R	R	0

$$E[X_a] = 1/4(0+1+1+0)$$
$$= 1/2$$

Theorem:

In any graph, there exists a cut such that at least half the edges cross the cut.

Proof: Pick a cut uniformly at random. I.e., for each vertex flip a fair coin to determine if it is in L or R.

Let X_a be the indicator R.V. for the event that edge $a = \{u,v\}$ crosses the cut. What is $E[X_a]$? Ans: 1/2.

Let X = #(edges crossing cut). So, $X = \sum_a X_a$.

By linearity of expectation, $E[X] = \sum_a E[X_a] = e/2$.

So, there must exist a cut with at least $e/2$ edges crossing it.

The Probabilistic Method

1. Define distribution D over objects (doesn't have to be uniform)
2. Define R.V. X(object) = value of object
3. Prove E[X] = v.
4. Conclude there must exist objects with value ≥ v.

A Puzzle

10% of the surface of a sphere is colored green, and the rest is colored blue. Show that now matter how the colors are arranged, it is possible to inscribe a cube in the sphere so that all of its vertices are blue.

Solution

Pick a random cube. (Note: any particular vertex is uniformly distributed over surface of sphere).

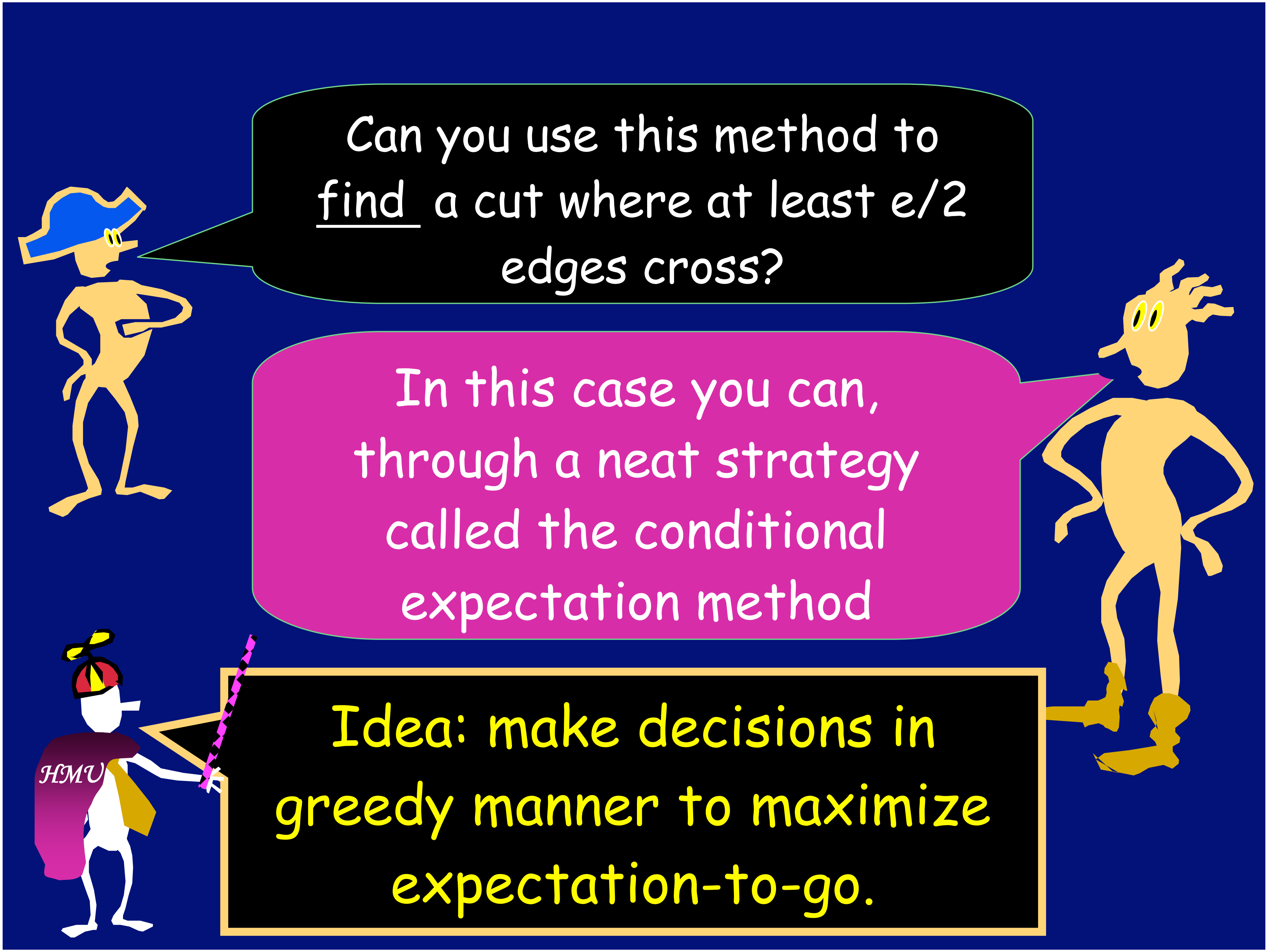
Let X_i = 1 if i^{th} vertex is blue, 0 otherwise

$E[X_i] = Pr(X_i=1) = 9/10$

Let $X = X_1 + X_2 + ... + X_8$

$E[X] = 8*9/10 > 7$

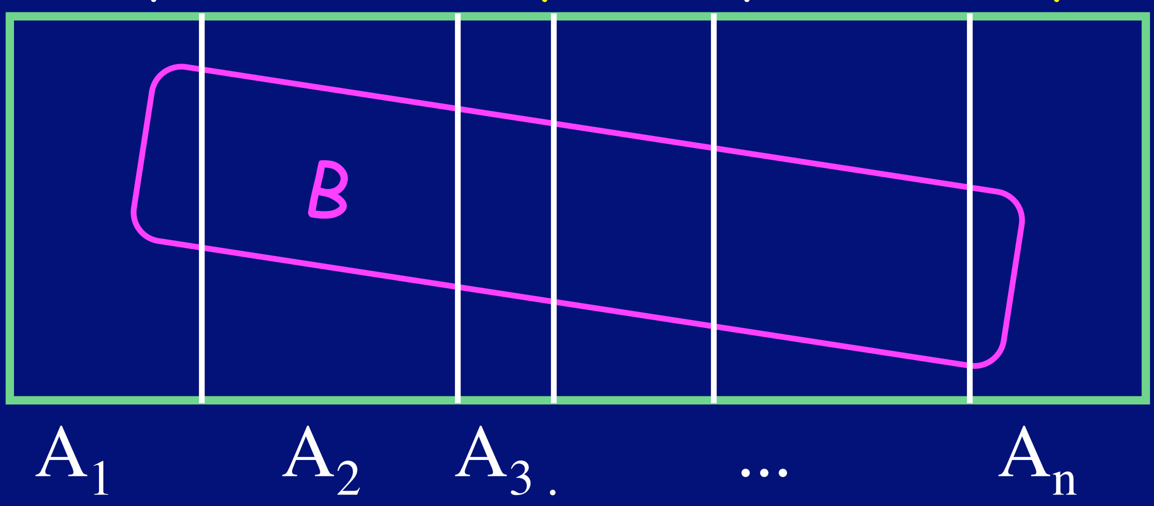
So, must have some cubes where $X = 8$.


Can you use this method to find a cut where at least e/2 edges cross?
In this case you can, through a neat strategy called the conditional expectation method
Idea: make decisions in greedy manner to maximize expectation-to-go.
HMU

First, a few more facts...

For any partition of the sample space S into disjoint events $A_1, A_2, \ldots, A_n$, and any event B,

$\Pr(B) = \sum_i \Pr(B \cap A_i) = \sum_i \Pr(B|A_i)\Pr(A_i)$.



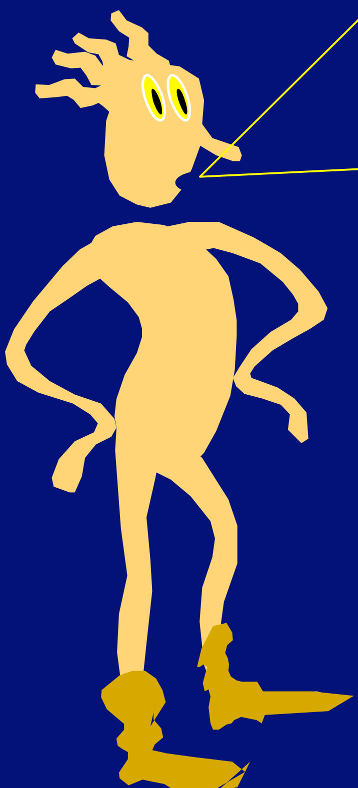
Def: Conditional Expectation

For a random variable X and event A, the conditional expectation of X given A is defined as:

Useful formula: for any partition of S into $A_1, A_2, \ldots$ we have: $E[X] = \sum_i E[X|A_i]\Pr(A_i)$.

Proof: just plug in $\Pr(X=k) = \sum_i \Pr(X=k|A_i)\Pr(A_i)$.

Recap of cut argument



Pick random cut.

- Let X_a=1 if a crosses, else X_a=0.
- Let X = total #edges crossing.
- So, $X = \sum_a X_a$.
- Also, $E[X_a] = 1/2$.
- By linearity of expectation,

$$E[X] = e/2.$$

Conditional expectation method

Say we have already decided fate of vertices 1,2,…,i-1. Let X = number of edges crossing cut if we place rest of vertices into L or R at random.

Let A = event that ver

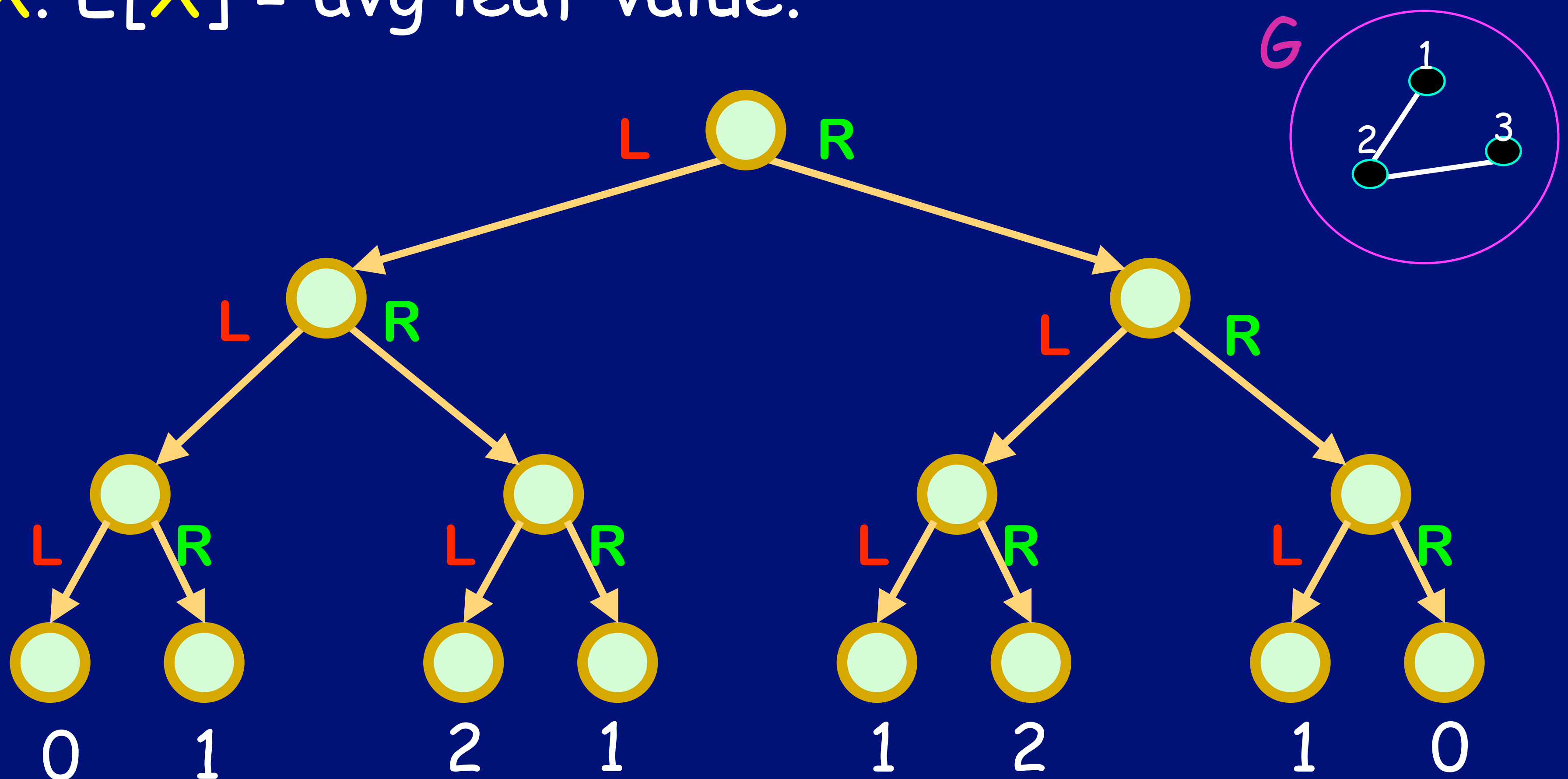
Invariant: the expectation-to-go is always at least e/2

So, $E[X] = E[X|A]*(1/2$

It can't be the case that both terms on the RHS are smaller than the LHS. So just put vertex i into side whose C.E. is larger.

Pictorial view (important!)

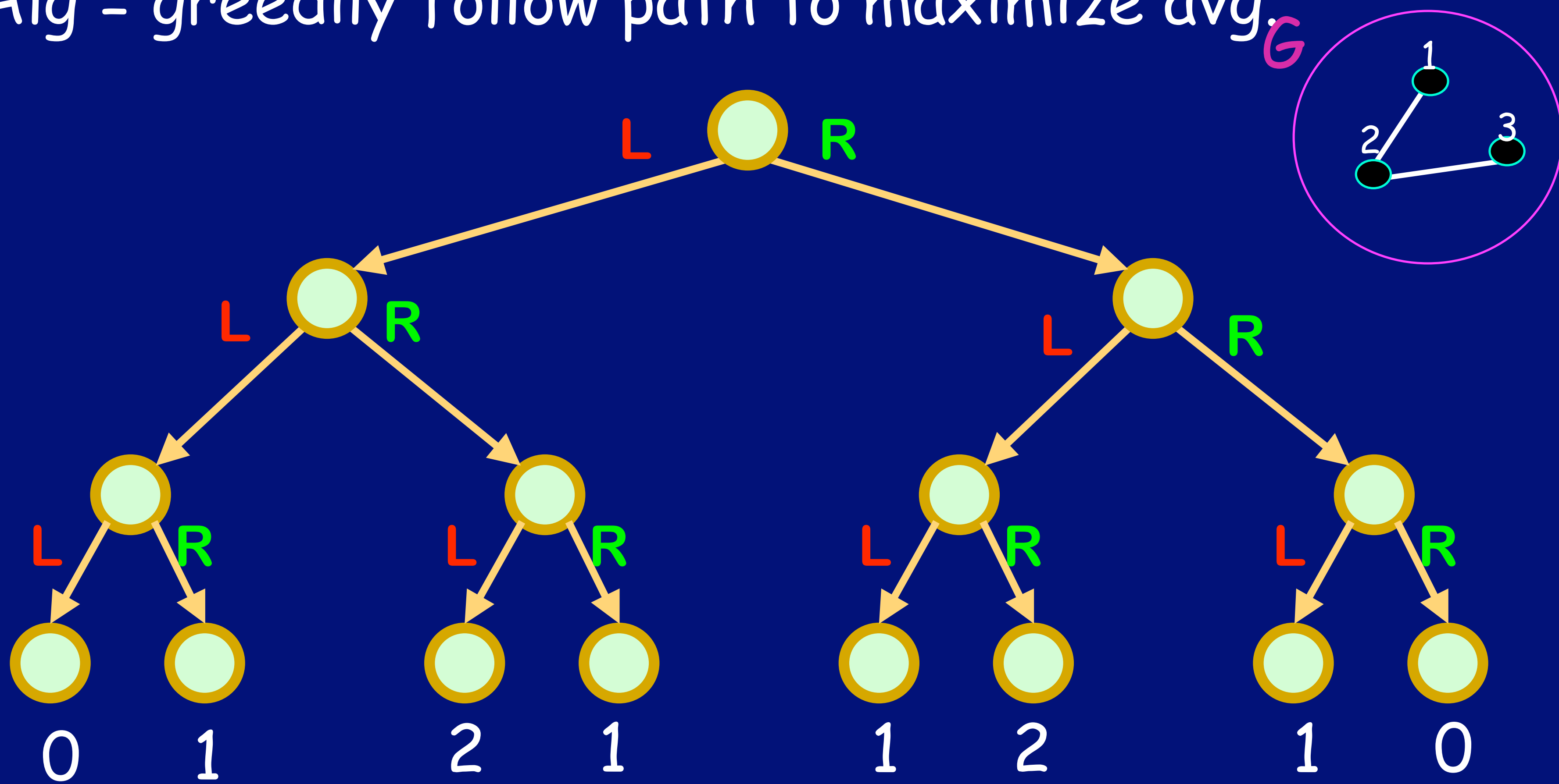
View S as leaves of choice tree. i^{th} choice is where to put vertex i. Label leaf by value of X. E[X] = avg leaf value.



Pictorial view (important!)

If A is some node (the event that we reach that node), then $E[X|A]$ = avg value of leaves below A.

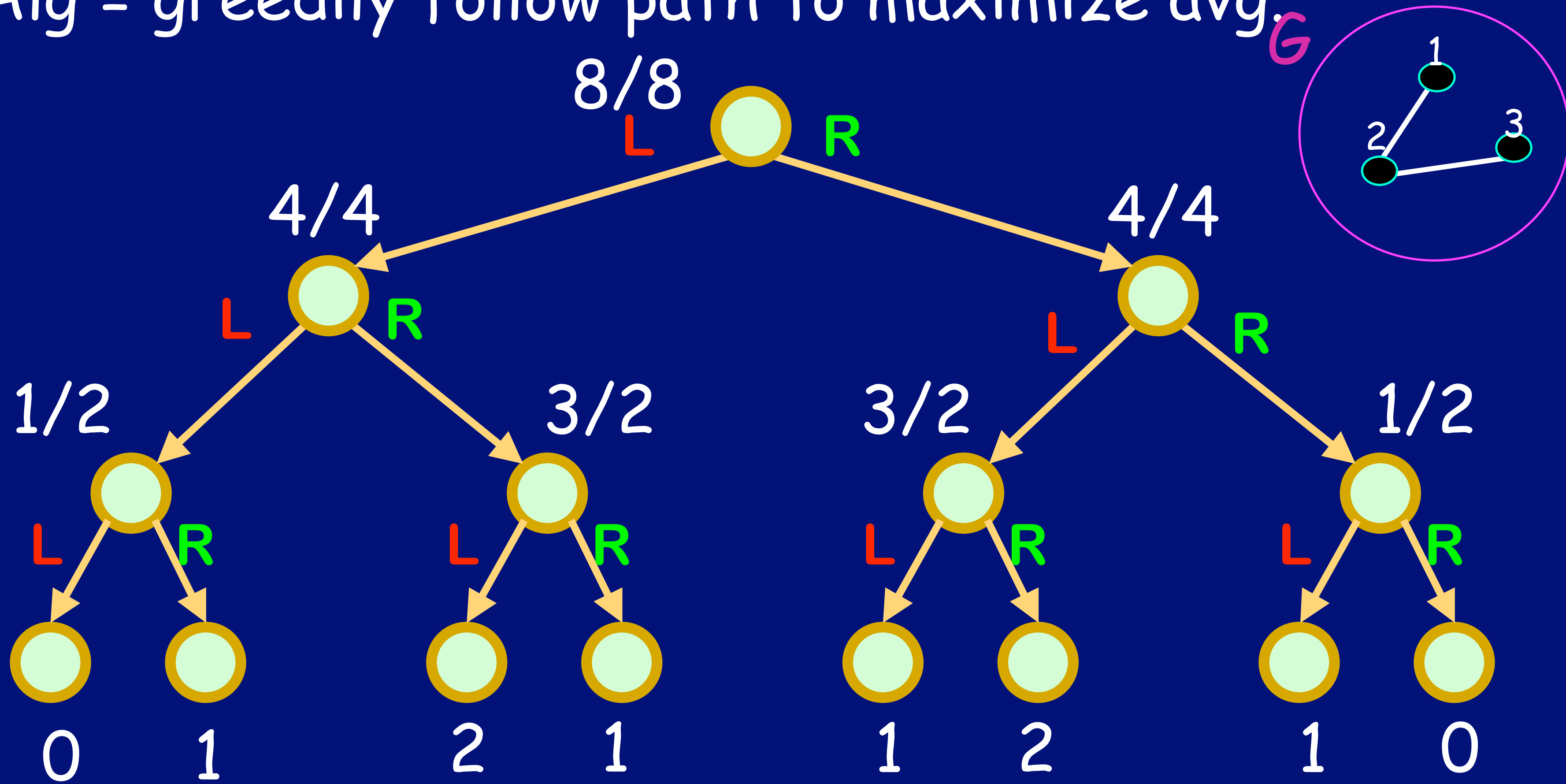
Alg = greedily follow path to maximize avg.



Pictorial view (important!)

If A is some node (the event that we reach that node), then $E[X|A]$ = avg value of leaves below A.

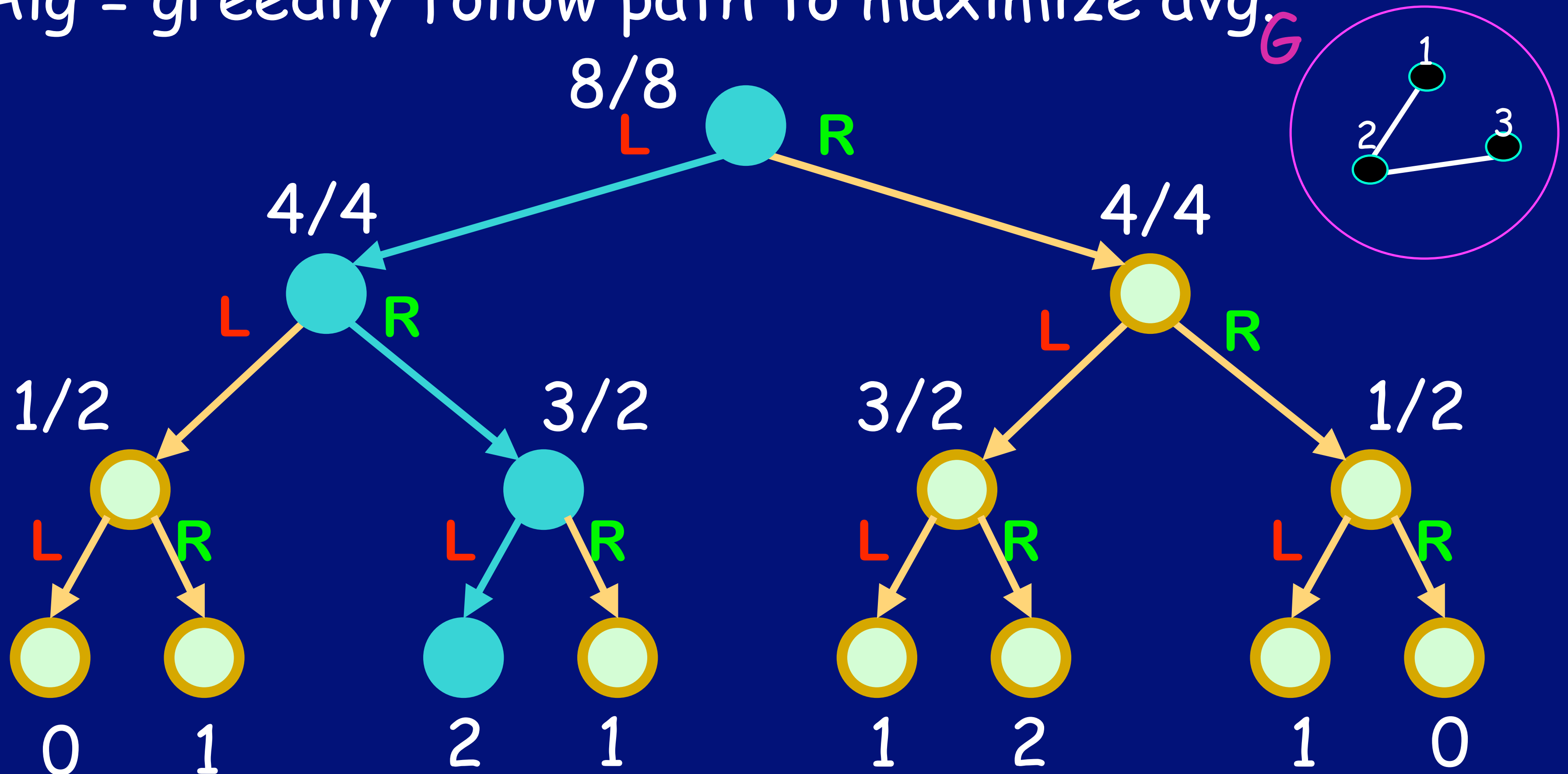
Alg = greedily follow path to maximize avg.



Pictorial view (important!)

If A is some node (the event that we reach that node), then $E[X|A]$ = avg value of leaves below A.

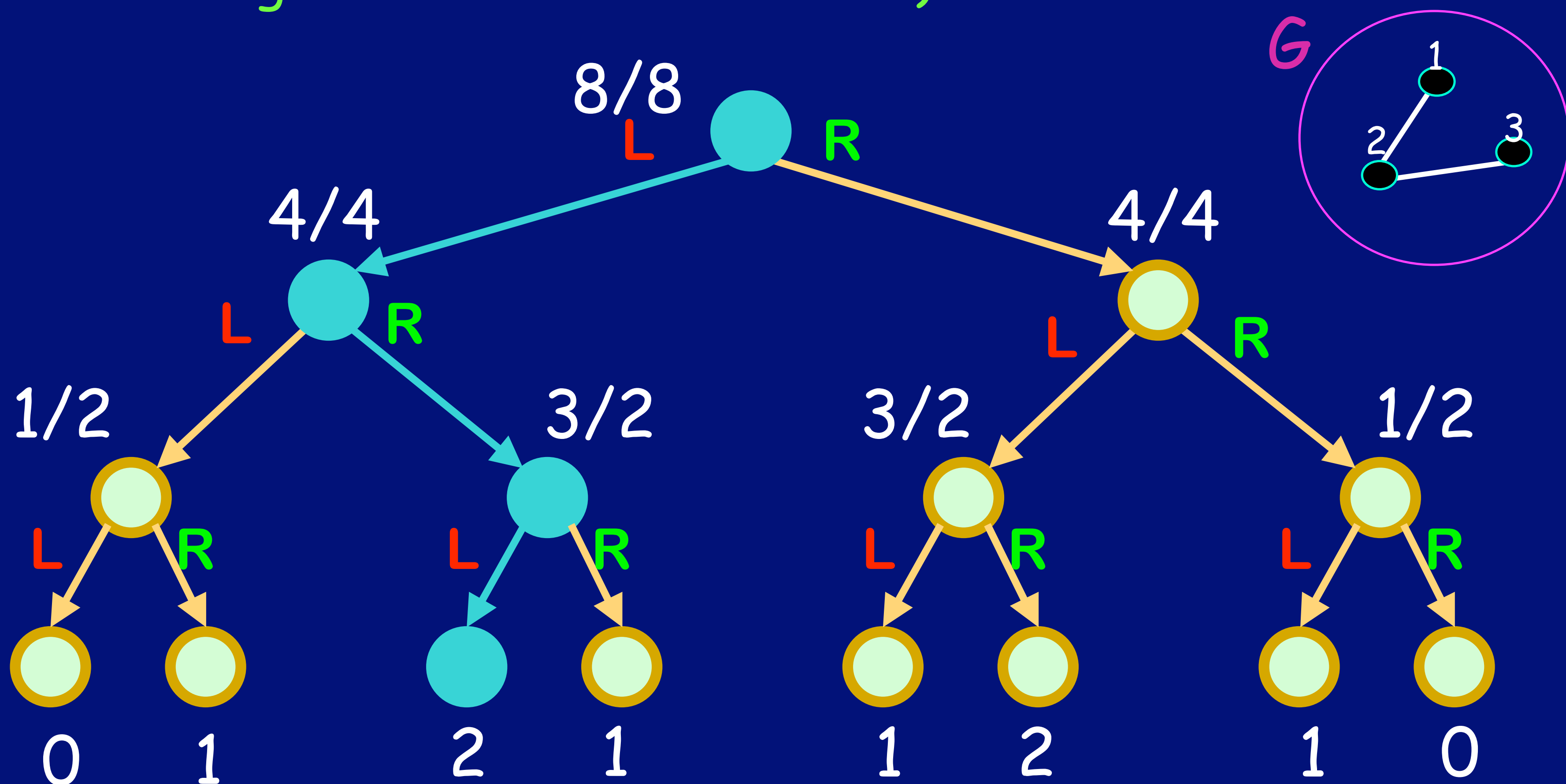
Alg = greedily follow path to maximize avg.



Pictorial view (important!)

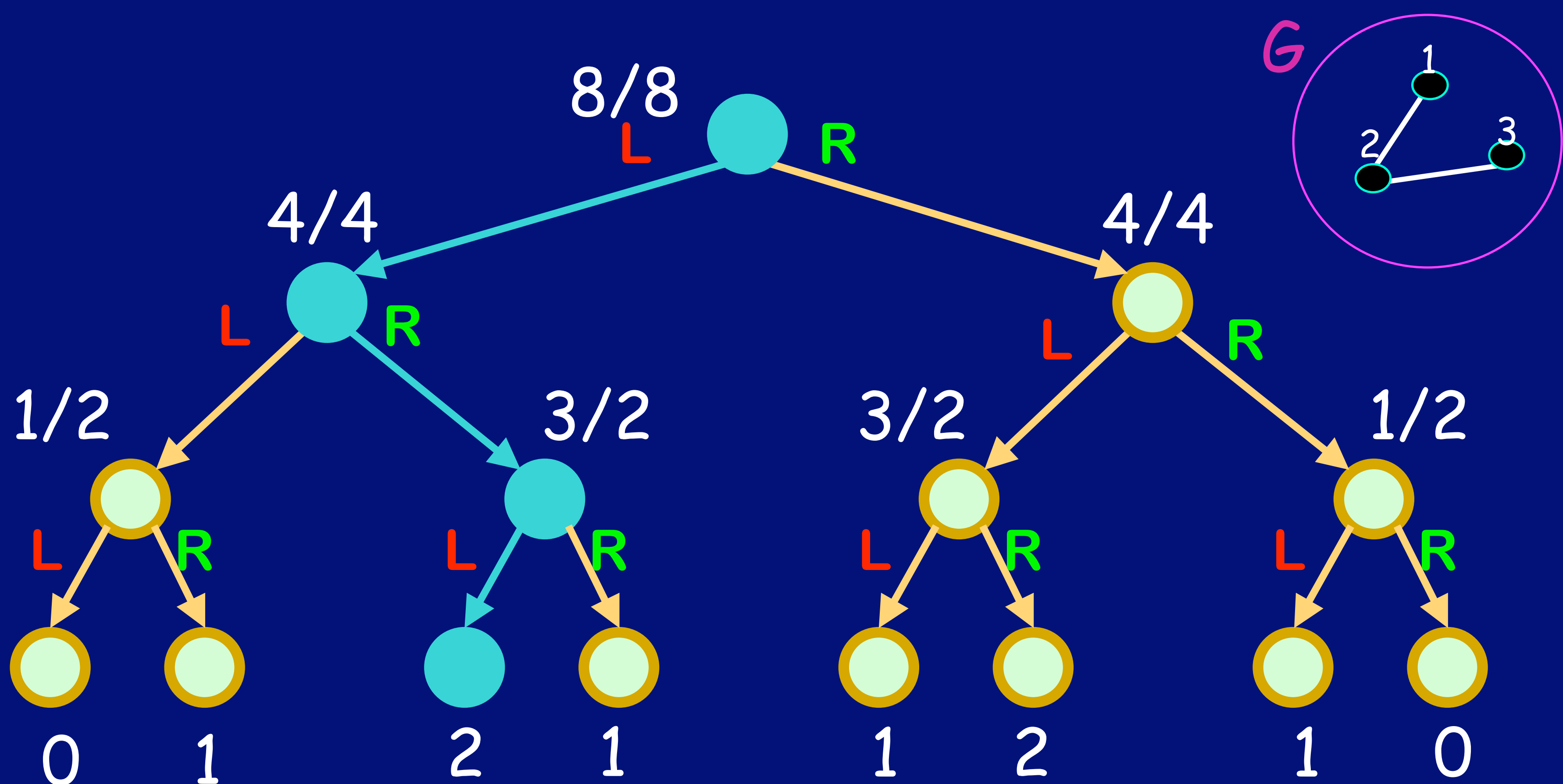
Linearity of expectation gives us a way of magically computing E[X|A] for any node A.

(Even though the tree has 2^n leaves)



Pictorial view (important!)

In particular, E[X|A] = (# edges crossing so far) + (# edges not yet determined)/2

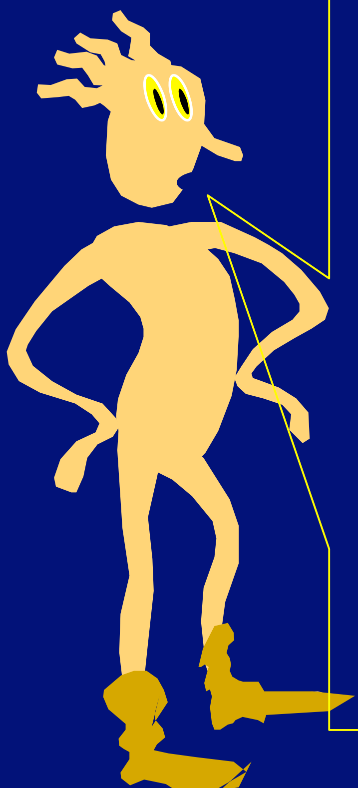


Conditional expectation method

In fact, our algorithm is just: put vertex i into the side that has the fewest of its neighbors so far.

(The side that causes the most of the edges determined so far to cross the cut).

But the probabilistic view was useful for proving that this works!



Often, though, we can't get an exact handle on these expectations. The probabilistic method can give us proof of existence without an algorithm for finding the thing.

In many cases, no efficient algorithms for finding the desired objects are known!

Constraint Satisfaction

Is there an assignment to Boolean variables $X_1, X_2, \ldots, X_5$ that makes this formula true?

$$\overbrace{(X_1 \text{ or } !X_2 \text{ or } X_4)}^{\text{clause}} \text{ and } (!X_3 \text{ or } \overbrace{!X_4}^{\text{literal}} \text{ or } X_5) \text{ and } (X_2 \text{ or } X_3 \text{ or } X_4)$$

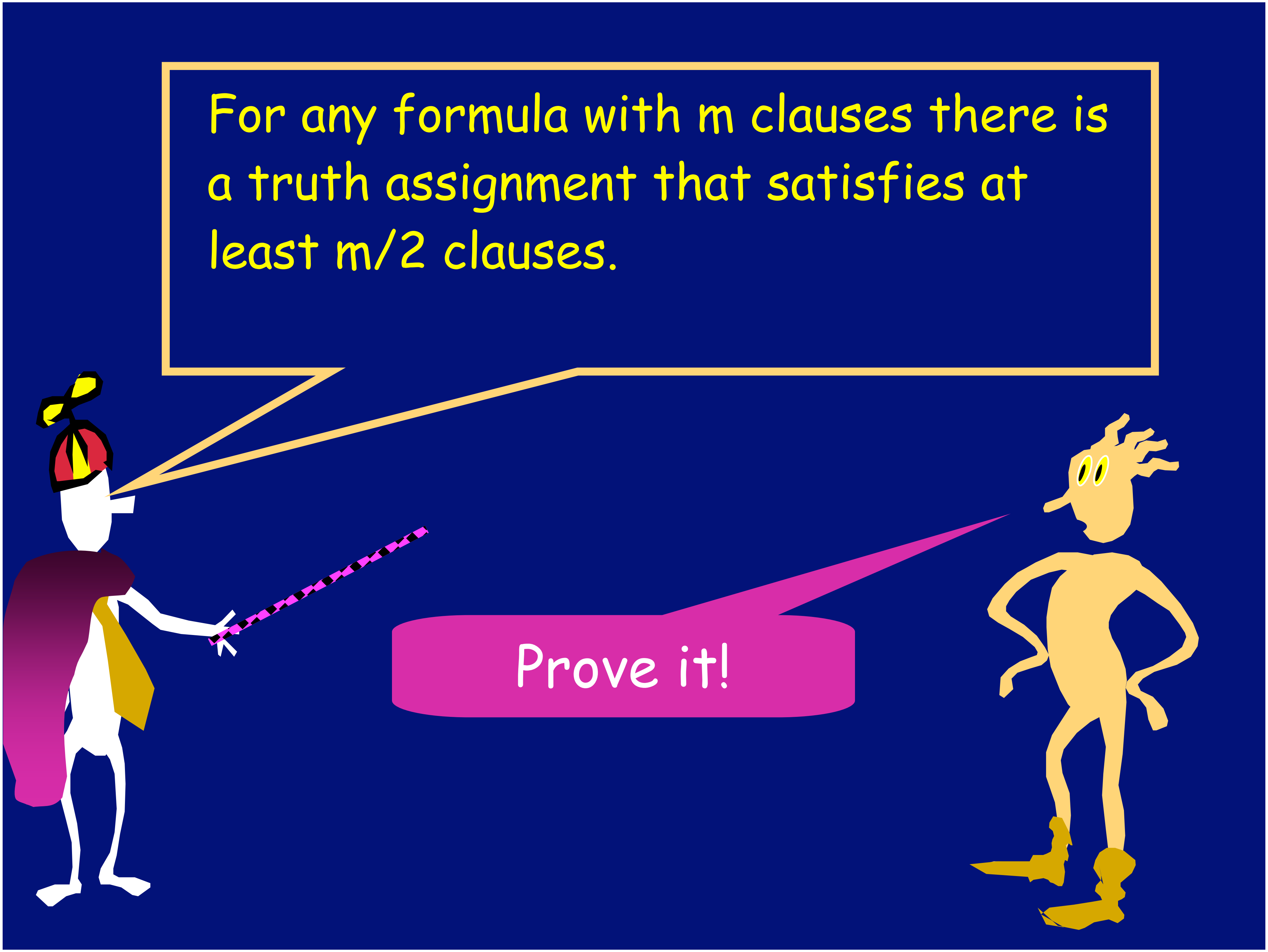
Clause is *satisfied* if at least one of its literals is true.

X_1=T, X_2=T, X_3=F, X_4=F, X_5=F

Constraint Satisfaction

In general, it's difficult to determine if a formula is satisfiable, or to maximize the number of satisfying clauses (more on this later…)

$(X_1 \text{ or } !X_2 \text{ or } X_4) \text{ and } (!X_3 \text{ or } !X_4 \text{ or } X_5) \text{ and } (X_2 \text{ or } X_3 \text{ or } X_4)$


For any formula with m clauses there is a truth assignment that satisfies at least m/2 clauses.
Prove it!

For any formula with m clauses there is a truth assignment that satisfies at least m/2 clauses.

- Make a random coin flip to decide if each variable should be T or F.

- Let Z_i=1 if the i^{th} clause is satisfied and Z_i=0 otherwise. If a clause has k vars, the chance it is *not* satisfied by this random assignment is $(1/2)^k$ (all literals must be F)

- So, the chance it *is* satisfied is $1-(1/2)^k = E[Z_i] \geq 1/2$.

- Therefore, $E[Z_1+Z_2+\ldots + Z_m] \geq m/2$

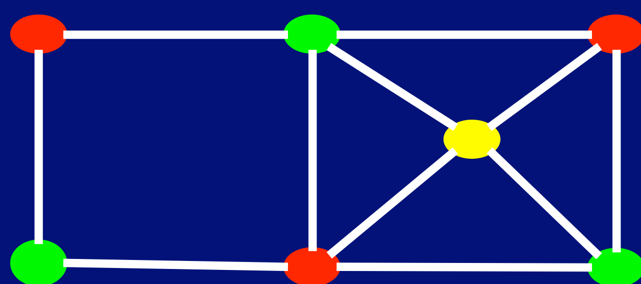
For any formula with m clauses there is a truth assignment that satisfies at least m/2 clauses.

If each clause has k literals, there must be an assignment that satisfies at least $1-(1/2)^k$ clauses (can use C.E. method to find such an assignment).

Independent Sets

An *independent set* in a graph is a set of vertices with no edges between them.

Note: a coloring of a graph divides it into independent sets.



So, a k-colorable graph on n vertices must have an independent set of size at least n/k (letterbox).

Theorem: If a graph G has n vertices and e edges, then it has an independent set with at least $n^2/4e$ vertices.

Let $d = 2e/n$ be the average degree.
Randomly take away vertices and edges:

1. Delete each vertex of G (together with its incident edges) with probability $1-1/d$

2. For each remaining edge remove it and one of its vertices.

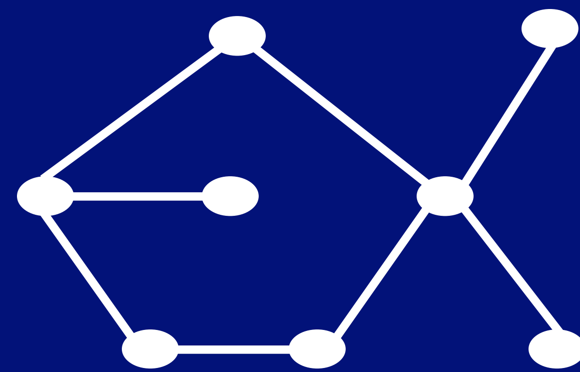
The remaining vertices form an independent set. How big is it expected to be?

Example

$n = 8$

$e = 8$

$d = 2e/n = 2$



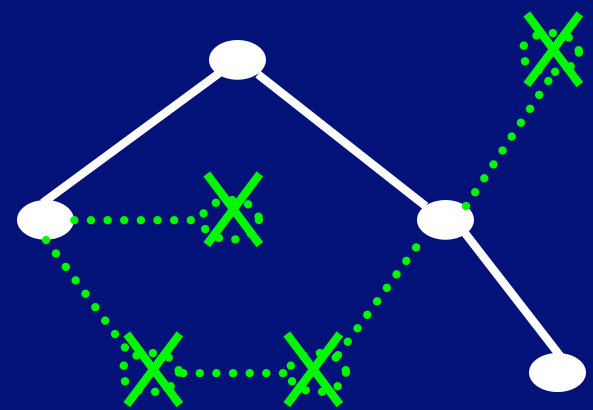
1. Delete each vertex of G (together with its incident edges) with probability $1-1/d$ = 1/2

Example

$n = 8$

$e = 8$

$d = 2e/n = 2$



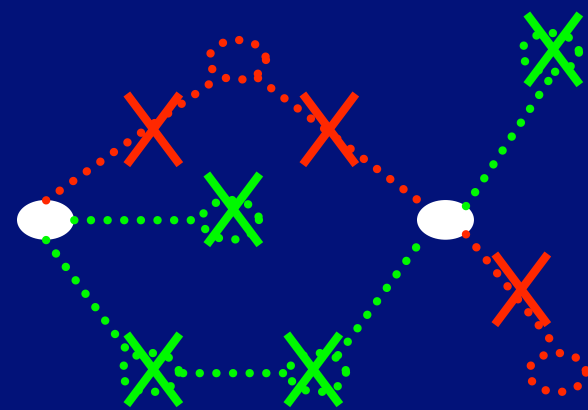
1. Delete each vertex of G (together with its incident edges) with probability $1-1/d = 1/2$
2. For each remaining edge remove it and one of its vertices.

Example

$n = 8$

$e = 8$

$d = 2e/n = 2$



1. Delete each vertex of G (together with its incident edges) with probability $1\text{-}1/d = 1/2$
2. For each remaining edge remove it and one of its vertices.

Theorem: If a graph G has n vertices and e edges, then it has an independent set with at least $n^2/4e$ vertices.

Let X be the number of *vertices* that survive step 1:

$$E[X] = n/d.$$

Let Y be the number of *edges* that survive step 1:

$$E[Y] = e(1/d)^2 = nd/2\ (1/d)^2 = n/2d.$$

Step 2 removes one at most Y vertices (one per surviving edge), so we're left with at least X-Y vertices.

$$E[X-Y] = n/d - n/2d = n/2d = n^2/4e$$

The Probabilistic Method

- what it is
- how to use it to prove theorems

Conditional expectation

- definition
- partitioning identity