CS5314
Randomized Algorithms

Lecture 18: Probabilistic Method
(De-randomization, Sample-and-Modify)

Objectives

- Introduce two topics:

  De-randomize by conditional expectation

  - provides a deterministic way to construct an object with some property

  Sample-and-modify

  - a more advanced technique to prove the existence of a certain object

De-randomization
(using Conditional Expectation)

- Let us revisit the large-cut problem
- Recall that this problem is NP-hard
- We have shown that there exists a cut of size at least m/2, and whose size is thus at least half of the largest cut

Question: Can we construct such an approximate large-cut explicitly?

Finding Large-Cut

- Let $G=(V, E)$ be a graph with n vertices and m edges
- Let $v_1, v_2, \ldots, v_n$ be the vertices in V
- Recall that if we partition V by placing each vertex randomly and independently into A and B, the expected size of the cut(A,B) is at least $m/2$

➔ Without loss of generality, we assume that v_1 is placed in A

Finding Large-Cut (2)

Consider the two cases how v_2 is placed

Knowing v_1 is in A, we can actually determine

E[size of cut(A,B) | v_2 is in A] , and

E[size of cut(A,B) | v_2 is in B] (how?)

Questions:

Can we make use of these values to decide where to place v_2 ?

Finding Large-Cut (3)

Observation:

If our target is to find a cut whose size is at least E[size of cut(A,B)], it cannot be wrong to place v_2 in the set whose corresponding expectation is larger

Proof: WLOG, suppose that

$$E[\text{size of cut}(A,B) \mid v_2 \text{ is in } B] \geq E[\text{size of cut}(A,B) \mid v_2 \text{ is in } A]$$

Proof

Then,

E[size of cut(A,B)]

= E[size of cut(A,B) | v_2 in A] Pr(v_2 in A) +
E[size of cut(A,B) | v_2 in B] Pr(v_2 in B)

$\leq$ E[size of cut(A,B) | v_2 is in B]

➔ Placing v_2 in B ensures at least one assignment of remaining vertices will have cut-size $\geq$ E[size of cut(A,B)]

Finding Large-Cut (4)

- Knowing v_1 is in A, and which set (say X_2) is better to place v_2, we can determine

  E[size of cut(A,B) | v_2 in X_2, v_3 in A] and

  E[size of cut(A,B) | v_2 in X_2, v_3 in B]

Question:

Can we make use of these values to decide where to place v_3 ?

Finding Large-Cut (5)

Ans. Definitely Yes !!

➔ We should place v_3 in the set that maximizes the above conditional expectation, because it ensures that one assignment of remaining vertices has cut-size ≥ E[size of cut(A,B)|v_2 in X_2]

➔ By our choice of X_2, such an assignment has cut-size ≥ E[size of cut(A,B)]

Finding Large-Cut (6)

- Continue the above process, we can decide where to place $v_2, v_3, \ldots, v_n$
- Let $X_2, X_3, \ldots, X_n$ be the corresponding set each are placed

Then, we must have

$$E[\text{size of cut } (A,B)]$$
$$\leq E[\text{size of cut } (A,B) \mid v_j \text{ in } X_j \text{ for } j=2,\ldots, n]$$

Finding Large-Cut (7)

Since

$$m/2 \leq E[\text{size of cut}(A,B)]$$

and

$$E[\text{size of cut}(A,B) \mid v_j \text{ in } X_j \text{ for } j=2,\ldots,n]$$

$$= \text{size of cut } (A,B)$$

when v_j is in X_j for $j = 2,\ldots, n$

➔ We obtain a cut (deterministically) whose size is at least $m/2$!!!

Sample-and-Modify

In Basic counting/Conditional expectation :

- we construct some probability space
- show that an object with some desired properties can be picked directly

In Sample-and-Modify :

- we first select an object randomly, but it may not have the desired properties
- then we modify the object to get the desired properties

Independent Set

Definition: An independent set of a graph G is a set of vertices with no edges between them

- Finding an independent set with largest number of vertices is NP-hard
- Let's use Sample-and-Modify to obtain a lower bound on size of the largest independent set

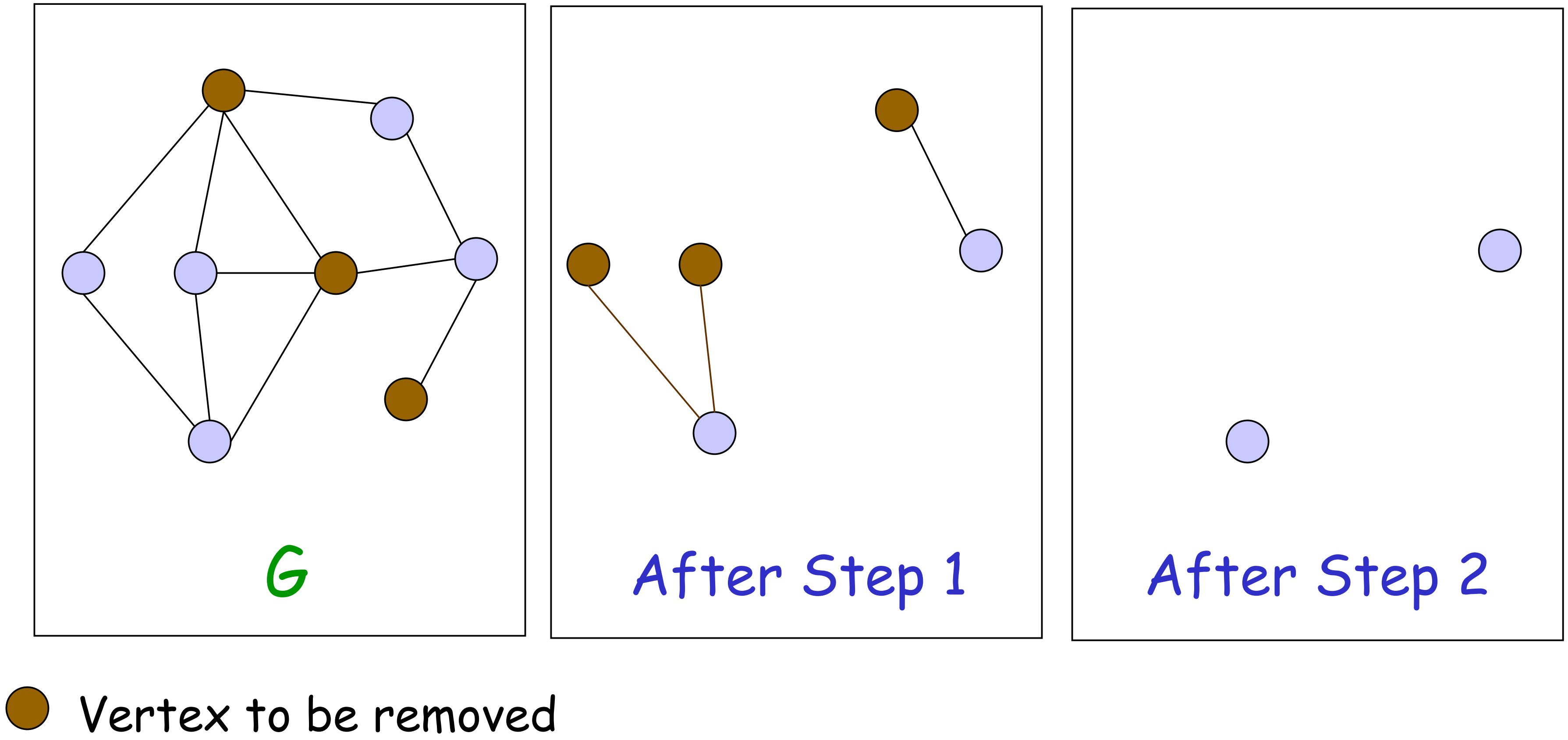
Independent Set (2)

Theorem: Let G be a graph with n vertices and m edges. Then G has an independent set with at least $n^2/(4m)$ vertices

Proof: Let $d = 2m/n$ = ave degree (wlog, assume each connected component has ≥ 3 vertices, so $d > 1$)

Consider the following:

1. Delete each vertex (and its incident edges) independently with probability $1 - 1/d$
2. For each remaining edge, remove it by removing randomly one of its vertex



- After Step 1 : may not be independent
- After Step 2 : must be independent (why?)

Proof (cont)

- Let X = # vertices after Step 1
- Let Y = # edges after Step 1
- Let Z = # vertices after Step 2

Target: Can we bound $E[Z]$?

Firstly,

$$E[X] = n/d, \quad E[Y] = m/d^2 = n/(2d)$$

Proof (cont)

Then, we have

$$Z \geq X - Y$$ ... (why not $Z = X - Y$?)

So, $E[Z] \geq E[X-Y]$

$= E[X] - E[Y]$

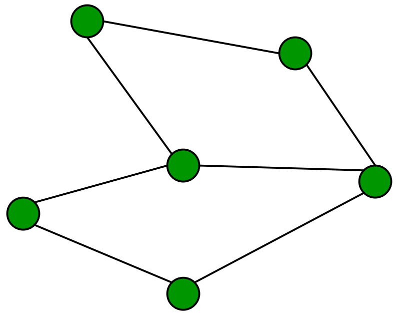
$= n/(2d) = n^2/(4m)$

➔ Theorem follows from expectation argument

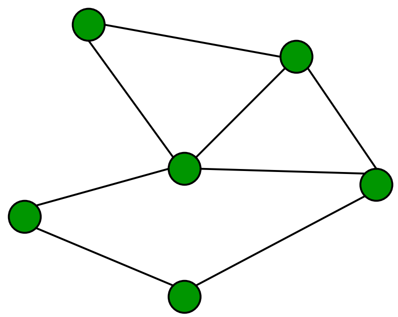
Graphs with Large Girth

Definition: The girth of a (simple) graph G is the length of the smallest cycle

E.g.,



girth = 4



girth = 3

Graphs with Large Girth (2)

- Intuitively, we expect graphs with large average degree to have small girth
- However, this is not necessarily true …

We start by showing the following theorem:

Theorem: For any integer $k \geq 3$, there is a graph with n vertices, at least $n^{1+1/k}/4$ edges, and girth at least k

How to prove?

Proof

Consider the following algorithm:

1. Set $p = n^{1/k-1}$
2. Select a graph G randomly from $G_{n,p}$
3. For any cycle that appears in G with length less than k, remove one edge randomly in that cycle

➔ the graph obtained after Step 3 have girth at least k

(Did you see that we are using the Sample-and-Modify?)

Proof (2)

It remains to find # edges in the final graph

First, let X = # edges in G

➔ $E[X] = n(n-1)p/2$

$= (1/2)\, n^{1+1/k}(1-1/n) \geq (1/3)\, n^{1+1/k}$

Next, let Y = # cycles in G with length $< k$

- observe that any specific cycle of length j occurs with probability p^j

Proof (3)

Since possible # length-j cycle

$= n(n-1)(n-2)\ldots(n-j+1)/(2j)$

➔ $E[Y] = \sum_{j=3 \text{ to } k-1} p^j n(n-1)(n-2)\ldots(n-j+1)/(2j)$

$\leq \sum_{j=3 \text{ to } k-1} p^j n^j$

$= \sum_{j=3 \text{ to } k-1} n^{j/k} < \sum_{j=3 \text{ to } k-1} n^{(k-1)/k}$

$< kn^{(k-1)/k} \leq n$ for large enough n

Proof (4)

Let Z = # edges after Step 3

Then, we have $Z \geq X - Y$ ... (why not $Z = X-Y$?)

So, $E[Z] \geq E[X-Y]$

$= E[X] - E[Y]$

$> (1/3)\, n^{1+1/k} - n$ for large enough n

$> (1/4)\, n^{1+1/k}$ for large enough n

➔ Theorem follows

Graphs with Large Girth (3)

- The previous theorem immediately gives the following corollary:

Corollary: For any integer $k \geq 3$ and any positive real d representing the average degree, there is a graph with n vertices, at least $nd/2$ edges, and girth at least k

Proof: Choose sufficiently large n such that

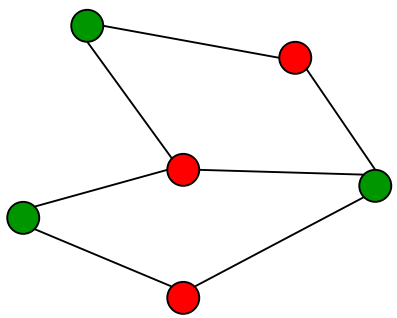
$$n^{1/k} > 2d$$

Graphs with Large Girth (4)

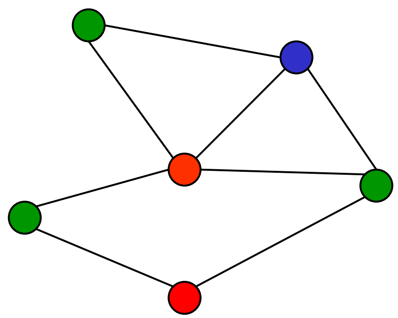
A related problem is as follows:

Let G be a (simple) graph

Definition: The chromatic number, $\chi(G)$, is the minimum # colors needed to color vertices of G such that no adjacent vertices have the same color



$\chi(G) = 2$



$\chi(G) = 3$

Large $\chi(G)$ and Large Girth

- Intuitively, we expect graphs with large chromatic number to have small girth
- However, this is not necessarily true…

The theorem below is due to Erdős (1959):

Theorem: For all $k \geq 3$ and positive c, there is a graph with $\chi(G)$ at least c and girth at least k

How to prove?

Proof

Consider the following algorithm:

1. Set $p = n^{1/k-1}$
2. Select a graph G randomly from $G_{n,p}$
3. For any cycle that appears in G with length less than k, remove one **vertex** randomly in that cycle

➔ the graph obtained after Step 3 have girth at least k

Proof (2)

Let Y = # cycles in G with length less than k

As shown before:

$$E[Y] = \sum_{j=3 \text{ to } k-1} p^j \, n(n-1)(n-2)\ldots(n-j+1)/(2j)$$

$$< kn^{(k-1)/k} = o(n) \quad \text{for large enough } n$$

Then, by Markov inequality,

$$\Pr(Y \geq n/2) = o(1) \qquad \text{for large enough } n$$

Proof (3)

- Next, we investigate the size of the largest independent set in G, as this will be related to the chromatic number

Let A = size of largest independent set in G

- For any x, if $A \geq x$, there must exists a group of x vertices such that no edges are between them
- By union bound, we have

$$\Pr(A \geq x) \leq C(n,x)\,(1-p)^{x(x-1)/2}$$

Proof (4)

When x is slightly large, say $\lceil (3 \ln n) / p \rceil$,

$$\Pr(A \geq x) \leq C(n,x)\,(1-p)^{x(x-1)/2}$$

$$\leq n^x\,(1-p)^{x(x-1)/2}$$

$$\leq n^x\,e^{-px(x-1)/2} = (n\,e^{-p(x-1)/2})^x$$

$$\leq (n\,e^{-px/2.5})^x \qquad \text{for large enough } n$$

$$\leq (n^{-0.2})^x = o(1)$$

Proof (5)

Then, for large enough n,

$\quad$ we have $\Pr(Y \geq n/2) \leq 0.1$

$\quad$ and $\Pr(A \geq \lceil (3 \ln n)/p \rceil) \leq 0.1$

This implies that, for large enough n

$\quad$ $\Pr(\,(Y < n/2) \cap (A < \lceil (3 \ln n)/p \rceil)\,)$

$\quad$ $\geq 0.8 > 0$ … (why?)

Proof (6)

That is, for large enough n,

we can select a graph G with less than $n/2$ cycles of "short" lengths, and

whose largest independent set is at most $3 \ln n / p$

➔ After removing one vertex from each "short" cycle in this graph G, # vertices in the resulting graph (after Step 3) $\geq n/2$

(What happens to the largest independent set?)

Proof (7)

Let G^* = resulting graph after Step 3

Let $A(G^*)$ = size of largest indep set of G^*

Obviously, $A(G^*) \leq A(G) \leq 3 \ln n / p$ ... (why?)

Also, $A(G^*) \geq |G^*| / \chi(G^*)$ ... (why?)

Combining, we get

$$\chi(G^*) \geq |G^*| / A(G^*) \geq (n/2) / (3 \ln n/p)$$

$$= n^{1/k}/(6 \ln n) \geq c \quad \text{for large enough } n$$

➔ G^* exists with girth $\geq k$, and $\chi(G^*) \geq c$