Today
- bit more on random walks on graphs
- Monte Carlo methods & approx counting

Random walks on undirected graphs

$G=(V,E)$ undirected graph

consider simple random walk on this graph:

$p_{ij} = \frac{1}{d_i} \quad \forall (i,j) \in E$

[MC is periodic iff graph is bipartite]
[but if so, consider lazy r.w.]

$$\boxed{\pi_i = \frac{d_i}{2m} \qquad m: \# \text{ of edges}}$$

$\sum_i \frac{d_i}{2m} = 1$

$\pi_j = \sum_i \pi_i p_{ij}$

$\frac{d_j}{2m} = \sum_{\substack{i \text{ s.t.} \\ (i,j)\in E}} \frac{d_i}{2m} \cdot \frac{1}{d_i}$

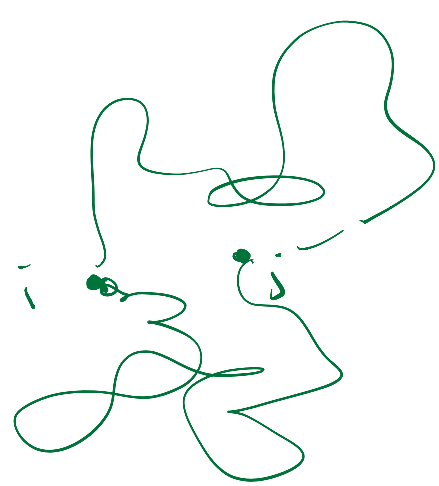
Some key quantities:

hitting time	$h_{ij} = E(T_{ij})$
commute time	$c_{ij} = h_{ij} + h_{ji}$
covertime	$C(G) =$ exp. time to visit all vertices

$C(G)$

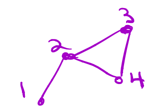
$h_{ii} = \frac{1}{\pi_i} = \frac{2m}{d_i}$

Lemma $\forall$ edge (i,j) $\quad h_{ij} + h_{ji} \le 2m$

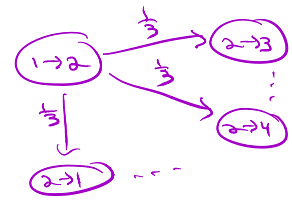


Pf: Consider corresponding random walk on directed edges (states are $i \to j$, $j \to i$ $\forall (i,j) \in E$) $\quad$ 2m states

transition probabilities: $q_{(i \to j,\, \ell \to r)} = \begin{cases} \frac{1}{d_j} & \ell = j \\ 0 & \text{o.w.} \end{cases}$



state at time t



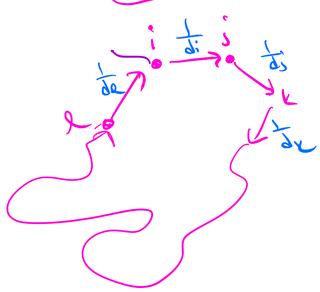
$Q = \left(q_{\vec{e}, \vec{e}'} \right)$ is doubly stochastic

col sum for $i \to j$ $= \sum_{\substack{k \text{ s.t.} \\ (k,i) \in E}} q_{k \to i,\, i \to j} = \sum_{\substack{k \text{ s.t.} \\ (k,i) \in E}} \frac{1}{d_i} = 1$

D.S. $\Rightarrow$ π is uniform, i.e. $\pi_{\vec{e}} = \frac{1}{2m}$ $\forall \vec{e}$

$\Rightarrow h_{i \to j,\, i \to j} = 2m$

$\boxed{h_{ij} + h_{ji}}$ in original r.w. $\leq h_{i \to j,\, i \to j} = 2m$ in other walk



(j) k, ..., i j

j k l ... i ... j i ... i j

Corollary

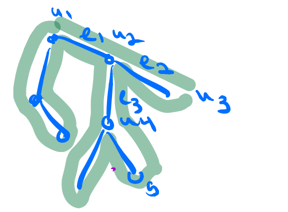
$\forall\ G=(V,E)$ connected nonbipartite (or lazy) $\quad C(G)$ = expected cover time of random walk on G $\leq \underline{2m(n-1)}$ $\quad O(mn)$

Pf: Let T be a spanning tree on G and let $e_1, \dots, e_{n-1}$ be the edges in the tree.
$(u_1,v_1) \quad (u_n, v_n)$

Consider doubled tree T^*, where each edge is duplicated once in each direction.

Every vertex has indegree = outdegree
$\Rightarrow$ has Euler tour say

$\vec{e}_1 \ \vec{e}_2 \ \dots \ \overleftarrow{e}_1 \ \dots$
$u_1 \to v_1 \quad v_1 \to v_2 \quad \dots \quad v_1 \to u_1 \dots$

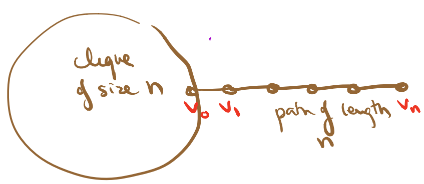


$u_1\ u_2\ v_3\ u_2\ u_4\ u_5\ u_4 \dots u_4\ u_2$

$$E(\text{cover time}) \leq \sum_{i=1}^{n-1} E(T_{u_i v_i} + T_{v_i u_i}) = (n-1)2m$$

Examples

$C(G) = \Theta(n^2)$



$m = O(n^2)$

$C(G) = O(n^3)$

Fact: Starting from v_1, $\Pr(\text{reach } v_n \text{ before returning to } v_0) = \frac{1}{n}$

$\frac{n-1}{n}$

bound above $\quad C(G) = O(n^3)$

$O(n \log n)$ — coupon collectors.

$\frac{n-i}{n} \qquad E(T \text{ to see new}) = \frac{n}{n-i}$

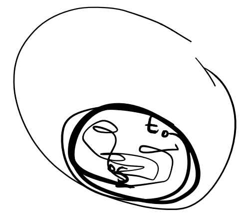
#balls $\rightarrow$ n bins till every bin has a ball.

Application: s-t connectivity

Given G undirected graph $s,t \in V$

determine if s & t are in same CC.

DFS / BFS $\left\{ \begin{array}{l} O(m) \text{ time} \\ O(n) \text{ space} \end{array} \right.$ keep track of all vertices search has visited so far



Observation: very simple randomized alg using log space

(input on separate read-only tape)

Simulate r.w. of length $4mn$ on G starting from s

$\Pr(\text{r.w. doesn't reach } t \text{ when } \exists \text{ path}) \le \frac{1}{2}$

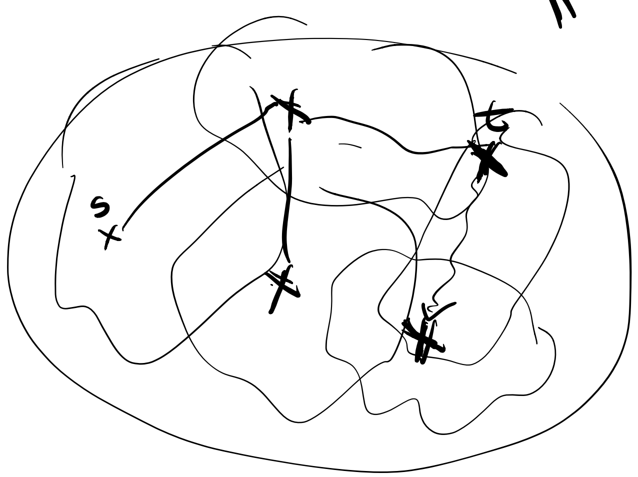
logarithmic factors being ignored

Both algorithms $\tilde{O}(mn)$ space·time

Can we interpolate?

$\tilde{O}(p)$ space $\quad \tilde{O}\left(\frac{mn}{p}\right)$ time?

Feige



random walks $\longleftrightarrow$ electrical networks.

Monte Carlo Methods

collection of tools for estimating values thru sampling & estimation

$\mu = \frac{\text{area } A}{\text{area of square}}$.

(ε, δ) Approximation

A randomized alg gives an (ε,δ) approx for value V if the output X of the alg satisfies

$$\Pr(|X - V| > \varepsilon |V|) \le \delta$$

Example

Sample indep random vars whose mean is quantity we want to estimate

Let $X_1, X_2, \ldots, X_m$ iid. Bernoulli with $E(X_i) = \mu$

If $m \ge \frac{3 \ln(\frac{2}{\delta})}{\varepsilon^2 \mu}$ then

$$\Pr\left(\left| \frac{1}{m} \sum_{i=1}^{m} X_i - \mu \right| \ge \varepsilon \mu \right) \le \delta$$

Pf: Chernoff bounds

Monte Carlo Thm

DNF Counting

Suppose want to know # satisfying assignments

$$(\bar{x}_1 \wedge x_2) \vee (x_2 \wedge x_3) \vee (x_1 \wedge x_2 \wedge \bar{x}_3 \wedge x_4) \vee (x_3 \wedge \bar{x}_4)$$

$(x_1 \vee x_2 \vee \bar{x}_3) \wedge \ldots$

$(\bar{x}_1 \wedge \bar{x}_2 \wedge x_3) \vee \ldots$

Obviously satisfying such a formula is easy

Counting # satisfying assignments is hard

If could do this, could solve $\cancel{X}$SAT

CNF formula φ $\xrightarrow{\text{DeMorgan}}$ DNF $\bar{\varphi}$

$\downarrow$

count # solns

If count $= 2^n$ then φ unsat $\leftarrow$

o.w. φ sat.

Problem actually #P-complete $\longrightarrow$ strong intractability

#P is counting analogue of NP

problems of form:
compute $f(x)$ where $f(x)$ is # solutions to problem x in NP

counting #sat assignments DNF formula   #P complete

Ham. cycles

matchings in bipartite graph.

P=NP

Approximate DNF Counting?

Obvious approach: sample random assignments, indep m times

$X_i = \begin{cases} 1 & \text{i}^{\text{th}} \text{ random assignment satisfies } \varphi \\ 0 & \text{o.w.} \end{cases}$

return $\frac{\sum_{i=1}^{m} X_i}{m} \cdot 2^n$ as estimate for # satisfying assignments

for (ε, δ) estimate need $m = \Omega\left(\frac{\ln(\frac{2}{\delta})}{\varepsilon^2 \mu}\right)$

$\uparrow$ fraction of assign satisfying.

μ could be exponentially small $\frac{n^2}{2^n}$

finding needle in a haystack. $n \quad n^2 \quad n^3$

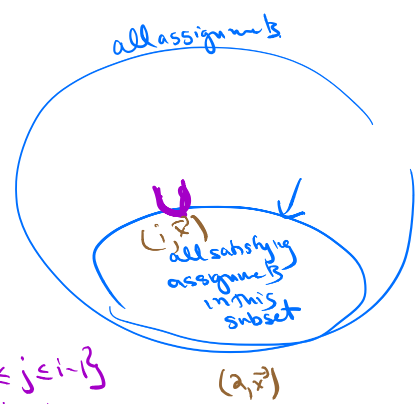
$\boxed{\exists \text{ FPRAS for DNF counting}}$

$\uparrow$ Fully polynomial randomized approx scheme

aka: a randomized alg for which, given an input x and any parameters ε & δ with $0 < \varepsilon, \delta < 1$, the alg outputs an (ε, δ) approx to $V(x)$ in time poly in $\frac{1}{\varepsilon}$, $\ln \frac{1}{\delta}$ & size of input

$\to$ in this example x is DNF formula, $V(x)$ is # satisfying assignments

$$\boxed{\bar{V} \text{ is } (\varepsilon, \delta) \text{ approx to } V \equiv \Pr\left(|\bar{V} - V| > \varepsilon V\right) \le \delta}$$



$\varphi = C_1 \vee C_2 \vee \dots \vee C_t$

SC_i : set of assignments satisfying clause i

S = set of all satisfying assignments
$= \{(i, \vec{x}) \mid \vec{x} \in SC_i, \ \vec{x} \notin SC_j \ \ 1 \le j \le i-1\}$ $\quad 1 \le i \le t$

$C_1 \vee C_2 \vee C_3 \vee C_4$

$(2, \vec{x}) \ (4, \vec{x})$

$U = \{(i, \vec{x}) \mid \vec{x} \in SC_i \quad 1 \le i \le t\}$

estimate $|S|$ by approximating $\frac{|S|}{|U|}$

M times, sample (i,x) from U & then check if its in S

$$|U| = \sum_{i=1}^{t} |SC_i|$$

$2^{\#\text{vars not in } C_i}$

$$\frac{|S|}{|U|} \geq \frac{1}{t}$$

$$M = \frac{3t \ln\left(\frac{2}{\delta}\right)}{\varepsilon^2} \text{ samples}$$

(ε, δ) approx

pick i with prob $\frac{|SC_i|}{\sum_j |SC_j|}$

setvars outside SC_i u.a.r

$$\Pr\left((i,x) \text{ selected}\right) = \frac{\cancel{|SC_i|}}{\sum_j |SC_j|} \cdot \frac{1}{\cancel{|SC_i|}} = \frac{1}{|U|}$$

DNF counting illustrates fundamental connection

sampling $\leftrightarrow$ counting

From approx sampling $\longrightarrow$ approx counting

$S \subseteq V$ I.S. if $\forall\, i,j \in S \;\; (i,j) \notin E$

Counting Independent Sets

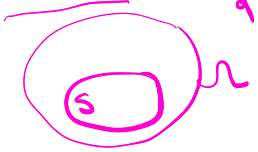
input $G=(V,E)$

output: estimate of $|\Omega(G)|$ # indep sets in graph.

Input: x, ε $\uparrow$ $G=(V,E)$ $\longrightarrow$ $\boxed{\text{FPAUS: Fully polynomial almost uniform sampler}}$ $\longrightarrow$ ε uniform sample w from $\underline{\Omega(x)}$ all IS's of G

runs in time polynomial in size of x & $\ln(\frac{1}{\varepsilon})$

$\forall S \quad \left| \Pr(w \in S) - \frac{|S|}{|\Omega|} \right| \le \varepsilon$



Want to show:

Given FPAUS for indep sets $\Longrightarrow$ can construct FPRAS for ISs

$G=(V,E)$, ε $\longrightarrow$ $\boxed{\text{FPRAS}}$ $\longrightarrow$ (ε, δ) approx to $|\Omega(G)|$

makes sequence of calls to FPAUS(x_1, ε_1) (x_2, ε_2) ...

runs in time poly in $\frac{1}{\varepsilon}$, size of input.

$G=(V,E)$ $\quad e_1, e_2, \ldots, e_m$ arbitrary ordering of edges

$E_i = \{e_1, e_2, \ldots, e_i\}$ $\quad G_i = (V, E_i)$ $\quad G_m = G$ $\quad G_0 = (V, \phi)$

$\Omega(G_i)$: set of ISs in G_i

$$|\Omega(G)| = \frac{|\Omega(G_m)|}{|\Omega(G_{m-1})|} \times \frac{|\Omega(G_{m-1})|}{|\Omega(G_{m-2})|} \times \cdots \times \frac{|\Omega(G_1)|}{|\Omega(G_0)|} \times \underbrace{|\Omega(G_0)|}_{2^n}$$

Let $r_i = \dfrac{|\Omega(G_i)|}{|\Omega(G_{i-1})|}$ $\qquad |\Omega(G)| = r_m \cdot r_{m-1} \cdots r_1 \cdot 2^n$

approx r_i by $\tilde{r}_i$ and output as estimate for $|\Omega(G)|$

$$X = 2^n \prod_{i=1}^{m} \tilde{r}_i$$

So need to bound $R = \prod_{i=1}^{m} \frac{\tilde{r}_i}{r_i}$

$\boxed{\text{Claim:}}$ If $\tilde{r}_i$ is $\left(\frac{\varepsilon}{2m}, \frac{\delta}{m}\right)$ approx to r_i $\forall i$
then X is (ε, δ) approx to $|\Omega(G)|$

Pf: $\Pr\left(|\tilde{r}_i - r_i| \le \frac{\varepsilon}{2m} r_i \right) \ge 1 - \frac{\delta}{m} \quad \forall i$

$$\Rightarrow r_i\left(1 - \frac{\varepsilon}{2m}\right) \le \tilde{r}_i \le r_i\left(1 + \frac{\varepsilon}{2m}\right) \qquad \text{w.p.} \ge 1 - \frac{\delta}{m}$$

$$\Rightarrow \left(1 - \frac{\varepsilon}{2m}\right) \le \frac{\tilde{r}_i}{r_i} \le \left(1 + \frac{\varepsilon}{2m}\right)$$

$$1 - \varepsilon \le \left(1 - \frac{\varepsilon}{2m}\right)^m \le \prod_{i=1}^{m} \frac{\tilde{r}_i}{r_i} \le \left(1 + \frac{\varepsilon}{2m}\right)^m \le 1 + \varepsilon \qquad \text{w.p.} \ge 1 - \delta$$

$$(1-\varepsilon) \prod_{i=1}^{m} r_i \le \prod_{i=1}^{n} \tilde{r}_i \le (1+\varepsilon) \prod_{i=1}^{m} r_i \qquad \text{w.p. } 1-\delta$$

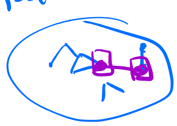
$$\equiv \Pr\left(\left| \prod_{i=1}^{m} \tilde{r}_i - \prod_{i=1}^{m} r_i \right| \ge \varepsilon \prod_{i=1}^{m} r_i \right) \le \delta$$

To get $\tilde{r}_i$ $\left(\frac{\varepsilon}{2m}, \frac{\delta}{m}\right)$-approx for r_i,

use FPAUS for ISs

Idea: (approx) sample indep sets in G_{i-1} & compute fraction of these that are indep in G_i

$$r_i = \frac{|\Omega(G_i)|}{|\Omega(G_{i-1})|}$$



G_{i-1} $\longrightarrow$ approx random I.S. in $\Omega(G_{i-1})$

Input: $G = (V, E)$, ε $\longrightarrow$ FPAUS: Fully polynomial almost uniform sampler $\longrightarrow$ ε uniform sample w from Ω: $\left| \Pr(w \in S) - \frac{|S|}{|\Omega|} \right| \le \varepsilon$

runs in time polynomial in inputsize (G) & $\ln\left(\frac{1}{\varepsilon}\right)$

For this to work, need $r_i = \frac{|\Omega(G_i)|}{|\Omega(G_{i-1})|}$ not too small (no needle in haystack)

Claim: $r_i \ge \frac{1}{2}$

G_i contains one extra edge (u, v)

so any I in G_{i-1} but not indep in G_i contains both u & v

So $\forall\ I$ in $G_{i-1} \setminus G_i \longrightarrow I \setminus v$ in $G_i \cap G_{i-1}$

$$|\Omega(G_{i-1} \setminus G_i)| \le |\Omega(G_i)|$$

$$r_i = \frac{|\Omega(G_i)|}{|\Omega(G_{i-1})|} = \frac{|\Omega(G_i)|}{|\Omega(G_i)| + |\Omega(G_{i-1} \setminus G_i)|} \ge \frac{1}{2}$$

$\Rightarrow$ with polynomially many calls to FPAUS, we can get good approx $\tilde{r}_i$ to r_i using Monte Carlo Thm

Two errors we need to bound:

1. FPAUS $\neq$ exact sampler so exp value $\neq r_i$

   But using, say, $\frac{\varepsilon}{6m}$ sampler $|E(\tilde{r}_i) - r_i| \leq \frac{\varepsilon}{6m}$

2. With samples we get approx to $E(\tilde{r}_i)$

   since r_i big $(\geq \frac{1}{2})$, $E(\tilde{r}_i)$ big $\Rightarrow$

   Monte Carlo Thm $\Rightarrow$ $O\left(\frac{m^2}{\varepsilon^2} \ln\left(\frac{1}{\delta}\right)\right)$ samples suffice

Thm: Given a FPAUS for sampling ISs, one can construct a FPRAS for ISs

Approach works for many "self-reducible" problems

Another example: counting # matchings in a graph

Again: $E = \{e_1, \ldots, e_m\}$

$G_i = (V, E_i)$ where $E_i = \{e_1, \ldots, e_i\}$

$$|M(G)| = \frac{|M(G_m)|}{|M(G_{m-1})|} \frac{|M(G_{m-1})|}{|M(G_{m-2})|} \cdots \frac{|M(G_2)|}{|M(G_1)|} |M(G_1)|$$

Like before: $\frac{|M(G_i)|}{|M(G_{i+1})|} \geq \frac{1}{2}$

Big question remains: how to construct approx sampler?