FPSAC/SFCA, Moscow
June 26-30, 2000

A generalized cover time for random walks on graphs.

Cyril Banderier. Algorithms Project
(INRIA Rocquencourt, France)
http://algo.inria.fr/banderier/
Robert P. Dobrow. Clarkson University
(Postdam, NY)
http://www.clarkson.edu/ dobrowb/

Supper Conjecture

(submitted during a supper at the meeting Analysis of Algorithms in June 1999, at Barcelona)

- m guests around a table
- probability p : pour water
- gives the water carafe randomly to his right or left neighbor, etc.

$T(p) :=$ # carafe moves before everyone has got some water.

Conjecture

$$pE[T(p)] \to 0 \times \infty = mH_m \text{ when } p \to 0 \ ?$$

$(H_m = 1 + \frac{1}{2} + \cdots + \frac{1}{m})$

What about $E[T(p)]$ if people give the carafe preferentially to their friends?

Trivial Pursuit for dummies

random walk on a graph

cover time:= first time that every vertex has been hit by the walk

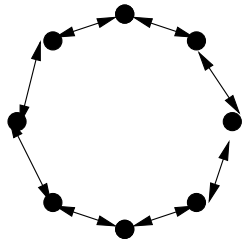
marking time:= first time that every vertex is marked (when the walk reaches vertex v_i, with probability p_i the vertex is *marked*)

cover time = coupon collector problem

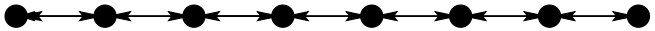
(Pokemon cards !)

marking time = two levels of randomness

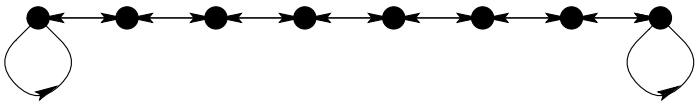
(Trivial Pursuit !)



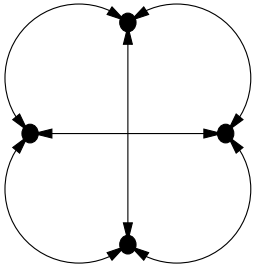
Cyrclic graph (m=8)



Linear graph with reflexion at buttoms (m=8)



Linear graph with loops at buttoms (m=8)



Complete graph (m=4)

Can you guess the marking time of these graphs ?

Notations

Graph

- G = directed connected graph with vertices $v_1, \ldots, v_m$
- Transition matrix A (a_{ij} is the probability of a transition from v_i to v_j).
- Stationary distribution: $\pi_1, \ldots, \pi_m$.
- Probability to mark the vertex v_i whenever it is visited: p_i $(i = 1, \ldots, m)$, $q_i := 1 - p_i$.

Operators

$[z^2]\,(1 + 4z^2 + z^3) = 4$.

For $m = 4$ and $\alpha = \{1, 2, 4\}$

$\zeta_\alpha\,(3u_1 + u_1u_3 + u_3^3 + u_4) = u_3^3$ and

$\sigma_\alpha\,(3u_1 + u_1u_2 + u_3^3 + u_4) = 3q_1 + q_1q_2 + 1 + q_4$.

Marking time on any graph

Theorem 1 (General formula for average marking time). *The average time for marking all the vertices of an arbitrary graph G is given by*

$$E[T(\{p_1,\dots,p_m\})] = \sum_{\alpha\neq\emptyset}(-1)^{|\alpha|+1}\sigma_\alpha \sum_{i=1}^{m} u_1(\mathrm{Id}-AU)^{-1}_{1i},$$

where A is the transition matrix, $U =$

Proof ${A^n}_{i,j}$:= probability of moving from vertex v_i to v_j in n steps.
$[u_1^{k_1}\cdots u_m^{k_m}](AU)^n_{i,j}$:= idem + vertex v_t is visited k_t times

$$F(z,u_1,\dots,u_m)=u_1\sum_{i=1}^{m}\sum_{k=0}^{\infty}z^k(AU)^k_{1,i}=u_1\sum_{i=1}^{m}(\mathrm{Id}-zAU)^{-1}_{1,i}.$$

marking all the vertices at least once $\Longrightarrow$ substitution $u_i^n \leftarrow 1-q_i^n$

$\Longrightarrow$ difference operator $\Delta_i f(u_i) := f(1)-f(q_i)$.

Generating function of the walks that marked *all* the vertices:

$$F^{\mathrm{all}}(z)=\Delta_1\Delta_2\dots\Delta_m F=\sum_{\alpha\subseteq\{1,\dots,m\}}(-1)^{|\alpha|}\sigma_\alpha F(z,u_1,\dots,u_m),$$

$$E[T(\{p_1,\dots,p_m\})]=\frac{\partial}{\partial z}_{|z=1}(1-z)F^{\mathrm{all}}(z).$$

A change of variable $1-z=t$ and a local development in $t=0$ gives

$$\frac{\partial}{\partial z}_{|z=1}\sigma_\alpha\left(\frac{(1-z)u_1}{\mathrm{Id}-zAU}\right)_{ij}=-\sigma_\alpha\left(\mathrm{Id}-AU\right)^{-1}_{ij}\quad\square$$

Remark. Setting the p_i's to 1 $\Rightarrow$ coupon collector problem.

Marking time on the complete graph

Theorem 2. *Let $\widetilde{G}$ be a graph with transition probabilities $\widetilde{a}_{i,j} := \pi_j$. The expected time $E[\widetilde{T}]$ for marking the whole graph $\widetilde{G}$ is $1/p$ times the time needed for visiting all the vertices*

$$E[\widetilde{T}] = \frac{1}{p} E[\textit{coupon collector on } \widetilde{G}],$$

and an inclusion-exclusion formula holds

$$E[\widetilde{T}] = \frac{1}{p} \sum_{\alpha \neq \emptyset} \frac{(-1)^{m-|\alpha|}}{-\zeta_\alpha(\pi_1 + \pi_2 + \cdots + \pi_m)},$$

where the ζ_α's replace some π_i by 0.

Proof Consider the $\pi_i's$, the q_i's, and the p_i's as formal variables, then

$$\text{Prob}(\widetilde{T} \leq n) = \{p_1^{>0}\} \dots \{p_m^{>0}\}(\pi_1 p_1 + \pi_1 q_1 + \cdots + \pi_m p_m + \pi_m q_m)^n,$$

where $\{p_1^{>0}\}p_1 p_2^n + 1 + p_2\}$ stands for the sum of monomials in p_1 of positive exponent. Thus, we have

$$\text{Prob}(\widetilde{T} \leq n) = \sum_{\alpha \subseteq \{1,\dots,m\}} (-1)^{|\alpha|}(\zeta_\alpha(\pi_1 p_1 + \pi_1 q_1 + \cdots + \pi_m p_m + \pi_m q_m))^n,$$

$$\sum \text{Prob}(\widetilde{T} \leq n) z^n = \sum_{\alpha \subseteq \{1,\dots,m\}} \frac{(-1)^{|\alpha|}}{1 - z(q + p\zeta_\alpha(\pi_1 + \pi_2 + \cdots + \pi_m))}.$$

Multiplying by $1 - z$ gives $\sum \text{Prob}(\widetilde{T} = n) z^n$
and differentiating in $z = 1$ gives the expected time

$$E[\widetilde{T}] = \sum_{\alpha \neq \emptyset} (-1)^{|\alpha|} \frac{1}{p(-1 + \zeta_\alpha(\pi_1 + \pi_2 + \cdots + \pi_m))} \quad \square$$

	graph G	graph $\widetilde{G}$
transition matrix	$A = \begin{pmatrix} 1/2 & 1/2 & 0 \\ 1/3 & 1/3 & 1/3 \\ 1/4 & 1/4 & 1/2 \end{pmatrix}$	$\widetilde{A} = \begin{pmatrix} 3/8 & 3/8 & 1/4 \\ 3/8 & 3/8 & 1/4 \\ 3/8 & 3/8 & 1/4 \end{pmatrix}$
stationary distribution	$(3/8,\ 3/8,\ 1/4)$	$(3/8,\ 3/8,\ 1/4)$
	$U = \begin{pmatrix} u_1 & 0 & 0 \\ 0 & u_2 & 0 \\ 0 & 0 & u_3 \end{pmatrix}$	$U = \begin{pmatrix} u_1 & 0 & 0 \\ 0 & u_2 & 0 \\ 0 & 0 & u_3 \end{pmatrix}$
red generating matrix of the walks	$(\mathrm{Id} - zAU)^{-1}$	$(\mathrm{Id} - z\widetilde{A}U)^{-1}$
expected covering time	$95/12 \simeq 7.91$	$29/5 \simeq 5.80$
expected marking time (closed form)	$\dfrac{435 + 494p + 187p^2 + 24p^3}{75p + 60p^2 + 9p^3}$	$29/5p^{-1}$
expected marking time (asymptotics)	$E[T(p)] = 29/5p^{-1} + O(1)$	$E[\widetilde{T}(p)] = 29/5p^{-1}$

Proposition 1. *The marking time T can be approximated by $\widetilde{T}$:*

$$E[\widetilde{T}] - E[T] = O(1), \qquad (p \to 0).$$

Proof In order to prove that

$$\sum_{\alpha\neq\emptyset}(-1)^{|\alpha|+1}\sigma_\alpha \sum_{i=1}^{m} u_1(\mathrm{Id}-\widetilde{A}U)^{-1}_{1i} - \sum_{\alpha\neq\emptyset}(-1)^{|\alpha|+1}\sigma_\alpha \sum_{i=1}^{m} u_1(\mathrm{Id}-AU)^{-1}_{1i} = O(1)$$

it is sufficient to check that

$$[p^{-1}]\sigma_\alpha \left(\mathrm{Id} - \widetilde{A}U\right)^{-1}_{ij} - [p^{-1}]\sigma_\alpha \left(\mathrm{Id} - AU\right)^{-1}_{ij} = O(1).$$

- $\alpha = \{1, \ldots, m\}$ eigenprojection
- $\alpha \neq \{1, \ldots, m\}$ perturbation of two noninvertible matrices

• $\alpha = \{1, \ldots, m\}$ $A = P + R$ (so $P^n = P = \widetilde{A}$ and $PR = RP = 0$)
$(1 - zA)^{-1} = \sum z^n (P + R)^n = \frac{\widetilde{A}}{1-z} + (\mathrm{Id} - zR)^{-1}$
with $(\mathrm{Id} - zR)^{-1}$ regular in $z = 1$ (by Perron–Frobenius theory) so
$(1 - zA)^{-1} = \frac{\widetilde{A}}{1-z} + O(1)$.

•$\alpha \neq \{1, \ldots, m\}$ perturbation of two noninvertible matrices

$$(\mathrm{Id} - A + \epsilon B)^{-1} = \frac{\widetilde{A}}{-\epsilon\lambda'(0)} + O(1) = (\mathrm{Id} - \widetilde{A} + \epsilon\widetilde{B})^{-1}$$

where B is A with its i-th column set to 0 whenever $i \in \alpha$ (so
$A - \epsilon B = \sigma_\alpha AU$),
and where $\lambda(\epsilon)$ is the perturbation of the eigenvalue 1, so
$\lambda'(0) = \sum_{i \in \alpha} \pi_i$. □

Probabilistic proof

One wants

$$E[T] = E[\widetilde{T}] + O(1).$$

Note

$$P(T \leq n) = \sum_{\lambda=(n_1,\dots,n_m)} p_\lambda c_\lambda,$$

by definition

$$E[T] = \sum_{n\geq 0}(1 - \text{Prob}(T \leq n)).$$

limit theorem on Markov Chains, p_λ follows a multidimensional Gaussian law $g(\lambda)$, thus for 'central interval" I and "central box" B

$$E[T] = \sum_{n\in I}\left(1 - \sum_{\lambda\in B} g(\lambda)c_\lambda\right) + o(p^{-1}).$$

I and B only rely on the stationary distribution $\Longrightarrow E[T] = E[\widetilde{T}] + o(p^{-1}) = E[\widetilde{T}] + O(1)$, as $E[\widetilde{T}]$ and $E[T]$ are *rational fractions* in p. □

Integral form

- previous formulae are impracticable as they comprise 2^m summands
- "integral formula" computes $E[T]$ to any precision in linear time

Theorem 3 (First order asymptotics. Integral form). *For any graph G with a stationary distribution $(\pi_1, \ldots, \pi_m)$, the expected marking time is*

$$E[T] = \frac{K}{p} + O(1), \qquad \text{where} \quad K = \int_0^\infty (1 - \prod_{j=1}^{m} (1 - \exp(-\pi_j x)))\, dx.$$

Proof set $p' := -\ln(1-p)$

$$E[T] = \sum_{n \in I} \left(1 - \sum_{\lambda \in B} g(\lambda) \prod_{j=1}^{m} 1 - \exp(-p' n_j \pi_j) \right) + O(1).$$

$$= \sum_{n \geq 0} (1 - \prod_{j=1}^{m} (1 - \exp(-p' n \pi_j))) + O(1) + o(p^{-1}).$$

Euler-Maclaurin formula with $f(x) := 1 - \prod_{j=1}^{m}(1 - \exp(-p'x\pi_j)$:

$$\sum_{n=0}^{\infty} f(n) - \int_0^{\infty} f(x)dx = \frac{f(0) + f(\infty)}{2} + \int_0^{\infty} (x - \lfloor x \rfloor - 1/2) f'(x)dx = O(1).$$

so

$$E[T] = \frac{1}{p'} \int_0^{\infty} (1 - \prod_{j=1}^{m} (1 - \exp(-\pi_j x))) \, dx + O(1) \text{ when } p' \to 0.$$

Since $p' = p + o(p)$, one has $E[T] = K/p + O(1)$. □

Balanced graphs

outdegree of each vertex = its indegree, each outgoing edge is equally likely.

Lemma 1 (Stationary distribution for balanced graphs).
For balanced graphs, one has $\pi_i = \frac{N_i}{\sum_1^m N_i}$
(N_i *is the number of incoming edges of vertex* v_i*).*

Proof N_{ij}:=#of edges from v_i to v_j, N_{*j}:=# incoming edges to v_j, N_{i*}:=# outgoing edges from v_i. The left eigenvector (for eigenvalue 1) satisfies

$$(\pi_1, \pi_2, \ldots, \pi_m) \begin{pmatrix} N_{11}/N_{1*} & \ldots & N_{1m}/N_{1*} \\ \vdots & \vdots & \vdots \\ N_{m1}/N_{m*} & \ldots & N_{mm}/N_{m*} \end{pmatrix} = (\pi_1, \pi_2, \ldots, \pi_m);$$

hence $\sum_i \pi_i \frac{N_{ij}}{N_{i*}} = \pi_j$. This equation is indeed satisfied by $\pi_i := \frac{N_{*i}}{\sum_1^m N_{*i}}$ when $N_{i*} = N_{*i}$. □

Theorem 4 (Balanced graphs). *For balanced graphs with $N := \sum N_i$ edges, one has*

$$E[T] = \frac{N}{p} \sum_{\alpha \neq \emptyset} \frac{(-1)^{m-|\alpha|}}{-\zeta_\alpha(N_1 + N_2 + \cdots + N_m)} + O(1),$$

where the substitutions ζ_α operate on the $\{N_j\}$.

Regular graphs

graphs whose *all* vertices have the same number of incoming and outgoing edges

Theorem 5 (Regular graphs). *When all the π_i's are equal (in particular if G is a regular graph and outgoing edges are chosen uniformly at random), one has $E[T] = \frac{mH_m}{p} + O(1)$.*

Proof

$$E[\widetilde{T}] = \sum_{i=1}^{m} (-1)^i \binom{m}{i} \frac{1}{p(-1 + \frac{m-i}{m})} = \frac{mH_m}{p}.$$

(forward finite difference operators and Euler's transform) □

size	graph	stationary distribution	expected cover time	expected marking time
$m = 1$	K_1	(1)	1	1
$m = 2$	K_2	(1/2, 1/2)	2	3
$m = 3$	K_3	(1/3, 1/3, 1/3)	4	$3H_3$=11/2
$m = 3$	L_3	(1/4, 1/2, 1/4)	5	19/3
$m = 4$	L_4	(1/6, 1/3, 1/3, 1/6)	10	99/10
$m = 4$	K_4	(1/4, 1/4, 1/4, 1/4)	13/2	$4H_4$=25/3
$m = 4$	C_4	(1/4, 1/4, 1/4, 1/4)	7	$4H_4$=25/3
$m = 4$	Y_4	(1/5, 2/5, 1/5, 1/5)	10	110/10
$m = 4$	T_4	(1/8, 3/8, 1/4, 1/4)	163/15	369/35
$m = 4$	Q_4	(3/10, 1/5, 1/5, 3/10)	69/10	62/7

Conclusion

We got, for any graph

- closed forms formulae for the marking time,
- asymptotics,
- practicable formula.

Easy behavior: *Complete graph with loops*
coupon collector time: mH_m
marking time: $mH_m p^{-1}$

General behavior: *any graph*
marking time: related to the stationary distribution.

Finally the Supper Conjecture holds for all *regular graphs*.

Open problem:

Estimate *your* knowledges in sports, history, ... and then say how long will a typical Trivial Pursuit game last for you ?