Journal of Integer Sequences, Vol. 11 (2008),
Article 08.4.4

Edge Cover Time for Regular Graphs

Roberto Tauraso
Dipartimento di Matematica
Università di Roma "Tor Vergata"
via della Ricerca Scientifica
00133 Roma
Italy
tauraso@mat.uniroma2.it

Abstract

Consider the following stochastic process on a graph: initially all vertices are uncovered and at each step cover the two vertices of a random edge. What is the expected number of steps required to cover all vertices of the graph? In this note we show that the mean cover time for a regular graph of N vertices is asymptotically $(N \log N)/2$. Moreover, we compute the exact mean cover time for some regular graphs via generating functions.

1 Introduction

The classical coupon collector's problem can be extended in many ways. In some variants that can be found in the literature (see, for example, [1, 2, 3, 4]) the objects to be collected are the vertices of a graph. There are various interesting collection processes (e.g., a random walk through the graph), but the following one does not seem to have been considered much. Let $\mathcal{G}$ be a connected graph with $N \geq 2$ vertices and $M \geq 1$ edges (no loops). An *edge covering* of $\mathcal{G}$ is a set of edges so that every vertex of $\mathcal{G}$ is adjacent to (or *covered* by) at least one edge in this set. Initially all vertices of the graph are uncovered and at each step we pick a random edge among all edges and we cover its two vertices. Let $\tau_{\mathcal{G}}$ be the *edge cover time*, i.e., the random variable that counts the number of steps required to cover all vertices of G. What is its expected value $E[\tau_{\mathcal{G}}]$?

Let $C(\mathcal{G}, k)$ be the number of edge coverings of $\mathcal{G}$ with exactly k edges. Then the probability that at the n-th step the whole graph is covered is given by

$$p(n) = \sum_{k=1}^{M} C(\mathcal{G}, k) \left\{ {n \atop k} \right\} \frac{k!}{M^n}.$$

Since the following identity holds

$$\left\{ {n \atop k} \right\} k! = \sum_{r=1}^{k} (-1)^{k-r} \binom{k}{r} r^n,$$

then

$$p(n) = \sum_{r=1}^{M} \widehat{C}(\mathcal{G}, r) \left(\frac{r}{M} \right)^n$$

where

$$\widehat{C}(\mathcal{G}, r) = \sum_{k=r}^{M} (-1)^{k-r} \binom{k}{r} C(\mathcal{G}, k).$$

The probability generating function is

$$P(x) = \sum_{n=1}^{\infty} p(n) x^n = \sum_{r=1}^{M} \widehat{C}(\mathcal{G}, r) \sum_{n=1}^{\infty} \left(\frac{rx}{M} \right)^n = \sum_{r=1}^{M-1} \widehat{C}(\mathcal{G}, r) \frac{rx}{M - rx} + \frac{x}{1-x}$$

because $\widehat{C}(\mathcal{G}, M) = C(\mathcal{G}, M) = 1$. In order to compute $E[\tau_{\mathcal{G}}]$ we define

$$Q(x) = \sum_{n=2}^{\infty} (p(n) - p(n-1)) x^n + p(1)x = P(x)(1-x).$$

Then

$$\begin{aligned} Q'(x) &= P'(x)(1-x) - P(x) \\ &= \left(\sum_{r=1}^{M-1} \widehat{C}(\mathcal{G}, r) \frac{Mr}{(M - rx)^2} + \frac{1}{(1-x)^2} \right)(1-x) - \sum_{r=1}^{M-1} \widehat{C}(\mathcal{G}, r) \frac{rx}{M - rx} - \frac{x}{1-x} \\ &= \left(\sum_{r=1}^{M-1} \widehat{C}(\mathcal{G}, r) \frac{Mr}{(M - rx)^2} \right)(1-x) - \sum_{r=1}^{M-1} \widehat{C}(\mathcal{G}, r) \frac{rx}{M - rx} + 1. \end{aligned}$$

Finally we are able to express the answer in finite terms (see [6]) and we obtain

$$E[\tau_{\mathcal{G}}] = Q'(1) = 1 - \sum_{r=1}^{M-1} \widehat{C}(\mathcal{G}, r) \frac{r}{M - r}. \tag{1}$$

In the next sections we will apply the above formula to several kind of graphs, after the generating function whose coefficients give $C(\mathcal{G}, k)$ has been determined. We decided to consider only regular graphs, so that no vertex is privileged with respect to the others.

Before we start, we would like to establish some bounds for $E[\tau_{\mathcal{G}}]$ when $\mathcal{G}$ is a generic d-regular graph with N vertices (and $dN/2$ edges). Since the graph is regular, at each step every vertex has the same probability to be covered. Hence if we assume that only one vertex of the chosen edge is covered, then the modified process is just the classical coupon collector's problem, and therefore its mean cover time NH_N is greater than $E[\tau_{\mathcal{G}}]$. A more precise asymptotic bound is given by the following theorem, which uses the probabilistic method (see, for example, [2]) .

Theorem 1.1. *Let $\mathcal{G}$ be a d-regular graph with N vertices and let $\tau_{\mathcal{G}}$ its cover time. Then for any $\alpha > 0$*

$$Pr\left(\left|\frac{\tau_{\mathcal{G}} - (N\log N)/2}{N}\right| \leq \alpha\right) \geq 1 - 2e^{-2\alpha} + o(1). \tag{2}$$

Moreover,

$$E[\tau_{\mathcal{G}}] \sim (N\log N)/2. \tag{3}$$

Proof. Let $A(v)$ be the event such that the vertex v is not covered after

$$f(N) = N(\log N + a)/2$$

steps with $a \in \mathbb{R}$. Since the probability that the vertex v is covered at any step is

$$p = d/(Nd/2) = 2/N,$$

it follows that

$$\Pr(A(v)) = \sum_{k > f(N)} (1-p)^{k-1}p = (1-p)^{f(N)} = \frac{e^{-a}}{N} + o(1/N).$$

If $v \neq w$ and $v - w$ is not an edge, then the probability that v or w are covered at any step is

$$p = 2d/(Nd/2) = 4/N.$$

Hence

$$\Pr(A(v) \cap A(w)) = (1-p)^{f(N)} = \frac{e^{-2a}}{N^2} + o(1/N^2).$$

On the other hand, if $v \neq w$ and $v - w$ is an edge then the probability that v or w are covered at any step is

$$p = (2d-1)/(Nd/2) = 4/N - 2/(Nd)$$

then

$$\Pr(A(v) \cap A(w)) = (1-p)^{f(N)} = \frac{e^{-a(2-1/d)}}{N^{2-1/d}} + o(1/N^{2-1/d}).$$

Let $X(v)$ be the indicator for the event $A(v)$ and let $X = \sum_v X(v)$ be the number of vertices v such that $A(v)$ occurs. We have the following estimates:

$$\begin{aligned}
E[X(v)] &= 1 \cdot \Pr(A(v)) = \frac{e^{-a}}{N} + o(1/N), \\
\mathrm{Var}[X(v)] &= E[(X(v)]^2) - E[X(v)]^2 = \frac{e^{-a}}{N} + o(1/N), \\
\mathrm{Cov}[X(v), X(w)] &= E[X(v) \cdot X(w)] - E[X(v)] \cdot E[X(w)] \\
&= \Pr(A(v) \cap A(w)) - \frac{e^{-2a}}{N^2} + o(1/N^2) \\
&= \begin{cases} o(1/N^2), & \text{if } v - w \text{ is not an edge;} \\ \dfrac{e^{-a(2-1/d)}}{N^{2-1/d}} + o(1/N^{2-1/d}), & \text{if } v - w \text{ is not an edge.} \end{cases}
\end{aligned}$$

Therefore

$$\begin{aligned}
E[X] &= \sum_v E[X(v)] = e^{-a} + o(1), \\
\mathrm{Var}[X] &= \sum_v \mathrm{Var}[X(v)] + \sum_{v \neq w} \mathrm{Cov}[X(v), X(w)] \\
&= e^{-a} + o(1) + Nd \left(\frac{e^{-a(2-1/d)}}{N^{2-1/d}} + o(1/N^{2-1/d}) \right) \\
&\quad + N(N-1-d)o(1/N^2) \\
&= e^{-a} + o(1).
\end{aligned}$$

So we can find an explicit upper and lower bound for $\Pr(X = 0)$, that is, the the probability that the cover time $\tau_{\mathcal{G}}$ is less than $N(\log N + a)/2$. By Chebyshev's inequality,

$$\Pr(X = 0) \le \Pr(|X - E[X]| \ge E[X]) \le \frac{\mathrm{Var}[X]}{(E[X])^2} = \frac{e^{-a} + o(1)}{e^{-2a} + o(1)} = e^a + o(1). \tag{4}$$

On the other hand

$$\Pr(X = 0) = 1 - \Pr(X > 0) \ge 1 - E[X] = 1 - e^{-a} + o(1). \tag{5}$$

Let $\alpha > 0$. By (4), if $a = -2\alpha$ then

$$\Pr(\tau_{\mathcal{G}} - (N \log N)/2 < -\alpha N) \le e^{-2\alpha} + o(1).$$

By (5), if $a = 2M$ then

$$\Pr(\tau_{\mathcal{G}} - (N \log N)/2 > \alpha N) = 1 - \Pr(\tau_{\mathcal{G}} - (N \log N)/2 < \alpha N) \le e^{-2\alpha} + o(1).$$

Finally

$$\begin{aligned}
\Pr(|\tau_{\mathcal{G}} - (N \log N)/2| < \alpha N) &= 1 - \Pr(\tau_{\mathcal{G}} - (N \log N)/2 < -\alpha N) \\
&\quad - \Pr(\tau_{\mathcal{G}} - (N \log N)/2 > \alpha N) \\
&\ge 1 - 2e^{-2\alpha} + o(1)
\end{aligned}$$

and Eq. (2) has been proved. Eq. (3) follows directly from (2). $\square$

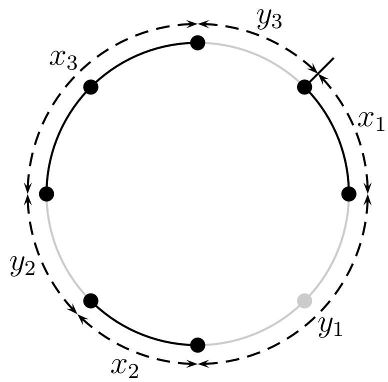
2 The cycle graph $\mathcal{C}_n$

The cycle $\mathcal{C}_n$ is a 2-regular graph with $N = n$ vertices placed around a circle and $M = n$ edges. In order to compute the number of ways $C(\mathcal{C}_n, k, v)$ such that k edges cover v vertices of $\mathcal{C}_n$, we choose one of the n vertices and, from there, we place clockwise the $v-k$ connected components. Let $x_i \ge 1$ be number of edges of the ith-component then these numbers solve the equation

$$x_1 + x_2 + \cdots + x_{v-k} = k.$$

Let $y_i \ge 1$ be number of edges of the gap between ith-component and the next one then these numbers solve equation

$$y_1 + y_2 + \cdots + y_{v-k} = n - k.$$



Hence n times the number of the all positive integral solutions of the previous equations gives $v-k$ times (the first component is labeled) the number $C(\mathcal{C}_n,k,v)$. Therefore

$$C(\mathcal{C}_n,k,v) = \frac{n}{v-k}\binom{k-1}{v-k-1}\binom{n-k-1}{v-k-1} = \frac{n}{k}\binom{k}{v-k}\binom{n-k-1}{v-k-1}$$

and the number of edge coverings with k edges is

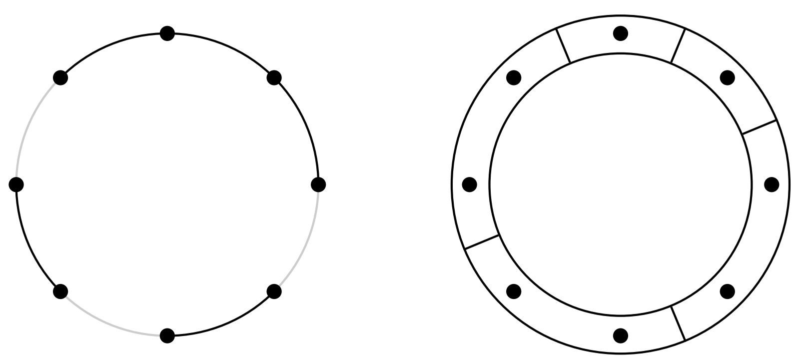
$$C(\mathcal{C}_n,k) = C(\mathcal{C}_n,k,n) = \frac{n}{k}\binom{k}{n-k} = [x^n y^k]\frac{2-xy}{1-yx-yx^2}.$$

The sequence is *triangular* with respect to the double index (n,k) since $C(\mathcal{C}_n,k)$ can be considered zero when $k>n$, and it appears in Sloane's *Encyclopedia* [7] as A113214.

It is interesting to note that the total number of edge coverings of $\mathcal{C}_n$ is the n-th Lucas number (A000032)

$$C(\mathcal{C}_n) = \sum_{k=1}^{n} C(\mathcal{C}_n,k) = \sum_{k=1}^{n}\frac{n}{k}\binom{k}{n-k} = L_n.$$

This is not a real surprise because the edge coverings of $\mathcal{C}_n$ are in bijective correpondence with the monomer-dimer tilings (no overlapping) of $\mathcal{C}_n$: replace any vertex covered by two edges with a monomer and then fill the rest with dimers.



The following identities are well known (see, for example, [5]) and we include the proofs here for completeness.

Lemma 2.1. *i) For any positive integer n*

$$1 - n\sum_{r=1}^{n-1}\frac{(-1)^r}{r}\binom{n-1}{r} = nH_n. \tag{6}$$

ii) For any positive integers n and p

$$\sum_{r=p}^{n-1}\frac{(-1)^{r-p}}{r}\binom{n-1-p}{r-p} = \frac{1}{p\binom{n-1}{p}} \tag{7}$$

Proof. As regards identity (6)

$$\int_0^1 \frac{(1-x)^{n-1}-1}{x}\,dx = \int_0^1 \sum_{r=1}^{n-1}(-1)^r\binom{n-1}{r}x^{r-1}\,dx = \sum_{r=1}^{n-1}\frac{(-1)^r}{r}\binom{n-1}{r},$$

and

$$\int_0^1 \frac{(1-x)^{n-1}-1}{x}\,dx = -\int_0^1 \frac{t^{n-1}-1}{t-1}\,dx = -\int_0^1 \sum_{r=0}^{n-2} t^r\,dt = -H_{n-1}.$$

Now identity (7),

$$\int_0^1 x^{p-1}(1-x)^{n-1-p}\,dx = \int_0^1 x^{p-1}\sum_{r=p}^{n-1}(-1)^{r-p}\binom{n-1-p}{r-p}x^{r-p}\,dx = \sum_{r=p}^{n-1}\frac{(-1)^{r-p}}{r}\binom{n-1-p}{r-p},$$

and on the other hand

$$\int_0^1 x^{p-1}(1-x)^{n-1-p}\,dx = B(p-1, n-1-p) = \frac{(p-1)!(n-1-p)!}{(n-1)!} = \frac{1}{p\binom{n-1}{p}}.$$

□

We are now able to find an explicit formula for the mean cover time for the cycle graph.

Theorem 2.2.

$$E[\tau_{\mathcal{C}_n}] = nH_n - n\sum_{p=1}^{\lfloor n/2\rfloor}\frac{\binom{n-p}{p}}{p\binom{n-1}{p-1}}.$$

Proof. By (1)

$$\begin{aligned}
E[\tau_{\mathcal{C}_n}] &= 1-\sum_{r=1}^{n-1}\widehat{C}(\mathcal{C}_n,r)\frac{r}{n-r} = 1-\sum_{r=1}^{n-1}\widehat{C}(\mathcal{C}_n,n-r)\frac{n-r}{r} \\
&= 1-\sum_{r=1}^{n-1}\frac{n-r}{r}\sum_{k=n-r}^{n}(-1)^{k-n+r}\binom{k}{n-r}\frac{n}{k}\binom{k}{n-k} \\
&= 1-n\sum_{r=1}^{n-1}\frac{1}{r}\sum_{k=n-r}^{n}(-1)^{k-n+r}\binom{k-1}{n-r-1}\binom{k}{n-k} \\
&= 1-n\sum_{r=1}^{n-1}\frac{1}{r}\sum_{p=0}^{r}(-1)^{r-p}\binom{n-p-1}{r-p}\binom{n-p}{p} \\
&= 1-n\sum_{r=1}^{n-1}\frac{(-1)^r}{r}\binom{n-1}{r}-n\sum_{p=1}^{n-1}\binom{n-p}{p}\sum_{r=p}^{n-1}\frac{(-1)^{r-p}}{r}\binom{n-p-1}{r-p}.
\end{aligned}$$

Therefore, by the previous lemma

$$E[\tau_{\mathcal{C}_n}] = nH_n - n\sum_{p=1}^{n-1}\frac{\binom{n-p}{p}}{p\binom{n-1}{p}} = nH_n - n\sum_{p=1}^{\lfloor n/2\rfloor}\frac{\binom{n-p}{p}}{p\binom{n-1}{p}}.$$

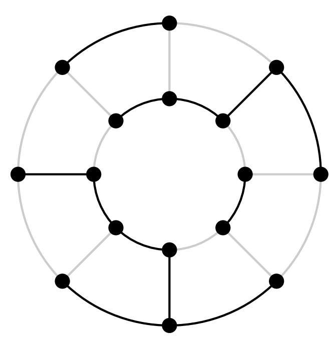
□

The exact values of $E[\tau_{\mathcal{C}_n}]$ for $N=n=2,3,4,5,6,7,8$ are

$$1,\frac{5}{2},\frac{11}{3},\frac{31}{6},\frac{67}{10},\frac{167}{20},\frac{151}{15}.$$

3 The cyclic ladder $\mathcal{C}_n\times\mathcal{K}_2$

The cyclic ladder is a 3-regular graph obtained by taking the graph cartesian product of the cycle graph $\mathcal{C}_n$ and the complete graph $\mathcal{K}_2$. It has $N=2n$ vertice (n on the outer circle and n in the inner circle) and $M=3n$ edges (n on each circle and the n rungs).



The number of coverings of $\mathcal{C}_n \times \mathcal{K}_2$ without rungs is L_n^2 because the inner and the outer circles are covered independently. Assume that the covering has $r \geq 1$ rungs then we label the first one and we let $x_i + 1 \geq 1$ be the number of edges (on one circle) between the i-th rung and the next one. Since the number of coverings of a linear graph with $x_i + 2$ vertices with the end vertices already covered is the Fibonacci number F_{x_i+3} (just the number of monomer-dimer tilings of the (x_i+2)-strip) then the number of coverings with $r \geq 1$ rungs is given by the r-convolution

$$(n/r) \sum_{x_1+\cdots+x_r=n-r} \prod_{i=1}^{r} F_{x_i+3}^2 \,.$$

Therefore

$$C(\mathcal{C}_n \times \mathcal{K}_2) = L_n^2 + \sum_{r=1}^{n} (n/r) \sum_{x_1+\cdots+x_r=n-r} \prod_{i=1}^{r} F_{x_i+3}^2 \,.$$

Now it is easy to find the generating function: since

$$h(x) = \sum_{n=0}^{\infty} L_n^2 x^n = \frac{4-7x-x^2}{(1+x)(1-3x+x^2)} \quad \text{and} \quad g(x) = \sum_{n=0}^{\infty} F_{n+3}^2 x^n = \frac{4+x-x^2}{(1+x)(1-3x+x^2)},$$

it follows that

$$\begin{aligned} f(x) &= h(x) + (xD)\left(\sum_{r=1}^{\infty} \frac{(xg(x))^r}{r}\right) = h(x) + (xD)\left(\log\left(\frac{1}{1-xg(x)}\right)\right) \\ &= \frac{4-15x-18x^2-x^3}{(1+x)(1-6x-3x^2+2x^3)} \\ &= 4+5x+43x^2+263x^3+1699x^4+10895x^5+69943x^6+448943x^7+o(x^7) \end{aligned}$$

and $C(\mathcal{C}_n \times \mathcal{K}_2) = [x^n]f(x)$ is the sequence A123304.

Letting

$$h(x,y) = \frac{4-(4y+3y^2)x-y^3x^2}{1-(y+y^2)x-(y^2+y^3)x^2+y^3x^3}$$

and

$$g(x) = \frac{1+2y+y^2+(-y+y^2+y^3)x-y^3x^2}{1-(y+y^2)x-(y^2+y^3)x^2+y^3x^3},$$

by a similar argument, we can show that $C(\mathcal{C}_n \times \mathcal{K}_2, k) = [x^n y^k]f(x,y)$ where

$$\begin{aligned} f(x,y) &= h(x,y) + \left(x\frac{\partial}{\partial x}\right)\left(\log\left(\frac{1}{1-xyg(x,y)}\right)\right) \\ &= \frac{4-(3y+9y^2+3y^3)x-(4y^2+10y^3+4y^4)x^2+(y^3-y^4-y^5)x^3}{(1+xy)(1-(2y+3y^2+y^3)x-(2y^3+y^4)x^2+(y^3+y^4)x^3)}. \end{aligned}$$

By (1), the exact values of $E[\tau_{\mathcal{C}_n \times \mathcal{K}_2}]$ for $N = 2n = 4, 6, 8, 10$ are

$$\frac{18}{5}, \frac{1919}{280}, \frac{788}{77}, \frac{334283}{24024}.$$

4 $\mathcal{K}_n$ and $\mathcal{K}_{n,n}$

There are two other important regular graphs whose edge coverings are counted by sequences contained in the Sloane's *Encyclopedia* [7]: the complete graph $\mathcal{K}_n$ and the complete bipartite graph $\mathcal{K}_{n,n}$. All the formulas can be verified by applying the inclusion-exclusion principle. The triangular sequence $C(\mathcal{K}_n, k)$ for $0 \le k \le M = \binom{n}{2}$ is A054548 and $C(\mathcal{K}_n)$ is A006129:

$$C(\mathcal{K}_n, k) = \sum_{j=0}^{n} (-1)^j \binom{n}{j} \binom{\binom{n-j}{2}}{k} \quad \text{and} \quad C(\mathcal{K}_n) = \sum_{j=0}^{n} (-1)^j \binom{n}{j} 2^{\binom{n-j}{2}}.$$

By (1), the exact values of $E[\tau_{\mathcal{K}_n}]$ for $N = n = 2, 3, 4, 5, 6$ are

$$1, \frac{5}{2}, \frac{19}{5}, \frac{671}{126}, \frac{97}{14}.$$

The triangular sequence $C(\mathcal{K}_{n,n}, k)$ for $0 \le k \le M = n^2$ is A055599 and $C(\mathcal{K}_{n,n})$ is A048291

$$C(\mathcal{K}_{n,n}, k) = \sum_{i=0}^{n} (-1)^i \binom{n}{i} \sum_{j=0}^{n} (-1)^j \binom{n}{j} \binom{(n-j)(n-i)}{k}$$

and

$$C(\mathcal{K}_{n,n}) = \sum_{j=0}^{n} (-1)^j \binom{n}{j} (2^{n-j} - 1)^n.$$

By (1), the exact values of $E[\tau_{\mathcal{K}_{n,n}}]$ for $N = 2n = 2, 4, 6, 8, 10$ are

$$1, \frac{11}{3}, \frac{1909}{280}, \frac{4687}{455}, \frac{144789}{10313}.$$

References

[1] M. Adler, E. Halperin, R. Karp and V. Vazirani, A stochastic process on the hypercube with applications to peer-to-peer networks, *Proc. STOC 2003*, 575–584.

[2] N. Alon and J. H. Spencer, *The Probabilistic Method*, 2nd Edition, Wiley, 2000.

[3] C. Cooper and A. Frieze, The cover time of random regular graphs, *SIAM J. Discrete Math.*, **18** (2005), 728–740.

[4] N. Dimitrov and C. Plaxton, Optimal cover time for a graph-based coupon collector process, *Proc. ICALP 2005*, 702–716.

[5] H. W. Gould, *Combinatorial Identities*, Morgantown, West Virginia, 1972.

[6] A. N. Myers and H. Wilf, Some new aspects of the coupon collector's problem, *SIAM J. Discrete Math.*, **17** (2003), 1–17.

[7] N. J. A. Sloane, *The On-Line Encyclopedia of Integer Sequences*, published electronically at http://www.research.att.com/~njas/sequences/.

2000 *Mathematics Subject Classification*: Primary 60C05.
Keywords: coupon collector's problem, graph, edge cover.

(Concerned with sequences A000032, A006129, A048291, A054548, A055599, A113214, and A123304.)

Received October 23 2006; revised version received May 24 2008; September 30 2008. Published in *Journal of Integer Sequences*, October 4 2008.

Return to Journal of Integer Sequences home page.