J. Appl. Prob. **51,** 247–261 (2014)
Printed in England

COVER TIMES AND GENERIC CHAINING

JOSEPH LEHEC,* *Université Paris-Dauphine*

Abstract

A recent result of Ding, Lee and Peres (2012) expressed the cover time of the random walk on a graph in terms of generic chaining for the commute distance. Their argument is based on Dynkin's isomorphism theorem. The purpose of this article is to present an alternative approach to this problem, based only on elementary hitting time estimates and chaining arguments.

Keywords: Markov chain; hitting time; cover time; generic chaining

2010 Mathematics Subject Classification: Primary 60J10
Secondary 60G15

1. Introduction

Let $(X_n)_{n\geq 0}$ be an irreducible Markov chain on some state space M. Given $A \subset M$, let

$$T(A) = \inf\{n \geq 0 \colon X_n \in A\}$$

be the first time the chain hits A and let

$$T_{\mathrm{cov}}(A) = \sup_{x\in A} T(x)$$

be the first time the chain X has visited every point of A. The cover time of A is, by definition,

$$\mathrm{cov}(A) = \sup_{x\in A}(\mathbb{E}_x T_{\mathrm{cov}}(A)),$$

where $\mathbb{E}_x$ stands for conditional expectation given $X_0 = x$ (similarly, $\mathbb{P}_x$ stands for conditional probability given $X_0 = x$). To avoid trivial situations, the chain is assumed to be positive recurrent throughout so that $\mathrm{cov}(A) < +\infty$ if and only if A is finite.

Using the strong Markov property it is easily seen that, given x, y, and z in M,

$$\mathbb{E}_x T(y) + \mathbb{E}_y T(z)$$

is the expectation (under $\mathbb{P}_x$) of the first time that the chain has visited y and z (in this order). This implies that

$$\mathbb{E}_x T(y) + \mathbb{E}_y T(z) \geq \mathbb{E}_x T(z).$$

Therefore, the commute time

$$d(x, y) = \mathbb{E}_x T(y) + \mathbb{E}_y T(x)$$

is a distance on M. This article deals with a problem dating back at least as far as [6]: can $\mathrm{cov}(A)$ be estimated in terms of the metric properties of (A, d)? An arguably definitive answer to this question has recently be given by Ding *et al.* [3]; their result is expressed in terms of generic chaining.

Received 13 February 2013; revision received 17 May 2013.
* Postal address: Université Paris-Dauphine, UMR CNRS 7534, Place de Lattre de Tassigny, Paris, 75016, France. Email address: lehec@ceremade.dauphine.fr

https://doi.org/10.1239/jap/1395771427 Published online by Cambridge University Press

1.1. The generic chaining

Estimating the supremum of a Gaussian process $(Y_s)_{s\in S}$ boils down to understanding the metric space (T, d), where $d(s,t) = (\mathbb{E}(Y_s - Y_t)^2)^{1/2}$ is the L_2 distance. This idea dates back to Kolmogorov (see [7]). Thanks to the works of Dudley, Fernique, and Talagrand (see [7] and references within), to name only the most important contributors, this idea has become a well-understood theory usually referred to as *generic chaining*. Let us describe it briefly and refer to [7] for details.

Throughout this article, we let $(N_n)_{n\geq 0}$ be the following sequence of integers:

$$N_0 = 1, \qquad N_n = 2^{2^n}, \quad n \geq 1. \tag{1}$$

Given a set S, a sequence $(\mathcal{A}_n)_{n\geq 0}$ of partitions of S is called *admissible* if $\mathcal{A}_{n+1}$ is a refinement of $\mathcal{A}_n$ and if $|\mathcal{A}_n| \leq N_n$ for every $n \geq 0$, where $|\mathcal{A}_n|$ is just the cardinality of $\mathcal{A}_n$. The cardinality condition implies, in particular, that $\mathcal{A}_0 = \{S\}$. Given a sequence of partitions $(\mathcal{A}_n)_{n\geq 0}$ of S and $s \in S$ we let $A_n(s)$ be the only element of $\mathcal{A}_n$ containing s.

Definition 1. Let (S, d) be a metric space. Set

$$\gamma_2(S,d) = \inf\left[\sup_{s\in S}\left(\sum_{n=0}^{+\infty} 2^{n/2}\Delta(A_n(s), d)\right)\right],$$

where the infimum is taken over all admissible partitions $(\mathcal{A}_n)_{n\geq 0}$ of S, and $\Delta(A, d)$ denotes the diameter of A.

Recall that a Gaussian process is a family $(Y_s)_{s\in S}$ of random variables such that every linear combination of the variables Y_s is Gaussian. The process is said to be centered if $\mathbb{E}Y_s = 0$ for every s. The fundamental result of Talagrand [7, Theorem 2.1.1] is as follows.

Theorem 1. *Let $(Y_s)_{s\in S}$ be a centered Gaussian process. Then*

$$\frac{1}{L}\gamma_2(S,d) \leq \mathbb{E}\sup_{s\in S} Y_s \leq L\gamma_2(S,d), \tag{2}$$

where L is a universal constant and d is the following distance on S:

$$d(s,t) = \sqrt{\mathbb{E}(Y_s - Y_t)^2}. \tag{3}$$

The upper bound is not specific to Gaussian processes, it applies to any centered process $(Y_s)_{s\in S}$ satisfying

$$\mathbb{P}(Y_s - Y_t \geq u) \leq \mathrm{e}^{-u^2/2d(s,t)^2}, \tag{4}$$

for all $s, t \in S$, for all $u > 0$, and for some distance d. Using a union bound it is not hard to see that a centered process for which (4) holds satisfies

$$\mathbb{E}\sup_{s\in A} Y_s \leq C\sqrt{\log|A|}\max_{s,t\in A} d(s,t), \tag{5}$$

for every finite subset A of S. The proof of the upper bound of (2) consists of applying this union bound repeatedly and at different scales.

The lower bound is another story; it is specific to Gaussian processes and much more difficult to prove. Roughly speaking, the argument relies on two properties: the concentration of the

https://doi.org/10.1239/jap/1395771427 Published online by Cambridge University Press

Gaussian measure and the Sudakov inequality. Let us state the latter; if $(Y_s)_{s\in S}$ is a centered Gaussian process then, for all finite subsets A of S,

$$\mathbb{E}\sup_{s\in A} Y_s \geq c\sqrt{\log|A|}\min_{s\neq t\in A} d(s,t), \tag{6}$$

where c is a universal constant and d is the L^2 distance (3).

1.2. The Ding *et al.* [3] theorem

Cover times satisfy inequalities analogous to (5) and (6) due to Matthews [6]: for any finite subset A of M, we obtain

$$\operatorname{cov}(A) \leq (1+\log|A|)\max_{x,y\in A}(\mathbb{E}_x T(y)), \qquad \operatorname{cov}(A) \geq \log|A| \min_{x\neq y\in A}(\mathbb{E}_x T(y)).$$

In view of these inequalities it seems natural to conjecture that the correct order of magnitude for $\operatorname{cov}(A)$ is

$$\gamma_1(A,d) = \inf\biggl[\sup_{x\in A}\biggl(\sum_{n=0}^{+\infty} 2^n \Delta(A_n(x),d)\biggr)\biggr],$$

rather than $\gamma_2(A,d)$ (recall that d is the commute distance $d(x,y) = \mathbb{E}_x T(y) + \mathbb{E}_y T(x)$). This is not quite correct. Here is the result of Ding *et al.* [3, Theorem 1.9].

Theorem 2. *If the Markov chain $(X_n)_{n\geq 0}$ is reversible (and if the state space M is finite) then*

$$\frac{1}{L}[\gamma_2(M,\sqrt{d})]^2 \leq \operatorname{cov}(M) \leq L[\gamma_2(M,\sqrt{d})]^2,$$

for some universal constant L.

Remark 1. Actually the inequality in Theorem 2 remains valid when M is infinite. Indeed, since $d(x,y)\geq 1$ when $x\neq y$, we then have $\gamma_2(M,\sqrt{d}) = +\infty$.

The correct order of magnitude $\gamma_2(M,\sqrt{d})^2$ is comparable to our wrong guess: clearly

$$\gamma_1(M,d) \leq [\gamma_2(M,\sqrt{d})]^2.$$

1.3. Purpose of the present article

The proof of Theorem 2 is very involved. In particular, it relies on Dynkin's isomorphism theorem which makes a connection between local times of the chain and the Gaussian free field associated to the chain. It may be interesting to have a simpler proof relying only on elementary hitting time estimates and on Talagrand's generic chaining. The purpose of this article is to provide such a proof.

Unfortunately, we fail to recover the whole of Theorem 2; here is what we prove.

Theorem 3. *If $(X_n)_{n\geq 0}$ is irreducible and positive recurrent, then*

$$\operatorname{cov}(M) \leq L[\gamma_2(M,\sqrt{d})]^2, \tag{7}$$

for some universal constant L. More generally, we have

$$\operatorname{cov}(A) \leq L[\gamma_2(A,\sqrt{d})]^2, \tag{8}$$

for every subset A of M.

Inequality (7) is slightly stronger than the upper bound of Theorem 2 since the chain is no longer assumed to be reversible. Besides, it is not clear whether the approach of Ding *et al.* [3] yields (8).

Theorem 4. *If, in addition, the chain $(X_n)_{n\geq 0}$ is reversible then*

$$\gamma_1(M, d) \leq L\,\mathrm{cov}(M), \tag{9}$$

where L is a universal constant. Again, we actually have

$$\gamma_1(A, d) \leq L\,\mathrm{cov}(A),$$

for every $A \subset M$.

Remark. The reversibility assumption is necessary. Indeed, consider the discrete torus $\mathbb{Z}_N$ and the Markov kernel given by

$$P(x, x+1) = 1, \quad \text{for all } x \in \mathbb{Z}_N.$$

Clearly, $d(x, y) = N$ for all $x \neq y$, which implies that

$$\gamma_1(T, d) \approx N \log(N).$$

On the other hand, $T_{\mathrm{cov}}(\mathbb{Z}_N) = N$ almost surely (whatever the starting point).

Since $\gamma_1(M, d) \leq [\gamma_2(M, \sqrt{d})]^2$, (9) is weaker than the lower bound of Theorem 2. Let us comment a little bit more on this. In order to compute $\gamma_1(M, d)$, we can restrict to partitions $(\mathcal{A}_n)_{n\geq 0}$ satisfying

$$\mathcal{A}_n = \{\{x\},\ x \in M\},$$

for $n \geq k$, where k is the only integer satisfying

$$N_{k-1} < |M| \leq N_k.$$

Then by convexity we get

$$\begin{aligned}\bigg(\sum_{n=0}^{\infty} 2^{n/2}\sqrt{\Delta(A_n(x), d)}\bigg)^2 &= \bigg(\sum_{n=0}^{k} 2^{n/2}\sqrt{\Delta(A_n(x), d)}\bigg)^2 \\ &\leq (k+1)\sum_{n=0}^{\infty} 2^n \Delta(A_n(x), d),\end{aligned}$$

for every $x \in M$, yielding

$$[\gamma_2(M, \sqrt{d})]^2 \leq C \log(\log|M|)\gamma_1(M, d),$$

for some universal C (provided that $|M| \geq 3$). Therefore, the estimate (9) differs from the correct order of magnitude by at most a factor of $\log(\log|M|)$. This is sharp; there is a Markov chain for which the gap is indeed $\log(\log|M|)$ (see Appendix A).

2. The upper bound

Since $(X_n)_{n\geq 0}$ is an irreducible, positive recurrent Markov chain, there is a unique invariant probability measure which we denote by π. The purpose of this section is to bound

$$\mathbb{E} \sup_{x\in M} T(x)$$

through a chaining argument. Since no estimate such as (4) is available for hitting times, the chaining procedure will be different from Talagrand's, and is taken from [2] and [4].

We need some more notation. Let

$$T^0(x) = 0, \qquad T^k(x) = \inf(n \geq T^{k-1}(x) + 1,\ X_n = x), \quad \text{for all } k \geq 1.$$

When the chain starts from x, the variable $T^k(x)$ is just the kth return time to x. Also, let

$$N_k = \sum_{n=0}^{k-1} \delta_{X_n}$$

be the empirical measure of the chain X. In other words, $N_k(x)$ is the number of visits to x before time k.

The following deviation estimate is due to Kahn *et al.* [4, Lemma 5.2].

Lemma 1. *Let $x \neq y$ in M. Then, for every $\varepsilon > 0$ and for every integer k,*

$$\mathbb{P}_x\bigg(N_{T^k(x)}(y) \leq (1-\varepsilon)\frac{k\pi(y)}{\pi(x)}\bigg) \leq \exp\bigg[-\frac{\varepsilon^2 k}{4\pi(x)d(x,y)}\bigg].$$

Let us sketch the argument we use to prove this result. Because of the strong Markov property, under $\mathbb{P}_x$ the variables

$$(N_{T^i(x)}(y) - N_{T^{i-1}(x)}(y))_{i\geq 1}$$

are independent and identically distributed. And it is a standard fact (see, for instance, [1, Chapter 2]) that their law is geometric: for every integer r

$$\mathbb{P}_x(N_{T^1(x)}(y) \geq r) = p_{xy}(1-p_{yx})^r,$$

where

$$p_{xy} = \mathbb{P}_x(T(y) \leq T^1(x)) = \frac{1}{\pi(x)d(x,y)}.$$

Therefore, Lemma 1 is a Hoeffding-type estimate for sums of independent geometric variables. We refer to [4] for the details.

Our next result is taken from Barlow *et al.* [2, p. 336].

Lemma 2. *Let A be a finite subset of M, let $z \in A$, and let k be an integer. Then*

$$\begin{aligned}\mathbb{E}_z T_{\text{cov}}(A) &\leq \frac{\mathbb{E}_z T^k(z)}{\mathbb{P}_z(T_{\text{cov}}(A) \leq T^k(z))} \\ &= \frac{k}{\pi(z)\mathbb{P}_z(T_{\text{cov}}(A) \leq T^k(z))}.\end{aligned}$$

Proof. Let
$$N = \inf(n \geq 1,\ T_{\text{cov}}(A) \leq T^{nk}(z)).$$
Then, by Wald's identity,
$$\mathbb{E}_z T_{\text{cov}}(A) \leq \mathbb{E}_z T^{Nk}(z) = \mathbb{E}_z(N)\mathbb{E}_z T^k(z).$$

On the other hand, if N is larger than n then the walk fails to cover A during any of the following intervals of time:
$$[0, T^k(z)), \quad [T^k(z), T^{2k}(z)), \quad \ldots, \quad [T^{(n-1)k}(z), T^{nk}(z)),$$
so that
$$\mathbb{P}_x(N > n) \leq \mathbb{P}_z(T_{\text{cov}}(A) \geq T^k(z))^n.$$
The result follows.

Barlow *et al.* [2] combined these two lemmas with a nice chaining argument. Although it is not written this way, their result is essentially the Dudley version of Theorem 3 [7, Theorem 1.2.1], i.e.
$$\operatorname{cov}(M) \leq L\biggl(\sum_{n=0}^{\infty} \mathrm{e}_n(M, \sqrt{d})2^{n/2}\biggr)^2,$$
where
$$\mathrm{e}_n(M, \sqrt{d}) = \inf_A\Bigl(\sup_{x\in M} \sqrt{d(x, A)}\Bigr)$$
(the infimum is taken over all subsets A of M satisfying $|A| \leq N_n$). This is weaker than Theorem 3. Indeed, swapping the supremum and the sum in the definition of γ_2, it is easily seen that
$$\gamma_2(M, \sqrt{d}) \leq C \sum_{n=0}^{\infty} \mathrm{e}_n(M, \sqrt{d})2^{n/2},$$
for some universal constant C. We show that it is possible to modify the chaining argument in Barlow *et al.* [2] to obtain Theorem 3.

Let z, x, and y be in M such that $x \neq y$ and let k and l be two integers larger than 1. Observe that
$$\begin{aligned}\mathbb{P}_z(T^l(y) > T^k(x)) &= \mathbb{P}_z(N_{T^k(x)}(y) \leq l-1)\\ &\leq \mathbb{P}_z(N_{T^k(x)}(y) - N_{T^1(x)}(y) \leq l-1)\\ &= \mathbb{P}_x(N_{T^{k-1}(x)}(y) \leq l-1).\end{aligned}$$

The last equality is a consequence of the strong Markov property. If $(l-1)/\pi(y) < (k-1)/\pi(x)$, applying Lemma 1 to $k-1$, $l-1$, and
$$\varepsilon = 1 - \frac{(l-1)\pi(x)}{(k-1)\pi(y)},$$
gives
$$\mathbb{P}_z(T^l(y) > T^k(x)) \leq \exp\biggl[-\biggl(\frac{k-1}{\pi(x)} - \frac{l-1}{\pi(y)}\biggr)^2 \bigg/ 4d(x, y)\frac{k-1}{\pi(x)}\biggr]. \tag{10}$$

This will be our key estimate. Lastly, we shall use the following elementary fact: if x and y are distinct elements of M then

$$\frac{1}{\pi(x)} = \mathbb{E}_x T^1(x) \le \mathbb{E}_x T(y) + \mathbb{E}_y T(x) = d(x, y).$$

Let us reformulate Theorem 3.

Proposition 1. *Let $A \subset M$, let $z \in A$, and let $(\mathcal{A}_n)_{n\ge 0}$ be an admissible sequence of partitions of A. Then*

$$\mathbb{E}_z(T_{\text{cov}}(A)) \le L\left(\sup_{x\in A}\sum_{n=0}^{\infty} 2^{n/2}\sqrt{\Delta(A_n(x))}\right)^2.$$

Recall that $A_n(x)$ denotes the only element of $\mathcal{A}_n$ containing x. Also, Δ denotes the diameter with respect to the commute distance.

Proof. Let $t_0(A) = z$, and for each n and for each $B \in \mathcal{A}_n$ let $t_n(B)$ be an arbitrary element of B. Given $x \in A$, we let $x_n = t_n(A_n(x))$. We can assume that A is finite and that

$$\mathcal{A}_n = \{\{x\},\ x \in A\}$$

for n large enough (the right-hand side of the desired inequality is equal to $+\infty$ otherwise). Therefore, $x_n = x$ eventually. Let

$$r_n(x) = \sup_{y\in A_n(x)} \sum_{k=n}^{+\infty} 2^{k/2}\sqrt{\Delta(A_k(y))}$$

and

$$k_n(x) = \lfloor 34\pi(x_n) r_n(x) r_0(x)\rfloor + 1,$$

where $\lfloor r\rfloor$ denotes the integer part of r. Observe that $r_n(x)$ and $k_n(x)$ depend only on $A_n(x)$. In particular, $k_0(x)$ depends on nothing. Also,

$$\begin{aligned} r_n(x) - r_{n+1}(x) &\ge 2^{n/2}\sqrt{\Delta(A_n(x))} \\ &\ge 2^{n/2}\sqrt{d(x_n, x_{n+1})}. \end{aligned}$$

We claim that, for every x and n,

$$\mathbb{P}_z(T^{k_{n+1}(x)}(x_{n+1}) > T^{k_n(x)}(x_n)) \le \mathrm{e}^{-2^{n+3}} \le \frac{1}{N_{n+3}}. \tag{11}$$

Indeed, if $x_n = x_{n+1}$ then $k_{n+1}(x) \le k_n(x)$ and the inequality is trivial. Otherwise, write

$$\begin{aligned} \frac{k_n(x) - 1}{\pi(x_n)} - \frac{k_{n+1}(x) - 1}{\pi(x_{n+1})} &\ge 34(r_n(x) - r_{n+1}(x)) r_0(x) - \frac{1}{\pi(x_n)} \\ &\ge 34\cdot 2^{n/2}\sqrt{d(x_n, x_{n+1})}\, r_0(x) - \frac{1}{\pi(x_n)}. \end{aligned}$$

Since $x_n \ne x_{n+1}$ and $\sqrt{d(x_n, x_{n+1})} \le r_0(x)$, we have

$$\frac{1}{\pi(x_n)} \le \sqrt{d(x_n, x_{n+1})}\, r_0(x).$$

https://doi.org/10.1239/jap/1395771427 Published online by Cambridge University Press

Therefore,

$$\frac{k_n(x)-1}{\pi(x_n)} - \frac{k_{n+1}(x)-1}{\pi(x_{n+1})} \geq (34\cdot 2^{n/2}-1)\sqrt{d(x_n,x_{n+1})}r_0(x)$$
$$\geq 33\cdot 2^{n/2}\sqrt{d(x_n,x_{n+1})}r_0(x).$$

Also,

$$\frac{k_n(x)-1}{\pi(x_n)} \leq 34 r_n(x) r_0(x) \leq 34 r_0(x)^2.$$

Since $33^2/(4\cdot 34)\geq 2^3$, combining (10) with the last two inequalities yields (11).

The number of possible couples (x_n, x_{n-1}) is at most $N_n N_{n+1}$. Recall the definition (1) of N_n and observe that $N_n^2 \leq N_{n+1}$ for all n. A union bound shows that the probability that there exists x and n such that

$$T^{k_{n+1}(x)}(x_{n+1}) \geq T^{k_n(x)}(x_n)$$

is at most

$$\sum_{n\geq 0}\frac{N_n N_{n+1}}{N_{n+3}} \leq \sum_{n\geq 0}\frac{1}{N_{n+2}} \leq \sum_{n\geq 4} 2^{-n} = \tfrac{1}{8}.$$

Therefore, with a probability of at least $\frac{7}{8}$, we have

$$T^{k_{n+1}(x)}(x_{n+1}) \leq T^{k_n(x)}(x_n),$$

for all x and n; hence,

$$T^{k_n(x)}(x_n) \leq T^{k_0(x)}(x_0) = T^{k_0}(z).$$

Since $x_n = x$ for n large enough and $k_n(x)\geq 1$, we obtain

$$\text{for all } x\in A, \qquad T(x) \leq T^{k_0}(z),$$

with a probability of at least $\frac{7}{8}$. In other words,

$$\mathbb{P}_z(T_{\text{cov}}(A) \leq T^{k_0}(z)) \geq \tfrac{7}{8}.$$

Together with Lemma 2, we get

$$\mathbb{E}_z T_{\text{cov}}(A) \leq \frac{8k_0}{7\pi(z)} \leq \frac{8}{7}\bigg(34 r_0^2 + \frac{1}{\pi(z)}\bigg).$$

Unless $A=\{z\}$, in which case $\text{cov}(A)=0$ and there is nothing to prove, we have $1/\pi(z)\leq \Delta(A)\leq r_0^2$. Therefore,

$$\mathbb{E}_z T_{\text{cov}}(A) \leq \frac{8\cdot 35}{7}\bigg(\sup_{x\in A}\sum_{n=0}^{\infty} 2^{n/2}\sqrt{\Delta(A_n(x))}\bigg)^2,$$

which is the required result.

3. The lower bound

We start this section with another definition; given $A \subset M$ let

$$\operatorname{cov}_-(A) = \min_{x\in A} \mathbb{E}_x T_{\operatorname{cov}}(A), \qquad \operatorname{cov}_+(A) = \max_{x\in A} \mathbb{E}_x T_{\operatorname{cov}}(A).$$

Note that the cover time of A, which was previously denoted by $\operatorname{cov}(A)$, is now denoted by $\operatorname{cov}_+(A)$ to avoid confusion with $\operatorname{cov}_-(A)$. In this section we prove the following result.

Proposition 2. *Let $(X_n)_{n\geq 0}$ be an irreducible, positive recurrent Markov chain on a discrete state space M. If the chain is reversible then, for every finite subset A of M,*

$$\gamma_1(A, d) \leq L(\operatorname{cov}_-(A) + \Delta(A, d)),$$

where L is a universal constant.

Remarks. (i) This result yields Theorem 4 since, clearly,

$$\operatorname{cov}_-(A) \leq \operatorname{cov}_+(A), \qquad \Delta(A, d) \leq \operatorname{cov}_+(A).$$

(ii) The term $\Delta(A, d)$ cannot be removed from the inequality. Indeed, if $M = \{0, 1\}$ and the transitions are given by the matrix

$$\begin{pmatrix} \varepsilon & 1-\varepsilon \\ \varepsilon & 1-\varepsilon \end{pmatrix},$$

then

$$\gamma_1(M, d) \geq \Delta(M, d) = \frac{1}{\varepsilon(1-\varepsilon)};$$

whereas $\operatorname{cov}_-(M) = \min(1/\varepsilon, 1/(1-\varepsilon))$.

3.1. Talagrand's growth condition

Recall the majorizing measure theorem, Theorem 1 If $(Y_s)_{s\in S}$ is a centered Gaussian process then

$$\gamma_2(S, d) \leq L\mathbb{E} \sup_{s\in S} Y_s,$$

where d is the L^2 distance (3). The proof of this result consists of showing (using Gaussian concentration and Sudakov's inequality) that the functional

$$A \mapsto \mathbb{E} \sup_{s\in A} Y_s$$

satisfies an abstract growth condition, and that such functionals dominate γ_2. Here is the definition of the growth condition adapted to the γ_1 situation (rather than γ_2).

Definition 2. (*Growth condition.*) Let (M, d) be a metric space. A functional $F\colon \mathcal{P}(M) \to \mathbb{R}_+$ is said to satisfy the growth condition with parameters $r > 1$ and $\tau \in \mathbb{N}$ if, for every step $n \in \mathbb{N}$ and every scale $a > 0$, the followings holds. Let $m = N_{n+\tau}$; for every sequence $H_1, \dots, H_m$ of nonempty subsets of M satisfying

1. $\Delta(\bigcup_{i\leq m} H_i) \leq ra$,
2. $d(H_i, H_j) \geq a$, for all $i \neq j$,

https://doi.org/10.1239/jap/1395771427 Published online by Cambridge University Press

3. $\Delta(H_i) \le a/r$, for all i,

we have
$$F\Big(\bigcup_{i\le m} H_i\Big) \ge a2^n + \min_{i\le m} F(H_i).$$

Theorem 5. *If F is nondecreasing for the inclusion and satisfies the growth condition with parameters r and τ then*

$$\gamma_1(M,d) \le L2^\tau(\Delta(M,d) + rF(M)),$$

where L is a universal constant.

We refer to [7, Section 1.3] for a proof of this theorem. The purpose of the rest of this section is to show that the functional
$$A \mapsto \mathrm{cov}_-(A)$$
is nondecreasing and satisfies the growth condition on (M,d) (where d is the commute distance) with universal parameters τ and r.

Lemma 3. *The functional $A \mapsto \mathrm{cov}_-(A)$ is nondecreasing for the inclusion.*

Proof. We use the strong Markov property. The shift operator is denoted by σ, i.e. for every integer k
$$\sigma_k(X_0, X_1, \dots) = (X_k, X_{k+1}, \dots).$$
Let $A \subset B$ and let $x \in B$. Then

$$T_{\mathrm{cov}}(B) \ge T(A) + T_{\mathrm{cov}}(A) \circ \sigma_{T(A)}.$$

In words: at time $T(A)$ the chain has yet to visit every point of $A \backslash \{X_{T(A)}\}$. By the strong Markov property
$$\begin{aligned}\mathbb{E}_x T_{\mathrm{cov}}(B) &\ge \mathbb{E}_x T(A) + \mathbb{E}_x[\mathbb{E}_{X_{T(A)}} T_{\mathrm{cov}}(A)]\\ &\ge \mathbb{E}_x T(A) + \mathrm{cov}_-(A)\\ &\ge \mathrm{cov}_-(A),\end{aligned}$$

which is the required result.

3.2. Variations on Matthews' bound

The following result is due to Matthews [6, Theorem 2.6].

Lemma 4. *Let A be a finite subset of M, let $a > 0$, and assume that $\mathbb{E}_x T(y) \ge a$ for every $x \ne y$ in A. Then*
$$\mathrm{cov}_-(A) \ge a \sum_{k=1}^{|A|-1} \frac{1}{k} \ge a\log(|A|).$$

Proof. Let $x \in A$. Assuming that $|A| \ge 2$ (otherwise the result is trivial) we have

$$\sum_{y\in A,\, y\ne x} \mathbb{P}_x(T_{\mathrm{cov}}(A) = T(y)) = 1.$$

https://doi.org/10.1239/jap/1395771427 Published online by Cambridge University Press

So there exists $y \in A$ such that

$$\mathbb{P}_x(T_{\text{cov}}(A) = T(y)) \geq \frac{1}{|A| - 1}. \tag{12}$$

Let $A' = A \setminus \{y\}$, let $S = T_{\text{cov}}(A')$, and let $T = T_{\text{cov}}(A)$. Clearly,

$$T = S + (T(y) \circ \sigma_S)\, \mathbf{1}_{\{S < T(y)\}},$$

where $\mathbf{1}_{\{\cdot\}}$ is the indicator function. By the strong Markov property,

$$\mathbb{E}_x T = \mathbb{E}_x S + \mathbb{E}_x[(\mathbb{E}_{X_S} T(y))\, \mathbf{1}_{\{S < T(y)\}}].$$

On the event $\{S < T(y)\}$ the point X_S is an element of A different from y. Therefore, $\mathbb{E}_{X_S} T(y) \geq a$. Together with (12) we obtain

$$\mathbb{E}_x T_{\text{cov}}(A) \geq \mathbb{E}_x T_{\text{cov}}(A') + \frac{a}{|A| - 1}.$$

An obvious induction on $|A|$ finishes the proof.

The following lemma is proved the same way.

Lemma 5. *Let $H_1, \ldots, H_m$ be nonempty subsets of M satisfying*

$$\mathbb{E}_x T(y) \geq a, \quad \textit{for all } (x, y) \in H_i \times H_j, \quad \textit{for all } i \neq j.$$

Then, for all $x \in \bigcup_{i \leq m} H_i$,

$$\mathbb{E}_x \max_{i \leq m} T(H_i) \geq a \log(m).$$

An additional application of the strong Markov property yields the following refinement of Lemma 5.

Proposition 3. *Let $H_1, \ldots, H_m$ be nonempty subsets of M satisfying $\mathbb{E}_x T_y \geq a$, for all $(x, y) \in H_i \times H_j$, for all $i \neq j$. Then*

$$\text{cov}_-\left(\bigcup_{i \leq m} H_i\right) \geq a \log(m) + \min_{i \leq m} \text{cov}_-(H_i). \tag{13}$$

Proof. Let $x \in \bigcup_{i \leq m} H_i$. Let $S = \max_{i \leq m} T(H_i)$ and $T = T_{\text{cov}}(\bigcup_{i \leq m} H_i)$. If $S = T(H_i)$ then at time S the chain has yet to visit every point of $H_i \setminus \{X_S\}$. Therefore,

$$T \geq S + \sum_{i=1}^{m} (T_{\text{cov}}(H_i) \circ \sigma_S)\, \mathbf{1}_{\{S = T(H_i)\}}.$$

Using the strong Markov property, we get

$$\begin{aligned}
\mathbb{E}_x T &\geq \mathbb{E}_x S + \sum_{i=1}^{m} \mathbb{E}_x[(\mathbb{E}_{X_S} T_{\text{cov}}(H_i))\, \mathbf{1}_{\{S = T(H_i)\}}] \\
&\geq \mathbb{E}_x S + \min_{i \leq m} \text{cov}_-(H_i).
\end{aligned}$$

Together with the previous lemma we get the result.

We are close to the desired growth condition. We would like to obtain the inequality (13) under the weaker hypothesis

$$d(x, y) = \mathbb{E}_x T(y) + \mathbb{E}_y T(x) \geq a, \quad \text{for all } x, y \in H_i \times H_j,\ i \neq j.$$

This is done in the next subsection. Roughly speaking, reversibility insures that, for a reasonable proportion of x and y, the hitting times $\mathbb{E}_x T(y)$ and $\mathbb{E}_y T(x)$ are of the same of order of magnitude.

3.3. Reversibility

Again, this part of the argument is taken from [4]. We start with a simple lemma concerning directed graphs. Given a directed graph $G = (V, E)$, a path of G is a sequence $x_1, \dots, x_m$ of vertices satisfying $(x_i, x_{i+1}) \in E$ for $i \leq m$. The length of such a path is defined to be m. An independent set is a subset A of V satisfying $(x, y) \notin E$ for all x, y in A.

Lemma 6. *If every path of G has length of at most m then G has an independent set of cardinality of at least $|V|/m$.*

This is a standard result, but we still sketch the argument used in the proof. It is easy to show by induction on m that G is then m-colorable: it is possible to map the vertices of G to $\{1, \dots, m\}$ in such a way that connected points have different images. Then, by the pigeon hole principle, at least $|V|/m$ vertices have the same image, which is the result.

From now on, the chain $(X_n)_{n\geq 0}$ is assumed to be reversible. Consequently, we have the following commuting property for hitting times.

Lemma 7. *For every sequence $x_1, \dots, x_m$ of elements of M we have*

$$\begin{aligned}
&\mathbb{E}_{x_1} T(x_2) + \cdots + \mathbb{E}_{x_{m-1}} T(x_m) + \mathbb{E}_{x_m} T(x_1) \\
&\quad = \mathbb{E}_{x_1} T(x_m) + \mathbb{E}_{x_m} T(x_{m-1}) + \cdots + \mathbb{E}_{x_2} T(x_1).
\end{aligned} \tag{14}$$

We refer to [5, Lemma 10.10] for a proof.

Corollary 1. *Let A be a subset of M and $a > 0$. If $\Delta(A, d) \leq 16a$ and if $d(x, y) \geq a$ for all $x \neq y$ in A, then there exists a subset A' of A satisfying*

- $|A'| \geq |A|/33$,
- $\mathbb{E}_x T(y) \geq a/4$ *for all $x \neq y$ in A'.*

Proof. We define a graph G with vertex set A by saying that the edge (x, y) is present if $x \neq y$ and $\mathbb{E}_x T(y) \leq a/4$. Let $x_1, \dots, x_m$ be a path of G. Then the inequalities

$$\mathbb{E}_{x_i} T(x_{i+1}) \leq \frac{a}{4}, \qquad \mathbb{E}_{x_{i+1}} T(x_i) \geq \frac{3a}{4}$$

and (14) give

$$\frac{(m-1)a}{4} + \mathbb{E}_{x_m} T(x_1) \geq \frac{3(m-1)a}{4} + \mathbb{E}_{x_1} T(x_m).$$

Together with the bound on the diameter of A, we obtain $m - 1 \leq 32$. Therefore, G has an independent set of cardinality at least $|A|/33$. This is our set A'.

3.4. The growth condition for the cover time

Proposition 4. *The functional* $A \mapsto \operatorname{cov}_-(A)$ *satisfies the growth condition with parameters* $r = 16$ *and* $\tau = 5$.

Proof. Let $n \in \mathbb{N}$, $a > 0$, and $m = N_{n+5}$. Let $H_1, \ldots, H_m$ satisfy

1. $\Delta(\bigcup_{i \le m} H_i) \le 16a$,
2. $d(H_i, H_j) \ge a$, for all $i \ne j$,
3. $\Delta(H_i) \le a/16$, for all $i \le m$.

Let $x_1, \ldots, x_m$ belong to $H_1, \ldots, H_m$, respectively. By the first two properties and Corollary 1, there exists a subset I of $\{1, \ldots, m\}$ satisfying

- $|I| \ge m/33$,
- $\mathbb{E}_{x_i} T(x_j) \ge a/4$, for every $i \ne j$ in I.

Let $i \ne j$ in I and let $(x, y) \in H_i \times H_j$. Then

$$\mathbb{E}_x T(y) \ge \mathbb{E}_{x_i} T(x_j) - \mathbb{E}_{x_i} T(x) - \mathbb{E}_y T(x_j) \ge \frac{a}{4} - \frac{a}{16} - \frac{a}{16} = \frac{a}{8}.$$

Proposition 3 gives

$$\operatorname{cov}_-\left(\bigcup_{i \in I} H_i\right) \ge \frac{a}{8} \log(|I|) + \min_{i \in I} \operatorname{cov}_-(H_i).$$

Since

$$|I| \ge \frac{N_{n+5}}{33} \ge \frac{N_{n+5}}{N_3} \ge N_{n+4} \ge \mathrm{e}^{8 \cdot 2^n},$$

we obtain

$$\operatorname{cov}_-\left(\bigcup_{i \le m} H_i\right) \ge a 2^n + \min_{i \le m} \operatorname{cov}_-(H_i),$$

which is the result.

Now, by Theorem 5 we obtain

$$\gamma_1(M, d) \le L(\operatorname{cov}_-(M) + \Delta(M, d)).$$

Obviously we can replace M by any subset A of M in this inequality: if a functional F satisfies the growth condition on (M, d) then it also satisfies it on (A, d).

Appendix A. The $\log(\log(|M|))$ gap

We have seen in Section 1 that, for any metric space (M, d),

$$[\gamma_2(M, \sqrt{d})]^2 \le C \log(\log|M|) \gamma_1(M, d). \tag{15}$$

We show in this appendix that this is sharp and that the example saturating the inequality can be chosen to be the state space of a reversible Markov chain equipped with the commute distance. The example is taken from [4] and was pointed out to the author by James Lee.

Let M be a rooted tree of depth D (large enough) satisfying

- nodes at depth $i \leq D - 1$ have $N_i + 1$ children,
- edges between depth i and depth $i + 1$ have multiplicity 2^i,

and let X be the random walk on this graph. The stationary measure is given by $\pi(x) = d(x)/2E$ for every x, where $d(x)$ is the number of edges (counted with multiplicity) starting from x and E is the total number of edges. Also, π is reversible. Let us compute the commute distance d. Because of the tree structure it is easily seen that

$$d(x, y) = \sum_{i=0}^{n-1} d(x_i, x_{i+1}), \tag{16}$$

where $x_0, \dots, x_n$ is the shortest path from x to y. Therefore, it is enough to compute $d(x, y)$ when x and y are neighbors, in which case we use the following formula:

$$\mathbb{P}_x(T(y) < T^1(x)) = \frac{1}{\pi(x) d(x, y)}$$

(see [1]). Because of the tree structure, $\mathbb{P}_x(T(y) < T^1(x))$ is just the transition probability from x to y. We obtain

$$d(x, y) = 2E \cdot 2^{-i},$$

when (x, y) is an edge between depth i and depth $i + 1$. When x and y are any two nodes of M, (16) then implies that

$$E \cdot 2^{-i+1} \leq d(x, y) \leq E \cdot 2^{-i+3}, \tag{17}$$

where i is the depth of their closest common ancestor.

Proposition 5. *There is a universal constant C such that*

$$\frac{DE}{C} \leq \gamma_1(M, d) \leq CDE, \tag{18}$$

$$\frac{D\sqrt{E}}{C} \leq \gamma_2(M, \sqrt{d}) \leq CD\sqrt{E}. \tag{19}$$

Since D is of the order of $\log(\log|M|)$, this shows that (15) is sharp (up to the constant).

Proof. Let us start with the upper bound of (18). It is more convenient to use the following definition for γ_1:

$$\gamma_1(M, d) = \inf \sup_{x \in M} \sum_{i=0}^{+\infty} 2^i d(x, M_i),$$

where the infimum is taken over every sequence $(M_i)_{i \in \mathbb{N}}$ of subsets of M satisfying the cardinality condition $|M_i| \leq N_i$ for every i. It is well known (see [7]) that this definition coincides with the one with partitions, up to a universal factor.

For $0 \leq i \leq D$ let S_i be the set of vertices of depth at most i. Using (17) we obtain $d(x, S_i) \leq E \cdot 2^{-i+3}$ for every $x \in M$. Therefore,

$$\sup_{x \in M} \sum_{i=0}^{+\infty} 2^i d(x, S_i) \leq E \sum_{i=0}^{D} 2^i 2^{-i+3} = 8E(D + 1).$$

Also, it is easily shown that

$$|S_i| \leq N_{i+3}.$$

The sequence $(S_i)_{n\in\mathbb{N}}$ does not quite satisfy the right cardinality condition, but this is not a big deal. If we shift the sequence by letting $M_0 = M_1 = M_2 = S_0$ and $M_i = S_{i-3}$ for $i \geq 3$, we still have

$$\sup_{x\in M} \sum_{i=0}^{+\infty} 2^i d(x, M_i) \leq CED,$$

for some universal C, which proves the upper bound of (18).

To prove the lower bound we need to show that the previous sequence of approximations is essentially optimal. Let $(M_i)_{i\geq 0}$ be a sequence of subsets of M satisfying $|M_i| \leq N_i$ for every i. A vertex x of depth $i \leq D-1$ has $N_i + 1$ children. So at least one of them, call it y, has the following property: neither y nor any of its offspring belong to M_i. Using this observation, we can construct inductively a sequence $x_0, x_1, \dots, x_D$, where x_0 is the root of M and such that

- x_{i+1} is a child of x_i,
- neither x_{i+1} nor any of its offspring belong to M_i,

for every $i \leq D-1$. Let $i \leq D-1$ and let $x \in M_i$. Since x is not an offspring of x_{i+1} we have $d(x, x_D) \geq E \cdot 2^{-i+1}$. Thus,

$$\sum_{i=0}^{\infty} 2^i d(x_D, M_i) \geq E \sum_{i=0}^{D-1} 2^i 2^{-i+1} = 2ED,$$

which proves the lower bound of (18).

Inequality (19) is proved in exactly the same way.

Acknowledgement

This work owes a lot to stimulating discussions the author had with Guillaume Aubrun.

References

[1] Aldous, D. and Fill, J. A. (2014). Reversible Markov chains and random walks on graphs. In preparation. Early drafts of chapters available at: http://www.stat.berkeley.edu/users/aldous/RWG/book.html

[2] Barlow, M. T., Ding, J., Nachmias, A. and Peres, Y. (2011). The evolution of the cover time. *Combin. Prob. Comput.* **20,** 331–345.

[3] Ding, J., Lee, J. R. and Peres, Y. (2012). Cover times, blanket times, and majorizing measures. *Ann. Math.* **175,** 1409–1471.

[4] Kahn, J., Kim, J. H., Lovász, L. and Vu, V. H. (2000). The cover time, the blanket time, and the Matthews bound. In *41st Annual Symposium on Foundations of Computer Science*, IEEE, Los Alamitos, CA, pp. 467–475.

[5] Levin, D. A., Peres, Y. and Wilmer, E. L. (2009). *Markov Chains and Mixing Times*. American Mathematical Society, Providence, RI.

[6] Matthews, P. (1988). Covering problems for Brownian motion on spheres. *Ann. Prob.* **16,** 189–199.

[7] Talagrand, M. (2005). *The Generic Chaining*. Springer, Berlin.