ALEA, Lat. Am. J. Probab. Math. Stat. 19, 755–767 (2022)
DOI: 10.30757/ALEA.v19-29

Covering a compact space by fixed-radius or growing random balls

David J. Aldous

Department of Statistics, 367 Evans Hall # 3860, U.C. Berkeley CA 94720
E-mail address: aldous@stat.berkeley.edu
URL: http://www.stat.berkeley.edu/users/aldous

Abstract. Simple random coverage models, well studied in Euclidean space, can also be defined on a general compact metric space S. In one specific model, "seeds" arrive as a Poisson process (in time) at random positions with some distribution θ on S, and create balls whose radius increases at constant rate. By standardizing rates, the cover time C depends only on θ. The value $\chi(S) = \min_\theta \mathbb{E}_\theta C$ is a numerical characteristic of the compact space S, and we give weak general upper and lower bounds in terms of the covering numbers of S. This suggests a future research program of improving such general bounds, and estimating $\chi(S)$ for familiar examples of compact spaces. We treat one example, infinite product space $[0,1]^\infty$ with the product topology. On a different theme, by analogy with the geometric models, and with the discrete coupon collector's problem and with cover times for finite Markov chains, one expects a "weak concentration" bound for the distribution of C to hold under minimal assumptions. We prove this as a simple consequence of a general result for increasing set-valued Markov processes.

1. Introduction

Analogs of the classical coupon collector's problem have been extensively studied in several different contexts. One context is geometric: covering by (for instance) random balls in Euclidean space. See Hall (1988) for now-classical results, and Penrose (2021) for references to recent work, in this broad area. The particular case of *balls* makes sense in any metric space, but apparently has not been studied in that generality. We will briefly discuss the fixed-radius setting (section 2), but we find it more interesting to examine what we will call the *growth model*. In this model (section 3), "seeds" arrive as a Poisson process (in time) at random positions with some distribution θ on a space S, and create balls whose radius increases at constant rate. We study the *cover time*, the time C at which the space is entirely covered. In Euclidean space, this is the growth model underlying the well-studied *Johnson-Mehl tessellation*, defined as the partition of $\mathbb{R}^d$ into the regions first covered by the ball emanating from a specific seed. The focus in that literature (see Møller (1992) for an extensive survey of that model, and Chiu et al. (2013) Chapter 9 for the broad account of random tessellation models) is on stochastic geometry properties of the cells of the tessellation. As part

Received by the editors January 29th, 2021; accepted January 24th, 2022.
2010 *Mathematics Subject Classification.* 60G55, 60G70.
Key words and phrases. Cover time, Johnson-Mehl tessellation, random coverage.

of that literature, the cover time C of a bounded subset of $\mathbb{R}^d$ has been studied in detail in Chiu (1995), and sharp asymptotic results are known. In this article we consider instead the setting of an arbitrary compact metric space (compactness ensures that the cover time C has finite expectation). In section 6 we give simple general bounds on mean cover times $\mathbb{E}_\theta C$. These bounds, if applied to the case of $\mathbb{R}^d$, are much weaker than those known via explicit calculation (see section 5.1). However our main purpose (section 5) is to observe that, with standardized rates, the "mean cover time from optimal seed distribution" value $\chi(S) = \min_\theta \mathbb{E}_\theta C$ is a numerical characteristic of any compact space S. We give weak general upper and lower bounds on $\chi(S)$ in terms of the covering numbers of S. In section 6 we apply these general bounds to one example, infinite product space $[0,1]^\infty$ with product topology. Relating $\chi(S)$ to other characteristics of S, and estimating $\chi(S)$ sharply for familiar examples of compact spaces, remain interesting and challenging open problems.

1.1. *Concentration of cover time distribution.* One intuitively expects that the "weak concentration" property of the coupon collector time T_n (that $\text{s.d.}(T_n)/\mathbb{E}T_n \to 0$ as $n \to \infty$) should extend quite generally to other cover time contexts, and should hold under minimal assumptions even when one cannot calculate the expectation explicitly. Indeed this is known to be true in the Markov chain context (see section 7). We will prove analogous results concerning weak concentration of the distribution of the cover time C. Part of our motivation is to spotlight two different general methods (known, but apparently not well known) for showing weak concentration in general settings without calculating the expectation of the cover time[1]. In each of sections 2 and 3 we specify a model (fixed-radius or growing random balls), recall the relevant general method, and show that a concentration bound is obtained very easily using that method.

Further discussion of models and methodology is deferred to section 7.

2. A concentration bound for covering with fixed radius random balls

Here we indicate how a concentration result for covering, obtainable on Euclidean space in sharp form by explicit calculation (see Hall (1988)), can be extended to weak bounds in a very general setting. Take a compact metric space (S,ρ). Let μ be a probability measure on S with full support, and for $r > 0$ define

$$\eta(r) := \inf_s \mu(\text{ball}(s,r)) > 0$$

where $\text{ball}(s,r) = \{s' : \rho(s,s') \le r\}$. Write $\sigma_1, \sigma_2, \ldots$ for i.i.d. random points of S from distribution μ. For fixed $r_0 > 0$ consider the random subset

$$\mathcal{R}_n = \mathcal{R}_n^{(r_0)} := \cup_{1 \le i \le n} \text{ball}(\sigma_i, r_0).$$

We call this the *fixed-radius model.* Consider the cover time

$$C = C^{(r_0)} := \min\{n : \mathcal{R}_n = S\} \tag{2.1}$$

for which compactness easily implies $\mathbb{E}C < \infty$. The probability that a given point s is in $\text{ball}(\sigma_i, r_0)$ equals $\mu(\text{ball}(s,r_0))$, and so the mean time until point s is covered equals $1/\mu(\text{ball}(s,r_0))$, which is at most $1/\eta(r_0)$. So to obtain a concentration result for C a natural assumption is that $\mathbb{E}C \gg 1/\eta(r_0)$, in other words that $\eta(r_0)\mathbb{E}C$ is large. Our result below is of that general form, but also involves the *dimension*-related quantity $d(r)$ defined as the smallest integer such that

$$\text{each ball of radius } r \text{ can be covered by } d(r) \text{ balls of radius } r/2. \tag{2.2}$$

Proposition 2.1. *In the fixed-radius model, for the cover time C at (2.1),*

$$\text{var}\left(\frac{C}{\mathbb{E}C}\right) \le \kappa \, \frac{d(r_0)}{\eta(r_0/2)\mathbb{E}C}$$

[1] Other than its order of magnitude.

for the absolute constant κ stated in Proposition 2.2 below.

We will derive Proposition 2.1 from a known general result, discussed as Proposition 2.2 below.

2.1. *The random subset cover bound.* Here we copy the setup and result directly from Aldous (1991). Let S_0 be a finite set. Let $\mathcal{Y}$ be a random subset of S_0, whose distribution is arbitrary subject to the requirement

$$\mathbb{P}(s \in \mathcal{Y}) > 0 \text{ for each } s \in S_0. \tag{2.3}$$

Let $\mathcal{Y}_1, \mathcal{Y}_2, \ldots$ be independent random subsets distributed as $\mathcal{Y}$. Let $\mathcal{R}_n$ be the range of this process: $\mathcal{R}_n = \cup_{i \le n} \mathcal{Y}_i$ and let C_{set} be the cover time

$$C_{set} := \min\{n : \mathcal{R}_n = S_0\}.$$

Note $\mathbb{E}C_{set} < \infty$ by (2.3) and finiteness of S_0. For any non-random subset $B \subset S_0$ let $c(B)$ be the mean cover time of B:

$$c(B) := \mathbb{E}C(B); \quad C(B) := \min\{n : \mathcal{R}_n \supseteq B\}.$$

Our bound involves the *terminal set*

$$\mathcal{T} := S_0 \setminus \mathcal{R}_{C_{set}-1}$$

that is the last uncovered portion of S_0.

Proposition 2.2 (Aldous (1991) Theorem 1)**.** $\operatorname{var}\left(\frac{C_{set}}{\mathbb{E}C_{set}}\right) \le \kappa \frac{\mathbb{E}c(\mathcal{T})}{\mathbb{E}C_{set}}$ *for an absolute constant κ.*

Though stated in Aldous (1991) for a finite state space S_0, Proposition 2.2 extends to continuous space, in particular our compact metric space S, with unchanged proof, except that now we need to replace assumption (2.3) by the assumption $\mathbb{E}C_{set} < \infty$.

A numerical value for κ is not stated in Aldous (1991) but the argument shows

> if $\frac{\mathbb{E}c(\mathcal{T})}{\mathbb{E}C_{set}} \le 1/8$ then the inequality in Proposition 2.2 holds for $\kappa = 4(1 + \sum_{n \ge 1}(n + 1)^2 2^{1-n}) = 92$.

The alternate case can be handled directly by using the obvious submultiplicativity property of C_{set}, that is

$$\mathbb{P}(C_{set} \ge t_1 + t_2) \le \mathbb{P}(C_{set} \ge t_1)\ \mathbb{P}(C_{set} \ge t_2). \tag{2.4}$$

Because $\mathbb{P}(C_{set}/\mathbb{E}[C_{set}] \ge 2) \le 1/2$, submultiplicativity implies that $C_{set}/\mathbb{E}[C_{set}]$ is stochastically dominated by $2G_{1/2}$ where $G_{1/2}$ has Geometric(1/2) distribution. It follows that $\operatorname{var}(C_{set}/\mathbb{E}[C_{set}]) \le \mathbb{E}(2G_{1/2})^2 = 6$. This implies

> if $\frac{\mathbb{E}c(\mathcal{T})}{\mathbb{E}C_{set}} \ge 1/8$ then the inequality in Proposition 2.2 holds for $\kappa = 6 \times 8 = 48$.

So in fact Proposition 2.2 is known to hold with $\kappa = 92$, but apparently no attempt has ever been made to optimize the constant.

Of course it may be difficult to analyze $\mathcal{T}$, and so one does not expect to obtain sharp bounds on specific models in this way. But Proposition 2.2 may be useful in obtaining order of magnitude bounds in general settings. In particular if there is some geometric or metric structure on the set and if the random subsets $\mathcal{Y}$ are small in diameter, then $\mathcal{T}$ must be small in diameter, so one needs only to bound $c(B)$ as a function of the diameter of B. The next section gives a simple illustration of that method.

2.2. *Proof of Proposition 2.1.* In the notation of Proposition 2.2, the terminal set $\mathcal{T}$ is such that $\mathcal{T} \subset \mathrm{ball}(s, r_0)$ for some $s \in S$, so

$$c(\mathcal{T}) \leq \sup_s \mathbb{E}C(\mathrm{ball}(s, r_0)).$$

The mean time until one of the random centers σ falls in a given ball of radius $r_0/2$ is at most $1/\eta(r_0/2)$. Note that a ball of radius $r_0/2$ is covered by any ball of radius r_0 whose center is in the former ball. So from the definition of dimension $d(r_0)$, for each s there are $d(r_0)$ points $s_1, \ldots, s_{d(r_0)}$ such that $\mathrm{ball}(s, r_0)$ is covered whenever each of $(\mathrm{ball}(s_i, r_0/2), 1 \leq i \leq d(r_0))$ contains at least one of the random centers σ, and so

$$\sup_s \mathbb{E}C(\mathrm{ball}(s, r_0)) \leq d(r_0)/\eta(r_0/2).$$

The result follows from Proposition 2.2.

3. The growth model and its concentration bound

Consider as before a compact metric space (S, ρ), a probability measure μ on S, but now introduce two rates $0 < \lambda < \infty$ and $0 < v < \infty$. Write $0 < \tau_1 < \tau_2 < \ldots$ for the times of a rate-λ Poisson process, and write $\sigma_1, \sigma_2, \ldots$ for i.i.d. random points of S from distribution μ. The verbal description

> seeds arrive at times of a Poisson process at i.i.d. random positions, and then create balls whose radius grows at rate v

is formalized as the set-valued *growth process*

$$\mathcal{X}(t) := \cup_{i: \tau_i \leq t} \, \mathrm{ball}\,(\sigma_i, v(t - \tau_i)). \tag{3.1}$$

We study the cover time

$$C := \min\{t : \ \mathcal{X}(t) = S\}$$

which is finite because $\mathbb{E}\tau_1 = 1/\lambda$ and so

$$1/\lambda \leq \mathbb{E}C \leq 1/\lambda + \Delta/v \tag{3.2}$$

where Δ is the diameter of S. To obtain a concentration bound it is natural to require that $\mathbb{E}C$ is large relative to the maximum expected time to cover any given single point, that is relative to

$$c^* := \max_{s \in S} \mathbb{E}C(s); \quad C(s) := \min\{t : \ s \in \mathcal{X}(t)\}.$$

It turns out this is the only requirement.

Proposition 3.1. *In the growth model* (3.1), $\mathrm{var}\left(\frac{C}{\mathbb{E}C}\right) \leq \frac{c^*}{\mathbb{E}C}$.

We will derive Proposition 3.1 from a known general result, discussed as Proposition 3.2 below. Note that the expectation of the number of balls covering v at time t equals $\int_0^t \mu(\mathrm{ball}(s, vu)) \; \lambda du$ and so from the Poisson property

$$\mathbb{P}(C(s) > t) = \exp\left(-\int_0^t \mu(\mathrm{ball}(s, vu)) \; \lambda du\right) \tag{3.3}$$

from which we can in principle obtain a formula for $\mathbb{E}C(s)$.

3.1. *A monotonicity bound for Markov chains.* We will adapt a result from Aldous (2016b). The setting there is a continuous-time Markov chain (X_t) on a finite state space Σ, where we study the hitting time

$$T := \inf\{t : \ X_t \in \Sigma_0\} \tag{3.4}$$

for a fixed subset $\Sigma_0 \subset \Sigma$. Assume

$$h(x) := \mathbb{E}_x T < \infty \text{ for each } x \in \Sigma \tag{3.5}$$

which holds in the finite case under the natural "reachability" condition. Assume also a rather strong "monotonicity" condition:

$$h(x') \leq h(x) \text{ whenever } x \to x' \text{ is a possible transition.} \tag{3.6}$$

Proposition 3.2 (Aldous (2016b))**.** *Under conditions* (3.5) *and* (3.6)*, for any initial state,*

$$\frac{\mathrm{var}\ T}{\mathbb{E}T} \leq \max\{h(x) - h(x') : \ x \to x' \textit{ a possible transition}\}.$$

Though stated in Aldous (2016b) for a finite state space Σ, the proof of Proposition 3.2 extends to the continuous space setting by simply replacing sums by integrals. We will write out the argument for our specific growth model, and then comment on its generality.

For our growth model $\mathcal{X}(t)$ at (3.1), the state space is the space of compact subsets x of the compact metric space S. Write

$$h(x) = \mathbb{E}_x C$$

and observe that the process $h(\mathcal{X}(t))$ is non-increasing. The only discontinuities of $h(\mathcal{X}(t))$ are at a time τ when a new seed arrives at a point σ, at which time there is a transition $x \to x \cup \{\sigma\}$ of $\mathcal{X}(t)$. Consider the martingale

$$M(t) := \mathbb{E}[C|\mathcal{X}(t)] = h(\mathcal{X}(t \wedge C)) + t \wedge C.$$

The Doob-Meyer decomposition of $M^2(t)$ into a martingale $Q(t)$ and a predictable process $a(t)$ is

$$M^2(t) - M^2(0) = Q(t) + \int_0^t a(\mathcal{X}(u))\ du$$

where

$$a(x) := \int (h(x) - h(x \cup \{\sigma\})^2 \mu(d\sigma). \tag{3.7}$$

Taking expectation at $t = \infty$ gives

$$\mathrm{var}\ C = \mathbb{E} \int_0^C a(\mathcal{X}(u))\ du.$$

The martingale property for $\mathbb{E}[C|\mathcal{X}(t)]$ corresponds to the identity

$$b(x) := \int (h(x) - h(x \cup \{\sigma\})\mu(d\sigma) = 1 \text{ for } x \neq S \tag{3.8}$$

and therefore

$$EC = \mathbb{E} \int_0^C b(\mathcal{X}(u)) du.$$

So

$$\begin{aligned} \frac{\mathrm{var}\ C}{\mathbb{E}C} &\leq \sup_x \frac{a(x)}{b(x)} && (3.9) \\ &\leq \sup_{x,\sigma} (h(x) - h(x \cup \{\sigma\}). && (3.10) \end{aligned}$$

Here (3.10) is the specific result we need for the growth process. But note that the argument works for essentially any Markov process (X_t) and stopping time T such that, writing $h(x) := \mathbb{E}_x T$, the process $h(X_t)$ is decreasing. In that general setting we obtain (3.9) for the functions analogous to $a(x)$ and $b(x)$ at (3.7) and (3.8) associated with the given process.

3.2. *Proof of Proposition 3.1.* To apply (3.10) to prove Proposition 3.1 it is enough to show that, for each pair (x, σ),

$$h(x) - h(x \cup \{\sigma\}) \leq \mathbb{E}C(\sigma). \tag{3.11}$$

But this holds by considering the natural coupling $(\mathcal{X}(t), \mathcal{X}'(t) = \mathcal{X}(t) \cup \text{ball}(\sigma, vt), t \geq 0)$ of the growth processes with $\mathcal{X}(0) = x, \mathcal{X}'(0) = x \cup \{\sigma\}$. In this coupling, for the time $C^*(\sigma)$ at which σ is reached by a ball of $\mathcal{X}(\cdot)$ whose seed arrived after time 0, we have (by the triangle inequality on S) that $\mathcal{X}(C^*(\sigma) + t) \supseteq \mathcal{X}'(t)$, and so the cover times for these two processes differ by at most $C^*(\sigma)$. But this $C^*(\sigma)$ is distributed as $C(\sigma)$ for the growth process started at the empty set, establishing (3.11).

4. Cover time bounds for the standardized growth model

Comparing the statements of Propositions 2.1 and 3.1 suggests that the growth model is more tractable for the study of covering. Intuitively this is because the behavior of the growth model is “smoother" in that it does not rely on the detailed geometry of the space (S, ρ) at the given distance r_0. In this section we record some simple observations.

We can “standardize" the growth model by choosing time and distance units to make $\lambda = v = 1$. Precisely, from a standardized process $\mathcal{X}^0(t)$ on (S, ρ^0) we can construct the non-standardized process as $\mathcal{X}(t) = \mathcal{X}^0(t/\lambda)$ on space (S, ρ) where $\rho(x, y) = \frac{v}{\lambda}\rho^0(x, y)$. Within this correspondence we have, for instance,

$$\mathbb{E}C = \tfrac{1}{\lambda}\mathbb{E}C^0, \quad \Delta = \tfrac{v}{\lambda}\Delta^0$$

enabling all the inequalities stated later in the standardized case to be transferred to the general case. For instance in Proposition 4.2 below, (b) becomes

$$\Delta = \tfrac{v}{\lambda}\Delta^0 \leq \kappa_1 \tfrac{v}{\lambda}(\mathbb{E}C^0)^2 = \kappa_1 \tfrac{v}{\lambda}(\lambda \mathbb{E}C)^2 = \kappa_1 v \lambda (\mathbb{E}C)^2. \tag{4.1}$$

We start with a simple relationship between the diameter Δ and $\mathbb{E}C$, for which we need the following elementary lemma for metric spaces.

Lemma 4.1. *If a connected compact metric space (S, ρ) can be covered by balls of radii $(r_i, 1 \leq i \leq m)$, then $\Delta \leq 2\sum_i r_i$.*

Proof: It is enough to prove this for open balls; then take the “open ball" result with $(r_i + \varepsilon)$ and let $\varepsilon \downarrow 0$ to get the “closed ball" result. Write $S = \cup_{1 \leq i \leq m} \text{ball}_{open}(s_i, r_i)$. Consider the graph G on vertices $(s_i, 1 \leq i \leq m)$ where (s_i, s_j) is an edge if and only if the open balls $\text{ball}_{open}(s_i, r_i)$ and $\text{ball}_{open}(s_j, r_j)$ have non-empty intersection, in which case $\rho(s_i, s_j) \leq r_i + r_j$. If the graph G is not connected, then there is a proper subset $V \subset \{s_i, 1 \leq i \leq m\}$ with complement V^c such that $\cup_{s_i \in V} \text{ball}_{open}(s_i, r_i)$ and $\cup_{s_i \in V^c} \text{ball}_{open}(s_i, r_i)$ are disjoint, contradicting connectedness of S. So G is connected, and so for any given pair (s_α, s_β) of vertices there is a path in G from s_α to s_β, and so $\rho(s_\alpha, s_\beta) \leq 2\sum_{1 \leq i \leq m} r_i - (r_\alpha + r_\beta)$. Then for any given pair s', s'' in S we can find such s_α and s_β with $\rho(s', s_\alpha) \leq r_\alpha$ and $\rho(s'', s_\beta) \leq r_\beta$, establishing the result. □

Proposition 4.2. *In the standardized growth model on a space (S, ρ),*
(a) $\mathbb{E}C \leq 1 + \Delta$.
(b) If S is connected then $\Delta \leq \kappa_1(\mathbb{E}C)^2$ for an absolute constant κ_1.

Proof: Part (a) is (3.2). For (b), consider

$$D(t) := 2\sum_i (t - \tau_i)^+.$$

At time t the state $\Xi(t)$ is the union of balls of radii $(t-\tau_i)^+$ and so by Lemma 4.1 we have

$$\Delta \le D(C).$$

We can rewrite $D(t)$ in terms of the Poisson counting process $(N(t))$ as $D(t) = 2\int_0^t N(u)du$ and then

$$\Delta \le \mathbb{E}D(C) = 2\int_0^\infty \mathbb{E}[N(t)1_{(t\le C)}]\ dt.$$

Using the Cauchy-Schwarz inequality

$$\Delta \le 2\int_0^\infty (t^2+t)^{1/2}\ \sqrt{\mathbb{P}(C\ge t)}\ dt. \tag{4.2}$$

Breaking the integral at 1 leads to

$$\Delta \le 2^{3/2}\left(1+\int_0^\infty t\ \sqrt{\mathbb{P}(C\ge t)}\ dt\right).$$

Reusing the submultiplicativity property (2.4) for this cover time C, combined with Markov's inequality $\mathbb{P}(C \ge e\mathbb{E}C) \le e^{-1}$, leads to an exponential tail bound

$$\mathbb{P}(C\ge t) \le \exp(1-\tfrac{t}{e\mathbb{E}C}) \tag{4.3}$$

and so

$$\Delta \le 2^{3/2}\left(1+e^{1/2}\int_0^\infty t\ \exp(-\tfrac{t}{2e\mathbb{E}C}\ dt\right).$$

The integral equals $(2e\ \mathbb{E}C)^2$, and because $\mathbb{E}C \ge 1$ from (3.2) we finally find that (b) holds for

$$\kappa_1 = 2^{3/2}(1+e^{1/2}(2e)^2) \approx 141.$$

We have not attempted to optimize this constant. □

Continuing with this standardization, consider a sequence of connected compact metric spaces $S = S^{(n)}$ and probability distributions $\mu = \mu^{(n)}$. Proposition 3.1 implies that as $n\to\infty$

$$\text{if } \tfrac{c^*}{\mathbb{E}C}\to 0 \text{ then } \tfrac{C}{\mathbb{E}C}\to 1 \text{ in } L^2. \tag{4.4}$$

Can we relate the hypothesis $c^*/\mathbb{E}C \to 0$ to other aspects of the spaces? Recall that c^* is in principle directly calculable from (3.3), whereas determining whether $\mathbb{E}C$ is of the same order, or larger order, than c^* requires some more detailed knowledge of the space S.

If the diameters $\Delta^{(n)}$ are bounded (as n increases) then by Proposition 4.2 the mean cover times $\mathbb{E}C^{(n)}$ are bounded; because $\mathbb{P}(C^{(n)} > t) \ge \exp(-t)$ the conclusion (and hence the assumption) of (4.4) is false. So we need study only the case $\Delta^{(n)} \to \infty$. Here is a simple example to show that the conclusion of (4.4) is not always true.

Example. Take $S^{(n)}$ to be the real line segment $[0,n]$ and $\mu^{(n)}(\{0\}) = 1-1/n$ and $\mu^{(n)}(\{n\}) = 1/n$. One easily sees that

$$n^{-1}C^{(n)} \to_d \min(1, \tfrac{1}{2}(1+\xi))$$

where ξ has Exponential(1) distribution.

In an opposite direction, we note a simple upper bound on $\mathbb{E}C/c^*$, that is a lower bound on $c^*/\mathbb{E}C$, in terms of the *covering number*

$$\mathrm{cov}(r) := \text{ minimum number of radius } r \text{ balls that cover } S\ . \tag{4.5}$$

Proposition 4.3. *In the standardized growth model,*

$$\frac{\mathbb{E}C}{c^*} \le \min_{a>0}[a + e(e+\log\mathrm{cov}(ac^*))].$$

Proof: As at (4.3) the submultiplicative property of $C(s)$ implies $\mathbb{P}(C(s) \geq t) \leq \exp(1 - \frac{t}{e\mathbb{E}C(s)})$. Applying this to the centers (s_i) of $\mathrm{cov}(r)$ covering radius r balls,

$$\mathbb{P}(\max_i C(s_i) \geq t) \leq e\ \mathrm{cov}(r) \exp(-\tfrac{t}{ec^*}).$$

Setting $t_0 := ec^* \log \mathrm{cov}(r)$,

$$\mathbb{E}[\max_i C(s_i)] = \int_0^\infty \mathbb{P}(\max_i C(s_i) \geq t)\ dt \leq t_0 + e \cdot ec^*.$$

Because $C \leq r + \max_i C(s_i)$ we have

$$\mathbb{E}C \leq r + ec^*(e + \log \mathrm{cov}(r)).$$

Setting $r = ac^*$ gives the stated bound. □

5. The minimum cover time

For the standardized growth model on connected compact (S, ρ), take two points s_1, s_2 which are diametrically opposite, that is $\rho(s_1, s_2) = \Delta$. Then the maximum of $\mathbb{E}_\mu C$ over μ equals $1 + \Delta$, attained by the measure μ degenerate at s_1. But what can we say about the *minimum* of $\mathbb{E}_\mu C$ over μ? In other words, there is a numerical characteristic of a compact metric space S defined by

$$\chi(S) = \min_\mu \mathbb{E}_\mu C.$$

This suggests a research program:
(i) Find general bounds relating $\chi(S)$ to other numerical characteristics of S
(ii) Estimate $\chi(S)$ for familiar examples of compact spaces.
In this article we make only a modest start on this program, via the general bounds in Propositions 5.1 and 5.3 below, and via analysis of an infinite product space in section 6.

Intuitively, $\chi(S)$ should be related to the covering numbers $\mathrm{cov}(r)$ at (4.5), and indeed we easily find upper and lower bounds, as follows. Given r, consider μ uniform on the centers $(s_i, 1 \leq i \leq \mathrm{cov}(r))$ of the covering radius-r balls. Then $C \leq r + \tau_{\mathrm{cov}(r)}$ where τ_n is the elementary coupon collector time with $\mathbb{E}\tau_n = n(1 + 1/2 + \ldots + 1/n) \leq (1 + \log n)n$. So we have established

Proposition 5.1.

$$\chi(S) \leq \min_{r>0} [r + \mathrm{cov}(r)(1 + \log \mathrm{cov}(r))].$$

For a bound in the opposite direction, observe first that for the Poisson counting process $(N(t), 0 \leq t < \infty)$ of seed arrival times,

Lemma 5.2. *If t_0 and c_0 are such that $\mathbb{P}(C > c_0) + \mathbb{P}(N(c_0) > t_0) < 1$ then $\mathrm{cov}(c_0) \leq t_0$.*

Proof: The assumption implies that the event $\{C \leq c_0, N(c_0) \leq t_0\}$ has non-zero probability; on that event we have

$$\mathrm{cov}(c_0) \leq N(C) \leq N(c_0) \leq t_0.$$

□

Applying Lemma 5.2 with $c_0 = 3\mathbb{E}C$ and $t_0 = 3c_0$ gives $\mathrm{cov}(3\mathbb{E}C) \leq 9\mathbb{E}C$. This is true for any μ and so, using monotonicity of $r \to \mathrm{cov}(3r)$, we have established

Proposition 5.3. *In the standardized growth model,*

$$\chi(S) \geq \sup\{r : \mathrm{cov}(3r) > 9r\}.$$

5.1. *Euclidean space.* As mentioned in the introduction, the growth process on $\mathbb{R}^d$ has been studied for its role in the construction of the Johnson-Mehl tessellation. Write C_L for the cover time of $[0, L]^d$ for the standardized process. In this case, $\mathrm{cov}(r) = (1 + o(1))c_d(L/r)^d$ for $r << L$, for a constant c_d. To apply our general results (Propositions 5.1 and 5.3) we want to choose r such that $\mathrm{cov}(r)$ is the same order of magnitude as r, which means we choose r to be of order $L^{\frac{d}{d+1}}$. This yields asymptotic bounds of the form

$$(1 - o(1))c'_d L^{\frac{d}{d+1}} \leq \chi(C_L) \leq (1 + o(1))c''_d L^{\frac{d}{d+1}} \ \log L \quad \text{as } L \to \infty.$$

However, in this setting the cover time has been analyzed much more precisely. For μ_0 uniform on $[0, L]^d$, Theorem 4 of Chiu (1995) shows that there are constants a_d such that

$$\mathbb{E}_{\mu_0} C_L \sim a_d L^{\frac{d}{d+1}} \ \log L \ \text{ as } L \to \infty$$

and moreover obtains a more refined limit distribution after appropriate rescaling. Intuitively, we expect that $\mathbb{E}_{\mu_0} C_L$ is very close to the minimum $\chi(C_L) := \min_\mu \mathbb{E}_\mu C_L$ but we have not attempted a proof.

6. An infinite-dimensional example

Here we study the infinite product space $S_\Delta := [0, \Delta]^\infty$ which, in the product topology, is compact and metrizable. Write $\mathbf{x} = (x_i, i \geq 1)$ for elements of S_Δ. We will use the metric

$$\rho(\mathbf{x}, \mathbf{y}) := \sum_{i \geq 1} 2^{-i} |x_i - y_i|$$

for which the diameter of S_Δ equals Δ. Write C_Δ for the cover time for the standardized growth model on S_Δ. In this example the general bounds from section 5 are sufficient to establish the (logarithmic scale) asymptotic behavior of $\chi(S_\Delta) := \min_\mu \mathbb{E}_\mu C_\Delta$ as $\Delta \to \infty$.

Theorem 6.1. $\log_2 \frac{\chi(S_\Delta)}{\Delta} \sim -\sqrt{2 \log_2 \Delta}$ *as* $\Delta \to \infty$.

To start the proof we need upper and lower bounds on $\mathrm{cov}(r)$.

Lemma 6.2. *On the product space* S_Δ*, for* $m \geq 2$
(a) $\mathrm{cov}((m + 2)2^{-m}\Delta) \leq 2^{m(m+1)/2}$.
(b) $\mathrm{cov}(2^{-m}\Delta) \geq (m - 1)! 2^{(m-1)(m-2)/2}$.

Proof: Given Δ and $j \geq 0$, the set

$$\mathcal{B}_\Delta(j) := \left\{ \frac{\Delta}{2^{j+1}}, \frac{3\Delta}{2^{j+1}}, \frac{5\Delta}{2^{j+1}}, \ldots, \frac{(2^{j+1} - 1)\Delta}{2^{j+1}} \right\}$$

consists of 2^j elements, and each point in $[0, \Delta]$ is within distance $\Delta/2^{j+1}$ from the closest point of $\mathcal{B}_\Delta(j)$. Now for $m \geq 1$ consider the subset $\mathcal{B}_\Delta$ of S_Δ defined by

$$\mathcal{B}_\Delta := \{\mathbf{x} = (x_i) : x_i \in \mathcal{B}_\Delta(m + 1 - i) \text{ for all } 1 \leq i \leq m, \ x_i = \Delta/2 \text{ for all } i > m\}.$$

The cardinality of $\mathcal{B}_\Delta$ equals $\prod_{i=1}^m 2^{m+1-i} = 2^{m(m+1)/2}$. For each element of S_Δ, the distance to the closest element of $\mathcal{B}_\Delta$ is at most

$$\sum_{i=1}^{m} 2^{-i}\Delta/2^{m+2-i} + \sum_{i=m+1}^{\infty} 2^{-i}\Delta/2 = (m + 2)\Delta/2^m.$$

This proves part (a). For part (b), define

$$F(z) := \mathbb{P}(\sum_{i \geq 1} 2^{-i} U_i \leq z) \text{ for i.i.d. } U[0, 1] \text{ summands } U_i.$$

Consider the uniform distribution on S_Δ, that is the distribution ν of i.i.d. Uniform$(0,\Delta)$ random variables $(V_i, 1 \le i < \infty)$. Write $\mathbf{c} = (\Delta/2, \Delta/2, \ldots)$. For $r \le \Delta/2$ and any $\mathbf{x} \in S_\Delta$,

$$\begin{aligned} \nu(\mathrm{ball}(\mathbf{x}, r)) &\le \nu(\mathrm{ball}(\mathbf{c}, r)) \\ &= \mathbb{P}\left(\sum_i 2^{-i}|V_i - \Delta/2| \le r\right) \\ &= \mathbb{P}\left(\sum_i 2^{-i}\frac{|V_i - \Delta/2|}{\Delta/2} \le \frac{2r}{\Delta}\right) \\ &= F(2r/\Delta) \end{aligned}$$

the final equality because $\frac{|V_i - \Delta/2|}{\Delta/2}$ has $U(0,1)$ distribution. So the entire space S_Δ cannot be covered by fewer than $1/F(2r/\Delta)$ balls of radius r, implying

$$\mathrm{cov}(r) \ge 1/F(2r/\Delta). \tag{6.1}$$

We now need to upper bound the function $F(z)$ as $z \downarrow 0$. Consider $z = 2^{-m}$ for $m \ge 1$, and note

$$F(2^{-m}) \le \mathbb{P}(\sum_{i=1}^m 2^{-i}U_i \le 2^{-m}). \tag{6.2}$$

Consider the event $A_m := \{\max_{1 \le i \le m} 2^{-i}U_i \le 2^{-m}\}$, for which

$$\mathbb{P}(A_m) = \prod_{i=1}^m (2^i 2^{-m}) = 2^{-m(m-1)/2}.$$

Conditional on A_m the sequence $(2^{-i}U_i/2^{-m}, 1 \le u \le m)$ has i.i.d. $U(0,1)$ distribution, and by a textbook exercise (Pitman (1993) Exercise 5.4.18d) the probability that the sum of m such random variables is less than 1 equals $1/m!$. So we obtain the exact formula

$$\mathbb{P}(\sum_{i=1}^m 2^{-i}U_i \le 2^{-m}) = \frac{2^{-m(m-1)/2}}{m!}. \tag{6.3}$$

From (6.1) we have $\mathrm{cov}(\Delta 2^{-m}) \ge 1/F(2^{-(m-1)})$ and then (b) follows via (6.2) and (6.3). $\square$

Proof of Theorem 6.1. Proposition 5.1 and Lemma 6.2(a) imply

$$\chi(S_\Delta) \le \min_{m \ge 1} [a(m) + b(m)], \tag{6.4}$$

where

$$a(m) := (m+2)\Delta/2^m; \quad b(m) := 2^{m(m+1)/2}(1 + \log 2^{m(m+1)/2}).$$

Choose $m(\Delta) \sim \sqrt{2(1-\varepsilon)\log_2 \Delta}$ for small $\varepsilon > 0$. Then

$$\log_2 \frac{a(m(\Delta))}{\Delta} \sim -m(\Delta); \quad \log_2 \frac{b(m(\Delta))}{\Delta} \sim -\varepsilon \log_2 \Delta$$

and using $m(\Delta)$ in (6.4) is enough to establish the upper limit in Theorem 6.1. Similarly, Proposition 5.3 and Lemma 6.2(b) imply

$$\chi(S_\Delta) \ge \sup_m \{2^{-m}\Delta/3 : (m-1)! 2^{(m-1)(m-2)/2} > 3 \cdot 2^{-m}\Delta\}. \tag{6.5}$$

Choose $m(\Delta) \sim \sqrt{2(1+\varepsilon)\log_2 \Delta}$ for small $\varepsilon > 0$. Then

$$\log_2 \frac{(m(\Delta)-1)! 2^{(m(\Delta)-1)(m(\Delta)-2)/2}}{\Delta} \sim \varepsilon \log_2 \Delta; \quad \log_2 \frac{3 \cdot 2^{-m(\Delta)}\Delta}{\Delta} \sim -m(\Delta)$$

and so the constraint in (6.5) is satisfied for large Δ. Using $m(\Delta)$ in (6.4) gives

$$\log_2 \frac{\chi(S_\Delta)}{\Delta} \geq -m(\Delta) \text{ for large } \Delta$$

establishing the lower limit in Theorem 6.1.

7. Discussion

7.1. *Open problems for the general growth model.*

- Are there easily checkable conditions to ensure that $c^*/\mathbb{E}C$ is small, so that Proposition 3.1 is informative?
- Can one improve the general upper and lower bounds on $\chi(S)$ in section 5? In particular, can $\chi(S)$ be more sharply related to some measure of *entropy* of the metric space S (see e.g. Leinster and Roff (2021) for possible notions of entropy)?
- On a general compact metric space there is no canonical definition of the *uniform* distribution, but any definition of *entropy* (as above) can be used to define *uniform* as *maximal entropy*. Analogously, one motivation for considering $\chi(S)$ is that the minimizing distribution gives a certain, explicitly probabilistic, definition that may serve as proxy for *uniform* is some senses.
- Estimate $\chi(S)$ for specific spaces, in some asymptotic sense analogous to our section 6 estimate for S_Δ. As well as other classical compact spaces familiar from analysis, one can consider a finite graph with edge lengths, with the metric of shortest route length. Moreover there are *random* metric spaces of contemporary interest in probability, such as the "mean-field model of distance" (Aldous and Steele (2004)), the Brownian CRT (Goldschmidt (2020)), or the Brownian map (Le Gall (2013)).
- For μ attaining the minimum $\min_\mu \mathbb{E}_\mu C$, do we always have weak concentration? That is, is there a function $\psi(\Delta) \downarrow 0$ as $\Delta \uparrow \infty$ such that on every connected compact metric space, for the standardized growth model,

  $$\mathrm{var}_\mu \left(\frac{C}{\mathbb{E}_\mu C} \right) \leq \psi(\Delta)$$

  for the minimizing μ?
- Is there an effective algorithmic procedure for finding a minimizing μ? This seems loosely similar to the well-studied *k-median problem* discussed in Charikar et al. (2002).
- If S is a compact group, with a metric invariant under the group action, then is the uniform (Haar) measure the minimizing measure?

Regarding the final problem above, it can be shown that, on the circle of integer circumference L, for the fixed-radius model with $r = 1/2$, the mean cover time for seed distribution μ uniform on L evenly-spaced points is smaller than that for μ uniform on the circle (the discrete analog is noted in Falgas-Ravry et al. (2020) Example 4.1). We do not know if this type of example is a counter-example in the growth model; if so, replace by an asymptotic ($\Delta \to \infty$) conjecture.

7.2. *Models for spread of information.* The growth model can be regarded as an extremely simplistic model for the spread of information or the spread of an epidemic, a field with a huge literature studying models on graphs or Euclidean space: for instance Draief and Massoulié (2010); Kiss et al. (2017); Riley et al. (2015). A related growth model in two dimensions, where seeds arrive (instead of as a constant-rate process) as a Poisson process whose rate is the current occupied area, is studied in Aldous (2013) and Chatterjee and Durrett (2011).

7.3. *Other uses of the two general bounds.* We have used two general methods – the random subset cover bound (Proposition 2.2) and the monotonicity bound (Proposition 3.2) – which are in principle applicable in very general covering-like contexts to establish weak concentration bounds in general settings without calculating the expectation of the covering time. We provide some history of these methods below, and speculate that there may be other applications not yet explored.

The random subset cover bound, Proposition 2.2, for general i.i.d. random subsets of a set, was given in Aldous (1991) as part of the proof of a weak concentration bound for the Markov chain cover time C_{MC}. In the Markov chain context, the i.i.d. subsets arise as excursions from a given state. In the result, the essential condition is that the maximum mean hitting time to any single state is $o(\mathbb{E}C_{MC})$. In that sense the bound is closely analogous to the bounds in this article. In the 30 years since Aldous (1991), study of random walk cover times has entered a more sophisticated phase based on the Ding et al. (2012) discovery of its connection with Gaussian free fields and Talagrand's theory of majorizing measures. In contrast, the program of using general results for i.i.d. random subsets as part of analysis of specific contexts within covering seems not to have been developed until the recent work of Falgas-Ravry et al. (2020). That paper discusses known results in combinatorial settings, develops new general results and applies them to several topics: connectivity in random graphs; covering a square with random discs; covering the edges of a graph by spanning trees, and matroids by bases; and random k-SAT.

The monotonicity bound, Proposition 3.2, was given in Aldous (2016b) as a tool for establishing weak concentration for first passage percolation times on general graphs. It was also used in Aldous (2016a) for weak concentration of the time of emergence of the giant component in bond percolation on general graphs. Both contexts involve hitting time of an increasing set-valued Markov process, as does our application in section 3.2.

Acknowledgements

I thank an anonymous referee for the reference to Chiu (1995) and thoughtful comments on exposition.

References

Aldous, D. The incipient giant component in bond percolation on general finite weighted graphs. *Electron. Commun. Probab.*, **21**, Paper No. 68, 9 (2016a). MR3564215.

Aldous, D. and Steele, J. M. The objective method: probabilistic combinatorial optimization and local weak convergence. In *Probability on discrete structures*, volume 110 of *Encyclopaedia Math. Sci.*, pp. 1–72. Springer, Berlin (2004). MR2023650.

Aldous, D. J. Threshold limits for cover times. *J. Theoret. Probab.*, **4** (1), 197–211 (1991). MR1088401.

Aldous, D. J. When knowing early matters: gossip, percolation and Nash equilibria. In *Prokhorov and contemporary probability theory*, volume 33 of *Springer Proc. Math. Stat.*, pp. 3–27. Springer, Heidelberg (2013). MR3070464.

Aldous, D. J. Weak concentration for first passage percolation times on graphs and general increasing set-valued processes. *ALEA Lat. Am. J. Probab. Math. Stat.*, **13** (2), 925–940 (2016b). MR3550985.

Charikar, M., Guha, S., Tardos, E., and Shmoys, D. B. A constant-factor approximation algorithm for the k-median problem. *J. Comput. System Sci.*, **65** (1), 129–149 (2002). MR1946291.

Chatterjee, S. and Durrett, R. Asymptotic behavior of Aldous' gossip process. *Ann. Appl. Probab.*, **21** (6), 2447–2482 (2011). MR2895421.

Chiu, S. N. Limit theorems for the time of completion of Johnson-Mehl tessellations. *Adv. in Appl. Probab.*, **27** (4), 889–910 (1995). MR1358899.

Chiu, S. N., Stoyan, D., Kendall, W. S., and Mecke, J. *Stochastic geometry and its applications.* Wiley Series in Probability and Statistics. John Wiley & Sons, Ltd., Chichester, third edition (2013). ISBN 978-0-470-66481-0. MR3236788.

Ding, J., Lee, J. R., and Peres, Y. Cover times, blanket times, and majorizing measures. *Ann. of Math. (2)*, **175** (3), 1409–1471 (2012). MR2912708.

Draief, M. and Massoulié, L. *Epidemics and rumours in complex networks*, volume 369 of *London Mathematical Society Lecture Note Series.* Cambridge University Press, Cambridge (2010). ISBN 978-0-521-73443-1. MR2582458.

Falgas-Ravry, V., Larsson, J., and Markström, K. Speed and concentration of the covering time for structured coupon collectors. *Adv. in Appl. Probab.*, **52** (2), 433–462 (2020). MR4123642.

Goldschmidt, C. Scaling limits of random trees and random graphs. In *Random graphs, phase transitions, and the Gaussian free field*, volume 304 of *Springer Proc. Math. Stat.*, pp. 1–33. Springer, Cham (2020). MR4043223.

Hall, P. *Introduction to the theory of coverage processes.* Wiley Series in Probability and Mathematical Statistics: Probability and Mathematical Statistics. John Wiley & Sons, Inc., New York (1988). ISBN 0-471-85702-5. MR973404.

Kiss, I. Z., Miller, J. C., and Simon, P. L. *Mathematics of epidemics on networks. From exact to approximate models*, volume 46 of *Interdisciplinary Applied Mathematics.* Springer, Cham (2017). ISBN 978-3-319-50804-7; 978-3-319-50806-1. MR3644065.

Le Gall, J.-F. Uniqueness and universality of the Brownian map. *Ann. Probab.*, **41** (4), 2880–2960 (2013). MR3112934.

Leinster, T. and Roff, E. The maximum entropy of a metric space. *Q. J. Math.*, **72** (4), 1271–1309 (2021). MR4350150.

Møller, J. Random Johnson-Mehl tessellations. *Adv. in Appl. Probab.*, **24** (4), 814–844 (1992). MR1188954.

Penrose, M. D. Random Euclidean coverage from within. *ArXiv Mathematics e-prints* (2021). arXiv: 2101.06306.

Pitman, J. *Probability.* Springer Texts in Statistics. Springer-Verlag New York (1993). ISBN 978-0-387-97974-8; 978-0-387-94594-1. DOI: 10.1007/978-1-4612-4374-8.

Riley, S., Eames, K., Isham, V., Mollison, D., and Trapman, P. Five challenges for spatial epidemic models. *Epidemics*, **10**, 68–71 (2015). DOI: 10.1016/j.epidem.2014.07.001.