BLANKET TIMES AND THE GAUSSIAN FREE FIELD

ZACHARY HAMAKER

ABSTRACT. The blanket time of a random walk on a graph G is the expected time until the proportion of time spent at each vertex approximates the stationary distribution. In 1996, Winkler and Zuckerman conjectured that the blanket time is of the same order as the cover time. Ding, Lee and Peres proved this conjecture in 2010 by relating both terms to the maximum of the Gaussian free field on G. We outline the connection between the blanket time and the Gaussian free field.

1. Covert Times and Blanket Times

We begin by defining the blanket time of a random walk, discussing an application, historical aspects, and its relationship to the cover time. Before doing so, we must review several important concepts. Let $G = (V, E)$ be an undirected graph with n vertices and m edges, and let $\{X_t\}$ be a random walk on G. For $\tau_{cov}^v = \min\{t : V \subset \{X_s\}_{s=0}^t\}$ where $X_0 = v$, we define the *cover time*

$$t_{cov} = \max_v \mathbb{E}(\tau_{cov}^v).$$

Here, τ_{cov}^v is the (random) exact time it takes for $\{X_t\}$ to visit every state when starting at v, while t_{cov} is the expected time to visit every state, starting from the worst possible v. Recall the *hitting time* $H_u(v)$ is the expected time for the random walk to travel from u to v. We state the best bound for the cover time:

Theorem 1.1 (Matthews' Theorem)**.** *For any G with $|V| = n$,*

$$\min_{u,v} H_u(v) n \log n \leq t_{cov} \leq \max_{u,v} H_u(v) n \log n.$$

The cover time can be viewed as a measurement of certain types of connectivity for the random walk. Familiar concepts in this vein include the hitting time and the *commute time* $\kappa_{uv} = H_u(v) + H_v(u)$, which measures the expected time to go from u to v and back. The blanket time time is a similar quantity, measuring the expected time it takes to visit every state approximately proportional to the stationary distribution. In order to define the blanket time, we must quantify the amount of time spent at each state. For the remainder of the paper, it is assumed that X_0 has some known initial distribution ν. The *local time* at v is

$$L_t^v = \frac{\mathbb{E}(\sum_{s=0}^t \mathbb{1}_{X_s=v})}{\pi_v}.$$

This quantity represents the proportion of time spent at v up to time t. Note

$$\mathbb{E}(\sum_{s=0}^t \mathbb{1}_{X_s=v}) \approx \pi_v t$$

Date: March 5, 2013.

for t large, so $L_t^v \to 1$ as $t \to \infty$. For $\delta \in (0,1)$, the *δ-blanket time* is

$$t_{bl}^{\delta} = \min\{t : \max_{u,v} \frac{L_t^u}{L_t^v} > \delta\}.$$

Since $L_t^u, L_t^v \to 1$, this must hold for large enough t.

We motivate the definition of the δ-blanket time with an example. We present an algorithm inspired by the Google page rank algorithm. Our graph is the internet, where each edge is represented by a link. Note this graph structure is directed, while our results will only apply to non-directed graphs (reversible Markov chains). We start a web surfer at a random website who clicks on links at random and record each site she visits. With some positive probability, she grows bored and goes to a new random website. This defines a Markov chain on the internet. A website v has *page rank* proportional to the amount of time such a walker spends on it, i.e. proportional to π_v. We wish to estimate this by examining the progression of one such surfer and ranking a website according to the number of visits. The blanket time represents the time we expect this estimate to become relatively accurate.

The blanket time was first introduced in [7] by Winkler and Zuckerman. Here, the following bold conjecture is made:

Conjecture 1. *Let G be a graph. For each $\delta \in (0,1)$, there exists C_δ depending solely on δ such that*

$$t_{bl}^{\delta} \le C_\delta t_{cov},$$

or equivalently

$$t_{bl}^{\delta} \asymp_\delta t_{cov}.$$

In [7], Winkler and Zuckerman verified the conjecture when $t_{cov} \asymp n \log n$. This is true for most graphs. Additionally, they verify the conjecture for cycles. In [4] Kahn, Kim, Lovasz and Vu show that

$$t_{bl}^{\delta} = O(t_{cov}(\log\log n)^2).$$

This result is probably good enough for Google, but not quite enough. Finally, in [2], the conjecture was proved by Ding, Lee and Peres.

Theorem 1.2. *For G a graph,*

$$t_{cov} \asymp m\mathbb{E}(\max_{v \in V} \eta_v)^2 \asymp_\delta t_{bl}^{\delta}$$

where $\{\eta_v\}$ is the Gaussian free field on G.

The remainder of this paper is devoted to explaining the presence of this middle term and outlining its role in the proof of this theorem.

2. The Gaussian connection to Blanket Time

We begin by arguing heuristically that the blanket time will be related to the maximum of some Gaussians for t large. First, let $\tau_{cov}^{\mathcal{O}}$ be the (random) time to cover G and return to the initial vertex. Note this is an excursion, hence $L_{\tau_{cov}^{\mathcal{O}}}^v$ is a random variable, and we can construct a sequence of identically distributed such variables within our random walk. Pick k large. By the Central Limit Theorem, the sum of k such variables will be approximately normal. The rate of convergence will depend on the variance of $L_{\tau_{cov}^{\mathcal{O}}}^v$. For any given v, we can pick k such that we expect the average to be as close as we like to the expected excursion length. However, in order to gain control of the ratios, we must have some control over the maximum of these Gaussian averages over all v. Additionally, since $L^u_{\tau^{\mathcal{O}}_{cov}}$ and $L^v_{\tau^{\mathcal{O}}_{cov}}$ are not independent, we must be able to account for some correlation between Gaussian random variables. We outline this connection, both in a more formal and fully rigorous fashion.

First, we introduce Gaussian processes. We say the random variables $\{g_x\}_{x \in S}$ form a *Gaussian process* if any finite linear combination $\sum_{i=1}^k c_i g_{x_i}$ has a Gaussian distribution. A key property of a Gaussian process is that it can be defined by specifying the covariance $K(x, y)$ of each $x, y \in S$ (see e.g. [1]). These processes are one of the major areas of study in probability. Important examples include Brownian motion, the Brownian bridge, the Ornstein-Uhlenbeck semi-group and the Gaussian free field. Because they are so well understood, we will gain an enormous amount of tools by relating our problem to the Gaussian free field of G.

For a given metric space (S, d), we can define a Gaussian process by specifying

$$\sqrt{\mathbb{E}(|g_x - g_y|^2)} = d(x, y).$$

By requiring $g_{x_0} = 0$ for some $x_o \in S$, we can recover the covariance matrix. Note the commute time κ clearly defines a metric on V. Additionally, since the effective resistance R is proportional to the commute time, we see it forms a metric as well. The *Gaussian free field* of a graph G is the Gaussian process $\{\eta_v\}_{v \in V}$ determined by specifying

$$\mathbb{E}(|\eta_u - \eta_v|^2) = R_{uv},$$

or equivalently the Gaussian process derived from the metric space $(V, \sqrt{R})$.

We now relate the Gaussian free field to the local time. Fix $v_0 \in V$ and let $\Gamma_{v_0}(u, v) = \mathbb{E}_x(L^v_{H_u(v)})$. It is a theorem that the Gaussian free field of G has covariance function $\Gamma_{v_0}(u, v)$ (this is not obviously symmetric). This connection be strengthened greatly. Define the *inverse local time* $\tau(t)$ of $v_0 \in V$ by

$$\tau(t) = \min\{s : L^{v_0}_s > t.$$

This is the first time s that the chain has spent t/π_{v_0} time at state v_0. If we consider instead the continuous random walk on G, we have the generalized second Ray-Knight isomorphism theorem, as stated in [3].

Theorem 2.1 (Isomorphism Theorem)**.**

$$\left\{ L^x_{\tau(t)} + \frac{1}{2}\eta_x^2 : x \in V \right\} =_d \left\{ \frac{1}{2}(\eta_x + \sqrt{2t})^2 : x \in V \right\}$$

where $=_d$ signifies equality in distribution.

This explicit formula is the key tool in proving, for the continuous walk, that

$$t_{bl} \leq A_\delta m(\mathbb{E}(\max \eta_x))^2$$

where $A_\delta \approx 1/(1-\delta)^2$. The basic idea of the proof is as follows. For t sufficiently large compared to $\mathbb{E}(\max \eta_x)$, we will have $L^x_{\tau(t)}$ and t much larger than η_x^2 for all x. Therefore, we can explicitly relate $L^x_{\tau(t)}$ and t via the isomorphism theorem and show that the error introduced by the remaining terms of the relation is negligible. In particular, we must show this holds for both $\min_x L^x_{\tau(t)}$ and $\max_x L^x_{\tau(t)}$, so that their ratio can be controlled. The tool used to do so the following concentration inequality (see e.g. [5]).

Lemma 2.2. *Let $\{\eta_x\}_{x\in S}$ be a Gaussian process and define $\sigma = \sup_x \sqrt{\mathbb{E}(\eta_x^2)}$. Then for all $\alpha > 0$,*

$$\mathbb{P}(|\sup_x \eta_x - \mathbb{E}(\sup_x \eta_x)| > \alpha) \leq 2\exp(-\alpha^2/2\sigma^2).$$

Upon completing this argument, the we must then show that the blanket time of the discrete and continuous walks are of the same order. This result is quite technical, though intuitively clear. Making this outline rigorous is a non-trivial endeavor, and we omit the proof.

3. Analogies to the cover time

The connection between the Gaussian free field and the cover time are significantly more challenging to understand. We paint an impressionistic picture of why one would expect the two to be related, and a few words on how the connection is demonstrated. We draw analogies between the Gaussian free field and the Matthews' bound Theorem 1.1. The upper bound from Theorem 1.1 can be related to a union bound for the Gaussian free field: compare

$$t_{cov} \leq \max_{u,v} H_u(v) n \log n$$

and

$$(\mathbb{E}(M))^2 \leq 2 \max_x \mathbb{E}(X_x^2) \log n$$

where $M = \max_x \eta_x$. The lower bound from Theorem 1.1 has a similarly strong analogy in the Gaussian setting.

Theorem 3.1 (Sudakov minorization)**.** *Let $\{\eta_x\}_{x\in S}\}$ be a Gaussian process with $|S| = n$ and $d(x,y) \geq \alpha$ for $x \neq y$. Then*

$$\mathbb{E}(M) \geq \alpha\sqrt{\log n}.$$

This result can be found in [6]. While these analogies are not so overt as to be instantly compelling, there are several more. One relates an entropic result to the cover time, while another relates concentration results of Gaussian processes to concentration results of local times. The actual proof is quite difficult, relying on an important result of Talagrand's known as majorizing measures.

We outline the strategy of the proof. Consider to vertices u and v close to each other with respect to R. If L_t^u is large, we would expect that a substantial proportion of the time between visits to u will be spent at v, to L_t^v should be large as well, i.e. L_t^u and L_t^v should be highly correlated. For u and v far apart with respect to the metric R, we expect that their local times at time τ_{cov} are nearly independent. The proof uses the isomorphism theorem, but with greater delicacy, as we must show some local time is exactly zero. This error is controlled by grouping the local times of vertices according to their distance with respect to R. However, it is not enough to group at the global level. We must construct such groupings at many scales. Majorizing measures is a technique for doing this optimally with Gaussian processes. The proof relies on constructing an analogy to this process for the local times, and using results from the literature on Gaussian processes to make this connection explicit.

References

[1] Patrick Billingsley. *Probability and measure*, volume 939. Wiley, 2012.

[2] Jian Ding, James R Lee, and Yuval Peres. Cover times, blanket times, and majorizing measures. In *Proceedings of the 43rd annual ACM symposium on Theory of computing*, pages 61–70. ACM, 2011.

[3] Nathalie Eisenbaum, Haya Kaspi, Michael B Marcus, Jay Rosen, and Zhan Shi. A ray-knight theorem for symmetric markov processes. *The Annals of Probability*, 28(4):1781–1796, 2000.

[4] Jeff Kahn, Jeong Han Kim, Laszlo Lovasz, and Van H Vu. The cover time, the blanket time, and the matthews bound. In *Foundations of Computer Science, 2000. Proceedings. 41st Annual Symposium on*, pages 467–475. IEEE, 2000.

[5] Michel Ledoux. *The concentration of measure phenomenon*, volume 89. Amer Mathematical Society, 2001.

[6] Michel Ledoux and Michel Talagrand. *Probability in Banach Spaces: isoperimetry and processes*, volume 23. Springer, 2011.

[7] Peter Winkler and David Zuckerman. Multiple cover time. *Random Structures Algor. 9 (1996), 403*, 411, 1996.