



Vol. 16 (2011), Paper no. 62, pages 1693–.

Journal URL

<http://www.math.washington.edu/~ejpecp/>

## Gaussian upper bounds for heat kernels of continuous time simple random walks

Matthew Folz\*

### Abstract

We consider continuous time simple random walks with arbitrary speed measure  $\theta$  on infinite weighted graphs. Write  $p_t(x, y)$  for the heat kernel of this process. Given on-diagonal upper bounds for the heat kernel at two points  $x_1, x_2$ , we obtain a Gaussian upper bound for  $p_t(x_1, x_2)$ . The distance function which appears in this estimate is not in general the graph metric, but a new metric which is adapted to the random walk. Long-range non-Gaussian bounds in this new metric are also established. Applications to heat kernel bounds for various models of random walks in random environments are discussed.

**Key words:** random walk, heat kernel, Gaussian upper bound, random walk in random environment.

**AMS 2010 Subject Classification:** Primary 60G50, 30K08; Secondary: 60K37.

Submitted to EJP on February 10, 2011, final version accepted August 10, 2011.

---

\*Department of Mathematics, The University of British Columbia, 1984 Mathematics Road, Vancouver, B.C., Canada, V6T 1Z2. [mfolz@math.ubc.ca](mailto:mfolz@math.ubc.ca). Research supported by an NSERC Alexander Graham Bell Canada Graduate Scholarship.

# 1 Introduction

Let  $\Gamma = (G, E)$  be an unoriented graph. We assume that  $\Gamma$  is connected, contains neither loops nor multiple edges, is locally finite, and countably infinite. Let  $d$  be the usual graph metric; given  $x, y \in G$ ,  $d(x, y)$  is equal to the number of edges in the shortest (geodesic) path between  $x$  and  $y$ . We write  $B(x, r) := \{y \in G : d(x, y) \leq r\}$  for the closed ball of radius  $r$  in the metric  $d$ .

We assume that  $\Gamma$  is a weighted graph, so that associated with each  $(x, y) \in G \times G$  is a non-negative edge weight  $\pi_{xy}$  which is symmetric ( $\pi_{xy} = \pi_{yx}$  for  $x, y \in G$ ) and satisfies  $\pi_{xy} > 0$  if and only if  $\{x, y\} \in E$ . The edge weights can be extended to a measure on  $G$  by setting  $\pi_x := \pi(\{x\}) := \sum_{y \in G} \pi_{xy}$  for  $x \in G$ , and this extends to all subsets of  $G$  by countable additivity.

Let  $(\theta_x)_{x \in G}$  be an arbitrary collection of positive vertex weights. We consider the continuous-time simple random walk  $(X_t)_{t \geq 0}$ , which has generator  $\mathcal{L}_\theta$ , given by

$$(\mathcal{L}_\theta f)(x) := \frac{1}{\theta_x} \sum_{y \sim x} \pi_{xy} (f(y) - f(x)).$$

Regardless of the choice of  $(\theta_x)_{x \in G}$ , the jump probabilities of these processes are  $P(x, y) = \pi_{xy}/\pi_x$ ; the various walks corresponding to different choices of  $(\theta_x)_{x \in G}$  will be time-changes of each other.

Two specific choices of the vertex weights  $(\theta_x)_{x \in G}$  arise frequently. The first is the choice  $\theta_x := \pi_x$ , which yields a process called the constant-speed continuous time simple random walk (CSRW). The CSRW may also be constructed by taking a discrete-time simple random walk on  $(\Gamma, \pi)$ , which we denote by  $(X_n)_{n \in \mathbb{Z}_+}$ , together with an independent rate 1 Poisson process  $(N_t)_{t \geq 0}$ ; the CSRW is the process  $Y_t := X_{N_t}$ .

The second choice,  $\theta_x \equiv 1$ , yields a stochastic process referred to as the variable-speed continuous time simple random walk (VSRW). This walk has the same jump probabilities as the CSRW, but instead of waiting for an exponentially distributed time with mean 1 at a vertex  $x$  before jumping, the VSRW waits for an exponentially distributed time with mean  $\pi_x^{-1}$ . As discussed in [4], the VSRW may explode in finite time.

Associated with the process  $(X_t)_{t \geq 0}$  is a semigroup  $(P_t)_{t \geq 0}$  defined by  $(P_t f)(x) := \mathbb{E}^x f(X_t)$ , and which possesses a density  $p_t(x, y)$  with respect to the measure  $\theta$ , defined by

$$p_t(x, y) := \frac{1}{\theta_y} \mathbb{P}^x(X_t = y).$$

This function is also called the *heat kernel* of the process  $(X_t)_{t \geq 0}$ .

We discuss here an alternative construction of the heat kernel which will be used in Section 3; this closely follows the discussion in [25]. Let  $(G_n)_{n \in \mathbb{Z}_+}$  be an increasing sequence of finite

connected subsets of  $G$  whose limit is  $G$ . Given  $U \subset G$ , we denote the first hitting time of  $U$  by  $T_U := \inf\{s \geq 0 : X_s \in U\}$ .

For each  $n \in \mathbb{Z}_+$ , we define the killed heat kernel  $p_t^{(G_n)}(x, y)$  by

$$p_t^{(G_n)}(x, y) := \frac{1}{\theta_y} \mathbb{P}^x(X_t = y, T_{G \setminus G_n} > t).$$

This object satisfies the following conditions:

$$\begin{cases} \frac{\partial}{\partial t} p_t^{(G_n)}(x, y) = (\mathcal{L}_\theta)_y p_t^{(G_n)}(x, y) & \text{if } x, y \in G_n, \\ p_t^{(G_n)}(x, y) = 0 & \text{if } x \in G \setminus G_n \text{ or } y \in G \setminus G_n, \\ p_t^{(G_n)}(x, y) \geq 0 & \text{for all } x, y \in G. \end{cases}$$

Furthermore, we have that for all  $x, y \in G$  and  $t > 0$  and  $n \in \mathbb{Z}_+$ ,

$$\begin{aligned} p_t^{(G_n)}(x, y) &\leq p_t^{(G_{n+1})}(x, y), \\ \lim_{n \rightarrow \infty} p_t^{(G_n)}(x, y) &= p_t(x, y). \end{aligned}$$

We will also need a distance function on  $G \times G$  which is adapted to the vertex weights  $(\theta_x)_{x \in G}$ ; this will be the metric which appears in our heat kernel estimates. In general, Gaussian upper bounds for the heat kernel do not hold if one only considers the graph metric, see Remark 6.6 of [4] for an example. Let  $d_\theta(\cdot, \cdot)$  be a metric which satisfies

$$\begin{cases} \frac{1}{\theta_x} \sum_{y \sim x} \pi_{xy} d_\theta^2(x, y) \leq 1 & \text{for all } x \in G, \\ d_\theta(x, y) \leq 1 & \text{whenever } x, y \in G \text{ and } x \sim y. \end{cases} \quad (1.1)$$

It is not difficult to verify that such metrics always exist. We write  $B_\theta(x, r) := \{y \in G : d_\theta(x, y) \leq r\}$  for the closed ball of radius  $r$  in the metric  $d_\theta$ ; it should be noted that  $B_\theta(x, r)$  may contain infinitely many points for some choices of  $x \in G$  and  $r > 0$ , or, equivalently, points arbitrarily far from  $x$  in the graph metric. Note that for the CSRW, the graph metric always satisfies both of the above conditions.

The use of metrics different from the graph metric in heat kernel estimates was initiated by Davies in [9], and this metric is similar to the metrics considered there. These metrics are closely related to the intrinsic metric associated with a given Dirichlet form; some details on the latter may be found in [18]. Recent work using similar metrics includes [4], [12], [15], and [20].

We will need the following condition:

**Definition:** A monotonically increasing function  $g : (a, b) \rightarrow (0, \infty)$  is  $(A, \gamma)$ -regular on  $(a, b)$  ( $A \geq 1, \gamma > 1, 0 \leq a < b \leq \infty$ ) if for all  $a < t_1 < t_2 < \gamma^{-1}b$ , the inequality

$$\frac{g(\gamma t_1)}{g(t_1)} \leq A \frac{g(\gamma t_2)}{g(t_2)}$$

holds. If  $a = 0$  and  $b = \infty$ , then we say that  $g$  is  $(A, \gamma)$ -regular.

For appropriate values of  $A$  and  $\gamma$ , this set of functions includes polynomial functions such as  $ct^{d/2}$ , exponential functions such as  $c \exp(Ct^\alpha)$ , and various piecewise combinations of  $(A, \gamma)$ -regular functions such as  $c_1 t^{d_1/2} \mathbf{1}_{(0, T]} + c_2 t^{d_2/2} \mathbf{1}_{(T, \infty)}$ , where  $c_1$  and  $c_2$  are chosen to ensure that the resulting function is continuous.

Our work will assume that one has already obtained on-diagonal upper bound for the heat kernel at two points  $x_1, x_2 \in G$ ; that is, there are functions  $f_1, f_2$  which are  $(A, \gamma)$ -regular on  $(a, b)$  such that, for all  $t > 0$  and  $i \in \{1, 2\}$ ,

$$p_t(x_i, x_i) \leq \frac{1}{f_i(t)}. \tag{1.2}$$

On-diagonal bounds such as (1.2) have been studied in considerable detail in both discrete and continuous settings, and follow from a variety of analytic inequalities, such as a Sobolev inequality [22], a Nash inequality [6], a log-Sobolev inequality [10], or a Faber-Krahn inequality [14]. Generally, these methods yield a uniform upper bound, valid for all  $x \in G$ . In the present setting of graphs, one may also use isoperimetric inequalities on general graphs, or volume growth estimates in the particular case of Cayley graphs of groups; details are in [2], [23], and [24].

In the context of Riemannian manifolds, Grigor'yan has shown that any Riemannian manifold  $M$  which satisfies an on diagonal upper bound at two points  $x, y \in M$  admits a Gaussian upper bound for the heat kernel  $q_t(x, y)$ . His result is as follows:

**Theorem A.** [13] *Let  $x_1, x_2$  be distinct points on a smooth Riemannian manifold  $M$ , and suppose that there exist  $(A, \gamma)$ -regular functions  $f_1, f_2$  such that, for all  $t > 0$  and  $i \in \{1, 2\}$ ,*

$$q_t(x_i, x_i) \leq \frac{1}{f_i(t)}. \tag{1.3}$$

*Then for any  $D > 2$  and all  $t > 0$ , the Gaussian upper bound*

$$q_t(x_1, x_2) \leq \frac{4A}{(f_1(\delta t)f_2(\delta t))^{1/2}} \exp\left(-\frac{d^2(x_1, x_2)}{2Dt}\right) \tag{1.4}$$

*holds, where  $\delta = \delta(D, \gamma)$ .*

One remarkable aspect of this result is that it only requires on-diagonal bounds at the points  $x_1$  and  $x_2$ . Prior to [13], there are several proofs of Gaussian upper bounds for the heat kernel on manifolds, but these papers involve more restrictive hypotheses on the underlying manifold, in addition to requiring on-diagonal heat kernel estimates which hold for all  $x \in G$ . In practice, the upper bounds (1.3) are often obtained from a uniform upper heat kernel bound using the techniques described previously, such as a Nash inequality. However, Theorem A leaves open the possibility of obtaining Gaussian upper bounds for  $q_t(x_1, x_2)$  using only the restricted information in (1.3).

For the discrete time SRW on  $(\Gamma, \pi)$ , one may again assume a uniform upper bound for the heat kernel, and obtain a Gaussian upper bound from it. This was done first by Hebisch and Saloff-Coste in [17] using functional-analytic techniques, and later by Coulhon, Grigor'yan, and Zucca in [7], using techniques analogous to the ones used by Grigor'yan in [13].

In discrete time, a SRW cannot move further than distance  $n$  in time  $n$ , and hence  $p_n(x, y) = 0$  whenever  $d(x, y) > n$ , whereas a continuous time random walk has no such constraint. For the CSRW on  $\mathbb{Z}$  with the standard weights, the heat kernel does not exhibit Gaussian decay if  $d(x, y) \gg t$  (see [5]), and as a result we will only attempt to obtain Gaussian upper bounds when  $d_\theta(x, y) \leq t$ . Non-Gaussian estimates applicable where  $d_\theta(x, y) \geq t$  will be discussed in Section 2, which adapt work of Davies from [8] and [9].

Our main result is a Gaussian upper bound for the heat kernel  $p_t(x, y)$  which is valid under mild hypotheses on  $(\Gamma, \pi)$  and  $(\theta_x)_{x \in G}$ .

**Theorem 1.1.** *Let  $(\Gamma, \pi)$  be a weighted graph, and suppose that there exists a constant  $C_\theta > 0$  such that the vertex weights  $(\theta_x)_{x \in G}$  satisfy  $\theta_x \geq C_\theta$  for each  $x \in G$ . Let  $f_1, f_2$  be  $(A, \gamma)$ -regular functions satisfying, for  $i \in \{1, 2\}$ ,*

$$\sup_{0 < t < \infty} \frac{f_i(t)}{e^{t^{1/2}}} \leq A. \tag{1.5}$$

*Suppose also that there exist vertices  $x_1, x_2 \in G$  such that for all  $t > 0$  and  $i \in \{1, 2\}$ ,*

$$p_t(x_i, x_i) \leq \frac{1}{f_i(t)}. \tag{1.6}$$

*Then there exist constants  $C_1(A, \gamma, C_\theta), C_2(\gamma), \alpha(\gamma) > 0$ , such that for all  $t \geq 1 \vee d_\theta(x_1, x_2)$ ,*

$$p_t(x_1, x_2) \leq \frac{C_1}{(f_1(at)f_2(at))^{1/2}} \exp\left(-C_2 \frac{d_\theta^2(x_1, x_2)}{t}\right).$$

**Remarks:**

1. There is no assumption of stochastic completeness on the process  $(X_t)_{t \geq 0}$ ; these heat kernel estimates hold even if  $(X_t)_{t \geq 0}$  has finite explosion time.

2. The main utility of this result is in settings where  $f_i(t)$  has polynomial growth, so that (1.5) is satisfied. Suppose that for  $i \in \{1, 2\}$ ,  $f_i(t) = f(t) := \exp(ct^\alpha)$  for some  $c, \alpha > 0$ . By Cauchy-Schwarz,  $p_t(x_1, x_2) \leq (p_t(x_1, x_1)p_t(x_2, x_2))^{1/2}$ , and hence  $p_t(x_1, x_2) \leq \exp(-ct^\alpha)$  for all  $t > 0$ . On the other hand, by Theorem 2.2, if  $C > 1$  and  $t \geq Cd_\theta(x_1, x_2)$ ,

$$p_t(x_1, x_2) \leq (\theta_{x_1} \theta_{x_2})^{-1/2} \exp\left(-c_1 \frac{d_\theta^2(x_1, x_2)}{t}\right).$$

If  $0 \leq x \leq y$  and  $0 \leq x \leq z$ , then  $x \leq (yz)^{1/2}$ , so for  $t \geq Cd_\theta(x_1, x_2)$ ,

$$\begin{aligned} p_t(x_1, x_2) &\leq (\theta_{x_1} \theta_{x_2})^{-1/4} \exp\left(-\frac{c}{2} t^\alpha - c_1 \frac{d_\theta^2(x_1, x_2)}{2t}\right) \\ &= \frac{c_2}{f(c_3 t)} \exp\left(-c_1 \frac{d_\theta^2(x_1, x_2)}{2t}\right), \end{aligned}$$

so that a Gaussian upper bound of the desired form can be obtained very easily. Moreover, as  $t \rightarrow \infty$ , it is the on-diagonal term which provides most of the decay in the heat kernel and not the Gaussian exponential factor.

Nevertheless, the growth condition (1.5) is satisfied in many applications (as long as  $A$  is taken sufficiently large). For example, it is typically satisfied for random walks on graphs of polynomial volume growth, super-polynomial but sub-exponential volume growth, or exponential volume growth satisfying a certain isoperimetric inequality [24].

3. Let us note that if  $f$  is  $(A_1, \gamma)$ -regular, and  $A_2 \geq A_1 \geq 1$ , then  $f$  is also  $(A_2, \gamma)$ -regular. Thus, as long as there exist  $A_1, A_2, A_3 \geq 1$  such that  $f_1$  is  $(A_1, \gamma)$ -regular,  $f_2$  is  $(A_2, \gamma)$ -regular, and  $\sup_{0 < t < \infty} \frac{f_i(t)}{e^{t^{1/2}}} \leq A_3$ , then for  $A = A_1 \vee A_2 \vee A_3$ ,  $f_1, f_2$  are  $(A, \gamma)$ -regular, and (1.5) is satisfied.

4. In many applications, one has a uniform on-diagonal heat kernel upper bound, that is, an estimate of the form

$$p_t(x, x) \leq \frac{1}{f(t)},$$

which is valid for all  $x \in G$  and all  $t > 0$ ; various techniques for obtaining such estimates were discussed earlier. However, in other cases, one may obtain a heat kernel upper bound of the form

$$p_t(x, x) \leq \frac{1}{V(x, ct^{1/2})},$$

which is valid for all  $x \in G$  and all  $t > 0$ , and where  $c > 0$  is independent of  $x$  and  $V(x, r) := \pi(B(x, r))$ . This particular on-diagonal upper bound is related to the condition of volume doubling; see [11]. Theorem 1.1 yields Gaussian upper bounds for the heat kernel even in the second situation, where one may have a different on-diagonal upper bound at each point of the graph.

The following is an immediate consequence of Theorem 1.1.

**Corollary 1.2.** *Let  $(\Gamma, \pi)$  be a weighted graph, and suppose that there exists a constant  $C_\theta > 0$  such that the vertex weights  $(\theta_x)_{x \in G}$  satisfy  $\theta_x \geq C_\theta$  for each  $x \in G$ . Let  $f$  be an  $(A, \gamma)$ -regular function satisfying (1.5). If for each  $t > 0$ , the uniform heat kernel condition*

$$\sup_{x \in G} p_t(x, x) \leq \frac{1}{f(t)}$$

*is satisfied, then there exist constants  $C_1(A, \gamma, C_\theta), C_2(\gamma), \alpha(\gamma) > 0$  such that for all  $x_1, x_2 \in G$ , and  $t \geq 1 \vee d_\theta(x_1, x_2)$ ,*

$$p_t(x_1, x_2) \leq \frac{C_1}{f(\alpha t)} \exp\left(-C_2 \frac{d_\theta^2(x_1, x_2)}{t}\right).$$

If  $f$  is only  $(A, \gamma)$ -regular on  $(T_1, T_2)$ , then we obtain a restricted version of Theorem 1.1:

**Theorem 1.3.** *Let  $(\Gamma, \pi)$  be a weighted graph, and suppose that there exists a constant  $C_\theta > 0$  such that the vertex weights  $(\theta_x)_{x \in G}$  satisfy  $\theta_x \geq C_\theta$  for each  $x \in G$ . Let  $f_1, f_2$  be  $(A, \gamma)$ -regular functions on  $(T_1, T_2)$  satisfying, for  $i \in \{1, 2\}$ ,*

$$\sup_{t \in (T_1, T_2)} \frac{f_i(t)}{e^{t^{1/2}}} \leq A.$$

*If there exist vertices  $v_1, v_2 \in G$  such that for all  $t \in (T_1, T_2)$  and  $i \in \{1, 2\}$ , the estimate*

$$p_t(v_i, v_i) \leq \frac{1}{f_i(t)}$$

*holds, then there exist constants  $C_1(A, \gamma, C_\theta), C_2(\gamma), \alpha(\gamma) > 0$  such that for all  $t > 0$  satisfying  $72\gamma^4 e^4 T_1^2 \vee 1 \vee d_\theta(v_1, v_2) < t < T_2$ ,*

$$p_t(v_1, v_2) \leq \frac{C_1}{(f_1(\alpha t)f_2(\alpha t))^{1/2}} \exp\left(-C_2 \frac{d_\theta^2(v_1, v_2)}{t}\right).$$

**Remarks:**

1. The primary use of this result is in the case that  $T_2 = \infty$ , in which case one obtains Gaussian upper bounds for all sufficiently large times. In random environments such as supercritical percolation clusters, the functions which appear in existing on-diagonal heat kernel upper bounds may not be  $(A, \gamma)$ -regular, but rather  $(A, \gamma)$ -regular on  $(T, \infty)$  for some  $T > 0$ ; Theorem 1.3 is useful for obtaining Gaussian upper bounds in this setting. Theorem 1.3 has also been used to obtain Gaussian heat kernel estimates for the random conductance model; see [1].

The structure of this paper is as follows. Section 2 establishes long-range, non-Gaussian heat kernel upper bounds for the heat kernel using the metric  $d_\theta$ , similar to earlier estimates of Davies in [8] and [9]. Section 3 proves a maximum principle, analogous to the one established in [13]; this is subsequently used to estimate a tail sum of the square of the heat kernel. The direct analogue of the maximum principle from [13] does not work in the setting of graphs, and additional restrictions are necessary in order to establish the maximum principle of this paper.

In Section 4, we estimate this tail sum further using a telescoping argument from [13]. In [13], this argument is iterated infinitely many times, but in the present setting the telescoping argument cannot be employed past a finite number of steps. At this point, it is necessary to use the heat kernel estimates of Section 2 to get a final estimate on the tail sum. In Section 5, this estimate of the tail sum is used to estimate a weighted sum of the square of the heat kernel, and in turn, this estimate is used in Section 6 to establish Theorem 1.1. Section 7 discusses the modifications to Section 4 which are necessary to prove Theorem 1.3. Finally, Section 8 discusses applications to random walks on percolation clusters, and how the results of this paper may be applied to existing work on random walks in random environments.

## 2 Long range bounds for the heat kernel

In this section, we establish non-Gaussian upper bounds for the heat kernel  $p_t(x, y)$  which are close to optimal in the space-time region where  $d_\theta(x, y) \geq t$ . These bounds are closely related to the long-range bounds found in [8] and [9], and are established using the same general techniques. These bounds hold for all  $x, y \in G$  and all  $t > 0$ , although they give results weaker than Gaussian upper bounds in the space-time region where  $d_\theta(x, y) \leq t$ .

**Theorem 2.1.** *If  $x_1, x_2 \in G$ , then for all  $t > 0$ ,*

$$p_t(x_1, x_2) \leq (\theta_{x_1} \theta_{x_2})^{-1/2} \exp \left( -\frac{1}{2} d_\theta(x_1, x_2) \log \left( \frac{d_\theta(x_1, x_2)}{2et} \right) - \Lambda t \right),$$

where  $\Lambda \geq 0$  is the bottom of the  $L^2$  spectrum of the operator  $\mathcal{L}_\theta$ .

*Proof.* By Proposition 5 of [8], for all  $x, y \in G$  and  $t > 0$ , we have the estimate



$$p_t(x, y) \leq (\theta_x \theta_y)^{-1/2} \inf_{\psi \in L^\infty(G)} \exp(\psi(x) - \psi(y) + c(\psi)t), \quad (2.1)$$

where  $c(\psi) := \sup_{x \in G} b(\psi, x) - \Lambda$ , and

$$b(\psi, x) := \frac{1}{2\theta_x} \sum_{y \sim x} \pi_{xy} (e^{\psi(y) - \psi(x)} + e^{\psi(x) - \psi(y)} - 2).$$

Fix  $x_1, x_2 \in G$ , set  $D := d_\theta(x_1, x_2)$  and, for  $\lambda > 0$ , define  $\psi_\lambda(x) := \lambda(D \wedge d_\theta(x, x_1)) \in L^\infty(G)$ . Using the triangle inequality for the metric  $d_\theta$  and the fact that the function  $g(t) := e^t + e^{-t} = 2 \cosh(t)$  is increasing for  $t \geq 0$ , we obtain

$$\begin{aligned} b(\psi_\lambda, x) &:= \frac{1}{2\theta_x} \sum_{y \sim x} \pi_{xy} (e^{\psi_\lambda(y) - \psi_\lambda(x)} + e^{\psi_\lambda(x) - \psi_\lambda(y)} - 2) \\ &\leq \frac{1}{2\theta_x} \sum_{y \sim x} \pi_{xy} (e^{\lambda d_\theta(x, y)} + e^{-\lambda d_\theta(x, y)} - 2). \end{aligned}$$

At this point, we use the inequality

$$e^s + e^{-s} - 2 \leq s^2 e^s,$$

which is valid for all  $s \geq 0$ . This gives

$$\begin{aligned} b(\psi_\lambda, x) &\leq \frac{1}{2\theta_x} \sum_{y \sim x} \pi_{xy} (e^{\lambda d_\theta(x, y)} + e^{-\lambda d_\theta(x, y)} - 2) \\ &\leq \frac{1}{2\theta_x} \sum_{y \sim x} \pi_{xy} (\lambda^2 d_\theta^2(x, y) e^{\lambda d_\theta(x, y)}) \\ &= \left( \frac{1}{\theta_x} \sum_{y \sim x} \pi_{xy} d_\theta^2(x, y) \right) \left( \frac{1}{2} \lambda^2 e^\lambda \right) \\ &\leq \frac{1}{2} \lambda^2 e^\lambda. \end{aligned}$$

Since this estimate holds uniformly in  $x$ , we have that

$$\sup_{x \in G} b(\psi_\lambda, x) \leq \frac{1}{2} \lambda^2 e^\lambda,$$

and

$$c(\psi_\lambda) := \sup_{x \in G} b(\psi_\lambda, x) - \Lambda \leq \frac{1}{2} \lambda^2 e^\lambda - \Lambda.$$

Set  $f(\lambda) := \frac{1}{2}\lambda^2 e^\lambda$ . Combining these estimates with (2.1), we get, for each  $\lambda > 0$ ,

$$\begin{aligned} p_t(x_1, x_2) &\leq (\theta_{x_1} \theta_{x_2})^{-1/2} \exp(\psi_\lambda(x_1) - \psi_\lambda(x_2) + c(\psi_\lambda)t) \\ &= (\theta_{x_1} \theta_{x_2})^{-1/2} \exp(-\lambda d_\theta(x_1, x_2) + c(\psi_\lambda)t) \\ &\leq (\theta_{x_1} \theta_{x_2})^{-1/2} \exp(-\lambda d_\theta(x_1, x_2) + f(\lambda)t - \Lambda t) \\ &= (\theta_{x_1} \theta_{x_2})^{-1/2} \exp\left(t \left(-\lambda \left(\frac{d_\theta(x_1, x_2)}{t}\right) + f(\lambda)\right) - \Lambda t\right). \end{aligned}$$

By optimizing over  $\lambda > 0$ , we have

$$p_t(x_1, x_2) \leq (\theta_{x_1} \theta_{x_2})^{-1/2} \exp\left(t \widehat{f}\left(\frac{d_\theta(x_1, x_2)}{t}\right) - \Lambda t\right),$$

where  $\widehat{f}$  is the Legendre transform of  $f$ , defined by

$$\widehat{f}(\gamma) := \inf_{\lambda > 0} (-\lambda \gamma + f(\lambda)).$$

Note that if  $f(\lambda) \leq g(\lambda)$  for all  $\lambda > 0$ ,  $\widehat{f}(\gamma) \leq \widehat{g}(\gamma)$ . Now, the function  $g(\lambda) := e^{2\lambda}$  satisfies  $f(\lambda) \leq g(\lambda)$  for all  $\lambda > 0$ , so

$$\widehat{f}(\gamma) \leq \widehat{g}(\gamma) = -\frac{\gamma}{2} \log\left(\frac{\gamma}{2e}\right).$$

Thus, applying this estimate to the preceding work gives

$$p_t(x_1, x_2) \leq (\theta_{x_1} \theta_{x_2})^{-1/2} \exp\left(-\frac{1}{2} d_\theta(x_1, x_2) \log\left(\frac{d_\theta(x_1, x_2)}{2et}\right) - \Lambda t\right),$$

which holds for all  $t > 0$ . □

One may also use these results to obtain a weak Gaussian upper bound for the heat kernel which does not use any information from on-diagonal bounds.

**Theorem 2.2.** *If  $x_1, x_2 \in G$ , then for  $t \geq d_\theta(x_1, x_2)$ ,*

$$p_t(x_1, x_2) \leq (\theta_{x_1} \theta_{x_2})^{-1/2} \exp\left(-\frac{d_\theta^2(x_1, x_2)}{2t} \left(1 - \frac{d_\theta(x_1, x_2)}{t}\right) - \Lambda t\right).$$

*Proof.* We proceed as in the proof of Theorem 2.1. Instead of using the inequality  $e^s + e^{-s} - 2 \leq s^2 e^s$ , we use the estimate

$$e^s + e^{-s} - 2 \leq s^2 \left(1 + \frac{s e^s}{6}\right),$$

which was used previously in [9]; we then obtain estimates similar to those above, except with  $f(\lambda) := \frac{1}{2} \lambda^2 \left(1 + \frac{\lambda e^\lambda}{6}\right)$ . In [9], Davies computes that

$$\widehat{(2f)}(\gamma) \leq -\frac{\gamma^2}{4} + \frac{\gamma^3}{8},$$

and since  $\widehat{f}(\gamma) = \frac{1}{2} \widehat{(2f)}(2\gamma)$ , we obtain

$$\widehat{f}(\gamma) \leq -\frac{1}{2} \gamma^2 + \frac{1}{2} \gamma^3 = -\frac{1}{2} \gamma^2 (1 - \gamma).$$

Inserting this estimate into the above yields

$$p_t(x_1, x_2) \leq (\theta_{x_1} \theta_{x_2})^{-1/2} \exp \left( -\frac{d_\theta^2(x_1, x_2)}{2t} \left(1 - \frac{d_\theta(x_1, x_2)}{t}\right) - \Lambda t \right),$$

as desired. □

### 3 Maximum Principle

For the remainder of the paper, we fix a set of vertex weights  $(\theta_x)_{x \in G}$  for which there exists  $C_\theta > 0$  with  $\theta_x \geq C_\theta$  for all  $x \in G$ , and an associated metric  $d_\theta$ , satisfying (1.1). We also fix an increasing set of finite connected subsets  $(G_n)_{n \in \mathbb{Z}_+}$  with limit  $G$ .

Let  $x_0 \in G$  be a point for which there exists a  $(A, \gamma)$ -regular function  $f$  satisfying (1.5) such that for  $t > 0$ ,

$$p_t(x_0, x_0) \leq \frac{1}{f(t)}.$$

We define  $u(x, t) := p_t(x_0, x)$ , and  $u^{(k)}(x, t) := p_t^{(G_k)}(x_0, x)$  for  $k \in \mathbb{Z}_+$ .

In this section, we will prove a maximum principle for the quantities

$$J_R^{(k)}(t) := \sum_{x \in G_k} (u^{(k)})^2(x, t) \exp(\xi_R(x, t)) \theta_x,$$

where  $\xi_R$  will be defined later. This will allow us to estimate various sums and weighted sums of  $u^2$ . One basic estimate which we will use repeatedly is, for any  $H \subset G$  and  $k \in \mathbb{Z}_+$ ,

$$\sum_{x \in H} (u^{(k)})^2(x, t) \theta_x \leq \sum_{x \in H} u^2(x, t) \theta_x \leq \sum_{x \in G} p_t(x_0, x) p_t(x, x_0) \theta_x = p_{2t}(x_0, x_0) \leq \frac{1}{f(2t)}, \quad (3.1)$$

using the symmetry and semigroup properties of the heat kernel.

The reason for considering the killed heat kernels  $p_t^{(G_k)}(x, y)$  is that the function  $u^{(k)}$  is finitely supported, and thus there is no difficulty in interchanging double sums. When

$$\sup_{x \in G} \frac{\pi_x}{\theta_x} = \infty,$$

$\mathcal{L}_\theta$  is not a bounded operator on  $L^2(\theta)$  (see [8] for a proof), and the interchange of sums in (3.2) is not straightforward. We also remark that there is in general no simple description of the domain of the associated Dirichlet form  $\mathcal{E}$  in this case.

Fix  $k \in \mathbb{Z}_+$ . Differentiating  $J_R^{(k)}(t)$  and using the fact that  $u$  is a solution to the heat equation on  $G_k$ , we get (writing  $u_x^{(k)}$  for  $u^{(k)}(x, t)$ ,  $\zeta$  for  $\exp \circ \xi$ , and  $\zeta_x$  for  $\zeta(x, t)$ ),

$$\begin{aligned} \frac{d}{dt} J_R^{(k)}(t) &= \sum_{x \in G} \left( \frac{\partial}{\partial t} u_x^{(k)} \right) (2u_x^{(k)} \zeta_x) \theta_x + \sum_{x \in G} \left( \frac{\partial}{\partial t} \zeta_x \right) (u_x^{(k)})^2 \theta_x \\ &= \sum_{x \in G_k} (\mathcal{L}_\theta u_x^{(k)}) (2u_x^{(k)} \zeta_x) \theta_x + \sum_{x \in G_k} \left( \frac{\partial}{\partial t} \zeta_x \right) (u_x^{(k)})^2 \theta_x. \end{aligned}$$

Note that  $\left( \frac{\partial}{\partial t} u_x^{(k)} \right) (2u_x^{(k)} \zeta_x) = (\mathcal{L}_\theta u_x^{(k)}) (2u_x^{(k)} \zeta_x)$  even if  $x_0 \notin G_k$  or  $x \notin G_k$ . By a Gauss-Green type calculation and using the fact that  $u_y^{(k)} = 0$  for  $y \in G \setminus G_k$ ,

$$\begin{aligned} \sum_{x \in G_k} (\mathcal{L}_\theta u_x^{(k)}) (2u_x^{(k)} \zeta_x) \theta_x &= \sum_{x \in G_k} \sum_{y \in G} (u_y^{(k)} - u_x^{(k)}) (2u_x^{(k)} \zeta_x) \pi_{xy} \\ &= \sum_{x \in G_k} \sum_{y \in G_k} (u_y^{(k)} - u_x^{(k)}) (2u_x^{(k)} \zeta_x) \pi_{xy} + \sum_{x \in G_k} \sum_{y \in G \setminus G_k} (u_y^{(k)} - u_x^{(k)}) (2u_x^{(k)} \zeta_x) \pi_{xy} \\ &= \sum_{x \in G_k} \sum_{y \in G_k} (u_y^{(k)} - u_x^{(k)}) (2u_x^{(k)} \zeta_x) \pi_{xy} + \sum_{x \in G_k} \sum_{y \in G \setminus G_k} (-u_x^{(k)}) (2u_x^{(k)} \zeta_x) \pi_{xy} \\ &\leq \sum_{x \in G_k} \sum_{y \in G_k} (u_y^{(k)} - u_x^{(k)}) (2u_x^{(k)} \zeta_x) \pi_{xy} \\ &= - \sum_{x, y \in G_k} (u_y^{(k)} - u_x^{(k)}) (u_y^{(k)} \zeta_y - u_x^{(k)} \zeta_x) \pi_{xy}. \end{aligned} \quad (3.2)$$

The equality (3.2) follows from interchanging the order of summation, which is permissible since  $u^{(k)}$  has finite support. Completing the square, we see that

$$\begin{aligned}
-\sum_{x,y \in G_k} (u_y^{(k)} - u_x^{(k)})(u_y^{(k)} \zeta_y - u_x^{(k)} \zeta_x) \pi_{xy} &= -\sum_{x,y \in G_k} \zeta_y (u_y^{(k)} - u_x^{(k)})^2 \pi_{xy} \\
&\quad - \sum_{x,y \in G_k} u_x^{(k)} (u_y^{(k)} - u_x^{(k)}) (\zeta_y - \zeta_x) \pi_{xy} \\
&\leq \frac{1}{4} \sum_{x,y \in G_k} (u_x^{(k)})^2 \frac{(\zeta_x - \zeta_y)^2}{\zeta_y} \pi_{xy}.
\end{aligned}$$

It follows that

$$\begin{aligned}
\frac{d}{dt} J_R^{(k)}(t) &\leq \frac{1}{4} \sum_{x,y \in G_k} (u_x^{(k)})^2 \frac{(\zeta_x - \zeta_y)^2}{\zeta_y} \pi_{xy} + \sum_{x \in G_k} \left( \frac{\partial}{\partial t} \zeta_x \right) (u_x^{(k)})^2 \theta_x \\
&= \sum_{x \in G_k} (u_x^{(k)})^2 \sum_{y \in G_k} \left( \frac{\theta_x}{\pi_x} \frac{\partial}{\partial t} \zeta_x + \frac{1}{4} \frac{(\zeta_x - \zeta_y)^2}{\zeta_y} \right) \pi_{xy} \\
&= \sum_{x \in G_k} (u_x^{(k)})^2 \zeta_x \sum_{y \in G_k} \left( \frac{\theta_x}{\pi_x} \frac{\partial}{\partial t} \zeta_x + \frac{1}{4} \left( \frac{\zeta_x^2 - 2\zeta_x \zeta_y + \zeta_y^2}{\zeta_x \zeta_y} \right) \right) \pi_{xy} \\
&= \sum_{x \in G_k} (u_x^{(k)})^2 \zeta_x \sum_{y \in G_k} \left( \frac{\theta_x}{\pi_x} \frac{\partial}{\partial t} \zeta_x + \frac{1}{2} (\cosh(\zeta_x - \zeta_y) - 1) \right) \pi_{xy}.
\end{aligned}$$

Given  $\lambda > 1$ , there exists  $K_\lambda < \infty$  so that the inequality

$$2 \cosh t - 2 \leq \lambda t^2 \tag{3.3}$$

holds for  $|t| \leq K_\lambda$ . Now, we define the distance function  $d_{R,\theta}(x) := (R - d_\theta(x_0, x))_+$ , and set

$$\xi_R(x, t) := -\frac{\delta d_{R,\theta}^2(x) + \varepsilon}{s - t}.$$

Here  $R \geq 0$ ,  $t > 0$ , and  $s = s(t) > t$  are parameters that will be allowed to vary, and  $\delta, \varepsilon > 0$  are parameters that will be fixed. For the rest of this paper, we will fix  $\lambda, \delta, \varepsilon$  so that the following conditions are satisfied:

$$\lambda > 1, \tag{3.4}$$

$$\delta < \frac{1}{\lambda}, \tag{3.5}$$

$$\varepsilon \geq \frac{\lambda \delta^2}{4(1 - \lambda \delta)}, \tag{3.6}$$

$$\frac{K_\lambda}{\delta} = 6\gamma e^2. \tag{3.7}$$

Let us show that such an assignment of constants is possible by exhibiting  $\lambda_0, \delta_0, \varepsilon_0$  which satisfy the above conditions. First, we choose  $\lambda_0 = 2$ , so that  $K_{\lambda_0} = 2.98\dots \leq 3$ ; this satisfies (3.4). Next, since  $\lambda_0$  and  $\gamma$  are known, we may define  $\delta_0$  through (3.7), and estimate

$$\delta_0 := \frac{K_{\lambda_0}}{6\gamma e^2} < \frac{1}{2\gamma e^2} < \frac{1}{\lambda_0},$$

so that (3.5) is also satisfied. We then choose  $\varepsilon_0$  to be

$$\varepsilon_0 := \frac{\lambda_0 \delta_0^2}{4(1 - \lambda_0 \delta_0)}.$$

Let us also note that (3.6) is equivalent to

$$\frac{4\varepsilon}{\lambda\delta(\delta + 4\varepsilon)} \geq 1. \quad (3.8)$$

Once  $\lambda, \delta$  and  $\varepsilon$  have been fixed, we have the following result:

**Lemma 3.1. (Maximum Principle)** *If conditions (3.4), (3.5), (3.6), (3.7) are satisfied, and  $R \geq 0$ ,  $t > 0$ , and  $s > t$  are chosen so that*

$$R - 6\gamma e^2(s - t) + \frac{1}{2} \leq 0, \quad (3.9)$$

then for each  $k \in \mathbb{Z}_+$ ,

$$\frac{\partial}{\partial t} J_R^{(k)}(t) \leq 0.$$

*Proof.* Given  $k \in \mathbb{Z}_+$  and  $x \in G_k$ , set

$$\phi^{(k)}(x) := \sum_{y \in G_k} \pi_{xy} \left( \frac{\theta_x}{\pi_x} \frac{\partial}{\partial t} \xi_x + \frac{1}{2} (\cosh(\xi_x - \xi_y) - 1) \right)$$

Suppose that for all  $x \in G_k$ , whenever  $y \sim x$  and  $y \in G_k$ ,  $|\xi_x - \xi_y| \leq K_\lambda$ . Using (3.3), the inequality  $|d_{R,\theta}^2(x) - d_{R,\theta}^2(y)| \leq 2d_{R,\theta}(x) + 1$ , and (3.4), (3.5), (3.6), and (3.8), we obtain

$$\begin{aligned}
\phi^{(k)}(x) &:= \sum_{y \in G_k} \pi_{xy} \left( \frac{\theta_x}{\pi_x} \frac{\partial}{\partial t} \xi_x + \frac{1}{2} (\cosh(\xi_x - \xi_y) - 1) \right) \\
&\leq \sum_{y \in G_k} \pi_{xy} \left( \frac{\theta_x}{\pi_x} \frac{d}{dt} \xi_x + \frac{\lambda}{4} (\xi_x - \xi_y)^2 \right) \\
&= (s-t)^{-2} \sum_{y \in G_k} \pi_{xy} \left( -\frac{\theta_x}{\pi_x} (\delta d_{R,\theta}^2(x) + \varepsilon) + \frac{\lambda \delta^2}{4} (d_{R,\theta}^2(x) - d_{R,\theta}^2(y))^2 \right) \\
&= (s-t)^{-2} \sum_{y \in G_k} \pi_{xy} \left( -\frac{\theta_x}{\pi_x} (\delta d_{R,\theta}^2(x) + \varepsilon) + \frac{\lambda \delta^2}{4} (d_{R,\theta}(x) - d_{R,\theta}(y))^2 (d_{R,\theta}(x) + d_{R,\theta}(y))^2 \right) \\
&\leq (s-t)^{-2} \sum_{y \in G_k} \pi_{xy} \left( -\frac{\theta_x}{\pi_x} (\delta d_{R,\theta}^2(x) + \varepsilon) + \frac{\lambda \delta^2}{4} d_{\theta}^2(x,y) (2d_{R,\theta}(x) + 1)^2 \right) \\
&= (s-t)^{-2} \left( -\theta_x (\delta d_{R,\theta}^2(x) + \varepsilon) + \sum_{y \in G_k} \pi_{xy} \frac{\lambda \delta^2}{4} d_{\theta}^2(x,y) (2d_{R,\theta}(x) + 1)^2 \right) \\
&= \frac{\lambda \delta^2}{4} (2d_{R,\theta}(x) + 1)^2 (s-t)^{-2} \theta_x \left( \frac{1}{\theta_x} \sum_{y \in G_k} d_{\theta}^2(x,y) \pi_{xy} - \frac{4}{\lambda \delta^2} \frac{\delta d_{R,\theta}^2(x) + \varepsilon}{(2d_{R,\theta}(x) + 1)^2} \right) \\
&\leq \frac{\lambda \delta^2}{4} (2d_{R,\theta}(x) + 1)^2 (s-t)^{-2} \theta_x \left( \frac{1}{\theta_x} \sum_{y \in G_k} d_{\theta}^2(x,y) \pi_{xy} - \inf_{u \geq 0} \frac{4}{\lambda \delta^2} \frac{\delta u^2 + \varepsilon}{(2u + 1)^2} \right) \\
&= \frac{\lambda \delta^2}{4} (2d_{R,\theta}(x) + 1)^2 (s-t)^{-2} \theta_x \left( \frac{1}{\theta_x} \sum_{y \in G_k} d_{\theta}^2(x,y) \pi_{xy} - \frac{4\varepsilon}{\lambda \delta (\delta + 4\varepsilon)} \right) \\
&\leq \frac{\lambda \delta^2}{4} (2d_{R,\theta}(x) + 1)^2 (s-t)^{-2} \theta_x \left( \frac{1}{\theta_x} \sum_{y \in G_k} d_{\theta}^2(x,y) \pi_{xy} - 1 \right) \\
&\leq 0.
\end{aligned}$$

Since

$$\frac{d}{dt} J_R^{(k)}(t) \leq \sum_{x \in G} (u_x^{(k)})^2 \zeta_x \phi^{(k)}(x),$$

we conclude that

$$\frac{d}{dt} J_R^{(k)}(t) \leq 0.$$

Now, let us analyze the inequality

$$|\xi_x - \xi_y| = \left| \frac{\delta (d_{R,\theta}^2(x) - d_{R,\theta}^2(y))}{s-t} \right| \leq K_{\lambda}.$$

As before, we have  $|d_{R,\theta}^2(x) - d_{R,\theta}^2(y)| \leq 2d_{R,\theta}(x) + 1$ , so this holds if

$$d_{R,\theta}(x) \leq \frac{K_\lambda}{2\delta}(s-t) - \frac{1}{2},$$

and, since  $d_{R,\theta}(x) \leq R$ , it certainly holds when

$$R - 6\gamma e^2(s-t) + \frac{1}{2} \leq 0.$$

which is precisely the condition in the statement of the Lemma. □

Now, for  $k \in \mathbb{Z}_+$ , we define

$$\begin{aligned} I_R^{(k)}(t) &:= \sum_{x \in G_k \setminus B_\theta(x_0, R)} (u^{(k)}(x, t))^2 \theta_x, \\ I_R(t) &:= \sum_{x \in G \setminus B_\theta(x_0, R)} u^2(x, t) \theta_x. \end{aligned}$$

By (3.1), all of these quantities are finite, and by monotone convergence,

$$\lim_{k \rightarrow \infty} I_R^{(k)}(t) = I_R(t). \quad (3.10)$$

The maximum principle allows us to estimate  $I$ , as follows:

**Lemma 3.2.** *Suppose that  $R_0 \geq R_1$ , and  $s > t_0 \geq t_1 > 0$  are such that  $R, s, t$  satisfy (3.9). Then*

$$I_{R_0}(t_0) \leq \exp\left(\frac{\varepsilon}{s-t_0}\right) I_{R_1}(t_1) + \exp\left(\frac{\varepsilon}{s-t_0}\right) \exp\left(-\frac{\delta(R_0 - R_1)^2 + \varepsilon}{s-t_1}\right) \frac{1}{f(2t_1)}.$$

*Proof.* First, since  $d_{R_0,\theta}$  vanishes outside of  $B_\theta(x_0, R_0)$ , for each  $k \in \mathbb{Z}_+$ ,

$$\begin{aligned} I_{R_0}^{(k)}(t_0) &:= \sum_{x \in G_k \setminus B_\theta(x_0, R_0)} (u^{(k)}(x, t_0))^2 \theta_x \\ &\leq \sup_{x \in G_k \setminus B_\theta(x_0, R_0)} \exp(-\xi_{R_0}(x, t_0)) \sum_{x \in G_k \setminus B_\theta(x_0, R_0)} (u^{(k)}(x, t_0))^2 \exp(\xi_{R_0}(x, t_0)) \theta_x \\ &\leq \exp\left(\frac{\varepsilon}{s-t_0}\right) \sum_{x \in G_k \setminus B_\theta(x_0, R_0)} (u^{(k)}(x, t_0))^2 \exp(\xi_{R_0}(x, t_0)) \theta_x \\ &\leq \exp\left(\frac{\varepsilon}{s-t_0}\right) J_{R_0}^{(k)}(t_0). \end{aligned}$$



Next, for  $\ell \in [t_1, t_0]$ ,

$$R_0 - 6\gamma e^2(s - \ell) + \frac{1}{2} \leq 0,$$

and so the maximum principle yields  $J_{R_0}^{(k)}(t_0) \leq J_{R_0}^{(k)}(t_1)$ , so that

$$\begin{aligned} I_{R_0}^{(k)}(t_0) &\leq \exp\left(\frac{\varepsilon}{s - t_0}\right) J_{R_0}^{(k)}(t_1) \\ &= \exp\left(\frac{\varepsilon}{s - t_0}\right) \left( \sum_{x \in G_k \setminus B_\theta(x_0, R_1)} + \sum_{x \in G_k \cap B_\theta(x_0, R_1)} \right) (u^{(k)}(x, t_1))^2 \exp(\xi_{R_0}(x, t_1)) \theta_x \\ &\leq \exp\left(\frac{\varepsilon}{s - t_0}\right) I_{R_1}^{(k)}(t_1) \\ &\quad + \exp\left(\frac{\varepsilon}{s - t_0}\right) \sup_{x \in G_k \cap B_\theta(x_0, R_1)} \exp(\xi_{R_0}(x, t_1)) \sum_{x \in G_k \cap B_\theta(x_0, R_1)} (u^{(k)}(x, t_1))^2 \theta_x \\ &= \leq \exp\left(\frac{\varepsilon}{s - t_0}\right) I_{R_1}^{(k)}(t_1) + \exp\left(\frac{\varepsilon}{s - t_0}\right) \exp\left(-\frac{\delta(R_0 - R_1)^2 + \varepsilon}{s - t_1}\right) \frac{1}{f(2t_1)}. \end{aligned}$$

The last three inequalities follow from bounding above the exponential weight  $\exp(\xi_{R_0}(x, t_1))$  by 1 (on  $G_k \setminus B_\theta(x_0, R_1)$ ), by using the inequality  $d_{R_0, \theta}(x) \geq R_0 - R_1$  (on  $G_k \cap B_\theta(x_0, R_1)$ ), and using (3.1).

Letting  $k \rightarrow \infty$  and using (3.10), we get

$$I_{R_0}(t_0) \leq \exp\left(\frac{\varepsilon}{s - t_0}\right) I_{R_1}(t_1) + \exp\left(\frac{\varepsilon}{s - t_0}\right) \exp\left(-\frac{\delta(R_0 - R_1)^2 + \varepsilon}{s - t_1}\right) \frac{1}{f(2t_1)},$$

which completes the proof of the Lemma. □

## 4 Further estimates for $I_R(t)$

In this section, we will prove the following estimate for  $I_R(t)$ :

**Lemma 4.1.** *Suppose that  $t_0 \geq R_0 \geq 1/2$ . There exist positive constants  $m_0, m_1, n_0, n_1, \alpha$ , which do not depend on either  $t_0$  or  $R_0$ , so that*

$$I_{R_0}(t_0) \leq m_0 \frac{1}{f(\alpha t_0)} \exp\left(-m_1 \frac{R_0^2}{t_0}\right) + n_0 \exp(-n_1 R_0).$$

In [13], a similar estimate is obtained without the  $n_0 \exp(-n_1 R_0)$  term, and is a key step in establishing Gaussian upper bounds. The condition (1.5) in the statement of Theorem 1.3 prevents the term  $n_0 \exp(-n_1 R_0)$  from dominating the ‘Gaussian term’  $m_0 \frac{1}{f(at_0)} \exp\left(-m_1 \frac{R_0^2}{t_0}\right)$ .

*Proof.* Given  $t_0 \geq R_0 \geq 1/2$ , we define sequences  $(t_j)_{j \in \mathbb{Z}_+}$ ,  $(s_j)_{j \in \mathbb{Z}_+}$ ,  $(R_j)_{j \in \mathbb{Z}_+}$  by

$$\begin{aligned} t_j &:= t_0 \gamma^{-j}, \\ s_j &:= 2t_j, \\ R_j &:= \left(\frac{1}{2} + \frac{1}{j+2}\right) R_0. \end{aligned}$$

Recall that  $\gamma > 1$  was seen first in the  $(A, \gamma)$ -regularity of the function  $f$ . Note that

$$\begin{aligned} R_j - R_{j+1} &\geq \frac{R_0}{(j+3)^2}, \\ s_j - t_{j+1} &= \left(2 - \frac{1}{\gamma}\right) t_j. \end{aligned}$$

As long as

$$R_j - 6\gamma e^2 (s_j - t_j) + \frac{1}{2} \leq 0, \quad (4.1)$$

then Lemma 3.2 gives

$$I_{R_j}(t_j) \leq \exp\left(\frac{\varepsilon}{s_j - t_j}\right) I_{R_{j+1}}(t_{j+1}) + \frac{1}{f(2t_{j+1})} \exp\left(\frac{\varepsilon}{s_j - t_j}\right) \exp\left(-\frac{\delta(R_j - R_{j+1})^2 + \varepsilon}{s_j - t_{j+1}}\right). \quad (4.2)$$

Let us analyze when (4.1) is satisfied. Let  $j^*$  denote the maximal  $j$  for which (4.1) holds. First,  $j^* \geq 0$ , since

$$R_0 - 6\gamma e^2 (s_0 - t_0) + \frac{1}{2} = R_0 - 6\gamma e^2 t_0 + \frac{1}{2} < 0$$

Using the definition of  $(R_j)_{j \in \mathbb{Z}_+}$ , we obtain

$$\frac{1}{4} \leq \frac{R_0}{2} < R_{j^*} \leq R_0,$$

and the maximality of  $j^*$  shows that

$$\begin{aligned} R_{j^*} &\leq 6\gamma e^2 t_{j^*}, \\ R_{j^*+1} &> 6\gamma e^2 t_{j^*+1} - \frac{1}{2}. \end{aligned}$$

Rearranging, we obtain

$$\begin{aligned} \frac{1}{6\gamma e^2} R_{j^*} &\leq t_{j^*} < \frac{1}{2e^2} R_{j^*}, \\ \frac{1}{12\gamma e^2} R_0 &< t_{j^*} < \frac{1}{2e^2} R_0. \end{aligned} \tag{4.3}$$

Applying (4.2) repeatedly yields

$$\begin{aligned} I_{R_0}(t_0) &\leq \prod_{k=0}^{j^*} \exp\left(\frac{\varepsilon}{s_k - t_k}\right) I_{R_{j^*}}(t_{j^*}) \\ &\quad + \sum_{k=0}^{j^*} \left( \prod_{\ell=0}^k \exp\left(\frac{\varepsilon}{s_\ell - t_\ell}\right) \right) \exp\left(-\frac{\delta(R_k - R_{k+1})^2 + \varepsilon}{s_k - t_{k+1}}\right) \frac{1}{f(2t_{k+1})} \\ &:= S_1 + S_2. \end{aligned}$$

The product in  $S_1$  may be estimated as follows:

$$\begin{aligned} S_1 &:= \prod_{k=0}^{j^*} \exp\left(\frac{\varepsilon}{s_k - t_k}\right) I_{R_{j^*}}(t_{j^*}) \\ &= \exp\left(\frac{\varepsilon}{t_0} \sum_{k=0}^{j^*} \gamma^k\right) I_{R_{j^*}}(t_{j^*}) \\ &\leq \exp\left(\frac{\varepsilon\gamma}{\gamma-1} \frac{1}{t_{j^*}}\right) I_{R_{j^*}}(t_{j^*}) \\ &\leq \exp\left(\frac{12\varepsilon\gamma^2 e^2}{(\gamma-1)R_0}\right) I_{R_{j^*}}(t_{j^*}) \\ &\leq \exp\left(\frac{24\varepsilon\gamma^2 e^2}{\gamma-1}\right) I_{R_{j^*}}(t_{j^*}). \end{aligned} \tag{4.4}$$

We will deal with the  $I_{R_{j^*}}(t_{j^*})$  term later. Continuing,

$$\begin{aligned}
S_2 &:= \sum_{k=0}^{j^*} \left( \prod_{\ell=0}^k \exp \left( \frac{\varepsilon}{s_\ell - t_\ell} \right) \right) \exp \left( -\frac{\delta(R_k - R_{k+1})^2 + \varepsilon}{s_k - t_{k+1}} \right) \frac{1}{f(2t_{k+1})} \\
&\leq \sum_{k=0}^{j^*} \exp \left( \frac{\varepsilon\gamma}{(\gamma-1)t_0} \gamma^k \right) \exp \left( -\frac{\delta(R_k - R_{k+1})^2 + \varepsilon}{s_k - t_{k+1}} \right) \frac{1}{f(2t_{k+1})} \\
&= \sum_{k=0}^{j^*} \exp \left( \frac{\varepsilon\gamma^2}{(\gamma-1)(2\gamma-1)t_0} \gamma^k \right) \exp \left( -\frac{\delta(R_k - R_{k+1})^2}{s_k - t_{k+1}} \right) \frac{1}{f(2t_{k+1})} \\
&\leq \sum_{k=0}^{j^*} \exp \left( \frac{\varepsilon\gamma^2}{(\gamma-1)(2\gamma-1)t_0} \gamma^k \right) \exp \left( -\frac{\delta\gamma}{(2\gamma-1)(k+3)^4} \frac{\gamma^k R_0^2}{t_0} \right) \frac{1}{f(2t_{k+1})}.
\end{aligned}$$

At this point, define  $\beta > 0$ , which depends only on  $\gamma > 1$ , by

$$\beta := \inf_{k \geq 0} \frac{\gamma^{k+1}}{(2\gamma-1)(k+2)(k+3)^4},$$

so that for  $k \geq 0$ ,

$$\beta(k+2) \geq \frac{\gamma^{k+1}}{(2\gamma-1)(k+3)^4}.$$

The  $(A, \gamma)$ -regularity of  $f$  gives, for  $0 \leq j \leq k$ ,

$$\frac{f(2t_j)}{f(2t_{j+1})} \leq A \frac{f(2t_0)}{f(2t_1)},$$

and multiplying these estimates together yields

$$\begin{aligned}
\frac{1}{f(2t_{k+1})} &\leq \frac{1}{f(2t_0)} \left( A \frac{f(2t_0)}{f(2t_1)} \right)^{k+1} \\
&= \frac{1}{f(2t_0)} \exp \left( (k+1) \log \left( A \frac{f(2t_0)}{f(2t_1)} \right) \right). \tag{4.5}
\end{aligned}$$

We remark that this is the only point in the proof where we use the  $(A, \gamma)$ -regularity of  $f$ .

Set  $L := \log \left( A \frac{f(2t_0)}{f(2t_1)} \right)$  and insert (4.5) into our earlier estimate for  $S_2$  to obtain

$$\begin{aligned}
S_2 &\leq \frac{1}{f(2t_0)} \sum_{k=0}^{j^*} \exp\left(\frac{\varepsilon\gamma^2}{(\gamma-1)(2\gamma-1)t_0} \gamma^k\right) \exp\left(-\delta\beta(k+2)\frac{R_0^2}{t_0}\right) \exp((k+1)L) \\
&\leq \frac{1}{f(2t_0)} \exp\left(\frac{\varepsilon\gamma^2}{(\gamma-1)(2\gamma-1)t_{j^*}}\right) \sum_{k=0}^{j^*} \exp\left(-\delta\beta(k+2)\frac{R_0^2}{t_0}\right) \exp((k+1)L) \\
&= \frac{1}{f(2t_0)} \exp\left(\frac{24\varepsilon\gamma^3 e^2}{(\gamma-1)(2\gamma-1)}\right) \exp\left(-\delta\beta\frac{R_0^2}{t_0}\right) \\
&\quad \times \sum_{k=0}^{j^*} \exp\left(-(k+1)\left(\delta\beta\frac{R_0^2}{t_0} - L\right)\right).
\end{aligned}$$

At this point, we divide into cases based on whether

$$\delta\beta\frac{R_0^2}{t_0} - L \geq \log 2$$

or not. If it is, then we have

$$\begin{aligned}
S_2 &\leq \frac{1}{f(2t_0)} \exp\left(\frac{24\varepsilon\gamma^3 e^2}{(\gamma-1)(2\gamma-1)}\right) \exp\left(-\delta\beta\frac{R_0^2}{t_0}\right) \sum_{k=0}^{j^*} \exp(-(k+1)\log 2) \\
&\leq \frac{1}{f(2t_0)} \exp\left(\frac{24\varepsilon\gamma^3 e^2}{(\gamma-1)(2\gamma-1)}\right) \exp\left(-\delta\beta\frac{R_0^2}{t_0}\right). \tag{4.6}
\end{aligned}$$

If not, then we can estimate  $S_2$  by

$$\begin{aligned}
S_2 &\leq I_{R_0}(t_0) \\
&\leq \sum_{x \in G} u^2(x, t_0) \theta_x \\
&\leq \frac{1}{f(2t_0)} \\
&\leq \frac{1}{f(2t_0)} \exp\left(-\delta\beta\frac{R_0^2}{t_0} + \log\left(A\frac{f(2t_0)}{f(2t_1)}\right) + \log 2\right) \\
&= \frac{2A}{f(2t_1)} \exp\left(-\delta\beta\frac{R_0^2}{t_0}\right). \tag{4.7}
\end{aligned}$$

It remains to estimate the quantity  $I_{R_{j^*}}(t_{j^*})$ . From Theorem 2.1, we have the following pointwise estimate of the heat kernel:

$$p_t(x, y) \leq (\theta_x \theta_y)^{-1/2} \exp \left( -\frac{1}{2} d_\theta(x, y) \log \left( \frac{d_\theta(x, y)}{2et} \right) \right)$$

Hence,

$$\begin{aligned} I_{R_{j^*}}(t_{j^*}) &:= \sum_{x \in G \setminus B_\theta(v_0, R_{j^*})} u^2(x, t_{j^*}) \theta_x \\ &\leq \sup_{x \in G \setminus B_\theta(x_0, R_{j^*})} u(x, t_{j^*}) \sum_{x \in G \setminus B_\theta(x_0, R_{j^*})} u(x, t_{j^*}) \theta_x \\ &\leq \sup_{x \in G \setminus B_\theta(x_0, R_{j^*})} u(x, t_{j^*}). \end{aligned}$$

At this point, note that if  $t > 0$  is fixed, the function

$$\phi_t(d) := \exp \left( -\frac{1}{2} d \log \left( \frac{d}{2et} \right) \right)$$

is nonincreasing for  $d \geq 2t$ . Since  $R_{j^*} > 2e^2 t_{j^*}$ , we get

$$\begin{aligned} I_{R_{j^*}}(t_{j^*}) &\leq \sup_{x \in G \setminus B_\theta(x_0, R_{j^*})} u(x, t_{j^*}) \\ &\leq C_\theta^{-1} \phi_{t_{j^*}}(R_{j^*}) \\ &\leq C_\theta^{-1} \phi_{t_{j^*}}(2e^2 t_{j^*}) \\ &= C_\theta^{-1} \exp(-e^2 t_{j^*}) \\ &\leq C_\theta^{-1} \exp \left( -\frac{1}{12\gamma} R_0 \right). \end{aligned} \tag{4.8}$$

This is the only point in the argument at which we explicitly use the fact that the vertex weights are bounded below.

Now, we can put all of our estimates together. Combining (4.4), (4.6), (4.7), (4.8) we have

$$I_{R_0}(t_0) \leq m_0 \frac{1}{f(\alpha t_0)} \exp \left( -m_1 \frac{R_0^2}{t_0} \right) + n_0 \exp(-n_1 R_0),$$

where the constants  $\alpha, m_0, m_1, n_0, n_1$  may be taken to be

$$\alpha := \frac{2}{\gamma}, \quad m_0 := \exp \left( \frac{24\epsilon\gamma^3 e^2}{(\gamma-1)(2\gamma-1)} \right) \vee 2A, \quad m_1 := \delta\beta,$$

$$n_0 := C_\theta^{-1} \exp\left(\frac{24\varepsilon\gamma^2 e^2}{\gamma - 1}\right), \quad n_1 := \frac{1}{12\gamma}.$$

The fact that  $\gamma - 1$  can be very close to 0 is a potential concern. In practice, one will often have the choice of several values of  $\gamma$ ; for example, if  $f(t) = t^\alpha$ , one may choose any  $\gamma > 1$ . One also has the option of using the fact that  $(A, \gamma)$ -regularity implies  $(A^{2^n}, \gamma^{2^n})$ -regularity to increase  $\gamma$  at the cost of increasing  $A$  (and hence  $m_0$ ) also. However, choosing  $\gamma$  excessively large will cause  $\alpha$  and  $n_1$  to be undesirably close to zero.

□

## 5 Estimating a weighted sum of $u^2$

For  $H \subset G$ , let us define the following weighted sum of  $u^2$ ,

$$\begin{aligned} E_{\kappa, D, H}(x_0, t) &:= \sum_{x \in H} u^2(x, t) \exp\left(\kappa \frac{(d_\theta(x, x_0) \wedge D)^2}{t}\right) \theta_x \\ &= \sum_{x \in H} p_t^2(x, x_0) \exp\left(\kappa \frac{(d_\theta(x, x_0) \wedge D)^2}{t}\right) \theta_x. \end{aligned}$$

**Lemma 5.1.** *There exist constants  $\kappa_0, C, \alpha_0 > 0$  such that for  $t \geq \frac{1}{2} \vee \frac{D}{2}$ ,*

$$E_{\kappa_0, D, G}(x_0, t) \leq \frac{C}{f(\alpha_0 t)}.$$

*Proof.* Fix  $t \geq \frac{1}{2} \vee \frac{D}{2}$ , and choose  $\kappa_0$  to satisfy the inequalities  $16\kappa_0 - m_1 < 0$ ,  $8\kappa_0 - n_1 < 0$ , where  $m_1, n_1$  are the constants in Lemma 4.1.

We define  $k^*$  to be the largest nonnegative integer so that  $2^{k^*} \leq \sqrt{t}$  (if there is no such nonnegative integer, set  $k^* = 0$ ), and partition  $G$  as  $\bigcup_{0 \leq j \leq k^*+1} A_j$ , where

$$\begin{aligned} A_0 &:= \{x \in G : d_\theta(x_0, x) \leq \sqrt{t}\}, \\ A_k &:= \{x \in G : 2^{k-1}\sqrt{t} < d_\theta(x_0, x) \leq 2^k \sqrt{t}\} \text{ for } 1 \leq k \leq k^*, \\ A_{k^*+1} &:= \{x \in G : d_\theta(x_0, x) > 2^{k^*} \sqrt{t}\}. \end{aligned}$$

We turn our attention to the quantities  $E_{\kappa_0, D, A_j}(x_0, t)$  for  $0 \leq j \leq k^* + 1$ , which satisfy

$$E_{\kappa_0, D, G}(x_0, t) = \sum_{j=0}^{k^*+1} E_{\kappa_0, D, A_j}(x_0, t). \quad (5.1)$$

On  $A_0$ , the exponential weight  $\exp\left(\kappa_0 \frac{(d_\theta(x, x_0) \wedge D)^2}{t}\right)$  is bounded above by  $e^{\kappa_0}$ , and hence

$$E_{\kappa_0, D, A_0}(x_0, t) \leq e^{\kappa_0} \sum_{x \in A_0} u^2(x, t) \theta_x \leq e^{\kappa_0} \frac{1}{f(2t)} \leq e^{\kappa_0} \frac{1}{f(\alpha t)}. \quad (5.2)$$

For  $1 \leq j \leq k^*$ , on  $A_j$ , the exponential weight  $\exp\left(\kappa_0 \frac{(d_\theta(x, x_0) \wedge D)^2}{t}\right)$  is bounded above by  $\exp(\kappa_0 4^j)$ . Since  $2^{j-1} \sqrt{t} \leq t$ , we may apply the bound of Lemma 4.1 to obtain

$$\begin{aligned} \sum_{j=1}^{k^*} E_{\kappa_0, D, A_j}(x_0, t) &\leq \sum_{j=1}^{k^*} \exp(\kappa_0 4^j) I_{2^j \sqrt{t}}(t) \\ &\leq \sum_{j=1}^{k^*} \exp(\kappa_0 4^j) \left( m_0 \frac{1}{f(\alpha t)} \exp(-m_1 4^j) + n_0 \exp(-n_1 2^{j-1} \sqrt{t}) \right) \\ &= m_0 \frac{1}{f(\alpha t)} \sum_{j=1}^{k^*} \exp((4\kappa_0 - m_1) 4^{j-1}) + n_0 \sum_{j=1}^{k^*} \exp(2^{j-1} (2\kappa_0 2^j - n_1 \sqrt{t})) \\ &\leq m_0 \frac{1}{f(\alpha t)} \sum_{j=1}^{k^*} \exp((4\kappa_0 - m_1) 4^{j-1}) + n_0 \sum_{j=1}^{k^*} \exp(2^{j-1} (4\kappa_0 - n_1) \sqrt{t}) \\ &\leq m_0 \frac{1}{f(\alpha t)} \sum_{j=1}^{k^*} \exp((4\kappa_0 - m_1) 4^{j-1}) \\ &\quad + n_0 \exp((4\kappa_0 - n_1) \sqrt{t}) \sum_{j=1}^{k^*} \exp((2^{j-1} - 1)(4\kappa_0 - n_1) \sqrt{t}) \\ &\leq m_0 \frac{1}{f(\alpha t)} \sum_{j=1}^{k^*} \exp((4\kappa_0 - m_1) 4^{j-1}) \\ &\quad + n_0 \exp((4\kappa_0 - n_1) \sqrt{t}) \sum_{j=1}^{k^*} \exp\left(\frac{1}{\sqrt{2}} (2^{j-1} - 1)(4\kappa_0 - n_1)\right) \\ &\leq m_0 T_0 \frac{1}{f(\alpha t)} + n_0 T_1 \exp((4\kappa_0 - n_1) \sqrt{t}), \end{aligned}$$

where



$$T_0 := \sum_{j=1}^{\infty} \exp((4\kappa_0 - m_1)4^{j-1}) < \infty,$$

$$T_1 := \sum_{j=1}^{k^*} \exp\left(\frac{1}{\sqrt{2}}(2^{j-1} - 1)(4\kappa_0 - n_1)\right) < \infty.$$

By (1.5), we know that

$$\exp((4\kappa_0 - n_1)\sqrt{t}) \leq \frac{A}{f((4\kappa_0 - n_1)^2 t)},$$

so that

$$\sum_{j=1}^{k^*} E_{\kappa_0, D, A_j}(x_0, t) \leq (m_0 T_0 + n_0 T_1 A) \frac{1}{f((\alpha \vee (4\kappa_0 - n_1)^2) t)}. \quad (5.3)$$

On  $A_{k^*+1}$ , the exponential weight  $\exp\left(\kappa_0 \frac{(d_\theta(x, x_0) \wedge D)^2}{t}\right)$  is bounded above by  $\exp\left(\kappa_0 \frac{D^2}{t}\right) \leq \exp(4\kappa_0 t)$ , since  $D \leq 2t$ . By definition, we have  $\frac{1}{2}\sqrt{t} < 2^{k^*} \leq \sqrt{t}$ , and hence another application of Lemma 3.1 gives

$$\begin{aligned} E_{\kappa_0, D, A_{k^*+1}}(x_0, t) &\leq \exp(4\kappa_0 t) I_{2^{k^*} \sqrt{t}}(t) \\ &\leq \exp(4\kappa_0 t) I_{t/2}(t) \\ &\leq m_0 \frac{1}{f(\alpha t)} \exp\left(4\kappa_0 t - m_1 \frac{t}{4}\right) + n_0 \exp\left(4\kappa_0 t - n_1 \frac{t}{2}\right) \\ &= m_0 \frac{1}{f(\alpha t)} \exp\left(\frac{1}{4}(16\kappa_0 - m_1)t\right) + n_0 \exp\left(\frac{1}{2}(8\kappa_0 - n_1)t\right) \\ &= m_0 \frac{1}{f(\alpha t)} \exp\left(\frac{1}{8}(16\kappa_0 - m_1)\right) + n_0 \exp\left(\frac{1}{2}(8\kappa_0 - n_1)t\right) \\ &\leq m_0 \frac{1}{f(\alpha t)} \exp\left(\frac{1}{8}(16\kappa_0 - m_1)\right) + n_0 \exp\left(\frac{1}{2\sqrt{2}}(8\kappa_0 - n_1)\sqrt{t}\right). \end{aligned}$$

By (1.5) again,

$$\exp\left(\frac{1}{2\sqrt{2}}(8\kappa_0 - n_1)\sqrt{t}\right) \leq \frac{A}{f(1/8 \cdot (8\kappa_0 - n_1)^2 t)},$$

and so

$$E_{\kappa_0, D, A_{k^*+1}}(x_0, t) \leq (m_0 \exp\left(\frac{1}{8}(16\kappa_0 - m_1)\right) + n_0 A) \frac{1}{f((\alpha \vee 1/8 \cdot (8\kappa_0 - n_1)^2) t)}. \quad (5.4)$$

Combining (5.1) with (5.2),(5.3), and (5.4) completes the proof. □

## 6 Gaussian upper bounds for the heat kernel

We are now ready to prove Theorem 1.1.

*Proof.* Let  $D := d_\theta(x_1, x_2)$  and assume that  $t \geq 1 \vee D$ . Then  $\frac{t}{2} \geq \frac{1}{2} \vee \frac{D}{2}$ , so we may apply Lemma 5.1 with the points  $x_1$  and  $x_2$  (for which we have (1.6)) to obtain positive constants  $c$  and  $\alpha$  such that, for  $t \geq 1 \vee D$ ,

$$\begin{aligned} E_{c,D,G}(x_1, t/2) &\leq \frac{C}{f_1(\alpha t/2)}, \\ E_{c,D,G}(x_2, t/2) &\leq \frac{C}{f_2(\alpha t/2)}. \end{aligned}$$

The truncated distance  $\rho_\theta(x, y) := d_\theta(x, y) \wedge D$  satisfies  $d_\theta^2(x_1, x_2) = \rho_\theta^2(x_1, x_2) \leq 2(\rho_\theta^2(x_1, x) + \rho_\theta^2(x, x_2))$  for all  $x \in G$ . By using the semigroup property and Cauchy-Schwarz combined with the above considerations, we obtain, for all  $t \geq 1 \vee D$ ,

$$\begin{aligned} p_t(x_1, x_2) &= \sum_{x \in G} p_{t/2}(x_1, x) p_{t/2}(x, x_2) \theta_x \\ &\leq \sum_{x \in G} p_{t/2}(x_1, x) \exp\left(c \frac{\rho_\theta^2(x_1, x)}{t}\right) p_{t/2}(x, x_2) \exp\left(c \frac{\rho_\theta^2(x_2, x)}{t}\right) \exp\left(-c \frac{\rho_\theta^2(x_1, x_2)}{2t}\right) \theta_x \\ &\leq (E_{c,D,G}(x_1, t/2) E_{c,D,G}(x_2, t/2))^{1/2} \exp\left(-c \frac{\rho_\theta^2(x_1, x_2)}{2t}\right) \\ &\leq \frac{C}{(f_1(\alpha t/2) f_2(\alpha t/2))^{1/2}} \exp\left(-c \frac{d_\theta^2(x_1, x_2)}{2t}\right), \end{aligned}$$

which completes the proof of Gaussian upper bounds for the heat kernel. □

## 7 Restricted $(A, \gamma)$ -regular functions

In Section 4, where we estimated the quantity  $I_R(t)$ , we assumed that  $t_0 \geq R_0 \geq 1/2$ , and used  $(A, \gamma)$ -regularity to obtain, for  $0 \leq k \leq j^*$ ,

$$\frac{1}{f(2t_{k+1})} \leq \frac{1}{f(2t_0)} \left( A \frac{f(2t_0)}{f(2t_1)} \right)^{k+1}.$$

This is the only point at which  $(A, \gamma)$ -regularity is used. It follows that if  $f$  is merely  $(A, \gamma)$ -regular on  $(T_1, T_2)$ , then for this inequality to hold, we must have  $T_1 < 2t_{j^*+1}$  and  $2t_1 < \gamma^{-1}T_2$ . Subsequently, in Section 5, we apply our bounds for  $I_R(t)$  with  $t = t_0$  and  $R = 2^j\sqrt{t}$ , for  $0 \leq j \leq \sup\{k \in \mathbb{Z} : 2^k \leq \sqrt{t}\} \vee 0$ . Using (4.3), and setting  $t_0 = t/2$  (where  $t \geq 1 \vee D$ ), we see that these inequalities hold when

$$\begin{aligned} T_1 &< \frac{1}{6\gamma^2 e^2} (t/2)^{1/2}, \\ T_2 &> 2(t/2). \end{aligned}$$

Rearranging, we have

$$\begin{aligned} t &> 72e^4\gamma^4 T_1^2, \\ t &< T_2, \end{aligned}$$

and applying these additional constraints yields Theorem 1.3.

## 8 Applications to random walks on percolation clusters

In this section, we show how Theorem 1.3 may be used to obtain Gaussian upper bounds for the CSRW on the infinite component of supercritical bond percolation on the lattice  $\mathbb{Z}^d$  equipped with the standard weights. A detailed description of percolation is given in [16]; a percolation cluster is a random connected subgraph of the lattice  $\mathbb{Z}^d$  obtained by deleting each edge independently with probability  $1 - p$  and keeping it otherwise. By fundamental results of percolation theory, there exists a critical probability  $p_c(d)$  such that for  $p > p_c(d)$  (i.e., the supercritical case), there is an a.s. unique infinite cluster; we consider the CSRW on this family of random graphs, which we denote by  $\mathcal{C}_{p,\infty}(\omega)$ .

For existing work on random walks on percolation clusters, including on-diagonal heat kernel estimates and invariance principles, see [21] and [3]. From now on, we fix  $p > p_c(d)$ , and write  $q_t^\omega(x, y)$  for the heat kernel of the CSRW on  $\mathcal{C}_{p,\infty}(\omega)$ ; the dependence on  $\omega$  of  $q_t^\omega(x, y)$  is a consequence of  $\mathcal{C}_{p,\infty}(\omega)$  being random. We denote the graph metric on  $\mathcal{C}_{p,\infty}(\omega)$  by  $d_\mathcal{C}$ . In [21], Mathieu and Remy proved the following on-diagonal heat kernel bound for the CSRW on  $\mathcal{C}_{p,\infty}(\omega)$ .

**Lemma 8.1.** [21] *There exist random variables  $N_x(\omega) < \infty$  and non-random constants  $c_1, c_2$  such that almost surely, for all  $x \in G$  and  $t > 0$ ,*

$$q_t^\omega(x, x) \leq \begin{cases} c_1 t^{-1/2} & \text{if } 0 < t \leq N_x(\omega), \\ c_2 t^{-d/2} & \text{if } N_x(\omega) < t. \end{cases}$$

The polynomial function  $f(t) := c_2 t^{d/2}$  is  $(A, \gamma)$ -regular on  $(N_x(\omega), \infty)$  for  $A = 1$ ,  $\gamma = 2$ , and hence an application of Theorem 1.3 shows that for  $t \geq C(N_x(\omega) \vee N_y(\omega)) \vee 1 \vee d_{\mathcal{C}}(x, y)$ , we have the Gaussian upper bound

$$q_t^\omega(x, y) \leq C_1 t^{-d/2} \exp\left(-C_2 \frac{d_{\mathcal{C}}^2(x, y)}{t}\right), \quad (8.1)$$

where  $C_1, C_2 > 0$  are non-random constants.

**Remarks:**

1. For the discrete time simple random walk on  $\mathcal{C}_{p, \infty}(\omega)$ , Gaussian upper bounds are obtained in [7] as an application of their discrete time heat kernel estimates. However, the bounds in [7] have a random constant  $C_1 = C_1(\omega)$  in (8.1). The reason is that [7] only considers functions which are  $(A, \gamma)$ -regular, and in general the function  $f(t) := c_1^{-1} t^{1/2} \mathbf{1}_{\{0 < t \leq N_x(\omega)\}} + c_2^{-1} t^{d/2} \mathbf{1}_{\{N_x(\omega) < t\}}$  is not  $(A, \gamma)$ -regular. The authors of [7] therefore bound  $f(t)$  by a smaller random function  $g(t) := d_1 t^{1/2} \mathbf{1}_{\{0 < t \leq N_x(\omega)\}} + d_2 t^{d/2} \mathbf{1}_{\{N_x(\omega) < t\}}$ , where  $d_1 = d_1(\omega)$  and  $d_2 = d_2(\omega)$  are random constants chosen to ensure that  $f \geq g$  and  $g$  is  $(A, \gamma)$ -regular.

2. Theorem 1.3 is also used in [1] to obtain Gaussian upper bounds for the heat kernel in the random conductance model; as in the case of supercritical percolation clusters, the function appearing in the on-diagonal heat kernel estimate of Proposition 4.1 of [1] is not  $(A, \gamma)$ -regular but rather  $(A, \gamma)$ -regular on  $(T, \infty)$  for some  $T > 0$ , so Theorem 1.3 yields Gaussian upper bounds for all sufficiently large times.

**Acknowledgement:** The author thanks his Ph.D supervisor, Martin Barlow, for suggesting this problem and providing helpful discussions and feedback.

**References**

[1] S. Andres, M. T. Barlow, J.-D. Deuschel, B. M. Hambly. Invariance principle for the random conductance model. Preprint.

[2] M. T. Barlow. Random Walks on Graphs. Unpublished manuscript.

[3] M. T. Barlow. Random walks on supercritical percolation clusters, *Ann. Probab.* **32** (2004), 3024-3084. MR2094438

[4] M. T. Barlow, J.-D. Deuschel. Invariance principle for the random conductance model with unbounded conductances, *Ann. Probab.* **38** (2010), 234-276. MR2599199

[5] M. T. Barlow, R. F. Bass. Random walks on graphical Sierpinski carpets. In: *Random walks and discrete potential theory*, ed. M Picardello, W. Woess, Cambridge University Press, 1999. MR1802425

- [6] E. A. Carlen, S. Kusuoka, D. W. Stroock. Upper bounds for symmetric Markov transition functions, *Ann. I. H. Poincaré-PR* (1987), 245-287. MR898496
- [7] T. Coulhon, A. Grigor'yan, F. Zucca. The discrete integral maximum principle and its applications, *Tohoku Math. J.* **57** (2005), 559-587. MR2203547
- [8] E. B. Davies. Large deviations for heat kernels on graphs, *J. London Math. Soc.* **47** (1993), 65-72. MR1200978
- [9] E. B. Davies. Analysis on graphs and noncommutative geometry, *J. Funct. Anal.* **111** (1993), 398-430. MR1203460
- [10] E. B. Davies. Explicit constants for Gaussian upper bounds on heat kernels, *Am. J. Math.* **109** (1987), 319-334. MR882426
- [11] T. Delmotte. Parabolic Harnack inequality and estimates of Markov chains on graphs, *Rev. Mat. Iberoam.* **15** (1999), 181-232. MR1681641
- [12] R. Frank, D. Lenz, D. Wingert. Intrinsic metrics for non-local symmetric Dirichlet forms and applications to spectral theory. Preprint.
- [13] A. Grigor'yan. Gaussian upper bounds for the heat kernel on arbitrary manifolds, *J. Differ. Geom.* **45** (1997), 33-52. MR1443330
- [14] A. Grigor'yan. Heat kernel upper bounds on a complete non-compact manifold, *Rev. Mat. Iberoam.* **10** (1994), 395-452. MR1286481
- [15] A. Grigor'yan, X. Huang, J. Masamune. On stochastic completeness of jump processes. To appear in *Math. Z.*
- [16] G.R. Grimmett. Percolation (2nd edition), Springer-Verlag, Berlin, 1999. MR1707339
- [17] W. Hebisch, L. Saloff-Coste. Gaussian estimates for Markov chains and random walks on groups, *Ann. Probab.* **21** (1993), 673-709. MR1217561
- [18] W. Hebisch, L. Saloff-Coste. On the relation between elliptic and parabolic Harnack inequalities, *Ann. I. Fourier* **51** (2001), 1437-1481. MR1860672
- [19] M. Keller, D. Lenz. Unbounded Laplacians on graphs: basic spectral properties and the heat equation, *Math. Model. Nat. Phenom.* **5** (2010), 198-224. MR2662456
- [20] M. Keller, D. Lenz. Dirichlet forms and stochastic completeness of graphs and subgraphs. To appear in *J. Reine. Angew. Math.*
- [21] P. Mathieu, E. Remy. Isoperimetry and heat kernel decay on percolation clusters, *Ann. Probab.* **32** (2004), 100-128. MR2040777
- [22] N. Th. Varopoulos. Hardy-Littlewood theory for semigroups, *J. Funct. Anal.* **63** (1985), 240-260. MR803094
- [23] N. Varopoulos, L. Saloff-Coste, T. Coulhon. Analysis and Geometry on Groups, Cambridge University Press, Cambridge, U.K., 1992. MR1218884

- [24] W. Woess. Random Walks on Infinite Graphs and Groups, Cambridge University Press, Cambridge, U.K., 2000. MR1743100
- [25] A. Weber. Analysis of the physical Laplacian and the heat flow on a locally finite graph, *J. Math. Anal. Appl.* **370** (2010), 146-158. MR2651136