arXiv:2611.15549v1 [math.OC] 11 Nov 2026

CONTROLLING AVALANCHE SIZES IN THE ABELIAN SANDPILE MODEL

A PREPRINT

Maike C. de Jongh[a] m.c.dejongh@utwente.nl

Antal A. Járai[b] a.jarai@bath.ac.uk

Richard J. Boucherie[a] r.j.boucherie@utwente.nl

M.N.M. van Lieshout[c,a] m.n.m.van.lieshout@cwi.nl

[a]Department of Applied Mathematics, University of Twente,
P.O. Box 217, 7500 AE Enschede, The Netherlands

[b]Department of Mathematical Sciences, University of Bath,
Claverton Down, Bath BA2 7AY, United Kingdom

[c]Centrum Wiskunde & Informatica,
P.O. Box 94079, 1090 GB Amsterdam, The Netherlands

November 13, 2026

ABSTRACT

The Abelian sandpile model is a classical framework for studying large cascading events, such as those arising in models of forest fires and earthquakes. This work studies the control of such avalanches through the targeted removal of sand grains. We formulate this sequential control problem as an average-cost Markov decision process in which, before each grain addition, a controller may remove a sand grain from a selected vertex. The objective is to balance the long-run expected avalanche size against the cost of intervention. We analyze local intervention policies, which remove grains from a prescribed set of vertices whenever possible, and establish structural results for the recurrent classes and stationary distributions of the induced Markov chains. When the intervention set satisfies a property called topple-freeness, we sharpen these results, derive expressions for the gain and bias functions, and give necessary and sufficient conditions for optimality on general finite graphs. Finally, we apply the general theory to several specific graph classes and obtain a complete characterization of optimal policies on the one-dimensional lattice, the complete graph and the star graph for every value of the intervention cost.

Contents

1 Introduction **2**
1.1 Related work 3
1.2 Organization 4

2 Model definition **4**
2.1 The Abelian sandpile model 4

2.2 The sandpile Markov decision process 6
2.3 The average-cost optimality equations 7

3 Local intervention policies on general graphs 8
3.1 The never-intervene policy 8
3.2 Structural results for local intervention policies 11
3.3 Topple-free intervention sets 14
3.3.1 Recurrent classes and stationary distributions under topple-freeness 14
3.3.2 Gain, bias and optimality 15

4 Applications to specific graph classes 22
4.1 The one-dimensional lattice 22
4.2 The two-dimensional lattice 30
4.3 The complete graph 31
4.4 The star graph 33

5 Discussion 34

1 Introduction

The Abelian sandpile model, introduced by Bak et al. [1, 2] and Dhar [3], provides a canonical framework for studying self-organized criticality. The model describes a dynamical system defined on a graph in which sand grains are successively added at randomly selected vertices. When the number of grains at a vertex exceeds its capacity, the vertex topples and distributes grains among its neighbors. This may cause topplings of other vertices, giving rise to an avalanche that can propagate across a large part of the graph. Grains leave the system at a sink vertex, playing the role of boundary. Critical behavior emerges in the model despite the absence of a parameter that can be tuned to drive the system towards a critical point. This mechanism, known as self-organized criticality (SOC) [4], has been associated with phenomena such as earthquakes [5], forest fires [6], fluctuations in financial markets [7], and evolutionary dynamics [8].

Since large-scale cascades can have destructive consequences in applied settings, a natural question is to what extent their occurrence or impact can be reduced by means of external control. In this paper, we use the framework of Markov decision processes (MDP) [9] to give an analysis of the above question in the context of the Abelian sandpile model, focusing on strategies that are optimal in the long-run. This extends the earlier work of de Jongh et al. [10], which studied the immediate rather than the long-run effect of interventions. While the Abelian sandpile model only shares broad qualitative features with real avalanches, we believe that its study provides useful insight into controlled SOC. The main difficulty in the analysis of SOC is the inherent strong spatial and temporal correlations of the cascades, and some aspects of these features are reasonably well understood at the mathematical level in the sandpile model. Namely, when the cost associated with an avalanche is *linear* in the number of topplings involved, Dhar's formula [3] relates expected toppling numbers to the Green's function of the underlying graph. This relation plays an important role in our proof of Theorem 3.1 and in the explicit calculations for specific graph classes in Section 4.

The setup of our work is as follows. We formulate the problem of controlling avalanche sizes through grain-removal interventions as an average-cost MDP. Before each grain addition, the controller may either take no action or remove one grain from a selected vertex. The cost function of the MDP is the sum of an immediate intervention cost β (when the controller decides to intervene) and the expected size of the subsequent avalanche, thereby capturing a trade-off. Since new sand grains arrive at each time step, our aim is to identify intervention strategies that achieve a sustained reduction in expected avalanche size relative to their intervention costs. The objective, therefore, is to minimize the long-run average cost per grain addition, or the *gain*. This formulation also allows us to exploit the stationary structure of the controlled sandpile dynamics to derive explicit expressions for the gain and bias and establish conditions for

optimality. Greater weight could instead be assigned to costs in the near future by using a discounted cost criterion [9]. This provides a different perspective that we do not consider in this work.

We study *local intervention policies* (see Definition 2.1), which remove grains from a prescribed set W of vertices whenever possible. We establish structural results for the recurrent classes and stationary distributions of the Markov chains induced by these policies. The special case $W = \emptyset$ corresponds to the *never-intervene policy*. We obtain stronger results when the intervention set W is *topple-free* under the policy, which means that in stationarity, no vertex in W topples during a transition. We give the precise definition in Definition 3.6 and identify several graph classes and intervention sets for which this property holds. Under the topple-free condition, we obtain sharper characterizations of recurrent classes and stationary distributions, as well as explicit expressions for the gain (i.e., the long-run average cost) and bias (i.e., the relative transient cost) functions (precise definitions of these are given in Section 2.3) and conditions for the optimality of the corresponding policy. In addition to these results being of relevance for the control problem, we believe they are also of interest from the point of view of the theory of Abelian sandpiles: under the never-intervene policy, we recover the usual sandpile dynamics, whereas the controlled Markov chains represent perturbations of the usual sandpile dynamics that preserve some, but need not retain all, of the structure of the original model.

One of our main contributions is the following theorem, stated here somewhat informally; see Theorem 3.10 for the precise version.

Theorem 1.1. *Consider a local intervention policy with intervention set W, and suppose that W is topple-free under this policy. There are explicit bounds $\underline{\beta}_W, \overline{\beta}_W \in \mathbb{R}$, depending only on W and the graph, such that if $\underline{\beta}_W \le \overline{\beta}_W$, then for every $\beta \in [\underline{\beta}_W, \overline{\beta}_W]$ the policy is optimal.*

The thresholds $\underline{\beta}_W, \overline{\beta}_W$ are expressed in terms of random walk quantities on the underlying graph.

As an application, we prove the following theorem, that identifies optimal policies for every $\beta \ge 0$ in one dimension. Consider the one-dimensional lattice $\{1, \dots, L\}$ with vertices 1 and L connected to the sink by a single edge. Let $K = \lfloor (L-1)/2 \rfloor$ and $W_k = \{k+1, \dots, L-k\}$ for $k = 0, \dots, K$. Define

$$\beta_k = \frac{k(k+1)}{2}, \qquad k = 0, \dots, K, \qquad \beta_{K+1} = \begin{cases} (L+1)^2/8, & \text{if } L \text{ is odd,} \\ L(L+2)/8, & \text{if } L\text{is even.} \end{cases}$$

Let π_0 be the always-intervene policy that removes the leftmost grain whenever the configuration is nonempty and for each $k = 1, \dots, K$, let π_k be any local intervention policy with intervention set W_k.

Theorem 1.2. *For each $k = 0, \dots, K$, the policy π_k is optimal if and only if $\beta \in [\beta_k, \beta_{k+1}]$. The never-intervene policy is optimal if and only if $\beta \ge \beta_{K+1}$.*

Our further contributions are the following. We derive general expressions for the gain and bias of the never-intervene policy and an exact threshold for its optimality on finite graphs. These results provide the foundation for the analysis of local intervention policies with general intervention sets. We apply the general theory for several graph classes, obtaining explicit quantities and graph-specific optimality results. For the complete graph and the star graph, these results yield a full characterization: for every value of β, an optimal policy can be found among the policies that either never intervene or always intervene.

1.1 Related work

Several intervention mechanisms have been studied for SOC systems. One line of research seeks to influence avalanche sizes by selecting the addition location of the next grain based on the current state of the system [11, 12, 13]. Related strategies deliberately trigger avalanches at vertices that are close to becoming unstable, aiming to prevent large-scale avalanches through more frequent, localized dissipation events [14, 15]. Since the addition site cannot be controlled in many applications, Qi and Pfenninger [16] instead introduce failure tolerance, which temporarily prevents the toppling of overloaded vertices. In a different model of cascading failure, Hou et al. [17] investigate how heterogeneous redistribution of the load of such vertices affects the size of the resulting avalanche. Control mechanisms for cascading

failures have also been studied in interdependent networks [18]. In a model of systemic risk in financial networks, Scala et al. [19] study targeted immunization policies that provide assistance to selected institutions.

The studies above rely primarily on numerical methods to investigate strategies for reducing avalanche sizes. A rigorous analysis of interventions was provided by De Jongh et al. [10], who consider removing sand grains from square-shaped connected regions consisting of vertices at maximal height. They quantify how such interventions affect the size of the next avalanche and identify optimal intervention locations that minimize its expected size. This paper extends their analysis from the effect of a single intervention on the next avalanche to the long-term performance of intervention strategies. We formulate this sequential control problem as an average-cost Markov decision process (MDP) [9].

MDPs have proved useful in a variety of decision-making problems involving spatially interacting or networked systems. For example, the framework has been used to derive optimal strategies for controlling droplet growth in the Ising model [20, 21, 22]. Other applications include the containment and suppression of wildfires [23, 24], epidemic control on social networks and under uncertain transmission dynamics [25, 26], the control of colloidal self-assembly processes for producing crystalline structures [27, 28] and optimal thinning strategies in forest management [29].

A related sequential control problem was studied by Cajueiro and Andrade [30], who introduce a dynamic programming formulation for the control of the directed Abelian Dhar-Ramaswamy model [31]. Before each addition of a new sand grain, the controller may either take no action or select a vertex at which to trigger an avalanche. The objective is to balance the cost of intervention against a penalty associated with avalanche size. Due to the exponential growth of the state space, an optimal strategy is computed numerically only for a 4×4 lattice. The resulting solution is then used as a benchmark for heuristics that can be applied to larger systems. Although our formulation shares a sequential decision-making perspective, the two control frameworks differ in several fundamental respects. While Cajueiro and Andrade [30] consider discounted control of a directed sandpile model through deliberately triggered avalanches, we consider average-cost control of the standard Abelian sandpile model through grain removal. Moreover, whereas their analysis of optimal policies is numerical and restricted to a fixed lattice size, our analysis yields rigorous structural results for general finite graphs, together with more detailed results for several specific graph classes.

1.2 Organization

The remainder of this paper is organized as follows. Section 2 formulates the control problem and provides the necessary background on the Abelian sandpile model and average-cost MDPs. Section 3 presents general results for local intervention policies, including sharper results under the topple-free condition. In Section 4, we apply the general theory to several specific graph classes, providing a complete characterization of the optimal policies on the one-dimensional lattice, the complete graph and the star graph for every value of β. In Section 5, we give some concluding remarks and directions for future research.

2 Model definition

In this section, we formally define the control problem studied in this work and provide the necessary background on the Abelian sandpile model and average-cost MDPs.

2.1 The Abelian sandpile model

We consider the Abelian sandpile model on a finite, undirected graph $\mathcal{G} = (V \cup \{s\}, E)$, with vertex set V, edge set E, and a distinguished vertex s called the sink. Let $\deg(x)$, $x \in V$, denote the degree of x in $\mathcal{G}$, including edges incident to the sink. We allow parallel edges, but assume that the graph contains no self-loops. Let b_{xy}, $x, y \in V$, denote the number of edges between x and y. A sandpile configuration on $\mathcal{G}$ is a function $\eta : V \to \mathbb{Z}_{\geq 0}$, where $\eta(x)$ denotes the height of the vertex $x \in V$, that is, the number of indistinguishable sand grains at x. We denote the set of all sandpile

configurations by $\mathcal{X}$. The total mass $m(\eta)$ of a sandpile $\eta \in \mathcal{X}$ is defined by

$$m(\eta) = \sum_{x \in V} \eta(x), \qquad \eta \in \mathcal{X}.$$

Let

$$V^+(\eta) = \{x \in V : \eta(x) > 0\}$$

denote the set of vertices containing at least one sand grain in the configuration $\eta \in \mathcal{X}$. A vertex $x \in V$ is stable in $\eta \in \mathcal{X}$ if $\eta(x) < \deg(x)$. A configuration is stable if each vertex is stable. Let the set of all stable sandpile configurations be denoted by Ω. We call a vertex $x \in V$ critical in $\eta \in \Omega$ if $\eta(x) = \deg(x) - 1$. A vertex $x \in V$ is noncritical if $\eta(x) < \deg(x) - 1$. For $i = 0, \ldots, \sum_{x \in V} (\deg(x) - 1)$, let

$$\Omega_i = \{\eta \in \Omega : m(\eta) = i\} \tag{1}$$

denote the set of stable configurations containing exactly i sand grains.

The sandpile Markov chain $(\eta_t)_{t \geq 0}$ on the state space Ω is defined by the following dynamics. At each discrete time $t = 0, 1, \ldots$, a vertex $x \in V$ is selected uniformly at random and a new sand grain is added at x. We denote the addition operator by $\alpha_x : \mathcal{X} \to \mathcal{X}$, where $\alpha_x \eta$ is obtained from η by adding one sand grain at x. If x becomes unstable due to this addition, then it topples, meaning that it sends one sand grain to each of its neighboring vertices. Any vertices that become unstable subsequently also topple, and this process continues until the configuration is stable. We refer to this procedure as the relaxation of the configuration. Sand grains that reach the sink during the relaxation are dissipated. Given a configuration $\eta \in \mathcal{X}$, we denote its stabilization by η°. A fundamental result due to Dhar [3], called the Abelian property, states that, starting from a configuration $\eta \in \Omega$ and adding a grain at a vertex x, the number of times each vertex $y \in V$ topples during the subsequent relaxation, denoted by $n_y(\eta, x)$, is independent of the order in which unstable vertices are toppled. In particular, $(\alpha_x \eta)^\circ$ is well-defined.

Let Δ denote the restriction of the graph Laplacian of $\mathcal{G}$ to $V \times V$, i.e.,

$$\Delta_{xy} = \begin{cases} \deg(x), & \text{if } x = y \in V; \\ -b_{xy}, & \text{if } x \neq y, \quad x, y \in V. \end{cases}$$

Identifying a configuration η with a row vector, the procedure of toppling a vertex x can be written as

$$\eta \to \eta - \Delta_{x,\cdot},$$

where $\Delta_{x,\cdot}$ denotes the row of Δ corresponding to x. The transition kernel of the sandpile Markov chain is given by

$$p(\eta, \eta') = \frac{1}{|V|} \sum_{x \in V} \mathbb{1}[(\alpha_x \eta)^\circ = \eta'], \qquad \eta, \eta' \in \Omega. \tag{2}$$

The sandpile Markov chain has a unique recurrent class and its stationary distribution, denoted by μ, is the uniform distribution on this class [3].

The main quantity of interest in this paper is the avalanche size, defined as the total number of topplings during the relaxation process. Our objective is to control the sandpile dynamics in such a way as to minimize the expected avalanche size. The avalanche size resulting from adding a grain at $x \in V$ to a configuration $\eta \in \Omega$ is therefore given by

$$n(\eta, x) = \sum_{y \in V} n_y(\eta, x).$$

Let $N(\eta)$ denote the size of the avalanche triggered by the next grain addition starting from a configuration $\eta \in \Omega$. Note that this is a random variable with distribution

$$\mathbb{P}(N(\eta) = k) = \frac{1}{|V|} \sum_{x \in V} \mathbb{1}[n(\eta, x) = k], \qquad k = 0, 1, \dots.$$

When we need to consider the full time-evolution $(\eta_t)_{t \geq 1}$, we distinguish the avalanche sizes as $N_1(\eta_1), N_2(\eta_2), \dots$. Here, conditional on $\eta_1 = \eta$, the random variable N_1 has the distribution of $N(\eta)$. The next configuration η_2 is a function of η_1 and the toppling numbers $(n_y(\eta_1, X_1))_{y \in V}$, where X_1 is the random location of the added grain. In general, denote the history of the process up to time t by $\mathcal{F}_t = \sigma(\eta_1, X_1, \dots, \eta_t, X_t)$. Then η_{t+1} is determined by η_t, X_t. On the event $\{\eta_{t+1} = \eta\}$, the random variable $N_{t+1}(\eta_{t+1})$ is a copy of $N(\eta)$ independent of $\mathcal{F}_t$.

2.2 The sandpile Markov decision process

We consider a setting in which the dynamics of the sandpile Markov chain can be controlled by means of local interventions. Specifically, before each new grain addition, an external decision maker may choose a vertex $x \in V$ from which to remove a sand grain at cost $\beta \geq 0$, or decide to refrain from intervening. The objective is to identify an optimal trade-off between minimizing the long-run expected avalanche size and the long-run average intervention cost. In order to analyze this optimization problem, we formulate it as a Markov decision process [9].

The state space of this MDP is the set of stable configurations Ω. The action space is given by

$$A = V \cup \{0\},$$

where action $a \in V$ corresponds to removing a sand grain from vertex a and action 0 corresponds to no intervention. The set of feasible actions in configuration $\eta \in \Omega$ is defined as

$$A(\eta) = V^+(\eta) \cup \{0\}.$$

Given a configuration $\eta \in \Omega$ and an action $a \in A(\eta)$, let η^a denote the post-decision configuration, defined by

$$\eta^a(x) = \begin{cases} \eta(x) - 1, & \text{if } a = x, \\ \eta(x), & \text{otherwise,} \end{cases} \qquad x \in V.$$

Thus, η^a is the configuration obtained immediately after applying action a, but before the addition of the new sand grain. The transition kernel $P : \Omega \times A \times \Omega \to [0, 1]$ of the MDP is given by

$$P(\eta' \mid \eta, a) = p(\eta^a, \eta'), \qquad \eta, \eta' \in \Omega, \quad a \in A(\eta),$$

where $p : \Omega \times \Omega \to [0, 1]$ is the transition kernel of the uncontrolled sandpile Markov chain defined in (2). The cost function $c : \Omega \times A \to \mathbb{R}$ consists of two components: the expected size of the next avalanche and the intervention cost. It is defined as

$$c(\eta, a) = \mathbb{E}[N(\eta^a)] + \beta \mathbb{1}[a \neq 0], \qquad \eta \in \Omega, \quad a \in A(\eta). \tag{3}$$

We assume that the decision maker follows a stationary deterministic policy, which is determined by a decision rule $d : \Omega \to A$ satisfying $d(\eta) \in A(\eta)$ for every $\eta \in \Omega$. Such a policy is denoted by $\pi = d^\infty$. Let Π denote the set of all stationary deterministic policies. Associated with a decision rule d, let $P_d : \Omega \times \Omega \to [0, 1]$ and $c_d : \Omega \to \mathbb{R}$ denote the transition kernel and cost function induced by d, namely,

$$P_d(\eta, \eta') = P(\eta' \mid \eta, d(\eta)), \quad \text{and} \quad c_d(\eta) = c(\eta, d(\eta)),$$

for $\eta, \eta' \in \Omega$. We refer to the Markov chain induced by π as the π-dynamics.

We focus on a particular class of policies that remove sand grains from a fixed set of vertices whenever possible. We call such policies *local intervention policies*.

Definition 2.1. Given a graph $\mathcal{G} = (V \cup \{s\}, E)$ and a subset $W \subseteq V$, a policy $\pi_W = (d_W)^\infty$ is called a local intervention policy with intervention set W if

$$d_W(\eta) \in \begin{cases} W \cap V^+(\eta) & \text{if } W \cap V^+(\eta) \neq \emptyset, \\ \{0\}, & \text{otherwise,} \end{cases} \qquad \eta \in \Omega.$$

Note that Definition 2.1 does not specify which occupied vertex in the set W is selected when there are multiple options. Hence, there may be different local intervention policies with the same intervention set. Whenever we explicitly specify the selection rule, we indicate this with a suitable superscript.

2.3 The average-cost optimality equations

To assess the performance of a policy $\pi \in \Pi$, we use the average-cost criterion. We consider repeated interventions in a system that is continually driven by the addition of new sand grains, without a terminal horizon or priority given to reducing earlier costs. Our objective is therefore to achieve a sustained reduction in expected avalanche costs relative to the intervention costs. The average-cost formulation also connects the control problem to the recurrent classes and stationary behavior of the perturbed sandpile Markov chain, which are of independent interest in the theory of the Abelian sandpile model. The average-cost function, or gain, of a general stationary deterministic policy $\pi = d^\infty \in \Pi$ is defined as

$$g^\pi(\eta) = \lim_{T\to\infty} \frac{1}{T} \mathbb{E}_\eta^\pi \left[\sum_{t=1}^{T} c_d(\eta_t) \right], \quad \eta \in \Omega, \tag{4}$$

where $(\eta_t)_{t\geq 1}$ denotes the sandpile configuration process under the policy π. Its law and initial configuration $\eta_1 = \eta$ are indicated by the superscript and subscript in the expectation. The sandpile MDP is multichain, meaning that the Markov chain induced by a stationary deterministic policy may have multiple recurrent classes. This occurs, for example, for the always-intervene policy on the star graph, see Lemma 4.13. Therefore, the gain may in general depend on the initial configuration.

Given a stationary deterministic policy $\pi \in \Pi$, let $R_1, \ldots, R_k$ denote the recurrent classes of the Markov chain induced by this policy. Let $\mu_{\pi,i}$ denote the stationary distribution on R_i. Moreover, for an initial sandpile configuration $\eta \in \Omega$, let $\delta_i(\eta)$ denote the probability that the chain eventually enters the recurrent class R_i. The gain can be written as [9]

$$g^\pi(\eta) = \sum_{i=1}^{k} \delta_i(\eta) \sum_{\eta' \in R_i} \mu_{\pi,i}(\eta') c_d(\eta'). \tag{5}$$

Let the bias of a policy $\pi = d^\infty \in \Pi$ be defined as[1]

$$h^\pi(\eta) = \lim_{T\to\infty} \mathbb{E}_\eta^\pi \left[\sum_{t=1}^{T} [c_d(\eta_t) - g^\pi(\eta_t)] \right]. \tag{6}$$

The bias can be interpreted as the expected cumulative deviation of the one-stage costs from the long-run average cost under policy π, starting from configuration η. Since the configuration space is finite, the gain g^π and bias h^π of a policy $\pi = d^\infty \in \Pi$ satisfy the following policy evaluation equations [9, Thm. 8.2.6]

$$(I - P_d) g^\pi = 0, \tag{7a}$$

$$c_d - g^\pi + (P_d - I) h^\pi = 0, \tag{7b}$$

where g^π, h^π, and c_d are regarded as column vectors indexed by the configuration space. The policy evaluation equations (7a) and (7b) uniquely determine the gain of policy π, whereas they determine the bias only up to addition

[1] The definition assumes that under policy π, the controlled Markov chain is aperiodic; otherwise the limit is replaced by a Cesaro-limit (see [9, Section 8.2.1]). However, we will be able to by-pass periodicity issues.

of an element of the null space of the matrix $I - P_d$. We call any function h^π satisfying the equations a bias function associated with the policy π.

A policy π^* is called average-cost optimal if it satisfies

$$g^{\pi^*}(\eta) \leq g^\pi(\eta), \qquad \text{for all } \eta \in \Omega, \quad \pi \in \Pi.$$

The optimal gain is

$$g^*(\eta) = \min_{\pi \in \Pi} g^\pi(\eta).$$

Since the state and action spaces are finite and the cost function is bounded, there exists an optimal stationary deterministic policy [9, Thms. 9.1.4 and 9.1.8]. This justifies restricting our attention to this class of policies.

Let $\mathcal{F}(\Omega)$ denote the space of real-valued functions on Ω. The optimal gain can be characterized by means of the following system of optimality equations [9, Thm. 9.1.3]:

$$\min_{a \in A(\eta)} \left\{ \sum_{\eta' \in \Omega} P(\eta'|\eta, a) g(\eta') - g(\eta) \right\} = 0, \tag{8a}$$

$$\min_{a \in B_\eta(g)} \left\{ c(\eta, a) - g(\eta) + \sum_{\eta' \in \Omega} P(\eta'|\eta, a) h(\eta') - h(\eta) \right\} = 0, \tag{8b}$$

where

$$B_\eta(g) = \left\{ a \in A(\eta) : \sum_{\eta' \in \Omega} P(\eta'|\eta, a) g(\eta') - g(\eta) = 0 \right\}.$$

The existence of a solution to these equations is established in [9, Thm. 9.1.4]. The function g in every solution equals the optimal gain g^* [9, Thm. 9.1.3]. Conversely, if $(g, h) \in \mathcal{F}(\Omega) \times \mathcal{F}(\Omega)$ is a pair of functions that satisfies the optimality equations (8a, 8b) and a decision rule $d^* \in D$ satisfies

$$P_{d^*} g = g,$$
$$d^* \in \arg\min_{d \in E(g)} \{c_d + P_d h\},$$

where $E(g) = \{d \in D : d(\eta) \in B_\eta(g)\}$, then the policy $\pi^* = (d^*)^\infty$ is average-cost optimal [9, Thm. 9.1.7].

3 Local intervention policies on general graphs

In this section, we study the controlled sandpile Markov chain under local intervention policies on general finite graphs. First, we analyze the local intervention policy for which the intervention set is empty, i.e., the policy that never intervenes. These results form the basis for the analysis of local intervention policies with general intervention sets. We provide structural results concerning the recurrent classes and stationary distributions of the Markov chains induced by local interventions. Then, we introduce the concept of topple-freeness and show that under this condition many of these results can be sharpened. Finally, under the topple-freeness condition, we provide explicit expressions for the gain and bias of local intervention policies and bounds on the immediate intervention cost β that guarantee their optimality.

3.1 The never-intervene policy

In this subsection, we analyze the local intervention policy $\pi_\emptyset = (d_\emptyset)^\infty$ with empty intervention set, which we call the *never-intervene policy*. Under this policy, the dynamics coincide with those of the uncontrolled sandpile Markov chain. We derive explicit expressions for its gain and bias and determine the exact threshold above which it is optimal.

Given a graph $\mathcal{G} = (V \cup \{s\}, E)$, let

$$G = \Delta^{-1}$$

denote the Green's function of $\mathcal{G}$. For any vertex $x \in V$, we define

$$\phi(x) = \sum_{y \in V} G(x, y). \tag{9}$$

The function ϕ admits the following probabilistic interpretation. Let

$$\tau_x = \inf\{t \geq 0 : X_t = x\}, \qquad x \in V \cup \{s\},$$

denote the first hitting time of vertex x of a continuous-time random walk $(X_t)_{t \geq 0}$ with jump rate 1 along each incident edge. This walk is the continuous-time Markov chain with generator $-\Delta$, killed when it reaches s. Then (see e.g. [32, Section 4.4, equation 4.21])

$$\phi(x) = \mathbb{E}_x[\tau_s]. \tag{10}$$

Theorem 3.1 expresses the gain and bias of the never-intervene policy in terms of the function ϕ.

Theorem 3.1. *For the policy $\pi_\emptyset = (d_\emptyset)^\infty$, where $d_\emptyset(\eta) = 0$ for all $\eta \in \Omega$, the gain is given by*

$$g^{\pi_\emptyset}(\eta) = \frac{1}{|V|} \sum_{x \in V} \phi(x), \qquad \eta \in \Omega. \tag{11}$$

Moreover, there exists a constant $C \in \mathbb{R}$ such that

$$h^{\pi_\emptyset}(\eta) = \sum_{x \in V} (\eta(x) - \deg(x) + 1)\phi(x) + C, \qquad \eta \in \Omega. \tag{12}$$

Proof. Let $g^{\pi_\emptyset}$ and $h^{\pi_\emptyset}$ be given by (11) and (12). Since $g^{\pi_\emptyset}$ is constant, (7a) holds immediately. It remains to verify that $g^{\pi_\emptyset}$ and $h^{\pi_\emptyset}$ satisfy (7b). From the definition of the stabilization process, we have

$$(\alpha_x \eta)^\circ(y) = (\alpha_x \eta)(y) - \sum_{z \in V} n_z(\eta, x)\Delta_{zy}, \qquad y \in V.$$

Hence, we obtain

$$\begin{aligned} h^{\pi_\emptyset}((\alpha_x \eta)^\circ) - h^{\pi_\emptyset}(\eta) &= \sum_{y \in V} \left[\mathbb{1}_x(y) - \sum_{z \in V} n_z(\eta, x)\Delta_{zy} \right] \phi(y) \\ &= \phi(x) - \sum_{z \in V} n_z(\eta, x) \sum_{y \in V} \Delta_{zy}\phi(y). \end{aligned}$$

Since $G = \Delta^{-1}$ and $\phi = G\mathbf{1}$, we have

$$\sum_{y \in V} \Delta_{zy}\phi(y) = 1, \qquad \text{for all } z \in V.$$

Hence,

$$h^{\pi_\emptyset}((\alpha_x \eta)^\circ) - h^{\pi_\emptyset}(\eta) = \phi(x) - n(\eta, x).$$

Thus, we obtain

$$((P_{d_\emptyset} - I)h^{\pi_\emptyset})(\eta) = \frac{1}{|V|} \sum_{x \in V} \phi(x) - \mathbb{E}[N(\eta)].$$

Using expression (11) and the fact that

$$c_{d_\emptyset}(\eta) = \mathbb{E}[N(\eta)] \quad \text{for all } \eta \in \Omega,$$

we conclude that

$$c_{d_\emptyset} - g^{\pi_\emptyset} + (P_{d_\emptyset} - I)h^{\pi_\emptyset} = 0.$$

Since the Markov chain induced by $\pi_\emptyset$ has a unique recurrent class, the policy evaluation equations determine the bias up to an additive constant. Hence, every bias function associated with $\pi_\emptyset$ has the form (12). □

Remark 3.2. Formula (11) can alternatively be obtained directly from the definition of the gain (4) and Dhar's formula [3].

Using the expressions (11) and (12) for the gain and bias of the never-intervene policy, we characterize the values of the immediate intervention cost β for which this policy is optimal.

Theorem 3.3. *Consider the policy $\pi_\emptyset = (d_\emptyset)^\infty$, where $d_\emptyset(\eta) = 0$ for all $\eta \in \Omega$. Policy $\pi_\emptyset$ is optimal if and only if*

$$\beta \geq \max\{\phi(a) : a \in V,\ \deg(a) \geq 2\}. \tag{13}$$

Proof. In order to prove sufficiency, we verify that $(g^{\pi_\emptyset}, h^{\pi_\emptyset})$ as defined in expressions (11) and (12) satisfy the average-cost optimality equations (8a) and (8b) whenever condition (13) holds. The validity of equation (8a) follows immediately from the fact that $g^{\pi_\emptyset}$ is constant over all sandpile configurations. We continue to prove equation (8b). Since $g^{\pi_\emptyset}$ is constant, the set $B_\eta(g^{\pi_\emptyset})$ equals $A(\eta)$ for every $\eta \in \Omega$. Observe that the expressions (11) and (12) satisfy the policy evaluation equation (7b) for policy $\pi_\emptyset$. Hence, action 0 attains the minimum in equation (8b) if

$$c(\eta, 0) + \sum_{\eta' \in \Omega} P(\eta'|\eta, 0) h^{\pi_\emptyset}(\eta') \leq c(\eta, a) + \sum_{\eta' \in \Omega} P(\eta'|\eta, a) h^{\pi_\emptyset}(\eta'), \qquad \text{for all } \eta \in \Omega, \quad a \in A(\eta). \tag{14}$$

Using the facts that

$$c(\eta, a) = \beta \mathbb{1}\{a \neq 0\} + c(\eta^a, 0),$$

and

$$P(\eta'|\eta, a) = P(\eta'|\eta^a, 0),$$

together with the policy evaluation equation (7b) for $\pi_\emptyset$, expression (14) is equivalent to

$$g^{\pi_\emptyset} + h^{\pi_\emptyset}(\eta) \leq \beta \mathbb{1}\{a \neq 0\} + g^{\pi_\emptyset} + h^{\pi_\emptyset}(\eta^a), \qquad \eta \in \Omega, \quad a \in A(\eta).$$

It follows that expression (14) holds if and only if

$$\beta \mathbb{1}\{a \neq 0\} \geq h^{\pi_\emptyset}(\eta) - h^{\pi_\emptyset}(\eta^a), \qquad \eta \in \Omega, \quad a \in A(\eta).$$

Since η and η^a differ only at site a, expression (12) ensures that the optimality equations hold whenever

$$\beta \geq \phi(a) \quad \text{for all } a \in V \text{ for which there exists } \eta \in \Omega \text{ with } \eta(a) > 0. \tag{15}$$

Since there exists a stable configuration $\eta \in \Omega$ with $\eta(a) > 0$ if and only if $\deg(a) \geq 2$, the above condition is equivalent to (13), which establishes sufficiency.

We proceed to prove necessity. Given $\eta \in \Omega$, $a \in A(\eta)$, and $x \in V$, the definition of stabilization yields

$$(\alpha_x \eta^a)^\circ(y) = \eta^a(y) + \mathbb{1}_x(y) - \sum_{u \in V} n_u(\eta^a, x) \Delta_{uy}.$$

Using the explicit expression (12) and the identity $(\Delta\phi)(u) = 1$, we obtain

$$h^{\pi_\emptyset}((\alpha_x \eta^a)^\circ) - h^{\pi_\emptyset}(\eta) = \phi(x) - n(\eta^a, x) - \phi(a) \mathbb{1}\{a \neq 0\},$$

where we set $\phi(0) = 0$. Therefore,

$$n(\eta^a, x) + \beta \mathbb{1}\{a \neq 0\} = \phi(x) - [h^{\pi_\emptyset}((\alpha_x \eta^a)^\circ) - h^{\pi_\emptyset}(\eta)] + (\beta - \phi(a)) \mathbb{1}\{a \neq 0\}.$$

Taking the expectation over the addition site now yields

$$c(\eta, a) = g^{\pi_\emptyset} + h^{\pi_\emptyset}(\eta) - \sum_{\eta' \in \Omega} P(\eta'|\eta, a) h^{\pi_\emptyset}(\eta') + (\beta - \phi(a)) \mathbb{1}\{a \neq 0\}. \tag{16}$$

Suppose that $\beta < \phi(a)$ for some $a \in V$ with $\deg(a) \geq 2$. Consider the stationary deterministic policy $\pi_a = (d_a)^\infty$, where

$$d_a(\eta) = \begin{cases} a, & \text{if } \eta(a) > 0, \\ 0, & \text{if } \eta(a) = 0. \end{cases} \tag{17}$$

Let R denote a recurrent class of the Markov chain induced by π_a. We claim that there is at least one configuration $\eta \in R$ with $\eta(a) > 0$. Indeed, suppose that $\eta(a) = 0$ for all $\eta \in R$. Then, for any $\eta \in R$, there is no intervention by the policy. If the next grain is added at a, which occurs with positive probability, we obtain a configuration η' with $\eta'(a) = 1$, since $\deg(a) \geq 2$. By assumption, $\eta' \notin R$, which contradicts the closedness of R. Let $\mu_{\pi_a, R}$ denote the stationary distribution of the Markov chain induced by π_a on R. Since R is irreducible and contains at least one configuration ζ satisfying $\zeta(a) > 0$, we have

$$\sum_{\zeta \in R} \mu_{\pi_a, R}(\zeta) \mathbb{1}\{\zeta(a) > 0\} > 0.$$

Taking expectations in (16) with respect to $\mu_{\pi_a, R}$ and using its invariance under the π_a-dynamics, we obtain

$$g^{\pi_a}(\eta) = g^{\pi_\emptyset} + \sum_{\zeta \in R} \mu_{\pi_a, R}(\zeta) \mathbb{1}\{\zeta(a) > 0\} (\beta - \phi(a)) < g^{\pi_\emptyset}(\eta), \qquad \eta \in R.$$

Thus, $\pi_\emptyset$ is not optimal whenever $\beta < \phi(a)$ for some $a \in V$ with $\deg(a) \geq 2$. This establishes necessity. □

3.2 Structural results for local intervention policies

In this subsection, we show that, although a local intervention policy perturbs the sandpile dynamics, part of the structure of the uncontrolled sandpile Markov chain is preserved. In particular, we relate the recurrent configurations of the controlled Markov chain to those of the uncontrolled sandpile dynamics on the graph obtained by identifying all vertices in W with the sink. For singleton intervention sets, we also establish a relation between the stationary distributions.

Given a graph $\mathcal{G} = (V \cup \{s\}, E)$, let

$$\tau_W = \inf\{t \geq 0 : X_t \in W\}, \qquad W \subseteq V \cup \{s\},$$

denote the first hitting time of the set W by a simple random walk $(X_t)_{t \geq 0}$. For a singleton set $\{x\}$, we simply write τ_x and we set $\tau_\emptyset = \infty$. Furthermore, for $W \subseteq V$ and $x \in V \setminus W$, let

$$p_W(x) = \mathbb{P}_x(\tau_W < \tau_s)$$

denote the probability that a simple random walk started at x hits W before hitting the sink.

Let $\widetilde{\mathcal{G}}_W$ denote the graph obtained from $\mathcal{G}$ by contracting all vertices in $W \cup \{s\}$ into a single sink vertex. Every edge from a vertex in $V \setminus W$ to a vertex in $W \cup \{s\}$ becomes an edge to the contracted sink in this procedure. Edges with both endpoints in $W \cup \{s\}$ are discarded. We refer to this graph as the contracted graph. Let $\widetilde{\Omega}_W$ denote the set of stable configurations on $\widetilde{\mathcal{G}}_W$, and let $\widetilde{R}_W$ denote the unique recurrent class of the uncontrolled sandpile Markov chain on $\widetilde{\mathcal{G}}_W$. We let ϕ_W denote the analogue of the function ϕ defined in expression (9) for the contracted graph $\widetilde{\mathcal{G}}_W$. Furthermore, let R_{π_W} denote the set of recurrent configurations of the Markov chain on $\mathcal{G}$ induced by the policy π_W, which may be the union of several recurrent classes. Given a stationary deterministic policy $\pi \in \Pi$ for the sandpile MDP on the graph $\widetilde{\mathcal{G}}_W$, we let $\widetilde{g}^\pi_W$ and $\widetilde{h}^\pi_W$ denote its gain and bias functions, respectively. For $\eta \in \Omega$, we write

$$\xi = \eta|_{V \setminus W} \in \widetilde{\Omega}_W$$

for the restriction of η to the graph $\widetilde{\mathcal{G}}_W$.

We first establish a relation between the recurrent configurations under a local intervention policy π_W and those of the uncontrolled sandpile model on the contracted graph $\widetilde{\mathcal{G}}_W$ in Theorem 3.4.

Theorem 3.4. *Given a graph $\mathcal{G} = (V \cup \{s\}, E)$ and a subset $W \subseteq V$, let $\pi_W = (d_W)^\infty$ denote a local intervention policy. We have*

$$\{\eta|_{V\setminus W} : \eta \in R_{\pi_W}\} \subseteq \widetilde{R}_W.$$

Proof. Consider a configuration $\eta \in R_{\pi_W}$ and let $\xi = \eta|_{V\setminus W}$. Let η' denote the result of one transition under the policy π_W, obtained from an intervention by the policy, addition at a vertex $x \in V$ and subsequent stabilization, i.e. $\eta' = (\alpha_x \eta^{d_W(\eta)})^\circ$. Recall that $n_y(\eta^{d_W(\eta)}, x)$ denotes the number of times a vertex $y \in V$ topples during this transition. We define

$$\gamma_u = \sum_{y\in W} n_y(\eta^{d_W(\eta)}, x) b_{yu}, \qquad u \in V \setminus W,$$

i.e., the total number of grains sent from W to u during the transition. By the Abelian property, the restriction $\xi' := \eta'|_{V\setminus W}$ can be obtained from ξ by first adding a grain at x if $x \notin W$, then adding γ_u grains at each vertex $u \in V \setminus W$ and finally stabilizing the resulting configuration on the contracted graph $\widetilde{\mathcal{G}}_W$. Letting α_u^k denote the operation of adding k grains to vertex u, for $k = 1, 2, \ldots$, and α_u^0 the identity operator, the configuration ξ' can be written as

$$\xi' = \left(\left(\prod_{u\in V\setminus W} \alpha_u^{\gamma_u} \right) \alpha_x^{\mathbb{1}\{x\notin W\}} \xi \right)^\circ .$$

Hence, the transition from ξ to ξ' is realized by a finite sequence of additions and relaxations on $\widetilde{\mathcal{G}}_W$. Since $\widetilde{R}_W$ is closed under addition and relaxation operators on $\widetilde{\mathcal{G}}_W$, it follows that

$$\xi \in \widetilde{R}_W \quad \Rightarrow \quad \xi' \in \widetilde{R}_W. \tag{18}$$

Now, suppose that the restriction $\xi \notin \widetilde{R}_W$. Since $\widetilde{R}_W$ is the unique recurrent class of the uncontrolled sandpile dynamics on $\widetilde{\mathcal{G}}_W$, there exists a finite sequence $x_1, \ldots, x_k \in V \setminus W$ such that,

$$(\alpha_{x_k} \cdots \alpha_{x_1} \xi)^\circ \in \widetilde{R}_W.$$

Consider the process on $\mathcal{G}$ under the π_W-dynamics. The same sequence of additions at $x_1, \ldots, x_k \in V \setminus W$ may cause topplings of vertices in the set W. For $y \in W$, let N_y denote the total number of topplings of y during these transitions. By the Abelian property, the resulting configuration restricted to $V \setminus W$ can be obtained by first performing the additions at $x_1, \ldots, x_k$ and stabilizing on $\widetilde{\mathcal{G}}_W$, then adding $\sum_{y\in W} N_y b_{yu}$ grains to each vertex $u \in V \setminus W$, and stabilizing again. Hence, it can be written as

$$\left(\left(\prod_{u\in V\setminus W} \alpha_u^{\sum_{y\in W} N_y b_{yu}} \right) (\alpha_{x_k} \cdots \alpha_{x_1} \xi)^\circ \right)^\circ .$$

Since $(\alpha_{x_k} \cdots \alpha_{x_1} \xi)^\circ \in \widetilde{R}_W$ and $\widetilde{R}_W$ is closed under addition and stabilization, this configuration also belongs to $\widetilde{R}_W$. Hence, we have shown that the set

$$\{\zeta \in \Omega : \zeta|_{V\setminus W} \in \widetilde{R}_W\}$$

is closed under the π_W-dynamics by (18) and that it can be reached from η with positive probability. Since $\xi \notin \widetilde{R}_W$, this set does not contain η. Therefore, there is a positive probability that the process leaves η and can never return, contradicting the recurrence of η. Thus, if $\eta \in R_{\pi_W}$, then $\eta|_{V\setminus W} \in \widetilde{R}_W$. □

The reverse inclusion in Theorem 3.4 does not hold in general, as the following counterexample demonstrates.

Counterexample to the reverse inclusion. Consider a graph $\mathcal{G} = (V \cup \{s\}, E)$, where $V = \{x, y\}$ and $E = \{(x, y), (y, s)\}$. Then, the process induced by the local intervention policy π_x has a unique recurrent class $R_{\pi_x} = \{(0, 1)\}$. Consequently,

$$\{\eta|_{V\setminus\{x\}} : \eta \in R_{\pi_x}\} = \{1\}.$$

On the other hand, the uncontrolled sandpile dynamics on the contracted graph $\widetilde{\mathcal{G}}_x$ has recurrent class $\widetilde{R}_x = \{0, 1\}$. This implies

$$\{\eta|_{V\setminus\{x\}} : \eta \in R_{\pi_x}\} \subsetneq \widetilde{R}_x.$$

Hence, the reverse inclusion in Theorem 3.4 does not hold in general. □

We proceed to investigate the stationary distributions of the Markov chains induced by local intervention policies with singleton intervention sets and relate them to those of the uncontrolled sandpile dynamics on the corresponding contracted graph. Let $\widetilde{\mu}_x$ denote the stationary distribution of the uncontrolled sandpile dynamics on the contracted graph $\widetilde{\mathcal{G}}_x$. A toppling of x in the original graph sends b_{xy} grains to each vertex $y \in V \setminus \{x\}$. Therefore, the effect of such a toppling on the restricted configuration, followed by stabilization on the contracted graph, is represented by the operator $B_x : \widetilde{\Omega}_x \to \widetilde{\Omega}_x$ defined as

$$B_x\xi = \left(\left[\prod_{y\in V\setminus\{x\}} \alpha_y^{b_{xy}}\right]\xi\right)^\circ.$$

Observe that the restriction of B_x to $\widetilde{R}_x$ is an element of the sandpile group $\mathcal{S}(\widetilde{\mathcal{G}}_x)$. Consequently, it generates a cyclic subgroup $\langle B_x\rangle \leq \mathcal{S}(\widetilde{\mathcal{G}}_x)$. We denote by $C_1, \ldots, C_N$ the orbits of B_x on $\widetilde{R}_x$. Since recurrent configurations can be identified with sandpile group elements, these orbits correspond to the cosets of the cyclic subgroup $\langle B_x\rangle$ in $\mathcal{S}(\widetilde{\mathcal{G}}_x)$. The following result shows that the projection of any stationary distribution μ_{π_x} onto $V \setminus \{x\}$ assigns equal mass to these orbits.

Theorem 3.5. *Given a graph $\mathcal{G} = (V \cup \{s\}, E)$ and a vertex $x \in V$, let μ_{π_x} denote any stationary distribution of the π_x-dynamics and let $\bar{\mu}_{\pi_x}$ denote the projection of μ_{π_x} onto $\widetilde{\Omega}_x$. Then, $\bar{\mu}_{\pi_x}$ assigns equal probability to the orbits $C_1, \ldots, C_N$ of B_x on $\widetilde{R}_x$, i.e.,*

$$\bar{\mu}_{\pi_x}(C_i) = \frac{1}{N}, \quad \textit{for all } i = 1, \ldots, N.$$

Proof. Let $\eta_0 \sim \mu_{\pi_x}$, and let $\xi_0 = \eta_0|_{\mathcal{G}\setminus\{x\}}$ denote its restriction to $\mathcal{G} \setminus \{x\}$. Observe that Theorem 3.4 implies that $\xi_0 \in \widetilde{\mathcal{R}}_x$. Consider the following coupling of the π_x-dynamics on $\mathcal{G}$ and the uncontrolled dynamics on the contracted graph $\widetilde{\mathcal{G}}_x$: at each time step, we select a vertex $y \in V$ uniformly at random. If $y \neq x$, we perform the transition under π_x in the first process, i.e., first remove a grain from x if possible, then add at y and relax. In the second process, we add at y and relax on $\widetilde{\mathcal{G}}_x$. If $y = x$, we perform the π_x-transition in the first process and do nothing in the second process. Let $\eta_t \in \Omega$ denote the configuration obtained after running the π_x-dynamics for t steps and let ξ_t denote the result of running the uncontrolled dynamics on the contracted graph $\widetilde{\mathcal{G}}_x$ under this coupling. Let k_t denote the total number of topplings of x up to time t in the evolution under the π_x dynamics. Then, by the Abelian property,

$$\eta_t|_{\mathcal{G}\setminus\{x\}} = B_x^{k_t}\xi_t.$$

This implies

$$\eta_t|_{\mathcal{G}\setminus\{x\}} \in C_i \Leftrightarrow \xi_t \in C_i, \quad i = 1, \ldots, N,$$

since C_i is invariant under B_x for each $i = 1, \ldots, N$. We now obtain

$$\bar{\mu}_{\pi_x}(C_i) = \lim_{T\to\infty} \mathbb{P}(\eta_T|_{\mathcal{G}\setminus\{x\}} \in C_i) = \lim_{T\to\infty} \mathbb{P}(\xi_T \in C_i) = \widetilde{\mu}_x(C_i) = \frac{|C_i|}{|\widetilde{\mathcal{R}}_x|},$$

since the stationary distribution of the second process is the same as that of the uncontrolled sandpile dynamics on the contracted graph. Since all cosets of a subgroup have the same cardinality, this yields

$$\bar{\mu}_{\pi_x}(C_i) = \frac{1}{N}, \quad \text{for all } i = 1, \dots, N.$$

□

3.3 Topple-free intervention sets

In this subsection, we introduce the notion of topple-freeness. We show that under this condition, the results in Section 3.2 can be considerably sharpened. For local intervention policies with topple-free intervention sets, we derive explicit expressions for the gain and bias and establish sufficient conditions for optimality.

Definition 3.6. Consider a graph $\mathcal{G} = (V \cup \{s\}, E)$ and a stationary deterministic policy π. A set $W \subseteq V$ is *topple-free* under π if, under every stationary distribution of the Markov chain induced by π, the probability that any vertex $x \in W$ topples during a transition is zero. In particular, we say that a vertex $x \in V$ is topple-free under π if the singleton set $\{x\}$ is topple-free under π.

3.3.1 Recurrent classes and stationary distributions under topple-freeness

Since vertices in a topple-free intervention set W do not topple during transitions from recurrent configurations of the Markov chain induced by the corresponding local intervention policy, more of the structure of the uncontrolled sandpile dynamics is preserved. We use this property to strengthen the structural results of the preceding subsection.

The following result shows that the reverse inclusion in Theorem 3.4 does hold under the topple-freeness condition. It shows that the projection of each recurrent class of the controlled dynamics onto the contracted graph equals the unique recurrent class of the uncontrolled dynamics on that graph.

Theorem 3.7. *Consider a graph $\mathcal{G} = (V \cup \{s\}, E)$ and a local intervention policy π_W with intervention set $W \subseteq V$. If W is topple-free under the policy π_W, then each recurrent class R of the Markov chain induced by π_W satisfies*

$$\{\eta|_{V\setminus W} : \eta \in R\} = \widetilde{R}_W. \tag{19}$$

Proof. Let R denote any recurrent class of the Markov chain induced by π_W. Let

$$\widetilde{R} = \{\eta|_{V\setminus W} : \eta \in R\},$$

i.e., the set of restrictions of configurations in R to $V \setminus W$. By Theorem 3.4, we have

$$\widetilde{R} \subseteq \widetilde{R}_W.$$

Consider a configuration $\eta \in R$ and let $\xi = \eta|_{V\setminus W}$ denote its restriction to $V \setminus W$. Since W is topple-free under π_W, no vertex in W topples during any transition from a configuration in R. Hence,

$$((\alpha_x \eta^{d_W(\eta)})^\circ)|_{V\setminus W} = (\alpha_x \xi)^\circ, \qquad \text{for all } x \in V \setminus W,$$

where addition and stabilization on the right-hand side are performed on the contracted graph $\widetilde{\mathcal{G}}_W$.

Since R is closed under the π_W-dynamics, it follows that $\widetilde{R}$ is closed under addition at every vertex $x \in V \setminus W$ and subsequent stabilization. Since $\widetilde{R}$ is a nonempty subset of $\widetilde{R}_W$ and the uncontrolled sandpile model on $\widetilde{\mathcal{G}}_W$ is irreducible on $\widetilde{R}_W$, this implies

$$\widetilde{R} = \widetilde{R}_W.$$

This proves the equality (19). □

If the vertex x is topple-free under π_x, then, we can draw a stronger conclusion about the stationary distribution than in Theorem 3.5. Indeed, in stationarity, no topplings of x contribute additional grains to the contracted graph. Therefore, the projected process follows the uncontrolled dynamics on the contracted graph, with an additional holding

step whenever the addition site is x. Consequently, every projected stationary distribution coincides with the stationary distribution of the uncontrolled dynamics on the contracted graph and is therefore uniform on $\widetilde{R}_x$, as stated in the following theorem.

Theorem 3.8. *Let $\mathcal{G} = (V \cup \{s\}, E)$ and let $x \in V$ be topple-free under the policy π_x. For any stationary distribution μ_{π_x} of the π_x-dynamics, let $\bar{\mu}_{\pi_x}$ denote its projection onto the configuration space of $\mathcal{G} \setminus \{x\}$. Then, $\bar{\mu}_{\pi_x}$ is the uniform distribution on $\widetilde{R}_x$.*

Proof. The fact that x is topple-free under the policy π_x implies that the number k_t in the proof of Theorem 3.5 equals 0 for each t. Hence, under the coupling introduced in this proof, we have

$$\eta_t|_{\mathcal{G}\setminus\{x\}} = \xi_t, \qquad \text{for all } t \geq 0.$$

Therefore,

$$\bar{\mu}_{\pi_x}(\xi) = \lim_{T\to\infty} \mathbb{P}(\eta_T \,|_{\mathcal{G}\setminus\{x\}} = \xi) = \lim_{T\to\infty} \mathbb{P}(\xi_T = \xi) = \widetilde{\mu}_x(\xi) = \frac{1}{|\widetilde{R}_x|}$$

for every $\xi \in \widetilde{R}_x$. □

3.3.2 Gain, bias and optimality

We first express the gain and bias of a local intervention policy π_W with topple-free intervention set W in terms of those of the never-intervene policy on the contracted graph. Theorem 3.10 uses these expressions to establish sufficient conditions for optimality of this policy.

Theorem 3.9. *Consider a graph $\mathcal{G} = (V \cup \{s\}, E)$ and a local intervention policy π_W with intervention set $W \subseteq V$. Suppose that W is topple-free under π_W. Then, for every $\eta \in \Omega$, the gain is given by*

$$g^{\pi_W}(\eta) = \frac{|V| - |W|}{|V|} \widetilde{g}_W^{\pi_\emptyset} + \frac{1}{|V|} \left(|W| + \sum_{x \in V\setminus W} p_W(x) \right) \beta, \tag{20}$$

where $\widetilde{g}_W^{\pi_\emptyset}$ denotes the gain of the never-intervene policy on $\widetilde{\mathcal{G}}_W$. Moreover, for each recurrent class R of the Markov chain induced by π_W, there exists a constant $C_R \in \mathbb{R}$ such that, for $\eta \in R$ and $\xi = \eta|_{V\setminus W}$,

$$h^{\pi_W}(\eta) = \widetilde{h}_W^{\pi_\emptyset}(\xi) + \left[\sum_{x\in W} \eta(x) + \sum_{x \in V\setminus W} p_W(x)\xi(x) \right] \beta + C_R, \tag{21}$$

where $\widetilde{h}_W^{\pi_\emptyset}$ denotes the bias of the never-intervene policy on $\widetilde{\mathcal{G}}_W$.

In the proof of this theorem, we are going to use a coupled random walks representation of the sandpile dynamics that we now describe (this is implicit in [33, Section 4], and reference [19] therein). Let us distinguish the grains in the pile, and whenever a toppling at $x \in V$ occurs, let us choose the destinations of the $\deg_G(x)$ redistributed grains according to a uniformly random permutation among all $(\deg_G(x))!$ possibilities. Whenever a grain moves, it takes a simple random walk step on G. The grains together, if observed only at the times they participate in a toppling, form a coupled system of simple random walks. In particular, the number of random walk steps a grain takes is exactly the number of topplings it was involved in, and the probability that a grain added at $x \in V \setminus W$ enters W equals $p_W(x)$. This coupling is useful to us, because the contributions of individual grains enter additively in the long-run cost, and hence the correlations between the walks do not play a role.

Proof of Theorem 3.9. From the definition of the cost function (3), it follows that g^{π_W} and h^{π_W} are of the form

$$g^{\pi_W}(\eta) = g_{\text{av}}^{\pi_W}(\eta) + g_{\text{int}}^{\pi_W}(\eta)\beta, \tag{22}$$

$$h^{\pi_W}(\eta) = h_{\text{av}}^{\pi_W}(\eta) + h_{\text{int}}^{\pi_W}(\eta)\beta, \tag{23}$$

where $g_{\text{av}}^{\pi_W}$ and $h_{\text{av}}^{\pi_W}$ arise from the avalanche costs, while $g_{\text{int}}^{\pi_W}$ and $h_{\text{int}}^{\pi_W}$ arise from the intervention costs. Let $(X_t^{\pi_W})_{t\geq 1}$ denote the Markov chain induced by the policy π_W, started from a recurrent configuration, and let $(\widetilde{X}_t^W)_{t\geq 1}$ denote a lazy version of the uncontrolled Markov chain on the contracted graph $\widetilde{\mathcal{G}}_W$ with holding probability $|W|/|V|$ at each state. We couple the two processes using the same sequence of additions. Whenever a sand grain is added at a vertex in W, we perform the addition only in X^{π_W}, while leaving $\widetilde{X}^W$ unchanged. Since W is topple-free under π_W and the process is started from a recurrent configuration, no vertex in W topples during a transition under π_W. Hence, under this coupling, the restriction of the controlled process to $V \setminus W$ evolves according to a lazy version of the uncontrolled sandpile dynamics on $\widetilde{\mathcal{G}}_W$, and the avalanche sizes in the two processes coincide at every time step. Since in the lazy uncontrolled process on $\widetilde{\mathcal{G}}_W$, an addition occurs with probability $(|V|-|W|)/|V|$, it follows that

$$g_{\text{av}}^{\pi_W}(\eta) = \frac{|V|-|W|}{|V|}\widetilde{g}_W^{\pi_\emptyset}(\xi), \qquad \eta \in R_{\pi_W}. \tag{24}$$

Likewise, we show that for any recurrent class R of the Markov chain induced by the policy π_W, there exists a constant C_R such that

$$h_{\text{av}}^{\pi_W}(\eta) = \widetilde{h}_W^{\pi_\emptyset}(\xi) + C_R, \qquad \eta \in R, \tag{25}$$

where $\xi = \eta|_{V\setminus W}$. To verify (25), observe that equation (7b) for the never-intervene policy on $\widetilde{\mathcal{G}}_W$ on the set of recurrent configurations reads

$$\left[\mathbb{E}[N(\xi)] - \widetilde{g}_W^{\pi_\emptyset} + \sum_{\xi'\in\widetilde{R}_W} P_{d_\emptyset}(\xi,\xi')\widetilde{h}_W^{\pi_\emptyset}(\xi') - \widetilde{h}_W^{\pi_\emptyset}(\xi)\right] = 0, \qquad \xi \in \widetilde{R}_W.$$

Let R denote a recurrent class of the Markov chain induced by π_W and let $\eta \in R$, with $\xi = \eta|_{V\setminus W}$. By Theorem 3.4, we have $\xi \in \widetilde{R}_W$. Since W is topple-free under π_W, we have

$$\mathbb{E}[N(\eta^{d_W(\eta)})] = \frac{|V|-|W|}{|V|}\mathbb{E}[N(\xi)], \tag{26}$$

$$\sum_{\eta' \,:\, \eta'|_{V\setminus W}=\xi'} P_{d_W}(\eta,\eta') = \frac{|V|-|W|}{|V|}P_{\widetilde{d}}(\xi,\xi') + \frac{|W|}{|V|}\mathbf{1}_{\{\xi'=\xi\}}, \qquad \xi' \in \widetilde{R}_W. \tag{27}$$

Combining relations (26) and (27) with (24), we obtain

$$\begin{aligned}
&\mathbb{E}[N(\eta^{d_W(\eta)})] - g_{\text{av}}^{\pi_W}(\eta) + \sum_{\eta'\in R} P_{d_W}(\eta,\eta')h_{\text{av}}^{\pi_W}(\eta') - h_{\text{av}}^{\pi_W}(\eta) \\
&= \frac{|V|-|W|}{|V|}\left[\mathbb{E}[N(\xi)] - \widetilde{g}_W^{\pi_\emptyset} + \sum_{\xi'\in\widetilde{R}_W} P_{\widetilde{d}}(\xi,\xi')\widetilde{h}_W^{\pi_\emptyset}(\xi') - \widetilde{h}_W^{\pi_\emptyset}(\xi)\right] = 0.
\end{aligned}$$

The left-hand side is exactly the policy evaluation equation $(7b)$. Indeed, since R is closed, the constant C_R in expression (25) cancels. Hence, for every recurrent class R, the function $h_{\text{av}}^{\pi_W}$ defined by (25) satisfies equation (7b).

Now, consider the functions $g_{\text{int}}^{\pi_W}$ and $h_{\text{int}}^{\pi_W}$ that arise from the intervention costs. Since W is topple-free under π_W, grains can only leave W through an intervention. Let $(\eta_t)_{t\geq 1}$ denote the process induced by policy π_W. Let

$$I_T = \sum_{t=1}^{T} \mathbb{1}\{d_W(\eta_t) \neq 0\}$$

denote the number of interventions up to time T and let

$$J_T = \sum_{t=1}^{T}\sum_{x\in W}[\eta_{t+1}(x) - (\eta_t^{d_W(\eta_t)})(x)]$$

denote the number of grains that enter W during the first T transitions. Furthermore, let

$$Z_t = \sum_{x \in W} \eta_t(x), \qquad t \geq 1,$$

denote the number of grains present in W at time t. Since grains can only leave W through an intervention and each intervention removes exactly one grain, we have

$$Z_{T+1} - Z_1 = J_T - I_T,$$

or equivalently,

$$I_T = J_T + Z_1 - Z_{T+1}.$$

Since

$$0 \leq Z_t \leq \sum_{x \in W} (\deg(x) - 1), \qquad \text{for all } t \geq 1,$$

the quantity $Z_1 - Z_{T+1}$ is bounded by the same constant for all $T \geq 1$. This implies

$$\lim_{T \to \infty} \frac{1}{T} \mathbb{E}[I_T] = \lim_{T \to \infty} \frac{1}{T} \mathbb{E}[J_T].$$

Thus, the long-run intervention rate equals the long-run rate at which grains enter W. An addition occurs directly in W with probability $|W|/|V|$. If a new grain is added at a vertex $x \in V \setminus W$, then, by the coupled random walks representation, the probability it enters W is given by $p_W(x)$.

Thus, we obtain

$$g_{\text{int}}^{\pi_W}(\eta) = \frac{1}{|V|} \left(|W| + \sum_{x \notin W} p_W(x) \right). \tag{28}$$

Let R denote a recurrent class of the Markov chain induced by π_W. Letting $\eta \in R$ and $\xi = \eta|_{V \setminus W}$, the expected future intervention cost generated by the sand grains that are present in η is given by

$$\left[\sum_{x \in W} \eta(x) + \sum_{x \notin W} p_W(x) \xi(x) \right] \beta.$$

Newly arriving sand grains contribute, in expectation, exactly $g_{\text{int}}^{\pi_W}$ per time step. This suggests that the intervention component of the bias is given by

$$h_{\text{int}}^{\pi_W}(\eta) = \sum_{x \in W} \eta(x) + \sum_{x \notin W} p_W(x) \eta(x) + C_R, \qquad \eta \in R. \tag{29}$$

We verify this expression using the policy evaluation equation (7b).

Let

$$H_{\text{int}}^{\pi_W}(\eta) = \sum_{x \in W} \eta(x) + \sum_{x \notin W} p_W(x) \eta(x), \qquad \eta \in \Omega. \tag{30}$$

We define the function $q_W : V \cup \{s\} \to \mathbb{R}$ as

$$q_W(x) = \begin{cases} 1, & \text{if } x \in W, \\ 0, & \text{if } x = s, \\ p_W(x), & \text{if } x \in V \setminus W. \end{cases}$$

Then, expression (30) can be written as

$$H_{\text{int}}^{\pi_W}(\eta) = \sum_{x \in V} \eta(x) q_W(x), \qquad \eta \in \Omega.$$

This implies

$$H_{\text{int}}^{\pi_W}(\eta^{d_W(\eta)}) - H_{\text{int}}^{\pi_W}(\eta) = -\mathbb{1}\{d_W(\eta) \neq 0\}. \tag{31}$$

For $\eta \in \Omega$ and $x \in V$, the definition of the stabilization process reads

$$(\alpha_x \eta)^\circ(y) = \eta(y) + \mathbb{1}_x(y) - \sum_{u \in V} n_u(\eta, x)\Delta_{uy}, \qquad y \in V.$$

Consequently,

$$H_{\text{int}}^{\pi_W}((\alpha_x\eta)^\circ) - H_{\text{int}}^{\pi_W}(\eta) = q_W(x) - \sum_{u \in V} n_u(\eta, x)(\Delta q_W)(u), \qquad \eta \in \Omega, \quad x \in V. \tag{32}$$

Now, let $\eta \in R$. Since W is topple-free under π_W, we have

$$n_u(\eta^{d_W(\eta)}, x) = 0, \qquad \text{for } u \in W, \quad x \in V.$$

Moreover, since

$$\begin{aligned} q_W(y) &= 1, \qquad y \in W, \\ q_W(s) &= 0, \qquad \text{and} \\ q_W(y) &= \frac{1}{\deg(y)} \sum_{z \sim y} b_{yz} q_W(z), \qquad y \in V \setminus W, \end{aligned}$$

we have

$$(\Delta q_W)(u) = \deg(u) q_W(u) - \sum_{z \sim u} b_{uz} q_W(z) = 0, \qquad u \in V \setminus W.$$

Therefore, using expression (32), we obtain

$$H_{\text{int}}^{\pi_W}((\alpha_x \eta^{d_W(\eta)})^\circ) - H_{\text{int}}^{\pi_W}(\eta^{d_W(\eta)}) = q_W(x). \tag{33}$$

Combining expressions (31) and (33) now yields

$$H_{\text{int}}^{\pi_W}((\alpha_x \eta^{d_W(\eta)})^\circ) - H_{\text{int}}^{\pi_W}(\eta) = q_W(x) - \mathbb{1}\{d_W(\eta) \neq 0\}.$$

Averaging over the addition vertex, we obtain

$$(P_{d_W} H_{\text{int}}^{\pi_W})(\eta) - H_{\text{int}}^{\pi_W}(\eta) = \frac{1}{|V|} \sum_{x \in V} q_W(x) - \mathbb{1}\{d_W(\eta) \neq 0\}. \tag{34}$$

Since

$$\frac{1}{|V|} \sum_{x \in V} q_W(x) = \frac{1}{|V|} \left(|W| + \sum_{x \notin W} p_W(x) \right) = g_{\text{int}}^{\pi_W},$$

expression (34) reduces to

$$\mathbb{1}\{d_W(\eta) \neq 0\} - g_{\text{int}}^{\pi_W} + (P_{d_W} H_{\text{int}}^{\pi_W})(\eta) - H_{\text{int}}^{\pi_W}(\eta) = 0,$$

which is exactly the policy evaluation equation (7b). This shows that the intervention component of the bias is given by (29).

Substituting (24) and (28) into (22) and (25) and (29) into (23) yields (21) and (20), respectively. □

Expression (20) shows that the gain of the local intervention policy π_W is the sum of the scaled gain of the never-intervene policy on the contracted graph and the long-run average intervention cost, determined by the probability that sand grains reach W before hitting the sink. Similarly, up to an additive constant, the bias is the sum of the bias of the never-intervene policy on the contracted graph and an intervention term depending on the configuration, in

which each grain is weighted by its probability of reaching W before the sink. Using these, we now establish sufficient conditions for the optimality of a local intervention policy.

Theorem 3.10. *Let $\pi_W = (d_W)^\infty$ be any local intervention policy with intervention set $W \subseteq V$. Suppose that W is topple-free under π_W. Let*

$$\underline{\beta}_W = \max_{\substack{a \in V \setminus W \\ \deg(a) \geq 2}} \frac{\phi_W(a)}{1 - p_W(a)}, \qquad \overline{\beta}_W = \min_{u \in W : F_W(u) > 0} \frac{1 + \sum_{v \in V \setminus W} b_{uv} \phi_W(v)}{F_W(u)},$$

where

$$F_W(u) = b_{us} + \sum_{v \in V \setminus W} b_{uv}(1 - p_W(v)).$$

We adopt the conventions that the maximum over the empty set equals 0, the minimum over the empty set equals ∞ and the ratio defining $\underline{\beta}_W$ is interpreted as $+\infty$ whenever $p_W(a) = 1$. If

$$\underline{\beta}_W \leq \overline{\beta}_W,$$

then π_W is optimal for every $\beta \in [\underline{\beta}_W, \overline{\beta}_W]$.

Proof. Define the function $q_W : V \cup \{0\} \to \mathbb{R}$ as

$$q_W(x) = \begin{cases} \beta, & \text{if } x \in W, \\ \phi_W(x) + p_W(x)\beta, & \text{if } x \in V \setminus W, \\ 0, & \text{if } x = 0, \end{cases}$$

where 0 denotes the action corresponding to no intervention, whereas s denotes the sink. Let

$$H_W(\eta) = \sum_{y \in V} \eta(y) q_W(y). \tag{35}$$

By expressions (21) and (12), applied to the contracted graph $\widetilde{\mathcal{G}}_W$, the function H_W coincides, up to an additive constant, with the bias h^{π_W} on each recurrent class of the Markov chain induced by π_W.

Let $\eta \in \Omega$ and $x \in V$. By the definition of the stabilization process, we have

$$(\alpha_x \eta)^\circ(y) = \eta(y) + \mathbb{1}_x(y) - \sum_{u \in V} n_u(\eta, x) \Delta_{uy}, \qquad y \in V. \tag{36}$$

Using expressions (36) and (35), we obtain

$$H_W\left((\alpha_x \eta)^\circ\right) - H_W(\eta) = q_W(x) - \sum_{u \in V} n_u(\eta, x)(\Delta q_W)(u). \tag{37}$$

For any $u \in V \setminus W$, we have

$$(\Delta \phi_W)(u) = 1, \qquad (\Delta p_W)(u) = 0.$$

Hence,

$$(\Delta q_W)(u) = 1, \qquad u \in V \setminus W.$$

Substituting this into (37) yields

$$H_W\left((\alpha_x \eta)^\circ\right) - H_W(\eta) = q_W(x) - \sum_{u \in W} n_u(\eta, x)(\Delta q_W)(u) - \sum_{u \in V \setminus W} n_u(\eta, x). \tag{38}$$

Now, let $\eta \in R_{\pi_W}$. Since W is topple-free under π_W, we have

$$n_u(\eta^{d_W(\eta)}, x) = 0, \qquad u \in W, \quad x \in V.$$

Therefore,

$$H_W\left(\left(\alpha_x \eta^{d_W(\eta)}\right)^\circ\right) - H_W\left(\eta^{d_W(\eta)}\right) = q_W(x) - n\left(\eta^{d_W(\eta)}, x\right). \tag{39}$$

Moreover, since $d_W(\eta) \in W$ whenever $d_W(\eta) \neq 0$ and $q_W(u) = \beta$ for every $u \in W$, expression (35) yields

$$H_W\left(\eta^{d_W(\eta)}\right) - H_W(\eta) = -\beta \mathbb{1}\{d_W(\eta) \neq 0\}. \tag{40}$$

Combining (39) and (40) now yields

$$H_W\left(\left(\alpha_x \eta^{d_W(\eta)}\right)^\circ\right) - H_W(\eta) = q_W(x) - n\left(\eta^{d_W(\eta)}, x\right) - \beta \mathbb{1}\{d_W(\eta) \neq 0\}.$$

Consequently,

$$n\left(\eta^{d_W(\eta)}, x\right) + \beta \mathbb{1}\{d_W(\eta) \neq 0\} = q_W(x) - \left[H_W\left(\left(\alpha_x \eta^{d_W(\eta)}\right)^\circ\right) - H_W(\eta)\right].$$

Using the definition of the cost function (3), we obtain, for every η belonging to a recurrent class of the Markov chain induced by π_W,

$$\begin{aligned} c_{d_W}(\eta) &= \mathbb{E}\left[N\left(\eta^{d_W(\eta)}\right)\right] + \beta \mathbb{1}\{d_W(\eta) \neq 0\} \\ &= \frac{1}{|V|}\sum_{x \in V} q_W(x) - \frac{1}{|V|}\sum_{x \in V}\left[H_W\left(\left(\alpha_x \eta^{d_W(\eta)}\right)^\circ\right) - H_W(\eta)\right] \\ &= \frac{1}{|V|}\sum_{x \in V} q_W(x) + H_W(\eta) - \left(P_{d_W} H_W\right)(\eta). \end{aligned}$$

Moreover,

$$\begin{aligned} \frac{1}{|V|}\sum_{x \in V} q_W(x) &= \frac{1}{|V|}\sum_{x \in V \setminus W} \phi_W(x) + \frac{1}{|V|}\left(|W| + \sum_{x \in V \setminus W} p_W(x)\right)\beta \\ &= \frac{|V| - |W|}{|V|}\widetilde{g}_W^{\pi_\emptyset} + \frac{1}{|V|}\left(|W| + \sum_{x \in V \setminus W} p_W(x)\right)\beta \\ &= g^{\pi_W}, \end{aligned}$$

where the second equality follows from Theorem 3.1, applied to $\widetilde{\mathcal{G}}_W$, and the final equality follows from (20). Hence,

$$c_{d_W}(\eta) = g^{\pi_W} + H_W(\eta) - (P_{d_W} H_W)(\eta) \tag{41}$$

on every recurrent class of the Markov chain induced by π_W.

Now, let $\pi = d^\infty$ denote an arbitrary stationary deterministic policy. For $\eta \in \Omega$, $a \in A(\eta)$, and $x \in V$, expression (35) implies

$$H_W(\eta^a) - H_W(\eta) = -q_W(a).$$

Combining this with expression (38), we obtain

$$\begin{aligned} H_W\left((\alpha_x \eta^a)^\circ\right) - H_W(\eta) &= q_W(x) - \sum_{u \in W} n_u(\eta^a, x)(\Delta q_W)(u) - \sum_{u \in V \setminus W} n_u(\eta^a, x) - q_W(a) \\ &= q_W(x) - n(\eta^a, x) + \sum_{u \in W} n_u(\eta^a, x)[1 - (\Delta q_W)(u)] - q_W(a). \end{aligned}$$

Rearranging gives

$$\begin{aligned} n(\eta^a, x) + \beta \mathbb{1}\{a \neq 0\} =& q_W(x) - [H_W((\alpha_x \eta^a)^\circ) - H_W(\eta)] \\ &+ \beta \mathbb{1}\{a \neq 0\} - q_W(a) + \sum_{u \in W} n_u(\eta^a, x)[1 - (\Delta q_W)(u)]. \end{aligned} \tag{42}$$

We now show that the final two terms in (42) are nonnegative when $\beta \in [\underline{\beta}_W, \overline{\beta}_W]$.

First, for $a = d(\eta)$, we have

$$\beta \mathbb{1}\{a \neq 0\} - q_W(a) = \begin{cases} 0, & \text{if } a = 0 \text{ or } a \in W, \\ \beta(1 - p_W(a)) - \phi_W(a), & \text{if } a \in V \setminus W. \end{cases}$$

This implies that, whenever

$$\beta \geq \underline{\beta}_W = \max_{a \in V \setminus W} \frac{\phi_W(a)}{1 - p_W(a)},$$

we have

$$\beta \mathbb{1}\{a \neq 0\} - q_W(a) \geq 0.$$

Next, let $u \in W$. Since $q_W(u) = \beta$, we have

$$\begin{aligned} (\Delta q_W)(u) &= \deg(u)\beta - \sum_{v \in W} b_{uv}\beta - \sum_{v \in V \setminus W} b_{uv}(\phi_W(v) + p_W(v)\beta) \\ &= \beta \left[b_{us} + \sum_{v \in V \setminus W} b_{uv}(1 - p_W(v)) \right] - \sum_{v \in V \setminus W} b_{uv}\phi_W(v) \end{aligned}$$

Consequently,

$$1 - (\Delta q_W)(u) = 1 + \sum_{v \in V \setminus W} b_{uv}\phi_W(v) - \beta \left[b_{us} + \sum_{v \in V \setminus W} b_{uv}(1 - p_W(v)) \right]$$

It follows that

$$1 - (\Delta q_W)(u) \geq 0, \qquad u \in W,$$

whenever

$$\beta \leq \overline{\beta}_W = \min_{u \in W : F_W(u) > 0} \frac{1 + \sum_{v \in V \setminus W} b_{uv}\phi_W(v)}{F_W(u)}.$$

Thus, if $\beta \in [\underline{\beta}_W, \overline{\beta}_W]$, then expression (42) yields

$$n(\eta^a, x) + \beta \mathbb{1}\{a \neq 0\} \geq q_W(x) - \left[H_W\left((\alpha_x \eta^a)^\circ\right) - H_W(\eta) \right].$$

Averaging over $x \in V$, we now obtain

$$c_d(\eta) \geq \frac{1}{|V|} \sum_{x \in V} q_W(x) + H_W(\eta) - (P_d H_W)(\eta) = g^{\pi_W} + H_W(\eta) - (P_d H_W)(\eta), \tag{43}$$

for every $\eta \in \Omega$ and $a \in A(\eta)$.

To finalize the proof, let $(\eta_t)_{t \geq 1}$ denote the Markov chain induced by the arbitrary stationary deterministic policy $\pi = d^\infty$. Summing inequality (43) over $t = 1, \ldots, T$ and taking expectations yields

$$\mathbb{E}_\eta^\pi \left[\sum_{t=1}^{T} c_d(\eta_t) \right] \geq T g^{\pi_W} + H_W(\eta) - \mathbb{E}_\eta^\pi [H_W(\eta_{T+1})].$$

Dividing by T and letting $T \to \infty$, using the fact that H_W is bounded, now gives

$$g^{\pi}(\eta) \geq g^{\pi_W} \qquad \eta \in \Omega.$$

Thus, we conclude that the policy π_W is optimal. □

4 Applications to specific graph classes

In this section, we apply the general results obtained in the preceding sections to four different graph classes: the one-dimensional lattice, the two-dimensional lattice, the complete graph and the star graph. Moreover, we discuss for each of these classes whether and under what conditions topple-freeness holds. For the one-dimensional lattice, the complete graph and the star graph, we characterize optimal policies for each value of the intervention cost.

4.1 The one-dimensional lattice

In this subsection, we apply the general results on local intervention policies with topple-free intervention sets to characterize the optimal policies on the one-dimensional lattice for all values of β.

Consider the one-dimensional lattice $\mathcal{G} = (V \cup \{s\}, E)$ with vertex set $V = \{1, \ldots, L\}$, where the boundary vertices 1 and L are connected to the sink by a single edge. The Green's function is given by [34, Theorem 3]

$$G(x,y) = \begin{cases} x(L+1-y)/(L+1), & \text{if } x \leq y, \\ y(L+1-x)/(L+1), & \text{if } y < x. \end{cases} \tag{44}$$

Using the definition (9), the function $\phi : V \to \mathbb{R}$ can be expressed as

$$\phi(x) = \sum_{y=1}^{L} G(x,y) = \frac{x(L+1-x)}{2} = \frac{L+1}{2} G(x,x), \quad x \in V. \tag{45}$$

Since $\phi(x)$ is a concave quadratic function of x, it is maximized at the central vertex when L is odd and at the two central vertices when L is even. Corollary 4.1 provides explicit expressions for the gain, bias and optimality threshold of the never-intervene policy on the one-dimensional lattice.

Corollary 4.1. *The gain of the never-intervene policy $\pi_\emptyset = (d_\emptyset)^\infty$ is given by*

$$g^{\pi_\emptyset}(\eta) = \frac{1}{12}(L+1)(L+2), \qquad \eta \in \Omega. \tag{46}$$

Furthermore, there exists a constant $C \in \mathbb{R}$ such that

$$h^{\pi_\emptyset}(\eta) = \sum_{x=1}^{L} (\eta(x) - 1)\phi(x) + C, \qquad \eta \in \Omega. \tag{47}$$

The policy $\pi_\emptyset$ is optimal if and only if

$$\beta \geq \begin{cases} (L+1)^2/8, & \textit{if } L \textit{ is odd}, \\ L(L+2)/8, & \textit{if } L \textit{ is even}. \end{cases}$$

Proof. The expressions for the gain and the bias follow immediately from Theorem 3.1 and expression (45). The optimality result follows immediately from Theorem 3.3 and expression (45). □

We now consider local intervention policies with nested intervention sets that are central intervals of the lattice. For each $k \in \{0, \ldots, \lfloor L/2 \rfloor\}$, let

$$W_k = \{k+1, \ldots, L-k\}$$

and let $\pi_k = (d_k)^\infty$ be any local intervention policy with intervention set W_k. When L is even, we have $W_{L/2} = \emptyset$, so $\pi_{L/2}$ coincides with the never-intervene policy.

To describe the recurrent classes of the Markov chains induced by these policies, we introduce the configurations $\bar{\eta}_j \in \Omega$, for $j = 0, \ldots, \lfloor L/2 \rfloor$, defined as

$$\bar{\eta}_j(i) = \begin{cases} 1, & \text{if } i \in \{1, \ldots, j\} \cup \{L - j + 1, \ldots, L\}, \\ 0, & \text{otherwise.} \end{cases}$$

The following theorem characterizes the unique recurrent class of the Markov chain induced by each policy π_k for $k \in \{1, \ldots, \lfloor L/2 \rfloor\}$. The case $k = 0$ is treated separately in Proposition 4.3.

Theorem 4.2. *For $k \in \{1, \ldots, \lfloor L/2 \rfloor\}$, let $\pi_k = (d_k)^\infty$ be any local intervention policy with intervention set W_k. Then the Markov chain induced by π_k has a unique recurrent class R_k, where*

$$R_k = \left\{ \eta \in \{0,1\}^L : \sum_{i=1}^{k} \eta(i) \geq k - 1, \sum_{i=L-k+1}^{L} \eta(i) \geq k - 1, \sum_{i=k+1}^{L-k} \eta(i) \leq 1 \right\} \setminus \{\bar{\eta}_{k-1}\}, \tag{48}$$

for $k = 1, \ldots, \lfloor (L-1)/2 \rfloor$ and, when L is even,

$$R_{L/2} = \{\eta \in \{0,1\}^L : \sum_{i=1}^{L} \eta(i) \geq L - 1\}. \tag{49}$$

Proof. If L is even and $k = L/2$, then $W_k = \emptyset$, so π_k is the never-intervene policy. The uncontrolled sandpile dynamics on a one-dimensional lattice has a unique recurrent class consisting of stable configurations with at most one empty vertex [35, Section 2]. This establishes expression (49).

Now, let $k \in \{1, \ldots, \lfloor (L-1)/2 \rfloor\}$. We first show that every recurrent configuration has at most one grain in the central strip $\{k + 1, \ldots, L - k\}$. Whenever this strip contains at least one grain, the policy removes one of these grains before the addition of the new sand grain. Suppose that a configuration has at most one grain in the central strip. Then the strip is empty immediately after the intervention prescribed by the policy. The subsequent addition of a new grain followed by stabilization can create at most one grain in the central strip. It follows that the set

$$\{\eta \in \Omega : \sum_{i=k+1}^{L-k} \eta(i) \leq 1\} \tag{50}$$

is closed under the dynamics induced by π_k.

Now suppose that a configuration contains more than one grain in the central strip. At every transition under policy π_k, one such grain is removed. The subsequent addition and stabilization can create at most one new grain in the strip. Hence, the number of grains in the central strip cannot increase. Moreover, if the configuration outside the central strip is not completely filled, there is a positive probability that no new grain ends up in the central strip during the subsequent addition and stabilization, so that the number of central grains decreases by one. If the configuration outside the strip is completely filled, a new grain necessarily appears in the central strip during the next transition. However, there is a positive probability that a vertex outside the central strip becomes empty, after which there is again a positive probability that the number of grains in the central strip decreases at the following transition. Thus, from every configuration containing more than one grain in the central strip, there is a finite path of positive probability to a configuration with one fewer grain in the central strip. Iterating this argument shows that the set (50) can be reached with positive probability in finitely many steps. Since this set is closed, every recurrent class of the Markov chain induced by π_k must be contained in (50).

Since the recurrent class is contained in (50), the central strip is empty immediately after an intervention. Moreover, the policy never removes grains from the two outside strips. Therefore, during the subsequent addition and stabilization, the dynamics restricted to each outside strip coincide with those of an uncontrolled sandpile model on a lattice of length k. Hence, the restriction of any recurrent configuration to either outside strip must belong to the

recurrent class of the uncontrolled sandpile model on a lattice of length k, which is given by [35]

$$\{\xi \in \{0,1\}^k : \sum_{i=1}^{k} \xi(i) \geq k-1\}.$$

It follows that every recurrent configuration has at most one empty vertex in each outside strip. Hence, every recurrent class is contained in

$$\left\{\eta \in \{0,1\}^L : \sum_{i=1}^{k} \eta(i) \geq k-1, \sum_{i=L-k+1}^{L} \eta(i) \geq k-1, \sum_{i=k+1}^{L-k} \eta(i) \leq 1\right\}. \tag{51}$$

We now show that from each configuration in (51), there exists a finite path of positive probability to the configuration $\bar{\eta}_k$. Let η denote a configuration in the set (51). Suppose first that at least one of the outside strips contains an empty vertex. We then successively add grains at the empty vertices in the outside strips. Before each addition, the policy removes a grain from the central strip whenever one is present. Hence, after at most two transitions, both outside strips are completely filled and the central strip is empty, so that the system has reached $\bar{\eta}_k$.

Now, suppose that both outside strips are already completely filled. If the central strip is empty, the configuration already equals $\bar{\eta}_k$. Otherwise, the central strip contains exactly one grain. The policy first removes this grain. Then, we add a grain at an arbitrary vertex in one of the outside strips. The resulting avalanche creates one empty vertex in that outside strip and one grain in the central strip. At the next transition, the policy removes the grain from the central strip, after which we add a grain at the newly created empty vertex. This yields $\bar{\eta}_k$.

Since each of these sequences of additions occurs with positive probability, it follows that $\bar{\eta}_k$ can be reached from every configuration in (51). Therefore, every recurrent class contains $\bar{\eta}_k$, which implies that the Markov chain induced by π_k has a unique recurrent class.

We proceed to show that the set of configurations accessible from $\bar{\eta}_k$ is precisely R_k, as defined in (48). To this end, we show that every configuration in (51) can be reached from $\bar{\eta}_k$, except for the configuration $\bar{\eta}_{k-1}$. Let η be a configuration in (48). If $\eta = \bar{\eta}_k$, the statement is obvious. Suppose that $\eta \neq \bar{\eta}_k$. We distinguish seven cases.

Case I: Suppose that the outside strips in η are completely filled. Since $\eta \neq \bar{\eta}_k$, there must be a grain in the central strip. In this case, η can be reached from $\bar{\eta}_k$ by simply adding a grain at the corresponding vertex in the central strip.

Case II: Suppose that η has one empty vertex in either the left or the right outside strip, but not at a vertex adjacent to the central strip, and that there is no grain in the central strip. Without loss of generality, suppose that the empty vertex is $i \in \{1, \ldots, k-1\}$, and thus lies in the left outside strip. Starting from $\bar{\eta}_k$ we first add a grain at vertex 1. After stabilization, this creates an empty vertex at k and a grain at $k+1$. The policy then removes the grain at $k+1$, after which we add a grain at vertex $k-i$. This moves the empty vertex to i, yielding the configuration η.

Case III: Suppose that η has one empty vertex at either the left or right neighbor of the central strip, and that there is no grain in the central strip. Without loss of generality, suppose that the empty vertex is k. Starting from $\bar{\eta}_k$, we again first add a grain at vertex 1, creating an empty vertex at k and a grain at $k+1$. After the policy removes the grain at $k+1$, we add a grain at an arbitrary vertex in the right outside strip. After stabilization, this creates an empty vertex in the right outside strip and a grain at $L-k$. The policy then removes the grain at $L-k$, after which we add a grain at the newly created empty vertex in the right outside strip. This results in the configuration η.

Case IV: Suppose that η has one empty vertex in either the left or the right outside strip and one grain in the central strip. Without loss of generality, suppose that the empty vertex is $i \in \{1, \ldots, k\}$, and thus lies in the left outside strip. Starting from configuration $\bar{\eta}_k$, we first add a grain at vertex $k+1-i$. After stabilization, this creates an empty vertex at i and a grain at vertex $k+1$. If the grain in the central strip of η is located at $k+1$, we have now reached η. Otherwise, the policy first removes the grain at $k+1$, after which we add the grain at the vertex in the central strip occupied in η, yielding configuration η.

Case V: Suppose that η has an empty vertex in both the left and the right outside strip, that at least one of these vertices is not adjacent to the central strip, and that there are no grains in the central strip. Without loss of generality, suppose that the empty vertices are $i \in \{1, \dots, k\}$ and $j \in \{L-k+2, \dots, L\}$, so that the empty vertex in the right outside strip is not adjacent to the central strip. Starting from configuration $\bar{\eta}_k$, we first add a grain at vertex $k+1-i$. After stabilization, this creates an empty vertex at i and a grain at $k+1$. The policy then removes the grain at $k+1$. Next, we add a grain at vertex L, which creates an empty vertex at $L-k+1$ and a grain at $L-k$. After the policy removes the grain at $L-k$, we add a grain at vertex $2L+2-j-k$. After stabilization, the empty vertex has moved from $L-k+1$ to j, yielding the configuration η.

Case VI: Suppose that η has an empty vertex in both the left and the right outside strip and that there is one grain in the central strip. Let the empty vertices be $i \in \{1, \dots, k\}$ and $j \in \{L-k+1, \dots, L\}$. Starting from configuration $\bar{\eta}_k$, we first add a grain at vertex $k+1-i$. After stabilization, this creates an empty vertex at i and a grain at vertex $k+1$. The policy then removes the grain at $k+1$, after which we add a grain at vertex $2L-k-j+1$. After stabilization, this creates an empty vertex at j and a grain at $L-k$. If $L-k$ is the occupied vertex in the central strip of η, we have reached η. Otherwise, the policy removes the grain at $L-k$, after which we add a grain at the vertex in the central strip that is occupied in η, yielding the configuration η.

Case VII: Finally, consider the configuration $\bar{\eta}_{k-1}$. We show that this configuration cannot be reached in a single step from any configuration in (51) under the dynamics induced by π_k. Since the recurrent class R_k is contained in (51), it then follows that $\bar{\eta}_{k-1}$ cannot be reached in a finite number of steps from R_k. Let η denote a configuration in (51) from which a transition to $\bar{\eta}_{k-1}$ is possible, and let $i \in V$ denote the corresponding addition location.

Suppose first that $i \in \{k+1, \dots, L-k\}$. Since the central strip is empty immediately after the intervention prescribed in η, addition at i and subsequent stabilization leaves a grain at i. Therefore, the new configuration cannot be $\bar{\eta}_{k-1}$.

Now, suppose that $i \in \{1, \dots, k\}$. For vertex k to be empty after the addition at i and subsequent stabilization, there are only two possibilities: either $i = 1$ and the left outside strip is completely filled, or $i \in \{2, \dots, k\}$ and the left outside strip has one empty vertex at $i-1$. In either case, addition at i and stabilization also creates a grain at vertex $k+1$. Hence, the resulting configuration cannot be $\bar{\eta}_{k-1}$. The same argument applies when $i \in \{L-k+1, \dots, L\}$. Thus, the configuration $\bar{\eta}_{k-1}$ is not reachable from any configuration in (51).

We now conclude that R_k is the unique recurrent class of policy π_k. We have shown that $\bar{\eta}_k$ belongs to every recurrent class, which implies that the recurrent class is unique. Furthermore, we have shown that this recurrent class is contained in (51). Cases I-VI show that each configuration in (51) except $\bar{\eta}_{k-1}$ is reachable from $\bar{\eta}_k$, whereas Case VII shows that $\bar{\eta}_{k-1}$ is not reachable. Thus, the unique recurrent class of policy π_k is precisely R_k. □

The above theorem establishes that for $k \in \{1, \dots, \lfloor L/2 \rfloor\}$, any local intervention policy $\pi_k = (d_k)^\infty$ with intervention set W_k induces a Markov chain with a unique recurrent class. The following example shows that for $k = 0$, this need not be the case. In Proposition 4.3, however, we show that it is possible to construct an always-intervene policy with a unique recurrent class and that always-intervene policies with multiple recurrent classes cannot be optimal.

Example Let $L \geq 4$ and consider the always-intervene policy $\pi_0^{\text{bnd}} = (d_0^{\text{bnd}})^\infty$, where

$$d_0^{\text{bnd}}(\eta) = \begin{cases} 1, & \text{if } \eta(1) = 1, \\ L, & \text{if } \eta(1) = 0 \text{ and } \eta(L) = 1, \\ \min\{i \in \{1, \dots, L\} | \eta(i) = 1\}, & \text{if } \eta(1) = \eta(L) = 0 \text{ and } \sum_{i=2}^{L-1} \eta(i) > 0, \\ 0, & \text{otherwise,} \end{cases} \qquad \eta \in \Omega.$$

We show that there exists a recurrent class that is contained in the set

$$C = \{\eta \in \{0,1\}^L : \sum_{i=1}^{L} \eta(i) = 2, \sum_{i=2}^{L-1} \eta(i) \geq 1\}.$$

It suffices to show that C is closed under the π_0-dynamics. Observe that any configuration in the set C contains at least one grain at an interior vertex. Since each configuration contains exactly two grains and the policy favors removal from the boundary vertices, there is exactly one grain at an interior vertex j after an intervention from any starting configuration in C. Suppose that the new grain is added at a vertex $k \neq j$. In that case, the new configuration is contained in C. If $k = j$, then the vertex j topples and the new configuration contains grains at $j-1$ and $j+1$. Since $L \geq 4$, at least one of these vertices is an interior vertex. Hence, the resulting configuration belongs to C. This shows that C is closed and therefore contains at least one recurrent class of the Markov chain induced by π_0. Since C is disjoint from the recurrent class R_0, it follows that the Markov chain induced by π_0 has at least two recurrent classes.

Proposition 4.3. *For the case $k = 0$, the following statements hold:*

1. *Under every always-intervene policy $\pi_0 = (d_0)^\infty$, the set*

$$R_0 = \{\eta \in \{0,1\}^L : \sum_{i=1}^{L} \eta(i) = 1\} \tag{52}$$

   is a recurrent class of the induced Markov chain.
2. *There exists an always-intervene policy $\pi_0 = (d_0)^\infty$ under which R_0 is the unique recurrent class of the induced Markov chain.*
3. *Let π_0 and π_0' be always-intervene policies whose induced Markov chains have a unique recurrent class and multiple recurrent classes, respectively. Then,*

$$g^{\pi_0}(\eta) = \beta \leq g^{\pi_0'}(\eta), \qquad \eta \in \Omega.$$

   The inequality above is strict for any η that belongs to a recurrent class $R \neq R_0$ of π_0'. Thus, π_0' is not optimal for any $\beta \geq 0$.

Proof.

1. It suffices to show that the set R_0 is closed and irreducible under the π_0-dynamics. For every $\eta \in R_0$, the policy removes the unique grain present in η. The subsequent addition results in another configuration in R_0. This implies that R_0 is closed. Moreover, every configuration in R_0 is reached during a single transition with probability $1/L$, regardless of the starting configuration in R_0. Hence, the set R_0 is irreducible. Thus, R_0 is a recurrent class of the Markov chain induced by any always-intervene policy.
2. Consider the always-intervene policy $\pi_0^{\text{left}} = (d_0^{\text{left}})^\infty$, where

$$d_0^{\text{left}}(\eta) = \begin{cases} 0, & \text{if } \eta(i) = 0 \text{ for all } i \in \{1, \ldots, L\}, \\ \min\{i \in \{1, \ldots, L\} : \eta(i) = 1\}, & \text{otherwise.} \end{cases}$$

   We show that under this policy, R_0 can be reached in a finite number of transitions from any starting configuration. It then follows from the first statement that R_0 is the unique recurrent class.
   Let η denote a configuration containing at least two grains and let j denote the position of its rightmost grain. We specify a sequence of transitions that occurs with positive probability and reaches the set R_0. Since the policy π_0^{left} always removes the leftmost grain, the grain at j remains after the intervention. Suppose that $j < L$. Then, we add the next grain at vertex $j+1$. By assumption, this vertex is empty, so the resulting configuration is stable and the rightmost occupied vertex is now $j+1$. Repeating this procedure, we reach a configuration whose rightmost grain

is at vertex L. During these transitions, the total number of grains remains constant. Once the rightmost grain is at vertex L, we add a grain at L after the intervention. After the resulting avalanche, exactly one grain is lost to the sink. Hence, the total number of grains decreases by one. This procedure can be repeated until a configuration is reached with exactly one grain, which happens in a finite number of transitions. Since each transition has positive probability of occurring, the set R_0 is accessible from every configuration. Thus, invoking the first statement, R_0 is the unique recurrent class of the Markov chain induced by π_0.

3. Within the recurrent class R_0, one sand grain is removed during each transition and no topplings occur. Therefore,

$$g^{\pi_0}(\eta) = g^{\pi_0'}(\eta) = \beta \qquad \text{for all } \eta \in R_0.$$

Let $R \neq R_0$ denote a recurrent class of the Markov chain induced by π_0'. Every configuration in R contains at least two grains. Therefore, during each transition, one grain is removed and the probability of a toppling is at least $1/L$. Therefore,

$$g^{\pi_0'}(\eta) \geq \beta + \frac{1}{L} > \beta \qquad \eta \in R.$$

By expression (5), the gain at any configuration $\eta \in \Omega$ is a convex combination of the stationary average costs on each of the recurrent classes. Since each recurrent class of the Markov chain induced by π_0' has stationary average cost at least β and R_0 is the unique recurrent class of the Markov chain induced by π_0, we obtain

$$g^{\pi_0'}(\eta) \geq \beta = g^{\pi_0}(\eta), \qquad \eta \in \Omega,$$

where the inequality is strict on every recurrent class $R \neq R_0$. Thus, the policy π_0' is not optimal for any $\beta \geq 0$.

□

Given a configuration $\eta \in \Omega$ and $k = 1, \ldots, \lfloor L/2 \rfloor$, we define the configurations

$$\eta_\ell^{(k)}(i) = \eta(k+1-i), \qquad \eta_r^{(k)}(i) = \eta(L-k+i), \qquad i = 1, \ldots, k.$$

These are the restrictions of η to the leftmost and rightmost k vertices, respectively, indexed from the interior towards the boundary. For $k = 0$, we regard $\eta_\ell^{(0)}$ and $\eta_r^{(0)}$ as configurations on the empty lattice. The following theorem gives the gain and the bias of the policies π_k, $k = 0, \ldots, (L-1)/2$.

Theorem 4.4. *For each $k \in \{1, \ldots, \lfloor (L-1)/2 \rfloor\}$, let $\pi_k = (d_k)^\infty$ be any local intervention policy with intervention set W_k. Let $\pi_0 = (d_0)^\infty$ denote an always-intervene policy whose induced Markov chain has the unique recurrent class R_0, defined in* (52)*. Then, for every $k \in \{0, \ldots, \lfloor (L-1)/2 \rfloor\}$, the gain is given by*

$$g^{\pi_k}(\eta) = \frac{k(k+1)(k+2)}{6L} + \frac{L-k}{L}\beta, \qquad \eta \in \Omega. \tag{53}$$

Moreover, for every $k \in \{0, \ldots, \lfloor (L-1)/2 \rfloor\}$, there exists a constant $C_k \in \mathbb{R}$ such that for every $\eta \in R_k$, the bias is given by

$$h^{\pi_k}(\eta) = \beta \sum_{x=k+1}^{L-k} \eta(x) + \sum_{i=1}^{k} \left[\frac{i(k+1-i)}{2} + \beta \left(1 - \frac{i}{k+1} \right) \right] (\eta_\ell^{(k)}(i) + \eta_r^{(k)}(i)) + C_k. \tag{54}$$

For odd L and $k = (L-1)/2$, expression (54) holds for every $\eta \in \Omega$.

Proof. We first show that for each $k \in \{0, \ldots, \lfloor (L-1)/2 \rfloor\}$, the set W_k is topple-free under the corresponding policy π_k. The result then follows by applying Theorem 3.9. For $k = 0$, the Markov chain induced by π_0 has a unique recurrent class given by R_0. Since each configuration in R_0 contains exactly one grain, which is removed by the policy before the next addition, no topplings occur during transitions from R_0. Therefore, the set $W_0 = V$ is topple-free under π_0. Now, consider $k \in \{1, \ldots, \lfloor (L-1)/2 \rfloor\}$. By the definition (48) of the recurrent class R_k of the Markov chain induced by π_k, each $\eta \in R_k$ has at most one sand grain in the set W_k. Thus, immediately after the intervention

by the policy, the set W_k is empty. Therefore, it is impossible for any vertex $x \in W_k$ to topple during the stabilization following the next addition. Hence, the set W_k is topple-free under the policy π_k for each $k \in \{0, \dots, \lfloor (L-1)/2 \rfloor\}$.

We now invoke Theorem 3.9. Let $\widetilde{g}^{\pi_\emptyset}_{W_k}$ and $\widetilde{h}^{\pi_\emptyset}_{W_k}$ denote the gain and bias of the never-intervene policy on the contracted graph $\widetilde{\mathcal{G}}_{W_k}$. For $k = 0$, the contracted graph $\widetilde{\mathcal{G}}_{W_k}$ has no non-sink vertices. We adopt the conventions $\widetilde{g}^{\pi_\emptyset}_V = 0$ and $\widetilde{h}^{\pi_\emptyset}_V = 0$. For any $k \in \{0, \dots, \lfloor (L-1)/2 \rfloor\}$, we have [36, Section 12.2]

$$p_{W_k}(k+1-i) = p_{W_k}(L-k+i) = 1 - \frac{i}{k+1}, \qquad i = 1, \dots, k. \tag{55}$$

Substituting this and $|V| = L$ and $|W| = L - 2k$ into expression (20), we obtain

$$g^{\pi_k}(\eta) = \frac{2k}{L}\widetilde{g}^{\pi_\emptyset}_{W_k} + \frac{L-k}{L}\beta.$$

For $k = 0$, this gives

$$g^{\pi_0}_L(\eta) = \beta, \qquad \eta \in \Omega,$$

in agreement with expression (53). For $k \geq 1$, this graph consists of two one-dimensional lattices of length k, each connected to the contracted sink. Since a new grain is added to either component with probability $1/2$, expression (46) yields (53).

Furthermore, the general bias formula (12) shows that the bias on the contracted graph is the sum of the biases on its two components, up to an additive constant. Substituting expressions (45), (47) and (55) into (21) yields expression (54).

Finally, suppose that L is odd and $k = (L-1)/2$, so that $W_k = \{z\}$. For every $\eta \in \Omega$, the vertex z is empty immediately after applying the intervention prescribed by π_k. If the next grain is added at z, no toppling occurs. Otherwise, only vertices in the connected component of occupied vertices containing the addition site in the post-decision configuration can topple, so z cannot topple. This implies that z never topples under policy π_k, even during transitions from transient configurations. Therefore, expression (54) satisfies the policy evaluation equation on the entire configuration space. Since the chain has a unique recurrent class, expression (54) holds for every $\eta \in \Omega$. □

Finally, Theorem 4.5 identifies an optimal local intervention policy for each value of β.

Theorem 4.5. *Let* $K = \left\lfloor \dfrac{L-1}{2} \right\rfloor$. *Define*

$$\beta_k = \frac{k(k+1)}{2}, \qquad k = 0, \dots, K,$$

and

$$\beta_{K+1} = \begin{cases} (L+1)^2/8, & \text{if } L \text{ is odd}, \\ L(L+2)/8, & \text{if } L \text{ is even}. \end{cases}$$

For each $k = 1, \dots, K$, *let* $\pi_k = (d_k)^\infty$ *be any local intervention policy with intervention set* W_k. *Let* π_0 *be an always-intervene policy whose induced Markov chain has unique recurrent class* R_0. *Then, for each* $k = 0, \dots, K$,

$$\pi_k \text{ is optimal} \quad \Longleftrightarrow \quad \beta \in [\beta_k, \beta_{k+1}].$$

Moreover,

$$\pi_\emptyset \text{ is optimal} \quad \Longleftrightarrow \quad \beta \geq \beta_{K+1}.$$

Proof. We first prove sufficiency. In the proof of Theorem 4.4, we established that W_k is topple-free under π_k for every $k = 0, \dots, K$. We can therefore invoke Theorem 3.10. For $k \geq 1$, the contracted graph $\widetilde{\mathcal{G}}_{W_k}$ consists of two paths of length k that are both connected to the contracted sink. For $i = 1, \dots, k$, we have, by (45),

$$\phi_{W_k}(k+1-i) = \phi_{W_k}(L-k+i) = \frac{i(k+1-i)}{2},$$

Combining this expression with (55), we obtain

$$\underline{\beta}_{W_k} = \max_{i\in\{1,\dots,k\}} \frac{(k+1-i)(k+1)}{2} = \frac{k(k+1)}{2} = \beta_k.$$

For $k = 0$, the maximum is taken over the empty set, so $\underline{\beta}_{W_0} = \beta_0 = 0$.

We proceed to evaluate the upper bound. First, suppose that $|W_k| \geq 2$. If $u \in W_k$ is not an endpoint of W_k, we have

$$F_{W_k}(u) = 0.$$

If u is an endpoint of W_k, then, using (55) with $i = 1$, we obtain

$$F_{W_k}(u) = \frac{1}{k+1}$$

and

$$1 + \sum_{v\in V\setminus W_k} b_{uv}\phi_{W_k}(v) = \frac{k+2}{2}.$$

Therefore,

$$\overline{\beta}_{W_k} = \frac{(k+1)(k+2)}{2} = \beta_{k+1}.$$

For even L, this includes the case $k = K = L/2 - 1$.

Suppose now that W_k is a singleton. This implies that L is odd and $k = K = (L-1)/2$. Let the central vertex be denoted by z. We have

$$F_{W_k}(z) = \frac{2}{K+1} \qquad \text{and } 1 + \sum_{v\in V\setminus W_K} b_{zv}\phi_{W_K}(v) = K+1.$$

Thus,

$$\overline{\beta}_{W_k} = \frac{(L+1)^2}{8} = \beta_{K+1}.$$

Theorem 3.10 now implies that π_k is optimal whenever $\beta \in [\beta_k, \beta_{k+1}]$.

For the never-intervene policy, we take $W = \emptyset$ and again invoke Theorem 3.10. Since $p_\emptyset = 0$ and $\phi_\emptyset = \phi$, we obtain

$$\underline{\beta}_\emptyset = \max_{x\in V} \phi(x) = \max_{x\in\{1,\dots,L\}} \frac{x(L+1-x)}{2} = \beta_{K+1}$$

and

$$\overline{\beta}_\emptyset = \infty.$$

Thus, $\pi_\emptyset$ is optimal whenever $\beta \geq \beta_{K+1}$.

We next prove necessity. Suppose that $\beta < \beta_k$, $k = 1, \dots, K$. By expression (53), we have

$$g^{\pi_k}(\eta) - g^{\pi_{k-1}}(\eta) = \frac{\beta_k - \beta}{L} > 0, \qquad k = 1, \dots, K.$$

Hence, if $\beta < \beta_k$, the policy π_{k-1} has strictly smaller gain than π_k for all $k = 1, \dots, K$.

Now, suppose that $\beta > \beta_{k+1}$, $k = 0, \dots, K-1$. By expression (53), we have

$$g^{\pi_k}(\eta) - g^{\pi_{k+1}}(\eta) = \frac{\beta - \beta_{k+1}}{L}.$$

Hence, if $\beta > \beta_{k+1}$, the policy π_{k+1} has strictly smaller gain than π_k.

Finally, using (46), we obtain both for odd and even L,

$$g^{\pi_K}(\eta) - g^{\pi_\emptyset}(\eta) = \frac{(L-K)(\beta - \beta_{K+1})}{L}.$$

It follows that π_K is not optimal when $\beta > \beta_{K+1}$ and $\pi_\emptyset$ is not optimal when $\beta < \beta_{K+1}$. This establishes necessity for each policy π_k, $k = 0, \dots, K$. □

4.2 The two-dimensional lattice

The two-dimensional lattice Consider the two-dimensional square lattice $\mathcal{G} = (V \cup \{s\}, E)$ with vertex set $V = \{1, \dots, L\}^2$, where every edge leaving V is connected to the sink s. The Green's function $G : V \times V \to \mathbb{R}$ is given by the following spectral representation [34, Section 5] (note that the formula in [34] differs by a factor of 4 from ours, since we consider the continuous-time random walk that takes jumps at rate 4, rather than the discrete-time walk)

$$G(x,y) = \frac{4}{(L+1)^2} \sum_{k=1}^{L} \sum_{\ell=1}^{L} \frac{\sin\left(\frac{\pi k x_1}{L+1}\right) \sin\left(\frac{\pi k y_1}{L+1}\right) \sin\left(\frac{\pi \ell x_2}{L+1}\right) \sin\left(\frac{\pi \ell y_2}{L+1}\right)}{4 - 2\cos\left(\frac{\pi k}{L+1}\right) - 2\cos\left(\frac{\pi \ell}{L+1}\right)}. \tag{56}$$

Combining (56) with definition (9) yields an explicit expression for the function ϕ, from which the gain and bias functions follow via (11) and (12). To determine the critical threshold of β for which the never-intervene policy is optimal, it remains to identify the maximizers of ϕ.

Corollary 4.6. *Consider the two-dimensional square lattice $\mathcal{G} = (V \cup \{s\}, E)$, where $V = \{1, \dots, L\}^2$ and every edge leaving V is connected to the sink s. The policy $\pi_\emptyset = (d_\emptyset)^\infty$, where $d_\emptyset(\eta) = 0$ for all $\eta \in \Omega$, is optimal if and only if*

$$\beta \geq \begin{cases} \phi\left(\left(\frac{L+1}{2}, \frac{L+1}{2}\right)\right), & \text{if } L \text{ is odd,} \\ \phi\left(\left(\frac{L}{2}, \frac{L}{2}\right)\right), & \text{if } L \text{ is even,} \end{cases}$$

where, in the even case,

$$\phi\left(\left(\frac{L}{2}, \frac{L}{2}\right)\right) = \phi\left(\left(\frac{L}{2}, \frac{L}{2}+1\right)\right) = \phi\left(\left(\frac{L}{2}+1, \frac{L}{2}\right)\right) = \phi\left(\left(\frac{L}{2}+1, \frac{L}{2}+1\right)\right).$$

Proof. To prove the statement, we make use of the probabilistic interpretation of the function ϕ as the expected first hitting time to the sink of a continuous-time random walk $(X_t)_{t \geq 0}$ with jump rate 1 along each edge, as given by (10). We write $X_t = (X_t^{(1)}, X_t^{(2)})$. Observe that the two coordinates are independent one-dimensional continuous-time simple random walks. Let

$$\tau^{(k)} = \inf\{t \geq 0 : X_t^{(k)} \notin \{1, \dots, L\}\}, \qquad k = 1, 2.$$

Then

$$\tau_s = \min\{\tau^{(1)}, \tau^{(2)}\}.$$

Therefore,

$$\phi(x_1, x_2) = \mathbb{E}_{(x_1,x_2)}[\tau_s] = \mathbb{E}_{(x_1,x_2)}[\min\{\tau^{(1)}, \tau^{(2)}\}] = \int_0^\infty \mathbb{P}_{x_1}(\tau^{(1)} > t)\mathbb{P}_{x_2}(\tau^{(2)} > t)dt.$$

Since $\mathbb{P}_i(\tau^{(k)} > t)$, $k = 1, 2$, $i = 1, \dots, L$, is nondecreasing as i moves toward the center, it follows that ϕ is maximized at the central vertex $((L+1)/2, (L+1)/2)$ if L is odd or at the four central vertices $(L/2, L/2)$, $(L/2, L/2+1)$, $(L/2+1, L/2)$ and $(L/2+1, L/2+1)$ if L is even. The result now follows from Theorem 3.3. □

By a numerical investigation, we found that intervention sets on the two-dimensional lattice need not be topple-free under the corresponding local intervention policies. Therefore, the results under the topple-freeness condition do not apply in general in this case.

4.3 The complete graph

Consider the complete graph $\mathcal{G} = (V \cup \{s\}, E)$, where $V = \{1, \dots, n\}$. Suppose that each vertex $x \in V$ is connected to the sink. In this case, the restriction of the graph Laplacian of $\mathcal{G}$ to $V \times V$ is given by

$$\Delta_{xy} = \begin{cases} n, & \text{if } x = y, \\ -1, & \text{if } x \neq y, \end{cases}$$

that is,

$$\Delta = (n+1)I_n - J_n,$$

where I_n denotes the $n \times n$ identity matrix and J_n the $n \times n$ all-ones matrix. This implies that the Green's function is given by

$$G = \Delta^{-1} = \frac{1}{n+1}(I_n + J_n),$$

or equivalently,

$$G(x,y) = \begin{cases} \dfrac{2}{n+1}, & \text{if } x = y, \\ \dfrac{1}{n+1}, & \text{if } x \neq y. \end{cases}$$

Using definition (9), we now obtain

$$\phi(x) = \sum_{y \in V} G(x,y) = \frac{2}{n+1} + (n-1)\frac{1}{n+1} = 1, \quad \text{for all } x \in V. \tag{57}$$

The gain and bias of the policy $\pi_\emptyset$ and the optimality threshold for never intervening in case of the complete graph are stated in Corollary 4.7.

Corollary 4.7. *The gain of the never-intervene policy $\pi_\emptyset = (d_\emptyset)^\infty$ is given by*

$$g^{\pi_\emptyset}(\eta) = 1, \qquad \eta \in \Omega.$$

Furthermore, there exists a constant $C \in \mathbb{R}$ such that

$$h^{\pi_\emptyset}(\eta) = \sum_{x \in V} (\eta(x) - n + 1) + C, \qquad \eta \in \Omega.$$

The policy $\pi_\emptyset$ is optimal if and only if $\beta \geq 1$.

Proof. The expressions for the gain and the bias follow immediately from Theorem 3.1 and expression (57). The optimality result follows immediately from Theorem 3.3 and expression (57). □

Corollary 4.8. *Let π_V denote an always-intervene policy. Its gain is given by*

$$g^{\pi_V}(\eta) = \beta, \qquad \eta \in \Omega$$

and it is optimal if and only if $\beta \leq 1$.

Proof. Under π_V the mass cannot increase from a nonempty configuration. Hence, within each recurrent class of the Markov chain induced by π_V, the mass must remain constant. Since every toppling dissipates one sand grain on the complete graph, no topplings can occur within a recurrent class. Therefore, V is topple-free under π_V. Moreover, one intervention occurs at every time step. This yields.

$$g^{\pi_V}(\eta) = \beta, \qquad \eta \in \Omega.$$

We now apply Theorem 3.10 with $W = V$. For the lower bound, we obtain

$$\underline{\beta}_V = 0.$$

For the upper bound, we have

$$F_V(u) = b_{us} = 1, \qquad u \in V$$

and hence,

$$\overline{\beta}_V = 1.$$

Thus, π_V is optimal whenever $\beta \leq 1$.

Conversely, it follows from Corollary 4.7 that the never-intervene policy has gain $g^{\pi_\emptyset}(\eta) = 1$ for every $\eta \in \Omega$. Therefore, the always-intervene policy π_V is not optimal if $\beta > 1$. □

Corollaries 4.8 and 4.7 provide a complete characterization of the optimality regimes in terms of β.

Corollary 4.9. *For $\beta \geq 1$, the never-intervene policy is optimal, whereas for $\beta \leq 1$, every always-intervene policy is optimal. In particular, when $\beta = 1$, both types of policies are optimal.*

We next investigate topple-freeness on the complete graph. The following Lemma shows that singleton intervention sets need not be topple-free in the complete graph. We checked numerically by an exhaustive search, that in the complete graph on $\{1, \ldots, n\} \cup \{s\}$ each vertex is topple-free when $1 \leq n \leq 6$. We show below that this is no longer true when $n \geq 7$.

Lemma 4.10. *If $n \geq 7$, no vertex $x \in V$ is topple-free under the single-vertex intervention policy π_x.*

Proof. Let $x \in V$ and let $\eta \in \mathcal{R}_{\pi_x}$. By sequentially adding particles to noncritical vertices in $V \setminus \{x\}$, we can reach a configuration $\overline{\eta}$ that satisfies

$$\overline{\eta}(y) = \begin{cases} k, & \text{if } y = x, \\ n-1, & \text{otherwise} \end{cases}$$

for some $k = 0, 1, \ldots, n-1$. Since η is a recurrent configuration and this sequence of additions occurs with positive probability, it follows that $\overline{\eta}$ is a recurrent configuration as well. We now show that, from every such configuration $\overline{\eta}$, there exists a finite sequence of transitions, occurring with positive probability, during which x topples.

First, suppose that $k \geq 2$. Choose any vertex $z \in V \setminus \{x\}$. Before adding a new grain at z, the policy reduces the height at x to $k-1$. Since all vertices in $V \setminus \{x\}$ have critical height, adding a particle at z causes every vertex in $V \setminus \{x\}$ to topple. As a result, the vertex x receives $n-1$ particles and reaches height $k + n - 2 \geq n$. This implies that x topples.

Now, consider $k \in \{0, 1\}$. In either case, after application of d_x, the height at x equals 0. Choose any vertex $z \in V \setminus \{x\}$ and add a sand grain at z. This causes every vertex in $V \setminus \{x\}$ to topple. Let the resulting configuration be denoted by η'. We have

$$\eta'(y) = \begin{cases} n-1, & \text{if } y = x, \\ n-2, & \text{if } y = z, \\ n-3, & \text{otherwise.} \end{cases}$$

Now, let y_1, y_2 denote distinct vertices in $V \setminus \{x, z\}$. Starting from η', perform four successive transitions with addition sites z, y_1, y_1, y_2 respectively. Note that none of these additions causes a toppling and before each addition one particle is removed from x. Since $n \geq 7$, the resulting configuration η'' is given by

$$\eta''(u) = \begin{cases} n-1, & \text{if } u \in \{z, y_1\}, \\ n-2, & \text{if } u = y_2, \\ n-5, & \text{if } u = x, \\ n-3, & \text{otherwise.} \end{cases}$$

Finally, add another sand grain at z. Before this addition, the policy reduces the height at x to $n-6$. Observe that this addition causes every vertex in $V \setminus \{x\}$ to topple. It follows that x reaches height $2n-7$. Since $n \geq 7$ implies $2n-7 \geq n$, the vertex x topples. Since the constructed sequence of transitions starts in a recurrent configuration and occurs with positive probability, x is not topple-free under π_x. □

4.4 The star graph

Consider the star graph $\mathcal{G} = (V \cup \{s\}, E)$, where $V = \{z, 1, \ldots, n\}$ and z denotes the center. Suppose that only the center is connected to the sink. Ordering the vertices as $z, 1, \ldots, n$, the restriction of the graph Laplacian of $\mathcal{G}$ to $V \times V$ is given by

$$\Delta_{xy} = \begin{cases} n+1, & \text{if } x = y = z, \\ -1, & \text{if } x = z \neq y, \quad \text{or } y = z \neq x, \\ 1, & \text{if } x = y \neq z, \\ 0, & \text{otherwise.} \end{cases}$$

Hence, the Green's function is given by

$$G(x,y) = \Delta^{-1}_{xy} = \begin{cases} 2, & \text{if } x = y \neq z, \\ 1, & \text{otherwise.} \end{cases}$$

Invoking definition (9), we now find

$$\phi(x) = \sum_{y \in V} G(x,y) = \begin{cases} n+1, & \text{if } x = z, \\ n+2, & \text{otherwise.} \end{cases} \tag{58}$$

Using this result, we now provide the gain and bias of the policy $\pi_\emptyset$, together with the threshold for never intervening, for the star graph.

Corollary 4.11. *The gain of the never-intervene policy $\pi_\emptyset = (d_\emptyset)^\infty$ is given by*

$$g^{\pi_\emptyset}(\eta) = n + 2 - \frac{1}{n+1}, \qquad \eta \in \Omega. \tag{59}$$

Furthermore, there exists a constant $C \in \mathbb{R}$ such that

$$h^{\pi_\emptyset}(\eta) = (\eta(z) - n)(n+1) + C, \qquad \eta \in \Omega. \tag{60}$$

The policy $\pi_\emptyset$ is optimal if and only if $\beta \geq n+1$.

Proof. The expressions for the gain and the bias follow immediately from Theorem (3.1) and expression (58). The optimality result follows immediately from Theorem 3.3 and expression (58). □

We next show that the center z is topple-free under the single-vertex intervention policy π_z.

Lemma 4.12. *The center z is topple-free under the policy π_z.*

Proof. After applying the decision rule d_z, the height at z is at most $n-1$. Since every leaf has degree 1, adding a grain at a leaf causes it to topple once and send one grain to z. Adding a grain at z also increases its height by one. Hence, during the addition and subsequent stabilization, the height at z increases by exactly one. This implies that z never topples under the policy π_z. □

Since every leaf has height zero in every stable configuration, the policy π_z coincides with the always-intervene policy π_V. The following lemma provides the gain of this policy.

Lemma 4.13. *The gain of the always-intervene policy $\pi_z = (d_z)^\infty$ is given by*

$$g^{\pi_z}(\eta) = \frac{n}{n+1} + \beta, \qquad \eta \in \Omega. \tag{61}$$

Proof. In Lemma 4.12, we established that z is topple-free under the policy π_z. Hence, we can apply Theorem 3.9 with $W = \{z\}$. On the contracted graph $\widetilde{\mathcal{G}}_z$, each of the n leaves is connected to the sink by one edge. Hence, every addition causes exactly one toppling. This yields

$$\widetilde{g}_z^{\pi_\emptyset} = 1.$$

Furthermore,

$$p_z(x) = 1, \qquad x \in V \setminus \{z\}.$$

Applying Theorem 3.9 now yields

$$g^{\pi_z}(\eta) = \frac{n}{n+1} + \beta, \qquad \eta \in \Omega.$$

□

We now establish a necessary and sufficient condition under which the always intervene policy π_z is optimal.

Corollary 4.14. *The always-intervene policy π_z is optimal if and only if*

$$\beta \leq n + 1.$$

Proof. By Lemma 4.12, the intervention set $W = \{z\}$ is topple-free under π_z. Hence, we can apply Theorem 3.10. Since each of the leaves has degree one, the maximum that defines the lower bound is taken over the empty set. This implies

$$\underline{\beta}_z = 0.$$

Moreover, we have for each leaf x,

$$p_z(x) = 1, \qquad \text{and } \phi_z(x) = 1.$$

Therefore, we obtain

$$\overline{\beta}_z = n + 1.$$

Thus, the policy π_z is optimal whenever $\beta \leq n + 1$.

To prove necessity, we compute, using expressions (59) and (61),

$$g^{\pi_z}(\eta) - g^{\pi_\emptyset}(\eta) = \beta - (n + 1).$$

Thus, when $\beta > n + 1$, the gain of the never-intervene policy is strictly smaller than that of π_z, which implies that π_z is not optimal. □

Corollaries 4.11 and 4.14 provide the following complete characterization of the optimal policies on the star graph.

Corollary 4.15. *The always-intervene policy is optimal for $\beta \leq n + 1$, whereas the never-intervene policy is optimal for $\beta \geq n + 1$. At $\beta = n + 1$, both policies are optimal.*

5 Discussion

In this paper, we formulated and analyzed the problem of controlling avalanche sizes through targeted removal of sand grains as an average-cost MDP that balances avalanche size against intervention cost. Within this formulation, we studied local intervention policies. We derived general structural results for such policies and sharpened these results under the condition of topple-freeness. In addition, for policies that satisfy this condition, we presented explicit expressions for the gain and bias and explicit bounds on β that guarantee their optimality. We applied our general

results to several specific graph classes, leading, among other things, to a complete characterization of the optimal policies on the one-dimensional lattice, the complete graph and the star graph for all values of β.

This work opens several interesting avenues for future research. First, the characterization of optimal policies on the one-dimensional lattice naturally raises the question whether similarly structured optimal policies exist for other graph classes. Our results show that the critical value of β above which the never-intervene policy is optimal is determined by the maximum of the function ϕ. On the one-dimensional lattice, as the intervention cost decreases, the optimal intervention sets expand to include increasingly many vertices with the largest values of ϕ. It is therefore natural to investigate whether similarly structured local intervention policies are optimal on other graph classes, such as higher-dimensional lattices. Moreover, it would be interesting to study the dependence of the size and shape of the corresponding intervention sets on the intervention cost.

Second, since several results rely on the property of topple-freeness, it would be interesting to determine whether this condition can be characterized in terms of the topology of the underlying graph or the spanning trees associated with recurrent sandpile configurations. Such graph-theoretic conditions could facilitate the identification of topple-free intervention sets to which our results apply.

Third, a possible direction for future research is to study the control problem under a discounted-cost criterion. Such a criterion assigns greater relative weight to the costs incurred in the near future, where the relative importance of future costs is determined by the discount factor. It therefore allows us to investigate the trade-off between reducing immediate costs and accounting for the subsequent evolution of the system. This would interpolate between minimizing immediate avalanche costs, as studied by de Jongh et al. [10], and minimizing the long-run average cost, as considered in the present work. Depending on the value of the discount factor, the resulting optimal policies may differ substantially from those obtained under the average-cost criterion. The discounted-cost criterion would be particularly relevant when interventions are intended to mitigate imminent avalanches.

Fourth, when an analytical solution is unavailable, approximate dynamic programming or reinforcement learning methods could be used to construct approximately optimal policies. The exact structural results obtained in this work could guide the design of reduced state and policy spaces. One possible approach is to aggregate configurations into a smaller collection of representative states or to represent them through a set of carefully selected features.

Finally, future research could consider alternative intervention mechanisms, such as controlling the addition site or temporarily modifying the toppling threshold. Different cost functions could also be studied. In this work, the cost function captures the trade-off between the immediate intervention cost and the expected avalanche size. Alternative objectives could instead penalize the probability that an avalanche exceeds a prescribed size or assign a nonlinear cost to large-scale avalanches. Such objectives may lead to policies that more strongly prioritize the prevention of catastrophic events.

Acknowledgments

MCdJ thanks AAJ and the Department of Mathematical Sciences at the University of Bath for their kind hospitality while conducting this research. MCdJ is also grateful to the Catharine van Tussenbroek Fund for financial support for her research visit to Bath.

References

[1] P. Bak, C. Tang, and K. Wiesenfeld. “Self-organized criticality: An explanation of the 1/f noise”. In: *Phys. Rev. Lett.* 59 (4 1987), pp. 381–384.

[2] P. Bak, C. Tang, and K. Wiesenfeld. “Self-organized criticality”. In: *Phys. Rev. A* 38 (1 1988), pp. 364–374.

[3] D. Dhar. “Self-organized critical state of sandpile automaton models”. In: *Phys. Rev. Lett.* 64 (14 1990), pp. 1613–1616.

[4] H. J. Jensen. *Self-Organized Criticality: Emergent Complex Behavior in Physical and Biological Systems*. Cambridge Lecture Notes in Physics. Cambridge University Press, 1998.

[5] C.H. Scholz. *The Mechanics of Earthquakes and Faulting*. 3rd ed. Cambridge University Press, 2019.

[6] B. Drossel and F. Schwabl. "Self-organized critical forest-fire model". In: *Phys. Rev. Lett.* 69 (11 1992), pp. 1629–1632.

[7] M. Bartolozzi, D.B. Leinweber, and A.W. Thomas. "Self-organized criticality and stock market dynamics: an empirical study". In: *Phys. A: Stat. Mech. Appl.* 350.2 (2005), pp. 451–465.

[8] P. Bak and K. Sneppen. "Punctuated equilibrium and criticality in a simple model of evolution". In: *Phys. Rev. Lett.* 71 (24 1993), pp. 4083–4086.

[9] M.L Puterman. *Markov decision processes: discrete stochastic dynamic programming*. John Wiley & Sons, 2014.

[10] M.C. de Jongh, R.J. Boucherie, and M.N.M. van Lieshout. *Optimal local interventions in the two-dimensional Abelian sandpile model*. 2026. arXiv: 2603.24459 [math.OC]. URL: https://arxiv.org/abs/2603.24459.

[11] J. Falk, M. Winkler, and W. Kinzel. "On the effect of the drive on self-organized criticality". In: *J. Phys. A: Math. Theor.* 48.40 (2015), p. 405003.

[12] P. Noël, C.D. Brummitt, and R.M. D'Souza. "Controlling Self-Organizing Dynamics on Networks Using Models that Self-Organize". In: *Phys. Rev. Lett.* 111 (7 2013), p. 078701.

[13] B. Parsaeifard and S. Moghimi-Araghi. "Controlling cost in sandpile models through local adjustment of drive". In: *Phys. A: Stat. Mech. Appl.* 534 (2019), p. 122185. ISSN: 0378-4371.

[14] D.O. Cajueiro and R.F.S. Andrade. "Controlling self-organized criticality in sandpile models". In: *Phys. Rev. E* 81 (1 2010), p. 015102.

[15] D.O. Cajueiro and R.F.S. Andrade. "Controlling self-organized criticality in complex networks". In: *Eur. Phys. J. B* 77.2 (2010), pp. 291–296.

[16] J. Qi and S. Pfenninger. "Controlling the self-organizing dynamics in a sandpile model on complex networks by failure tolerance". In: *Europhys. Lett.* 111.3 (2015), p. 38006.

[17] Y. Hou et al. "Overload cascading failure on complex networks with heterogeneous load redistribution". In: *Phys. A: Stat. Mech. Appl.* 481 (2017), pp. 160–166. ISSN: 0378-4371.

[18] M. Turalska and A. Swami. "Greedy control of cascading failures in interdependent networks". In: *Sci. Rep.* 11.3276 (2021).

[19] A. Scala et al. "Mitigating cascades in sandpile models: an immunization strategy for systemic risk?" In: *Eur. Phys. J. Special Topics* 225.10 (2016), pp. 2017–2023.

[20] M.C. de Jongh, R.J. Boucherie, and M.N.M. van Lieshout. "Controlled stochastic growth of structures in the zero-temperature Ising model". In: *Physica A: Statistical Mechanics and its Applications* (2026). In press. DOI: 10.2139/ssrn.6383335. URL: https://ssrn.com/abstract=6383335.

[21] M.C. de Jongh, S. Spitoni, and E.N.M. Cirillo. "Optimal strategies for the growth of dual-seeded lattice structures". In: *Physica D: Nonlinear Phenomena* (2026), p. 135470. ISSN: 0167-2789. DOI: https://doi.org/10.1016/j.physd.2026.135470. URL: https://www.sciencedirect.com/science/article/pii/S0167278926003660.

[22] S. Baldassarri and M.C. de Jongh. *Optimal strategies for controlled growth in metastable Kawasaki dynamics*. 2026. arXiv: 2603.18903 [math.OC]. URL: https://arxiv.org/abs/2603.18903.

[23] R.N. Haksar and M. Schwager. "Distributed Deep Reinforcement Learning for Fighting Forest Fires with a Network of Aerial Robots". In: *Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems*. 2018, pp. 1067–1074.

[24] R.N. Haksar et al. "Controlling Heterogeneous Stochastic Growth Processes on Lattices with Limited Resources". In: *Proceedings of the 58th IEEE Conference on Decision and Control*. 2019, pp. 1315–1322.

[25] A. Nasir. "Three-Layer Model for the Control of Epidemic Infection over Multiple Social Networks". In: *SN Appl. Sci.* 5.152 (2023).

[26] J. Song, W. Yang, and C. Zhao. "Decision-Dependent Distributionally Robust Markov Decision Process Method in Dynamic Epidemic Control". In: *IISE Trans.* 56.4 (2024), pp. 458–470.

[27] X. Tang, M.A. Bevan, and M.A. Grover. "Markov Decision Process Based Time-Varying Optimal Control for Colloidal Self-Assembly". In: *IFAC-PapersOnLine*. Vol. 49. 7. 2016, pp. 430–435.

[28] X. Tang, M.A. Bevan, and M.A. Grover. "The Construction and Application of Markov State Models for Colloidal Self-Assembly Process Control". In: *Mol. Syst. Des. Eng.* 2.1 (2017), pp. 78–88.

[29] M.N.M. van Lieshout. "Optimal decision rules for marked point process models". In: *Stoch. Environ. Res. Risk Assess.* 38 (Aug. 2024), pp. 1–11. DOI: 10.1007/s00477-024-02769-1.

[30] D.O. Cajueiro and R.F.S. Andrade. "Dynamical programming approach for controlling the directed Abelian Dhar-Ramaswamy model". In: *Phys. Rev. E* 82 (3 2010), p. 031108.

[31] D. Dhar and R. Ramaswamy. "Exactly solved model of self-organized critical phenomena". In: *Phys. Rev. Lett.* 63 (16 1989), pp. 1659–1662. DOI: 10.1103/PhysRevLett.63.1659. URL: https://link.aps.org/doi/10.1103/PhysRevLett.63.1659.

[32] R. Durrett. *Essentials of stochastic processes*. Third. Springer Texts in Statistics. Springer, Cham, 2016, pp. ix+275. ISBN: 978-3-319-45613-3; 978-3-319-45614-0. DOI: 10.1007/978-3-319-45614-0. URL: https://doi.org/10.1007/978-3-319-45614-0.

[33] D. Dhar. "Theoretical studies of self-organized criticality". In: *Phys. A* 369.1 (2006), pp. 29–70. ISSN: 0378-4371. DOI: 10.1016/j.physa.2006.04.004. URL: https://doi-org.ezproxy1.bath.ac.uk/10.1016/j.physa.2006.04.004.

[34] F. Chung and S.T. Yau. "Discrete Green's Functions". In: *J. Comb. Theory Ser. A.* 91.1 (2000), pp. 191–214. ISSN: 0097-3165. DOI: https://doi.org/10.1006/jcta.2000.3094. URL: https://www.sciencedirect.com/science/article/pii/S0097316500930942.

[35] F. Redig. "Mathematical aspects of the abelian sandpile model". In: *Mathematical statistical physics*. Elsevier B. V., Amsterdam, 2006, pp. 657–729. ISBN: 978-0-444-52813-1; 0-444-52813-X. DOI: 10.1016/S0924-8099(06)80051-X. URL: https://doi.org/10.1016/S0924-8099(06)80051-X.

[36] C.M. Grinstead and J.L. Snell. *Introduction to Probability*. 2nd. Providence, Rhode Island: American Mathematical Society, 1997. ISBN: 978-0821807491. URL: https://bookstore.ams.org/iprob/.